

Decoding the Genetics of Inherited Retinal Diseases in Ireland



**Trinity
College
Dublin**

The University of Dublin

A thesis submitted to the University of Dublin for
the degree of Doctor of Philosophy

2019

Adrian Dockery

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work, except where otherwise stated.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed

Adrian Matthew Dockery B.Sc., M.Sc.

September 2018

Publication History

Several articles that have been published or are currently under review have been incorporated into the body of this thesis, as per the thesis submission guidelines. Articles have been segmented into their most relevant thesis chapters for greater readability. Articles in their entirety have been included in the Appendix.

Appendix 1.

Dockery A, Stephenson K, Keegan D, Wynne N, Silvestri G, Humphries P, et al. Target 5000: Target Capture Sequencing for Inherited Retinal Degenerations. Genes. 2017 Nov 3;8(11):304.

* Elements of this publication appear in Chapters 1 and 4.1

Appendix 2.

Farrar GJ, Carrigan M, Dockery A, Millington-Ward S, Palfi A, Chadderton N, et al. Toward an elucidation of the molecular genetics of inherited retinal degenerations. Hum Mol Genet. 2017 Aug 1;26(R1):R2–11.

* Elements of this publication appear in Chapters 1, 4.1 and 5.

Appendix 3.

Stephenson K**, Dockery A**, Wynne N, Carrigan M, Farrar GJ, Kenna P, Keegan D. Multimodal Imaging in a Pedigree of X-Linked Retinoschisis with a Novel RS1 Variant. BMC Medical Genetics (2018) 19:195

* Elements of this publication appear in Chapter 4.2.

** These authors contributed equally.

Sources of Funding

Vision research at the Ocular Genetics Unit at TCD is supported by awards from Fighting Blindness Ireland (FB16FAR), the Health Research Board of Ireland (HRA_POR-2013-376; HRA_POR-2015-1140), the Medical Research Charities Group (MRCG-2012-4; MRCG-2013-8; MRCG-2016-14), the Irish Research Council (IRC), Science Foundation Ireland (SFI 11/PI/1080) and the European Research Council (ERC-2012-AdG 322656-Oculus).

Collaborative Efforts

Clinical evaluations and patient sampling were undertaken primarily by the ophthalmological teams at two clinical sites: Royal Victoria Eye and Ear Hospital, Dublin, Ireland (Dr. Paul Kenna and Dr. Niamh Wynne); the Mater Misericordiae University Hospital and Mater Private Hospital (Dr. David Keegan and Dr. Kirk Stephenson). Bioinformatic scripts were designed by, or with the guidance of, Dr. Matthew Carrigan. Several undergraduate students assisted in the optimisation of reactions to verify the presence and segregation of candidate variants found during the Target 5000 sequencing effort (Marlena Mucha, Avril Watson and Claire Heery). Ciara Shortall assisted in the sequencing and interpretation of variants found as part of the *RPGR* ORF15 experiments. Cloning, transfection and purification of AAV viral preparations for the AAV sequencing studies was driven by Dr. Mary Reilly, Dr. Naomi Chadderton and Dr. Sophia Millington-Ward.

Summary

The primary aim of this PhD was to prepare, process and analyse samples as part of the Target 5000 study. The objective of Target 5000 is to provide a genetic diagnosis for the approximate 5000 people in Ireland who are affected by inherited retinal diseases (IRDs). IRDs represent a clinical and genetic heterogeneous group of disorders. To achieve meaningful genotype-phenotype analyses for rare genetic conditions, it is necessary to collect data from sizable populations.

Genetic information can provide accurate diagnoses, and at times, can be informative regarding prognoses and can aid in directing optimal therapeutic intervention. As part of this PhD thesis, next generation sequencing technologies have been utilised to provide a picture of the genetic architecture of IRDs in the Irish population.

The first results chapter of the thesis, using data from nearly 1,000 IRD patients sequenced for up to 254 IRD genes, provides an overview of which genes and which mutations are prevalent in the Irish IRD population as well as identifying over 30 previously unreported novel IRD mutations.

The involvement of novel mutation in the *RS1* gene in a specific form of ocular disease in a large Irish pedigree is then explored in the second results chapter, along with the utility of various in silico tools to interrogate the potential pathogenicity of this novel variant.

During the course of these PhD studies, a wide array of case histories has resulted in novel and or important genetic and clinical findings from the perspective of IRDs. The third results chapter provides information on a collection of IRD case studies to establish the substantial novelty and value of information emanating from a study such as Target 5000.

The final research chapter in the PhD thesis links the genetic pathogenesis of IRDs to the field of gene therapy. One of the many core issues relating to gene therapy is the quality of the virus and potency of the therapeutic. By applying NGS-based sequencing technologies in a similar manner to that utilised for the main body of the thesis studies, the integrity of AAV preparations has been investigated.

Acknowledgements

Prof. G. Jane Farrar

I could not ask for a more ideal supervisor. Jane has consistently supported me academically and emotionally throughout my PhD. She has always seen potential in me that I often failed to see myself. For that, I will be eternally grateful.

Dr. Matthew Carrigan

Matt is not only one of the most intelligent people I have ever encountered but he is also one of the kindest. Matt has always gone above and beyond the call of duty in terms of mentorship. Even though he has moved on from the realm of science into machine learning, he continues to make time for me and the project that we both have invested so heavily in.

The Farrar Lab Postdocs

Thanks for all the sound advice over the years. I know I am very lucky to have so much knowledge and expertise all under one roof. It was Dr. Sophia Millington-Ward's lecture during my M.Sc. in Molecular Medicine that peaked my interest in gene therapies and the eye, thanks for saving me from taking up a boring desk job.

My Fellow PhD Candidates

Thanks Danny and Ciara for keeping me sane over the years. I am very grateful for all our brain storming sessions and for all the other sessions of the alcoholic variety.

Clinical Teams

Paul, Niamh, Kirk and David, thank you for your hard work and dedication, but most of all for your patience as I slowly came to grips with all of the ophthalmological terminology. This project wouldn't exist without you.

My Lovely Editors

Thank you, Natalie, Jeff and Bennett, for trudging through my horrendous writing and polishing this thesis into the some-what readable babble it is now.

SS Students

Marlena, Avril and Claire, I loved every minute of teaching you guys and I think you all had fun too. I am very proud of how well you all progressed and of course very grateful for all the hard work you guys put into optimising reactions for the project.

Friends

Mei-ling and Bee, I'm not in a coma, this thesis just took me ages to write. Pints soon, I promise.

Niamh

Nine years of studentship and student-related poverty was a lot to ask you to put up with. I'll be a grown-up soon I promise. Thanks for being so supportive and understanding. I look forward to all the weekends off hanging out with you and the many adventures in our future.

Orla

You were the first scientist in our family and I always aspired to be more like you. You always took such a genuine interest in my work and never stopped encouraging me even after you got sick. You also never stopped hassling me to start writing this thesis. I'm sorry you weren't there to see it in the end, I really hope this would have made you proud. I love you and miss you every day big sis.

Contents

1.1	An Introduction to Target 5000	3
1.2	An Introduction to the Anatomy and Cells of the Retina.....	5
1.3	An Introduction to Inherited Retinal Degenerations	10
1.4	An Introduction to Next Generation Sequencing	16
1.4.1	First Generation Sequencing	16
1.4.2	Next Generation Sequencing (NGS).....	19
1.5	Next-Generation Sequencing of IRD Populations.....	38
2	Materials and Methods	45
2.1	Ethical Approval	45
2.2	Patient Assessment Sites	45
2.3	Clinical Assessment	45
2.4	Sample Collection from Patients	46
2.5	DNA Extraction from Blood.....	47
2.6	DNA Extraction from Saliva	47
2.7	RNA Extraction from Blood.....	47
2.8	DNA Shearing for Next-Generation Sequencing	48
2.9	DNA Purification Using Magnetic Beads.....	48
2.10	Size-selection of DNA Products Using Magnetic Beads	50
2.11	Homemade DNA Binding Beads	50
2.11.1	Homemade DNA Storage Buffer	51
2.11.2	Homemade DNA-Binding Bead Solution	51
2.11.3	DNA Binding Bead Solution (Preparation)	52
2.12	Quantification of DNA.....	53
2.13	KAPA Library Preparation.....	54
2.14	KAPA HyperPlus Library Preparation	55
2.15	Indexing Adapters	55
2.16	Nimblegen Target Capture Enrichment	57

2.17	Variant Verification by Direct Sequencing	58
2.18	Amplification of <i>RPGR</i> ORF15	61
2.19	Mammalian Cell Triple Transfection	65
2.20	Recombinant AAV Harvest and Purification.....	66
2.20.1	Harvest	66
2.20.2	Caesium Chloride Gradient	67
2.20.3	Purification I	68
2.20.4	Purification II	69
2.20.5	Purification III	69
2.21	Library Preparation for Next Generation Sequencing of AAVs.....	70
3	Variant Interpretation and Annotation	73
3.1	Introduction.....	73
3.2	Data Analysis	73
3.3	Read Alignment	76
3.4	Duplicate Read Filtering	77
3.5	Local Realignment	77
3.6	Sequence Data Quality Control.....	78
3.7	Sequence Tidying	85
3.8	Variant Annotation	85
3.9	Pathogenicity Prediction Models for Nonsynonymous Variants	87
3.10	Pathogenicity Prediction Models for Splice Variants.....	89
3.11	Detection of Structural Variants	90
3.12	Selecting Variants of Interest	91
3.13	Guidelines for Interpretation of Sequence Variants.....	93
3.14	Objectives of Target 5000.....	101
4	Results Chapters	102
4.1	Summary of Target 5000 Findings	102
4.1.1	Introduction	102

4.1.2	Methods	103
4.1.3	Results.....	104
4.1.4	Discussion.....	132
4.2	Retinoschisis.....	141
4.2.1	Introduction	141
4.2.2	Methods	143
4.2.3	Results.....	145
4.2.4	Discussion.....	158
4.2.5	Concluding Remarks	164
4.3	Case Studies	165
4.3.1	Chapter Introduction.....	165
4.3.2	Chapter Methods.....	166
4.3.3	Case 1: Routine Rhodopsin	166
4.3.4	Case 2: FLVCR1, A Wolf in NR2E3's Clothing.....	169
4.3.5	Case 3: Using Genotyping to Refine Stargardt Prognosis.	176
4.3.6	Case 4: An X-Linked Conundrum	180
4.3.7	Case 5: Circumnavigating the Pitfalls of NGS	187
4.3.8	Case 6: From Bench to Bedside, An RPE65 Story.	192
4.3.9	Concluding Remarks	198
4.4	AAV Analysis by NGS.....	200
4.4.1	Introduction	200
4.4.2	Methods	206
4.4.3	Results.....	207
4.4.4	Discussion.....	224
5	Discussion.....	230
5.1	Deciphering the Genetic Pathogenesis of Unresolved IRDS	230
5.2	A Diagnostic Imperative Driven by Developments in Gene Therapy	236
5.3	Challenges for Clinical Incorporation of NGS Technologies	238

6	References.....	245
7	Appendices.....	283

Abbreviations

AAV	Adeno-associated virus
ACMG	American college of medical genetics
AF	Autofluorescence
AMD	Age-related macular degeneration
AO	Adaptive optics
BBS	Bardet-biedl syndrome
BWA	Burrows-Wheeler alignment
CD	Cone dystrophy
CNV	Copy number variant
CRD	Cone-rod dystrophy
CSNB	Congenital stationary night blindness
ddNTPs	Dideoxynucleotide triphosphates
dNTPs	Deoxynucleotide triphosphates
EOSRD	Early onset severe retinal degeneration
ERG	Electroretinogram
ETDRS	Early treatment diabetic retinopathy study
FEVR	Familial exudative vitreoretinopathy
FFM	Fundus flavimaculatus
GA	Glaucoma
iPS	Induced pluripotent stem-cell
IRD	Inherited retinal degeneration
ISCEV	International society for clinical electrophysiology of vision
ITR	Inverted terminal repeat
LCA	Leber's congenital amaurosis
MD	Macular dystrophy
MLPA	Multiplex ligation-dependent probe amplification
MNP	Multiple nucleotide polymorphism
NAbs	Neutralising antibodies
NGS	Next generation sequencing
nsSNV	Nonsynonymous single nucleotide variants
OA	Optic atrophy
OCT	Optical coherence tomography
OMIM	Online Mendelian Inheritance in Man (database)
PCR	Polymerase chain reaction
RCD	Rod-cone dystrophy
RD	Rod dystrophy
RP	Retinitis pigmentosa
SBL	Sequencing by ligation
SBS	Sequencing by synthesis
SGS	Second-generation sequencing
SMRT	Single molecule real time (sequencing)
ssSNV	Splice site single nucleotide variants
STGD1	Stargardt disease

SV	Structural Variant
USH	Usher syndrome
VMD	Vitelliform macular dystrophy
WES	Whole exome sequencing
WGS	Whole genome sequencing
ZMW	Zero mode waveguide

Introduction

1.1 An Introduction to Target 5000

Inherited retinal degenerations (IRDs) represent the most frequent cause of vision loss in people of working age. As a result, these conditions have a highly significant impact on quality of life, while also posing health-related costs and loss of income. IRDs are an extremely heterogeneous set of conditions associated with the loss of retinal function, and as a group, represent one of the most genetically diverse hereditary conditions.

Over 260 genes to date have been implicated in the syndromic and non-syndromic IRDs (1), with a wide range of clinical presentations and rates of progression. As this is a diverse set of conditions with frequently overlapping clinical presentations, it is typically divided into large sub-categories, primarily by specific regions or cell types affected, such as rod photoreceptors, cone photoreceptors or, for example, peripheral versus macular regions of the retina. Retinitis Pigmentosa [OMIM: #268000] (RP) is the most common form of IRD, is extremely genetically heterogeneous and affects as many as 1 in 3000 individuals (2,3). The disease is typically characterized by progressive loss of rod photoreceptor cells, followed by the gradual death of cone photoreceptors and generally involves characteristic features such as pigmentary deposits in the peripheral retina and attenuation of retinal vessels. In contrast, some other forms of IRD can be extremely rare and have a single gene aetiology; gyrate atrophy, for example, is estimated to affect roughly one in a million people (4).

IRDs are currently thought to affect approximately 2.5 million people globally. The vast majority of these individuals have received a diagnosis based on clinical phenotype alone,

rather than a genetic diagnosis, if they have been formally diagnosed at all. Clinical trials are in progress for a number of IRDs, however most such trials require patients to have a known causative mutation to participate.

In this introductory chapter of the thesis, the underlying principles of Target 5000 are introduced from both a retinal cell biology and a next generation sequencing (NGS) perspective. Target 5000 is an ongoing NGS-based study, which aims to genetically characterise a large national cohort of IRD patients. The most common method chosen for IRD genetic screening is targeted NGS. Although whole-exome sequencing offers the potential to locate disease-causing mutations in novel genes, in practice diagnosis rates in whole-exome and targeted-sequencing studies are similar (5), suggesting that the coding regions responsible for the majority of IRDs have been located. Although whole genome analysis has the potential to discover non-coding disease-causing mutations, the difficulty involved with data interpretation and cost associated with the study increase dramatically.

1.2 An Introduction to the Anatomy and Cells of the Retina

The cell biology of the eye is complex and highly organised (Figure 1.1). The front of the eye receives and focuses light towards the back of the eye. The retina is the name given to the layer of cells responsible for reception and transduction of photonic signals. The retina requires several layers of characteristically distinct neurons in order to achieve this feat (Figure 1.2). Any aberration of these cell types could potentially prevent the entire signalling cascade from functioning normally.

Photoreceptors are primarily broken down into two main types, rod photoreceptor cells and cone photoreceptor cells. Rods are far more abundant than cones, with approximately 91 million rod cells in the human retina compared to roughly 4.5 million cone cells (6). Rod cells are abundant almost entirely throughout the retina, whereas cone cells are almost exclusively found in the macula and more specifically in the fovea centralis, this area of the retina is responsible for the highest resolution of sight and central vision (Figure 1.3).

Rods function as low light receptors which are responsible for night vision capabilities (scotopic vision) but have little or no capacity to detect colour, which is believed to be why we see in greyscale in the dark. The cones function to perceive bright light (photopic vision) and colour, although it has also been shown that rods have a very limited capacity to recognise hue in scotopic conditions (7).

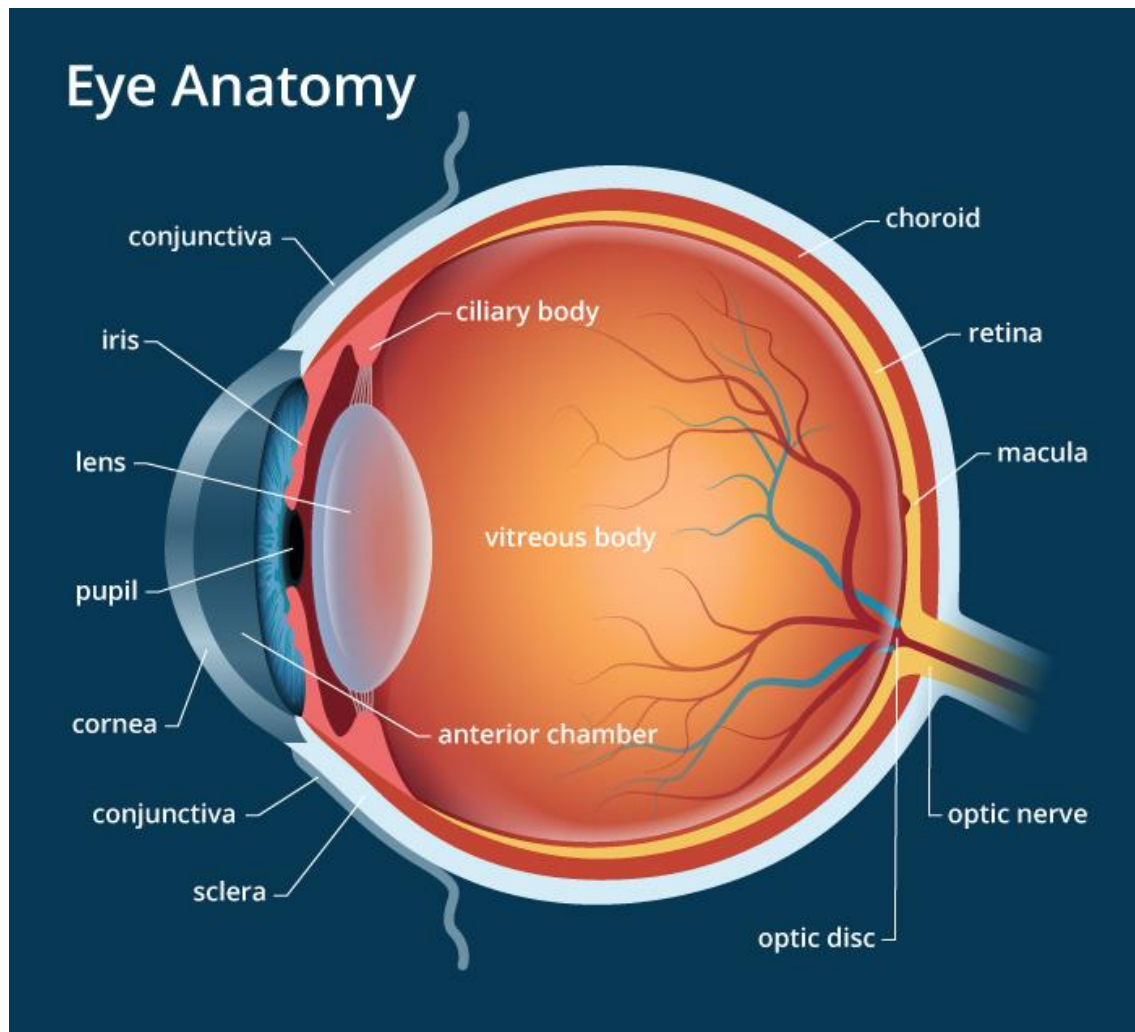


Figure 1.1. *Diagram of the human eye. Arrows indicate the various structures and compartments required for fully function vision. Figure taken from (8).*

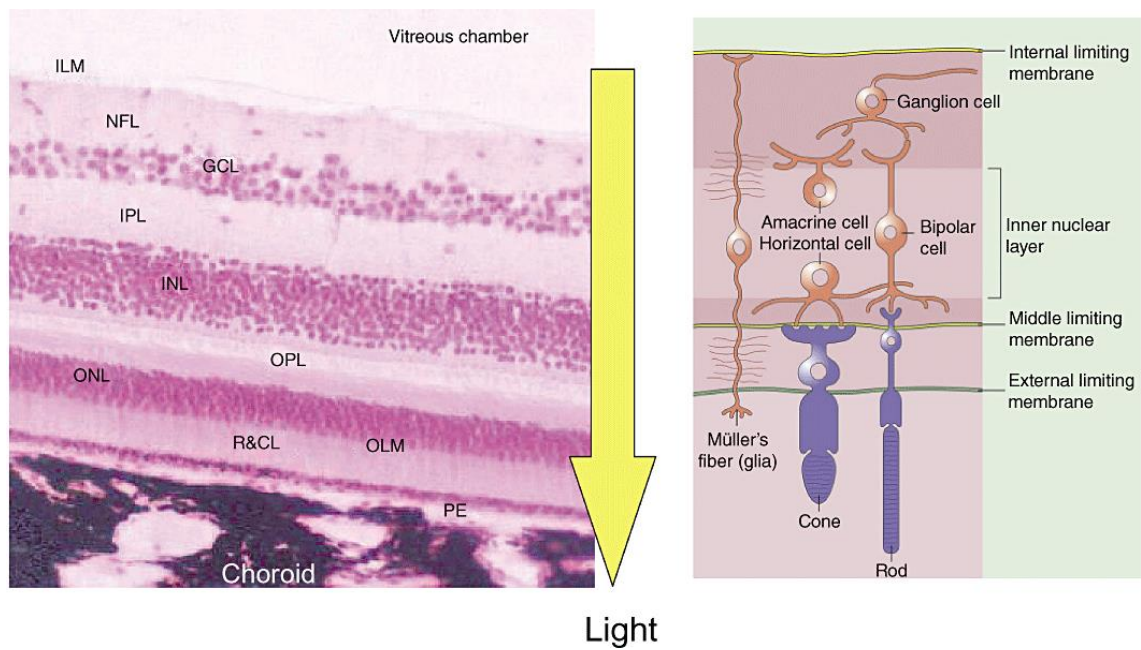


Figure 1.2. Cell composition of the retina. In order of light reception: ILM, inner-limiting membrane; NFL, nerve fibre layer; GCL, ganglion cell layer; IPL, inner-plexiform layer; INL, inner-nuclear layer; OPL, outer-plexiform layer; ONL, outer-nuclear layer; OLM, outer-limiting (external) membrane; R&CL, rod and cone (photoreceptor) layer; PE, (retinal) pigment epithelium. Illustration taken from Willoughby et al. (9).

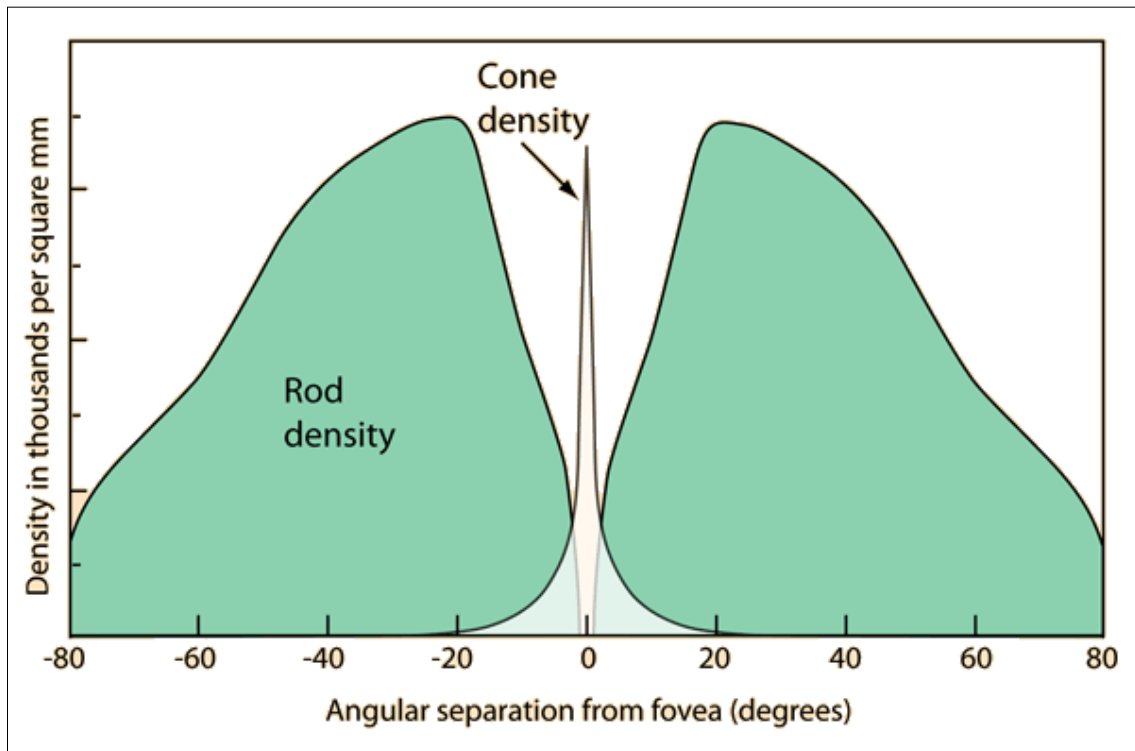


Figure 1.3. *Distribution of rod and cone photoreceptor cells in the retina. The graph illustrates the abundant presence of rods throughout the retina with the exception of the fovea, whilst cones are almost exclusively centred on the fovea centralis. Diagram taken from Desai and Alibhai (10).*

Although this is an over-simplification of the functionality of photo-transduction in the retina, this basic understanding of primary differences between photoreceptor cells can provide an incredibly useful insight into the initial steps taken to clinically assess retinal conditions. A patient's own account of changes in visual fields, visual acuity and light perception are usually accurate indicators of affected regions of the retina.

For example, if a patient has reported poor or no functional night vision (nyctalopia) and peripheral visual field loss, this is more fitting with the functionality and distribution of rod photoreceptor cells. However, if a patient reports light sensitivity (photophobia), blurring of central vision or loss of colour distinction, we know that these are more logically associated with the spatial distribution and established purpose of cone photoreceptor cells (11). These indications can then be followed up with an investigation of ophthalmoscopic characteristics such as peripheral bony spicules and choroid thinning in rod degenerations such as Retinitis Pigmentosa (5), or in the case of a cone degeneration, macular atrophy (10).

1.3 An Introduction to Inherited Retinal Degenerations

Significant progress in the last 30 years has been made in elucidating the pathogenesis of inherited retinal diseases (IRDs). Currently, over 260 genes have been implicated as causative of IRDs (Figure 1.4). Variation in any of the approximate 260 genes may result in varying degrees of visual impairment. The umbrella term IRD accounts for many well-known retinal pathologies including Retinitis Pigmentosa (RP), Lebers congenital amaurosis (LCA), Usher syndrome, congenital stationary night blindness (CSNB), Vitelliform macular dystrophy (VMD) and Stargardt disease amongst many others. The majority of these inherited conditions cause a progressive degeneration of photoreceptor cells initially presenting as a visual impairment, but eventually resulting in complete blindness. The severity and age of onset of these conditions is not only dependent on the specific condition, or even gene but can somewhat depend on the specific variant causative of the condition (12).

RP is an exceptionally diverse condition, in terms of both heritability and genetic causation that presents with a very recognisable pathology. It is estimated that RP constitutes up to 90% of all inherited retinal degenerations cases globally if both syndromic and non-syndromic forms of RP are considered (13). The most frequent forms of inheritance, autosomal dominant, autosomal recessive and X-linked, have been associated with over 80 genes to date (5). However, RP has also been shown to result from digenic inheritance (14), mitochondrial variants (15) and even rarely from cases of uniparental disomy (16).

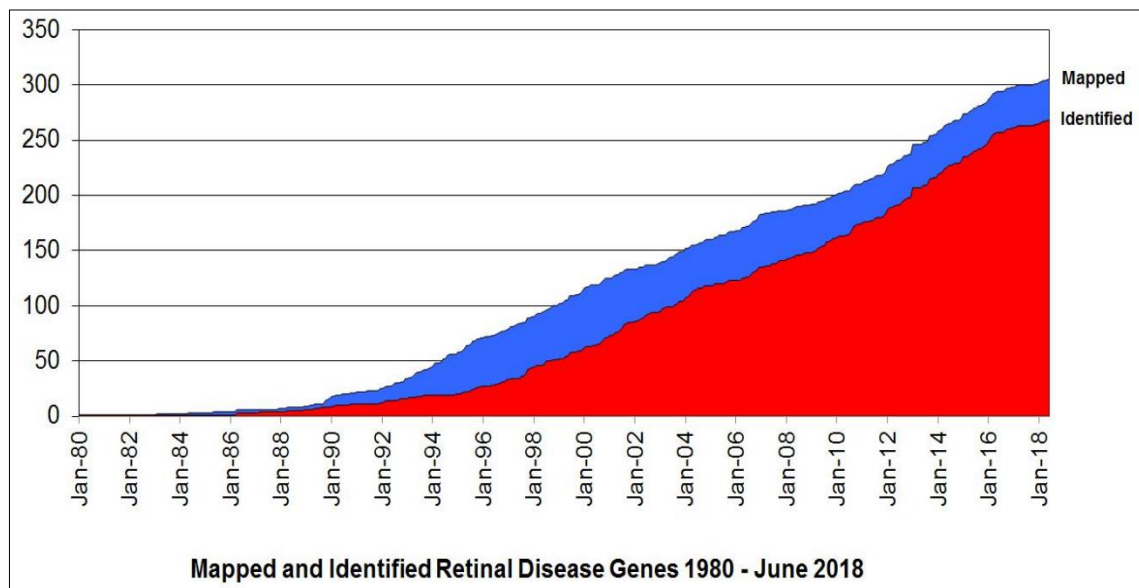


Figure 1.4. *The mapping and identification of genes associated with inherited retinal degenerations (IRDs). The illustration shows that the discovery of IRD-associated genes continues to grow with each subsequent year. Image taken from sph.uth.edu/RetNet (1).*

This level of genetic heterogeneity combined with phenotypic overlap between genotypes makes implicating a causative gene during clinical diagnosis incredibly challenging. Although particularly applicable to RP diagnoses, this is not only true for RP, but several other conditions also such as LCA (17), Bardet-Biedl (18) Syndrome and Usher Syndrome (19) (see also Chapter 4.1 for further details of conditions).

To further complicate matters, many IRD genes are associated with multiple phenotypes and levels of severity. For example, *ABCA4* variants are associated with two very different phenotypes, RP and Stargardt Disease. Whereas RP is the most common form of rod photoreceptor degeneration, Stargardt Disease is the most common form of juvenile macular degeneration (20). Symptomatically and biologically, these two conditions are readily distinguished from one another, particularly after an ophthalmic exam. However, genetically, the overlap means that a definitive diagnosis would not be possible without the appropriate clinical information. *ABCA4* is one of the largest genes associated with IRDs, and although studies have already tried to identify why a variant may cause one condition and not the other (21), it is likely that this issue will not be fully elucidated until a more comprehensive strategy, such as whole genome sequencing, is employed.

ABCA4 is far from the only gene to engage in such overlapping condition associations. Many genes are associated with multiple retinal pathologies. Of the over 80 genes associated with RP (5), 9 are linked to familial exudative vitreoretinopathy (FEVR), a condition which causes the blood flow to be reduced to the edges of the retina leading to progressive loss of vision. 15 genes are implicated in CSNB, 20 genes are seen in macular dystrophy (MD) cases and 33 genes are involved in cone/cone-rod dystrophy (CD/CRD)

(Figure 1.5) (22). Cone-rod dystrophy describes when both cones and subsequently rod photoreceptors are deteriorating.

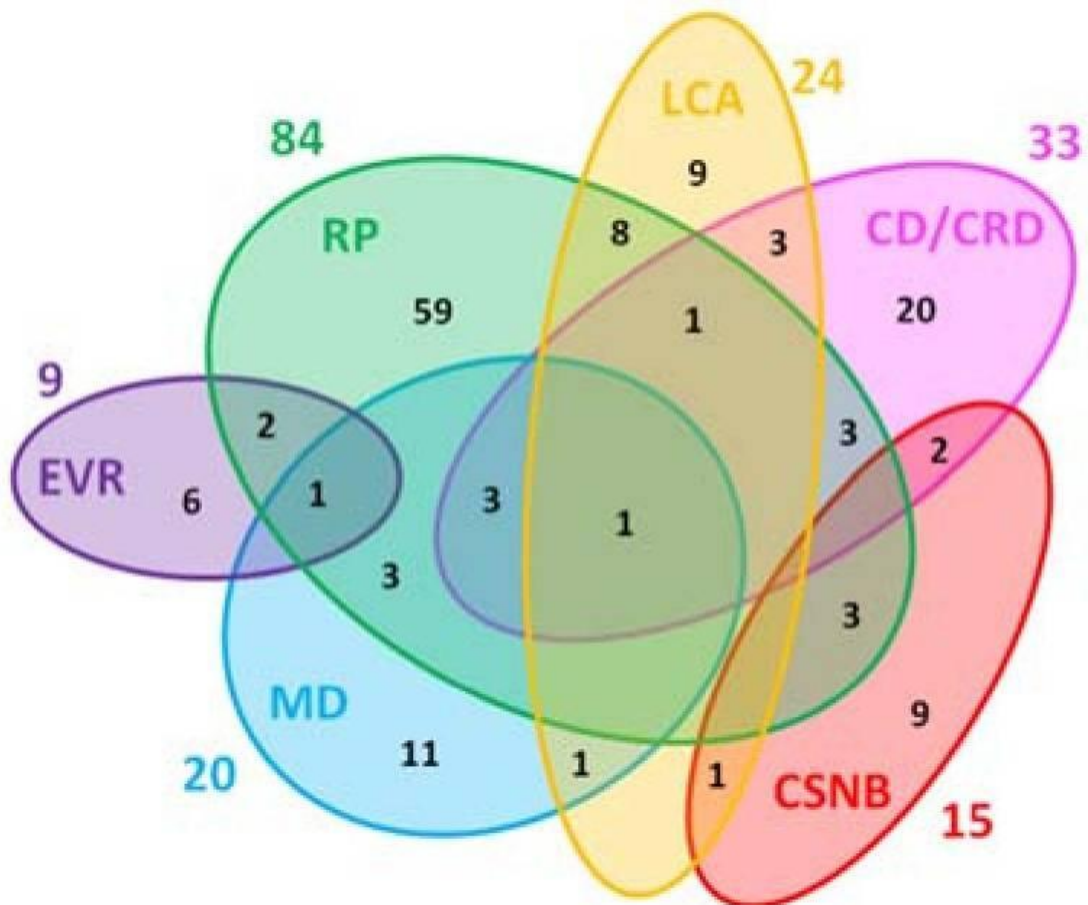


Figure 1.5. Venn diagram illustrating the genetic heterogeneity of common inherited retinal diseases. Numbers located outside the colour matched sections correspond to the number of genes specific to that condition. CD/CRD: cone dystrophy/cone-rod dystrophy; CSNB: congenital stationary night blindness; EVR: exudative vitreoretinopathy; LCA: Leber congenital amaurosis; MD: macular dystrophy; RP: retinitis pigmentosa. Image taken from Cremers et al. (22).

Additionally, some genes associated with IRDs are known to be variable in their manifestation of characteristic symptoms. One such example of this is the *BBS1* gene. Pathogenic variants in this gene are linked to Bardet-Biedl Syndrome and also non-syndromic RP. The most frequently observed pathogenic *BBS1* variant in the Irish population is NM_024649.4:c.1169T>G, p.Met390Arg (23). This variant has been seen to segregate in both the syndromic and non-syndromic conditions associated with this gene. This is similar to the syndromic and non-syndromic manifestations of *USH2A* variants that can either result in Usher Syndrome Type 2 or non-syndromic RP (24).

Historically, genes were identified by linkage analysis in sufficiently large pedigrees and followed by positional cloning. This led to the discovery of some of the first IRD associated pathogenic variants in the *OAT* (25), *RHO* (26) and *CHM* (27) genes. However, given the current knowledge of genetic and phenotypic diversity present in the field of IRDs a more multiplexed approach is required.

Although by definition, rare diseases affect less than 1 in every 2,000 people (28), the reality is that when taken together the frequency seems quite common. It is estimated that there are currently between 25-30 million Americans living with a rare disease (29). In the UK and Europe, it is estimated that 1 in 17 will be affected by some rare disease in their lifetime, which equates to 3.5 million in the UK and approximately 30 million across Europe (28). 80% of rare diseases have genetic associations already identified (30).

In the current age of next generation sequencing (NGS) and gene therapies, we are more equipped than ever to not only accurately diagnose rare disease, but to also potentially treat them. Given the previously discussed multiple categories of heterogeneity associated

with IRDs, all potential gene candidates must be analysed in order to find a true potential causative variant responsible for that individual's specific pathology. Although sequencing technologies have advanced dramatically in the last 40 years, very few methods have been made completely redundant. The various technological advances in the field will be discussed in the next section as will the particular niches that each methodological approach still fills today.

1.4 An Introduction to Next Generation Sequencing

1.4.1 First Generation Sequencing

Although sequencing reactions have evolved massively in the last 50 years, the fundamental principles remain the same. The introduction of Chain Termination reactions represented the dawn of the sequencing revolution, which began in 1977 (31). This “Sanger Sequencing” used standard deoxyribonucleotide triphosphates (dNTPs) spiked with smaller fractions of dideoxy and arabinonucleoside versions of standard dNTPs, which would cause the amplification by DNA polymerase to terminate when incorporated. The prematurely terminated strands would then represent the modified base that was incorporated at that point.

These various fragment lengths could then be sized by separation using gel-electrophoresis. This method enabled the innovators of this technology to create the first ever whole genome sequence, bacteriophage ϕ X174, a 5375 base pair (bp) genome. Just 5 years later, the same technology was used to sequence bacteriophage λ consisting of 48,501 bp (32). However, a notable drawback of this technology was that it required four reactions, one per terminating base, per sample that underwent sequencing.

Automation of this technology represented another giant leap forward in the sequencing field. This was first introduced by the addition of unique fluorophores to each terminating nucleotide (33), so that fluorometric measurements could be taken by a computer (Figure 1.6), which later enabled this technology to become fully automated (34). Several years later the ABI Prism 3700 was released and resulted in a thousand-fold decrease in cost

per base sequencing (35). Before it was surpassed, the Sanger sequencing method achieved one last great accolade, it was the primary technology used as part of the Human Genomic Project to sequence the first ever draft of the human genome in 2003. It was also used to sequence the first privately-funded project to sequence an individual (36). Although more elaborate and sophisticated methods followed this technology, it is important to note that Sanger sequencing is still largely incorporated into any NGS-based workflow due to the importance of verification of variants.

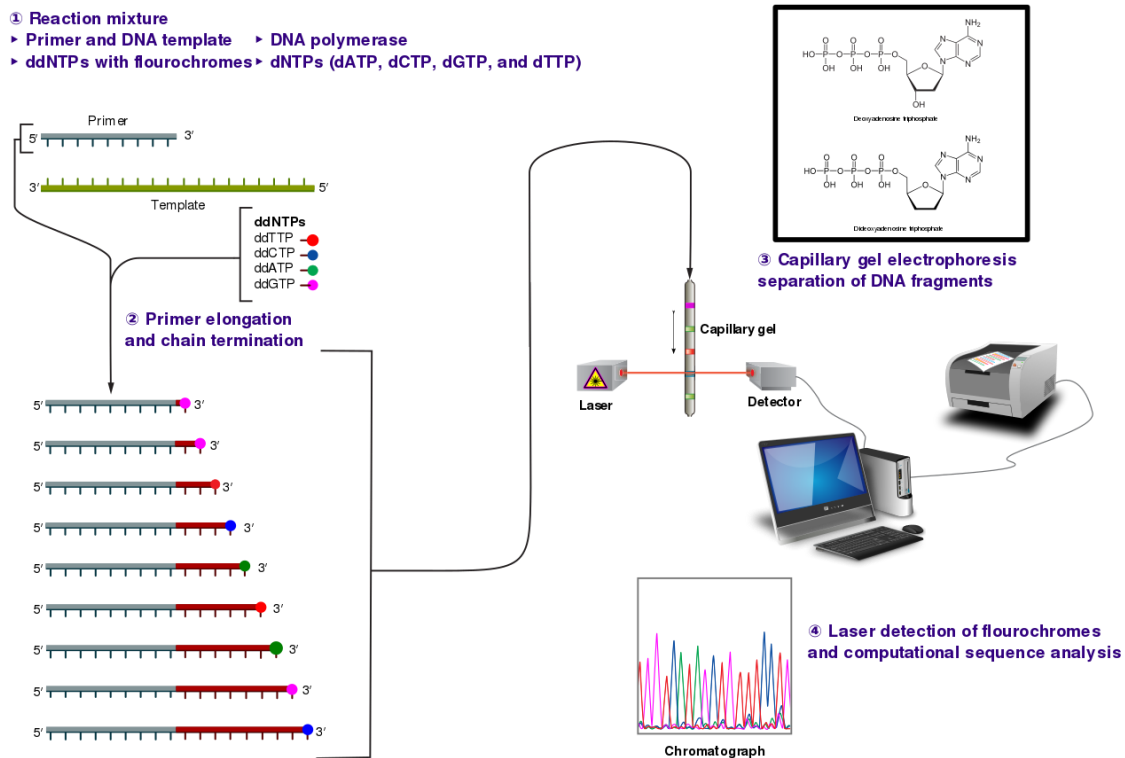


Figure 1.6. The fluorophore-based Sanger sequencing method. Step 1: A sequencing primer binds the template DNA. Step 2: The polymerase elongates the primer-initiated sequence using a pool of dNTPs with a smaller fraction of fluorophore-tagged ddNTPs. Step 3: The fragments are separated by size using capillary gel electrophoresis and imaged using laser excitation. Step 4: The detector records the emissions from the fluorophores post-excitation and this data is compiled into a chromatograph. Image taken from Estevezj (37).

1.4.2 Next Generation Sequencing (NGS)

NGS is the term most commonly used for any sequencing technology that takes advantage of massively paralleled or multiplexed sequencing. The Illumina Genome Analyzer was the first big game changer in this field. It massively reduced the time it took to complete large genome projects by its capability to sequence multiple fragments simultaneously. For comparison, The Human Genome Project took approximately 15 years to sequence one genome *de novo* using almost exclusively single read technology, whereas the Illumina Genome Analyzer completed sequencing and alignment of Dr. James Watson's entire genome in only two months (38).

NGS technology is now evolving at a rapid rate, coinciding with diminishing related costs. The original Human Genome Project cost \$3 billion over a period of 15 years to complete the first human genome sequence (39). It is now forecast that a human genome may cost as little as \$100 in the near future, using the massive sequencing capacity of Illumina's NovaSeq platform (40).

1.4.2.1 Key Terminology of Next Generation Sequencing

Library

An NGS library is the sum of template fragments of DNA/cDNA to be sequenced. A library's source material is changed depending on the application or investigation being carried out. Common sources for library preparation include specific amplified regions (e.g. a gene of interest), RNA transcripts, the exons of specific genes associated with a condition (e.g. Target 5000), all of the known exons or even whole genomes.

Adapters

These are short nucleotide fragments which are designed to be complimentary to those fragments tiled in the sequencing platform. During library preparation, these sequences are ligated to the input DNA so that they can be read during sequencing.

Indexes/ Barcodes

Indices are usually incorporated into adapter sequences and they are essentially known unique identifiers that can be used to multiplex and demultiplex samples as required. The ability to multiplex samples during library preparation is not only more time efficient but also more economical. However, when data is being generated from a sample pool, it is vital to be able to demultiplex the pool in order to identify variants belonging to each sample (Figure 1.7).

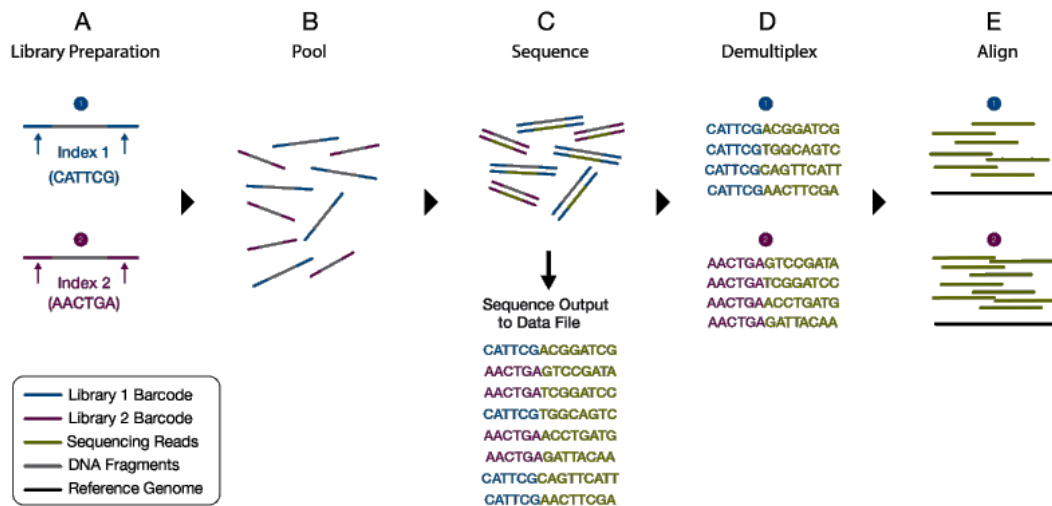


Figure 1.7. Library multiplexing using unique indices. (A) Unique index sequences are ligated to different samples during library preparation. (B) Samples are pooled together into the same library. (C) Sample pool is sequenced during a single instrument run and generates a single file. (D) A demultiplexing algorithm allows each read to be assigned to its correct sample origin. (E) Each set of reads belonging to each sample can then be separately analysed or aligned to the appropriate sequence. Image taken from www.illumina.com (41).

Reads

A read is the sequence produced from a single DNA molecule reaction that results from sequencing. Read lengths and quantities will vary greatly depending on the method used to create a library and the chemistry employed by the sequencing platform.

Single and Paired Reads

Once indexed, a DNA molecule can be read by a sequencer in one (single) or two (paired) directions. For example, if a particular kit and platform are capable of 300 cycles of reactions, this can be used to create reads that are 300 bp long in one direction or instead, it can read 150 bp of the fragment from both ends.

Coverage/ Read Depth

This is the number of reads which align to that position in the reference sequence (e.g. Human Reference Genome Hg38). This figure helps determine how successful sequencing at this position was and can also be used to detect structural variants ranging from single base pair insertions to whole gene deletions.

1.4.2.2 Chemistry of Next Generation Sequencing

By and large, NGS utilises ligation or synthesis methods (sequencing by synthesis) in parallel to produce enormous amounts of data. This in turn was implemented in a couple of different ways. Firstly, the addition of nucleotides could be by a continuous or discontinuous method. If the continuous method was employed, the excitation of the phosphate-labelled nucleotides has to be constantly monitored and sufficiently confined that changes could be detected by pyrosequencing (42). The discontinuous method uses reversible terminators and chemically cleavable fluorescent ddNTPs to image lots of molecules at once and then progress to the next cycle of additions (43).

The discontinuous method could also be employed using single molecule sequencing or ensemble approaches. The ensemble approach involves generating clusters of colonies with identical capture sequences on the applicable surface, usually tiled surface or beads (Figure 1.8). The notable difference between the continuous and discontinuous methods is the resulting read lengths and quantities. Continuous sequencing methods favours fewer reads of longer lengths, generally over 1000 bp, varying with technology employed, whilst the discontinuous method is better used to generate a larger quantity of reads with shorter lengths, generally 500 bp or less.

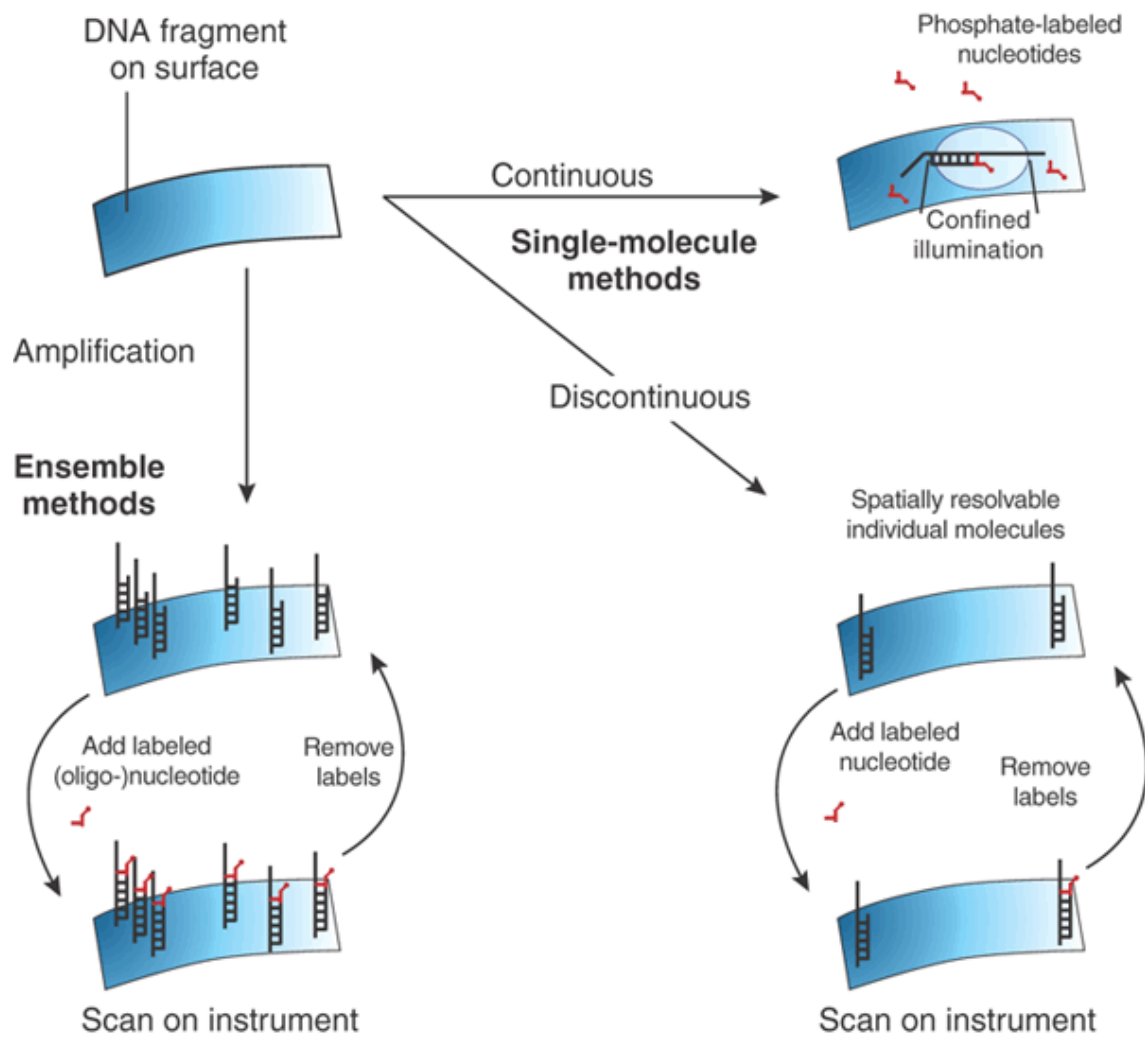


Figure 1.8. The different approaches of sequencing by synthesis. Sequencing can occur by the Continuous or Discontinuous methods. The Discontinuous methods can also be modified to incorporate Ensemble clustering. Illustration taken from Fuller et al. (44).

1.4.2.3 General Workflow of Next Generation Sequencing

Irrelevant of the precise methodology employed, all modern NGS platforms require some degree of highly complex library preparation in order to ensure high quality sequencing results (Figure 1.9). Figure 1.9 demonstrates a typical NGS workflow that takes advantage of target enrichment. For certain applications, such as mitochondrial sequencing, it may be deemed more useful to enrich for the desired sample prior to library preparation by utilising such methods as differential centrifugation, nuclear DNA depletion or selective mitochondrial amplification. However, for most target capture panels, it is more feasible and economical to enrich for the desired regions during library preparation.

Template material can be any source of DNA or RNA that one wishes to investigate. For the Target 5000 study, source material was generally patient blood or saliva samples, from which DNA could be extracted. For target capture protocols there is effectively no difference between either source of sample material, however it should be noted that if whole genome analysis were to be pursued, DNA extracted from blood would be more preferential as there is a lesser chance of bacterial contaminants relative to a DNA sample extracted from saliva. A fact which, of course, could be utilised to great benefit under different applications, such as bacterial detection in hospitals (45,46).

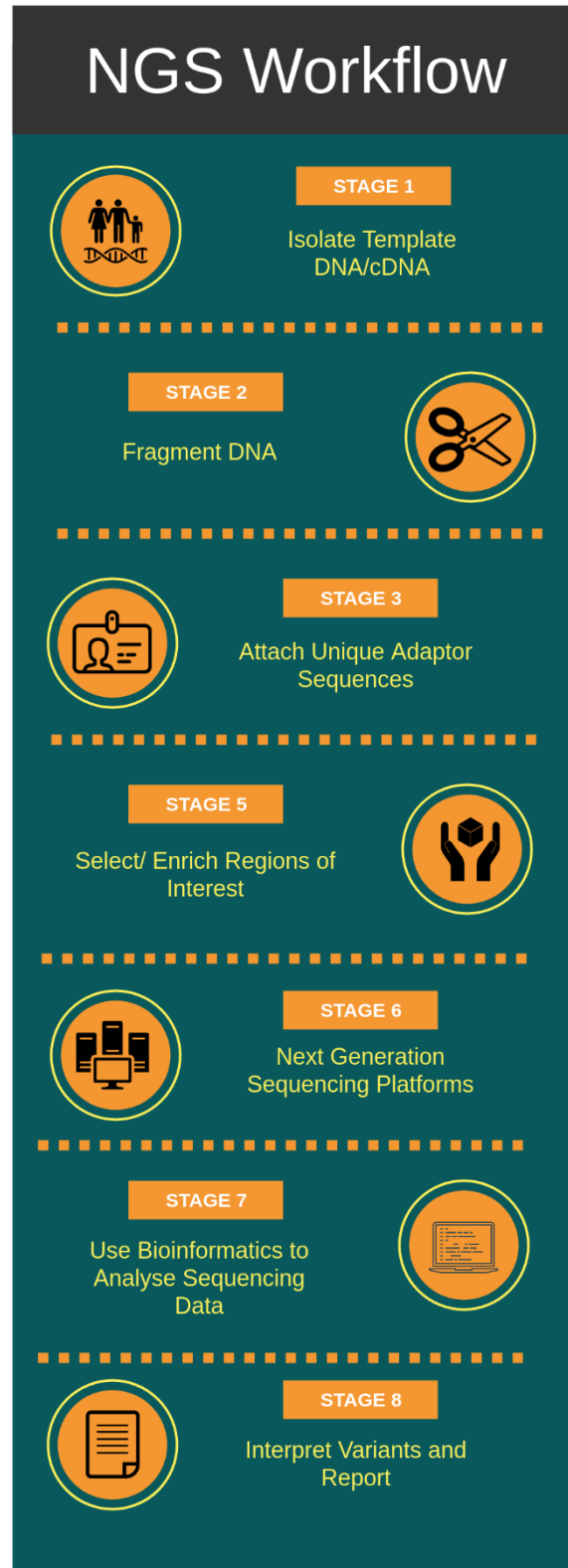


Figure 1.9. Work flow of next generation sequencing. This target enrichment approach is utilised in the Target 5000 project (author's own illustration).

Fragmentation is a key step in library preparation and is of particular importance for shorter-read based platforms. Broadly, fragmentation can occur by physical or biological means. Physical fragmentation is the traditional method and still the most popular, using sonication to break the DNA. Sonication can be applied to double and single stranded DNA. Newer biological methods use enzymes such as Fragmentase to digest large double stranded DNA sequences into smaller fragments in a time-dependent manner. Other methods which can be applied to double or single stranded DNA, allow for the DNA to be simultaneously fragmented and tagged with adapter sequences using enzymes that mimic the actions of transposons (47,48).

Adapter sequences can then be ligated to fragmented DNA to allow sequencing-by-synthesis to occur. Adapters minimally include homologous sequences that allow them to bind to the immobilised fragments within the sequencing platform. Adapters can also include indexing sequences that enable multiplexing of samples during library preparation and demultiplexing of samples bioinformatically.

Target enrichment is generally performed by one of two methods, amplification methods or hybridisation methods. Amplification methods involve some form of PCR to enable the regions of interest to out-compete the undesirable regions. The alternative is hybridisation capture methods, whereby conjugated probes can hybridize to regions of interest, and once hybridised the conjugated probes can be bound to specific beads allowing the remainder of the unwanted sample material to be washed away (Figure 1.10).

The NGS library will then be prepared and ready to load onto the appropriate platform. The different technologies used by the various industry leaders will be discussed in greater detail in the next section.

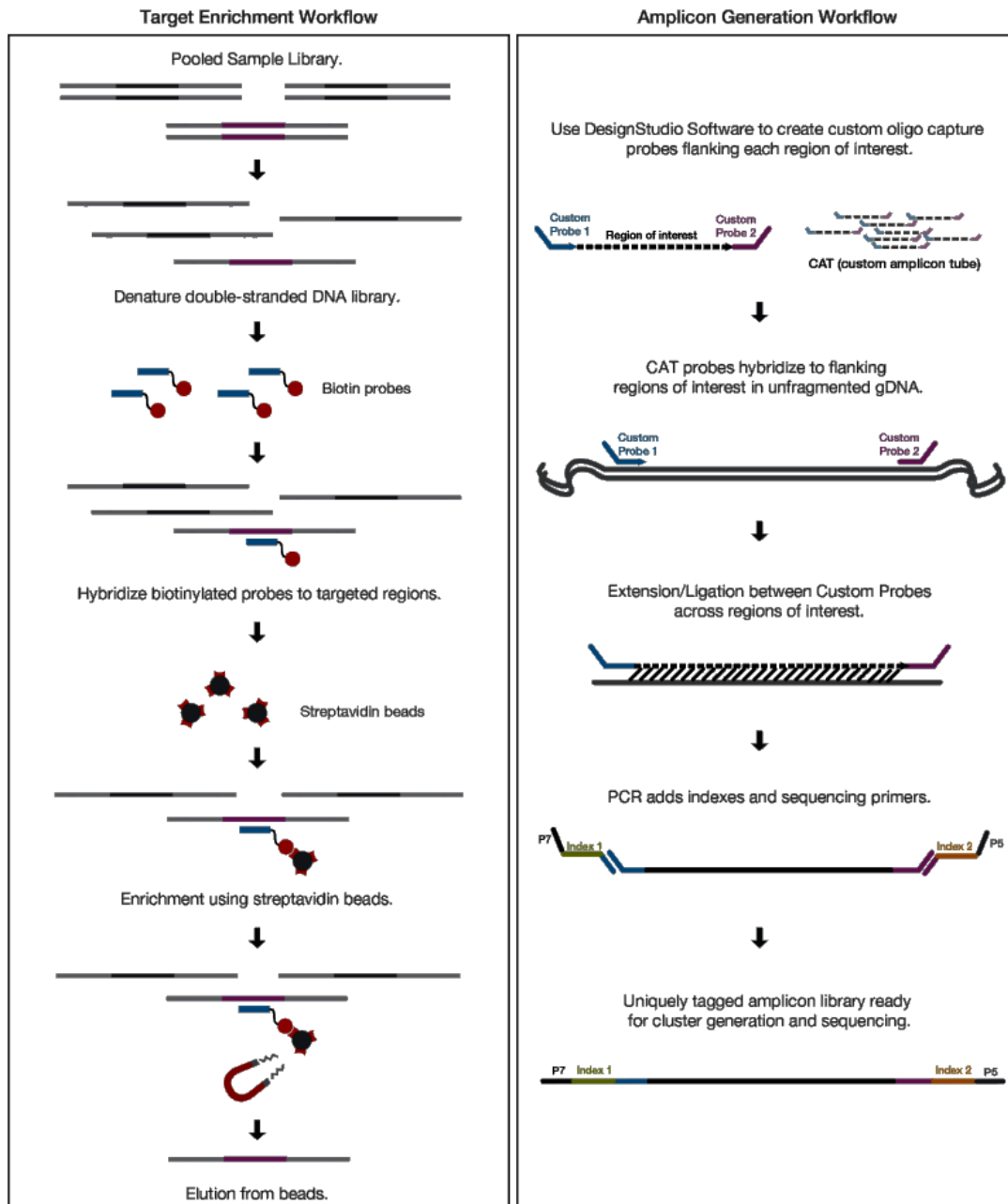


Figure 1.10. Two methods of target enrichment during library preparation. The first method (left panel) uses conjugated probes to selectively bind regions of interest. The second method (right panel) uses amplification to selectively target and tag regions of interest. Image taken from www.illumina.com (41).

1.4.2.4 Next-Generation Sequencing Platforms

NGS or second-generation sequencing (SGS) platforms in recent years have been dominated by three commercial entities: Illumina, Roche and Life Technologies. Their attributes and specific sequencing methods are outlined below.

Illumina Inc.

Illumina have been credited with much of the cost reductions associated with sequencing in the last 15 years. In 2009 they revealed a whole genome service that could achieve a coverage of 30x for \$48,000 (49) which was attributed to the successful integration of the then recently acquired Solexa Inc., who owned patented technologies such as cluster/colony sequencing. A year later that cost had more than halved (50) and now the company has set its sights on producing the first \$100 genome (40). Illumina are reported to hold the vast majority of the market share with a reported revenue of over \$2.75 billion in 2017, compared to the next largest company Thermo Fisher Scientific/ Life Technologies, who had a revenue of approximately \$400 million in 2017 (51).

Illumina, since their inception, have created a large number of sequencing platforms, each tailored to different sequencing capacity requirements. These platforms range from the iSeq 100 model which can produce 4 million reads per run, to the NovaSeq model, which can produce 20 billion reads per run (52). Despite the vast range of sequencing capacity, these platforms all share similar sequencing by synthesis (SBS) protocols (Figure 1.11).

The fragmented and indexed library flows over the optically transparent flow cell, which is essentially a solid surface with probes homologous to the adapter sequences attached

during library preparation. These template fragments bind the passing library samples and bridge amplification is subsequently initiated. This generates hundreds of millions of clusters with each cluster containing roughly 1000 replicates from the same template on the flow cell (53).

After cluster formation by bridge amplification, the library is enzymatically linearised and one at a time, fluorophore-labelled reversible terminating ddNTPs are added. The tagged additions to the clusters are then imaged and the cycle begins again (54).

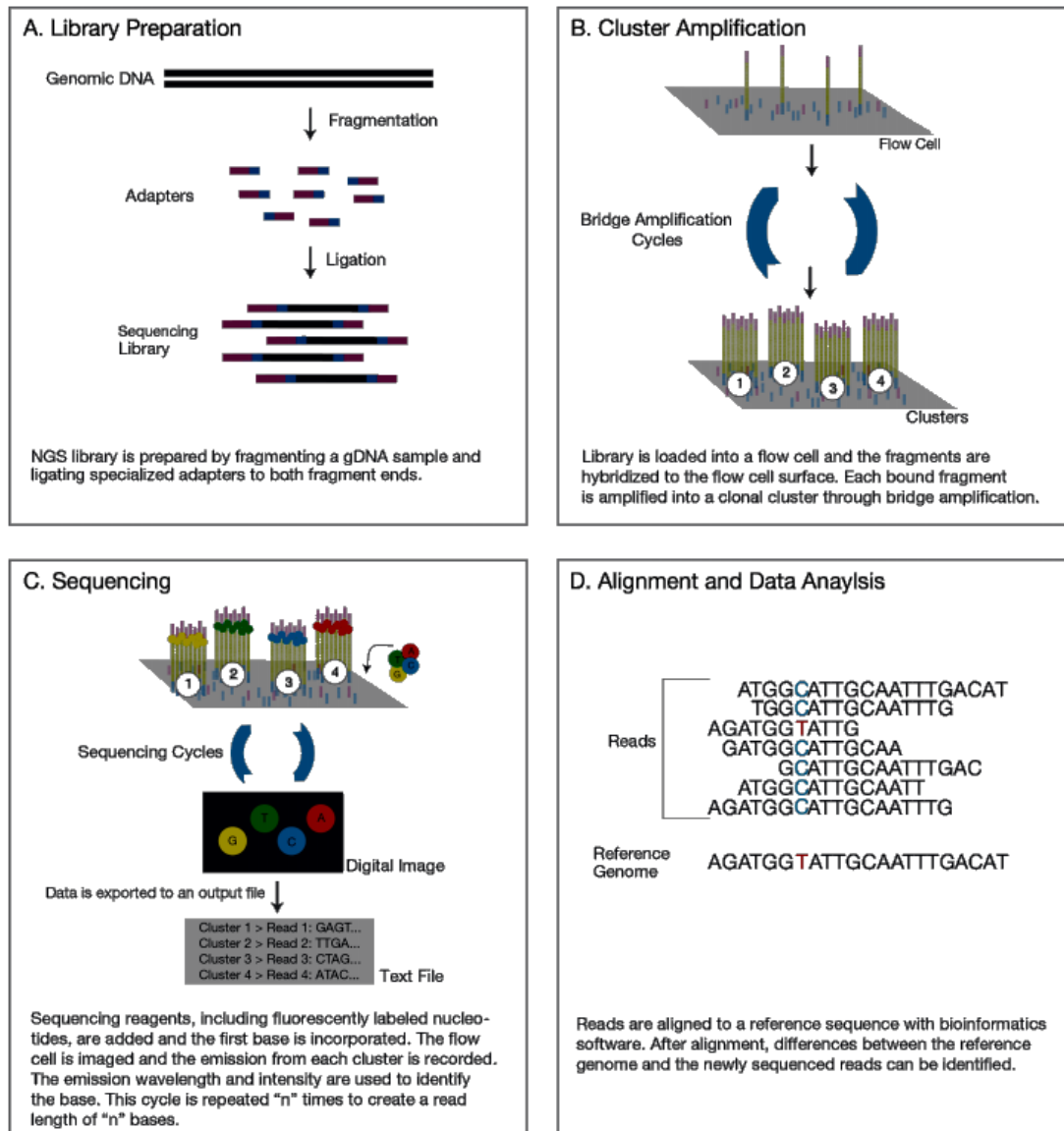


Figure 1.11. *Illumina NGS. Steps B and C occur on the Illumina platform whilst steps A and D are controlled by the person performing NGS. Image taken from www.illumina.com (41).*

Life Technologies Corporation

The Life Technologies Corp. was formed when Applied Biosystems Inc. and Invitrogen Corporation merged. The merger was subsequently purchased by Thermo Fisher Scientific. Life Technologies are best known for their series of sequencing platforms known as SOLiD (Sequencing by Oligonucleotide Ligation and Detection). This sequencing by ligation (SBL) differs from the SBS method because the specificity of the resulting data is based on the incorporated ligase enzyme as opposed to a polymerase.

Template DNA is still prepared by fragmentation, adapter attachment and clonal PCR. However, in this protocol PCR occurs by beads in emulsion (emPCR) and the outcome is that each bead represents the clonal amplification of a single template molecule (54). The SBL method was incredibly accurate, but was overshadowed by the incredibly short read lengths (85 bp), which limited its applications (55).

Roche

In recent years, Roche were one of the three largest suppliers of sequencing platforms (56). Roche bought the company 454 Life Sciences in 2007 after 454 used their technology to sequence the genome of Dr. James Watson (38). The 454-technology employed very similar methodology as Life Technologies Corp., with the variation of using light-based detection of pyrosequencing as opposed to pH change. Phosphate groups released from the sequencing reaction enabled the transformation of luciferin to oxyluciferin which generates detectable light (57).

The 454 systems were regarded as remarkably fast but realistically too expensive for most consumers (54). The platform was also shown to exhibit a high error rate when analysing

sequences containing uniform stretches ($n > 6$) of the same base, as this produced a disproportionate light emission to the actual number of bases (56).

In 2013 Roche announced they would shut down their 454-NGS platform support by 2016, after they had become non-competitive in the market. In the same year they announced a partnership with Pacific Biosciences of California Inc., (PacBio), investing \$75 million in their product development (58).

Other Technologies

There are now “Third Generation Sequencing” (TGS) approaches available. Even the best NGS approaches still had the short-comings of relatively small read lengths and also the issue of amplification bias, due to the common clonal amplification methods. The TGS methods that deal directly with the NGS drawbacks revolve around single molecule real time (SMRT) sequencing. Two companies in particular have developed platforms that reportedly perform SMRT sequencing very well.

Oxford Nanopore Technologies

Oxford Nanopore Technologies have released their minION sequencer which is not much larger than the average USB stick. A nanopore with an internal diameter of 1 nm is used to analyse DNA sequences as they pass through the pore. As the DNA molecule passes through the nanopore each base can be detected by a change in ionic current, this in turn can be translated into signals which can be further processed to call bases (Figure 1.12).

Each sequencer contains a flow cell with between 500-2000 active nanopores. Collectively, these nanopores could sequence hundreds of kilobases per second. It is also

claimed by Oxford Nanopore that this technology will eventually be capable of sequencing the whole human genome in 15 minutes (59). The unique cost effectiveness and mobility of this platform has already made it an essential device in the field of tracing infectious transmissions (60–62). However, this technology is currently associated with a massive error rate, 38.2%, which renders it unusable for many clinical applications (63). Sequencing of small genomes may mask this error rate somewhat where sufficient read depth may add confidence to base calls, however with a genome as large as our own, the low read depth would essentially render the data untrustworthy.

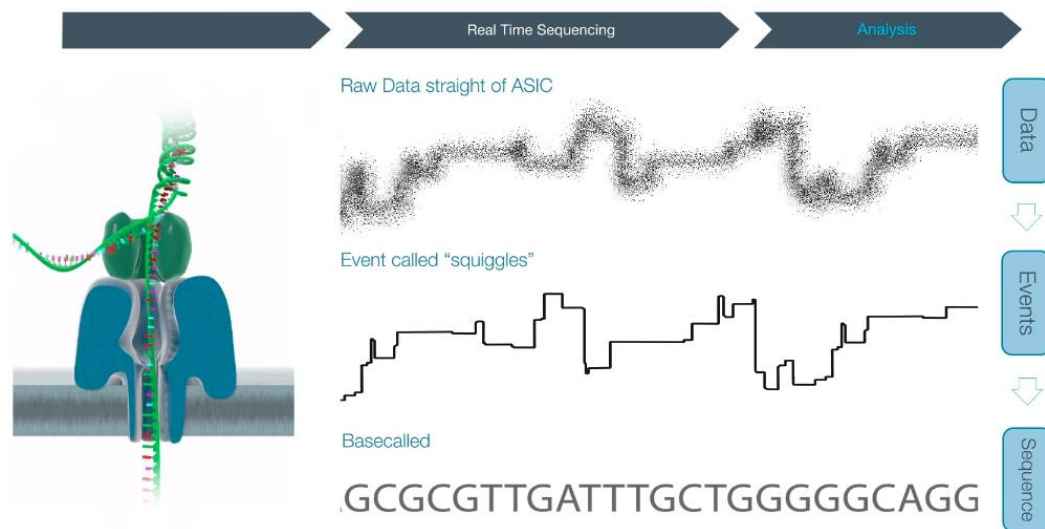


Figure 1.12. The workings of the Oxford Nanopore technology. The image on the left shows DNA passing through the nanopore. The traces on the right demonstrate the data output from this process. The produced signal is cleaned up and interpreted as sequence data. Image taken from www.nanoporetech.com (64).

Pacific Biosciences (PacBio)

In broad terms, PacBio's SMRT sequencing is not dissimilar to that of Oxford Nanopore's approach. PacBio do not use biological pores, instead they create a limited space using nanotechnology. These artificial pores, zero mode waveguides (ZMWs), are approximately 70 nm diameter x 100 nm depth and there are roughly 150,000 ZMWs per chip surface. They each have sufficient space only for their high-fidelity enzyme and a single strand DNA template (65). Within this limited space, the incorporation of a fluorescent-tagged nucleotide can be readily detected. During sequencing the activity of the enzyme is closely monitored at all times.

This method can produce read lengths of approximately 10 Kb but has an error rate of 15%. This error rate is attributed to unincorporated tagged-bases present during readings that create signal noise despite not being utilised by the polymerase. This detection method can also exhibit issues with appropriate quantification of incorporated bases (66), similar to the errors observed in the light-detection method utilised in the Roche platforms.

1.5 Next-Generation Sequencing of IRD Populations

Ophthalmology research has made “a quantum leap” in terms of progress due to the implementation of NGS (67). NGS is now routine in many countries for the detection and diagnosis of IRD patients (23,68–82). Most of these are using target capture sequencing in combination with an Illumina platform due to superior read depth and cost effectiveness (83). This strategy is capable of detecting candidate mutations in the majority of IRD patients sequenced and boasts similar detection rates to whole-exome studies (this will be discussed more later in the section).

However, as the technology becomes cheaper and more accessible, exome and whole genome sequencing, in conjunction with emerging methods to define the function of candidate mutations in regulatory and intronic regions will undoubtedly progress our understanding of the molecular genetics of IRDs to a new level. This is also true for the capture and characterisation of copy number variations (CNVs) which have been implicated as causative of some forms of IRD but go largely undetected by target capture sequencing (84).

The situation for IRDs is further complicated by the diversity of clinical presentations that can be caused by mutations even within a single IRD gene, as well as overlapping clinical phenotypes which may be the result of mutations in entirely distinct genes, making it impossible for a diagnosis to be given in the majority of cases on the basis of disease phenotype alone (68,85–87). By way of example, in a recent NGS study of an Irish IRD patient cohort, new disease phenotypes were associated with the *GNAT1* and *SLC24A1* genes (85). In both cases a severe homozygous mutation in a known congenital

stationary night blindness (CSNB) gene caused a late-onset form of RP involving photoreceptor loss.

Furthermore IRD phenotypes can be influenced by modifier loci (88–90). Given the limitations associated with the clinical diagnosis of specific forms of IRD, it is no surprise that the majority of patients with IRDs do not know the variant(s) responsible for their condition. Recent advances in the development of gene-specific therapies for IRDs has made this issue all the more pressing, since this offers, for the first time, the prospect of modulating, halting or reversing the degeneration associated with some of these conditions (91), however, such therapies cannot, obviously, be administered without a precise knowledge of the underlying mutation to be treated.

High throughput NGS technologies offer an opportunity to rapidly characterize causative mutations in the growing number of IRD genes that have now been identified. These technologies have greatly accelerated DNA sequencing by providing a means of simultaneously sequencing many small fragments of DNA from different regions of the genome in a single reaction, and moreover by barcoding patient DNA to enable sequencing of samples from multiple patients in parallel. The short sequence reads from such fragments are then reconstructed by comparing the sequence to a reference genome; hg38 being currently the reference genome of choice for many such studies (hg38; GCA_000001405.15).

NGS can be undertaken on the whole genome or limited to expressed (coding) sequences (whole exome) or targeted to particular regions of the genome (employing target capture panel NGS); where genes of interest are, in essence, captured and sequenced. This form

of sequencing has advanced at an exponential pace, with the cost per megabase of sequence halving regularly (92) and set to reduce further with the introduction of the new NovaSeq series from Illumina, among other innovations. IRDs have an ideal profile for NGS studies (69,83,93), representing a group of Mendelian conditions where many (to date approx. 260 genes), but not all of the causative genes, have been characterized. Given this scenario, NGS studies enable identification of known, or novel mutations in known genes and the stratification of IRD patient cohorts into those whose genetic pathogenesis has been resolved, and those for whom new genes, or variants in regulatory sequences, splice variants or CNVs associated with disease genes, may be causative of disease and therefore additional analyses deploying whole exome or whole genome sequencing would be appropriate.

A variety of NGS studies, many employing a target panel-based NGS approach focused on sequencing the exons of known IRD genes, have been undertaken (68–82,93) (Table 1.1 (5)). These studies support the view that IRDs represent an exemplar group of disorders for the application of panel-based NGS or WES as effective tools for detection of causative mutations. From such studies, some common patterns in findings have been observed, as have some unique findings, which at present, remain specific to individual studies. Of note, using target-capture based NGS, approximately 50–60% of IRD patients were found to carry disease causing or likely disease-causing mutations in many studies.

Table 1.1. *Table of representative studies using NGS to investigate IRDs. Table taken from Farrar et al. (5).*

IRD patient numbers	Patient origin (predominant)	NGS technology	Pathogenic/likely pathogenic mutations (%)	Reference
100	Dutch	Target capture 111 genes	~ 50%	Neveling et al. 2012
20	France	Target capture 254 genes	57%	Audo et al. 2012
50	UK	Target capture 105 genes	50%	O'Sullivan et al. 2012
50	UK	Target capture 73 genes	53% (early onset IRD; 25% overall)	Shanks et al. 2013
192	US	Target capture (GEDI) 214 genes	51%	Consugar et al. 2015
105	eyeGENE® database	Target capture 195 genes	49.5%	Ge et al. 2015
82	N Ireland	Target capture 186 genes	60%	Zhao et al. 2015
88	Italy	Target capture 72 genes	59.1%	Bernardis et al. 2016
309	Ireland	Target capture 216 genes	57%	Carrigan et al. 2016
537	Worldwide	Target capture 105 genes	51%	Ellingford et al. 2016
47	Spain	Target capture 75 genes	57.4%	Perez-Carro et al. 2016
88	European	Target capture & WES	61%	Weisschuh et al. 2016
58	Swiss	WES	64%	Tiwari et al. 2016
33	Mixed	WES	55%	de Castro-Miró et al. 2016
59	Spain	WES	71.8%	Riera et al. 2017
266	Dutch	Target panel & WES	52%	Haer-Wigman et al. 2017

Given additional family studies indicating segregation of the disease, and replication of the findings in a certified diagnostic laboratory, these patients would be deemed to be categorised as ‘resolved’ in terms of the genetic aetiology of the retinal pathology. Among recent examples of capture-panel NGS for IRDs, is a study of 537 IRD patients in which the variants identified were deemed to account, or likely account for the disease in 51% of cases and a recent study of Italian IRD patients, in which 59% of cases were resolved, similar to the identification levels obtained in many other NGS studies (68,70).

Detection rates were significantly affected by the range of conditions under study, the number of genes included in the capture panel, as well as whether the rate was corrected based on previous screening of the same population, making precise comparisons of detection rates between studies challenging. Nonetheless, there is a consistent finding that between 25-50% of cases were not solved by targeted sequencing or WES.

It is of interest that studies employing WES, rather than capture-panel NGS, have resulted in the identification of the underlying genetic defect in similar levels of IRD patients ~60% (71) to capture panel sequencing, although in a recent study of Spanish IRD patients, WES resulted in successful identification of pathogenic variants in approximately 70% of patients (72). The similarity in levels of resolution using capture-panels sequencing and WES suggests that there may not be many as yet unidentified new IRD genes.

Levels of mutation detection achieved for IRDs that are less genetically heterogeneous, such as choroideremia, involving solely the CHM gene, and Stargardt disease, predominantly caused by *ABCA4* mutations, were typically significantly higher (23,94).

However, even within such comparatively homogenous forms of IRD, the level of intragenic mutational heterogeneity present is still being characterised. For example, in a recent study in which 148 pathogenic or likely pathogenic mutations in the *ABCA4* gene were identified, about a third of these (n=48) represented new Stargardt- associated disease alleles (95).

There is a need for cohorts of IRD patients to receive accurate diagnoses. With each newly recruited patient there develops a burden on clinical, scientific and counselling professionals, a burden which may largely go unresolved without the discovery of causative underlying genetic components. The detection of pathogenic variants allows for more accurate diagnosis and potentially a more defined prognosis. Even without potentially facilitating access to the many ongoing clinical trials for IRDs (www.clinicaltrials.gov), there may still at times be the potential of empowerment in terms of slowing disease progression and providing information on likely modes of inheritance, information relevant to many family members and prospective parents.

Such an example is dietary intervention. For Stargardt disease there is evidence to suggest that reducing vitamin-A intake or use of vitamin-A substitutes may slow disease progression (96). For patients presenting with gyrate atrophy and cystoid macular edemas, reducing protein intake along with pyridoxine supplementation was shown to have a notable effect on improvement and even reverse the condition (97). Therefore, emphasising the fact that clinical trial participation should not be the only metric for success in these studies is important and that accurate genetic testing can result in an improvement of quality of life by alternative treatments for some IRDs and moreover can

be of important prognostic value and can provide essential information for genetic counselling and reproductive decisions.

2 Materials and Methods

2.1 Ethical Approval

Ethical approval for this study was awarded by the Research and Medical Ethics committee of the Royal Victoria Eye and Ear Hospital, Dublin, Ireland (13-06-2011: HRA-POR201097) and by the Institutional Review Board of the Mater Misericordiae University Hospital and Mater Private Hospital, Dublin, Ireland (MMUH IRB 1/378/1358) prior to commencement. All work was carried out in accordance with the approved guidelines. All patients have given written informed consent before recruitment to the study. No patients under 18 years of age were included in the study prior to the approval of amendments made to 13-06-2011: HRA-POR201097 which specified the terms of their inclusion in the project.

2.2 Patient Assessment Sites

Probands and, where applicable, other family members were primarily assessed at the Research Foundation of the Royal Victoria Eye and Ear Hospital (Dublin, Ireland), the Mater Misericordiae University Hospital (Dublin, Ireland) and The Royal Victoria Hospital (Belfast, Northern Ireland). Patients were referred from general practitioners and optometrists from all over the island of Ireland.

2.3 Clinical Assessment

Best-corrected visual acuity was assessed using revised 2000 Early Treatment Diabetic Retinopathy Study (ETDRS) charts (Precision Vision, La Salle, IL, USA). Colour vision was examined using the Lanthony desaturated D-15 panel (Gulden Ophthalmics, Elkins

Park, PA, USA) under standardised lighting conditions. Goldmann perimetry was used to assess the peripheral visual fields to the IV4e, I4e and 04e targets.

Full-field electroretinograms were performed according to the International Society for Clinical Electrophysiology of Vision (ISCEV) standards (98) using a Roland Consult RETI-port retiscan (Brandenburg an der Havel, Germany).

Fundus colour and autofluorescence photography was performed using a Topcon CRC50DX (Topcon Great Britain Ltd., Berkshire, England) or Optos Daytona (Optos plc, Dunfermline, Scotland). Spectral domain optical coherence tomography was performed using a Cirrus HD-OCT (Carl Zeiss Meditec, Berlin, Germany).

2.4 Sample Collection from Patients

Once a conclusion was made that the patient may potentially have a retinal condition with an underlying genetic cause, a biological sample would be requested in order to facilitate inclusion in the Target 5000 project. A 5 mL blood sample would be requested but where that was physically unobtainable or deemed undesirable by the patient, a saliva sample would be taken instead. When an RNA blood sample was to be taken, a DNA blood sample would be taken first as per the PAXgene, PreAnalytiX (Hombrechtikon, Switzerland) protocol.

2.5 DNA Extraction from Blood

5 mL of whole blood was collected from each patient and stored in BD Vacutainer blood collection tubes at -20°C. 2 mL of whole blood was used per DNA extraction. Extractions were performed using Qiagen (Hilden, Germany) QIAamp DNA Blood Midi Kits utilising the Spin Protocol. Although not suggested in the protocol, samples were thoroughly vortexed prior to lysis incubation at 70°C. Samples were incubated at 70 °C for lysis for ten minutes, removed for vortexing, and replaced in the incubator for at least a further 5 minutes. This was found to yield superior quantities of DNA.

2.6 DNA Extraction from Saliva

4mLs of patient saliva samples were collected using DNA Genotek (Ottawa, Canada) ORAcollect DNA tubes. DNA was isolated using the prepIT-L2P kit from the same company. When possible, DNA was isolated in parallel with whole blood samples using QIAamp DNA Blood Midi Kit, this substitution did not notably affect DNA yields or quality.

2.7 RNA Extraction from Blood

Patient RNA blood samples were collected using PreAnalytiX's PAXgene Blood RNA Tubes (IVD). RNA was isolated using the PAXgene Blood RNA kit (IVD). The original protocol calls for the use of a discard tube to collect and dispose of the first few millilitres of patient blood, however this blood instead was collected in Vacutainer blood collection tubes and utilised for the extraction of a DNA sample from the same patient.

2.8 DNA Shearing for Next-Generation Sequencing

8µg of isolated patient DNA was diluted to a uniform volume of 200 µL in Diagenode (Denville, United States) Bioruptor Plus Microtubes to maximise the efficiency of sonication. Two rotors were employed simultaneously to enable the paralleled fragmentation for up to 24 samples. Sonication of samples consisted of cycles of 30 seconds sonication and 30 seconds rest. This process took approximately 40 minutes to achieve fragment sizes in the region of 250 bp. The sonication bath was replaced with ice cold water every 10 minutes to ensure over-heating did not occur.

2.9 DNA Purification Using Magnetic Beads

Samples were purified using Beckman Coulter (California, United States) Agencourt AMPure XP beads. AMPure XP is a commercially prepared solution of carboxylate-modified magnetic beads, poly (ethylene glycol) and NaCl. 0.9 volumes of homogeneous AMPure XP were added to each sample that required purification. If maximum DNA recovery was required, up to 1.8 volumes of beads were used.

This solution was then adequately mixed and given 10 minutes to facilitate sufficient electrostatic coupling of the carboxylic groups on the beads to the DNA of interest. The sample and beads were then placed on a magnetic stand to allow for the separation of the bead-bound DNA and the remaining buffer. Once clear, the remaining buffer was removed and discarded. The bead-bound DNA, still on the magnetic surface, was then washed with 180 µLs of freshly prepared 80% ethanol with nuclease free water (measured separately and then mixed as the sum of their parts will not equal the resulting volume). This ethanol wash was left for 30 seconds to allow any disturbed beads to settle back

towards the magnet. The ethanol solution was then removed and replaced with another 180 μ Ls of freshly prepared 80% ethanol.

After allowing the beads to settle back to the magnet, this solution was removed, and the tube was removed from the magnetic surface. Appropriate elution buffer* was then added. The beads were then adequately mixed and given 10 minutes to facilitate the dissociation between DNA and the beads. The solution was then placed back on the magnetic stand and allowed to separate (approximately 2 minutes). The solution could then be removed without the carry-over of beads contaminants and placed in a new receptacle. When used in a 96-well format, this method provided a time-efficient and reproducible method of achieving purified DNA for a multitude of applications.

* For long term storage, traditional TE buffer was used (contains 1 mM EDTA). For reactions bound for sequencing platforms, low-EDTA TE buffer was used (contains 0.1 mM EDTA). For sequencing library preparation, 10mM Tris-HCl was used. For dehydration reactions, as used in the target capture protocol, nuclease-free water was used.

2.10 Size-selection of DNA Products Using Magnetic Beads

DNA is aggregated and bound to magnetic beads more readily in a hydrophobic environment (provided by PEG 8000) when salt ions (NaCl) are employed to neutralise the negative charge of DNA's phosphate backbone (99). However, DNA of higher molecular mass is more sensitive to precipitation at lower concentrations of PEG and NaCl than smaller fragments (100). This effectively means that when a small ratio of beads to sample material is used, DNA fragments of longer length will be selectively bound. Although the standard ratio of 0.9 volumes of beads would be sufficient to remove any undesirable primer sequences, the volume of beads would have to be reduced further if only larger fragments (above 200-300 bp) were desired.

Equally, this method was utilised to selectively remove fragments above the desired size. This was done by using much smaller ratios of beads, 0.1 - 0.2 volumes, and allowing the beads to bind the sample DNA. However, instead of continuing with the washing of the bound DNA and discarding the remaining solution, the desired DNA would remain in the unbound fraction and could simply be transferred to a new receptacle. The larger fragments bound to the beads could then be separately purified or discarded if not required.

2.11 Homemade DNA Binding Beads

Over the course of this project, use of the AMPure XP product proved immensely costly and was later replaced by a homemade solution of DNA binding magnetic beads, formulation recommend from (101). The applications of the beads however, as mentioned above, remained unchanged.

2.11.1 Homemade DNA Storage Buffer

When required, this was made from the following base ingredients.

Ingredients for 50 mL, pH 8.0 @ 25 °C:

Nuclease-free water	48.564 mL
Tris base, 1 M	0.500 mL
Disodium EDTA, 0.1 M	0.500 mL
Tween 20, 10% (v/v)	0.250 mL
HCl, 1 N	0.186 mL

2.11.2 Homemade DNA-Binding Bead Solution

When required, this was made from the following base ingredients.

Ingredients for 50 mL, pH 8.0 @ 25 °C:

NaCl, 5 M	25.000 mL
PEG 8000, 50% (w/v)	20.000 mL
Nuclease-free water	3.582 mL
Sera-Mag bead suspension*	1.000 mL
Tris base, 1 M	0.500 mL
Disodium EDTA, 0.1 M	0.500 mL
HCl, 1 N	0.168 mL
Tween 20, 10% (v/v)	0.250 mL

*GE Healthcare (Illinois, United States), Sera-Mag SpeedBead Carboxylate-Modified Magnetic Particles (Hydrophobic).

2.11.3 DNA Binding Bead Solution (Preparation)

Sera-Mag beads were vortexed thoroughly to ensure homogeneity. 1 mL of beads were transferred to a fresh 1.5 mL microcentrifuge tube, this was done quickly as beads will readily settle to the bottom of the bottle. A magnetic stand was used to separate beads from the remaining solution (this takes approximately 1 minute). Once the remaining solution was clear, it was removed, leaving the magnet-bound particles behind. The microcentrifuge tube was then removed from the magnetic stand and the bead fraction was resuspended in 1 mL DNA storage buffer (described above). Samples were mixed by vortexing for at least 10 seconds. If liquid had gathered in the microcentrifuge tube lid, samples were very briefly centrifuged and carefully vortexed. Samples were again placed on the magnetic stand to allow for separation of magnet-binding beads from the remaining solution. The remaining solution was then removed. This washing process was repeated for a total of 3 washes with the bead fraction remaining magnet-bound on the 3rd wash. During the binding stage, the previously specified volumes of Tris, EDTA, NaCl, HCl and water were amalgamated in a 50 mL tube. Once the clear supernatant was removed from the bead fraction in the microcentrifuge tube, beads were resuspended in 1 mL of the solution created in the 50 mL tube. The new bead solution was vortexed to homogenise. This 1 mL sample was then incorporated back into the remaining solution in the 50 mL tube. The 50 mL tube was vortexed for 30 seconds to ensure homogenisation. 20 mL of PEG 8000, 50% (w/v) was then added to the 50 mL tube. This was done slowly as the solution is very viscous and aggressive pipetting would result in inaccuracy. The final ingredient, Tween 20, 10% (v/v), was then added. This solution was mixed thoroughly until homogeneous. The solution was then stored at 4°C.

2.12 Quantification of DNA

The NanoDrop ND1000 spectrophotometer delivers fast and accurate quantification from as little as 1 μL of sample. Its 280/260 value output is a quick indication of sample contamination from phenol or other organic components from traditional DNA extractions. The short-comings of this device is firstly, it is not accurate at low concentrations of DNA ($<20 \text{ ng}/\mu\text{L}$). Secondly, it will detect all nucleotides in the sample, including free floating and undesired carry-over from PCR reactions. This is not an issue if the sample has undergone purification prior to quantification.

The Qubit fluorometer uses an integrating dye to selectively identify double stranded DNA. The kits for this device can be used to quantify concentrations as low as $10 \text{ pg}/\mu\text{L}$, however the sample preparation time is significantly longer than the NanoDrop, particularly for larger sample sets. Readings will also gradually fluctuate over time.

The Agilent TapeStation system is an automated platform for sensitive electrophoresis. Its detection range is 35 – 1000 bp and has the advantage of quantification and fragment sizing. This undoubtedly makes it the most powerful of the quantification devices, however the cost of reagents are significantly more than the other methods and sample sizes are limited to 15 samples per run with a DNA ladder, 16 without a ladder.

When utilised efficiently, these three combined methods can provide quick and accurate qualitative and quantitative information key to the quality control of your sample.

2.13 KAPA Library Preparation

This kit/protocol was used initially to prepare sequencing libraries and was used largely as per suppliers' recommendations. Briefly, DNA samples were sheared as per protocol above. The resulting DNA was purified using DNA-binding magnetic beads. The DNA was eluted in 10 mM Tris-HCl solution as enzymes used as part of this kit are sensitive to alterations of ion concentrations and thus no solution containing EDTA could be used as input for this protocol as it would result in the chelation of divalent ions. 1 µg of patient DNA in 50 µL, fragmented to an average size of roughly 250 bp would be used as input.

An End-Repair solution was used to ensure blunt-ends of the fragmented DNA. After a brief incubation, the resulting DNA was purified from this reaction using magnetic beads. The eluted DNA was added to an A-Tailing solution to ensure 3'-adenylation of the blunted DNA. Again, the reaction was incubated briefly, and the resulting product was purified by the magnetic bead method. This enabled the modified DNA to be adhered to adapters via a ligation reaction. These single-index adapters allowed for the eventual pooling of up to 24 patient samples in a single sequencing reaction.

The post-ligation product was cleaned twice using magnetic beads. The second purification also served as a size-selection step as recommended by manufacturer. The size-selection protocol was modified slightly to optimise the output fragment range. The initial bead volume was increased by 10% to 55 µL. This allowed for slightly more exclusion of undesired longer fragments of DNA. The second volume of beads was increased by 25% to 25 µL. This allowed for greater retention of product following purification.

The size-selected DNA was then used as template for the first PCR cycle. The number of PCR cycles were reduced to prevent over-amplification and replication-induced errors in the resulting sequences. Samples were then ready for target capture using Nimblegen enrichment.

2.14 KAPA HyperPlus Library Preparation

The HyperPlus protocol introduced some very welcome changes compared to its older, more time-consuming counterpart. Input DNA no longer required prior shearing due to the introduction of the Fragmentase enzyme. The enzymatic fragmentation introduced a more flexible and rapid method for achieving the desired fragment size.

This protocol also introduced more robust enzymes throughout, as several tedious purification steps were removed. End-Repair and A-Tailing steps now occur simultaneously and require no product purification before adapter ligation. This ultimately has resulted in a further reduction in PCR amplification, as less sample material is lost during the many purification steps. Therefore, a more patient-representative sequence can be carried through to sequencing.

2.15 Indexing Adapters

Initially NGS library sample pools were limited by the available number of unique indexing adapters. With Roche, this was 24 unique single index adapters. This number was well suited to the Illumina MiSeq Platform as it provided ample coverage across the target captured regions.

In later studies dual-index adapters were used, 8x i5 indices and 12x i7 indices, allowing for 96 unique pairings of adapters. The implementation of the dual-indices did not exclude samples from use with the MiSeq, in fact it made the multiplexing and sample number more flexible. Greater or fewer than 24 samples could now be pooled into the same sequencing run and the limiting factor then becomes the sequencing output capacity of the platform. There was also now more variability in the indices to select from if only running a small number of samples. This was an issue particularly with single indices as “index-hopping” could occur in a fraction of samples if the index diversity was low enough. Very specific pairings of single-indices would have to be used together if the multiplex pool contained 4 or less samples.

Another important consideration in the change from single to dual-index adapter is the use of blocking oligonucleotides. Blocking oligonucleotides are utilised in NGS library preparation to prevent the daisy chaining effect, which occurs when homologous regions of the adapters erroneously bind to each other during hybridisation. The negative impact of a daisy chain effect is that hybrid capture purification results in the simultaneous capture of off-target regions. For Roche KAPA library preparations, specific blocking oligonucleotides are available for each single index adapter as well as a generic blocking oligonucleotide. To negate the daisy chaining effect in dual-indexed samples, a pool of generic blocking oligonucleotides, xGen Universal Blocking Oligos (IDT) were used at the same concentration as the combined specific and generic blocking oligonucleotides used for the single indexed samples.

The major advantage of the implementation of dual-index adapters was that 96 samples could now be run simultaneously on larger sequencing platforms such as the Illumina

HiSeq 2500 Platform. Despite the sample pool containing 4 times more samples, the maximum output per run capabilities of the HiSeq 2500 compared to the MiSeq were now also 4 times greater (on Rapid Run mode per single flow cell, 2 x 100bp, for HiSeq 2500) (52). This meant that 96 samples could now be run in parallel, reducing preparation time, sequencing time and cost. When provided with the sequences, IDT were able to produce the dual-index adapters for a fraction of the cost that Illumina sold them for, economies of scale also allowed for a reduced cost as these requirements were unlikely to change again for the remainder of the project.

2.16 Nimblegen Target Capture Enrichment

Broadly, this protocol follows the guidelines outlined in SeqCap EZ Library SR User's Guide version 5.0 and continues on from the KAPA sequencing library preparation from patient gDNA. At this point patient DNA has been uniquely indexed and amplified. DNA is then quantified using a combination of electrophoresis methods (Agilent D1000 Tapestation) spectrophotometry (NanoDrop 1000) and fluorometry (Qubit). Accurate quantification at this point was vital to ensuring equal sequencing of each sample. Nimblegen probes were used to capture desired regions of the genome. Nimblegen probes (SeqCap EZ Choice Library) were chosen over their competitors due to their superior coverage and evenness of capture (102). Initial enrichment was done using Agilent probes but these were ultimately replaced by the superior Nimblegen probes.

When quantified, the indexed DNA can be pooled. Despite the upgrade to dual-indexed adapters during the project, the remainder of the protocol was optimised for 24-sample pools and that remained unchanged after brief efforts to further customise this segment of the protocol yielded sub-optimal results in terms of probe hybridisation efficacy.

Biotinylated hybridisation probes were incubated with DNA pools for approximately 60-70 hours as opposed to the suggested 16-20 hours, this was seen to increase hybridisation efficiency in the past and also found to be true in other similar studies (102). The hybridised DNA was then bound by streptavidin-conjugated magnetic beads, allowing the sample to be washed in order to remove all patient DNA that did not fall within the regions of interest. If not washed effectively, the limited sequencing capacity for that sample would be spread across the whole genome, potentially producing an unusable, shallow read depth across regions of interest.

The 24-sample library pools were then amplified. Again, the recommended cycle number was reduced to best avoid issues of erroneous replication and amplification bias. DNA pools were then purified and quantified using previously discussed methods.

2.17 Variant Verification by Direct Sequencing

Any variants of interest, that could potentially be deemed causative, were checked for segregation in the other available members of the pedigree. This included all members recruited to the study at that time, both affected and unaffected.

For point mutations, the approximate 300bp region surrounding the variant was amplified by PCR. Primers were designed using Primer3 software. Primer oligonucleotides were synthesised by Sigma-Aldrich (Missouri, United States). Template sequences were taken from the Hg38 reference genome. Variants required personalised optimisation of primers and amplification criteria however, deviations from the conditions stated here (Tables 2.1 and 2.2) were minor, if any. Q5 Polymerase (New England Biolabs, Massachusetts,

United States) was chosen as the enzyme of choice for standard PCR experiments as it boasts an extremely low replication error rate, at least 280-fold lower than standard Taq enzyme (103).

Products from optimised PCR experiments were purified using ExoSAP-IT PCR Product Cleanup Reagent (Thermo Fisher Scientific, Massachusetts, United States) or the previously described DNA-binding magnetic beads. DNA was then quantified, and sequencing of these products was outsourced to Eurofins Genomics (Konstanz, Germany). Given that many of the variants of interest were suspected to be present heterozygously, presence or absence of variants were analysed in sequence output traces and plain text files were largely ignored unless the variant of interest was X-linked and the patient was male.

Table 2.1. *Components and concentrations of standard PCR reactions for direct sequencing.*

COMPONENT	25 µl REACTION	50 µl REACTION
5X Q5 Reaction Buffer	5 µl	10 µl
10 mM dNTPs	0.5 µl	1 µl
10 µM Forward Primer	1.25 µl	2.5 µl
10 µM Reverse Primer	1.25 µl	2.5 µl
Template DNA	variable	variable
Q5 High-Fidelity DNA Polymerase	0.25 µl	0.5 µl
Nuclease-Free Water	to 25 µl	to 50 µl

Table 2.2. *Thermocycler conditions of standard PCR reactions for direct sequencing.*

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds
35 Cycles	98°C 62°C 72°C	10 seconds 30 seconds 30 seconds
Final Extension	72°C	2 minutes
Hold	4-10°C	

2.18 Amplification of *RPGR* ORF15

Countless optimisations of PCR conditions were unsuccessful when amplifying the repetitive region of exon 15 of the *RPGR* gene with Q5 and other polymerases. Amplification was successful, specific and reproducible with HotFire Polymerase (Solis Biodyne, Tartu, Estonia) as recommended by Dr. Kinga M. Bujakowska (Massachusetts Eye and Ear) (via personal communication).

PCR Primers F3 (5'-GACTAAACCCATAATATCCAAATCCA-3') and R6 (5'-GCCAAAATTTACCAGTGCCTCCTAT-3') were used for amplification of the target region. The conditions of this reaction are outlined below (Table 2.3 and 2.4). PCR products were purified for sequencing using ExoSAP-IT reagents. Sequencing of the amplified product was completed by tiling together results of sequencing reactions using the following sequencing primers: R4 (5'-CTCTCCTTCCTCCTTTTCAC-3'); R7 (5'-CCTTCCTCCTCTTCCCCCTCA-3'); R8 (5'-TCCTTCCTCCTCTTCCCCCTCCCA-3'); R9 (5'-CCCTGTGTGTTAGTAACTGAC-3'); F10 (5'-GGAAATGGGAAAGAGCAGAGGTC-3'); R5 (5'-ACTGGCCATAATCGGGTCACAT-3'). These primers are shown to align to *RPGR* ORF15 in Figure 2.1.

Sequencing of the PCR products was carried out by Eurofins Genomics. Eurofins Genomics also provide a “Power Read Upgrade” service. This service is tailored for harder to sequence DNA reads. Most reactions yielded high quality results from the standard service, however crucially, results from sequencing reactions using sequencing primers R8 and R9 were greatly improved when the additional “Power Read Upgrade: Repetitive Region” service.

Additionally, due to the repetitive nature of this sequence, aligning sequences became notably error-prone, especially when attempting to align multiple query sequences against the reference genome for exon 15. MAFFT's (104) scoring algorithm appears to be faster and more capable of performing this action than the more popular Clustal Omega (105). MAFFT's web tool allows for modifications of the scoring mechanism of alignments, including the gap creation and gap extension penalties which also proved informative.

Table 2.3. *Components and concentrations of RPGR Exon 15 PCR reactions for direct sequencing.*

COMPONENT	25 µl REACTION
10X Buffer B2	2.5
MgCL2 (25mM)	1.5
dNTP (20mM)	0.25
Solution S (x3)	7.5
Primer F3 (10uM)	0.75
Primer R6 (10uM)	0.75
H2O	6.5
HotFire Polym	0.25
DNA (20ng/ul)	5

Table 2.4. *Thermocycler conditions of RPGR Exon 15 PCR reactions for direct sequencing.*

STEP	TEMP	TIME
Initial Denaturation	95°C	15 minutes
45 Cycles	95°C 60°C 72°C	45 seconds 60 seconds 3 minutes 30 seconds
Final Extension	72°C	10 minutes
Hold	4-10°C	

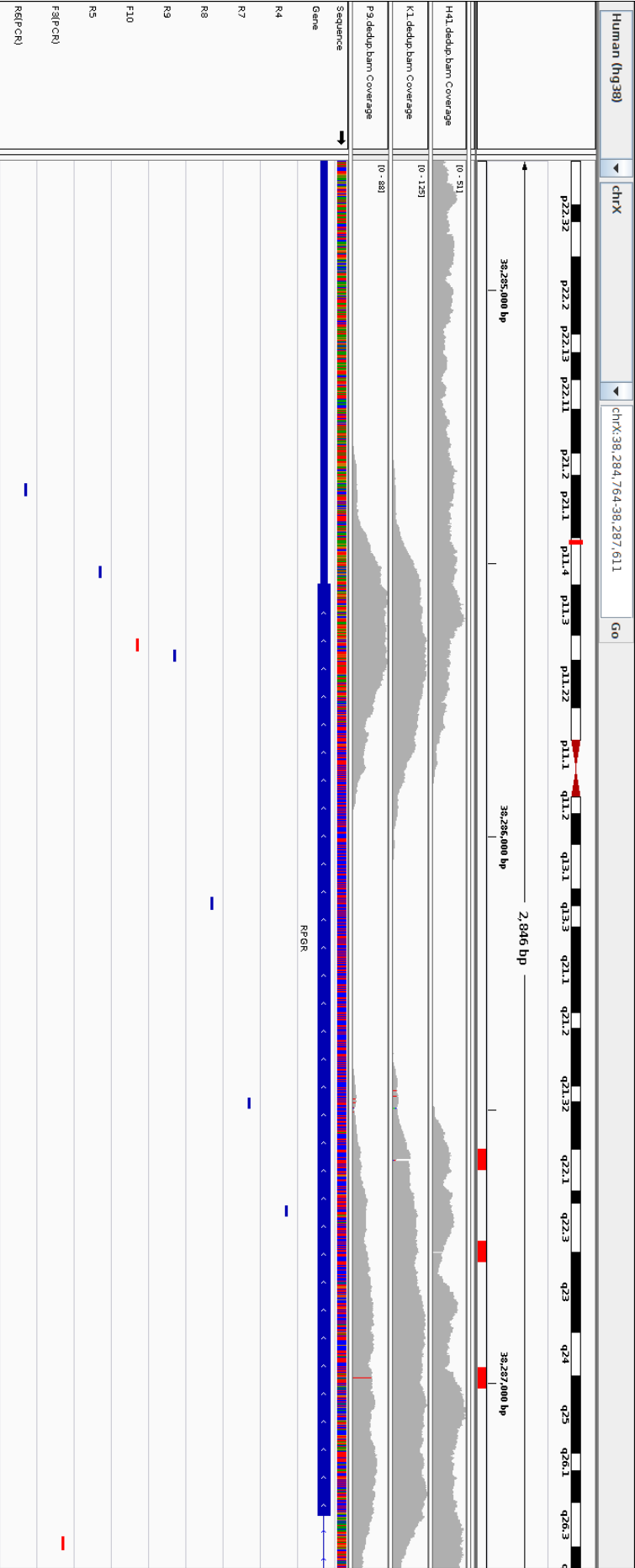


Figure 2.1. Sequencing primers for RPGR ORF15 as aligned to the reference genome as viewed in IGV (Integrative Genomics Viewer). The collections of sample sequence reads are from affected RPGR patients as detected by NGS to illustrate the utility of these reactions for standard variant verification purposes also. Small blue lines in the bottom panels indicate the primers were aligned to the sense strand and the red equivalent of these represents alignment to the antisense strand.

2.19 Mammalian Cell Triple Transfection

1500 ng of ITR containing plasmid, 1500 ng AAV serotype plasmid and 3000 ng pHelper plasmid were diluted with transfection buffer to give a final volume of 100 mL. This dilution was aliquoted into 10 mL fractions in 10 universal 30 mL tubes.

One at a time, each of the 30 mL white cap tubes were labelled with the current time plus 8 minutes and 10 mL of transfection reagent was added. After 5 minutes incubation of the transfection reagent, buffer and plasmid solution, 10 x 150 mm plates were removed of previously prepared confluent HEK293 cells from the 37°C, 5% CO₂ incubator.

8 minutes following addition of the plasmid/buffer mix to the first tube of solution, 2 mL were decanted onto each of the labelled HEK293 plates. Plates were returned to the incubator. Steps were repeated for subsequent tubes (#2 - #10) until all 100 plates of HEK293 cells were transfected and returned to the incubator set at 37°C, 5% CO₂. All plates were incubated for a period of 5 hours.

After 5 hours of incubation, the plates of transfected cells were removed from the incubator. The media was aspirated and very carefully replaced with 20 mL serum free medium, this was to ensure the monolayer of cells was not disturbed. Plates were subsequently returned to the incubator and allowed to incubate for an additional 60 hours.

2.20 Recombinant AAV Harvest and Purification

This work was conducted by fellow researchers in the lab, primarily Dr. Naomi Chadderton, but has been included here to provide an overview of the preparation process.

2.20.1 Harvest

100 mL of 37°C AAV lysis buffer was prepared. In batches of ten plates, media was aspirated and collected in a sterile 2 L container and cell monolayers were collected on one side of each plate. 14 mL of 37°C AAV lysis buffer was used to resuspend and collect the total cells from ten plates. The cell suspension was then transferred to a 50 mL universal tube and incubated on ice. This process was repeated an additional nine times until all the transfected cells in the 100 plates were harvested. When the original stock of lysis buffer was depleted, supernatant from earlier harvested pools was used to resuspend and collect the remaining cells. This process created four 50 mL balanced tubes and 2 L of collected cell media. The cell media was sealed and stored at 4°C.

The four 50 mL balanced tubes were then frozen at -80°C. When completely frozen, the four tubes were thawed at 37°C. This cycle of freeze-thaw was repeated twice more for a total of three cycles. After the final thaw, the four tubes were centrifuged for 30 minutes at 3000 rpm. The supernatant from the tubes was then carefully decanted into the vessel containing the 2 L of cell harvest media under sterile conditions. 550 mL of 40% PEG 8000, 2.5 M NaCl solution was added to the cell supernatants and resuspended by several inversions. This solution was incubated at 4°C overnight.

The total solution was decanted into 6 x 500 mL polypropylene bottles and centrifuged at 2500 g and 4°C for 30 minutes. The resulting pellets were resuspended in a total of 30 mL of PBS solution. 12 µl benzonase solution was added and the solution was incubated at 37°C for 30 min. 1.5 mL of 10% (w/v) sodium deoxycholate in PBS was added to give a final concentration of 0.05 % and the solution was returned to the 37°C incubator for an additional 10 min. 4.5 mL of PBS was added, and the solution was placed on ice.

2.20.2 Caesium Chloride Gradient

40.825 g CsCl was weighed into a 50 mL tube, ddH₂O was used to bring the volume to 50 mL. This tube was labelled as 1.6 CsCl solution. 27.415 g CsCl was weighed into a 50 mL tube, ddH₂O was used to bring the volume to 50 mL. This tube was labelled as 1.4 g/mL³ CsCl solution. 16.344 g CsCl was weighed into a 50 mL tube and added to the previously prepared AAV sample. The solution was mixed thoroughly by inverting five times and the solution was returned a container of ice.

Three polyallomer centrifuge tubes were labelled with the name of the AAV preparation and date, #4 was labelled as the balance tube. Firstly, 12 mL of the 1.4 g/mL³ CsCl solution was aliquoted into each tube. 12 mL of the 1.6 g/mL³ CsCl solution was then incredibly slowly pipetted, taking care to avoid air bubbles, into the bottom of the first tube, settling below the 1.4 g/mL³ CsCl solution. The procedure was repeated for each of the four tubes.

12 mL of the AAV/CsCl solution was carefully pipetted onto the top of the first tube ensuring that the surface remained undisturbed. This procedure was repeated two more times. 5.448 g CsCl was added to 12 mL ddH₂O, mixed to dissolve and added to the

surface of the fourth, balance tube. The four tubes were transferred to SW28 buckets and centrifuged at 28,000 rpm and 4°C for 18 hours without brake.

2.20.3 Purification I

The first tube was placed in the clamp stand with a rack of Dnase/Rnase-free eppendorfs racked below. Using a 16 G needle the side of the tube was pierced approximately 2.5 cm from the bottom with a single, slightly downward motion. The first racked eppendorf was aligned with fluid flow. 1 mL aliquots were collected until fluid flow stopped. This was repeated for the other two centrifuge tubes containing AAV samples.

6 µL samples were taken from each aliquot and read on the refractometer. Readings were marked on lids. The samples which fell between 3.373 and 3.368 were identified and marked. The fractions within this range were retained in the eppendorf rack. Two polyallomer centrifuge tubes were used to pool the contents of the centrifuge tubes of interest. Where necessary, the pooled samples were balanced with 1.4 g/mL³ CsCl solution and sealed. The balanced pooled tubes were centrifuged at 59,000 rpm and 4°C for 18 hours without brake.

2.20.4 Purification II

The first tube was placed in the clamp stand with a rack of Dnase/Rnase-free eppendorfs racked below. Using a 16 G needle the side of the tube was pierced approximately 0.5 cm from the bottom with a single, slightly downward motion. An 18 G needle was used to lightly pierce the top of the tube until fluid began to flow. 0.5 mL aliquots were collected until fluid pour stopped. This process was repeated for the second pooled sample tube. Again, 6 μ L samples were taken from each aliquot and read on the refractometer. The fractions within the range of 3.373 and 3.368 were retained from each sample. One polyallomer centrifuge tube used to pool the samples of interest. A second tube of 1.4 g/mL³ CsCl solution was used as a balance. Both tubes were centrifuged at 59,000 rpm and 4°C for 18 hours without brake.

2.20.5 Purification III

The first tube was placed in the clamp stand with a rack of Dnase/Rnase-free eppendorfs racked below. Using a 16 G needle the side of the tube was pierced approximately 0.25 cm from the bottom with a single, slightly downward motion. An 18 G needle was used to lightly pierce the top of the tube until fluid began to flow. 0.5 mL aliquots were collected until fluid pour stopped. This process was repeated for the second pooled sample tube. Again, 6 μ L samples were taken from each aliquot and read on the refractometer. The fractions within the range of 3.373 and 3.368 were retained from each sample.

3 0.5 mL slide-a-lyzers were labelled, attached to a float and placed in an autoclaved 1 L beaker containing 400 mL PBS with a magnetic stirrer bar for 5 minutes to hydrate. Fraction 3.705 and the fractions either side were each then placed into a separate hydrated

slide-a-lyser using a 1 mL syringe and 18 G needle. The corner of slide-a-lyser used was marked with a black spot. Slide-a-lyzers were returned to PBS and placed on magnetic stirrer at 4°C for a minimum of 1 hour. Liquid was replaced with fresh 400 mL PBS solution and returned to 4°C for a minimum of 1 hour. This process of liquid replacement and incubation was repeated eight more times.

Using a 1 mL syringe and 18 G needle, 0.5 mL air was taken up and gently introduced to a slide-a-lyser using a clean corner. Slide-a-lyser was orientated to best collect dialysed AAV containing fraction. Fraction was transferred to a DNase/RNase free 1.5 mL eppendorf and 25 µL was decanted to a 0.6 mL RNase/DNase free eppendorf. Both were frozen at -80°C. Fraction extraction steps repeated for the other two dialysed fractions.

2.21 Library Preparation for Next Generation Sequencing of AAVs

The standard KAPA library preparation for NGS is incompatible with AAV due to the single stranded nature of AAV DNA. Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences, Michigan, United States) was used to prepare the AAV libraries. This was used with the 1S Plus Dual Indexing Kit, as standard dual-index adapters were incompatible with the library kit due to the special truncated adapters required during the protocol.

24 AAVs of various compositions (serotype, promoter and insert length) were diluted (4 µL purified virus and 11 µL H₂O) treated to remove exogenous DNA from the viral capsid coats (Tables 2.5 and 2.6) and the treated samples were then incubated with protease to release the viral ssDNA (Tables 2.7 and 2.8). DNA was then sheared with sonication using 2 cycles of 30 seconds sonication with 30 seconds off-period between cycles.

Library preparation was performed as per manufacturer's recommendations. Briefly, samples were denatured by heat to enable better performance of the Adaptase enzyme which performs tailing and ligation of truncated adapters to the 3' ends of fragmented DNA. Following this, an extension step is used to form double stranded DNA and ensure incorporation of truncated adapters by a primer extension reaction. Ligation is then used to add the second truncated adapter to the 5' end of original strand of template ssDNA. An Indexing PCR step is then used to incorporate full length adapters to both ends of the target DNA. DNA is then purified and quantified. Samples were sequenced using an Illumina 2500 platform with 2 x 100 bp pair-ended sequencing.

Table 2.5. *Components of DNase ssDNA reactions for NGS library preparation.*

COMPONENT	50 μ l REACTION
Diluted Virus	15
DNase Enzyme	1
DNase Buffer	5
H2O	29

Table 2.6. *Thermocycler conditions of DNase ssDNA reactions for NGS library preparation.*

STEP	TEMP	TIME
Enzymatic Step	37°C	30 minutes
Enzyme Deactivation	70°C	10 minutes
Hold	4-10°C	

Table 2.7. *Components of Proteinase ssDNA reactions for NGS library preparation.*

COMPONENT	51 μ l REACTION
Proteinase K	1
DNase Reaction	50

Table 2.8. *Thermocycler conditions of Proteinase ssDNA reactions for NGS library preparation.*

STEP	TEMP	TIME
Enzymatic Step	50°C	60 minutes
Enzyme Deactivation	95°C	20 minutes
Hold	4-10°C	

3 Variant Interpretation and Annotation

3.1 Introduction

NGS experiments produce extraordinary amounts of data, the processing and analysis of which will determine its interpretability. This chapter describes the general workflow of handling and manipulating raw sequencing data to ultimately calling candidate variants for a given condition. The following section will provide an overview of data acquisition, quality control, sequence filtering and the implementation of predictive models and guidelines for interpretation of variants.

3.2 Data Analysis

Illumina platforms were exclusively used for the target capture libraries in the Target 5000 project. Analysis of the data acquired from NGS platforms is computationally intensive to process. In Illumina platforms, the data are first acquired by imaging the tiles on the flow cell to detect nucleotide binding in the relevant clusters. These data are then interpreted as nucleotides and added to the growing string of data produced from that cluster. The final product of this process is a string of nucleotides if single read cycles are used, or two strings of nucleotide data from opposite ends if paired-end methods are employed.

These data can now be thought of as something akin to a FASTA format text file output (106). However, the data is not seen to be infallible, so information regarding the quality of base calls is also attached, which can be used to determine the confidence at which that

base was correctly called. This type of file format is called FASTQ, due to its similar nature to FASTA, but with the additional quality score information attached (107). This process occurs entirely within the sequencing machinery and as such the raw data used to compose the output file are not readily acquired. It is possible to view the cluster imaging data (Figure 3.1) and the quality control data acquired throughout the process (Figure 3.2).

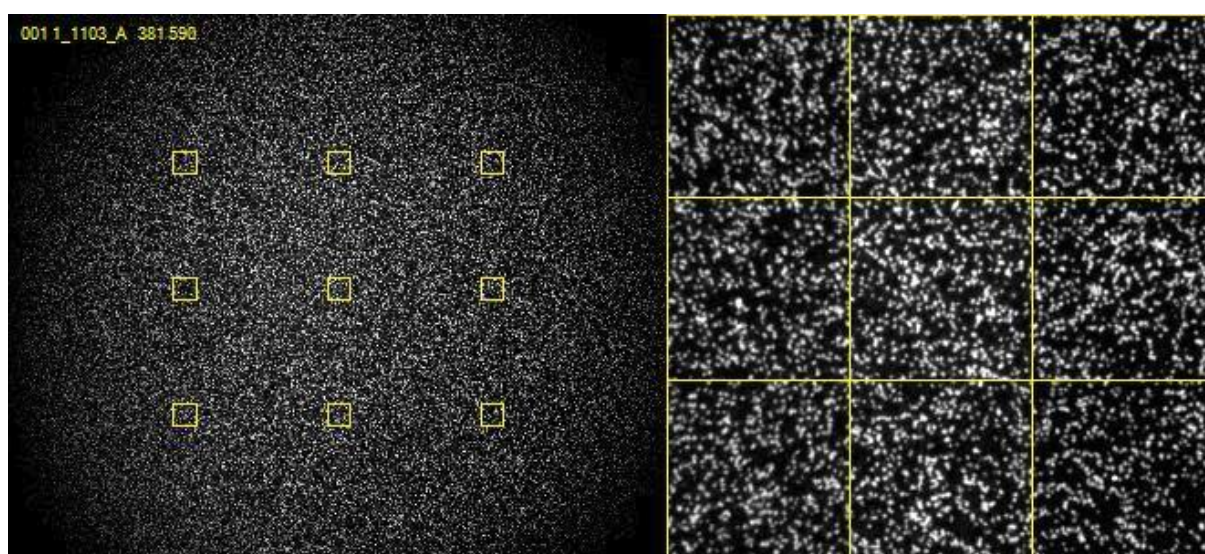


Figure 3.1. Image of cluster analysis that occurs on the Illumina MiSeq Platform. In this image, all fragments with an adenine addition are illuminated. Image acquired as part of the Target 5000 study.

Indexing QC

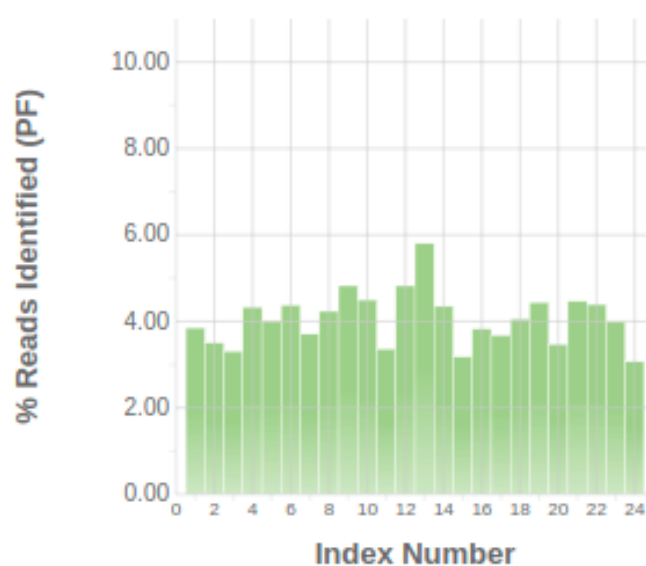


Figure 3.2. Example of Illumina MiSeq in-house quality control. Each indexed sample accounts for the data of one Target 5000 patient. This graph illustrates the approximate equal sequencing capacity allocated to each sample by virtue of accurate sample pooling prior to target capture enrichment, with the sum of all samples equating to 100%.

3.3 Read Alignment

At this point, the FASTQ files from the sequencing platform have been downloaded and are ready for analysis. The reads will appear to be random fragments of DNA without order or usable information in terms of variation. Therefore, the first thing to be undertaken is to map the reads to the relevant reference sequence or genome. In the case of Target 5000, human reference genome “hg19” (GRCh37) (108) was initially utilised and this was then subsequently replaced by the more current human reference genome “hg38” (GRCh38) (109). All data originating from sequencing runs prior to the switch were remapped to the newer genome.

This is arguably the most computationally intensive task in sequence analysis as the reads will contain no information regarding their appropriate position in the genome. To tackle this issue, many bioinformatic tools have been developed to assist with this task. Burrows-Wheeler Alignment (BWA) (110) was the alignment tool of choice for Target 5000. BWA uses the data from both paired ends to align against the reference sequence. This step emphasises the importance of size selection of products to be sequenced. If a sequencing kit capable of 2 x 100 bp reads is used, then the desired product length would be between 300 to 400 bp in length. This allows for a more tactical, broader capture of the desired region as fragments bound by capture probes in the exonic region will also capture a small fraction of the exon-intron boundaries. Alternatively, if the products to be sequenced are smaller than the total of the paired region (i.e. <200 bp) then the result will be a smaller region captured with higher depth.

3.4 Duplicate Read Filtering

Duplicate reads occur due to intrinsic errors in library preparation. In a sample with a diverse sequence there should be little or no enrichment of duplicate reads. However, this will still occur if the sequenced sample is of low-diversity or contains lots of homologous regions. Despite this, when amplification is incorporated into a library preparation protocol, amplification bias is more likely to occur and logically the bias will increase with additional rounds of amplification required. Such is the case when very limited sample input material is used, for example during single cell sequencing.

To circumnavigate this bias, Picard, a collective term for a diverse package of useful bioinformatic tools, was used to remove reads that were deemed to be duplicates (111). This tool is capable of receiving arguments that enables the reading of the output files from BWA's alignment files (.bam) and rewrites the files with the duplicate reads removed (.dedup.bam).

3.5 Local Realignment

FreeBayes is a genetic variant detector that uses Bayesian statistics to find deviations in sequence data from the reference sequence (112). FreeBayes uses short read alignments for a multiple number of samples and a reference genome to determine variation such as SNPs, indels and MNPs (multiple-nucleotide polymorphisms/polygenic sites). This use of short haplotype alignments is superior for variant detection at multiallelic loci. This

method is particularly useful for well-described genomic regions as it can account for known variants, single nucleotide variants as well as copy number variants, as prior information. This batch analysis method was deemed to resolve far more false indel calls than other local indel realignment tools such as GATK (113). GATK uses precise read alignment that fits the sample read data to the reference sequence with the least number of mismatches and fewest sequence gaps.

3.6 Sequence Data Quality Control

FastQC (114) was used to analyse various output statistics to ensure that the sequence data obtained was of high quality. Useful output measures from this tool include per base sequence quality (Figure 3.3a and 3.3b), per sequence quality score (Figure 3.4a and 3.4b), per sequence GC content, sequence length distribution and presence of duplicate sequences. Other useful metrics are also available from the FastQC output files.

Per base sequence quality is illustrated as a Box Whisker plot for each sample. Each position is annotated with the median value, inter-quartile range, the lower (10%) and upper (90%) data ranges, and the mean quality. The colour-coded nature of this output allows for rapid assessment of sequencing issues, from good quality (green) to poor quality (red). It is commonly accepted that read quality will deteriorate as read length increases (Figure 3.3a). It is also important to note that if the quality of a sequencing run drops below an acceptable threshold after a given number of bases, the sequencing data can be bioinformatically parsed to remove the sub-standard data. This process is called trimming, AfterQC (115) is just one of many programs optimised to perform this and

many other filtering tasks. Figure 3.3b typifies the high-quality data that is obtained as part of the Target 5000 study. Phred scores are used to determine data quality. Phred scores operate on a logarithmic scale such that a Phred score of 10 would imply the probability of 1 in 10 chance that a base was called incorrectly (90%). The maximum score for sequencing platforms is generally a Phred score of 30, implying a 1 in 1000 chance that a base has been miscalled (99.9%).

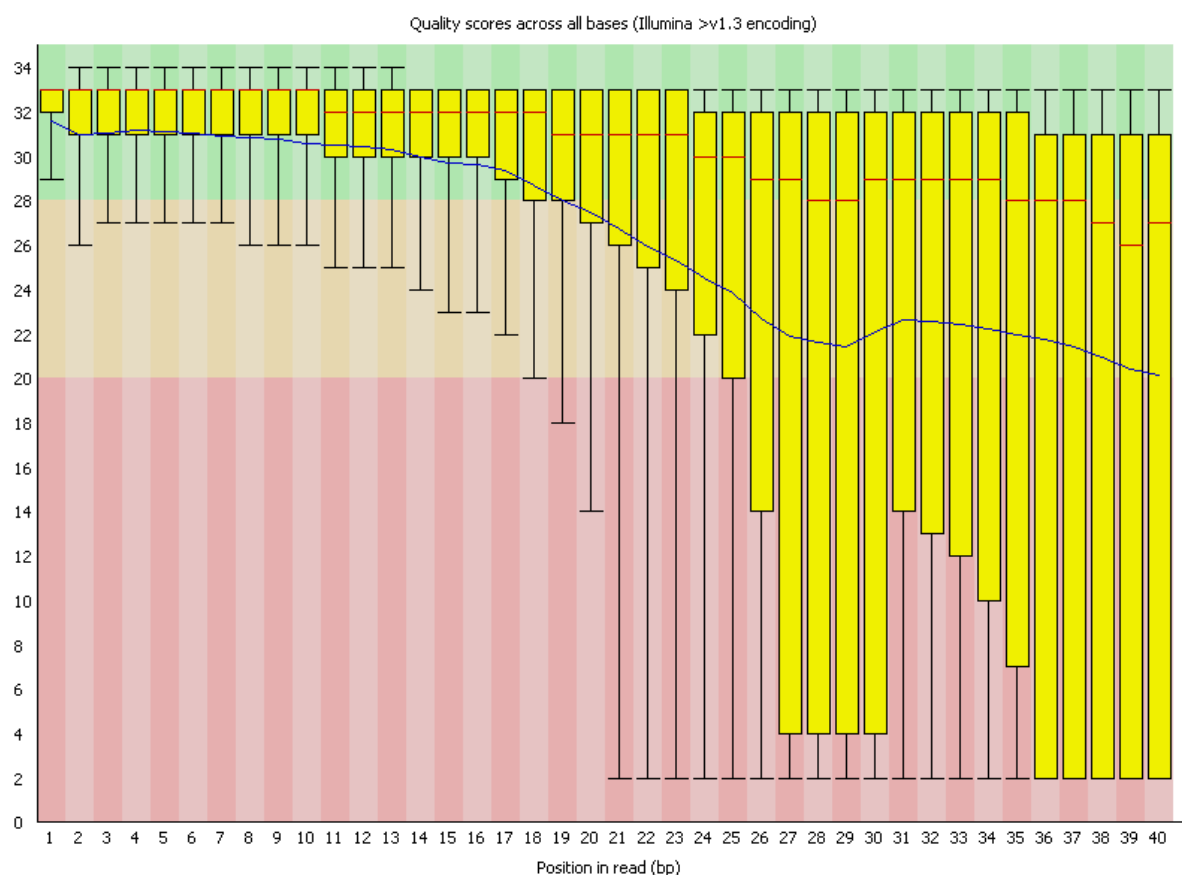


Figure 3.3a. Sample data available from FastQC (114) that illustrates typical per base sequence quality scores. Quality scores are listed on the Y-axis and base position is listed on the X-axis. Quality scores are grouped colour metrically from good (green) to poor (red). For each base the median value is illustrated by a straight horizontal red line, while the mean value is illustrated by the continuous blue line. The inter-quartile range (25-75%) is depicted by a yellow box. The upper and lower whisker bars illustrate the 10% and 90% points.

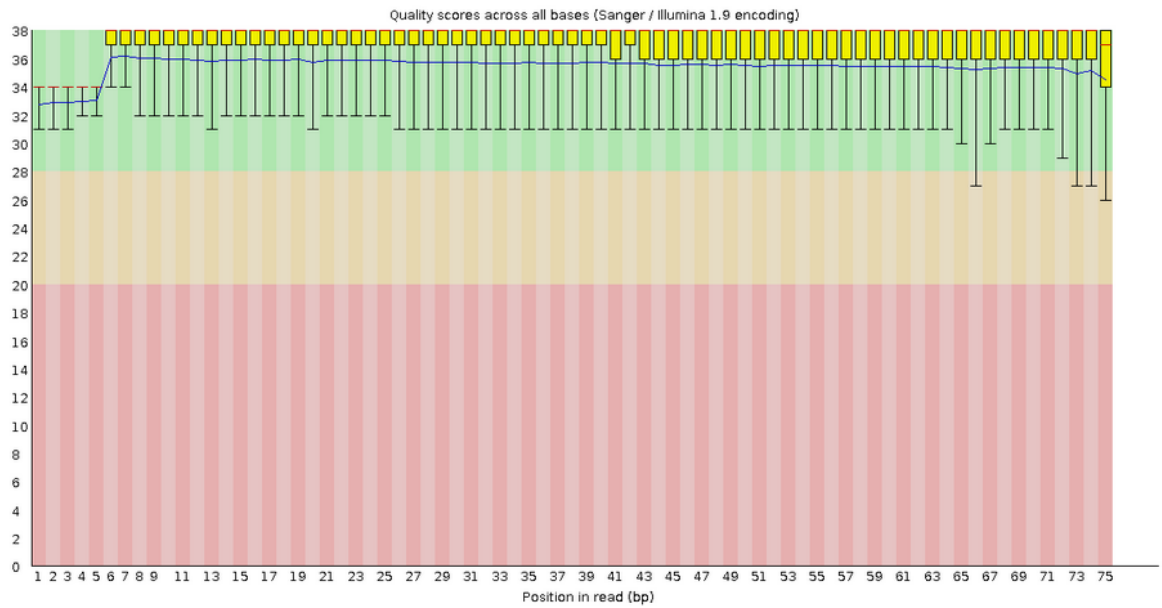


Figure 3.3b. Target 5000 sample data (patient D47) depicting typical per base sequence quality scores. Quality scores are listed on the Y-axis and base position is listed on the X-axis. Quality scores are grouped colour metrically from good (green) to poor (red). For each base the median value is illustrated by a straight horizontal red line, while the mean value is illustrated by the continuous blue line. The inter-quartile range (25-75%) is depicted by a yellow box. The upper and lower whisker bars illustrate the 10% and 90% points.

Per sequence quality score can provide an overall verification of a high-quality run. Where a per sequence quality score is bi-modal (Figure 3.4a), this can be an indication that a portion of the sequencing flow cell performed sub-optimally. This is known to occur when an air bubble becomes trapped in the flow cell prior to sample loading. This issue can be evaluated with the per-tile quality scores (if available) to assess whether a section of the flow cell under performed. Another reason for receiving an underwhelming quality score may be related to the issue of diminishing returns per base sequenced as can be seen above (Figure 3.3a). High quality data should contain very few reads with a score of 20 or less (Figure 3.3b).

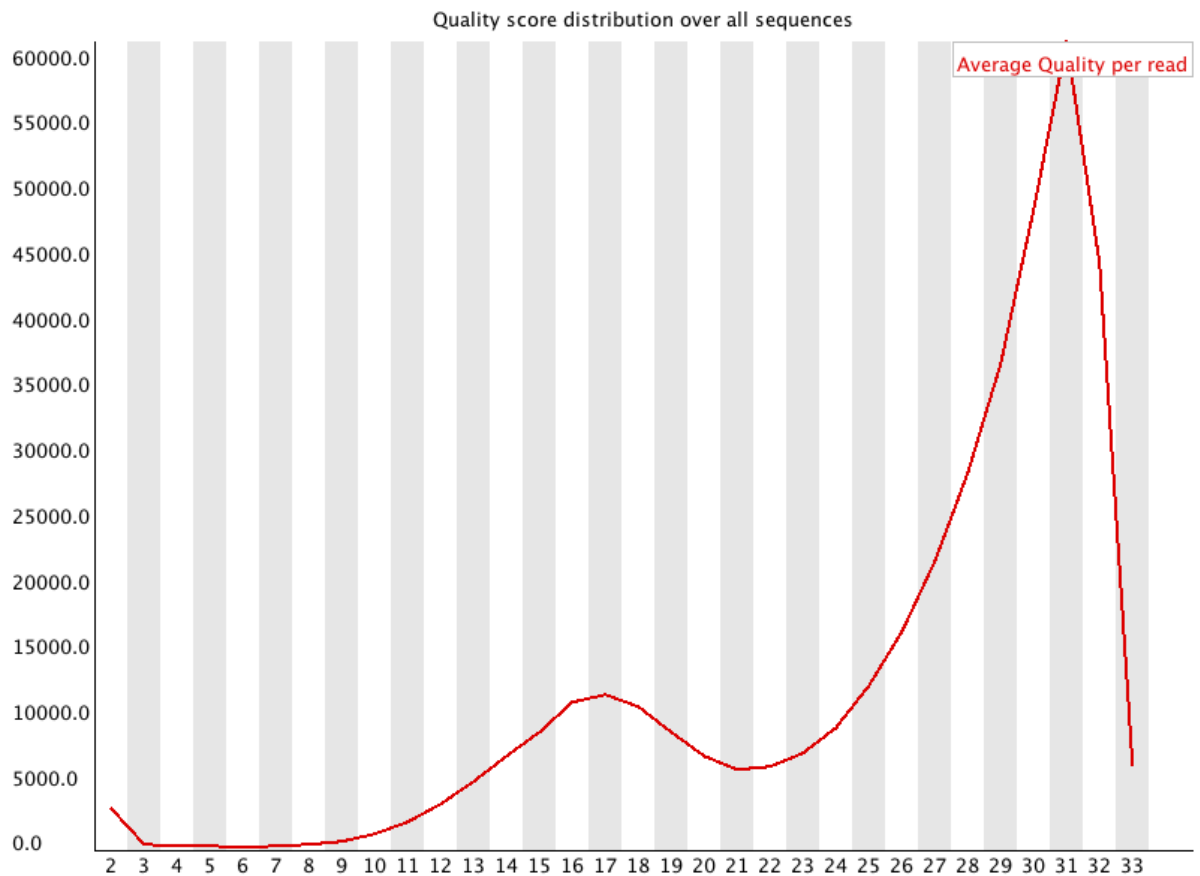


Figure 3.4a. Sample data available from FastQC (114) that illustrates a bi-modal graph of per sequence quality scores. The Y-axis depicts the number of reads while the X-axis lists the quality scores. This graphical pattern is typical of a sequencing run that contains both high and poor-quality data.

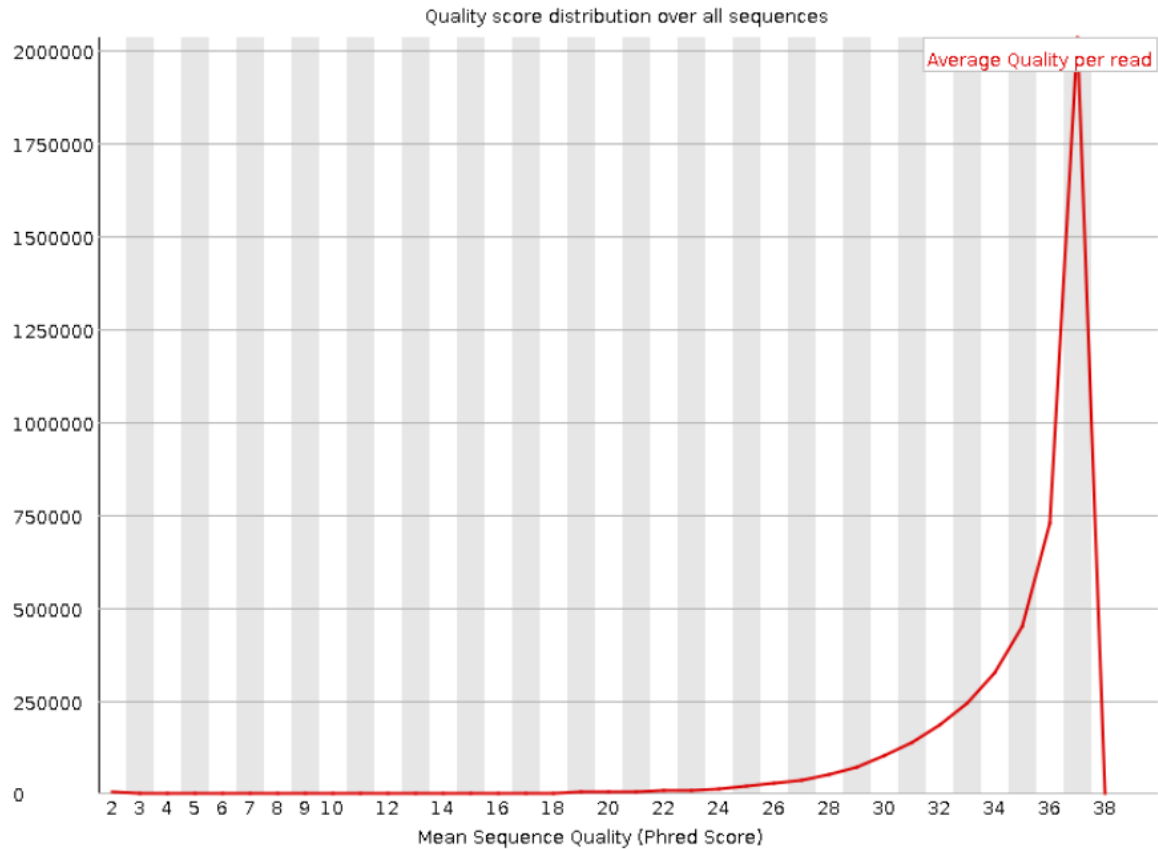


Figure 3.4b. Target 5000 patient data (patient X2) as output by FastQC (114) that illustrates a single peak graph of per sequence quality scores. The Y-axis depicts the number of reads while the X-axis lists the quality scores. This graphical pattern is typical of a sequencing run that solely contains high quality data.

3.7 Sequence Tidying

Now the data has been aligned, filtered and checked for errors. The next step is to ensure the sequence data is compliant with the field's recommendations. The output of the FreeBayes process generates Variant Call Format (.vcf) files. These variants called can then be filtered by read depth and quality. This ensures that filtered variants are called with confidence and are not likely to be present due to sequencing errors or contamination. Subsequently, true variants can now be normalised. This process ensures that all indels detected are left-aligned as per the "3' Rule" established in the Sequence Variant Nomenclature Guidelines (116).

This rule ensures that where multiple positions in the reference sequence could be inferred as changed, reports of such changes are consistent across multiple sources. For example, if the reference sequence contains "CAAAG" and the sequence data obtained reveals a single "A" deletion, the sequence will read "CAAG". Under the Human Genome Variation Society's 3' Rule, the most 3' position possible of the reference sequence is to be arbitrarily assigned to be the source of variation. This prevents multiple variant reports from being generated with alternative base substitutions when the actual base alternation cannot be determined.

3.8 Variant Annotation

Now that high-confidence, correctly positioned variants have been established they can now be annotated. The level of annotation required is entirely dependent on the end user's preference. Nonetheless, there are some annotations that are universally useful. GATK's VariantAnnotator tool will attach dbSNP identifiers to variants. Although SNP is the

abbreviation for single nucleotide polymorphism, dbSNP is a database of all short sequence variation (117).

These identifiers contain links to a variant's genotype and allele frequency in a given population. Depending upon the variant's records, it may also include information regarding its associated records of clinical significance (ClinVar) (118). If the variant has been significantly associated with a phenotype from previous GWAS reports, it may also have a dbGaP identifier (119). Afterwards, SnpEFF can be used to attach details of affected overlapping genes and their relevant protein changes if any (120). SnpSift (121) can be used to attach various additional lines of detail such as specific population frequencies. It can also link variants to another database, dbNSFP, if desired (122).

3.9 Pathogenicity Prediction Models for Nonsynonymous Variants

When processing NGS data, the discovery of a massive number of novel variants is guaranteed. The pathogenic potential of these variants must therefore be measured by a quantifiable method to be able to interpret the significance of these novel mutations in a time efficient manner. This computational automation of analysis was initiated primarily by two dominant programs ten years ago, PolyPhen-2 and SIFT (123,124). Briefly, PolyPhen-2 assessed variants based on their likelihood to affect protein structure based on amino acid substitutions, while SIFT scored amino acid substitutions based on the conservation levels of the original amino acid throughout evolution. Although these programs are still maintained and upgraded (125), some studies have shown the limited value of these methods (126,127).

In more recent years, variant predictors have moved towards ensemble models to predict functional impact of variants. These models typically incorporate several programs, such as PolyPhen-2 and SIFT and many others, to contribute to an overall metric. In 2015, a study was conducted that assessed the discriminatory prowess of 18 different methods of scoring deleterious mutations (128). 11 of these programs were based on functional prediction, 3 were based on conservation scores and 4 were ensemble methods. Their findings showed that ensemble methods that incorporated methods that scored specific protein features not only outperformed other ensemble methods but also performed better than any of their component methods (128).

Having gathered all the data for the best performing individual programs and ensemble tools, the authors used this information to develop their own ensemble method, superior

to any of the tools they had tested. This led to the creation of MetaLR (128). This tool also considers speed of analysis as several other predictive tools were criticised for their limited processing ability with few noted exceptions such as SIFT, Polyphen-2 and MutationTaster who each are capable of batch queries and fast processing. MetaLR and dbNSFP both have integrated scores from their component and associated programs so that values can be inserted rather than calculated, thus saving large quantities of time and computational effort. For these reasons, MetaLR has been our pathogenicity predictor of choice for Target 5000 since its release.

The Mendelian Clinically Applicable Pathogenicity (M-CAP) scoring program was the second ensemble predictor to be implemented in the bioinformatic pipeline for Target 5000 (129). M-CAP works in a very similar manner to MetaLR but with much tighter constraints on filtering. Effectively, M-CAP scores are consulted in our pipeline when there is a variant of unknown significance that is only moderately likely to be pathogenic as determined by MetaLR. M-CAP aims to misclassify no more than 5% of pathogenic variants by aggressively filtering variants it considers benign. The aim of this tool is to readily determine benign variants with high confidence. This does so by using amino acid conservation calculations with a novel gradient boosting method acquired from Mendelian disease related training data sets. M-CAP has been shown to significantly reduce the misclassification of rare variants by five-fold compared to MetaLR (129). Although M-CAP has been shown to be more sensitive than MetaLR, MetaLR's scoring methods also takes into account specific protein function as a metric in its analysis and thus still serves as an important computational tool to rely on.

REVEL (130) is the third and last ensemble predictive tool to be incorporated into our analysis pipeline for scoring nonsynonymous protein coding substitutions. REVEL is composed of many of the same constitutive programs as MetaLR, however, unlike MetaLR, REVEL has been trained on entirely different datasets to those of its component tools. The outcome of this training is marginally superior true positive detection rates compared to MetaLR. REVEL claims to have superior detection rates for rare variants with a population frequency under 5%. By utilising MetaLR, M-CAP and REVEL together, this takes advantage of the three most successful ensemble programs for detecting pathogenic variants by acknowledging their different methods, stringencies and training.

3.10 Pathogenicity Prediction Models for Splice Variants

Regulatory *cis*-elements account for a large proportion of what we know to be conserved mammalian sequences in our genome (131). Non-coding variants and splice mutations have the potential to not only affect expression levels of the encompassing gene, but also have the capacity to disrupt correct intron splicing and form aberrant proteins. Several intronic variants have already been proven to be causative in retinal degenerations (95,132). In certain cases, where the pathogenic intronic variant can be successfully detected, a treatment may be developed to correct the disruption it causes (133).

For these reasons, the detection of pathogenic splice variants is as important as detecting nonsynonymous protein coding variants. For this task two bioinformatic tools have been incorporated into the analysis to predict the pathogenicity of novel splice variants. Currently, our analysis includes the use of databases with precalculated scores of

pathogenicity and a bioinformatic tool that predicts the pathogenicity of novel splice variants.

dbNSFP is a comprehensive database that records scores from several pathogenicity predicting tools for reported non-synonymous and splice site single nucleotide variants (nsSNVs, ssSNVs). As of 2016, the database contained nearly 83 million nsSNVs and ssSNVs (122). The database also incorporates an additional data source, dbSCSNV (134) which has generated the predicted deleterious effect of 15 million theoretical ssSNVs that could potentially occur in known splice junctions. Combined, this accounts for the approximate 10 intronic nucleotide positions either side of a given exon (135).

SPIDEX is a tool that was created to analyse any variant, intronic or exonic for its potential to impact splicing (136). An analysis of human variant data from over 10,000 exons known to exhibit erroneous splicing was conducted; over 1,300 sequence features were determined to affect accurate splicing, 300 of which were novel. The SPIDEX code has shown that reported pathogenic single nucleotide changes in the intronic space that are more than 30 bases away from a splice site are scored 9 times more likely to affect splicing than common variants within the same region. However, to the best of this author's knowledge, there has been very little functional analysis to verify these findings.

3.11 Detection of Structural Variants

In addition to standard variant detection, structural variants were called by a separate pipeline, based on the tools LUMPY (137), which was used to identify structural variants via split reads or unusual paired read alignments, and CoNIFER (138), which was used

to identify structural variants based on aberrant read depths. Results from these tools were combined and filtered to output a final, high-quality list of putative structural variants and associated confidence scores for each sample.

3.12 Selecting Variants of Interest

At the beginning of the Target 5000 project candidate variants were chosen based on a logical set of rules. These rules ensured that there was ample evidence present to suggest that the variant(s) of interest selected had the potential to be causative for the observed condition.

Population Frequency

When dealing with rare Mendelian diseases, a population frequency of greater than 5% is highly unlikely. To assess if this is the case, each variant was searched for in several global databases of population frequencies. Traditionally, ExAc (the Exome Aggregation Consortium) (139) and the 1000 Genomes Project (140) were key references for this. More recently all these data and much more have been harvested into a meta-database called gnomAD (the Genome Aggregation Database) (141). This resource currently contains data from 123,136 exome sequences and 15,496 whole genome sequences. Throughout the Target 5000 project, these databases have been probed for the populations with the highest frequencies containing the variant of interest. This has helped to filter out many polymorphisms but has failed to detect any populations with the probed variant on many occasions, most likely due to the rare nature of the conditions and their associated causative variants.

Sample Frequency

Much like population frequency, sample data from the project can reveal population specific variants. This is particularly useful when dealing with a population (for example, Ireland) that has not been well sampled in previous global population studies. This means that variants that appear to be rare or not previously detected in population databases may be high frequency variants in the sample cohort. These data were initially uninformative due to the low sample size, but as this project has grown and continues to expand this will undoubtedly become an increasingly valuable resource. Currently, the Target 5000 database enables users to view all patients who possess a queried variant. This allows for the exclusion of candidature of that variant where found in unaffected individuals if associated with dominant inheritance or if found homozygously in unaffected individuals when associated with recessive inheritance patterns.

Computational Evidence of Pathogenicity

The bioinformatic tools used for this process have already been outlined. It is important to note that each program comes with recommended thresholds and scores to utilise when interpreting their respective values. When evaluating a novel variant, there had to be some computational evidence of pathogenicity before it could be considered an appropriate candidate variant of causation.

Phenotype Correlation

The variant(s) of interest had to be recognised as affecting a gene that had been previously associated with the condition reported in the clinical assessment or a similar clinical condition. Two primary databases were used to explore this. RetNet, a collection of gene symbols and the associated studies that link them to the observed retinal phenotypes (1).

OMIM, a database of human Mendelian conditions (142). The novel variant(s) also had to segregate appropriately with the phenotype where there was already an established genotype-phenotype association. New phenotype associations were considered only when there was sufficient segregation and biological relevance to suggest it was plausible.

Impact on Functional Domains

Nonsynonymous variant predictors reliance on conservation statistics have already been discussed (Section 3.9). Furthermore, databases of pathogenic variants such as ClinVar (118) and UniProt's Humsavar (143) were consulted to determine if the candidate variant was positioned in a mutational hotspot as this may be an indication that a key protein domain, essential for proper function, has been disrupted. Frameshift mutations and missense variants that resulted in premature stop codons were considered to disrupt any domain in which they were located or when found to precede a domain unless located at the extreme 3' terminus of a protein coding sequence.

3.13 Guidelines for Interpretation of Sequence Variants

In late 2015, The American College of Medical Genetics and Genomics (ACMG), The Association for Molecular Pathology (AMP) and The College of American Pathologists (CAP) formed a panel of experts to outline guidelines for interpreting the pathogenicity of sequence variants relating to Mendelian disorders (144). The goal of the panel was to formulate criteria such that a variant of unknown significance could be classified as "pathogenic", "likely pathogenic", "uncertain significance", "likely benign" or "benign". These comprehensive guidelines replaced the earlier reasonings for classification in Target 5000.

The lines of evidence used to classify variants as “pathogenic” or “likely pathogenic” are shown in Tables 3.1 and 3.2. Evidence that would suggest a variant is “benign” or “likely benign” is shown in Table 3.3. These guidelines point towards categories of evidence that may be available through various sources. These sources include population data (BA1, PM2, PS4), computational and predictive data (BP4, BP1, PP3, PM4, PM5, PS1, PVS1), functional data (BS3, PP2, PM1, PS3), segregation data (BS4, PP1), *de novo* data (PM6, PS2), allelic data (BP2, PM3) and other data (BP5, PP4). Caveats and notable pitfalls of utilising such evidence is also noted with the relevant points in Tables 3.1-3.3.

Table 3.4 outlines how the recommended combinations of lines of evidence can be used to determine a classification ranging from “pathogenic” to “benign”. Where there is notable conflict between evidence for and against the pathogenicity of a variant, the classification of “unknown significance” is recommended. These are guidelines, and although it is highly encouraged for clinical geneticists and associate team members to follow these as an industry standard, it has been shown that there was discordance between several institutes that reported on the same variant set. A year after the guidelines were introduced, a study was conducted across multiple sites using a set of nearly 100 variants spanning all classifications. After review there was a near 30% discordance rate between sites when classifying variants (145). This shows there is inherent variability when interpreting how this evidence can be implemented. However, it is nevertheless inarguably a useful resource of potential lines of evidence that may otherwise be overlooked.

Interestingly, the same study (145) also recorded the frequency at which each line of evidence was used when determining the classification of the near 100 variants. The four

most frequent codes used were PM2, PP3, BS1 and PVS1. All of these lines of evidence signify intuitive uses of population, computational and functional data that were likely implemented prior to any formal guidelines, much like the guiding principles of variant calling used in Target 5000 prior to the ACMG recommendations.

Table 3.1. ACMG Guidelines for using very strong and strong evidence for interpreting pathogenicity. Modified from (144).

Very strong evidence of pathogenicity

PVS1 Null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease.

Caveats:

- Beware of genes where LOF is not a known disease mechanism
- Use caution interpreting LOF variants at the extreme 3' end of a gene
- Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact
- Use caution in the presence of multiple transcripts

Strong evidence of pathogenicity

PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change

Example: Val->Leu caused by either G>C or G>T in the same codon

Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level

PS2 *De novo* (both maternity and paternity confirmed) in a patient with the disease and no family history

Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, *etc.* can contribute to non-maternity

PS3 Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product

Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well-established

PS4 The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls

Note 1: Relative risk (RR) or odds ratio (OR), as obtained from case-control studies, is >5.0 and the confidence interval around the estimate of RR or OR does not include 1.0. See manuscript for detailed guidance.

Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in

multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.

Table 3.2. ACMG Guidelines for using moderate and supporting evidence.
Modified from (144).

Moderate evidence of pathogenicity

- PM1 Located in a mutational hot spot and/or critical and well-established functional domain (*e.g.* active site of an enzyme) without benign variation
- PM2 Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes or ExAC
Caveat: Population data for indels may be poorly called by next generation sequencing
- PM3 For recessive disorders, detected in *trans* with a pathogenic variant
Note: This requires testing of parents (or offspring) to determine phase
- PM4 Protein length changes due to in-frame deletions/insertions in a non-repeat region or stop-loss variants
- PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before
Example: Arg156His is pathogenic; now you observe Arg156Cys
Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level
- PM6 Assumed *de novo*, but without confirmation of paternity and maternity

Supporting evidence of pathogenicity

- PP1 Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease
Note: May be used as stronger evidence with increasing segregation data
- PP2 Missense variant in a gene that has a low rate of benign missense variation and where missense variants are a common mechanism of disease
- PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc)
Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.
- PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic aetiology

PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

Table 3.3. ACMG Guidelines for using evidence for non-causation.

Modified from (144).

Stand-Alone evidence of lack of impact

BA1 Allele frequency is above 5% in Exome Sequencing Project, 1000 Genomes, or ExAC

Strong evidence of lack of impact

BS1 Allele frequency is greater than expected for disorder

BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder with full penetrance expected at an early age

BS3 Well-established *in vitro* or *in vivo* functional studies shows no damaging effect on protein function or splicing

BS4 Lack of segregation in affected members of a family

Caveat: The presence of phenocopies for common phenotypes (*i.e.* cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.

Supporting evidence of non-causation

BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease

BP2 Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder; or observed in *cis* with a pathogenic variant in any inheritance pattern

BP3 In-frame deletions/insertions in a repetitive region without a known function

BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc)

Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion.

BP4 can be used only once in any evaluation of a variant

BP5 Variant found in a case with an alternate molecular basis for disease

BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation

BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site and the nucleotide is not highly conserved

Table 3.4. *Rules for combining criteria to classify sequence variants*

Pathogenic

- (i) 1 Very strong (PVS1) AND
 - (a) ≥ 1 Strong (PS1–PS4)
 - OR
 - (b) ≥ 2 Moderate (PM1–PM6)
 - OR
 - (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5)
 - OR
 - (d) ≥ 2 Supporting (PP1–PP5)
- (ii) ≥ 2 Strong (PS1–PS4) OR
- (iii) 1 Strong (PS1–PS4) AND
 - (a) ≥ 3 Moderate (PM1–PM6)
 - OR
 - (b) 2 Moderate (PM1–PM6) AND ≥ 2 Supporting (PP1–PP5)
 - OR
 - (c) 1 Moderate (PM1–PM6) AND ≥ 4 supporting (PP1–PP5)

Likely pathogenic

- (i) 1 Very strong (PVS1) AND 1 moderate (PM1–PM6) OR
- (ii) 1 Strong (PS1–PS4) AND 1–2 moderate (PM1–PM6) OR
- (iii) 1 Strong (PS1–PS4) AND ≥ 2 supporting (PP1–PP5) OR
- (iv) ≥ 3 Moderate (PM1–PM6) OR
- (v) 2 Moderate (PM1–PM6) AND ≥ 2 supporting (PP1–PP5) OR
- (vi) 1 Moderate (PM1–PM6) AND ≥ 4 supporting (PP1–PP5)

Benign

- (i) 1 Stand-alone (BA1) OR
- (ii) ≥ 2 Strong (BS1–BS4)

Likely benign

- (i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) OR
- (ii) ≥ 2 Supporting (BP1–BP7)

Uncertain significance

- (i) Other criteria shown above are not met OR
- (ii) the criteria for benign and pathogenic are contradictory

3.14 Objectives of Target 5000

The Target 5000 study aims to provide genetic testing for the estimated 5,000 people in Ireland who are affected by a genetic retinal condition. Prior to Target 5000, most IRD patients in Ireland obtained only a clinical diagnosis. This study aims to work closely with the clinical teams involved in the project to refine these diagnoses and to also deliver accurate genetic diagnostics. While clinical trials are in progress for patients with inherited retinal degenerations (IRDs), many such trials require patients to have a known causative mutation to participate in these trials. Target 5000 aims to provide this where possible.

The IRD patient cohort used in the study has been obtained via a collaborative network of ophthalmologists whereby if an IRD is suspected, given consent, a DNA sample is taken and provided to a central laboratory for genetic analysis. The study seeks to detect previously identified, together with as yet undiscovered, pathological mutations in a panel of known retinal degeneration genes utilising target capture next generation sequencing (NGS). Where a candidate variant is identified by NGS, additional family members are recruited to verify presence and segregation of the variant with the relevant condition. The study to date includes data from roughly 1000 IRD patients.

4 Results Chapters

4.1 Summary of Target 5000 Findings

4.1.1 Introduction

There are an estimated 5000 people in Ireland who are affected by an inherited retinal degeneration (IRD). It is the goal of this study, through genetic diagnosis, to better enable these 5000 individuals to obtain a clearer understanding of their condition and improved access to potentially applicable therapies. In this chapter of the thesis the current findings from a target capture next-generation sequencing study of approximately 1000 patients IRD from over 620 pedigrees currently situated in Ireland is shown. IRD patients were sequenced for over 200 IRD genes and candidate disease causing variants identified.

In addition, we demonstrate how processes can be implemented to retrospectively analyse patient datasets for the detection of structural variants in previously obtained sequencing reads. Pathogenic or likely pathogenic mutations were detected in 72% of IRD pedigrees tested. In this chapter we report nearly 30 novel mutations including three large structural variants. The population statistics related to our findings are presented by condition and credited to their respective candidate gene mutations. Re-diagnosis rates of clinical phenotypes after genotyping are discussed. Possible causes of failure to detect a candidate mutation are evaluated. The inclusion of future additional aspects to this project, with a specific emphasis on analysis of structural variants and non-coding pathogenic variants, are expected to increase detection rates further and thereby produce an even more comprehensive representation of the genetic landscape of IRDs in Ireland.

4.1.2 Methods

Applicable Sections in the General Methods Chapter

Details of patient recruitment and assessment can be found in Chapters 2.1-2.3. Sample collection and preparation is outlined in Chapters 2.4-2.6, 2.8-2.9 and 2.16. Variant verification is described in Chapter 2.17.

Chapter-specific Methods

Gyrate Atrophy Inversion Confirmation: Four primers were designed to cover the two breakpoints of the homozygous inversion in the *OAT* gene, located at NC_000010.11:g.124405527 and NC_000010.11:g.124422152. Primers OAT-1 (forward: 5' -GGTAACCTGGATCCGGAACA-3') and OAT-2 (reverse: 5' -CTTCTGGGGTAGGACG TTGT-3') covered the breakpoint in exon 5 of *OAT* and primers OAT-3 (forward: 5' -GCGAGGGGTT TCACATCATC-3') and OAT-4 (reverse: 5' -GTTGGTGTTTCTCTGGCCTG-3') targeted the breakpoint upstream of the *OAT* gene. In the *OAT* gene of unaffected individuals, the pairing of primers OAT-1 and OAT-2 would form a product, as would the pairings of primers OAT-3 and OAT-4.

4.1.3 Results

Presentation and Solve Rates

Probands, in addition to additional family members where available, were sampled from 623 pedigrees in Ireland. The initial criteria for inclusion in the study included being over 18 years of age and a full clinical examination that implicated an IRD as a likely cause of visual impairment. The ethical approval from one of the clinical sites, Royal Victoria Eye and Ear Hospital (Dublin, Ireland) has since been amended to include the recruitment and testing of paediatric samples.

A chart depicting the frequency of each condition at clinical presentation as of the end of 2017 is represented in Figure 4.1.1. A current list of genes included as part of the targeted sequencing has been provided in the Figure 4.1.2. A pathogenic or likely pathogenic variant has been detected in over 72% of pedigrees analysed. In addition, in approximately 9% of cases one single candidate mutation could be detected and the gene in question has previously been associated with a recessively inherited IRD; it is likely that at least some of these patients will carry a second mutation in this gene that was not detected, such as a deep intronic or structural variant (95). This information is further depicted in Figure 4.1.3. A large number of likely pathogenic and novel mutations have also been found as part of this study (Table 4.1.1).

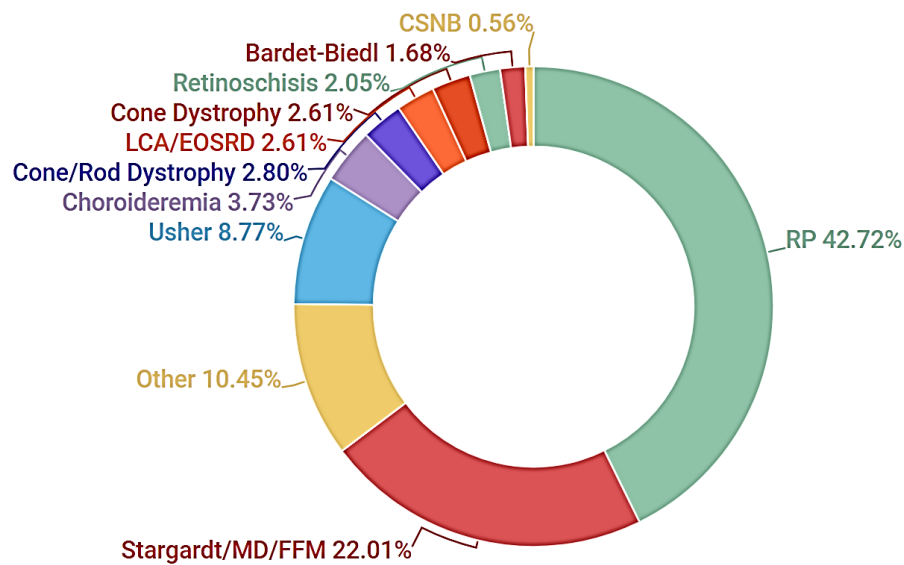


Figure 4.1.1. Clinical presentation of all inherited retinal degeneration (IRD) pedigrees included in the study as of the end of 2017. Abbreviations are listed clockwise as they appear above. RP: retinitis pigmentosa; Stargardt: Stargardt disease; MD: macular dystrophy; FFM: fundus flavimaculatus; Usher: Usher Syndrome; LCA: Leber congenital amaurosis; EOSRD: early-onset severe retinal dystrophy; Bardet-Biedl: Bardet-Biedl syndrome; CSNB: congenital stationary night blindness. Image taken from Dockery et al. (23).

A	BBS9	CNGB3	GNAT1	K	NEK2	PEX1	RDH12	T	WFS1
ADIPOR1	BEST1	CNNM4	GNAT2	KCNJ13	NEUROD1	PEX2	RDH5	TEAD1	WFS1
ABCA4	C	COL11A1	GNB3	KCNV2	NMNAT1	PEX7	RGR	TIMM8A	WHRN
ABCC6	C12orf65	COL2A1	GNPTG	KIAA1549	NPHP1	PGK1	RGS9	TIMP3	Z
ABHD12	C1QTNF5	COL9A1	GPR179	KIF11	NPHP3	PHYH	RGS9BP	TMEM126A	ZNF408
ACBD5	C21orf2	CRB1	GRK1	KIZ	NPHP4	PITPNM3	RHO	TMEM216	ZNF423
ADAM9	C2orf71	CRX	GRM6	KLHL7	NR2E3	PLA2G5	RIMS1	TMEM237	ZNF513
ADAMTS18	C8orf37	CSPP1	GUCA1A	L	NR2F1	PLK4	RLBP1	TOPORS	
ADGRA3	CA4	CTBP2	GUCA1B	LAMA1	NRL	PNPLA6	ROM1	TREX1	
ADGRV1	CABP4	CTNNA1	GUCY2D	LCA5	NYX	POC1B	RP1	TRIM32	
ADIPOR1	CACNA1F	CYP4V2	H	LRAT	O	POMGNT1	RP1L1	TRNT1	
AGBL5	CACNA2D4	D	HARS	LRIT3	OAT	PRCD	RP2	TRPM1	
AHI1	CAPN5	DHDDS	HGSNAT	LRP5	OFD1	PRDM13	RP9	TSPAN12	
AIPL1	CC2D2A	DHX38	HK1	LZTFL1	OPA1	PROM1	RPE65	TTC8	
ALMS1	CDH23	DMD	HMCN1	M	OPA3	PRPF3	RPGR	TTL5	
ARL2BP	CDH3	DRAM2	HMX1	MAK	OPN1LW	PRPF31	RPGR	TPPA	
ARL3	CDHR1	DTHD1	I	MAPKAPK3	OPN1MW	PRPF4	RPGRIP1	TUB	
ARL6	CEP164	E	IDH3B	MERTK	OPN1SW	PRPF6	RPGRIP1L	TUBGCP4	
ASRGL1	CEP250	EFEMP1	IFT140	MFN2	OTX2	PRPF8	RS1	TUBGCP6	
ATF6	CEP290	ELOVL4	IFT172	MFRP	P	PRPH2	RTN4IP1	TULP1	
ATXN7	CERKL	EMC1	IFT27	MFSD8	PANK2	PRPS1	S	U	
B	CFH	EXOSC2	IMPDH1	MIR204	PAX2	R	SAG	UNC119	
BBIP1	CHM	EYS	IMPG1	MKKS	PCDH15	RAB28	SDCCAG8	USH1C	
BBS1	CIB2	F	IMPG2	MKSI	PCYT1A	RAX2	SEMA4A	USH1G	
BBS10	CLN3	FAM161A	INPP5E	MTTP	PDE6A	RB1	SLC24A1	USH2A	
BBS12	CLRN1	FLVCR1	INVS	MVK	PDE6B	RBP3	SLC25A46	V	
BBS2	CLUAP1	FSCN2	IQCB1	MYO7A	PDE6C	RBP4	SLC7A14	VCAN	
BBS4	CNGA1	FZD4	ITM2B	N	PDE6G	RCBTB1	SNRNP200	W	
BBS5	CNGA3	G	J	NBA5	PDE6H	RD3	SPATA7	WDPCP	
BBS7	CNGB1	GDF6	JAG1	NDP	PDZD7	RDH11	SPP2	WDR19	

Figure 4.1.2. Current Target 5000 target gene panel (254 genes).

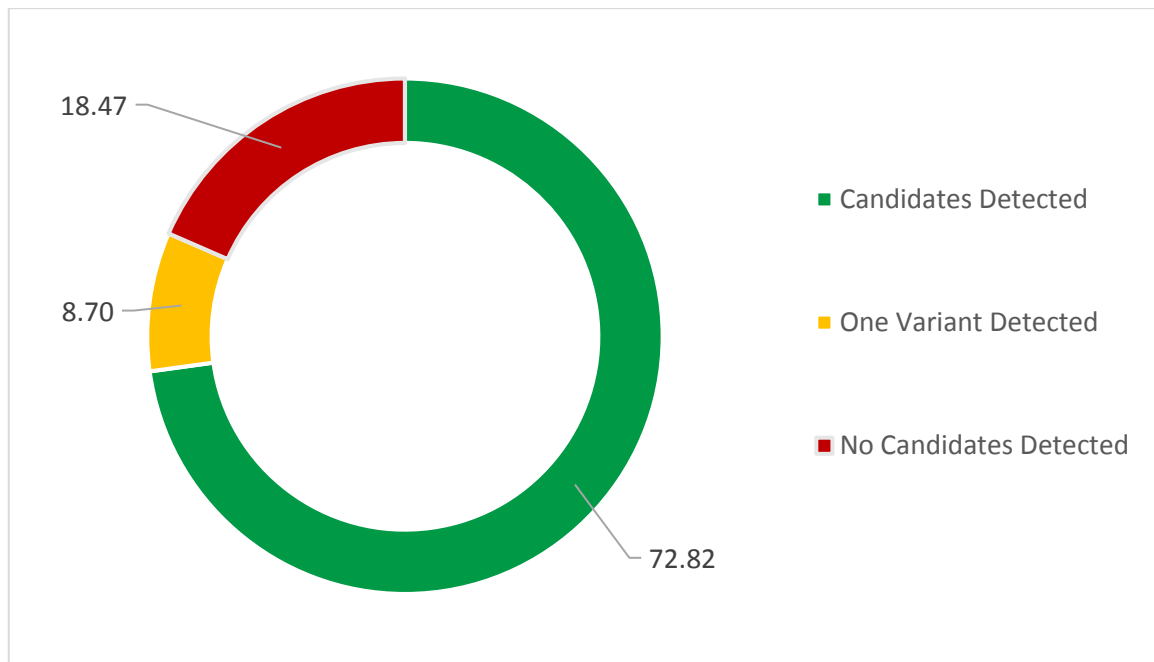


Figure 4.1.3. Current Target 5000 pedigree solve rate percentages. Target capture next generation sequencing for predominantly the exonic regions of up to 254 IRD-associated genes and a small number of known pathogenic variant harbouring intronic regions was employed to achieve this.

Table 4.1.1. Table of Novel Likely Pathogenic Variants. Missense mutations had to show segregation in pedigrees of at least 3 members, two of whom had to be affected. This table does not include the novel large structural variants that are reported in previous sections. The mutations in this table have not been previously associated with IRDs to the best of our knowledge.

Gene	Condition	Transcript ID	Genomic Location	Nucleotide Change	Protein Change	MetaLR Score	SpideX Z-Core	Observed with:
ABCA4	Stargardt Disease	NM_000350.2	g.1:94007721	c.5917delG	p.Val1973fs	None	None	p.Arg290Trp
ABCA4	Stargardt Disease	NM_000350.2	g.1:94098827	c.735T>G	p.Tyr245 *	None	-3.542	p.Asn1868Ile
ABCA4	Stargardt Disease	NM_000350.2	g.1:94030459	c.4320delT	p.Phe1440fs	None	None	p.Asn1868Ile
ADGRV1	USH Type II	NM_032119.3	g.1:91153391	c.18795delA	p.Leu6265fs	None	None	p.Arg5772 *
BBST1	Bardet-Biedl	NM_024649.4	g.11:66531658	c.1614delC	p.Leu539fs	None	None	p.Met390Arg
CNCA3	Cone Dystrophy	NM_001298.2	g.2:98396398	c.1228C>T	p.Arg410Trp	0.9441	None	p.Arg410Trp
ELOVL4	Stargardt Disease	NM_022726.3	g.6:79916758	c.789delTAACCTTmsAACT	p.Phe265fs_Asn264fs	None	None	/
MYO7A	USH Type I	NM_000260.3	g.11:77202351	c.5095C>T	p.Gln169 *	None	-3.017	p.Ala2009fs
MYO7A	USH Type I	NM_000260.3	g.11:77208775	c.6025delG	p.Ala2009fs	None	None	p.Gln1699 *
NR2E3	Retinitis Pigmentosa (Recessive)	NM_014249.3	g.15:71813635	c.994G>T	p.Glu332 *	None	None	p.119-2A>C
NYX	Congenital Stationary Night Blindness	NM_022567.2	g.X:41474475	c.1022A>C	p.Asp341Ala	0.7224	None	/
OPA1	Optic Atrophy	NM_130837.2	g.3:193614740	c.53_62delTGAAACACACAG	p.Val18fs	None	None	/
PRPF31	Retinitis Pigmentosa (Dominant)	NM_015629.3	g.19:54123748	c.528-1G>T	/	None	-2.131	/
PRPF8	Retinitis Pigmentosa (Dominant)	NM_006445.3.2	g.17:1653571	c.6337_6339delAAG	p.Lys2113del	None	None	/
RHO	Retinitis Pigmentosa (Dominant)	NM_000539.3	g.3:139531026	c.512C>A	p.Pro171Gln	0.4368	0.875	/
RP1	Retinitis Pigmentosa (Recessive)	NM_006269.1	g.8:534621124	c.160delG	p.Val54fs	None	None	p.Val54fs
RP2	Retinitis Pigmentosa (X-Linked)	NM_006915.2	g.X:46853795	c.425delA	p.Asn142fs	None	None	/
RPGR	Retinitis Pigmentosa (X-Linked)	NM_001034853.1	g.X:38266761	c.2236_2237delGA	p.Glu746fs	None	None	/
RPGRIPI	Leber Congenital Amaurosis	NM_020366.3	g.14:21328427	c.2899C>T	p.Gln967 *	None	-3.491	p.Gln967 *
R51	Retinoschisis	NM_000330.3	g.X:18644535	c.416delA	p.Gln139fs	None	None	/
R51	Retinoschisis	NM_000330.3	g.X:18647190	c.326+1G>A	/	None	-2.194	/
USH2A	Retinitis Pigmentosa (Recessive)	NM_206933.2	g.1:215647520	c.14791+2T>A	/	None	-2.921	p.Pro4090Thr
USH2A	USH Type II	NM_206933.2	g.2015680269	c.12172_12172delCTGinsTAAG	p.Leu4058fs	None	None	p.Glu767fs

Stargardt Disease

Stargardt disease (STGD1; OMIM #248200) is the most common inherited macular dystrophy with an estimated incidence ranging from one in 8,000 to one in 10,000 (146). Typical clinical features include bilateral central vision loss with macular atrophy and flecks in the retinal pigmentary epithelium. The age of onset and severity is dependent on the intrinsic pathogenicity of the causative mutations however most disease appears evident within the first two decades of life (147).

Stargardt disease is the second most common IRD observed in our study. *ABCA4* was unsurprisingly the most frequently observed gene associated with STGD1 or its phenotypically similar counterpart, fundus flavimaculatus. *ABCA4* was deemed the candidate gene in 80 pedigrees across these two conditions. At least two *ABCA4* candidate variants were detected in 67 of these pedigrees (84%). Several novel likely pathogenic mutations were observed in this study relating to STGD1, NM_000350.2: c.5917delG, p.Val1973fs; c.735T>G, p.Tyr245*; c.4320delT, p.Phe1440fs. These novel *ABCA4* mutations were all observed in separate Stargardt pedigrees, segregating with the condition in each family and *in trans* with known pathogenic *ABCA4* mutations.

Despite the relative genetic homogeneity of STGD1, *ABCA4* still represents a challenge in terms of diagnostics due to the suspected prevalence of many deep intronic pathogenic mutations (73). However, as many of these IRD diagnostic studies progress, it is likely that the increasing body of data will make it possible to more accurately estimate the pathogenicity of variants discovered during sequencing.

One example of this is the reclassification of the hypomorphic *ABCA4* variant p.Asn1868Ile (c.5603A>T) (148). This variant was previously dismissed as benign due

to its background relatively high population frequency of 7%. This p.Asn1868Ile variant was observed in 111 sequenced patients across 95 different pedigrees, including 4 patients who were homozygous for the mutation. Some homozygous patients also cited a strong family history of age-related macular degeneration (AMD).

Although initially filter out of our variant calling pipeline due to its highest population frequency, the inclusion of this single variant in the analysis has aided in the assessment of an additional 19 STGD1 patients in our IRD patient cohort. The majority of these 19 patients were clinically diagnosed with a milder form of STGD1 or a late-onset macular degeneration but only a single pathogenic or likely pathogenic *ABCA4* mutation could previously be detected. The observation in the current study that the p.Asn1868Ile variant is associated with a clinically distinguishable milder form of STGD1 is consistent with the associated phenotypes outlined by Zernant and colleagues (148).

ELOVL4 and *PROM1* also contributed to a small number of pedigrees that presented as Stargardt-like dominant macular dystrophies. Novel likely pathogenic mutations were detected in *ELOVL4* during cohort analysis (Table 4.1.1). However, the most common candidate mutations associated with dominant maculopathy in our participants were found in the *BEST1* gene. Notably *ABCA4* variants accounted for over 70% of all candidate variants across all macular degeneration associated pedigrees at the end of 2017 (Figure 4.1.4). This includes pedigrees that presented with an AMD-like phenotype who were included in the study due to the strong family history of the condition.

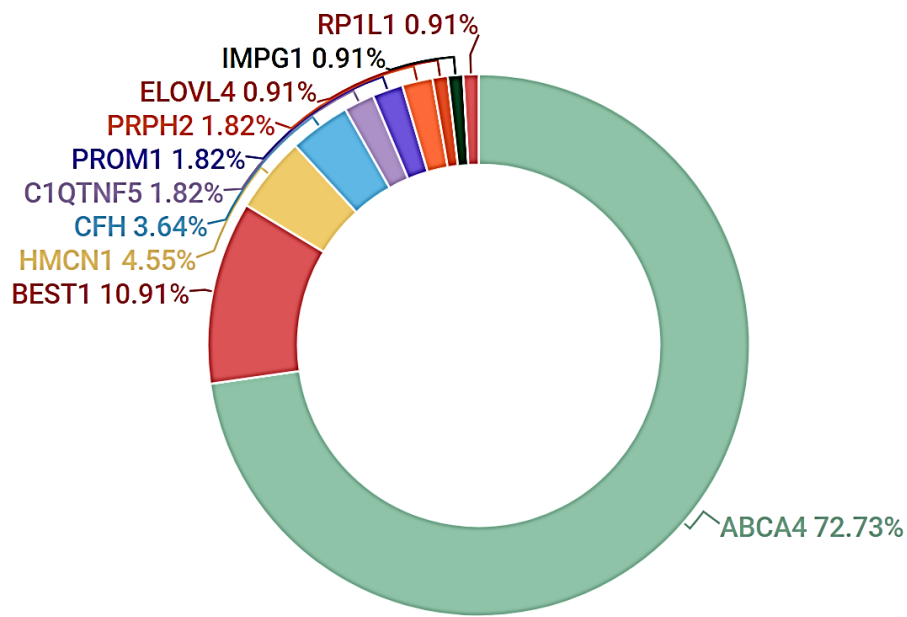


Figure 4.1.4. Gene candidates (and percentages) observed in the IRD grouping encompassing all macular-associated conditions including Stargardt disease, fundus flavimaculatus (FFM), macular dystrophy (MD) and age-related macular degeneration-like conditions (AMD-like). Image taken from Dockery et al. (23)

Retinitis Pigmentosa

RP was the most common clinical diagnosis for participants in the current study, accounting for nearly 40% of the total pedigrees sequenced. This figure encompasses the various modes of inheritance (autosomal dominant, autosomal recessive and X-linked) and phenotypes (typical, inverse and paravenous, among others) associated with RP.

The most common gene candidate found for dominant RP in the study was *RHO* (OMIM #180380). This accounts for over 12% of all RP pedigrees involved in the study and almost 30% of pedigrees diagnosed with dominant RP. This figure is an under representation of the prevalence of *RHO*-linked RP in this IRD population, as several pedigrees involved in prior single-gene studies on *RHO* were excluded from the current study, as causative mutations had already been established (149–151).

The predominant pathogenic mutations observed are NM_000539.3:c.533A>G, p.Tyr178Cys; c.620T>G, p.Met207Arg, however, numerous novel likely pathogenic mutations have also been detected. One variant found in a sufficiently large pedigree to verify segregation is shown in Table 4.1.1. In prior single-gene studies, the *RHO* mutations p.Tyr178Cys, p.Met207Arg and p.Thr94Ile were also commonly observed in the Irish IRD population.

USH2A (OMIM #608400) was the most frequently observed gene candidate for recessive RP. Pathogenic or novel mutations in this gene were observed in over 30% of recessive RP resolved cases. Although many *USH2A* pathogenic mutations are associated with Usher syndrome, none of the pedigrees which were initially clinically diagnosed with

recessive RP showed any other syndromic traits. This finding of non-syndromic retinal degenerations associated with mutations in *USH2A* is consistent with observations based on other *USH2A* cohorts (152).

In one Irish IRD pedigree with two affected individuals diagnosed with recessive RP, a large heterozygous deletion was detected in *USH2A* (Figure 4.1.5) by employing the structural variants analysis pipeline described in the Methods section. This deletion was approximately 9 kb in size and spans the first 150 amino acids of the coding region, genomic coordinates (hg38): g.chr1:216421919-216428002. This deletion encompasses a significant part of *USH2A*, including the start codon, transcriptional start site and part of the first exon. As such, it is highly likely to cause complete loss of function.

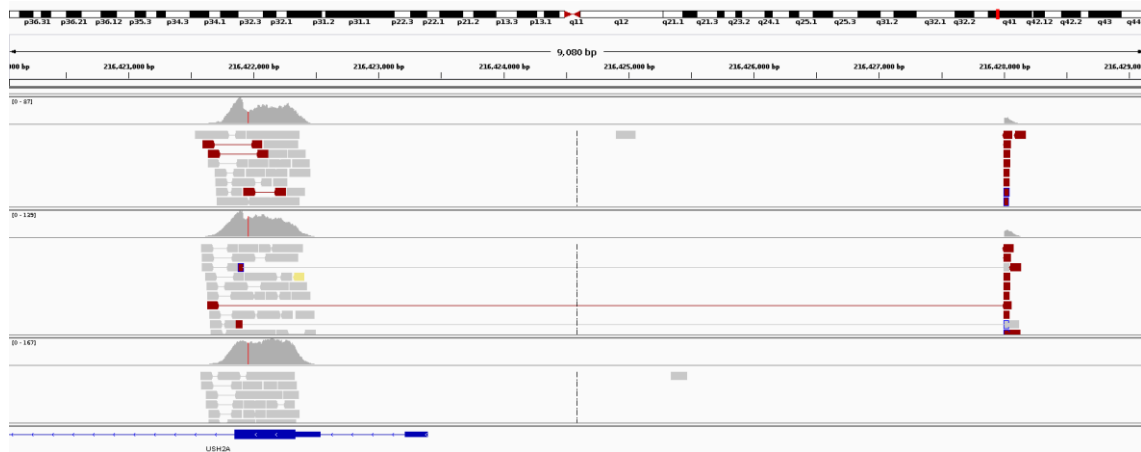


Figure 4.1.5. Sequence data from retinitis Pigmentosa (RP) patients with large USH2A deletions. The top two panels of reads are from the two affected patients who presented with RP. The bottom panel of reads is from an unaffected unrelated patient. Depicted above is the evidence of split reads (red) and a decline in read-depth in the affected patients relative to a control. Image taken from Dockery et al. (23).

The 9 kb mutation in *USH2A* was detected alongside a reported pathogenic mutation further downstream in the gene in both affected patients in this pedigree, NM_206933.2:c.2276G>T, p.Cys759Phe. An unaffected sibling from this pedigree was directly sequenced for both mutations and was found to be negative for the large deletion (Figure 4.1.6a) and positive for the pathogenic single base mutation, p.Cys759Phe (Figure 4.1.6b) confirming that the two *USH2A* mutations observed in this pedigree are found on different alleles. The region surrounding the breakpoint in exon 2 was a long pyrimidine run, encoding a series of serine and proline residues. The region surrounding the breakpoint upstream of *USH2A* was very similar, although one purine nucleotide was present in the surrounding 25 bp. It is plausible that these sequence similarities facilitated the formation of the deletion.

RP1 (OMIM #603937) mutations have been found in both dominantly and recessively inherited forms of RP. *RP1*-RP was first described as a dominantly inherited condition (153) and then subsequently found to also be associated with a recessive mode of inheritance several years later (154). Although there are quite a number of missense pathogenic mutations associated with *RP1*-RP (155), only a single missense mutation was observed in the current IRD cohort, p.Thr373Ile.

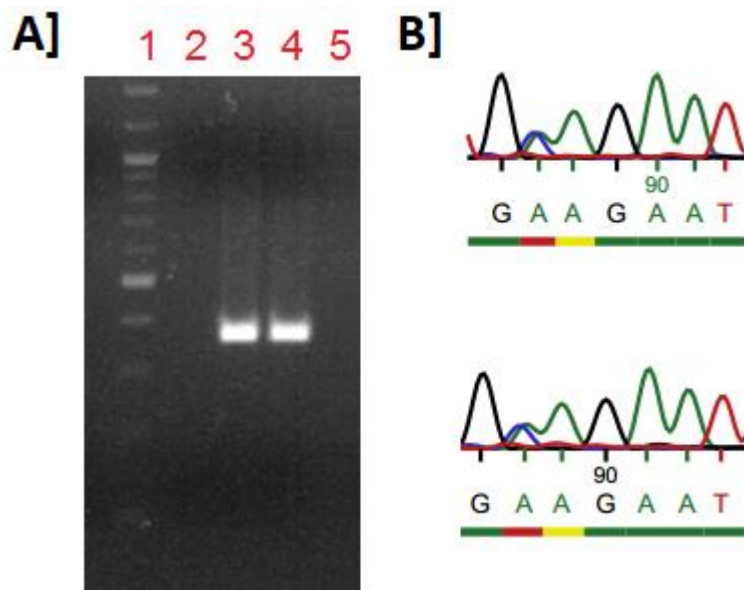


Figure 4.1.6. Verification of USH2A variants. **A]** Gel confirmation of USH2A deletion. Primers were designed to target the region of deletion in affected patients C41 and E15. A product of the desired size (~300bp) would only be formed if the deletion was present. The sample from the unaffected sibling in the same pedigree, P36, did not form a product. A product was formed from the samples of C41 and E15. Lane 1: 100 bp Ladder; Lane 2: Unaffected Patient P36; Lane 3: Affected Patient C41; Lane 4: Affected Patient E15; Lane 5: No template control **B]** Sanger sequencing trace of USH2A mutation, p.Cys759Phe. The top trace is unaffected patient P36 and the bottom trace is affected patient E15. It is clear that both patients share the same heterozygous pathogenic point mutation.

This *RP1* mutation was found heterozygously in 34 different IRD patients from our study. As this mutation has only been observed to be causative of an IRD when homozygous or as a compound heterozygote (154), it could not be deemed causative in any of the cases in this current study as a second pathogenic mutation was not detected in the *RP1* exonic sequence. All other candidate recessive or dominant mutations in *RP1*-RP deemed to be likely causative of the observed phenotype were frameshift mutations, including a novel recessive mutation, NM_006269.1:c.160delG, p.Val54fs (see Table 4.1.1).

Of the X-linked RP cases in which a candidate variant was detected, 80% carried known or likely pathogenic variants in the *RPGR* gene (OMIM #312610). The majority of these causative *RPGR* mutations were frameshift mutations, including one novel mutation, NM_001034853.1: c.2236_2237delGA, p.Glu746fs. Given that the repetitive region of *RPGR*'s ORF15 tends to negatively impact NGS efficacy (74), pathogenic mutations were rarely called in this region unless both read quality and depth were of a high standard. A small number of pedigrees also tested positive for candidate mutations in *RP2* (OMIM #300757), including a novel frameshift mutation, NM_006915.2:c.425delA, p.Asn142fs. All candidate genes found in RP pedigrees as part of this study are depicted in Figure 4.1.7.

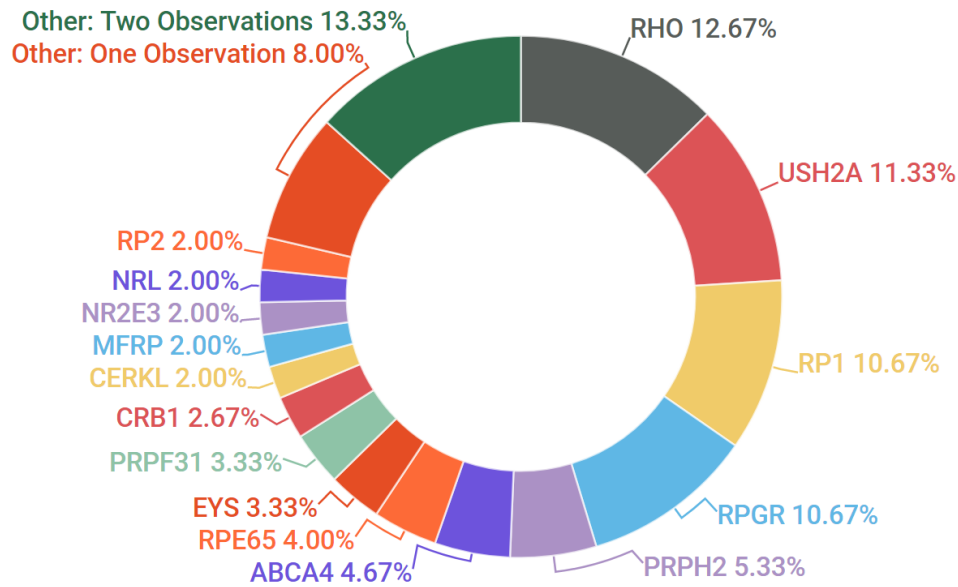


Figure 4.1.7. Gene candidates (and percentages) observed in cases of RP across all forms of inheritance. Genes with two observations as candidates: *CRX*, *IFT140*, *KLHL7*, *PDE6A*, *PDE6B*, *PROM1*, *PRPF8*, *ROM1*, *SLC24A1*, *SNRNP200*. Genes with single observations as candidates: *AHI1*, *C2orf71*, *CNGA1*, *CNGB1*, *GNAT1*, *GUCA1B*, *HK1*, *IMPDH1*, *OFD1*, *PRPF6*, *RDH12*, *TULP1*. Image taken from Dockery et al. (23).

Bardet-Biedl Syndrome

Although only 8 pedigrees in this study were initially clinically diagnosed with Bardet-Biedl Syndrome (BBS; OMIM #209900), an additional 19 pedigrees were rediagnosed as BBS, the majority of which presented clinically as simplex or recessive RP as no other syndromic signs characteristic of BBS were obvious to the clinical staff initially. This rediagnosis was prompted by genetic findings and where possible, confirmed by the ophthalmologists associated with the project. A number of these cases were subsequently found to have had polydactyly, which had typically been surgically removed at a young age. Additionally, a smaller number of patients were seen to have more prominent intellectual disabilities and indications of obesity upon subsequent annual clinical check-ups.

The most frequent mutation detected in the patient population was *BBS1* (OMIM #209901) NM_024649.4:c.1169T>G, p.Met390Arg. This mutation was observed homozygously in 17 pedigrees, and heterozygously in another 4 pedigrees. It was also the most likely BBS mutation to clinically present as simplex RP. The findings are consistent with other studies carried out with Northern European IRD patient cohorts (156,157). Five additional BBS pedigrees revealed candidate mutations across four other known BBS genes: *BBS4* (OMIM #600374), *BBS9/PTHB1* (OMIM #607968), *BBS10* (OMIM #610148) and *BBS16/SDCCAG8* (OMIM #613524).

Usher Syndrome

Usher Syndrome (USH) is diagnosed by the concurrent incidence of auditory impairment and RP. Three subtypes of USH were detected in our patient cohort, Type 1 (OMIM #276900), Type 2 (OMIM #276901) and Type 3 (OMIM #276902). These subtypes can be broadly identified by severity of hearing loss: congenital deafness, loss of hearing and progressive loss of hearing respectively. 47 pedigrees in the study were clinically diagnosed with some form of USH and at least one candidate variant was found in 44 pedigrees (94%). Pathogenic or likely pathogenic candidate mutations in *USH2A* alone were detected in 50% of all USH cases, which accounts for the majority of USH Type 2 cases. The most frequent candidate gene for USH Type 1 was *MYO7A* (OMIM #276903). Together, *USH2A* and *MYO7A* accounted for over 70% of the pathogenic mutations detected across all USH subtypes (Figure 4.1.8).

A large structural variant was also detected in an USH pedigree. In this instance, a patient presented with USH Type 1 (Figure 4.1.9) and no variants in USH associated genes were detected. However, once a detailed analysis was undertaken with our structural variants pipeline (Chapter 3.11), it became apparent that a large homozygous deletion spanning the first four exons of *USH1C* was present, (Figure 4.1.10). These findings were verified by PCR experiments designed to amplify the first four exons of *USH1C* (Figure 4.1.11).

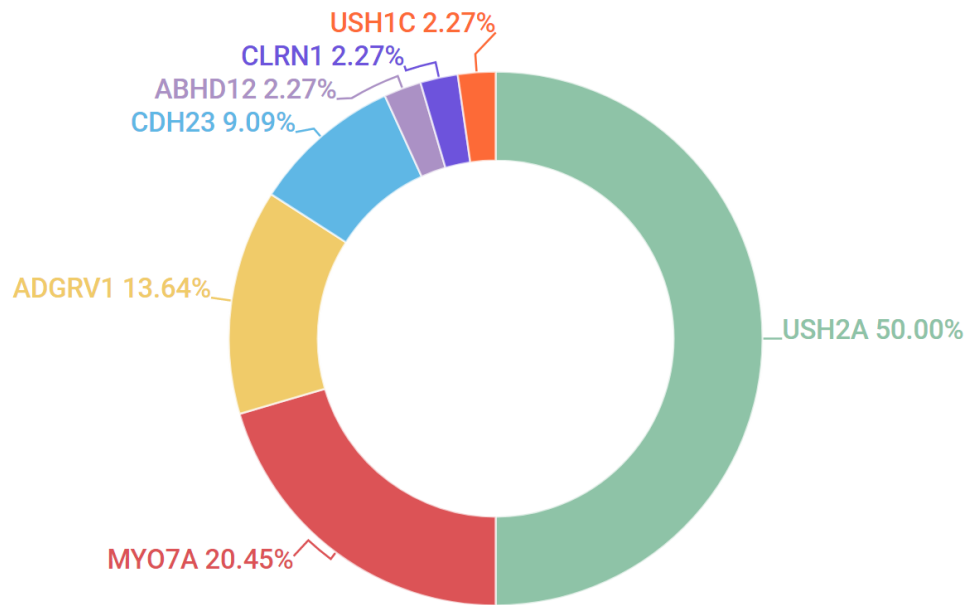


Figure 4.1.8. *Gene candidates (and percentages) found across all subtypes of Usher Syndrome. Image taken from Dockery et al. (23).*

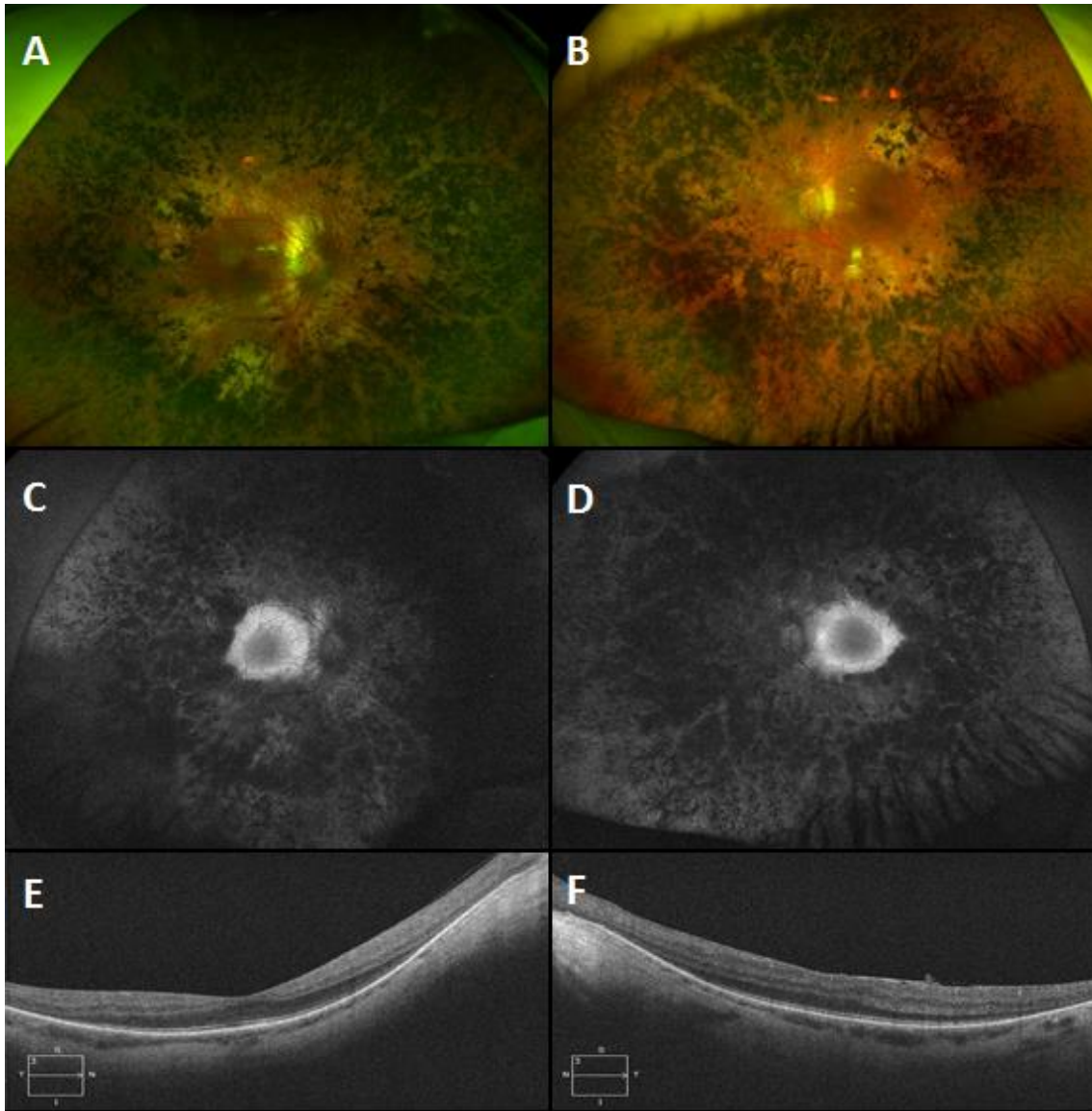


Figure 4.1.9. Montage of clinical images from an Usher syndrome patient with a large homozygous deletion in *USH1C*. Colour images of the right (**A**) and left (**B**) eyes of a 44-year-old male patient with Usher syndrome type 1 which show symmetrical circumferential mid-peripheral bone spicule pigmentation and RPE atrophy with relative preservation of the maculae. The corresponding autofluorescence images of the right (**C**) and left (**D**) eyes demonstrate peripheral hypoautofluorescence and preserved macular autofluorescence. Note the symmetrical hyperautofluorescent rings, which correspond to the extent of the preserved macular photoreceptors on the Optical Coherence Tomography (OCT) scans (**E**: right eye and **F**: left eye). Image taken from Dockery et al. (23).

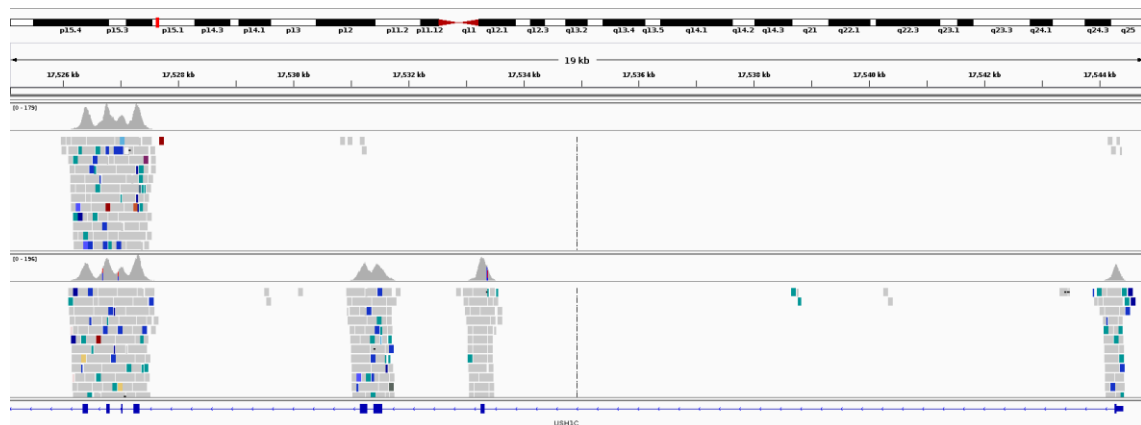


Figure 4.1.10. Large homozygous deletion spanning the first four exons of *USH1C*. In the top panel of reads from the patient sample, it is clear that there was no sequencable template in this region compared to the control (bottom panel of reads). Image taken from Dockery et al. (23).

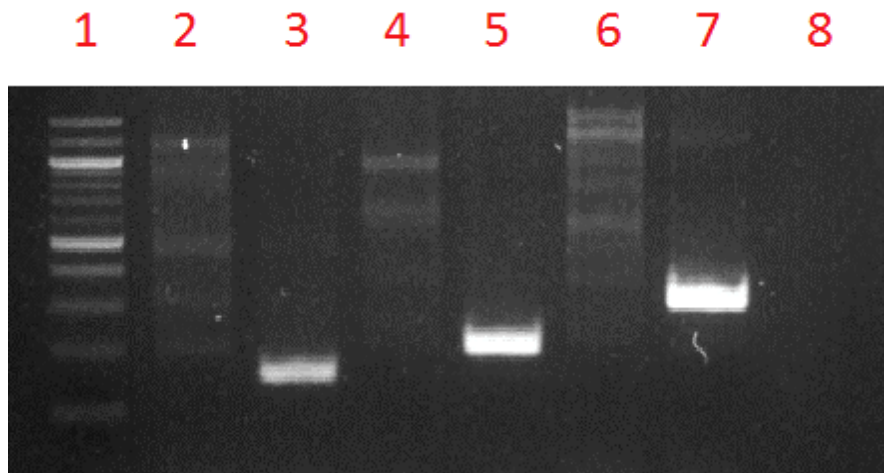


Figure 4.1.11. Analysis of USH1C deletion by use of PCR products. Primers were designed to amplify each exon believed to be deleted in Patient 1363. If the sequence for the exon was present a product would be formed. Exons 3 and 4 were designed as a single amplicon due to their proximity. The unaffected individual is positive for all of the USH1C exons. Patient 1363 is negative for all USH1C exons. Lane 1: 100 bp Ladder; Lane 2: Patient 1363 with primers for exon 1; Lane 3: Unaffected Patient with primers for exon 1; Lane 4: Patient 1363 with primers for exon 2; Lane 5: Unaffected Patient with primers for exon 2; Lane 6: Patient 1363 with primers for exons 3 + 4; Lane 7: Unaffected Patient with primers for exons 3 + 4. Lane 8: No template control.

Gyrate Atrophy

Gyrate atrophy/gyrate atrophy of choroid and retina (GA/GACR; OMIM #258870) is an extremely rare condition linked with a single gene aetiology, *OAT*. It is estimated that the global incidence of GA is less than one in a million, with the exception of Finland which has an estimated incidence of one in 50,000 (4). In some regions of the world, such as Australia, the condition is sufficiently rare that the first case of GA has only been reported in the last few years (158).

The *OAT* gene encodes a mitochondrial matrix enzyme, ornithine aminotransferase, which is involved in many biologically fundamental pathways such as the urea cycle, biosynthesis of creatine and proline metabolism (159). GA has a range of consequential phenotypes associated with it, primarily cystoid macular edema. It is for this reason that an accurate diagnosis is essential as these conditions are becoming increasingly treatable, both through dietary restrictions and medical intervention (97).

Here we report a novel homozygous inversion of the *OAT* gene. To the best of our knowledge, this is the first large inversion to be reported in this gene. The patient presented in the clinic with an obvious GA phenotype (Figure 4.1.12) and following analysis with the in-house structural variant calling pipeline (Chapter 3.11) it was apparent from the breakpoints that a 16.6 kb region had been inverted. The region starts approximately 3 kb upstream of the gene and ends in the middle of exon 5 of the gene (Figure 4.1.13 and Figure 4.1.14).

These findings were subsequently confirmed by alternating primer pairs designed to target the breakpoint regions of this inversion with unaffected patient DNA used as

controls. In the specific case of the homozygous inversion reported here, no product would result from pairings design against the reference genome for these breakpoints OAT-1 + OAT-2, OAT-3 + OAT-4. However, products in this case could only be formed when primers OAT-1 and OAT-3 were paired and likewise with primers OAT-2 and OAT-4, which would not generate any product when unaffected patient DNA was used as a template. This strategy is shown diagrammatically in Figure 4.1.15 and experimentally in Figure 4.1.16.

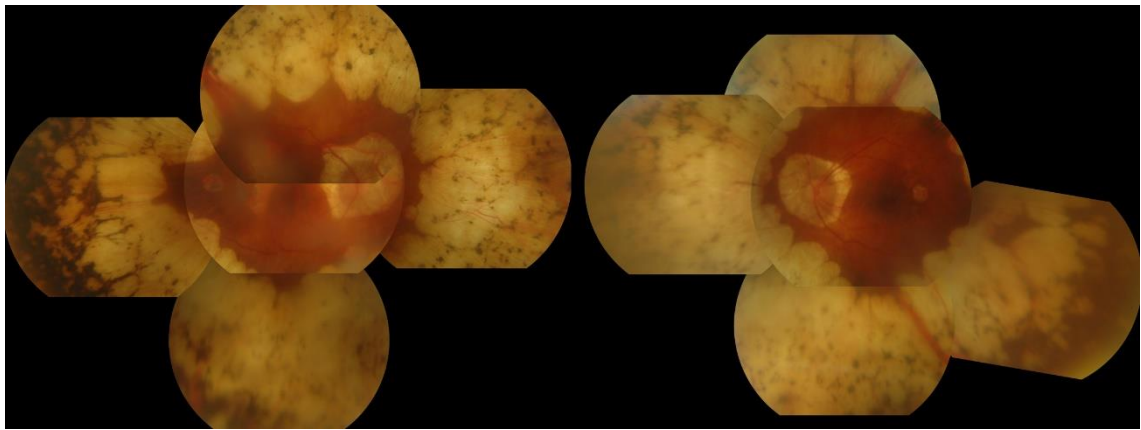


Figure 4.1.12. *Montage of fundus photography of an affected gyrate atrophy (GA) patient with an OAT inversion. Note the marked retinal and choroidal atrophy affecting the mid-peripheral fundus of each eye with relative sparing of the central macular area. The preserved macular areas show the scalloped edges which are characteristic of the condition. Image taken from Dockery et al. (23).*

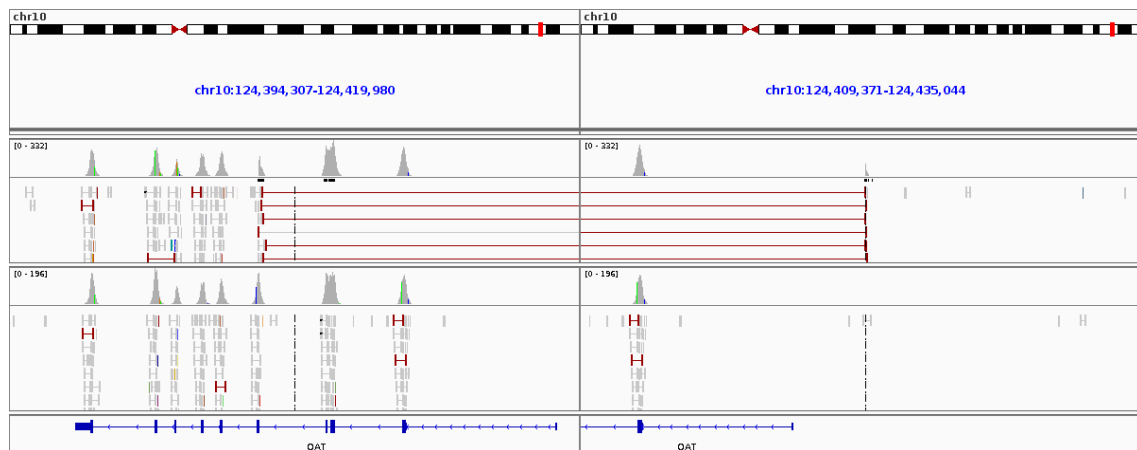


Figure 4.1.13. Sequencing results from the gyrate atrophy patient aligned to the reference genome. The top collection of sequence reads is from the affected GA patient and the bottom section of reads are from an unaffected patient sequenced in parallel. Split reads (red) from the affected patient can be seen to be spanning a 16.6 kb region of the OAT gene as viewed in IGV (Integrative Genomics Viewer). Image taken from Dockery et al. (23).

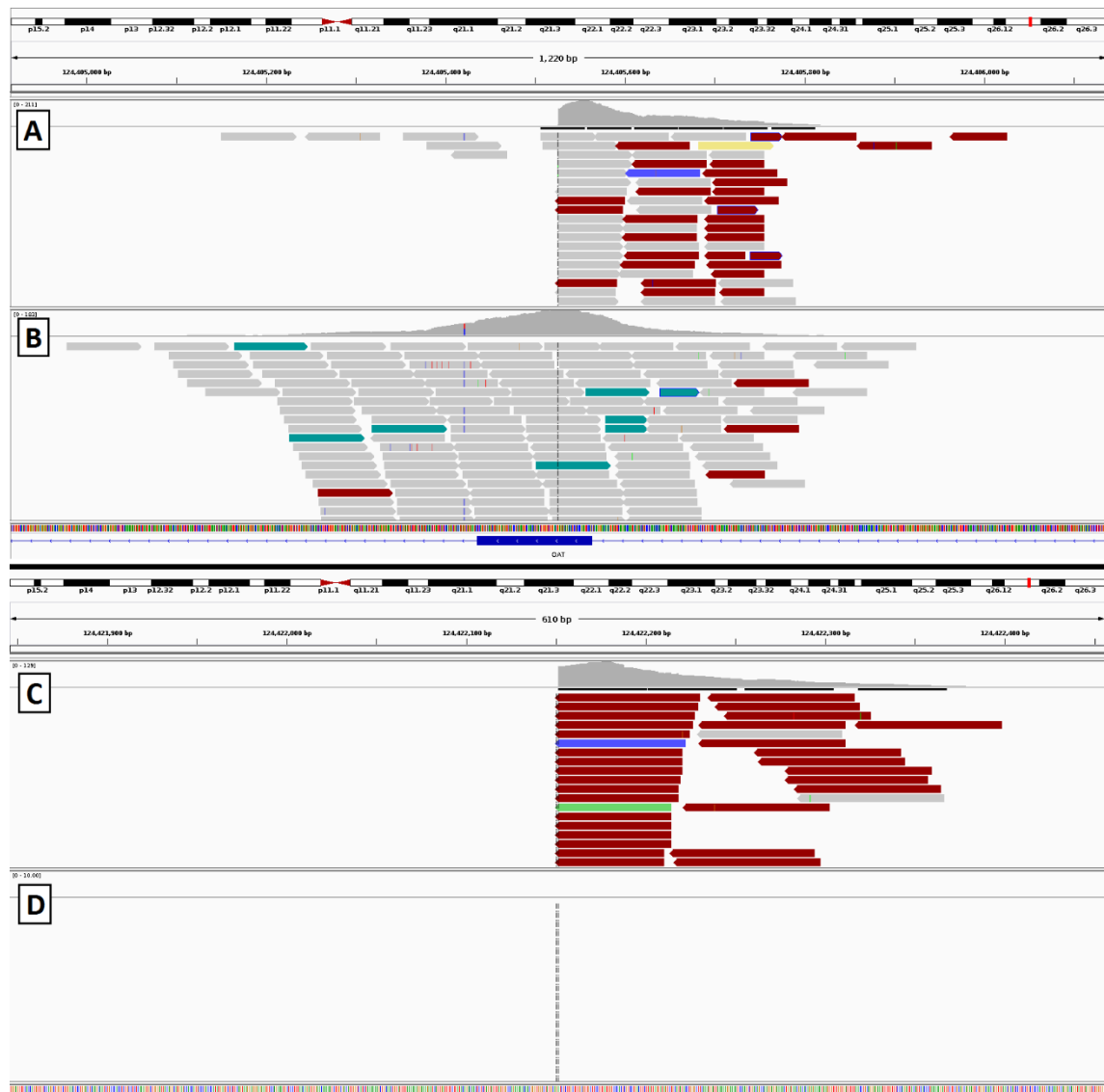


Figure 4.1.14. Sequencing results from the gyrate atrophy patient depicting alignment of reads at each breakpoint. (A, B) show reads surrounding the exonic breakpoint in the affected patient (A) and an unaffected control (B). (C, D) show the distal breakpoint. This region is not covered by our capture panel, but in the affected patient sample (C) alignments can be seen from library fragments that span the inversion breakpoint. In the control patient (D), no aligned reads are seen, as expected. Image taken from Dockery et al. (23).

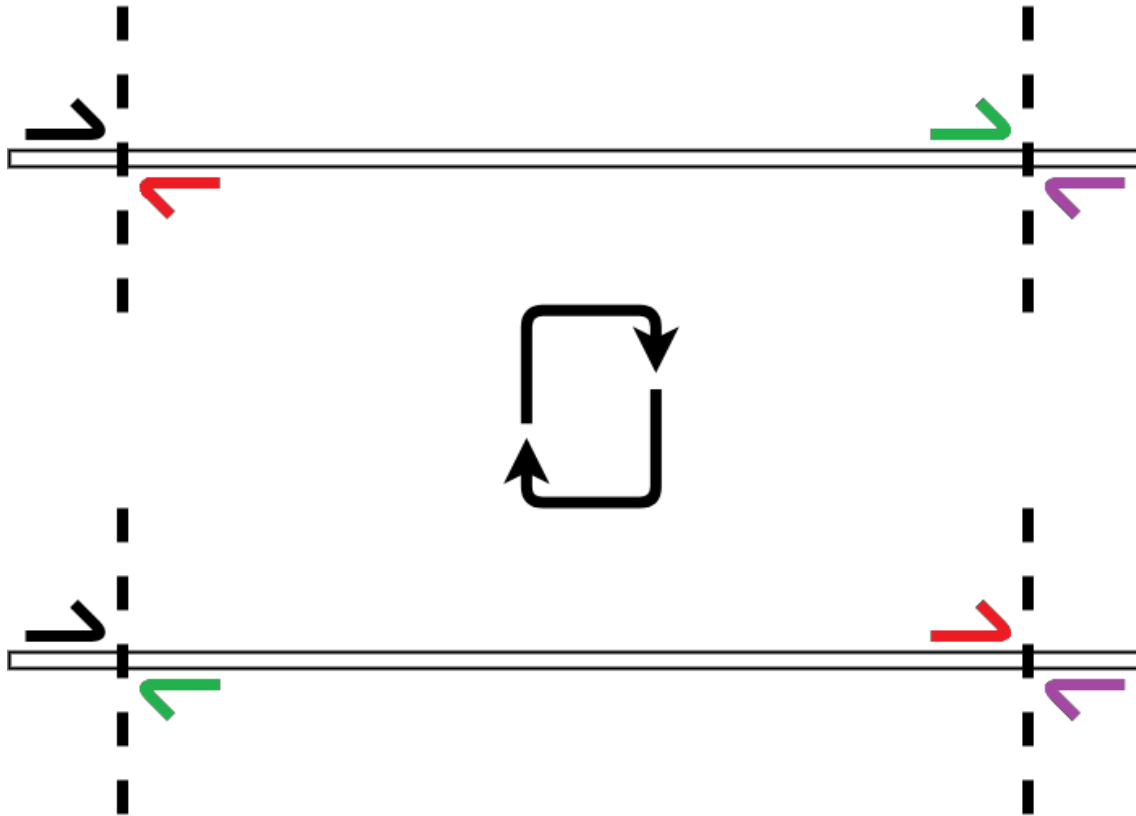


Figure 4.1.15. An illustration of the PCR strategy designed to detect a large homozygous inversion. Primers are indicated by colour; OAT-1 (black), OAT-2 (red), OAT-3 (green) and OAT-4 (purple). Broken lines indicate genomic breakpoints. Circular arrows indicate an inversion event. The top illustration depicts how primers would anneal around the breakpoints in a control sample. The bottom illustration shows how the internal primers (OAT-2 and OAT-3) are shuffled in an inversion event to form new primer pairings, detectable by PCR analysis.

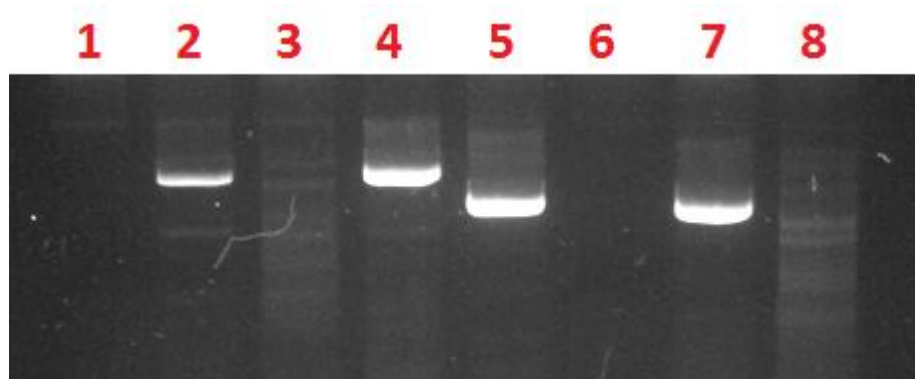


Figure 4.1.16. Confirmation of an OAT inversion using strategic PCR design. If the wildtype OAT sequence is present, the primer sets 1 + 2 and 3 + 4 can be used to generate a PCR product. If the inversion has occurred primer sets 1 + 3 and 2 + 4 can be used to generate of a PCR product. Unaffected patient is positive for products of a wildtype sequence. Patient H58 is positive only for inversion-specific products. Lane 1: Patient H58 Primers 1 and 2; Lane 2: Patient H58 Primers 1 and 3; Lane 3: Patient H58 Primers 1 and 2; Lane 4: Patient H58 Primers 2 and 4; Lane 5: Unaffected Patient Primers 1 and 2; Lane 6: Unaffected Patient Primers 1 and 3; Lane 7: Unaffected Patient Primers 3 and 4; Lane 8: Unaffected Patient Primers 2 and 4.

Novel Variants

Several novel likely pathogenic variants were detected in the course of this study across all conditions (Table 4.1.1). Pathogenicity of novel missense mutations was predicted using the various bioinformatics tools outlined in the Variant Analysis section (Chapter 3.9-3.11) including MetaLR, M-CAP and SPIDEX. Segregation analysis was undertaken for pedigrees sufficiently large to do so where available. A number of the variants reported here (Table 4.1.1) had already been allocated a dbSNP ID (*CNGA3* p.Arg410Trp, rs137852608; *RS1* c.326+1G>A, rs281865346) as they had been detected in previous sequencing studies, most likely as part of population studies not specifically focused on the detection of IRDs. Furthermore, those variants that had been previously identified contained no record of known pathogenicity or of the phenotype of the individual in which they were identified.

4.1.4 Discussion

To date, as part of Target 5000, approximately 20% of the Irish IRD patient population has been sequenced (nearly 1,000 patients), providing the first national-scale overview of the IRD landscape. The study offers not only a chance to discover new pathogenic variants in known IRD genes but represents a vital initial step in the genetic characterisation of patients to provide them with information regarding the underlying genetic pathogenesis of their disease.

Previously, we have reported the identification of over 40 novel variants in a smaller cohort of IRD patients (85,93) and here we describe an additional 23 novel mutations and

3 novel structural variants, totalling nearly 70 novel IRD mutations discovered as part of this study. As this is an ongoing study, many candidate variants have also been detected that hold the potential to be novel pathogenic variants, but additional segregation analysis is required to confirm this. Several mutations that have been previously reported such as *RHO*, p.Met207Arg (149), have presented in multiple pedigrees in this study. It is likely that some or all of these pedigrees are distantly related and current analysis is ongoing to verify this.

Significantly, the genetic pathogenesis of some previously ambiguous disease phenotypes has also been resolved, most notably the milder, late-onset phenotype of Stargardt disease that is associated with the p.Asn1868Ile *ABCA4* mutation. It is also important to note that recently other researchers have found this variant to have very low penetrance when observed *in trans* with known “severe” *ABCA4* variants (160).

The original findings of the p.Asn1869Ile study were also further investigated in the same research group with a clinical re-assessment of 27 patients who were found to have the mild variant *in trans* with a severe variant (161). They found that these patients consistently showed foveal sparing and milder phenotypes particularly in comparison to affected family members who did not possess the mild variant. Another mild variant, c.4253+43G>A, has been observed to be enriched in a Stargardt population relative to controls (162). This is likely to be the first hypomorphic deep-intronic variant to be associated with Stargardt disease. Our data however, would not imply that this variant is either hypomorphic, pathogenic or enriched in Stargardt disease patients in our cohort.

Given the variability of clinical presentation, ranging from the more frequent-milder and late-onset to the rarer-asymptomatic phenotype, it is possible that this variant may act as a risk allele for environmental insults or alternatively, may go largely undetected in genetic studies of Stargardt disease due to its preponderance to phenocopy the characteristics of AMD.

AMD is the most common cause of visual impairment in the developed world (163). In our own study, patients presenting with AMD and no family history are largely not recruited to the study. However, given the potential for misdiagnosis due to this phenocopy issue, it is possible that some of these simplex cases may have mild Stargardt-like phenotypes. This could not be said for certain without the appropriate genetic testing. However, it is also important to consider the reverse of this argument, that variants currently associated with mild, late-onset macular conditions are not sufficiently pathogenic to illicit a disease phenotype and that AMD patients for example may be misdiagnosed as atypical Stargardt disease (160).

NGS-based genetic diagnoses of IRD patients in this cohort prompted a clinical re-evaluation for many patients, predominantly from simplex RP to BBS, caused by the p.Met390Arg mutation in the *BBS1* gene, patients often presenting with subtle additional phenotypes due to, for example, early intervention for polydactyly.

Many challenges still remain for the application of NGS technologies in diagnostic medicine. Ambiguous disease phenotypes and the presence of disease genes that may be associated with multiple IRDs and different modes of inheritance can make achieving a robust diagnosis particularly difficult. The presence of stretches of repetitive sequence in

some IRD genes can also make it difficult to confidently call variants in relevant portions of the genome, which we anticipate may mask some disease-causing variants from analysis.

For example, an approximately 800 bp region in the centre of *RPGR* ORF15 shows a sharp drop in mapped reads due to the repetitive nature of the sequence (74). Given that ORF15 has been implicated in cases of X-linked RP in the past (164), we anticipate that some undiagnosed patients with X-linked retinitis pigmentosa are likely to harbour mutations in this region of the gene, and efforts are underway to augment the protocol we employ to improve coverage of this region. This will be discussed in greater detail in Chapter 4.3.

In future sequencing panels, it would be desirable to successfully incorporate an augmented protocol that enhances the success of sequencing in this region. Previous studies have shown that processes can be implemented prior to sample preparation for sequencing (165) to aid sequencing this region. However, for the time being, this genomic region will be interrogated separately until integration is optimised. This parallel investigation method has been shown to effectively discover novel mutations in previously unresolved cases in this project (Chapter 4.3) and also in other studies (166).

The RP-associated *RP1* gene also presents some challenges in relation to genetic diagnostics. The pathogenicity and mode of inheritance of a novel mutation in *RP1* is difficult to determine as mutations in this gene have previously been associated with both dominant and recessively inherited disease. This is an issue that has been discussed in a number of other *RP1* studies (167–169).

In a recent study, a meta-analysis of previously reported *RP1* pathogenic mutations was undertaken to link the impact of each variant to the functional region of the protein. Difficulties remain in identifying new *RP1* mutations as dominant- or recessive-acting, however, as regions of the gene predominately associated with dominant RP were found to also harbour mutations associated with recessive rod-cone dystrophies (170).

Interestingly, *RP1* is somewhat unique in that it is relatively small as a whole gene. This means that the additional incorporation of intronic sequencing into a target capture panel would not occupy the same sequencing capacity that many other genes would, given the same approach. In fact, this approach has been explicitly warned against in studies that have attempted this for larger genes such as *ABCA4*, *USH2A* and *CEP290* (171). Currently, likely pathogenic or pathogenic *RP1* variants that affect splicing are absent from the ClinVar database (172). This modification will likely be incorporated into future Target 5000 target capture gene panels in the future.

Methods for NGS data analysis are undoubtedly evolving quite rapidly. Research into the effects of splice site mutations and their respective functional impacts is already providing significant insights into the effect(s) of previous potentially overlooked variants in IRD datasets (173–175). Additionally, it is becoming increasingly commonplace for NGS studies of IRD populations to incorporate some form of detection or analysis of copy number variants (CNVs) (75,84). It has been shown in these studies that up to 18% of previously unsolved IRD pedigrees can be resolved with the detection of pathogenic CNVs.

In the current study we describe the bioinformatics methodologies employed to retrospectively analyse datasets to detect CNVs and sequence breakpoints that are present within the captured exonic regions assessed by target capture NGS (Chapter 3.11). Adopting these methodologies, we have identified three IRD pedigrees carrying three separate large structural variants; a heterozygous large deletion in the *USH2A* gene, a homozygous large deletion in the *USH1C* gene and a homozygous large inversion in the *OAT* gene. The structural variants observed using this approach were identified in genes as diverse as the conditions themselves involving gyrate atrophy (*OAT*), retinitis pigmentosa (*USH2A*) and Usher syndrome (*USH1C*). The large deletions in the *CHM* gene found in this IRD cohort have been previously reported (93) and will be further discussed in a later chapter (Chapter 4.3).

These findings serve to emphasise the importance of implementing analysis systems that enable detection of large-scale deletions and inversions in all IRD patients, as currently, we have observed that 100% of these rare SV events correlate with an IRD gene-associated pathogenic phenotype. Although split-read and read-depth analysis of short-read capture data, as performed in our study, is less sensitive to structural variants than similar methods applied to whole genome sequencing (WGS) data, it has the great advantage that the data it requires is already generated as part of standard sequencing pipelines and requires no additional experimentation.

It is highly likely that other SVs may be present in our cohort but remained undetected as their breakpoints lay outside the exonic regions targeted by the capture panel. This was partly solved by the use of read-depth analysis instead, as was successfully applied in the case of the *USH1C* deletion, which had no exonic breakpoints, but this method struggles

to detect structural variants that do not span several exons. Despite these limitations, we have demonstrated the utility of this approach for IRD diagnostics by generating clinically actionable results even from past datasets, and we recommend that it be used as a ‘stopgap’ measure to improve diagnosis rates in similar projects before more comprehensive studies involving WGS can be performed.

Thus far, during the course of the study, genetic analysis of IRD patients has identified candidate mutations in approximately 72% of cases. The diagnostics rate obtained is in line with other NGS studies (5,68,71–73). The growing body of data from NGS studies of IRDs similar to this one should facilitate the formation of better correlations between genotype and phenotype. As research in parallel studies such as natural histories of IRDs (176,177) and functional analysis of modifier loci (174) continue, this information in conjunction with NGS data will undoubtedly contribute to improvements in detecting pathogenic genetic variants responsible for IRDs, as well as providing insights regarding prognoses for some IRDs and importantly may also facilitate the future delivery of gene-specific treatments to the applicable patient populations.

Non-coding variants such as splice-affecting variants, either proximal or distal to canonical splice sites, are also likely to represent a significant fraction of the unobserved disease-causing variants (95,175). Previous studies have identified deep intronic variants that lead to intronic sequence being incorrectly retained in the mature mRNA as relevant to IRDs (178). These variants are highly likely to be missed by current studies, as very few capture panels target introns, and the interpretation of deep intronic variants is complicated as these regions are less constrained by purifying selection, leading to large numbers of observed variants. Despite the successful identification and thorough

functional analysis of some pathogenic deep intron mutations such as *CEP290* c.2991+1655A>G (179), targeted whole gene sequencing has been reported with lower than anticipated detection rates for certain IRD genes (171).

Recent advances in functional analysis of intronic variants have helped to accurately assess their pathogenicity potential. This is primarily referring to the use of patient-derived photoreceptor progenitor cells to assess *in vitro* the aberrant splicing that may occur as a result of their observed mutation (174). Several antisense oligonucleotide studies have also shown to be very effective in correcting such variants in several IRD genes to date, including *USH2A* (180), *ABCA4* (181) and *CHM* (182).

In addition to recent direct observations implicating some intronic variants, strong indirect evidence of unobserved disease-causing variants in known IRD genes exists. Whole-exome studies have very similar detection rates to studies focused merely on IRD-associated genes (73), implying that coding mutations in unsequenced genes do not represent a large fraction of unobserved disease-causing mutations. Furthermore, recessive pedigrees that could not be solved in this study with a panel of 254 genes were significantly enriched for single mutations in disease-relevant genes, strongly suggesting the presence of second, as yet unobserved intronic mutations.

Despite the sources of as yet ‘missed’ variation causative of IRDs, the results of this study so far highlight the vast levels of genetic heterogeneity inherent in IRDs in the Irish population and the significant value of a target capture NGS-based genetic evaluation for diagnostic purposes providing candidate disease causing mutations in approximately 70%

of IRD patients by sequencing of exons of this panel of up to 254 IRD genes. This has been clearly exemplified by the clinical re-categorisation of the disease pathology for several patients (for example, RP as BBS), the value of detecting pathogenic large structural variants and the continued reanalysis of patient datasets for emerging, previously undetected common pathogenic variants (*ABCA4*, p.Asn1868Ile) all of which were driven by NGS-based genetic data analysis. Future and ongoing studies, with a particular focus on structural variants and non-coding disease-causing variants, are likely to increase mutation detection rates further and yield an even more complete picture of the genetic architecture of IRDs in Ireland.

4.2 Retinoschisis

This work is currently in review for publication with the title “Multimodal Imaging in a Pedigree of X-Linked Retinoschisis with a Novel *RS1* Variant”. The clinical work shown was conducted by Dr. Kirk Stephenson and Dr. David Keegan of The Mater Private Hospital, Dublin, Ireland and by Dr. Niamh Wynne and Dr. Paul Kenna of The Royal Victoria Eye and Ear Hospital, Dublin, Ireland.

4.2.1 Introduction

This chapter will focus on the discovery and follow-up of a novel variant detected as part of Target 5000. It is focused on the application of this process for a pedigree with X-Linked Retinoschisis (OMIM: 312700, *XLRS1*) and the clinical and genetic workup of these patients for potential new therapies and possible future participation in appropriate trials now emerging for this form of IRD.

XLRS1 is a rare IRD (1:5,000 – 25,000 (183)) due to variants in the *RS1* gene encoding the retinoschisin protein (OMIM: 300839, Xp22.1). *XLRS1* is a mutationally heterogeneous disorder with over 160 known missense and nonsense variants (184), the majority of mutations falling in exons 4-6, the structurally important discoidin domain of the retinoschisin protein (185). Congenital poor vision in males is due to macular schisis and results in outer retinal atrophy with age. As novel gene therapies are in clinical trials (186,187), to optimise options for patients, confirmation of the presence of an *RS1* variant is essential and moreover must be supplemented with various clinical assays to assess the stage of disease and potential suitability for treatment.

With wide variability in disease phenotype between and within families (188), clinical diagnosis can be challenging. Multimodal imaging has been an excellent addition for confirming disease phenotype and guiding molecular genetic testing. Optical Coherence Tomography (OCT) imaging may detect intraretinal cystic spaces (189), which may appear in multiple retinal planes. Fundus autofluorescence (AF) imaging can detect subtle patterns of change in the retinal pigment epithelium (RPE), which allows staging of severity. Microperimetry can delineate areas of residual function (190). Thus, multimodal imaging can be used to demonstrate maintenance of anatomy which is of diagnostic and prognostic value and also may determine suitability for novel treatments.

The *RS1* genomic location was first implicated as pathogenic using multipoint linkage analysis of Dutch families affected by XLRS1 (191). Locus localisation was later refined and the *RS1* gene was identified by positional cloning. Mapping and expression analysis led to the discovery of the then novel transcript (XLRS1) encoding the retinoschisin protein (192).

Retinoschisin, a 224 amino acid protein, is expressed in photoreceptors and bipolar cells in the retina and has been shown to be crucial for cell adhesion during retinal development (193). It has been suggested that retinoschisin may also be involved in mitogen-activated protein kinase signalling and apoptosis in the retina (193) and recently has been shown to influence Na/K-ATPase signalling and localisation (194). Employing cryo-electron microscopy, it has been discovered that functional retinoschisin forms a dimer of octamer rings comprising a hexadecamer. Given this intricate structure, a mutation that affects the folded structure of the protein is likely to inhibit the proper assembly of the 16-subunit oligomer (195,196) as evidenced by the large number of reported pathogenic mutations

(184). Pathogenic *RS1* mutations are typically associated with XLRS1, although some mutations show greater phenotypic variability than others, including greater ranges of onset age and severity of condition (197).

Candidate *RS1* variants have been detected in 16 patients across 7 pedigrees to date as part of the current Target 5000 study. The variant of focus in this chapter, *RS1* (NM_000330.3) c.413C>A,p.Thr138Asn, is the most prevalent across affected patients, mostly due to the exceptional participation of one family. This variant was deemed the candidate variant in a total of 10 affected retinoschisis patients across 2 pedigrees. Other novel *RS1* candidates found as part of this study include: c.416delA,p.Glu139fs (23); c.326+1G>A (23); c.626G>A, p.Arg209His.

4.2.2 Methods

Approval Process

Ethics and patient recruitment processes are outlined in Chapters 2.1-2.3. All study participants signed an informed consent form to participate. The 2 children were recruited via RVEEH while all adults were recruited via MMUH/MPH.

Clinical Phenotyping

Members of a large pedigree with a clinical diagnosis of XLRS were invited to participate given informed consent. Nine affected males (7 adults, 2 children), one carrier female and one unaffected male attended the recruiting hospitals as specified above to take part in the IRD registry.

Phenotyping included an ophthalmic and medical history, pedigree mapping, and dilated ophthalmic examination. Colour photography and autofluorescence (AF, Optos plc, Scotland) and spectral domain optical coherence tomography (SD-OCT, Cirrus HD-OCT, Carl Zeiss Meditec AG, CA, USA) was undertaken. Multimodal imaging was assessed, and the macular findings were categorised into their appropriate stages, assessed by eye and not entire patient.

Next Generation Sequencing

Six of the affected family members were analysed by NGS, in an attempt to determine any genetic elements that may contribute to the variability observed during clinical assessment. The proband was analysed with an exonic panel of 218 genes and the subsequent five affected family members were analysed for an exonic panel of 254 genes.

Polymerase Chain Reaction and Sanger Sequencing

To validate the *RS1* variant identified by NGS, an amplicon for direct Sanger sequencing was designed incorporating the variant. The sequence used for primer design was Human reference transcript NM_000330.3. Forward primer: 5'-GCAGATGATCCACTGTGCTG – 3'. Reverse primer: 5' - TTTCTTGGGAGGTGGAGATG – 3'. PCR procedures were performed, and sanger sequenced as described in Chapter 2.17. PCR products were purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) prior to sanger sequencing.

Protein Modelling

In silico generated 3D models of single subunit wildtype RS1 (NM_000330.3) and the p.Thr138Asn mutant were generated using Iterative Threading ASSEmbly Refinement,

I-TASSER (198). Polymer structures of RS1 were obtained from the Protein Data Bank (PDB, ID#3JD6) (199). The predicted effect of single point mutations on protein stability was measured using STRUM (200). Protein alignments were generated using Clustal Omega (201).

4.2.3 Results

Mean patient age of affected adult males ($n=7$) was 67 ± 5.38 years. The recruited carrier female (daughter) was 47 years of age. Mean age of the 2 affected grandsons was 13 ± 1.41 years (Figure 4.2.1). Mean visual acuity for adult affected males was LogMAR 0.78 ± 0.16 for right eyes and 0.74 ± 0.28 for left eyes. Two eyes were not assessable due to one case of dense cataract (amblyopic eye) and one enucleation (due to previous retinal surgery, presumed post vitreous haemorrhage or detachment). Only one of 14 adult eyes developed either a retinal detachment or haemorrhage previously (indeterminable). Mean visual acuity of affected grandsons was LogMAR 0.2 for right eyes and 0.3 for left eyes.

Macular changes varied in severity, which showed no correlation to patient age. A severity staging was proposed by Tsui and Tsang (202), outlined in Table 1. Of 12 assessable adult eyes, 1 was in Stage 1 (8%), 6 were in Stage 2 (50%) and 5 were in Stage 3 (42%). The predominantly advanced stage of disease here is likely due to advanced patient age. Both affected grandsons were in Stage 1 disease. Confirmation of this *RS1* variant's pathogenicity will theoretically be of most benefit for the younger generation or possibly an older patient with a milder phenotype as they hold the greatest potential for reversible macular changes. Corresponding images from this series of patients are in accordance with this and are shown in Figure 4.2.2.

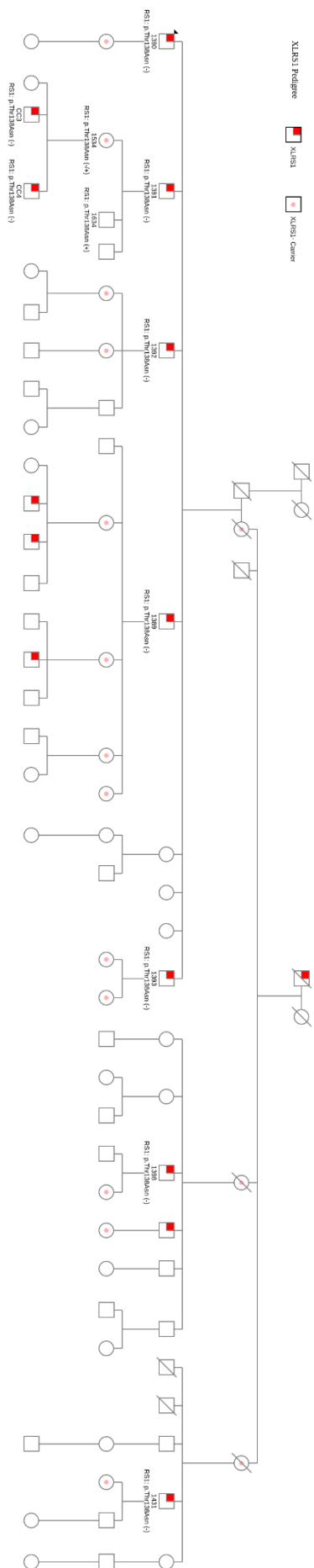


Figure 4.2.1. Pedigree of 5 generations of an XLRIS1 family. Individuals marked with a red square were confirmed to be clinically affected. Individuals within the pedigree tree have been annotated with their RS1 genotype where investigated.

Table 4.2.1. *The clinical staging of Retinoschisis (202)*

Stage	Colour	AF	SD-OCT
1	Macular schisis, classically radial (A)	Macular pigments displaced between schitic areas (B)	≥1 intraretinal schisis plane (C)
2	RPE pigment mottling (D)	Hyperautofluorescent ring at macula (E)	Collapse of schisis plane(s) (F)
3	Macular atrophy (G)	Large area of central hypoautofluorescence with surrounding hyperautofluorescent ring (H)	Outer retinal atrophy (I)

Figure 4.2.2 demonstrates the clinical staging of XLRS using multimodal imaging. Patient 1 (74y) demonstrates the features of Stage 1 (mild) disease with BCVA LogMAR 0.6. There are subtle pigmentary changes on colour photography (Fig 4.2.2A), with displacement of macular pigment leading to radial pattern of hyper- vs hypo-autofluorescence (Figure 4.2.2B). The most striking feature is on OCT with intraretinal cystic spaces (Figure 4.2.2C) in the inner plexiform and inner nuclear layers.

Patient 2 (60y) shows Stage 2 (intermediate) disease with BCVA of LogMAR 0.9. Colour photography reveals a hypopigmented ring at the macula, surrounding an area of normal pigmentation (Figure 4.2.2D). Autofluorescence confirms this as hypoautofluorescent with a surrounding ring of hyperautofluorescence and a preserved iso-autofluorescent foveal area (E). OCT reveals the absence/collapse of intraretinal cystic spaces but lacking complete outer retinal atrophy (Figure 4.2.2F). Incidental vitreo-foveal traction is noted in this case.

Patient 3 (71y) has stage 3 (advanced) disease with BCVA of LogMAR 1.0. The colour photograph demonstrates macular atrophy (Figure 4.2.2G), which appears as hypoautofluorescence with a hyperautofluorescent ring (Figure 4.2.2H). This corresponds to outer retinal atrophy without schitic intraretinal space on OCT (Figure 4.2.2I).

The mean age of affected adult patients in this cohort whose individual eyes were in each stage of XLRS was: Stage 1 74y (n=1), Stage 2 65.6y (n=6) and Stage 3 68.8y (n=5).

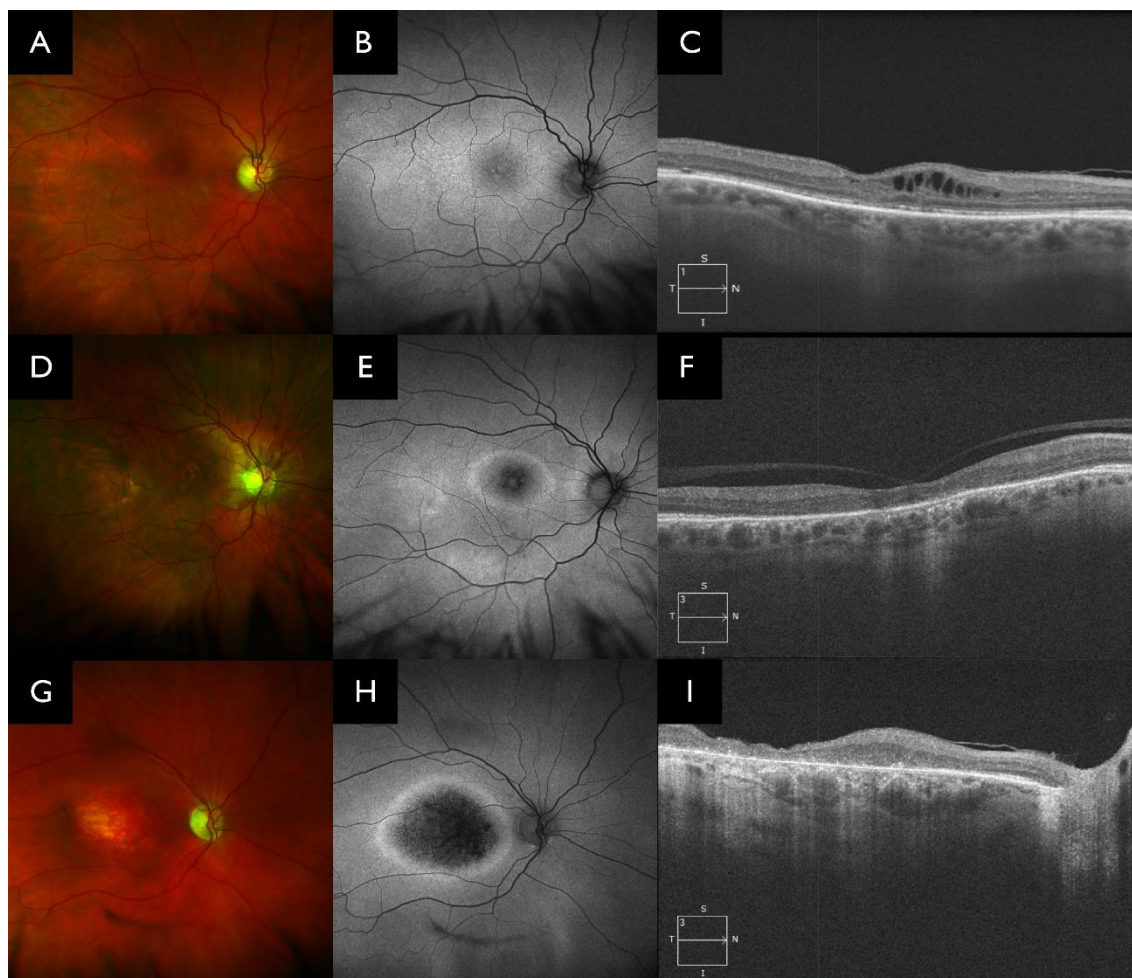


Figure 4.2.2. *Phenotypic features of selected cases from this XLRS1 pedigree. Colour photographs (left column), autofluorescence (middle column), and SD-OCT (right column) of the right eyes of 3 patients in Stage 1 (top row), Stage 2 (middle row) and Stage 3 (bottom row). Further description is available in Table 4.2.1 and the ‘Results’ section.*

The exonic regions from a panel of 218 IRD genes were sequenced by NGS in the proband from this pedigree. An additional 5 affected males were sequenced with a subsequent panel of the exonic regions of 254 IRD genes. Analysis of this NGS data led to the identification of a novel missense variant in the *RS1* gene within the pedigree, c.413C>A, p.Thr138Asn (93). This variant falls within exon 5 of the discoidin domain (185). This novel variant was chosen for further investigation as it was observed to segregate with disease in this family and was detected in a gene that is strongly associated with the observed phenotype (OMIM #312700).

No other candidate mutations were detected as part of the NGS investigation. 5 of the 6 family members who were examined by NGS were found to harbour a single *BBS10* (NM_024685.3) c.411G>T, p.Gln137His variant. This gene is associated with recessive Bardet-Biedl syndrome (OMIM #610148) and was deemed inconsequential as a second variant was not detected and no phenotypical manifestations associated with this condition were reported in any of the investigated patients. Further analysis of each of the members of the family by PCR amplification of *RS1* and Sanger sequencing demonstrated that each affected male was found to have a hemizygous nonsynonymous mutation in exon 5 of the *RS1* gene (NM_000330.3:c.413C>A, p.Thr138Asn). In addition, the single carrier female analysed was heterozygous for the same *RS1* mutation, while the unaffected male was hemizygous for the reference base at that position (Figure 4.2.3).

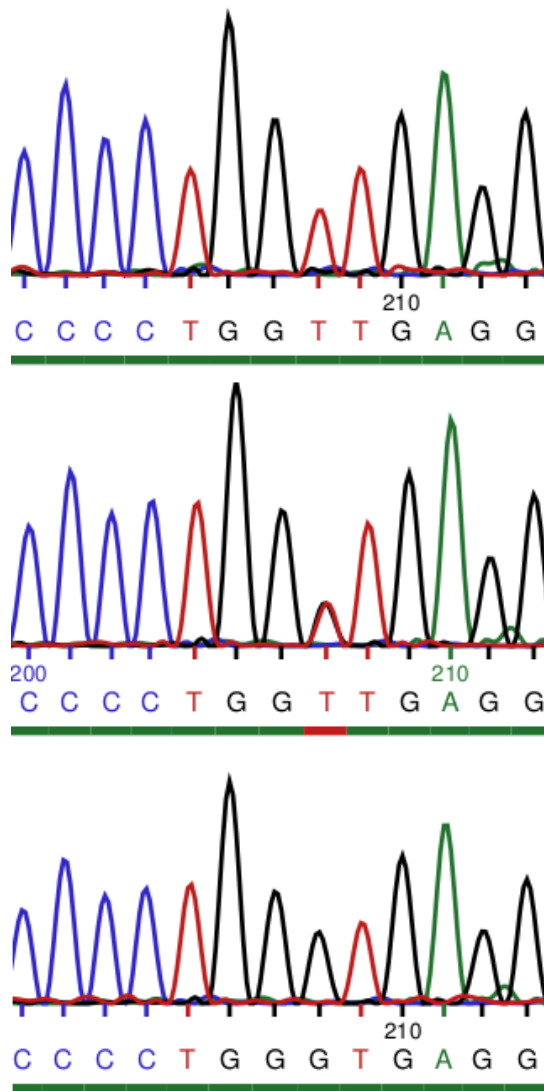


Figure 4.2.3. Sanger sequencing of RS1-exon 5 polymerase chain reaction (PCR) products from members of the pedigree. Top: Sanger sequencing result for affected male (ID# 1392) showing a T nucleotide trace at sequencing trace position 208. Middle: Sanger sequencing result for carrier female (ID#1534) showing a heterozygous result at sequencing trace position 207, revealing traces from both signals of wild-type G allele and the mutant T allele. Bottom: Sanger sequencing result for unaffected male (ID# 1634) showing a G nucleotide trace at sequencing trace position 208.

The p.Thr138Asn *RS1* mutation involves a polar to polar amino acid substitution. Given that this novel, previously unreported mutation is located within the critical discoidin domain of the retinoschisin protein, where many pathogenic mutations have been identified previously, additional *in silico* analysis was undertaken to explore the potential pathogenicity of this *RS1* mutation on protein structure and conformation.

The destabilising effect of this mutation as predicted *in silico* is shown below in terms of a prediction of the changes to fold stability (Table 4.2.2), where a negative delta delta G (ddG) score is indicative of a destabilising effect. This change is also represented visually as theoretical structural changes to the retinoschisin monomeric protein structure (Figure 4.2.4). Notable secondary structural changes are identified with arrows in Figure 4.2.4. The complexity of the hexadecameric structure of retinoschisin is shown in Figure 4.2.5 for reference.

Table 4.2.3 displays the comparative output scores for a known pathogenic *RS1* mutation and this novel mutation from three highly regarded variant pathogenicity prediction tools. As seen in the respective columns of the tools, both variants have been deemed likely pathogenic by all three tools. Also, the novel variant described in this study is scored more likely to be pathogenic by two out of three of the methods listed.

Table 4.2.2. Output results of the STRUM program (200). The change in stability scores between the mutant and wild type RS1 protein is given as delta delta G (ddG)

Position	Wild Type	Mutant Type	ddG
138	T	N	-0.78

Table 4.2.3. A direct comparison of the novel mutation, p.Thr138Asn, and a known pathogenic variant in RS1. Both variants were analysed with the same bioinformatic pipeline and have been scored by the same computational tools.

Variant	Nucleotide	Protein	DBSNP ID	MetaLR	M-CAP	REVEL	ClinVar Report
Novel	c.413C>A	p.Thr138Asn	None	0.9742	0.8775	0.929	None
Known	c.636G>A	p.Arg209His	rs281865362	0.9446	0.9233	0.796	Pathogenic

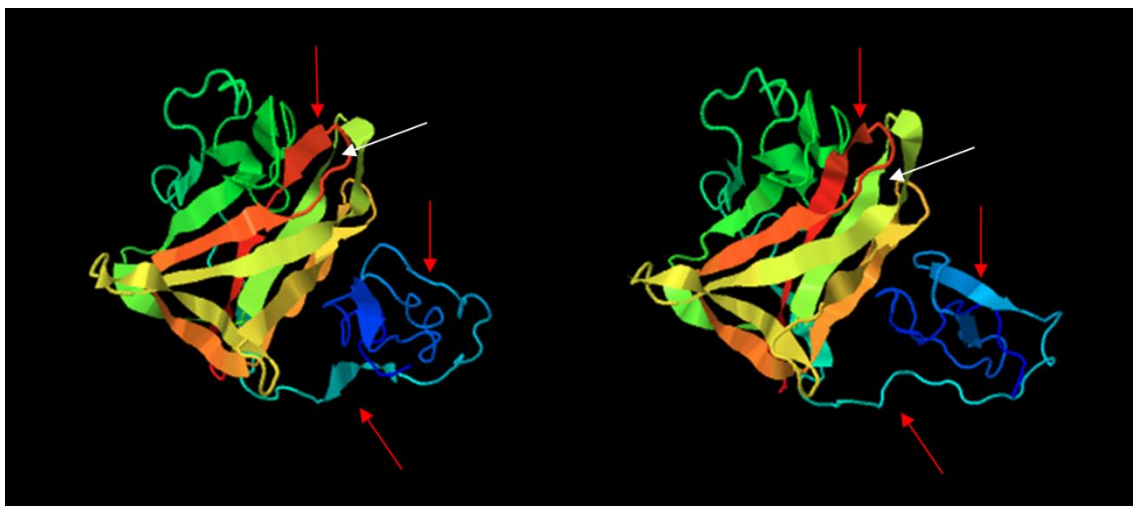


Figure 4.2.4. Protein models of reference RS1 (left) and mutant T138N (right) proteins generated by the I-TASSER program. Red arrows denote obvious structural changes to protein structures. A white arrow highlights the position of the amino acid change at codon 138.

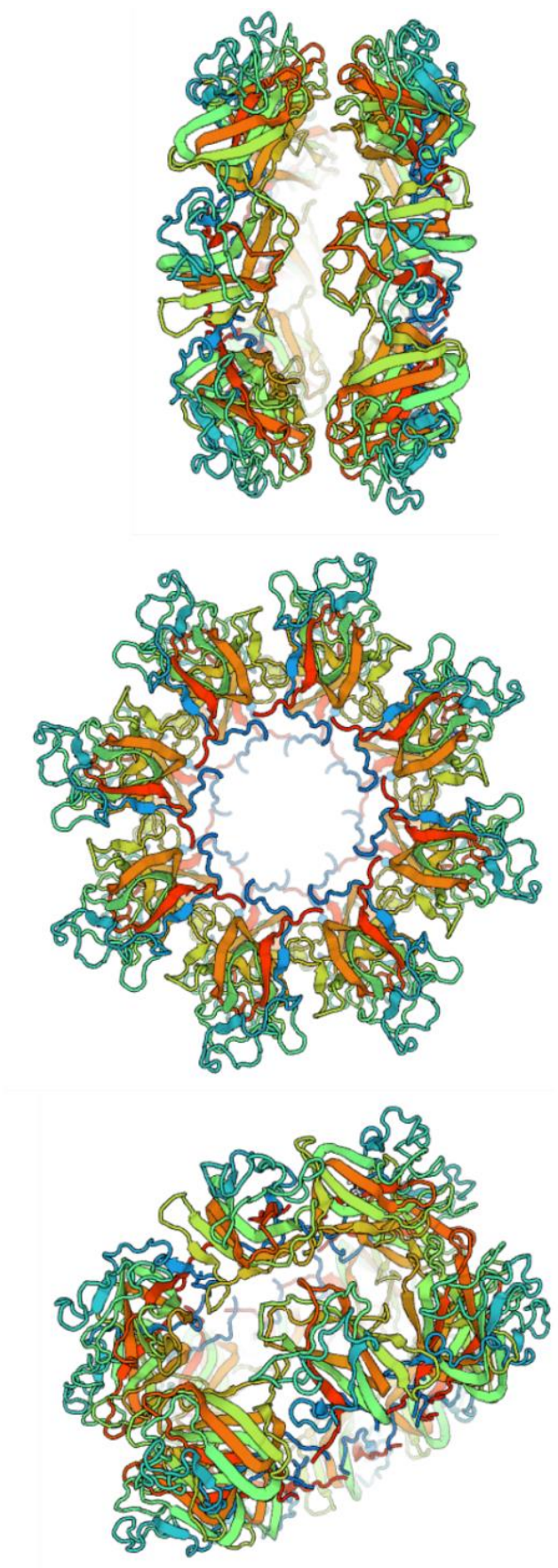


Figure 4.2.5. Different views of the back-to-back octamer rings of wild type RS1 obtained from the Protein Data Bank. (Left) Objective view. (Middle) Plan view. (Right) Perspective view.

The ACMG guidelines (144) have also been implemented to classify the novel variant. The relevant information has been outlined in Table 4 along with supporting evidence where appropriate (193,195,203). The guidelines state that there is sufficient evidence to classify this variant as pathogenic as there is at least one strong (PS4) and three moderate (PM1, PM2 and PM5) lines of evidence.

Table 4.2.4. *The ACMG guidelines applicable to the novel variant in question, RS1 p.Thr128Asn.*

Category Code	Qualifying Criteria	Relevance to Case
PS4	The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls.	This variant is found in a sufficient number of relatives to qualify it as strong evidence of pathogenicity (11).
PM1	Located in a mutational hot spot and/or critical and well-established functional domain (e.g. active site of an enzyme) without benign variation.	Located in well-established functional domain as reported by several studies (14,23).
PM2	Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes or ExAC.	Absent from every databased queried.
PM5	Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before.	Novel mutation at same amino acid as previously reported p.Thr138Ala (HGMD Database) (24).
PP1	Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease.	Segregation in a family in a gene known to cause disease.
PP3	Multiple lines of computational evidence support a deleterious effect on the gene or gene product.	Multiple lines of computational evidence presented here (Table 3).
PP4	Patient's phenotype or family history is highly specific for a disease with a single genetic aetiology.	Patient phenotype is highly specific for a disease.

4.2.4 Discussion

Retinoschisin has been shown to be an essential protein required for faithful retinal adhesion (185). There is some debate over which retinal plane is affected by the schisis, which has been documented on OCT at the level of the photoreceptor inner segments, the inner nuclear layer and the nerve fibre layer (204,205). Decreased retinoschisin protein function may, and likely does, cause multiple planes of schisis.

There is significant variability in the disease phenotype between the two eyes of the same individual, between different members of this XLR1 pedigree and, as previously documented, between unrelated individuals (188,197). This variability appears greater with age with some patients progressing through the clinical stages at different rates, possibly due to acquired/environmental factors. It has been noted by several other studies that age clearly contributes to the slowly progressive phenotype (206–208). Although this variability is noted, there are clinical characteristics that can guide the clinician to this diagnosis (e.g. pedigree, peripheral retinoschisis, lack of subretinal flecks or vitelliform changes) which are supported by a likely-causative *RS1* variant and lack of other known pathogenic genes. This supports the bimodal presentation of severity associated with XLR1 (209).

Severe cases of the disease (approximately 30%) are detected in infancy with poor fixation and/or strabismus. Milder cases (approximately 70%) are detected at school entry or in adulthood (210). This series does highlight that the underlying genetic variant on the *RS1* gene is not predictive of phenotypic severity for this, or indeed any of the other, genetic variant(s) that have been described (197). Thus, in assessing patients for

trial/intervention suitability both a genetic and full phenotypic analysis is required. For example, gene therapy will not be beneficial if there is significant structural change (i.e. stage 3). However, the combination of a stage 1 or 2 case with confirmed genotype would suggest a possible role for genetic manipulation/replacement. Our results also indicate that patient selection for such a therapeutic should not be restricted by age as we have observed phenotypes ranging from mild (Patient 1) to severe (Patient 3) in age-matched patients of advanced age (Figure 4.2.2). As our knowledge of the impact of multiple variants increases we may better predict the molecular effects of each on retinal structure and function thus tailoring our approach to intervention further.

The clinical findings may be subtle in early stage or mild forms of disease. Classically, radial cystic maculopathy at the fovea is seen in 98-100% of cases (211) (Fig 4.2.1A-C). Ancillary tests have proven useful, particularly in subtle cases. An electronegative electroretinogram (ERG) is suggestive of, but not specific, to XLRS1 (212) as this feature may be seen in acquired disease of the inner retina (e.g. central retinal artery occlusion). Microperimetry may assist in detection of subclinical disease (190). Late stage disease is a diagnostic challenge; however, clinical examination, family history and genotyping as outlined above will aid in differentiating the diagnosis.

Colour photography, Autofluorescence and Optical Coherence Tomography are non-invasive and accessible tests useful in determining the clinical stage of XLRS1. This is relevant, for current and upcoming gene therapy trials (186,187), in selecting those cases most likely to benefit from intervention (i.e. early to intermediate disease, Stage 1/Stage2

(211)). Ancillary tests, such as adaptive optics (AO) imaging and microperimetry, can complement assessment of cone structure/spacing in those who are deemed early/intermediate stage (213,214).

To our knowledge, this specific mutation is currently unreported in any clinical database available including the Leiden University's Open Variation Database for Retinoschisis (LOVD 3.0 (215)) and represents a novel and likely pathogenic mutation given the clear segregation in this pedigree and the significant predicted effects on protein structure.

Recent studies have outlined the importance of certain key residues in correctly assembling a functional RS1 oligomer. Several key residue positions necessary to ensure proper folding of the retinoschisin protein have been identified (195). Of note, amino acids at positions 137 and 139 were highlighted in that study as fundamental to forming a beta-sheet required in the interaction between subunits essential in the formation of their functional octamer ring structure. It is possible therefore, that an amino acid substitution at position 138 may provide some steric hindrance, that in turn, would interfere with the relationship between different subunits on the same octamer ring.

Such interference could in principle have disastrous consequences for the formation and stability of the oligomeric RS1 complex. This hypothesis is further supported by *in silico* analysis of the p.Thr138Asn mutation. Computational analysis was utilised to predict protein folding changes in the mutant protein (using I-TASSER); physiochemical properties, position specific conservation and secondary structure prediction scores were used to determine the destabilisation effect(s) of the protein substitution at position 138 (using STRUM) where significant changes in the above were predicted for the

p.Thr138Asn mutation. It is evident that the variant of focus in this study does affect the predicted protein folding configuration and secondary structures as illustrated in Figure 4.2.4.

The prediction scores in Table 4.2.3 are incredibly useful for assessing the confidence in which a variant can be called deleterious or not. However, they are confidence scores and not pathogenicity scores, so although the scores listed would be generally considered to be quite high, these scores cannot indicate the detrimental effect that this mutation may have relative to a similar amino acid substitution elsewhere in the protein. Increased bioavailability of functional retinoschisin protein may correlate to the lasting preservation of the retinal layers however to assess this dosage effect, each variant would have to be tested individually to be assessed for function and half-life.

This is a particularly difficult task for retinoschisin as it has traditionally proven to be quite difficult to produce in useful quantities (195). Currently we must rely on *in silico* predictors of a variant's effect on protein structure. This includes protein modelling (Figures 4.2.4 and 4.2.5) where possible and also ddG scores (Table 4.2.2), to predict the destabilising effects that a mutation may have on a protein. However, when all lines of evidence are taken together, this variant is classified as pathogenic under the guidelines outlined by the ACMG (144) (Table 4.2.4).

In XLRS1, age is not an absolute factor, particularly in adulthood, when evaluating severity of disease and the likely benefit from novel gene therapies. The individuals in this XLRS1 pedigree exhibited no clear correlation between patient age and 1) stage of disease, 2) visual acuity, 3) central retinal thickness or 4) age at onset despite a shared

novel *RS1* gene variant. Advanced cases can prove to be a diagnostic challenge as many inherited and acquired macular diseases follow a final common pathway of outer retinal atrophy. Each case must be judged on its potential (anatomically within Stage 1 or 2 and genetically possessing a pathogenic *RS1* variant) to respond to novel therapies. Furthermore, the variability in clinical presentations between family members with the same mutation suggests the possible involvement of genetic modifier loci and or environmental effect(s).

While there is variability in phenotype between relatives affected by the same genetic variant, the significance is that the clinician should not be discouraged from the diagnosis due to these differences but rather consider them within the spectrum of the 3 described clinical stages. Confirmation of the extent of severity (clinical stages) and a pathogenic *RS1* variant open a potential avenue of treatment for those in Stage 1 or 2. This information then allows appropriate genetic testing for relatives at risk with the associated possibility of disease-modifying treatment.

There is a possibility that there may be some form of modifier or digenic effect which could account for the observed phenotypical variability. With this in mind, the coding regions of all known genes associated with retinal degenerations were subjected to sequence analysis in six affected individuals in this pedigree. The presence of multiple pathogenic variants within a single patient is extremely rare but has been previously observed, even in our own cohort. Although, as previously mentioned, this method unfortunately did not detect any likely genetic cause of phenotypic variability between the affected patients in this pedigree. Of note, this analysis was limited to known IRD-associated genes and any potential genetic modifiers located elsewhere in the genome

would not have been detected. It is also important to acknowledge that a person's environment may also be unique. In this way, the hazards and toxins a person may be exposed to can modulate the manifestation of a condition.

Linkage analysis in mouse models of XLR51 have identified a region, named by the authors as *Mor1*. This chromosomal region has been found to harbour a variant capable of nullifying the detrimental effects of retinoschisis haploinsufficiency (216). To the best of our knowledge, this effect has not yet been investigated with respect to human genetics.

The retinoschisin protein is widely expressed throughout the layers of the retina in childhood and more specifically in rod and cone photoreceptor cells in adulthood and is seen to act as a cellular adhesive when functioning correctly. However, other proteins are also capable of fulfilling similar adhesive functions, for example adhereins and cadherins in adherens junctions or claudins and occludin in tight junctions. Given that some of these other proteins belong to large protein families, it is a possibility that variation in relative composition of these cellular adhesive protein cocktails may have a more notable contribution to a variable phenotype in the absence of functional retinoschisin.

This variability creates a diagnostic challenge in advanced XLR51 disease. The clinical clues in conjunction with adjunctive testing, confirmed by robust genetic assessment (i.e. outruling other causes and confirming a pathogenic XLR51 variant) confirm the diagnosis, thus empowering those affected by providing them with a diagnosis and in principle, bringing them closer to accessing novel gene therapies and preparing younger generations for those potential therapeutic options.

The diagnostic value of ERG assessments, in particular the characteristic decrease in b-wave activity in combination with suggestive clinical findings are powerful tools for diagnosing younger patients. However, the diagnostic value of these tests can decrease with patient age. In this chapter of the thesis the utility of NGS and sanger-based genetic analysis is described together with the value of multimodal imaging in this condition to A] confirm diagnosis, B] monitor progression, C] assess clinical stage and thus relevance to novel therapeutic options even in patients of advanced age. Autofluorescence and OCT are non-invasive, well-tolerated and readily available in most clinical centres, sadly in Ireland this is not the case for genetic testing. Standardisation of clinical assessment and staging criteria, perhaps in conjunction with a central European reading centre, would allow detection of individuals with modifiable XLRs1 and guidance towards clinical trial participation. In parallel, an increase in the availability of genetic testing for IRD patients will be essential to provide patients with a clear diagnosis and optimal clinical management programme. The results described in this chapter clearly highlight the value of studies such as Target 5000, a number of which are on-going, particularly in Europe and the US (23,68–82,93) (Chapter 1: Table 1.1).

4.2.5 Concluding Remarks

In this study NSG-based and Sanger-based sequencing have resulted in the identification of a novel segregating variant in the RS1 gene in a large XLRs1 pedigree. Detailed clinical and genetic analysis have been undertaken on the members of this pedigree and the associated novel variant providing strong evidence that this previously unreported RS1 variant is the likely pathogenic mutation in this family. In XLRs1, potentially

reversible macular changes can be detected with the use of multimodal imaging as described above. This, in combination with confirmation of a pathogenic *RS1* variant, can detect individuals with this condition who may be able to benefit from novel gene therapies which are currently in clinical trials. Of note the study highlights the inherent value of studies such as Target 5000 in helping to define the diagnosis, prognosis and treatment plans for IRD patients.

4.3 Case Studies

4.3.1 Chapter Introduction

The results presented in this chapter of the thesis are focused on the discovery and follow-up of various variants that were detected as part of Target 5000 in IRD patients and family members. As this is a selected sample set of case studies from many, these cases will range from a pedigree-based follow-up and segregation of known pathogenic variants in IRD genes, to the assessment of potentially pathogenic novel previously unreported variants and indeed the progression of members of one IRD pedigree into a small molecule clinical trial.

The research discussed in this chapter emphasises the importance of the additional body of work required after variant detection. This work includes segregation analysis and feedback to referring ophthalmologists, both key components in determining diagnoses correctly.

4.3.2 Chapter Methods

Details of patient recruitment and assessment can be found in Chapters 2.1-2.3. Sample collection and preparation is outlined in Chapters 2.4-2.6, 2.8-2.9 and 2.16. Variant verification is described in Chapter 2.17. Detail of sequencing the ORF15 region of *RPGR* can be found in Chapter 2.18.

4.3.3 Case 1: Routine Rhodopsin

The first autosomal dominant RP gene was localised using genetic linkage studies to human chromosome 3q in a large Irish pedigrees almost 30 years ago and subsequently pathogenic rhodopsin variants were discovered in IRD cohorts (217–219). These discoveries led to the recruitment of several additional large Irish and European IRD families who became involved in a number of subsequent linkage studies and analyses (150,151,220–222). These studies helped to determine many of the pathogenic *RHO* variants that have been since detected worldwide. Given rhodopsin's predominant association with autosomal-dominant inheritance, variants in this gene can often involve a more simplistic verification and segregation process given the presence of multiple affected individuals frequently in such families.

Although our detection rates of *RHO* variants are likely to be underrepresented in our Target 5000 statistics, due to the prior involvement of large Rhodopsin families in linkage studies which were then excluded from Target 5000, many *RHO* variants are still found

in other pedigrees deemed to be unrelated to the previously investigated pedigrees. Some of these variants, namely c.533A>G (p.Tyr178Cys), c.541G>A (p.Glu181Lys), and c.620T>G (p.Met207Arg), are shown in Figure 4.3.1. These three variants were collectively detected across 12 different pedigrees to date. Figure 4.3.1 illustrates how several different variants can be detected by a single optimised PCR reaction. As is the case in Figure 4.3.1, whereby a single optimised reaction encompassing a region of approximately 500 bp can span at least three pathogenic variants. Figure 4.3.1 also illustrates how three different proximal variants can differentially impact the theoretical folding of the resulting protein domain as generated by the I-TASSER suite (198).

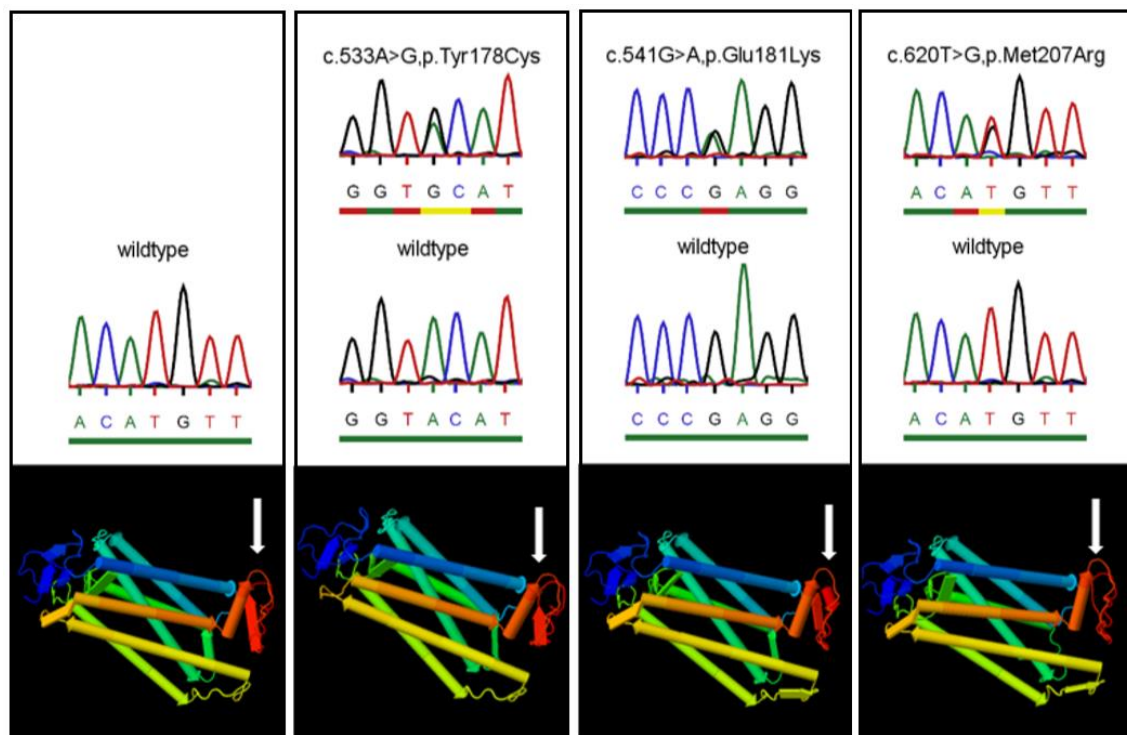


Figure 4.3.1. *Mutational analysis of various pathogenic RHO variants detected as part of Target 5000. Each mutation is shown above its associated predicted model with affected regions highlighted by a white arrow.*

Rhodopsin belongs to a superfamily of 7 transmembrane G-protein–coupled receptors. Rhodopsin is comprised of an apoprotein (opsin) that is covalently linked to a conjugated chromophore (11-*cis*-retinal), which is derived from vitamin A. When light is absorbed, the chromophore isomerizes into all-*trans*-retinal, with subsequent rhodopsin conformational changes leading to the activation of one of the primary phototransduction cascades in rod photoreceptors (223).

Variants surrounding the *RHO* p.Met207 region form part of the fifth transmembrane domain of rhodopsin. This domain has been shown to be vital to retinal binding in bovine rhodopsin (224). It is likely that undesirable changes in the protein folding of this region result in steric hindrance that prevents the apoprotein from successfully binding retinal. This impaired binding capacity is currently the reasoning for the pathogenicity of the aforementioned three variants located in this genomic region and found in this protein domain.

Rhodopsin makes for a challenging target for gene therapy as it is so highly expressed in the retina (225). However, positive results have been reported when targeting the endogenous rhodopsin for suppression and simultaneously replacing it with a codon-optimised version of *RHO* (226,227).

4.3.4 Case 2: *FLVCR1*, A Wolf in *NR2E3*'s Clothing.

This IRD case is focused on the application of our segregation process for a pedigree in which two genetic sources of IRD are present, exhibiting variants in both the *NR2E3* and *FLVCR1* genes. Mutations in both genes can cause IRDs which fall within the spectrum of Retinitis Pigmentosa (RP: OMIM #268000). Presented in this section of the chapter is the clinical and genetic examination of these patients and the implications that these results may have for potential application of new therapies and the future participation of patients in potential clinical trials that may emerge for these forms of IRD.

Feline leukaemia virus subgroup C cellular receptor 1 (*FLVCR1*: OMIM #609144) is a transmembrane protein involved in erythropoiesis and heme transport (228). *FLVCR1* was first discovered for its role in aplastic anaemia in domestic cats and subsequently erythroblast destruction in vitro (229). This was also the first study that theorised that *FLVCR1* was a receptor for an organic anion and identified its component domains as strikingly similar to other ancient Major Facilitator Superfamily (MFS) members. In a later study it was shown this receptor protein provided the first description of a mammalian heme transporter (230). The same group proceeded to investigate this effect in knockout mice models, which resulted in embryonically lethality (228).

Initially, aberrations in this gene have been associated with a neurological syndrome, posterior column ataxia with retinitis pigmentosa (PCARP: OMIM #609033) (231–233) and more recently a specific splice variant (c.1092 + 5G>A) has been reported multiple times to be associated with non-syndromic RP (71,234,235). In the current study we

report the first evidence of a protein coding *FLVCR1* variant (c.1022A>G, p.Tyr341Cys) being implicated in RP.

RP is an inherited retinal degeneration that is clinically identifiable by the presence of nyctalopia, constriction of peripheral visual field and typically a progressive rod-cone photoreceptor degeneration (OMIM #268000). PCARP is a systemic condition that presents as RP in the eye but is also associated with impairments of the nervous system, connective tissue, musculature among other pathologies (236). RP is the most common clinical diagnosis for participants in the Target 5000 study, the clinical diagnosis of RP accounts for nearly 40% of total pedigrees enrolled to date (23).

Figure 4.3.2 depicts the direct sequencing verification of the novel *FLVCR1* variant as well as the population frequencies as found in the gnomAD database (141). The population frequencies emphasise the rarity of this variant globally as it was undetected in the majority of populations sampled. Figure 4.3.3 illustrates the pedigree tree where the novel variant was first detected. Five members of this family presented with the initial clinical diagnosis of RP. Upon sequencing of the proband, two variants were discovered in the *NR2E3* gene. Confirmation testing was then undertaken for the other 4 affected members and any unaffected members of the family willing to participate. The remaining affected patients in this pedigree were not found to have compound heterozygous *NR2E3* like the proband, although of note, they were each homozygous for the novel *FLVCR1* variant.

The genetic findings were reported back to the referring clinical team. The referring ophthalmologists subsequently reviewed the fundus imagery, noting a detectable

difference in the proband's retinopathy compared to the other RP patients within the same family. These findings are presented in Figure 4.3.4.

Figure 4.3.4 typifies the phenotype observed with the homozygous *FLVCR1* p.Tyr341Cys genotype and compares it to the phenotype seen in the proband (compound heterozygous *NR2E3* variants). The *FLVCR1* genotype appears to present with a phenotype that resembles classical RP, with characteristics such as masses of bony spicules, attenuated blood vessels, waxy disc pallor and preserved maculae. The *NR2E3*-related phenotype however, illustrates the nummular deposits in the mid-periphery, healthy optic disc and normal blood vessels.

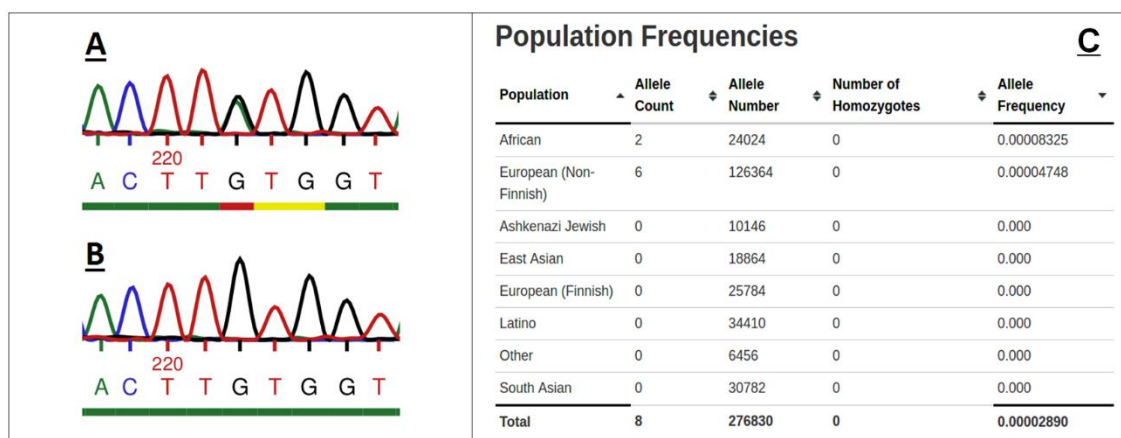


Figure 4.3.2. Direct sequencing of the novel *FLVCR1* variant, c.1022A>G, p.Tyr341Cys and current global population frequencies of the variant. A) Confirmation sequencing of the variant in the proband, verifying their status as heterozygous for the mutation. B) Confirmation sequencing of the variant in the maternal aunt of the proband, verifying their status as homozygous for the mutation. C) Population frequencies for the c.1022A>G, p.Tyr341Cys variant as they appear in the gnomAD database. This emphasises the rarity of the variant globally but notably the complete absence of detection of this mutation in all but the African and (Non-Finnish) European cohorts.



A black arrow indicates the proband in the F2 generation. Genotypes are depicted above only where they differed from the reference sequence. All available members of this pedigree have had confirmation sequencing for all three variants of interest.

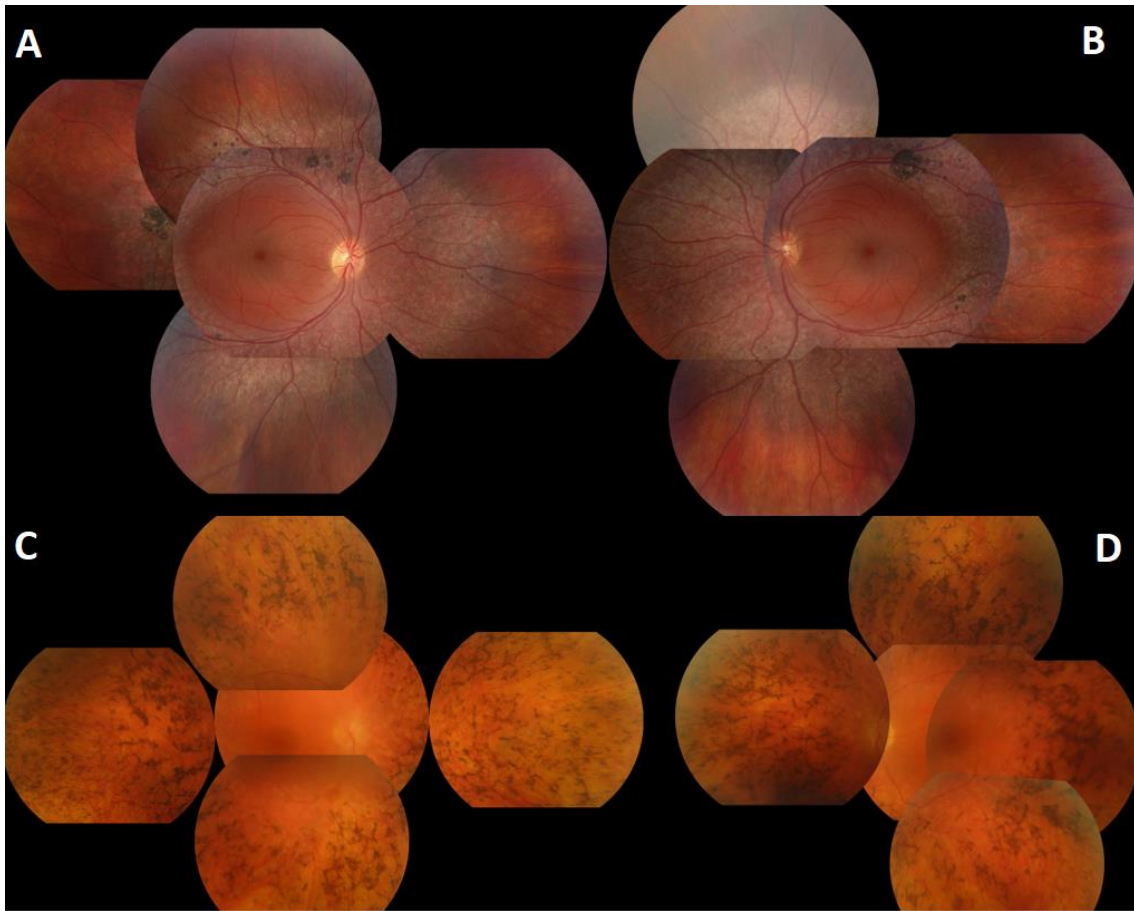


Figure 4.3.4. Montage of fundus images from RP spectrum patients with differential genotypes. The *NR2E3*-related phenotype (A- right eye and B- left eye) demonstrates the nummular, coin-like, deposits in the mid-periphery. This phenotype also shows healthy optic disc and normal blood vessels. The *FLVCR1* genotype results in a phenotype (C- right eye and D- left eye) that appears to bear a resemblance to classical RP, with features such as masses of bony spicules, attenuated blood vessels and waxy disc pallor. Both phenotypes seem to involve relative preservation of the maculae, more prominently seen in the *NR2E3* phenotype.

Since its initial detection, the novel *FLVCR1* variant p.Tyr341Cys, has been observed and deemed to segregate with the condition in two additional pedigrees. Each of the affected individual in these pedigrees are homozygous for the variant and no compound heterozygotes involving this variant have been detected yet as part of this study. It is possible that these seemingly distinct families share a common ancestry.

It is notable this variant is only the second disease-associated variant to be found in *FLVCR1* and to associate with RP and without posterior column degeneration (235). It is also the first protein coding variant found in this gene to be affiliated with non-syndromic RP. The age of onset for symptoms of ataxia has been reported as typically in the third decade of life, in our Target 5000 cohort, the oldest patient to present with this novel variant is currently in her sixth decade of life. It is unclear yet as to whether the disease pathology associated with this variant will remain completely non-syndromic throughout the entirety of a patient's lifespan or will it result in a milder, later onset of additional symptoms compared to other pathogenic variants found in this gene.

Fortunately, the small size of this gene (2.6 kb) in principle makes it suitable for inclusion in AAV vectors for use as a gene therapy. Such a therapy may help to alleviate the toxic effects of intracellular free-heme (231). The serotype of AAV could be chosen based on the presence or absence of systemic phenotypes. For example, if only the retina was to be targeted, an AAV 2/5 or 2/8 might be ideal, on the other hand, if the whole central nervous system was to be targeted then AAV-B1 would be the ideal candidate (237). Although many other bespoke serotypes are now emerging as a result of site-directed evolution of AAV serotypes (238–241). Given the recent results here on the role of *FLVCR1* in this

form of non-syndromic RP, *FLVCR1* becomes an increasingly interesting candidate for exploration of an AAV-mediated gene therapy.

4.3.5 Case 3: Using Genotyping to Refine Stargardt Prognosis.

Bull's eye maculopathy represents a small percentage of macular degenerations. It is a condition commonly characterised by concentric annular macular dystrophy, essentially a visible ring-shaped pattern of damage around the macula. This phenotype is strongly affiliated with the long-term use of drugs for autoimmune conditions such as arthritis or lupus, namely chloroquine or hydroxychloroquine (242,243).

Importantly, some genetic element is also suspected to be involved- traditionally this condition, when inherited, has been considered to be autosomal dominant in inheritance (244). However, more recent evidence has indicated that a specific variant in the *ABCA4* gene, c.5882G>A,p.Gly1961Glu, is associated with recessive Stargardt disease phenocopying bull's eye maculopathy (245). The variant appears to dominate the resulting phenotype when found as a compound heterozygote with other known pathogenic variants in *ABCA4* (245). The condition is reported to present initially as the formation of parafoveal lesions which over time result in a sudden parafoveal collapse which can be seen as thinning of the parafoveal region. Despite this collapse, there is significant foveal sparing associated with this phenotype which ultimately results in a more stable, milder condition (245).

Primed with the knowledge of this association, the clinical records of IRD patients who tested positive for *ABCA4*, p.Gly1961Glu Stargardt disease were probed to see if the same

could be found true for patients in our IRD cohort. Many patient records queried noted not only the distinctive Bull's eye pattern, but also a relatively later onset, with first incidence of glare reported in third decade of life with visual acuity and colour vision impairment in the subsequent decade. Progression of the condition was also typically reported as subjective. When the status of the periphery was commented on in reports, most observed reports suggested peripheral preservation of the retina. Interestingly, some reports also suggested a differential diagnosis of atrophy due to drug toxicity (data not shown).

In our database, a total of 11 patients with multiple pathogenic *ABCA4* variants carried at least one copy of the *ABCA4* p. Gly1961Glu variant. Many patients who were found to harbour this variant also carried two additional *ABCA4* variants (Table 4.3.1). This additional variation may contribute to the atypical severity observed in some patients (for example, Patient #2, Table 4.3.1). The association of this particular phenotype with the *ABCA4* p. Gly1961Glu variant was evident in 5/8 available reports with the remaining three reports noting abnormalities of the foveae and maculae of these patients.

It is evident that as genotyping studies expand globally, more associations can be made to assist in the refinement of clinical diagnoses. This particular phenotype represents a well-established association with drug toxicity which may mask underlying genetic causation. This phenocopy issue may also erroneously impact patients who suffer from long-term autoimmune conditions. Standard procedures call for the immediate cessation of chloroquine or hydroxychloroquine use if this phenotype is observed (246).

Currently it is believed that retinal toxicity occurs in 2% of patients in the first ten years of drug use but rises quickly to 20% with 20 years of use (247). This phenocopy dilemma represents a potential use of *ABCA4* pre-screening to prevent the unnecessary termination of medical intervention for lifelong autoimmune conditions. This issue also highlights the value of thorough clinical and genetic assessment, including accompanying conditions and current or previous use of medication.

It is also possible that the combination of genetic susceptibility and drug exposure can account for the varying levels of severity observed in patients using chloroquine or hydroxychloroquine. Noupou *et al.* have theorised that the basic pH of both chloroquine and hydroxychloroquine may be sufficient to disrupt the degradative capacity of lysosomes in the RPE space and that the suboptimal clearance of photoreceptor outer segments ultimately leads to the formation and aggregation of lipofuscin, a hallmark trait of Stargardt disease (245,248).

In conclusion, the study highlights the real value in obtaining detailed clinical and genetic histories and that only upon detailed evaluation of a patient's retinal and non-retinal disorders and associated drug regimes, will we progress towards precision medicine, and indeed in this case, towards precision ophthalmology.

Table 4.3.1. A tabular representation of patients found to harbor the ABCA4, c.5882G>A,p.Gly1961Glu variant and its clinically reported association with a bull's eye maculopathy phenotype.

Patient	Clinical Diagnosis	Bullseye Phenotype	Variant 1	Variant 2	Variant 3
1	Stargardt	Yes	c.5882G>A,p.Gly1961Glu	c.4139C>T,p.Pro1380Leu	
2	Advanced Macular Dystrophy	Abnormal	c.5882G>A,p.Gly1961Glu	c.6098T>G,p.Leu2033Arg	c.2549A>G,p.Tyr850Cys
3	Stargardt	Yes	c.5882G>A,p.Gly1961Glu	c.3481C>A,p.Arg1161Ser	c.5603A>T,p.Asn1868Ile
4	Stargardt	Yes	c.5882G>A,p.Gly1961Glu	c.5312+1G>A,None	
5	Stargardt	Abnormal (absent foveal reflexes)	c.5882G>A,p.Gly1961Glu	c.2852T>C,p.Ile951Thr	c.5603A>T,p.Asn1868Ile
6	Cone Dystrophy	Yes	c.5882G>A,p.Gly1961Glu	c.4577delC,p.Thr1526fs	
7	Macular Dystrophy	Abnormal	c.5882G>A,p.Gly1961Glu	c.6221G>A,p.Gly2074Asp	
8	Stargardt	Report Not Available	c.5882G>A,p.Gly1961Glu	c.5882G>A,p.Gly1961Glu	c.5113C>T,p.Arg1705Trp
9	Stargardt	Report Not Available	c.5882G>A,p.Gly1961Glu	c.4363T>C,p.Cys1455Arg	
10	Stargardt	Report Not Available	c.5882G>A,p.Gly1961Glu	c.3113C>T,p.Ala1038Val	c.1622T>C,p.Leu541Pro
11	Stargardt	Yes	c.5882G>A,p.Gly1961Glu	c.5714+5G>A,None	c.587C>T,p.Pro196Leu

4.3.6 Case 4: An X-Linked Conundrum

This case study deals with a concurrence of RP and choroideremia within one pedigree. Intriguingly, this is not the first pedigree whereby RP and choroideremia have led to a difficulties in delivering accurate diagnoses (249). Choroideremia is always associated with X-linked inheritance, while some forms of RP involve X-linked inheritance. Both have initial presentations of nyctalopia progressing to visual field loss. Choroideremia has a more characteristic presentation of white fundal reflex and chorioretinal degeneration (OMIM #303100). However, these changes can be much more difficult to detect if the patient is heavily pigmented (249).

The proband in this family presented with RP with a suspected X-linked or dominant pattern of inheritance. The proband was sequenced with target capture NGS and no likely variant candidates were established. A second affected member of the pedigree was then sequenced after capture probes were switched from Agilent to Nimblegen after the consistent observation of superior sequence coverage with the Nimblegen probes. The improved coverage was sufficient to cover an additional 500 bp into the repetitive region of *RPGR* ORF15 and to detect a two base-pair deletion (Figure 4.3.5). Figure 4.3.5 also incidentally illustrates the difference in sequence coverage achieved by switching targeting enrichment probes (as discussed in Chapter 2.16): top panel- Agilent; middle and bottom panels- Nimblegen.

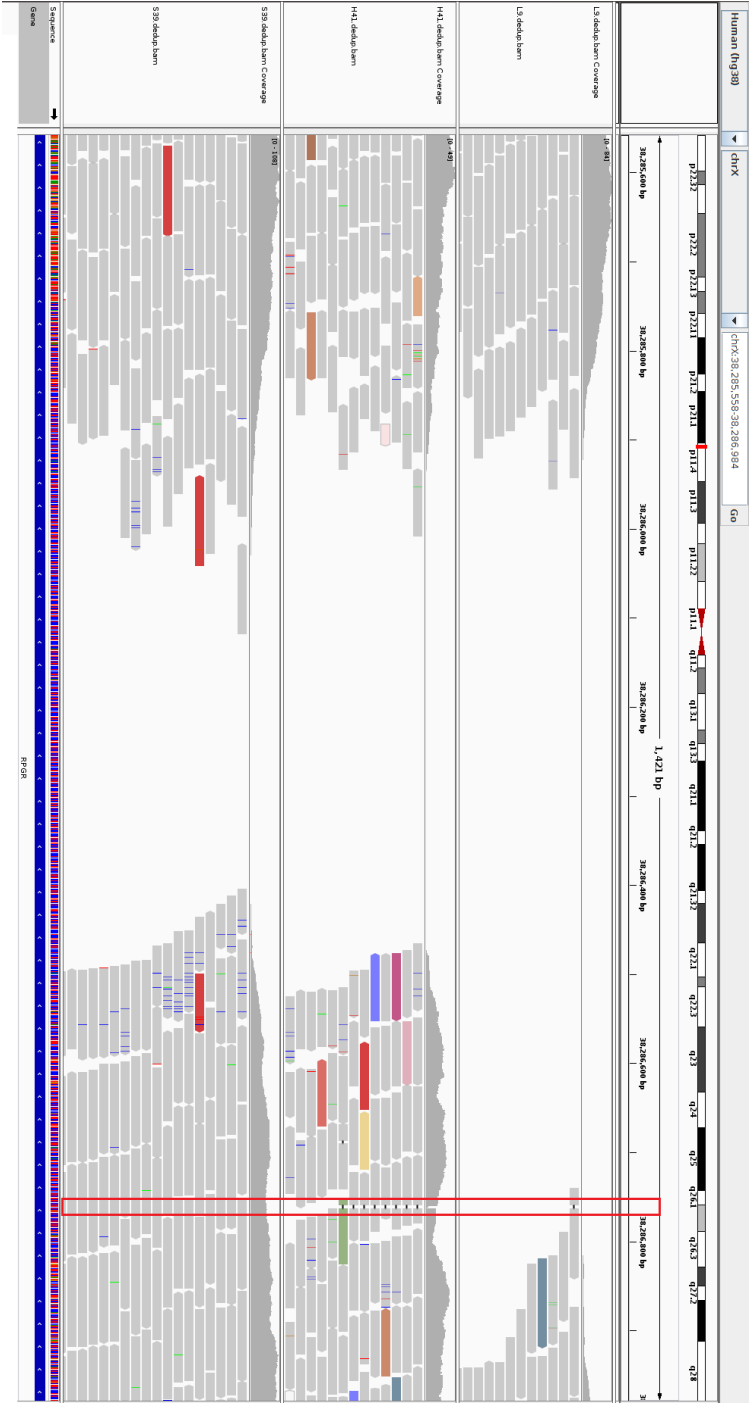


Figure 4.3.5. A pileup of reads from three affected individuals from the same pedigree as seen on IGV software. L9 (RP), H41 (RP) and S39 (choroideremia). A red box highlights the segregating RPGR two base-pair deletion. It can also be seen that S39 does not carry the deletion.

This sequence was sufficiently repetitive that amplicons generated by standard PCR protocols would routinely fail and produce massive amounts of non-specific products. Figure 4.3.5 shows that even though the section surrounding the repetitive zone can be sequenced successfully, all reads drop-off to zero in proximity of the difficult sequence. The incorporation of an improved PCR protocol (Chapter 2.18) specifically designed to amplify *RPGR* ORF15, enabled segregation analysis for this variant to be undertaken for this pedigree (Figure 4.3.6). When these nucleotide deletions are translated into protein sequence and aligned to the reference *RPGR* sequence, it was observed that there was a large truncation event as a result of the deletion event (Figure 4.3.7).

However, there was still no underlying genetic cause established for the presence of choroideremia in this family at this point. Once an affected (choroideremia) patient had been sequenced with NGS it quickly became apparent that there was a massive *CHM* deletion in that section of the pedigree (Figures 4.3.8 and 4.3.9). The deletion spans the entirety of the *CHM* gene and at least 40 kb either side of the gene. The exact boundaries of this deletion are still under investigation, currently it is known that the deletion spans at least 270 kb.

Figure 4.3.8 typifies the results observed when multiple regions within and surrounding *CHM* were amplified. The two *CHM* affected males showed no amplifiable template in any of these reactions, whereas all RP-affected males in the pedigree should consistent presence of *CHM* and neighbouring sequence. This is not the first report of massive deletions encompassing the *CHM* region (250,251), but these additional incidences may suggest that possibly there is significant fragility associated with this particular region of the X chromosome.

ORF15	CTTCTCCCTCCTTCTCTTCTTCTCTCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT
H41	CTTCTCCCTCCTTCTCTTCTTCTCT--TCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT
L9	CTTCTCCCTCCTTCTCTTCTTCTCT--TCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT
R43	CTTCTCCCTCCTTCTCTTCTTCTCT--TCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT
S71	CTTCTCCCTCCTTCTCTTCTTCTCT--TCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT
S72	CTTCTCCCTCCTTCTCTTCTTCTCT--TCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT
I40	CTTCTCCCTCCTTCTCTTCTTCTCTTCTCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT
S39	CTTCTCCCTCCTTCTCTTCTTCTCTTCTCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT
S40	CTTCTCCCTCCTTCTCTTCTTCTCTTCTCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT
S41	CTTCTCCCTCCTTCTCTTCTTCTCTTCTCTGTCTCCCTCCTCTTCTTCTCCTTCTCCAT

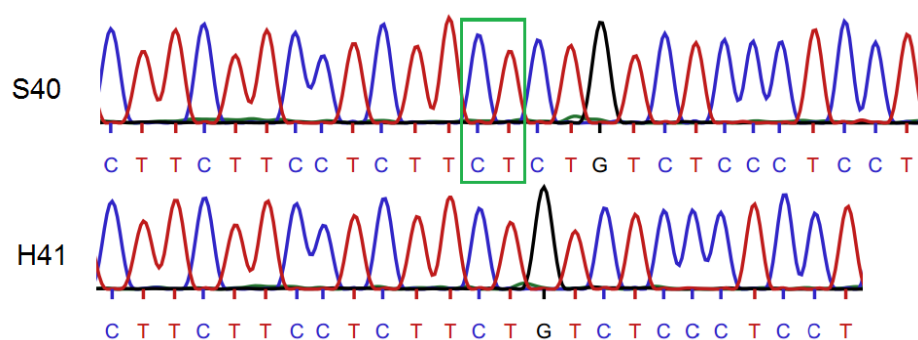


Figure 4.3.6. An alignment of sanger sequencing reads resulting from RPGR ORF15 amplification. The alignment shows the two base-pair deletion segregates with RP in this pedigree: ORF15- Hg38 reference sequence; H41- RP Affected male; L9- RP Affected male; R43- RP Affected male; S71- RP Affected male; S72- RP Affected male; I40- Unaffected male; S39- Choroideremia affected male; S40- Unaffected female; S41- Choroideremia affected male.

wtRPGR 2bpDe1RPGR	SFSMRRTLPIEGTLGLSACFLPNSVFPRCSERNLQESVLSEQDLMQPEEPDYLLDEMTK SFSMRRTLPIEGTLGLSACFLPNSVFPRCSERNLQESVLSEQDLMQPEEPDYLLDEMTK *****
wtRPGR 2bpDe1RPGR	EAEIDNSSTVESLGETTDILNMTHIMSLNSNEKSLKLSPVQKQKKQQTIGELTQDTALTE EAEIDNSSTVESLGETTDILNMTHIMSLNSNEKSLKLSPVQKQKKQQTIGELTQDTALTE *****
wtRPGR 2bpDe1RPGR	NDDSDYEEMSEMKEGKACKQHVSQGIFMTQPATTIEAFSDDEEVEIPEEKEGAEDSKGNG NDDSDYEEMSEMKEGKACKQHVSQGIFMTQPATTIEAFSDDEEVEIPEEKEGAEDSKGNG *****
wtRPGR 2bpDe1RPGR	IEEQEVEANEENVKVHGGGRKEKTEILSDDLTDKAEVSEGKAKSVGEAEDGPEGRGDGTCE IEEQEVEANEENVKVHGGGRKEKTEILSDDLTDKAEVSEGKAKSVGEAEDGPEGRGDGTCE *****
wtRPGR 2bpDe1RPGR	EGSSGAEHWQDEEREKGEKDKGRGEMERPGEGEKELAEKEEWKKRDGEEQEQKEREQGHQ EGSSGAEHWQDEEREKGEKDKGRGEMERPGEGEKELAEKEEWKKRDGEEQEQKEREQGHQ *****
wtRPGR 2bpDe1RPGR	KERNQEMEEGEEEEHGEGEEEEGDREEEEEKEGEGKEEGEGEEVEGEREKEEGERKKEER KERNQEMEEGEEEEHGEGEEEEGDRRGRREGG--RRERGR-----RRGRSGGRT----- ***** . . * : * * * . . * *
wtRPGR 2bpDe1RPGR	AGKEEKGEEEGDQGEGEREEETEGRGEEKEEGGEVEEGGEREEEEEGEGEEEEEG -----
wtRPGR 2bpDe1RPGR	EGEEEEEGEGEEEEEGKGEEEEEGEGEEEEEGEGEGEEEEEGEGEGEEEEEGEGEEEEEG -----
wtRPGR 2bpDe1RPGR	EGEGEEEEEGEGEEEEEGKGEEEEEGEGEEEEEGEGEGEGEGEGEGEGEEEEEGEGEEEEEG -----
wtRPGR 2bpDe1RPGR	EGEGEGEEEEEGEGEEEEEGEGEEEEEGEGEEEEEGEGEEEEEGEGEEEEEGEGEEEEEGEG -----
wtRPGR 2bpDe1RPGR	VEGEVEGEEGEGEGEEEEEGEGEEEREKEGEGEENRRNREEEEEEEGKYQETGEEENERQ -----
wtRPGR 2bpDe1RPGR	DGEYKVKVSKIKGSVKYKHKTYQKKSVTNTQGNGKEQRSKMPVQSKRLLKNPSPGSKKF -----
wtRPGR 2bpDe1RPGR	WNNVLPHYLELK -----

Figure 4.3.7. Alignment of normal RPGR protein sequence versus the 2 base-pair deletion equivalent sequence. The reference sequence (wtRPGR) can be seen to extend far beyond the truncated variant sequence (2bpDe1RPGR).



Figure 4.3.8. *PCR products resulting from amplification of CHM exon 4. Lanes (L): L0-500 bp marker; L1- A33 (RPGR carrier); L2-H41 (RP affected); L3-I40 (unaffected); L4-I41 (RPGR carrier); L5-L9 (RP affected); L6-R43 (RP affected); L7-S39 (Choroideremia affected); L8-S40 (Choroideremia carrier); L9-S41 (Choroideremia affected); L10-S71 (RP affected); L11-S72 (RP affected); L12 (No template control).*

4.3.7 Case 5: Circumnavigating the Pitfalls of NGS

Mutations in the ORF15 region of the *RPGR* gene are estimated to account for 60-70% of all X-linked RP patients (252,253). In terms of total RP incidence, it is estimated to account for 10-20% of all RP cases (164). This statistic emphasises the imperative for reproducible sequencing in the ORF15 zone. As described in the previous section (Chapter 4.3.6), a novel two base-pair deletion has been described to be associated with X-linked RP in an Irish IRD pedigree. That variant was located at the edge of the portion of ORF15 amenable to sequencing. The advantage of that scenario was that the NGS data allowed for an indication of a potential variant.

However, many mutations are suspected to be located deep within the repetitive sequence in ORF15, far beyond that which is detectable by standard NGS. This includes many potential aberrant splice acceptors sites as ORF15 is rich with AG dinucleotides which could potentially interact with the many upstream GT splice donor sites (164). The high conservation rate of this C-terminal domain amongst mammals also suggests that it holds an important function in retinal biology (254). Interestingly, the *RPGR* isoform that dominates expression in the retina is the only isoform to incorporate ORF15 (255).

This case study entails an example of deep ORF15 sequencing to detect a novel variant in two X-linked RP patients that is currently undetected in any control samples that have also undergone deep ORF15 analysis (Figure 4.3.10). This single base-pair deletion, *RPGR* c.2894delA, p.Glu965fs, logically disrupts the remainder of the reading frame and results in a truncated protein (Figure 4.3.11). This region has been previously shown by multiple studies to be crucial for appropriate functioning of the protein (256,257). Two

affected males from this pedigree- D38 and I14, both tested positive for this deletion. I14 was amplified and sequenced in a follow-up study to verify this result. This deletion was absent from all unaffected patient samples used as controls and all population databases queried.

ORF15 poses a particularly challenging task for identifying candidate variants. The repetitive nature of the ORF15 sequence not only makes it difficult to authentically amplify but also makes it impossible for short-read sequences from NGS to faithfully align to. Yet there is clearly an unmet need in standard NGS to detect potentially pathogenic variants in this region which can occur relatively frequently (256). Currently in this study ORF15 is investigated almost independently of NGS. It may be a quixotic endeavour to attempt to merge both protocols as the chemistry involved with both protocols have each proven sensitive to change and are both currently optimised to perform their individual tasks. However, some researchers have cited great success using the integrated method (165).

Early canine studies have shown that RPGR-RP is amenable to intervention with AAV gene therapy as long as photoreceptors are still present. The *RPGR* transgene could correct photoreceptor structure and reverse opsin mislocalisation in treated areas (258). This group have since further optimised this therapy and proven efficacy in primate models also (259). Great strides have been made in this therapy since the initial canine treatment. Codon-optimised ORF15 transgenes have recently been tested in multiple mouse models with great efficacy, the optimised transgene shows massively increased expression levels and faithful recapitulation of the ORF15 protein. It has also shown to have a good safety profile and to be capable of rescuing the disease phenotype (253).

Another study has also shown that the restoration of as little as 20% of the wildtype RPGR expression with a modified *RPGR* transgene was sufficient to recapitulate normal phenotypes (260). To summarise, variants in *RPGR* hold great potential for treatment with AAV-based therapy and it is unsurprising that there is currently three clinical trial in the recruitment stage for such studies (261).

ORF15	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
A20	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
A29	TCCCCTTCTCCTTCTCTTCCCCCTA ~~C~~ CCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
A36	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
BA3	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
B32	TCCCCTTCTCCTT ~~A~~ CTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
C15	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
CB48	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
CB55	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
E46	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
F15	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
F43	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
F53	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
F64	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
F79	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
F82	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
F93	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
F96	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
H28	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
H79	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
H81	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
I24	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
I40	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
I49	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
J80	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
O45	TCCCCTTCTCCTTCTCTTCCCCCTC ~~C~~ TCTCTCCTTCTCCTCTTCCCCCTCCCATTCT
O65	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
P46	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
R38	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
R54	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
H41	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
K1	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
P9	TCCCCTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
C50	T ~~C~~ A ~~C~~ CTTCTCCTTCTCTTCCCCCTCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
D38	T ~~C~~ A ~~C~~ CTTCTCCTTCTCTTCCCC-CCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
I14	TCCCCTTCTCCTTCTCTTCCCC-CCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT
I14	TCCCCTTCTCCTTCTCTTCCCC-CCCCCTTCTCCTTCTCCTCTTCCCCCTCCCATTCT

Figure 4.3.10. An alignment of reference RPGR sequence versus the results of deep ORF15 sequencing samples. The reference sequence (RPGR, NM_001034853.1) can be seen to consistently align to the unaffected samples (A20-C50) at the nucleotide position where D38 and the two separate preparations of I14 appear to harbour a single nucleotide deletion, as depicted by a “-“, surrounded by a green box. Many single base misalignments of unknown significance can also be observed in this sequence region, as depicted by red boxes.

4.3.8 Case 6: From Bench to Bedside, An RPE65 Story.

This case study provides an overview of the path from variant discovery to patient therapy as seen in an Irish IRD pedigree. *RPE65* is involved in the retinoid cycle that recycles retinoids throughout the RPE and photoreceptor cells (Figure 4.3.12). *RPE65* is a gene traditionally associated with autosomal recessive LCA, a severe and clinically distinct phenotype (262–264). This has been expanded to include less severe autosomal recessive retinopathies outside of the clinical description of LCA. Luxturna, the first ocular gene therapy approved by the FDA, is for biallelic *RPE65* retinal disease (265).

However in 2011, in contrast to the above, it was shown that in multiple Irish pedigrees a specific variant in *RPE65*, c.1430A>G (p.Asp477Gly), was associated with a very different inheritance pattern and phenotype (266). This amino acid position has been shown to be highly conserved across multiple species (Table 4.3.2). The disease phenotype was comparatively much milder and closely resembled the clinical manifestations of choroideremia. In addition, all affected members of the tested pedigrees also had identical haplotypes in the surrounding genomic region, implying a common ancestor.

This variant is one that continues to be detected in the ongoing Target 5000 study. Currently there are 14 confirmed affected individuals and a further 6 patients from these pedigrees that are pending analysis but likely to harbour this variant. These 20 affected patients span 7 pedigrees although likely originate from a single source as observed in

previous studies (266). An example of one such pedigree segregation is shown in Figure 4.3.13.

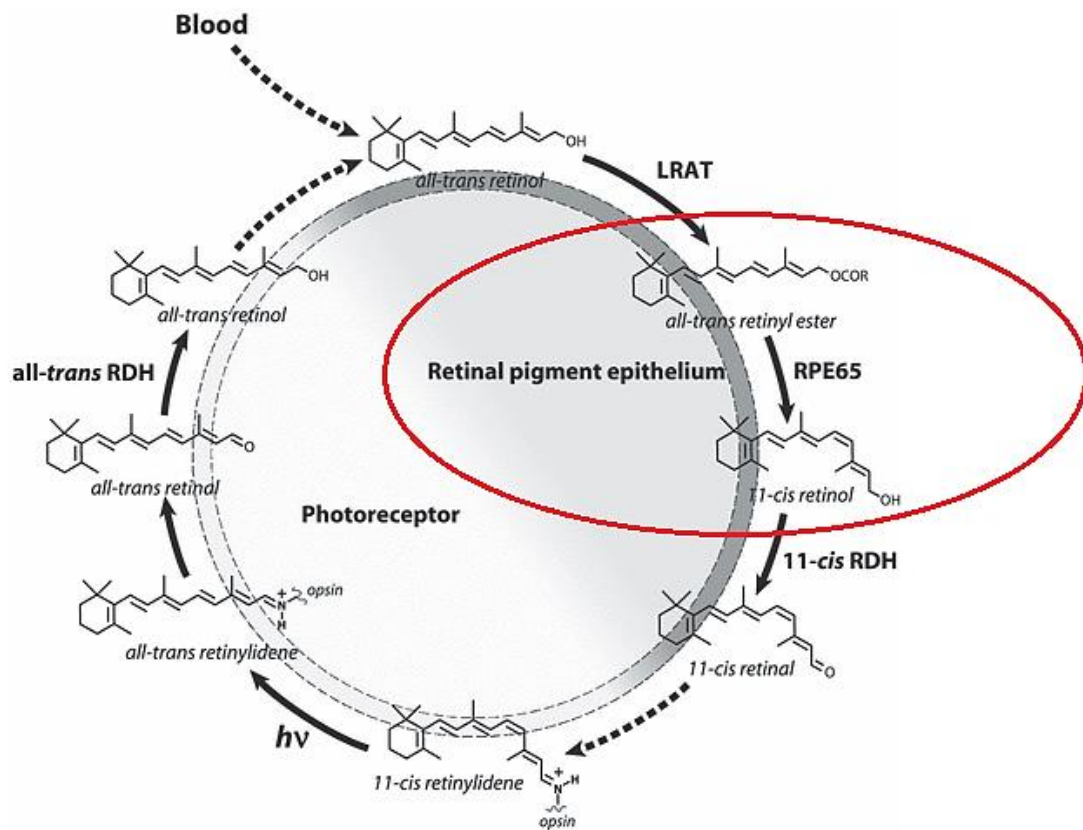


Figure 4.3.12. An illustration of the retinoid cycle. The substrate, product and site of activity of RPE65 have been highlighted in red. Illustration adapted from Nikolaeva et al. (267).

Table 4.3.2. Conservation Analysis of the RPE65 p.D477 amino acid. This amino acid is a very highly conserved across many distantly related species. Table taken from Bowne et al. (266).

<i>Organism</i>	<i>Amino acid sequence</i>
Human	LCKLVKTKETWWWQEPDSYPSEPIFVSHPDALEEDD
Chimp	LCKLVKTKETWWWQEPDSYPSEPIFVSHPDALEEDD
Macaque	LCKLVKTKETWWWQEPDSYPSEPIFVSHPDALEEDD
Rhesus	LCKLVKTKETWWWQEPDSYPSEPIFVSHPDALEEDD
Pig	LCKLVKTKETWWWQEPDSYPSEPIFVAHPDALEEDD
Mouse	LCKMNVKTKETIWMWQEPDSYPSEPIFVSQPDAL EEDD
Rat	LCKLVKTKETIWMWQEPDSYPSEPIFVSQPDAL EEDD
Cow	LCKLVKTKETWWWQEPDSYPSEPIFVSHPDALEEDD
Cat	LCKLVKTKETWWWQEPDSYPSEPIFVSHPDALEEDD
Dog	LCKLVKTKETWWWQEPDSYPSEPIFVSHPDALEEDD
Elephant	LCKLVKTKETWWWQEPDSYPSEPIFVSHPDAMEEDD
Chicken	–CKLVKTKETWWWQEPDSYPSEPIFVSHPDALEEDD
Zebrafish	FCKLVKSKETWIWQEPDAYPSEPLFVQSPDAEDED

Barcode	Genotype	Current Phenotype
O43	+/-	Affected
P26	+/-	Affected
P27	+/-	Affected
Z6	+/-	Affected
Z7	+/-	Affected
Z8	+/-	Affected
O44	+/+	Unaffected
O45	+/+	Unaffected
O46	+/+	Unaffected
P28	+/+	Unaffected
R35	+/+	Unaffected

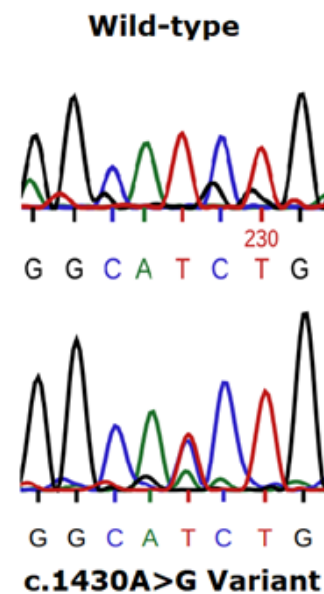


Figure 4.3.13. An illustration of a patients harbouring the RPE65 p.D477G variant. In this pedigree each individual's phenotype corresponds to the expected genotype (left image). Also shown is a typical sanger sequencing trace for the sequence of an unaffected (wild-type) genotype, O46 (right image, top). The variant sequence shown below is that from patient Z6 (right image, bottom). Sequences were read from the sense strand so the observed change above is T>G (GGCA[T>G]CTG). Data shown was generated from the Target 5000 project.

This variant has since been detected in other population studies alongside the clinical presentation of similar phenotypes, largely described as a choroideremia phenocopy (268,269). However, until recently the exact disease mechanism caused by this variant has evaded researchers. It was initially believed that this variant produced an abnormal protein structure that may affect enzymatic function (266).

Recently it has been shown in a mouse model that although this variant is not located in a critical functional domain, the variant induces sufficient change to alter the physiochemical properties of the p.D477 loop to produce an aggregation-prone surface. This surface gains the aberrant function of enabling abnormal protein-protein interactions (270). One hypothesised type of interaction is with ubiquitin ligases, which likely results in proteasomal degradation as seen in several other *RPE65* mutations (270,271). This theory may support the variability observed phenotypically as severity would then become dependent on several other factors such as ubiquitination rates and proteasome response.

Despite the remarkable success of the gene therapy treatment of biallelic *RPE65*-retinopathies with Luxturna, an AAV-based therapy (272), the application of this treatment to cases of p.D477G would be counter-intuitive as it appears to be a dominant-negative mutation and expressing more standard *RPE65* protein may possibly contribute to the previously discussed aggregation of protein.

An alternative therapeutic intervention has been explored for some Target 5000 patients with the p.D477G variant. A synthetic retinoid, to account for the reduced activity of RPE65, has been used to restore balance in the retinoid cycle. QLT091001 (9-*cis* retinyl acetate), is a prodrug that is transformed by hydrolysis *in vivo* to 9-*cis*-retinol which can act as a surrogate 11-*cis*-retinal, therefore re-establishing the vital biochemical component of the retinoid cycle (Figure 4.3.12) (273).

QLT091001 is an oral treatment that, when taken daily for 7 days, has shown benefit in the majority of patients within six months as part of an open-label, Phase 1b, proof-of-concept study (273,274). In the study, 80% of patients were shown to respond positively to this treatment with 60% showing an increase in visual fields and 60% with improved visual acuity. There were no serious adverse effects reported as part of the study. This trial showed promising data that treatment of patients with an *RPE65* p.D477G genotype are likely to respond well to the safe treatment of orally-administered synthetic retinoids. All of the clinical studies referred to in this section of the chapter were undertaken by Dr. Paul Kenna. The work-up of the genotypes in additional pedigrees were undertaken as part of this thesis work as presented above. The clinical trial data is referred to here to emphasise the many valuable outcomes that may be achieved by a study such as Target 5000.

4.3.9 Concluding Remarks

The results presented in this chapter of the PhD thesis demonstrate some of the typical and atypical genetic findings in IRD pedigrees identified as part of the Target 5000 study. The study also emphasises the importance of accurate clinical phenotyping and genotyping of IRD patients. In particular, case studies involving multiple pathogenic genotypes exemplify the importance of thorough genetic testing prior to therapeutic recruitment as, for example, a therapy designed to rectify pathogenic variants in *FLVCR1* will be of little benefit to a patient with pathogenic variants in *NR2E3*.

In addition, the study provides insights in terms of the Irish population of mutations in the ORF15 region of the *RPGR* gene which can be causative of the most common form of X-Linked RP. The need for ever-improving methods to sequence structurally complex region of the genome with repetitive sequences is elegantly illustrated by the *RPGR* ORF15 study.

Furthermore, the results also highlight that different variants in the same gene, such as *RPE65*, may require different therapeutic interventions and that robust genotyping should expedite patients, when deemed appropriate, accessing such interventions. The results presented here provide a snapshot of many of the different issues that have been encountered during the course of the Target 5000 study. Perhaps one of the most important issues for genotyping of IRDs as discussed in Chapter 4.1, is the presence of previously unreported novel variants in IRD patient cohorts and the need for global databases to aid in establishing their pathogenicity or otherwise, as well as the various in silico and

functional assays (also referred to in Chapter 4.1) that may be used to try to establish pathogenicity. The availability of a genetic test can remove the need for other more invasive, expensive, and time-consuming investigations, especially when a prognosis is consistent and well described.

4.4 AAV Analysis by NGS

4.4.1 Introduction

In this chapter the evaluation of the packaging efficiency of three serotypes of adeno-associated viruses (AAVs), AAV 2/2, AAV 2/5 and AAV 2/8 will be discussed. AAVs are single stranded DNA viruses that are approximately 4.7 kb in length. They differ from the naturally occurring adenovirus due to their reliance on co-infection with a helper virus to replicate. AAVs have become the vector of choice for gene delivery in many clinical trials in the field of IRDs (186,187,275).

As a target organ, the eye has multiple characteristics that make it an attractive option for gene therapy. Firstly, the presence of the blood-retina barrier functions as a protective layer (276). This reduces the risk of therapies designed to perform within the eye from becoming systemically introduced. Also, the eye itself is a highly compartmentalised organ. This enables the opportunity to exploit the distinct spaces within the eye to allow for increased accuracy when delivering therapies to specific tissues or areas of the eye, including incorporating cell-specific promoters for added specificity (277).

Many recombinant AAV vectors have been developed for specific use in the eye (278,279), among other tissues (280,281). These recombinant viruses can combine various capsids (protein coats) and inverted terminal repeats (ITRs: used for primase-independent second-strand synthesis) from different serotypes to generate hybrid vectors. This can be readily achieved as only ITRs are required *in cis* of the desirable sequence, other structural and packaging components; the *REP* and *CAP* genes can be delivered *in*

trans by co-transfection on separate plasmids. In principle such recombinant AAV vectors allow an additional layer of experimental elegance as different serotypes of AAV have been shown to selectively transduce specific cell types within the eye (282–284).

Additionally, intraocular administration of AAV typically shows attenuated immune responses when compared to systemic injections due to the unique immune response modulating the environment of the eye (285). This is particularly significant when considering the nature of viral transgene expression. Viral dose, serotype, the nature of the transgene, the efficacy and longevity of the therapeutic all become key considerations when contemplating factors that may influence an immune response against the foreign virus.

Although there are alternatives to AAV available in the realm of viral gene therapies, in many cases for ocular indications these tend to be overall less desirable. However, some viruses have characteristics that may be preferable to AAV for some gene therapies. For example, lentiviral vectors have a major advantage over AAV therapies as they can package significantly more DNA than AAV vectors. RetinoStat (Oxford BioMedica) was the first ocular gene therapy using a lentiviral vector to be delivered to humans and has reported a good safety profile to date (286,287).

This additional packaging capacity has also been used to express multiple genes simultaneously by use of a bicistronic vector (288). However, the wide range of transduction characteristics of AAV serotypes make it attractive for achieving a cell specific transgenic expression (289), especially with respect to gene therapies for retinal

degenerations. Furthermore, technologies such as directed evolution of AAV vectors are greatly expanding the array of AAV serotypes available for such purposes (241,290,291).

For some target cells types in the retina lentiviral vectors appear to be far less efficient, for example, when it comes to photoreceptor transduction rates compared with AAV (292). In addition, the safety and longevity of AAV vectors in the eye has been well documented in many preclinical studies and more recently in clinical studies (293–296).

Despite a study showing that up to 8.9 kb could be incorporated into a single rAAV 2/5 that was found to be of functional benefit (297), an attempt to replicate this work by a separate group has shown that AAV2, AAV5 and AAV8 seem incapable of packaging sequences over 5067 bp (298). It has also been found that AAV vectors containing larger inserts (>5.3 kb) are more likely to be degraded by the host cell's proteasome and transduce cells less efficiently (299). To overcome the packaging limitations of AAV, several approaches have been used, some of which employ dual AAV vectors in an attempt to achieve expression of gene inserts greater than 5 kb.

The Fragmented approach involves single or dual vectors that produce fragments of the desired insert sequence. Fragment sizes are determined by the vector's ability to package the insert. Regions of homology within the fragments can then be used to resolve the fragments into their desired complete size (300). It has been since revealed that this method was unintentionally used to package the first 8.9 kb insert that subsequently proved difficult to replicate (301). The methods by which this resolution occurs are illustrated in Figure 4.4.1. Although success has been shown in achieving full length products using the fragmented method, other studies have shown a lack of reproducibility

and faithful resolution which suggests that this may not currently be the optimal approach in a clinical setting due to the potential of creating mutant transgenes (302).

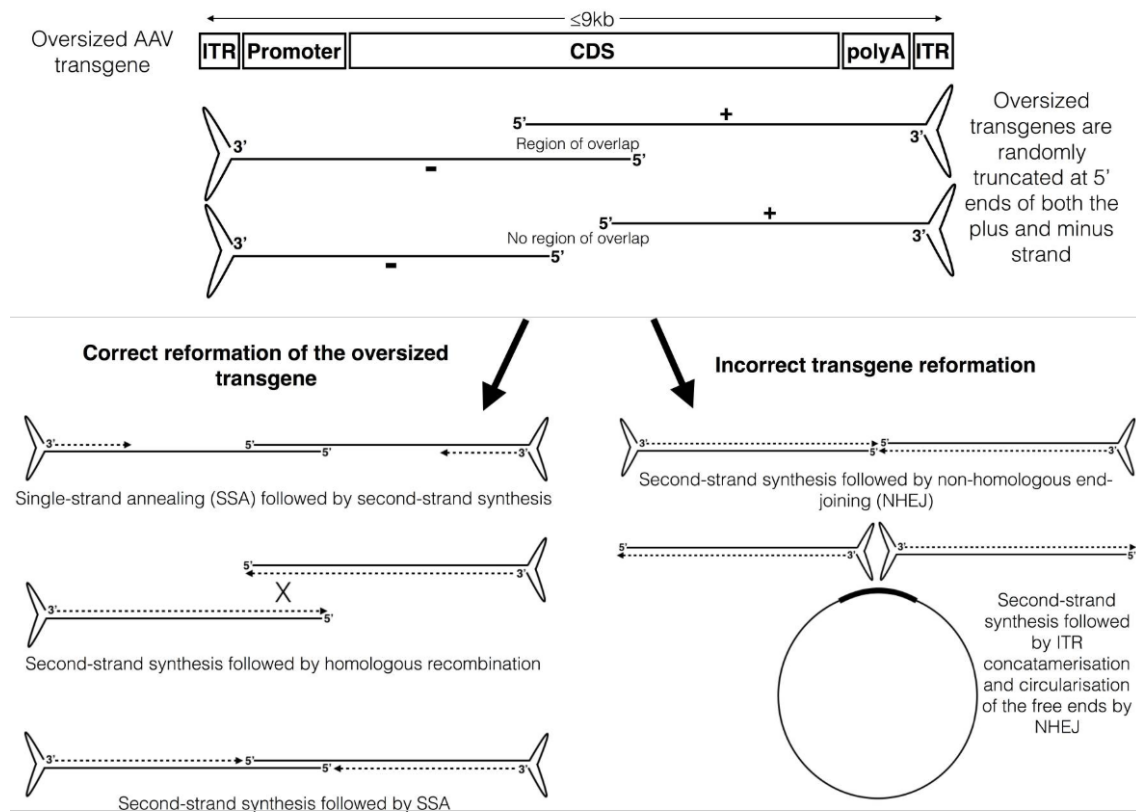


Figure 4.4.1. Transgene structure for an AAV gene therapy. AAV capsids carry single stranded DNA to be delivered to the host cell. Inside the nucleus the single-stranded transgene is converted into a double-stranded structure either by second-strand synthesis or by single-strand annealing of complementary transgene strands. Correct reformation of the transgene requires some overlap of single stranded DNA (bottom left), incorrect reformation of the transgene occurs if single strands don't sufficiently overlap or if they are reformed in the incorrect orientation (bottom right). ITR = inverted terminal repeat; CDS = coding sequence; polyA = polyadenylation signal; AAV = adeno-associated virus. Illustration taken from McClements and MacLaren (303).

The Overlapping approach is a modification of the Fragmented approach using dual vectors. In the Overlapping approach, the partial packaging of the AAV is still considered, however each of the two vectors contain different ends of the desired transgene with a pre-designed region of overlapping sequence. The overlapping sequence is usually a section of protein coding region of the desired transgene (304). This has been successfully used to treat mouse models of muscular dystrophy by generating a “minidystrophin” transgene of 7.3 kb with an overlapping region of 372 nucleotides (305).

Alternatively, transgenes could be stitched together by use of splice sites. This does not require any overlapping regions and is called *Trans*-splicing. The method works by exploiting AAV’s ITRs tendency to be concatemerized (306). This results in multiple AAV inserts incorporating into the same product. Splice sites are then used to ignore the resulting integrated ITR sequence (situated in the middle of the desired transgene). This can also result in the undesired orientation of constituent segments (307). Also efficiency of this method is highly variable but can be improved with careful splice site design (308).

Lastly, the *Trans*-splicing and Overlapping methods can be combined to form what is known as the Hybrid method. This has been shown to be the most efficient dual vector method (309). By utilising both concatemerization and homology recombination events, this method ensures that splicing will remove any redundant incorporated sequence resulting from the hybridisation of the two transgene components. The region of homology has been shown to predominantly influence efficiency and that the *Trans*-splicing method may only come into effect when the Overlapping method is inactive (304).

In addition to a variety of studies focused on generating AAV vectors with larger transgenes, much research has also been undertaken on optimising the production and purity of AAV vectors which have an optimally sized transgene (approx. 4.7kb) (299,310–312). In this regard, over-sized plasmid backbones have been employed in the production of AAV to try to eliminate packaging of the plasmid rather than the therapeutic transgene (313).

Furthermore, manufacturing protocols now incorporate steps to eliminate or minimise the presence of empty particles which can otherwise dominate the AAV vector population from a viral preparation (314,315). It is of note that to date it would seem that there has been less thorough investigation of the nature of the actual transgenes present in the AAV population. For example, does this population of packaged vectors comprise solely full-length transgene inserts or a mixed population of inserts including truncated transgenes?

Given the importance of such issues for the efficacy of AAV-mediated gene therapies and the evaluation of viral dose, among other issues, it was deemed of significant value to begin to address this question using the next generation sequencing (NGS) technologies established for Target 5000 and the many AAV preparations available from an on-going ocular gene therapy programme in the team.

In this study whole-viral genome NGS was used to observe the preferential packaging patterns of various recombinant in-house prepared AAVs with numerous substituted component elements. Subcategories of rAAVs were investigated for impact on packaging patterns due to insert size, relative vector backbone size and serotype. The results

presented below provide some interesting initial insights into this important topic for AAV-mediated ocular gene therapies.

4.4.2 Methods

See Chapters 2.19-2.21 for details of preparation of AAVs and library preparation of AAV samples for next generation sequencing. All AAV preparations were undertaken by Dr. Naomi Chadderton and other members of the Farrar laboratory.

AAV Bioinformatic Analysis

Reference sequences for alignment purposes were generated based on the sequence obtained during the cloning strategies undertaken as part of rAAV production. Sequences were aligned using alignment tools from BWA (110). Alignments were sorted and indexed using samtools (316). Read mapping was visualised using IGV (317). Depth of coverage was determined by the appropriate tools recommended by GATK (318) after the necessary file headers were amended as required using Picard tools (111).

Graphical Representations

For the purpose of generating graphs that could be readily interpreted, all depth of coverage graphs were created using normalised read counts. Reads were normalised by dividing the per-base coverage by the total coverage. Graphs dependent on relative ratios were calculated using total read counts for the specified regions unless otherwise stated. “Insert” size refers to the total sequence between ITRs that is designed to be encapsidated. “Backbone” size refers to the total sequence between ITRs that is not designed to be encapsidated. “Transgene” refers to the specific target of interest located within the

“insert” region, such as the protein-coding element of a replacement gene or a micro-RNA sequence.

4.4.3 Results

Five subgroups of the rAAVs were chosen for analysis by sequencing with the intention of investigating the effects of serotype and relative insert size on packaging efficiency. Each subgroup contained identical or near-identical provirus plasmid vectors.

Subgroup 1 was first chosen as the member rAAVs were identical in size and sequence, differing only by capsid type. The serotypes were AAV 2/2, AAV 2/5 and AAV 2/8 (Figure 4.4.2). The transgene in question contained a replacement protein coding sequence (an IRD-related wildtype cDNA) that could potentially be used to restore absent or reduced levels of that protein in a mutant model. The insert size was approximately 4.4 kb and the vector backbones were roughly 2.7 kb. Although AAV 2/5 and 2/8 had similar transgene packaging efficiencies, AAV 2/2 appeared, from this initial analysis, to have a superior selective capacity for preferentially packaging the insert and transgene relative to the undesired vector backbone incorporation (Figure 4.4.2).

Upon close inspection of sequencing depth of coverage analysis, it can also be seen that AAV 2/5 seem to preferentially package the region in close proximity to the Left ITR in abundance when compared to the other two serotypes (Figure 4.4.3). It can also be seen that the highest sequencing depth across all three serotypes covers the region of the transgene insert. Lastly, all three AAV serotypes exhibited high selectivity for

preferentially packaging the desired insert as opposed to the functionally redundant plasmid backbone sequence. This specificity will become more apparent in later sections where the relative ratio of insert to backbone sequence packaging is less favourable.

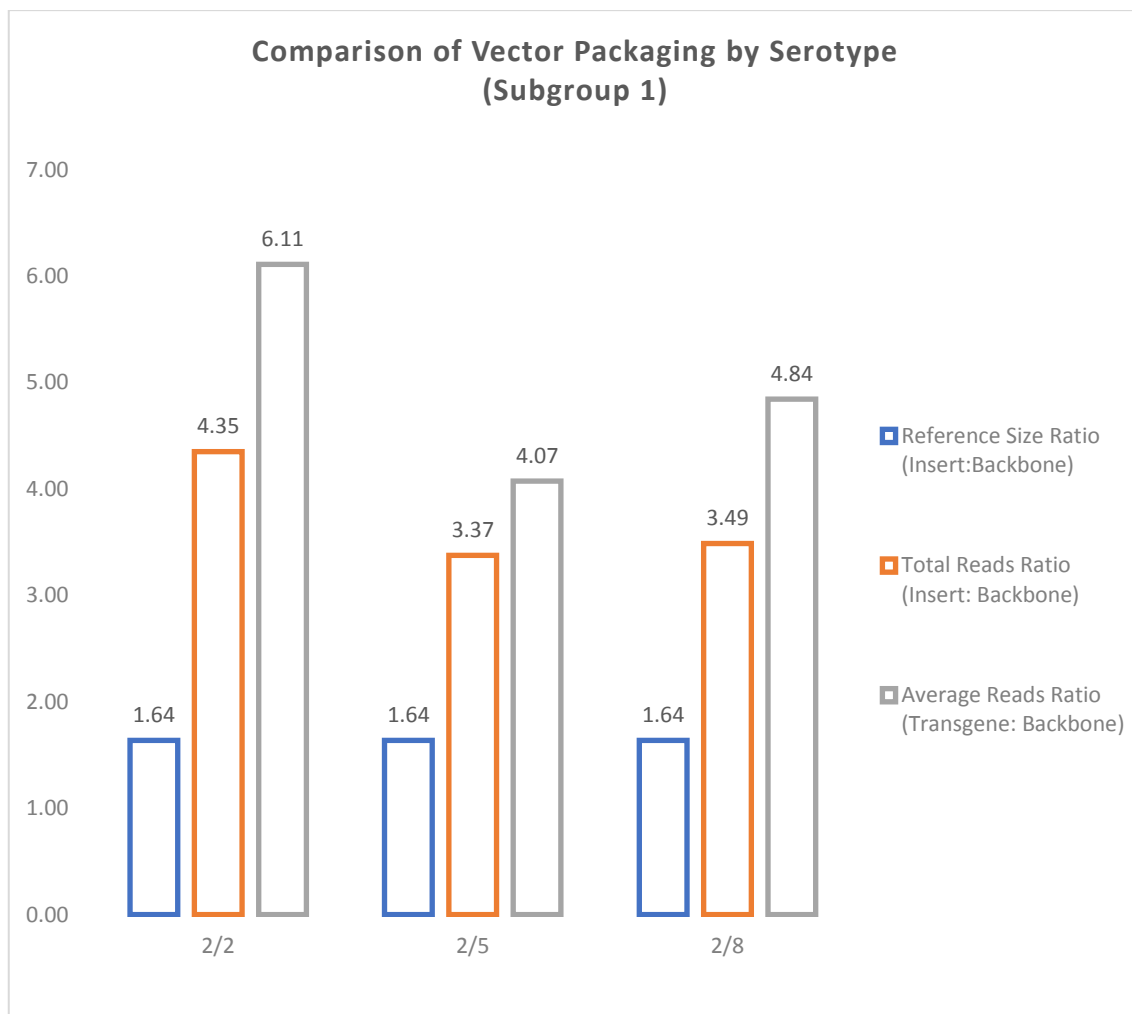


Figure 4.4.2. A graphical representation of relative ratios of vector packaging grouped by AAV serotype (see *Methods: Graphical Representation* for additional details). Each sample has an identical insert size, backbone size and transgene sequence.

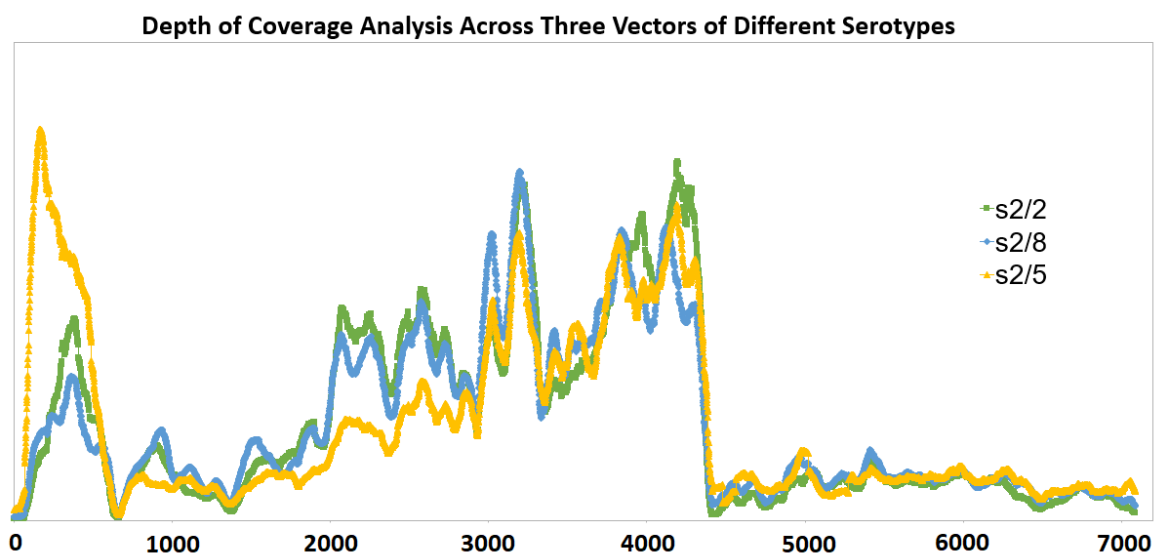


Figure 4.4.3. A relative depth analysis graph of coverage of identical vectors with different serotypes in Subgroup 1, the sequence reads at each base are represented relative to the total number of reads for that sample (see Methods: Graphical Representation for additional details). Approximate locations of key sequence features (numbers indicate basepairs as per reference sequences): Left ITR (0-150); CMV Enhancer (150-500); CBA Promoter (500-850); CBA Intron (850-1850); Transgene (1850-4400); Right ITR (4400-4550); Vector Backbone (4550-7100).

Subgroup 2 was selected for their similar criteria as Subgroup 1. Subgroup 2 however did not contain a protein coding replacement transgene sequence, instead the customised insert contained a microRNA sequence therefore samples in Subgroup 2 differed only by capsid type. The serotypes were AAV 2/5 and AAV 2/8 (Figure 4.4.4). The approximate insert and backbone sizes in this subgroup were similar to those in Subgroup 1, 4.4 kb and 2.7 kb respectively.

Subgroup 2 shows a notable difference in the serotypes respective ability to selectively package the desirable portion of the transgene insert as opposed to vector backbone, despite being equal in all other regards (Figure 4.4.5). AAV 2/8 shows ideal uniform coverage across the transgene insert and near absence of vector backbone incorporation. AAV 2/5 however shows spikes of coverage with relatively lower uniformity across the transgene region and relatively more backbone coverage in comparison to its AAV 2/8 counterpart.

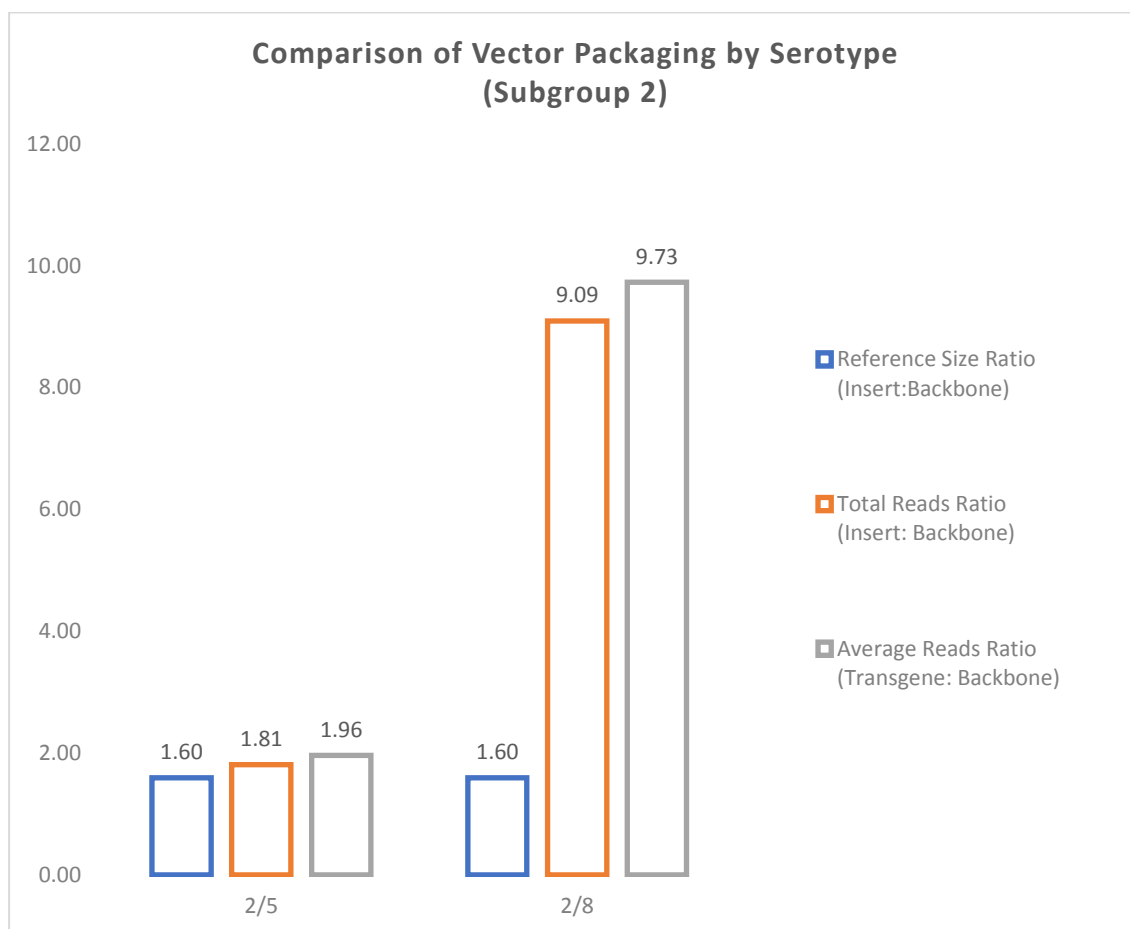


Figure 4.4.4. A graphical representation of relative ratios of vector packaging grouped by AAV serotype in Subgroup 2 (see Methods: Graphical Representation for additional details). Both samples have an identical insert size, backbone size and sequence.

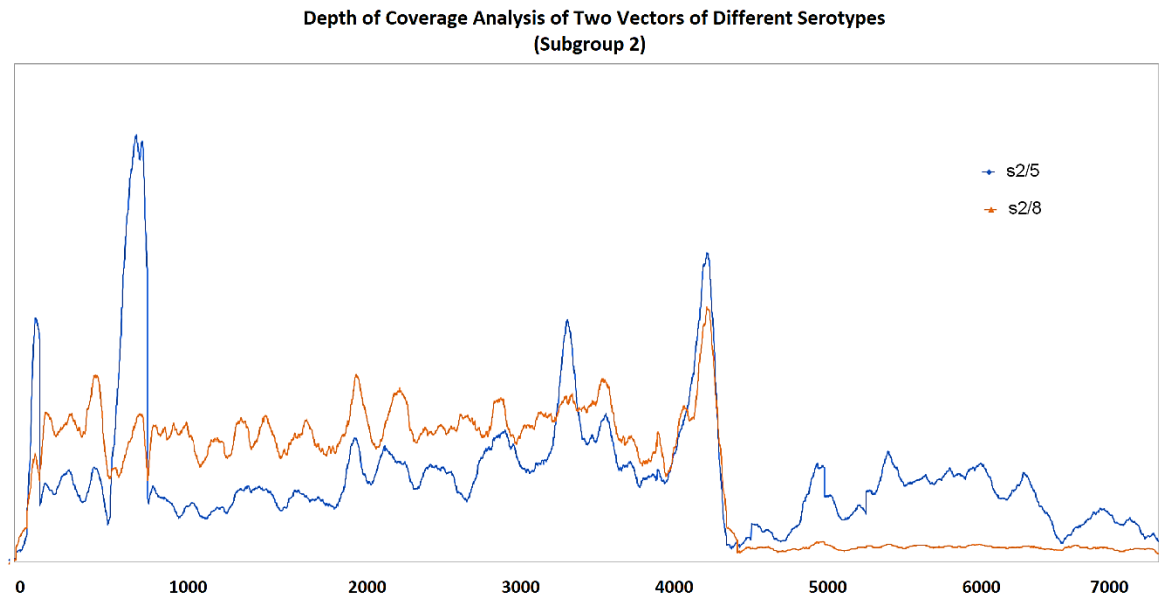


Figure 4.4.5. A relative depth analysis graph of coverage of identical vectors with different serotypes in Subgroup 2, the sequence reads at each base are represented relative to the total number of reads for that sample (see *Methods: Graphical Representation* for additional details). Approximate locations of key sequence features (numbers indicate basepairs as per reference sequences): Left ITR (0-150); H1 Promoter (350-600); CBA Promoter (500-850); Transgene (2700-4000); Right ITR (4350-4500); Vector Backbone (4500-7100).

The candidates in Subgroup 3 differed slightly from the previous two subgroups. Subgroups 1 and 2 each were composed of similar size vector components whereas Subgroup 3 possessed a relatively shorter insert sequence. The insert size was approximately 2.4 kb and the backbone size was roughly 2.9 kb. The two samples in Subgroup 3 were identical in size and near-identical in sequence, differing almost exclusively by capsid type. The serotypes were AAV 2/5 and AAV 2/8 (Figure 4.4.6). The sequence variation in this group was due to the substitution of a targeting micro-RNA (2/5) sequence with a non-targeting (2/8) control sequence.

Subgroup 3 shows a prominent difference between samples in terms of coverage of desired insert as opposed to vector backbone (Figures 4.4.6 and 4.4.7). Although both serotypes showed a favourable ratio of coverage of insert to backbone, AAV 2/8 appears to be more selective than its 2/5 equivalent, very similar to the data observed in Subgroup 2.

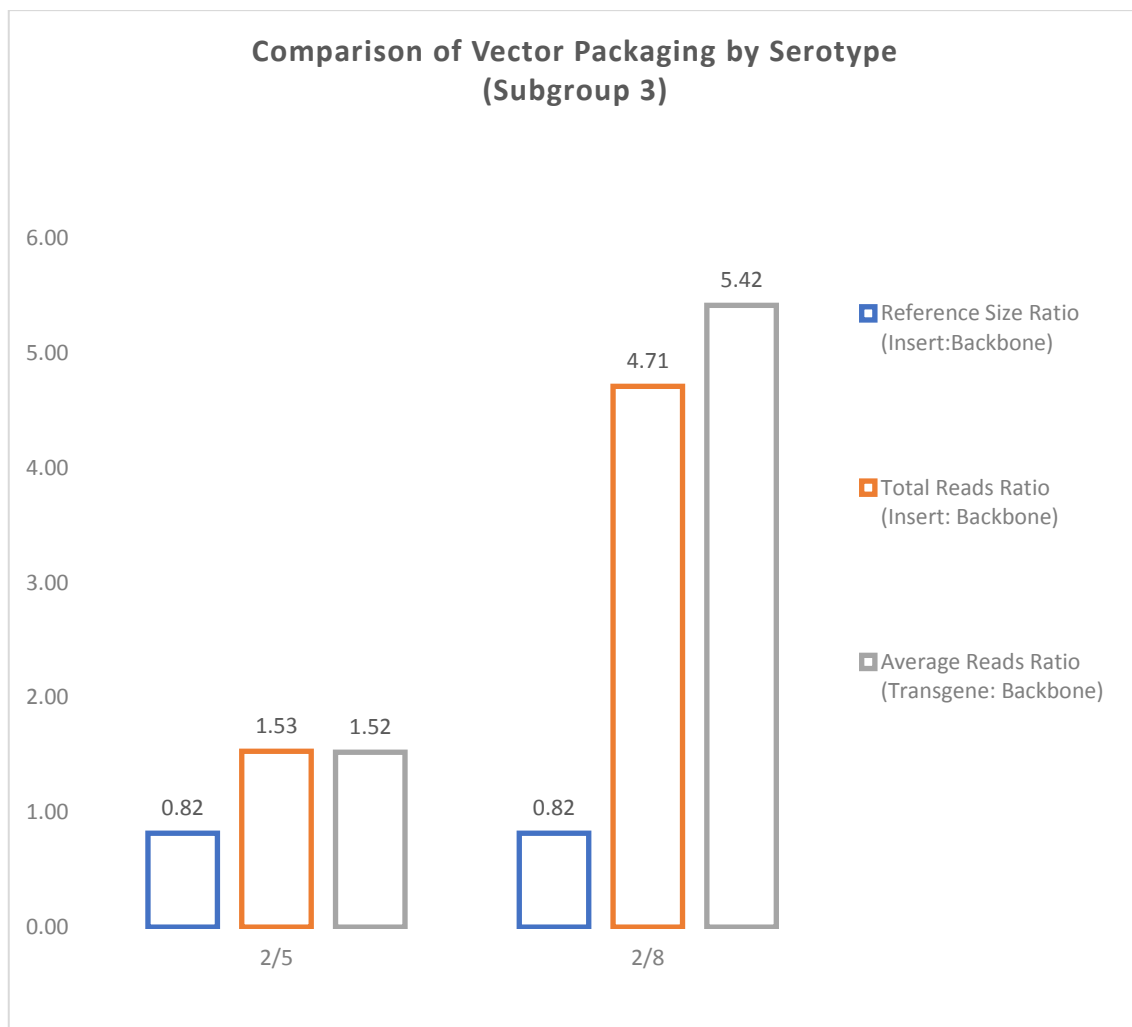


Figure 4.4.6. A graphical representation of relative ratios of vector packaging grouped by AAV serotype in Subgroup 3 (see Methods: Graphical Representation for additional details). Both samples have an identical insert size, backbone size and near-identical sequences.

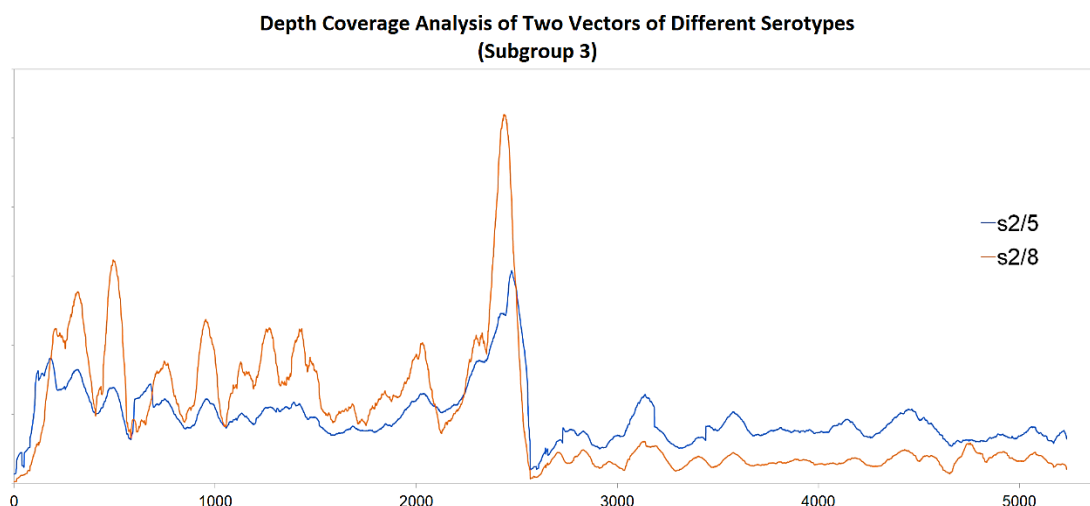


Figure 4.4.7. A relative depth analysis graph of coverage of near-identical vectors with different serotypes in Subgroup 3, the sequence reads at each base are represented relative to the total number of reads for that sample (see Methods: Graphical Representation for additional details). Approximate locations of key sequence features (numbers indicate basepairs as per reference sequences): Left ITR (0-150); (Non-/Targeting Antisense Sequence (350-600); CMV Promoter (900-1550); Transgene EGFP (1550-2300); Right ITR (2550-2700); Vector Backbone (2700-5400).

The samples in Subgroup 4 were identical vectors of containing a transgene, designed with the intention of restoring function where lost in mutant models. Both proviral vectors were identical in sequence. The insert size was approximately 3.5 kb and the backbone size roughly 2.9 kb. The serotypes were AAV 2/5 and AAV 2/8 (Figure 4.4.8). Subgroup 4 produced remarkably different coverage ratios compared to the other subgroups. Subgroup 4 showed a counterintuitive superior coverage of backbone vectors in comparison to the desired insert (Figures 4.4.8 and 4.4.9). Both serotypes displayed this pattern of coverage, however 2/5 displayed a slightly higher ratio of insert to backbone coverage compared to 2/8, 0.85 to 0.51 respectively.

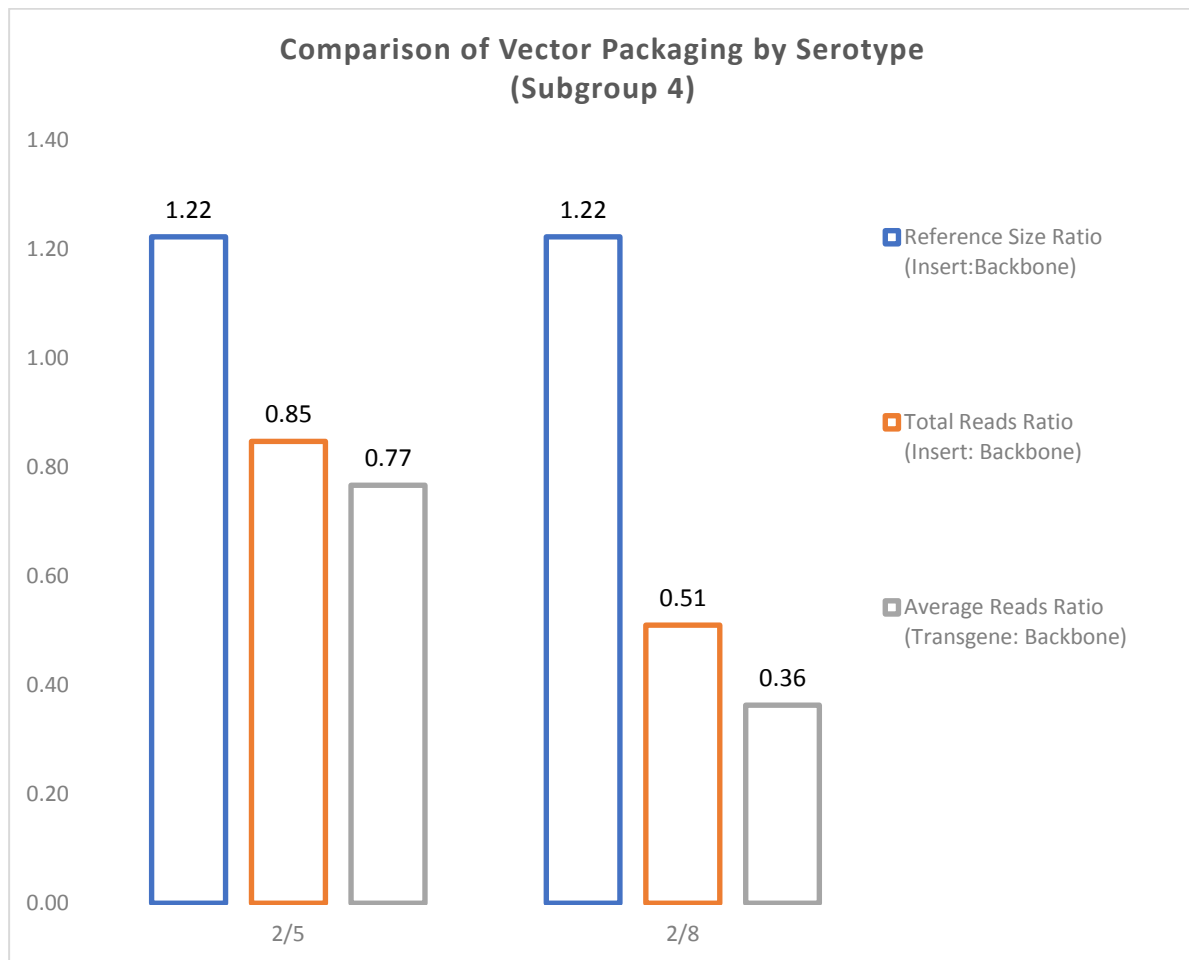


Figure 4.4.8. A graphical representation of relative ratios of vector packaging grouped by AAV serotype in Subgroup 4 (see Methods: Graphical Representation for additional details). Both samples have an identical insert size, backbone size and identical sequences.

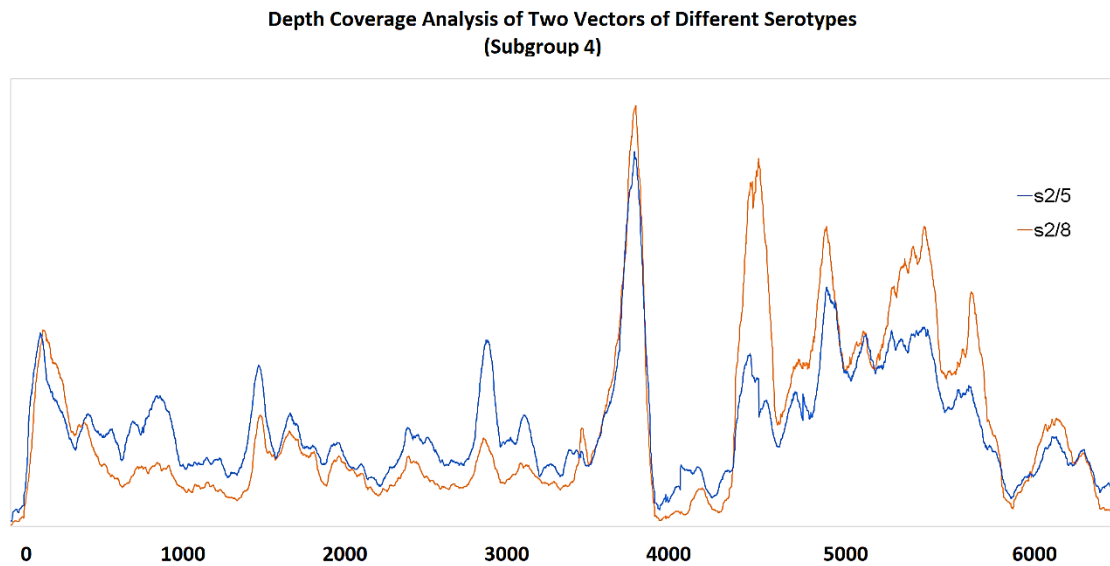


Figure 4.4.9. A relative depth analysis graph of coverage of identical vectors with different serotypes in Subgroup 4, the sequence reads at each base are represented relative to the total number of reads for that sample (see Methods: Graphical Representation for additional details). Approximate locations of key sequence features (numbers indicate basepairs as per reference sequences): Left ITR (0-150); Promoter (300-2100); Transgene (2150-3200); Right ITR (3700-3850); Vector Backbone (3850-6400).

The samples in Subgroup 5 were unique amongst the other subgroups as the backbone used in the pro-viral vector was substantially larger than that used for the others. The insert sequence was designed for gene replacement. The insert size for these vectors was approximately 2.7 kb and the backbone sequence roughly 7 kb. This backbone sequence is over 2 times larger than the backbone used in the other subgroups. This strategy was employed to investigate whether or not this would encourage faithful packaging of the insert as opposed to the backbone sequence. The serotypes investigated in this subgroup were AAV 2/2 and AAV 2/5. Figures 4.4.10 and 4.4.11 illustrate the superior proficiency of AAV 2/2 over AAV 2/5. The results of the subgroups have been collated into Table 4.4.1 for ease of comparison.

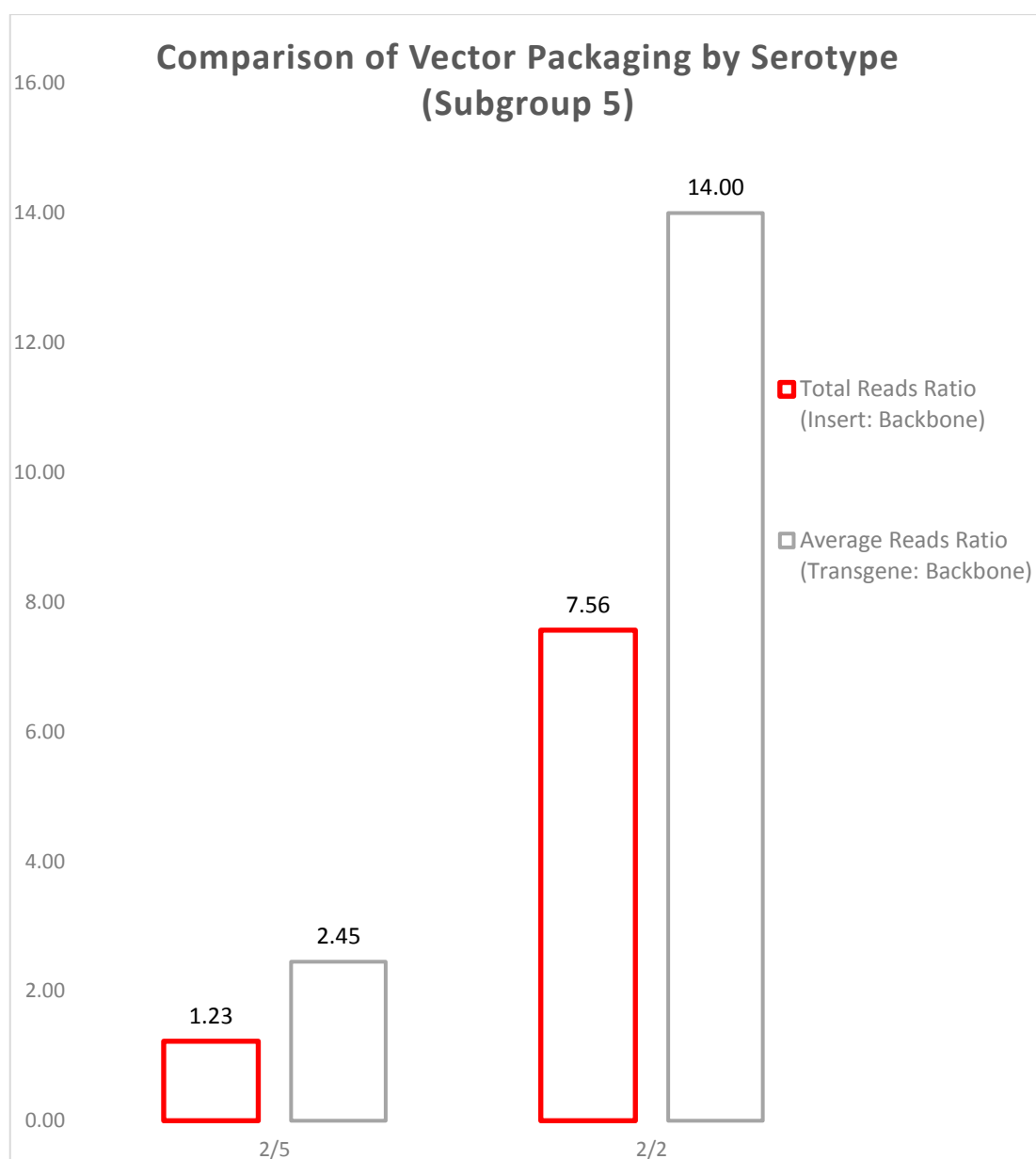


Figure 4.4.10. A graphical representation of relative ratios of vector packaging grouped by AAV serotype in Subgroup 5 (see Methods: Graphical Representation for additional details). Both samples have an identical insert size, backbone size and identical sequences.

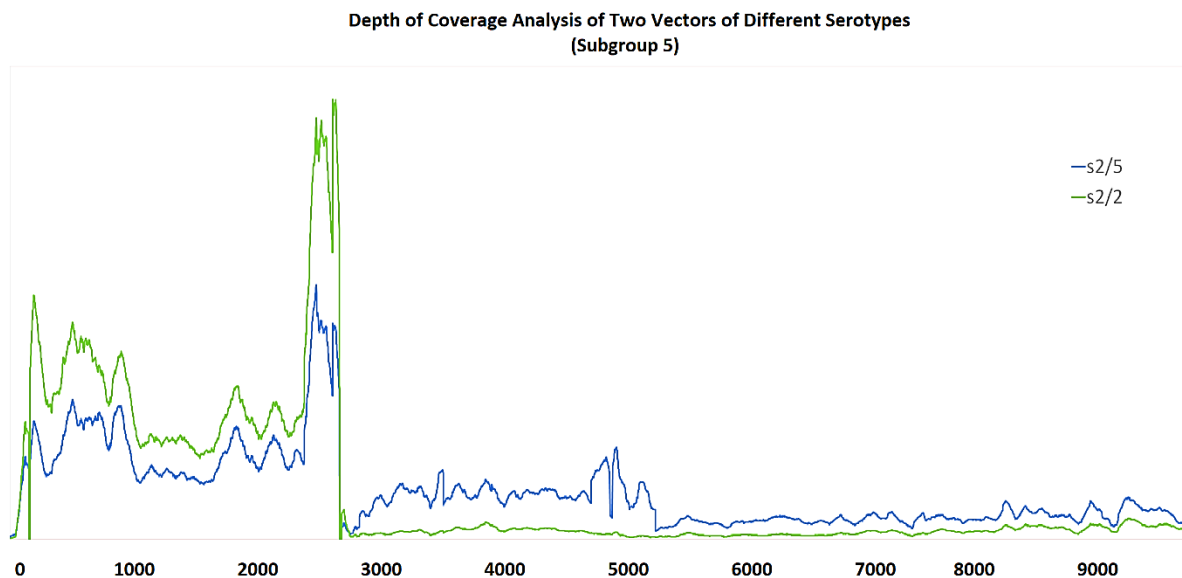


Figure 4.4.11. A relative depth analysis graph of coverage of identical vectors with different serotypes in Subgroup 5, the sequence reads at each base are represented relative to the total number of reads for that sample (see Methods: Graphical Representation for additional details). Approximate locations of key sequence features (numbers indicate basepairs as per reference sequences): Left ITR (0-150); Promoter (250-750); Transgene (950-2450); Right ITR (2750-2900); Vector Backbone (2900-9800).

Table 4.4.1.1. *Compilation of data from 5 AAV subgroups. MCS is a smaller backbone, whereas CHOP is a much larger backbone*

Subgroup	Promoter	Serotype	Backbone	Size (bp)	Insert Size (bp)	Backbone Size (bp)	Size Ratio (Insert: Backbone) in Reference	Total Insert: Backbone	Gene : Backbone (Average Reads)	% Reads in Backbone
1	CBA	2/2	MCS	7086	4400	2686	1.64	4.35	6.11	11.45
1	CBA	2/8	MCS	7086	4400	2686	1.64	3.49	4.84	13.85
1	CBA	2/5	MCS	7086	4400	2686	1.64	3.37	4.07	14.22
2	Rho	2/5	MCS	7091	4360	2731	1.60	1.81	1.96	24.51
2	Rho	2/8	MCS	7091	4360	2731	1.60	9.09	9.73	6.06
3	H1	2/5	MCS	5254	2364	2890	0.82	1.53	1.52	38.59
3	H1	2/8	MCS	5254	2364	2890	0.82	4.71	5.42	17.46
4	Rho	2/8	MCS	6423	3533	2890	1.22	0.51	0.36	56.30
4	Rho	2/5	MCS	6423	3533	2890	1.22	0.85	0.77	43.94
5	CMV	2/5	CHOP	10011	2600	7411	0.35	1.23	2.45	44.74
5	CMV	2/2	CHOP	10011	2600	7411	0.35	7.56	14.00	11.64

4.4.4 Discussion

Many studies to date have illustrated the fact that AAV particles are likely to naturally contain products other than what we expect, if anything at all (319). These studies have used a battery of tests to detect and attempt to rectify these issues. Electron-microscopy with the correct staining procedure can be used for detection of full and empty particles (310). This technique can also be used in combination with ion-exchange chromatography to purify production-scale batches of AAV across a multitude of serotypes (311). Infrared fluorescence imagery can be also used to detect the empty particles (312). Several studies have found empty capsids to be present at shockingly high rates of 50 to 98% (319–321), although refined AAV purification methodologies have been employed to try to minimise this (313).

The purity of these AAV preparations remain an important issue as empty capsids have the potential to interfere with successful gene therapies in several ways. Empty vectors could potentially exacerbate any existing susceptibility to trigger an unnecessary immune response that exists in a host (310). Prior AAV studies have attributed some obstruction of their treatment to immunogenic responses, as this was deemed to effect the longevity of the treatment (322,323). This included the generation of capsid-specific neutralising antibodies (NAbs) which would likely affect any subsequent treatment with that same capsid in the future. Biologically, these empty capsids can compete for transduction of target cells, limiting the efficacy of any AAV therapy with empty capsid contamination (324). Empty capsids can also induce aggregation of capsids which is also likely to negatively affect potential therapies (314).

Unfortunately, empty capsids are not the only issue of contamination of AAV particles. It is evident from the preliminary data shown in this chapter that faithful packaging of vector inserts remains an important issue yet to be resolved. This may potentially be a more worrying issue as truncated or mispackaged sequences are not applicable to many of the detection methods used for recognising empty vectors. Our use of short-read, pair-ended NGS has suggested a preponderance of all AAV serotypes tested to preferentially package from the 3' end of the transgene inserts. This behaviour also appeared to occur independent of vector size.

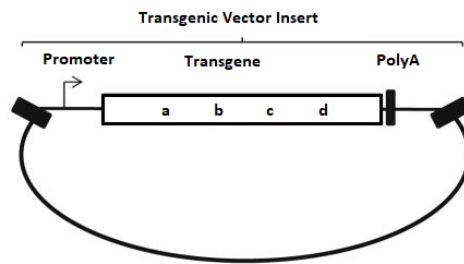
Only AAV 2/5, when utilised with a transgene insert size of approximately 4.5 kb showed an irregular increase of coverage depth at the 5' end of the insert (Subgroups 1 and 2). This preferential use of the 3' ITR for integration has also been observed in long-read NGS from another study (325). This may also possibly be a result of vector preparation, as cloning into sites near the ITRs is likely to disrupt the D-sequences of the ITRs which are directly related to the success of encapsidation of viral sequences (326).

There are some theories among the AAV community as to why this may occur. Truncated sequences are believed to be caused, in part, by incomplete encapsulation (325,327). This appears to occur in the same fashion that attempting encapsulation of inserts greater than 5 kb may occur (Figure 4.4.1). This results in the fragmentation of products at the 5' end of synthesis. This fragmentation would therefore limit the success of transgene recapitulation once transduced into the target cell. This limitation is suspected to be entirely dependent on the abundance of homology between fragmented pairs (Figure

4.4.12), as homology has been observed to be the preferred method of second strand resolution previously (304).

Another interesting observation is that 2/5 in comparison to other serotypes of similar nature, had a preponderance to package the backbone region of the vector. This is particularly evident in Subgroup 5, where the coverage of 2/5 and 2/8 across the insert appear strikingly similar, yet 2/5 contained significantly more reads aligning to the backbone sequence than 2/8. It has been shown in other studies that 2/5 possessed a superior ability to package oversized inserts when compared to other serotypes and this may be the result of a similar mechanism that results in erroneous packaging with smaller inserts (297).

A) AAV Plasmid Vector



B) AAV Reassembly

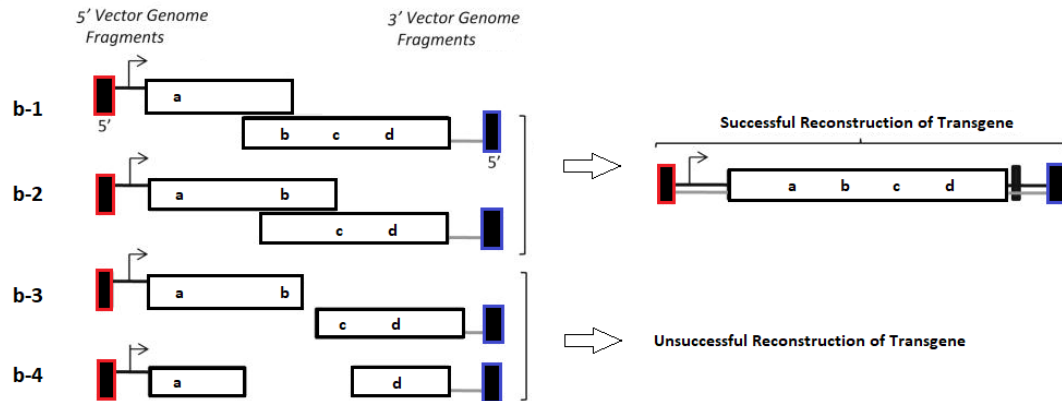


Figure 4.4.12. An illustration of successful and unsuccessful AAV reassembly events. A) An illustration of a typical transgenic vector with components a-d representing elements usually found in a transgenic insert such as a promoter or polyA signal. B) A theoretical display of transgene reconstruction based on single strand annealing. Reassembly in this scenario has accounted for the observed preferential encapsulation of 3' transgene sequence. Scenarios b-1 and b-2 show sufficient fragment overlap to enable successful reassembly whereas situations b-3 and b-4 do not. Image repurposed from Hirsch et al.'s illustration of dual vector transgene packaging (328).

There is also some evidence to suggest that replication can “stall” at hairpin structure in sequences (325,329) and that similar stalling events may occur in promoter sequences, of both eukaryotic and prokaryotic origin, at the replication fork barriers. This can result in hotspots for recombination (325,330). In the case of hairpins, the secondary structure may be misinterpreted almost as ITRs, resulting in the generation of truncated inserts (329).

Previous attempts at NGS for AAVs have been severely limited by their small sample size and lack of diversity, revealing approximately 2% or less of any contaminating sequence (325,327). Here we observe several preparations of AAVs designed for multiple purposes. In each subgroup it was observed that AAV 2/5 contained proportionally more undesirable sequence when compared to other samples in that subgroup, with the only exception of Subgroup 4, whereby uncharacteristic ratios of backbone sequence were encapsidated. It is interesting to note that whilst most AAV preparations showed preferential packaging of the sequence between the ITRs, AAV2/5 consistently showed comparatively higher depth of coverage across the first approximate 5 kb of any sequence, this is particularly evident when a sufficiently large vector backbone was used, Subgroup 5 (Figure 4.4.11).

Taken together, the observations of fragmented genomes and AAV 2/5 observed propensity to package a larger quantity of sequence, this demonstrates a potential for optimisation by the same methods currently being applied to dual-vector strategies. Although limited by packaging capacity in a single vector, hybrid and self-complementary designs may maximise faithful reassembly for smaller therapeutic methodologies such as antisense oligonucleotides targeting pathogenic variants. This may

also take advantage of the self-complementary strategy's bonus of reduced lag time for expression to occur post injection (331).

One of the key findings of this very preliminary study is that there is currently an unmet need for accurate detection and quantification of the contents of filled or partially-filled AAV capsids. Methods such as optical density measurements may be informative in terms of contaminating empty capsids but will not be able to distinguish between capsids with the desired sequence and those containing erroneous sequence or partial sequence. Our results have also indicated that Q-PCR may provide very misleading titre values depending on the specific preparation and probe design.

For example, probes designed against the 3' of the transgene and the 5' of the backbone sequence are very likely to overestimate the titre value of AAV containing full length transgenes albeit providing an accurate view of viral particle dose (assuming empty particles have been eliminated). NGS may not be available to every laboratory currently so an interim measure may be to increase the quantity and coverage of probes across transgenes used to Q-PCR viral genomes for quantification. Given the data presented in this chapter, it is possible that quantification of viral genomes and therefore therapeutic doses could be miscalculated by a significant factor depending on the positioning of Q-PCR probes.

Future work relating to this study will ideally involve the establishment of more accurate AAV alignments. This study has been confined partially by the availability of reference sequences common to multiple AAV samples, that allow for alignment and coverage-based interpretations. It may be more informative to perform a "*de novo* genome

assembly” for each sample. This may reveal more AAV traits that have been currently overlooked, however successful *de novo* genome assembly usually depends on the combination of both long- and short-read length NGS data (332). (332). Furthermore, more detailed investigation of baseline information, for example, the inherent variability between different AAV preparations of the same AAV serotype with the same insert will aid in the future interpretation of results using different AAV serotypes and or different inserts.

Most importantly, the data presented here can only be interpreted as a pseudo-behavioural study for AAV encapsidation. It was largely a comparative study of how similar AAVs differentially encapsidated pro-viral vectors. This data should not be extrapolated to the functional capabilities of any of the discussed viral genomes. It will be critical to evaluate these findings with regard to the relevant efficacy data that results from the experimental implementation of these samples.

In summary, this pilot study serves to highlight the value of NGS in interrogating AAV quality and in trying to elucidate the parameters determining efficient packaging of full-length transgenes, an extremely important issue which can potentially influence the efficacy of AAV therapies and the viral dose of such therapies.

5 Discussion

5.1 Deciphering the Genetic Pathogenesis of Unresolved IRDS

We are now in the era of genomic medicine where genetic information can provide accurate and robust diagnoses, and at times, can be informative regarding prognoses and

can aid in directing optimal therapeutic intervention. This progress has been underpinned by rapid advances in DNA sequence technologies and data analyses tools (Chapters 1.4-1.5). In no clinical field has the impact of these advances been so clearly felt as that of inherited retinal disorders (IRDs). We now understand that more than 300 genes are implicated in IRDs where a few decades ago we naively believed there to be one recessive, one X-linked and one dominant gene (1). As part of this PhD thesis, NGS technologies have been utilised to provide a picture of the genetic architecture of IRDs in the Irish population.

The first results chapter (Chapter 4.1) of the thesis, using data from nearly 1,000 IRD patients sequenced for up to 254 IRD genes, provides an overview of which genes and which mutations are prevalent in the Irish IRD population as well as identifying over 30 previously unreported novel IRD mutations (23). The involvement of a novel mutation in the *RS1* gene in a specific form of ocular disease in a large Irish pedigree is then explored in Chapter 4.2, along with the utility of various in silico tools to interrogate the potential pathogenicity of this novel variant (Dockery/Stevenson et al., manuscript under review). Indeed, issues regarding how best to categorise novel variants of unknown significance are discussed in both Chapters 4.1-4.2. During the course of these PhD studies, a wide array of case histories has resulted in novel and or important genetic and clinical findings from the perspective of IRDs. Chapter 4.3 provides information on a carefully selected group of IRD case histories to demonstrate the significant novelty and value of information emanating from a study such as Target 5000. It is the intention that some of these will be further expanded into publications in the near future.

The final research chapter in the PhD thesis links the elucidation of the genetic pathogenesis of IRDs to the ever-growing field of gene therapy. Certainly, there are now over 30 adeno associated virus (AAV)-mediated gene therapies in clinical trial for IRDs (333) and as stated earlier in the thesis, as of December 2017, Luxturna was the first ocular gene therapy approved for any IRD (265). One of the many key issues relating to gene therapy is the quality of the virus and efficacy of the therapeutic. By applying NGS-based sequencing technologies in a similar manner to that utilised for the main body of the thesis studies, the integrity of AAV preparations has been interrogated particularly with respect to the size and integrity of the packaged transgene. Some interesting preliminary findings have demonstrated that at times a significant proportion of AAV vectors do not contain full length transgenes, an issue that can influence ‘perceived’ viral titre / dose and indeed efficacy. It is planned to expand this work and submit it for publication.

While great strides have been made in elucidating the genetic architecture of IRDs including in the Irish population the subject of this thesis, it is apparent from the body of research detailed in the thesis, that while approximately 60% plus of IRDs have a clear candidate variant that may be causative of disease, currently the genetic pathogenesis of IRDs remains unresolved in ~40% of cases (Chapters 1.5 and 4.1). IRD mutation detection may remain elusive for a variety of reasons, possibly due to inadequate sequence coverage, inappropriate filtering of data (and loss of the pathogenic variant), variants in intronic sequences which affect splicing (such sequences do not form part of most IRD capture panels to date), structural variants such as inversions and duplications which can be difficult to resolve with NGS, regulatory variants or new, as yet

uncharacterized, IRD genes and which therefore were not included in capture panels, among many other potential causes.

However, from some NGS studies there are indicators that support the premise that, within the known IRD genes included in many target capture panels, additional significant levels of pathological genetic variants are present but as yet remain undetected (84,171,334,335). For example, it has been found that there can be a preponderance of patients with one causative mutation in a recessive IRD gene; such over-representation of heterozygotes compared to a control population has been observed in a number of NGS studies, including our own study of the Irish IRD population, for example, for the *ABCA4* gene causative of Stargardt disease (148,162). The majority of clinically defined Irish Stargardt cases who did not have two *ABCA4* mutations in the coding sequence, had a single *ABCA4* mutation, 60% of unresolved Stargardt pedigrees.

Recent studies have focused on revealing the genetic variants underlying such observations. Various tools have been developed to detect copy number variations (CNVs) in NGS datasets and have been employed to successfully detect CNVs in IRD patient cohorts (84,336,337). Indeed, CNVs were identified in a recent WES study in approx. 10% of the 60 IRD patients assessed by exon coverage data analysis and confirmed by PCR (336). Alternative methods for CNV detection, that previously have been employed extensively in diagnostic laboratories for disorders other than IRDs, include quantitative PCR (qPCR), multiplex ligation-dependent probe amplification (MLPA) and comparative genomic hybridization (338,339).

Using the latter, Van Cauwenbergh and colleagues developed a custom microarray (arrEYE) with coding and noncoding sequence from 166 known and candidate IRD genes and 196 noncoding RNAs for CNV detection in IRD patients (337); the CNV detection rate obtained with arrEYE from a first study of 57 IRD patients was 3.5%. In a recent study, 18% of unresolved cases (n = 28) were resolved by CNV mapping (84). It is clear that CNVs may represent a significant contributor to the unresolved cases of IRDs and that methods of CNV analysis, both computational and experimental, should, where possible be included in future IRD studies, which has not always been the case thus far.

In addition to CNVs, a proportion of the remaining unresolved IRD cases may be caused by variants that affect RNA splicing and thereby contribute to disease. Mutations that may have an impact on pre-mRNA splicing can be predicted using various *in silico* tools such as Human Splicing Finder (www.umd.be/HSF/) among others and can be confirmed by transcript analysis. The Midigene approach of analysis uses large genomic fragments to investigate the effect of potential splice variants with subsequent transcription analysis *in vitro* (340). Photoreceptor progenitor cells derived from IRD patient fibroblasts have also proven very informative in terms of splicing analysis (174). RNA sequencing technologies are greatly enabling high throughput transcriptomics and undoubtedly will be applied to a greater extent to explore splice variants in IRDs in the future.

A transcriptomic approach to assess the effects of IRD mutations on splicing may in principle be readily adopted for ubiquitously expressed IRD genes, and for tissue specific IRD genes, using iPS cells differentiated into appropriate cell lineages as per the example above. One approach to identify intronic variants with effects on RNA splicing in IRD patients involves deep intronic sequencing. For example, in a recent study of the *USH2A*

gene, the most frequent cause of Usher syndrome type II (USH2) involving RP and sensorineural hearing loss, an analysis of the whole 800 kb of the *USH2A* gene was prompted by the identification of patients with an USH2 phenotype and a single exonic mutation in the *USH2A* gene. Deep sequencing of intronic sequences enabled the identification of candidate splice mutations and subsequent functional validation of these mutations using reporter minigene assays leading to the resolution of 3 out of 5 USH2 patients with a single exonic mutation (341).

In another study, the functional effect of a commonly observed mutation in the *ABCA4* gene (the c.5461-10T>C variant) causative of Stargardt disease was analysed using patient-derived fibroblasts reprogrammed into induced pluripotent stem (iPS) cells and then differentiated into photoreceptor progenitor cells. This *ABCA4* variant was found to induce skipping of exon 39 or exon 39 and 40 in the mature transcript again using a minigene assay (174). Recently homozygosity mapping and whole genome sequencing revealed a variant deep in intron 18 of the *PROM1* gene causative of a recessive cone-rod dystrophy, resulting in the inclusion of a pseudo-exon in the mutant transcript, which was functionally validated again using a minigene assay (342). Detection of pathogenic IRD intronic variants will elude many of the current capture panel NGS strategies for IRDs which are focused solely on exons and hence the extension of screening studies in the future to include intronic sequences will aid in elucidating what proportion of IRD cases involve aberrant splicing.

In addition to CNVs and splice variants as a source of as yet ‘unresolved’ pathogenic IRD mutations, variants in regulatory sequences, such as, promoter sequences or miRNAs, and their associated target sites, may also be implicated in some forms of IRD. Indeed a

mutation in the seed region of microRNA-204 has been found to segregate in a family with autosomal dominantly inherited retinal dystrophy and bilateral coloboma (343), a 5'UTR sequence in the *NMNAT1* gene has been implicated in a form of LCA (344) and recently a single base mutation in the promoter region of the *CHM* gene has been implicated as causative of choroideremia (345), all highlighting the potential role of regulatory mutations in some forms of IRD. The implementation of more extensive NGS strategies in the future will aid in characterising such IRD regulatory mutations.

In conclusion, while significant advances in high throughput genomic and transcriptomic NGS has greatly facilitated an elucidation of much of the genetic architecture of IRD patient cohorts, substantial levels of unresolved IRD cases still remain. Recently it has become evident that a significant proportion of these will be accounted for by CNVs, splicing defects and regulatory variants. It still remains to be established within these unresolved cases, what proportion will be caused by new, as yet uncharacterised, retinal disease genes.

5.2 A Diagnostic Imperative Driven by Developments in Gene Therapy

There is an obvious rationale for pursuing such studies to their conclusion, in that gene therapies for a growing number of ocular disorders are now in clinical trial (292,346–349). To date, approximately 300 IRD patients have been treated with ocular gene therapies (www.clinicaltrials.gov). In some of these trials, a continued retinal degeneration has been observed (350) and the parameters determining this still need to be elucidated. It is of note however, that an adeno-associated virus (AAV)-RPE65 therapy (voretigene neparvovec, Luxturna) has successfully progressed through to Phase III

clinical trial (CHOP/Spark Therapeutics; www.sparktx.com), as has an AAV-ND4 therapy for Leber hereditary optic neuropathy (LHON) (349).

Data from many ocular gene therapy trials support the view that intraocular delivery of AAV is well tolerated in the human eye and hence represents a safe platform for gene delivery (346–349). A large number of gene therapies employing AAV vectors are also at the stage of preclinical evaluation in animal IRD models, in which benefit has been demonstrated (349,351–355). Thus far many gene therapies have been directed towards recessive forms of IRD, although strategies such as suppression and replacement or genome editing are being considered to address the 30% of IRDs that are dominantly inherited (351,356).

The efficacy of the therapeutic approaches will be determined in part by features specific to individual IRDs. For example, the target cell type within the retina, whether photoreceptors, RPE cells or retinal ganglion cells (349,351–355), among others, will influence vector choice and route of administration. Strategies such as directed evolution are also expediting the generation of AAV serotypes with a predilection for specific retinal cell types (357,358). Additionally, the severity of disease and whether retention of a relatively intact target cell population is maintained over years, thereby providing a significant timeframe for therapeutic intervention, will also greatly influence therapeutic efficacy.

For severe degenerative IRDs, where few or no photoreceptors remain, optogenetic therapies may provide an alternative option by enabling remaining cells to become photosensitive (359,360). While AAV has shown substantial promise for a number of

ocular indications, preclinical and clinical studies have also been undertaken with a variety of other non-viral and viral vectors, for example, nanoparticles and lentiviral vectors amongst others (361,362). Moreover, methodologies for systemic delivery of potentially therapeutic compounds into the retina by modulation of permeability at the inner blood-retina barrier have been extensively tested in animal systems and have employed AAV vectors (363).

Given the rapid pace at which therapies are being developed, it is imperative to establish the genetic architecture of IRDs in different national and ethnic groups, in principle thereby facilitating participation of patients in future human clinical trials or access in the future to marketed gene therapies. In parallel with genotyping of IRDs, it is vital that patients are clinically profiled over time, as this will greatly augment our understanding of the natural histories of these ocular disorders and genotype-phenotype correlations. Knowledge regarding the natural history of a disorder is of direct value to patients, and moreover, will in principle facilitate future patient participation in clinical trials and will aid in informing the choice of appropriate primary and secondary endpoints for clinical trial design. Therefore, as such, identification of the disease-causing mutations is of immediate clinical relevance to the entire IRD patient population.

5.3 Challenges for Clinical Incorporation of NGS Technologies

As discussed in a previous section (Chapter 1.4.2), although there are similarities and parallels between library preparation and sequencing platforms, it is the lack of uniformity that limits the application of this technology, particularly for clinical use. The large variation in both sample handling and sequence data acquisition means that

standardising the quality control and quality assurance aspects of this technology may prove especially difficult. It is likely that some instruments may require further optimisation before they will be incorporated officially into clinical practice.

The data produced from sequencing experiments is not only massive in size but also may require some advanced knowledge of bioinformatics in order for it to be interpreted correctly and utilised to its full potential. This would not only necessitate the acquisition of large specialised servers but also the training and allocation of personnel capable of managing these and the associated data.

The issue of variant annotation and interpretation has made significant progress in recent years, most prominently influenced by the publication of the ACMG Guidelines (144), which outlines specific lines of evidence that should be considered when assigning the confidence of pathogenicity or benignity to a variant, as outlined in Chapter 3.13. However, a challenge that remains is that some variants, which may be disease causing, will continue to be assigned a ‘unresolved’ status due to as yet ineffective interrogation methods, although a number may indeed be causative. The implementation of functional studies for such variants may help to resolve this, although such work is costly and therefore may not be realised for many variants.

Once a variant has been appropriately interpreted, there then is the issue of reporting this variant to the required audience. The complexity of this scenario varies greatly with the confidence in the variant’s pathogenicity and the clinical consequences of reporting that variant. Clinically, establishing a potentially causative variant could impact not only the

diagnosis, but also the prognosis, the availability of therapeutics options and indeed can have implications for offspring from the patient.

Whole exome and whole genome sequencing also inherently come with an increased risk of discovering incidental findings. For the most part such findings may not be essential to report depending on the associated ethical approval. However, any findings that are concurrently associated with medical implications and that are clinically actionable may cause ethical dilemmas and therefore may also carry a responsibility to report such results back to the patient. The ACMG have also written guidelines which refer to the discovery of secondary findings such as these. The term “secondary findings” has been deemed more appropriate as “incidental” implies that that target was unintentionally included in the study, which is rarely the case (364). The secondary findings described in these guidelines primarily apply to a specific compiled list of genes whereby variants would be clinically actionable. These suggested guidelines also propose an option be given to the patient to opt-out of receiving information regarding the secondary findings applicable to their genetic test.

It is also exceptionally vital to consider that genetic information rarely impacts a single individual. Genetic information can not only affect the family planning decisions of those tested but may also lead to the inference of information about that individual’s family members. In Ireland, as of the National Rare Disease Plan 2014-2018 (365), when family members are known to be potential carriers for rare diseases these family members are entitled to pre-conception genetic screening and counselling to better inform them of the risks and facts of the situation. Sadly, in reality this is an underfunded and understaffed resource in Ireland. In the superiorly resourced UK, recruitment has been largely

successful for risk factor genetic screens when a family member has shown to be affected by a certain genetic condition. The NHS in the UK have recently announced at the start of September that they have sequenced their 81,179th genome and thanked their population base for their eagerness to participate (366).

This can understandably result in very distressing outcomes for the relatives of tested individuals. Nevertheless, in a recent melanoma risk factor study, 95% of participants reported that they did not regret receiving the information about genomic risk score (367). However, it was also found that 60% of family members of Huntington's Disease patients in India did not want the results of their test to be publicised (368). Taken collectively, this could be an indication that genomic risk stratification on a population-scale could be the next step forward and would be well tolerated by the public if the results were managed appropriately. This era of genomic testing could involve screening strategies and personalised prevention plans, which could possibly fill the interim gap between current medicine and generally available personalised medicine.

Although genetic screening is not a new concept, some potential harms associated with misdiagnosis have been illuminated in the past which may have impeded further implementation of similar studies. Previously, a UK population screen was found to over-diagnose patients with a breast cancer risk. It has been extrapolated that one cancer death would be prevented for every three individuals who were over-diagnosed and treated (369). However, in a similar population screen conducted in Denmark, it was estimated that for every woman over-diagnosed, the lives of two to three women would be saved (370). These studies demonstrate the inter-variability possible in such a strategy.

The most immediate and possibly the largest road block for implementing NGS/TGS in clinical practice may be financing the associated cost. NGS may be getting cheaper with each passing year, although this does not guarantee that all health services will be accommodating in picking up the bill for what may be considered a non-essential expenditure. This will undoubtedly have implications for certified laboratories, government policy makers and private health care providers. The policy put forward would also have to be deemed agreeable by healthcare professionals, clinical geneticists and lastly, the insurance industry (371). One cancer-based report suggests costs of over \$1 million per year to clinically sequence a small cancer gene panel in 800 patients (372).

One could argue that the value of patient sequence data will only increase over time. This is primarily for two reasons - we not only gain a greater understanding of the frequency of variation we observe in the human genome, but also more evidence to support the damaging impact that the variation may have. This information can be used to create more efficient, evidence-based therapeutic strategies. A fantastic practical example of this is the development of the exon skipping gene therapy developed by ProQR, QR-421a, for conditions associated with the *CEP290* and *USH2A* genes (373). An estimated 16,000 patients in the Western world experience vision loss due to pathogenic variants in the exon 13 region of *USH2A*. This therapeutic is currently in the pre-clinical stage of testing but holds enormous promise for many patients without an option of alternative treatment.

Recently the NIH in the United States have endorsed a study which outlines all of the useful clinical applications of detecting foetal DNA in maternal plasma (374). This is not limited to basic variant testing but also includes the investigation of RNA, epigenetic modifications and copy number variants (375). The benefits to those enrolled patients to

participate in such tests, as part of a national health care system, could be potentially massive as this type of early detection screening could allow for a far wider therapeutic window than ever previously available. Not only does the prenatal detection of the genetic source of a disease allow for diagnostic testing, it also allows appropriate counselling of parents and family members. The importance of offering the choice of prenatal diagnosis to families and couples in this circumstance is not easily quantified but must be comprehended and highlighted.

Although it is possible that some technologies around NGS have perhaps reached their pinnacle of performance, new innovative ways to utilise NGS are emerging constantly. Recently, tissue from mouse brain and human breast cancer was analysed by a new type of RNA-sequencing, called “Spatial Transcriptomics” (376). This type of sequencing uses fixed tissue to detect differences in mRNA levels in a 2D plane, essentially providing superior information regarding gene expression patterns. This application could help improve our understanding of disease manifestation and progression. Armed with this knowledge, gene therapists could potentially formulate better strategies for target regions and optimal time of delivery. However, caution is required as interpreting the spatial resolution of these data has many limitations including sample preparation, data quality and data processing (377).

In summary, these are exciting times for the field of variant identification. Coinciding with the rapid evolution of NGS, IRD investigations have entered an age of data acquisition on an unprecedented scale. As this data eventually becomes compiled and appropriately interpreted, it is likely that many new associations and pathomechanisms will be finally elucidated. We are now capable of DNA, RNA and protein interrogation

far beyond anything that was available to our predecessors. The onus is now on researchers today to continue this voyage of discovery and illumination which started with linkage studies and radioisotopes in dark rooms, because with great power comes great responsibility (378).

6 References

1. RetNet - Retinal Information Network [Internet]. [cited 2018 Jun 22]. Available from: <https://sph.uth.edu/RetNet/>
2. Ferrari S, Di Iorio E, Barbaro V, Ponzin D, Sorrentino FS, Parmeggiani F. Retinitis Pigmentosa: Genes and Disease Mechanisms. *Curr Genomics*. 2011 Jun;12(4):238–49.
3. Ali MU, Rahman MSU, Cao J, Yuan PX. Genetic characterization and disease mechanism of retinitis pigmentosa; current scenario. *3 Biotech* [Internet]. 2017 Aug [cited 2018 Sep 14];7(4). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5515732/>
4. Bangal S, Bhandari A, Dhaytadak P, Gogri P. Gyrate atrophy of choroid and retina with myopia, cataract and systemic proximal myopathy: A rare case report from rural India. *Australas Med J*. 2012 Dec 31;5(12):639–42.
5. Farrar GJ, Carrigan M, Dockery A, Millington-Ward S, Palfi A, Chadderton N, et al. Toward an elucidation of the molecular genetics of inherited retinal degenerations. *Hum Mol Genet*. 2017 Aug 1;26(R1):R2–11.
6. Purves D, Augustine GJ, Fitzpatrick D, Katz LC, LaMantia A-S, McNamara JO, et al. Anatomical Distribution of Rods and Cones. *Neurosci 2nd Ed* [Internet]. 2001 [cited 2018 Jul 8]; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK10848/>
7. Zele AJ, Cao D. Vision under mesopic and scotopic illumination. *Front Psychol* [Internet]. 2015 Jan 22 [cited 2018 Jul 8];5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4302711/>
8. Human Eye Anatomy - Parts of the Eye Explained [Internet]. All About Vision. [cited 2019 Apr 23]. Available from: <https://www.allaboutvision.com/resources/anatomy.htm>
9. Willoughby Colin E, Ponzin Diego, Ferrari Stefano, Lobo Aires, Landau Klara, Omid Yadollah. Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function – a review. *Clin Experiment Ophthalmol*. 2010 Aug 5;38(s1):2–11.
10. Desai S, Alibhai AY. Cone Dystrophy. In: *Atlas of Retinal OCT: Optical Coherence Tomography* [Internet]. Elsevier; 2018 [cited 2018 Aug 11]. p. 154. Available from: <http://linkinghub.elsevier.com/retrieve/pii/B9780323461214000704>

11. Huet RAC van, Estrada-Cuzcano A, Banin E, Rotenstreich Y, Hipp S, Kohl S, et al. Clinical Characteristics of Rod and Cone Photoreceptor Dystrophies in Patients With Mutations in the C8orf37 Gene. *Invest Ophthalmol Vis Sci*. 2013 Jul 1;54(7):4683–90.
12. Kaway CS, Adams MKM, Jenkins KS, Layton CJ. A Novel ABCA4 Mutation Associated with a Late-Onset Stargardt Disease Phenotype: A Hypomorphic Allele? *Case Rep Ophthalmol*. 2017 Mar 9;8(1):180–4.
13. Ayuso C, Millan JM. Retinitis pigmentosa and allied conditions today: a paradigm of translational research. *Genome Med*. 2010 May 27;2(5):34.
14. Böhm S, Riedmayr LM, Nguyen ONP, Gießl A, Liebscher T, Butz ES, et al. Peripherin-2 and Rom-1 have opposing effects on rod outer segment targeting of retinitis pigmentosa-linked peripherin-2 mutants. *Sci Rep*. 2017 May 24;7(1):2321.
15. Lemoine S, Panaye M, Rabeyrin M, Errazuriz-Cerda E, Mousson de Camaret B, Petiot P, et al. Renal Involvement in Neuropathy, Ataxia, Retinitis Pigmentosa (NARP) Syndrome: A Case Report. *Am J Kidney Dis Off J Natl Kidney Found*. 2018 May;71(5):754–7.
16. Thompson DA, McHenry CL, Li Y, Richards JE, Othman MI, Schwinger E, et al. Retinal dystrophy due to paternal isodisomy for chromosome 1 or chromosome 2, with homoallelism for mutations in RPE65 or MERTK, respectively. *Am J Hum Genet*. 2002 Jan;70(1):224–9.
17. Thompson JA, Roach JND, McLaren TL, Lamey TM. A Mini-Review: Leber Congenital Amaurosis: Identification of Disease-Causing Variants and Personalised Therapies. In: *Retinal Degenerative Diseases* [Internet]. Springer, Cham; 2018 [cited 2018 Aug 12]. p. 265–71. (Advances in Experimental Medicine and Biology). Available from: https://link.springer.com/chapter/10.1007/978-3-319-75402-4_32
18. Chandrasekar SP, Namboothiri S, Sen P, Sarangapani S. Screening for mutation hotspots in Bardet-Biedl syndrome patients from India. *Indian J Med Res*. 2018 Feb;147(2):177–82.
19. Kurtenbach A, Hahn G, Kernstock C, Hipp S, Zobor D, Stingl K, et al. Usher Syndrome and Color Vision. *Curr Eye Res*. 2018 Jul 17;0(0):1–7.
20. What is Stargardt Disease? | blindness.org [Internet]. [cited 2018 Aug 12]. Available from: <https://www.blindness.org/stargardt-disease>

21. Shastri BS. Evaluation of the common variants of the ABCA4 gene in families with Stargardt disease and autosomal recessive retinitis pigmentosa. *Int J Mol Med*. 2008 Jun;21(6):715–20.
22. Cremers FPM, Boon CJF, Bujakowska K, Zeitz C. Special Issue Introduction: Inherited Retinal Disease: Novel Candidate Genes, Genotype–Phenotype Correlations, and Inheritance Models. *Genes* [Internet]. 2018 Apr 14 [cited 2018 Aug 12];9(4). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5924557/>
23. Dockery A, Stephenson K, Keegan D, Wynne N, Silvestri G, Humphries P, et al. Target 5000: Target Capture Sequencing for Inherited Retinal Degenerations. *Genes*. 2017 Nov 3;8(11):304.
24. Rivolta C, Sweklo EA, Berson EL, Dryja TP. Missense mutation in the USH2A gene: association with recessive retinitis pigmentosa without hearing loss. *Am J Hum Genet*. 2000 Jun;66(6):1975–8.
25. Mitchell GA, Brody LC, Looney J, Steel G, Suchanek M, Dowling C, et al. An initiator codon mutation in ornithine-delta-aminotransferase causing gyrate atrophy of the choroid and retina. *J Clin Invest*. 1988 Feb;81(2):630–3.
26. McWilliam P, Farrar GJ, Kenna P, Bradley DG, Humphries MM, Sharp EM, et al. Autosomal dominant retinitis pigmentosa (ADRP): localization of an ADRP gene to the long arm of chromosome 3. *Genomics*. 1989 Oct;5(3):619–22.
27. Cremers FP, van de Pol DJ, van Kerkhoff LP, Wieringa B, Ropers HH. Cloning of a gene that is rearranged in patients with choroideraemia. *Nature*. 1990 Oct 18;347(6294):674–7.
28. What is a rare disease? [Internet]. [cited 2018 Aug 12]. Available from: <http://www.raredisease.org.uk/what-is-a-rare-disease/>
29. FAQs About Rare Diseases | Genetic and Rare Diseases Information Center (GARD) – an NCATS Program [Internet]. [cited 2018 Aug 12]. Available from: <https://rarediseases.info.nih.gov/diseases/pages/31/faqs-about-rare-diseases>
30. Clinicians & Researchers [Internet]. Rare Diseases Ireland. [cited 2018 Aug 12]. Available from: <http://rdi.ie/menu/clinicians-and-researchers/>
31. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977 Dec;74(12):5463–7.

32. Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB. Nucleotide sequence of bacteriophage λ DNA. *J Mol Biol.* 1982 Dec 25;162(4):729–73.
33. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. Fluorescence detection in automated DNA sequence analysis. *Nature.* 1986;321(6071):674–9.
34. Hunkapiller T, Kaiser RJ, Koop BF, Hood L. Large-scale and automated DNA sequence determination. *Science.* 1991;254(5028):59–67.
35. Chan EY. Advances in sequencing technology. *Mutat Res Mol Mech Mutagen.* 2005 Jun 3;573(1):13–40.
36. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The Diploid Genome Sequence of an Individual Human. *PLOS Biol.* 2007 Sep 4;5(10):e254.
37. Estevezj. English: The Sanger (chain-termination) method for DNA sequencing. (1) A primer is annealed to a sequence, (2) Reagents are added to the primer and template, including: DNA polymerase, dNTPs, and a small amount of all four dideoxynucleotides (ddNTPs) labeled with fluorophores. During primer elongation, the random insertion of a ddNTP instead of a dNTP terminates synthesis of the chain because DNA polymerase cannot react with the missing hydroxyl. This produces all possible lengths of chains. (3) The products are separated on a single lane capillary gel, where the resulting bands are read by a imaging system. (4) This produces several hundred thousand nucleotides a day, data which require storage and subsequent computational analysis. [Internet]. 2012 [cited 2018 Jul 27]. Available from: <https://commons.wikimedia.org/wiki/File:Sanger-sequencing.svg>
38. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature.* 2008 Apr;452(7189):872–6.
39. Human Genome Project Completion: Frequently Asked Questions [Internet]. National Human Genome Research Institute (NHGRI). [cited 2018 Jul 28]. Available from: <https://www.genome.gov/11006943/human-genome-project-completion-frequently-asked-questions/>
40. Susan Tousi and Illumina's Path to the \$100 Genome [Internet]. Sync Magazine. 2017 [cited 2018 Jul 28]. Available from: <https://sync-magazine.com/2017/susan-tousi-illumina-one-hundred-dollar-genome/>

41. Sequencing Technology | Sequencing by synthesis [Internet]. [cited 2018 Jul 29]. Available from: <https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>
42. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005 Sep 15;437(7057):376–80.
43. Guo J, Xu N, Li Z, Zhang S, Wu J, Kim DH, et al. Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc Natl Acad Sci U S A*. 2008 Jul 8;105(27):9145–50.
44. Fuller CW, Middendorf LR, Benner SA, Church GM, Harris T, Huang X, et al. The challenges of sequencing by synthesis. *Nat Biotechnol*. 2009 Nov;27(11):1013–23.
45. Peacock SJ, Parkhill J, Brown NM. Changing the paradigm for hospital outbreak detection by leading with genomic surveillance of nosocomial pathogens. *Microbiology* [Internet]. 2018 Jul 27 [cited 2018 Jul 29]; Available from: <http://sci-hub.tw/http://mic.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000700/#tab2>
46. Kitsios GD, Fitch A, Manatakis DV, Rapport SF, Li K, Qin S, et al. Respiratory Microbiome Profiling for Etiologic Diagnosis of Pneumonia in Mechanically Ventilated Patients. *Front Microbiol* [Internet]. 2018 [cited 2018 Jul 29];9. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.01413/full>
47. Picelli S, Björklund ÅK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res*. 2014 Dec;24(12):2033–40.
48. Wang Q, Gu L, Adey A, Radlwimmer B, Wang W, Hovestadt V, et al. Tagmentation-based whole-genome bisulfite sequencing. *Nat Protoc*. 2013 Oct;8(10):2022–32.
49. Bio-IT World [Internet]. [cited 2018 Jul 29]. Available from: <http://www.bio-itworld.com>

50. Dolan KA. Illumina CEO On The Lunch That Launched Him [Internet]. Forbes. [cited 2018 Jul 29]. Available from: 2010/09/16/careers-life-sciences-intelligent-technology-illumina
51. Top 10 Sequencing Companies | The Lists [Internet]. GEN. [cited 2018 Jul 30]. Available from: <https://www.genengnews.com/the-lists/top-10-sequencing-companies/77901078>
52. Sequencing Platforms | Compare NGS platforms (benchtop, production-scale) [Internet]. [cited 2018 Jul 29]. Available from: <https://www.illumina.com/systems/sequencing-platforms.html>
53. Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, et al. Next-generation sequencing and its applications in molecular diagnostics. *Expert Rev Mol Diagn.* 2011;11(3):333–43.
54. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of Next-Generation Sequencing Systems [Internet]. BioMed Research International. 2012 [cited 2018 Jul 29]. Available from: <https://www.hindawi.com/journals/bmri/2012/251364/>
55. Pillai S, Gopalan V, Lam AK-Y. Review of sequencing platforms and their applications in pheochromocytoma and paragangliomas. *Crit Rev Oncol Hematol.* 2017 Aug;116:58–67.
56. Metzker ML. Sequencing technologies — the next generation. *Nat Rev Genet.* 2010 Jan;11(1):31–46.
57. Froehlich T, Heindl D, Roesler A. Miniaturized, high-throughput nucleic acid analysis [Internet]. US20100248237A1, 2010 [cited 2018 Aug 3]. Available from: <https://patents.google.com/patent/US20100248237/en>
58. Speights K. Pacific Biosciences Rockets on Roche Diagnostics Deal [Internet]. The Motley Fool. 2013 [cited 2018 Jul 30]. Available from: <https://www.fool.com/investing/general/2013/09/25/pacific-biosciences-rockets-on-roche-diagnostics-d.aspx>
59. Check Hayden E. Nanopore genome sequencer makes its debut. *Nat News* [Internet]. [cited 2018 Aug 5]; Available from: <http://www.nature.com/news/nanopore-genome-sequencer-makes-its-debut-1.10051>
60. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, et al. Long-read metagenomics reveals cryptic and abundant marine viruses. *bioRxiv.* 2018 Jun 12;345041.

61. Todd SM, Settlage RE, Lahmers KK, Slade DJ. *Fusobacterium* Genomics Using MinION and Illumina Sequencing Enables Genome Completion and Correction. *mSphere*. 2018 Aug 29;3(4):e00269-18.
62. Faria NR, Kraemer MUG, Hill S, Jesus JG de, Aguiar RS de, Iani FCM, et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *bioRxiv*. 2018 Apr 16;299842.
63. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif*. 2015 Mar;3:1–8.
64. MinION [Internet]. [cited 2018 Aug 5]. Available from: <https://nanoporetech.com/products/minion>
65. Ambardar S, Gupta R, Trakroo D, Lal R, Vakhlu J. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J Microbiol*. 2016 Dec 1;56(4):394–404.
66. Mardis ER. Next-Generation Sequencing Platforms. *Annu Rev Anal Chem*. 2013 Jun 12;6(1):287–303.
67. Bolz HJ. [Next-Generation Sequencing: A Quantum Leap in Ophthalmology Research and Diagnostics]. *Klin Monatsbl Augenheilkd*. 2017 Mar;234(3):280–8.
68. Bernardis I, Chiesi L, Tenedini E, Artuso L, Percesepe A, Artusi V, et al. Unravelling the Complexity of Inherited Retinal Dystrophies Molecular Testing: Added Value of Targeted Next-Generation Sequencing. *BioMed Res Int*. 2016;2016:6341870.
69. Consugar MB, Navarro-Gomez D, Place EM, Bujakowska KM, Sousa ME, Fonseca-Kelly ZD, et al. Panel-based Genetic Diagnostic Testing for Inherited Eye Diseases is Highly Accurate and Reproducible and More Sensitive for Variant Detection Than Exome Sequencing. *Genet Med Off J Am Coll Med Genet*. 2015 Apr;17(4):253–61.
70. Ellingford JM, Barton S, Bhaskar S, O'Sullivan J, Williams SG, Lamb JA, et al. Molecular findings from 537 individuals with inherited retinal disease. *J Med Genet*. 2016 Nov;53(11):761–7.
71. Tiwari A, Bahr A, Bähr L, Fleischhauer J, Zinkernagel MS, Winkler N, et al. Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. *Sci Rep*. 2016 29;6:28755.

72. Riera M, Navarro R, Ruiz-Nogales S, Méndez P, Burés-Jelstrup A, Corcóstegui B, et al. Whole exome sequencing using Ion Proton system enables reliable genetic diagnosis of inherited retinal dystrophies. *Sci Rep*. 2017 09;7:42078.
73. Haer-Wigman L, van Zelst-Stams WA, Pfundt R, van den Born LI, Klaver CC, Verheij JB, et al. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur J Hum Genet EJHG*. 2017;25(5):591–9.
74. Audo I, Bujakowska KM, Léveillard T, Mohand-Saïd S, Lancelot M-E, Germain A, et al. Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J Rare Dis*. 2012 Jan 25;7:8.
75. de Castro-Miró M, Tonda R, Escudero-Ferruz P, Andrés R, Mayor-Lorenzo A, Castro J, et al. Novel Candidate Genes and a Wide Spectrum of Structural and Point Mutations Responsible for Inherited Retinal Dystrophies Revealed by Exome Sequencing. *PloS One*. 2016;11(12):e0168966.
76. Neveling K, Collin RWJ, Gilissen C, van Huet RAC, Visser L, Kwint MP, et al. Next-generation genetic testing for retinitis pigmentosa. *Hum Mutat*. 2012 Jun;33(6):963–72.
77. O’Sullivan J, Mullaney BG, Bhaskar SS, Dickerson JE, Hall G, O’Grady A, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J Med Genet*. 2012 May;49(5):322–6.
78. Shanks ME, Downes SM, Copley RR, Lise S, Broxholme J, Hudspith KA, et al. Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease. *Eur J Hum Genet*. 2013 Mar;21(3):274–80.
79. Ge Z, Bowles K, Goetz K, Scholl HPN, Wang F, Wang X, et al. NGS-based Molecular diagnosis of 105 eyeGENE® probands with Retinitis Pigmentosa. *Sci Rep [Internet]*. 2015 Dec 15 [cited 2018 Sep 17];5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4678898/>
80. Zhao L, Wang F, Wang H, Li Y, Alexander S, Wang K, et al. Next-generation sequencing-based molecular diagnosis of 82 retinitis pigmentosa probands from Northern Ireland. *Hum Genet*. 2015 Feb;134(2):217–30.
81. Perez-Carro R, Corton M, Sánchez-Navarro I, Zurita O, Sanchez-Bolivar N, Sánchez-Alcudia R, et al. Panel-based NGS Reveals Novel Pathogenic Mutations in Autosomal Recessive Retinitis Pigmentosa. *Sci Rep [Internet]*.

- 2016 Jan 25 [cited 2018 Sep 17];6. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4726392/>
82. Weisschuh N, Mayer AK, Strom TM, Kohl S, Glöckle N, Schubach M, et al. Mutation Detection in Patients with Retinal Dystrophies Using Targeted Next Generation Sequencing. PLoS ONE [Internet]. 2016 Jan 14 [cited 2018 Sep 17];11(1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4713063/>
 83. Siemiatkowska AM, Collin RWJ, den Hollander AI, Cremers FPM. Genomic approaches for the discovery of genes mutated in inherited retinal degeneration. Cold Spring Harb Perspect Med. 2014 Jun 17;4(8).
 84. Bujakowska KM, Fernandez-Godino R, Place E, Consugar M, Navarro-Gomez D, White J, et al. Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. Genet Med Off J Am Coll Med Genet. 2017;19(6):643–51.
 85. Carrigan M, Duignan E, Humphries P, Palfi A, Kenna PF, Farrar GJ. A novel homozygous truncating GNAT1 mutation implicated in retinal degeneration. Br J Ophthalmol. 2016 Apr;100(4):495–500.
 86. Shankar SP, Hughbanks-Wheaton DK, Birch DG, Sullivan LS, Conneely KN, Bowne SJ, et al. Autosomal Dominant Retinal Dystrophies Caused by a Founder Splice Site Mutation, c.828+3A>T, in PRPH2 and Protein Haplotypes in trans as Modifiers. Invest Ophthalmol Vis Sci. 2016 Feb;57(2):349–59.
 87. Tee JJL, Smith AJ, Hardcastle AJ, Michaelides M. RPGR-associated retinopathy: clinical features, molecular genetics, animal models and therapeutic options. Br J Ophthalmol. 2016;100(8):1022–7.
 88. Vollrath D, Yasumura D, Benchorin G, Matthes MT, Feng W, Nguyen NM, et al. Tyro3 Modulates Mertk-Associated Retinal Degeneration. PLoS Genet [Internet]. 2015 Dec 11 [cited 2018 Sep 17];11(12). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4687644/>
 89. Rose AM, Shah AZ, Venturini G, Krishna A, Chakravarti A, Rivolta C, et al. Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa. Sci Rep [Internet]. 2016 Jan 19 [cited 2018 Sep 17];6. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4725990/>

90. Conley SM, Stuck MW, Watson JN, Naash MI. Rom1 converts Y141C-Prph2-associated pattern dystrophy to retinitis pigmentosa. *Hum Mol Genet*. 2017 Feb 1;26(3):509–18.
91. Pierce EA, Bennett J. The Status of RPE65 Gene Therapy Trials: Safety and Efficacy. *Cold Spring Harb Perspect Med* [Internet]. 2015 Sep [cited 2018 Sep 17];5(9). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4561397/>
92. Lecuit M, Eloit M. The potential of whole genome NGS for infectious disease diagnosis. *Expert Rev Mol Diagn*. 2015;15(12):1517–9.
93. Carrigan M, Duignan E, Malone CPG, Stephenson K, Saad T, McDermott C, et al. Panel-Based Population Next-Generation Sequencing for Inherited Retinal Degenerations. *Sci Rep*. 2016 14;6:33248.
94. Riveiro-Alvarez R, Lopez-Martinez M-A, Zernant J, Aguirre-Lamban J, Cantalapiedra D, Avila-Fernandez A, et al. Outcome of ABCA4 disease-associated alleles in autosomal recessive retinal dystrophies: retrospective analysis in 420 Spanish families. *Ophthalmology*. 2013 Nov;120(11):2332–7.
95. Schulz HL, Grassmann F, Kellner U, Spital G, Rütther K, Jägle H, et al. 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. *Invest Ophthalmol Vis Sci*. 2017 Jan;58(1):394–403.
96. Federspiel CA, Bertelsen M, Kessel L. Vitamin A in Stargardt disease—an evidence-based update. *Ophthalmic Genet*. 2018 Jun 25;1–5.
97. Heller D, Weiner C, Nasie I, Anikster Y, Landau Y, Koren T, et al. Reversal of cystoid macular edema in gyrate atrophy patients. *Ophthalmic Genet*. 2017;38(6):549–54.
98. Robson AG, Nilsson J, Li S, Jalali S, Fulton AB, Tormene AP, et al. ISCEV guide to visual electrodiagnostic procedures. *Doc Ophthalmol Adv Ophthalmol*. 2018;136(1):1–26.
99. Cheng C, Jia J-L, Ran S-Y. Polyethylene glycol and divalent salt-induced DNA reentrant condensation revealed by single molecule measurements. *Soft Matter*. 2015 May 6;11(19):3927–35.

100. Lis JT, Schleif R. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. *Nucleic Acids Res.* 1975 Mar;2(3):383–9.
101. SPRI bead mix - OpenWetWare [Internet]. [cited 2018 Jul 15]. Available from: https://openwetware.org/wiki/SPRI_bead_mix
102. García-García G, Baux D, Faugère V, Moclyn M, Koenig M, Claustres M, et al. Assessment of the latest NGS enrichment capture methods in clinical context. *Sci Rep.* 2016 Feb 11;6:20948.
103. DNA Polymerase Selection Chart | NEB [Internet]. [cited 2018 Jul 19]. Available from: <https://international.neb.com/tools-and-resources/selection-charts/dna-polymerase-selection-chart>
104. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* [Internet]. [cited 2018 Jul 25]; Available from: <https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbx108/4106928>
105. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* 2013 Jul 1;41(W1):W597–600.
106. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A.* 1988 Apr;85(8):2444–8.
107. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 2010 Apr;38(6):1767–71.
108. GRCh37 - hg19 - Genome - Assembly - NCBI [Internet]. [cited 2018 Aug 14]. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/
109. GRCh38 - hg38 - Genome - Assembly - NCBI [Internet]. [cited 2018 Aug 14]. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/
110. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009 Jul 15;25(14):1754–60.
111. Picard Tools - By Broad Institute [Internet]. [cited 2018 Jun 22]. Available from: <http://broadinstitute.github.io/picard/>

112. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. ArXiv12073907 Q-Bio [Internet]. 2012 Jul 17 [cited 2018 Jun 25]; Available from: <http://arxiv.org/abs/1207.3907>
113. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011 May;43(5):491–8.
114. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. [cited 2018 Aug 18]. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
115. Chen S, Huang T, Zhou Y, Han Y, Xu M, Gu J. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics*. 2017 Mar 14;18(3):80.
116. Sequence Variant Nomenclature [Internet]. [cited 2018 Aug 19]. Available from: <http://varnomen.hgvs.org/recommendations/DNA/variant/delins/>
117. Kitts A, Phan L, Ward M, Holmes JB. The Database of Short Genetic Variation (dbSNP) [Internet]. National Center for Biotechnology Information (US); 2014 [cited 2018 Aug 19]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK174586/>
118. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014 Jan 1;42(Database issue):D980–5.
119. Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, et al. NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res*. 2014 Jan 1;42(Database issue):D975–9.
120. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012 Jun;6(2):80–92.
121. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front Genet* [Internet]. 2012 Mar 15 [cited 2018 Aug 19];3. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3304048/>

122. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat.* 37(3):235–41.
123. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010 Apr;7(4):248–9.
124. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073–81.
125. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc.* 2016 Jan;11(1):1–9.
126. Min L, Nie M, Zhang A, Wen J, Noel SD, Lee V, et al. Computational Analysis of Missense Variants of G Protein-Coupled Receptors Involved in the Neuroendocrine Regulation of Reproduction. *Neuroendocrinology.* 2016;103(3–4):230–9.
127. Flanagan SE, Patch A-M, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomark.* 2010 Aug;14(4):533–7.
128. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet.* 2015 Apr 15;24(8):2125–37.
129. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet.* 2016;48(12):1581–6.
130. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet.* 2016 Oct 6;99(4):877–85.
131. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature.* 2011 Oct 12;478(7370):476–82.
132. Sheck L, Davies WIL, Moradi P, Robson AG, Kumaran N, Liasis AC, et al. Leber Congenital Amaurosis Associated with Mutations in CEP290, Clinical Phenotype, and Natural History in Preparation for Trials of Novel Therapies. *Ophthalmology.* 2018 Jun;125(6):894–903.

133. Dulla K, Aguila M, Lane A, Jovanovic K, Parfitt DA, Schulkens I, et al. Splice-Modulating Oligonucleotide QR-110 Restores CEP290 mRNA and Function in Human c.2991+1655A>G LCA10 Models. *Mol Ther Nucleic Acids*. 2018 Jul 23;12:730–40.
134. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res*. 2014 Dec 16;42(22):13534–44.
135. Jian X, Liu X. In Silico Prediction of Deleteriousness for Nonsynonymous and Splice-Altering Single Nucleotide Variants in the Human Genome. In: Reeves A, editor. *In Vitro Mutagenesis* [Internet]. New York, NY: Springer New York; 2017 [cited 2018 Aug 24]. p. 191–7. Available from: http://link.springer.com/10.1007/978-1-4939-6472-7_13
136. Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RKC, et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science*. 2015 Jan 9;347(6218):1254806.
137. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 2014 Jun 26;15(6):R84.
138. Krumm N, Sudmant PH, Ko A, O’Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Res*. 2012 Aug;22(8):1525–32.
139. ExAC Browser [Internet]. [cited 2018 Aug 21]. Available from: <http://exac.broadinstitute.org/>
140. 1000 Genomes | A Deep Catalog of Human Genetic Variation [Internet]. [cited 2018 Aug 21]. Available from: <http://www.internationalgenome.org/>
141. gnomAD browser [Internet]. [cited 2018 Aug 21]. Available from: <http://gnomad.broadinstitute.org/>
142. OMIM - Online Mendelian Inheritance in Man [Internet]. [cited 2018 Aug 22]. Available from: <https://www.omim.org/>
143. humsavar.txt [Internet]. [cited 2018 Aug 24]. Available from: <https://www.uniprot.org/docs/humsavar>
144. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and

Genomics and the Association for Molecular Pathology. *Genet Med Off J Am Coll Med Genet*. 2015 May;17(5):405–24.

145. Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet*. 2016 Jun 2;98(6):1067–76.

146. Strauss RW, Ho A, Muñoz B, Cideciyan AV, Sahel J-A, Sunness JS, et al. The Natural History of the Progression of Atrophy Secondary to Stargardt Disease (ProgStar) Studies: Design and Baseline Characteristics: ProgStar Report No. 1. *Ophthalmology*. 2016 Apr;123(4):817–28.

147. Michaelides M, Hunt DM, Moore AT. The genetics of inherited macular dystrophies. *J Med Genet*. 2003 Sep;40(9):641–50.

148. Zernant J, Lee W, Collison FT, Fishman GA, Sergeev YV, Schuerch K, et al. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet*. 2017 Jun;54(6):404–12.

149. Farrar GJ, Findlay JB, Kumar-Singh R, Kenna P, Humphries MM, Sharpe E, et al. Autosomal dominant retinitis pigmentosa: a novel mutation in the rhodopsin gene in the original 3q linked family. *Hum Mol Genet*. 1992 Dec;1(9):769–71.

150. al-Jandal N, Farrar GJ, Kiang AS, Humphries MM, Bannion N, Findlay JB, et al. A novel mutation within the rhodopsin gene (Thr-94-Ile) causing autosomal dominant congenital stationary night blindness. *Hum Mutat*. 1999;13(1):75–81.

151. Jane Farrar G, Kenna P, Redmond R, Shiels D, McWilliam P, Humphries MM, et al. Autosomal dominant retinitis pigmentosa: A mutation in codon 178 of the rhodopsin gene in two families of celtic origin. *Genomics*. 1991 Dec 1;11(4):1170–1.

152. Lenassi E, Vincent A, Li Z, Saihan Z, Coffey AJ, Steele-Stallard HB, et al. A detailed clinical and molecular survey of subjects with nonsyndromic USH2A retinopathy reveals an allelic hierarchy of disease-causing variants. *Eur J Hum Genet EJHG*. 2015 Oct;23(10):1318–27.

153. Pierce EA, Quinn T, Meehan T, McGee TL, Berson EL, Dryja TP. Mutations in a gene encoding a new oxygen-regulated photoreceptor protein cause dominant retinitis pigmentosa. *Nat Genet*. 1999 Jul;22(3):248–54.

154. Khaliq S, Abid A, Ismail M, Hameed A, Mohyuddin A, Lall P, et al. Novel association of RP1 gene mutations with autosomal recessive retinitis pigmentosa. *J Med Genet*. 2005 May;42(5):436–8.
155. Zhang X, Chen LJ, Law JP, Lai TYY, Chiang SWY, Tam POS, et al. Differential pattern of RP1 mutations in retinitis pigmentosa. *Mol Vis*. 2010 Jul 15;16:1353–60.
156. Castro-Sánchez S, Álvarez-Satta M, Cortón M, Guillén E, Ayuso C, Valverde D. Exploring genotype-phenotype relationships in Bardet-Biedl syndrome families. *J Med Genet*. 2015 Aug;52(8):503–13.
157. Cox KF, Kerr NC, Kedrov M, Nishimura D, Jennings BJ, Stone EM, et al. Phenotypic expression of Bardet-Biedl syndrome in patients homozygous for the common M390R mutation in the BBS1 gene. *Vision Res*. 2012 Dec 15;75:77–87.
158. Moloney TP, O'Hagan S, Lee L. Ultrawide-field fundus photography of the first reported case of gyrate atrophy from Australia. *Clin Ophthalmol Auckl NZ*. 2014;8:1561–3.
159. Ramesh V, McClatchey AI, Ramesh N, Benoit LA, Berson EL, Shih VE, et al. Molecular basis of ornithine aminotransferase deficiency in B-6-responsive and -nonresponsive forms of gyrate atrophy. *Proc Natl Acad Sci U S A*. 1988 Jun;85(11):3777–80.
160. Runhart EH, Sangermano R, Cornelis SS, Verheij JBGM, Plomp AS, Boon CJF, et al. The Common ABCA4 Variant p.Asn1868Ile Shows Nonpenetrance and Variable Expression of Stargardt Disease When Present in trans With Severe Variants. *Invest Ophthalmol Vis Sci*. 2018 Jul 2;59(8):3220–31.
161. Collison FT, Lee W, Fishman GA, Park JC, Zernant J, McAnany JJ, et al. CLINICAL CHARACTERIZATION OF STARGARDT DISEASE PATIENTS WITH THE p.N1868I ABCA4 MUTATION. *Retina Phila Pa*. 2018 Sep 7;
162. Zernant J, Lee W, Nagasaki T, Collison FT, Fishman GA, Bertelsen M, et al. Extremely hypomorphic and severe deep intronic variants in the ABCA4 locus result in varying Stargardt disease phenotypes. *Cold Spring Harb Mol Case Stud*. 2018 Aug;4(4).
163. Maugeri A, Barchitta M, Mazzone MG, Giuliano F, Agodi A. Complement System and Age-Related Macular Degeneration: Implications of Gene-

Environment Interaction for Preventive and Personalized Medicine. *BioMed Res Int*. 2018;2018:7532507.

164. Megaw RD, Soares DC, Wright AF. RPGR: Its role in photoreceptor physiology, human disease, and future therapies. *Exp Eye Res*. 2015 Sep;138:32–41.

165. Li J, Tang J, Feng Y, Xu M, Chen R, Zou X, et al. Improved Diagnosis of Inherited Retinal Dystrophies by High-Fidelity PCR of ORF15 followed by Next-Generation Sequencing. *J Mol Diagn JMD*. 2016;18(6):817–24.

166. Huang X-F, Wu J, Lv J-N, Zhang X, Jin Z-B. Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: a complementary approach to clinical genetic diagnostic testing. *Genet Med Off J Am Coll Med Genet*. 2015 Apr;17(4):307–11.

167. Kabir F, Ullah I, Ali S, Gottsch ADH, Naeem MA, Assir MZ, et al. Loss of function mutations in RP1 are responsible for retinitis pigmentosa in consanguineous familial cases. *Mol Vis*. 2016;22:610–25.

168. Méndez-Vidal C, Bravo-Gil N, González-Del Pozo M, Vela-Boza A, Dopazo J, Borrego S, et al. Novel RP1 mutations and a recurrent BBS1 variant explain the co-existence of two distinct retinal phenotypes in the same pedigree. *BMC Genet*. 2014 Dec 14;15:143.

169. Al-Rashed M, Abu Safieh L, Alkuraya H, Aldahmesh MA, Alzahrani J, Diya M, et al. RP1 and retinitis pigmentosa: report of novel mutations and insight into mutational mechanism. *Br J Ophthalmol*. 2012 Jul;96(7):1018–22.

170. El Shamieh S, Boulanger-Scemama E, Lancelot M-E, Antonio A, Démontant V, Condroyer C, et al. Targeted next generation sequencing identifies novel mutations in RP1 as a relatively common cause of autosomal recessive rod-cone dystrophy. *BioMed Res Int*. 2015;2015:485624.

171. González-Del Pozo M, Martín-Sánchez M, Bravo-Gil N, Méndez-Vidal C, Chimenea Á, Rodríguez-de la Rúa E, et al. Searching the second hit in patients with inherited retinal dystrophies and monoallelic variants in ABCA4, USH2A and CEP290 by whole-gene targeted sequencing. *Sci Rep*. 2018 Sep 6;8(1):13312.

172. rp1[gene] - ClinVar - NCBI [Internet]. [cited 2018 Sep 27]. Available from: <https://www.ncbi.nlm.nih.gov/clinvar>

173. Bax NM, Sangermano R, Roosing S, Thiadens AAHJ, Hoefsloot LH, van den Born LI, et al. Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. *Hum Mutat.* 2015 Jan;36(1):43–7.
174. Sangermano R, Bax NM, Bauwens M, van den Born LI, De Baere E, Garanto A, et al. Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T→C Mutation in Stargardt Disease. *Ophthalmology.* 2016;123(6):1375–85.
175. Cornelis SS, Bax NM, Zernant J, Allikmets R, Fritsche LG, den Dunnen JT, et al. In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. *Hum Mutat.* 2017;38(4):400–8.
176. Savoie BT, Ferrone PJ. COMPLICATED CONGENITAL RETINOSCHISIS. *Retin Cases Brief Rep.* 2017 Winter;11 Suppl 1:S202–10.
177. Dimopoulos IS, Freund PR, Knowles JA, MacDonald IM. THE NATURAL HISTORY OF FULL-FIELD STIMULUS THRESHOLD DECLINE IN CHOROIDEREMIA. *Retina Phila Pa.* 2018;38(9):1731–42.
178. Littink KW, Pott J-WR, Collin RWJ, Kroes HY, Verheij JBG, Blokland EAW, et al. A novel nonsense mutation in CEP290 induces exon skipping and leads to a relatively mild retinal phenotype. *Invest Ophthalmol Vis Sci.* 2010 Jul;51(7):3646–52.
179. Valkenburg D, van Cauwenbergh C, Lorenz B, van Genderen MM, Bertelsen M, Pott J-WR, et al. Clinical Characterization of 66 Patients With Congenital Retinal Disease Due to the Deep-Intronic c.2991+1655A>G Mutation in CEP290. *Invest Ophthalmol Vis Sci.* 2018 Sep 4;59(11):4384–91.
180. Slijkerman RW, Vaché C, Dona M, García-García G, Claustres M, Hetterschijt L, et al. Antisense Oligonucleotide-based Splice Correction for USH2A-associated Retinal Degeneration Caused by a Frequent Deep-intronic Mutation. *Mol Ther Nucleic Acids.* 2016 Nov 1;5(10):e381.
181. Albert S, Garanto A, Sangermano R, Khan M, Bax NM, Hoyng CB, et al. Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. *Am J Hum Genet.* 2018 Apr 5;102(4):517–27.
182. Garanto A, van der Velde-Visser SD, Cremers FPM, Collin RWJ. Antisense Oligonucleotide-Based Splice Correction of a Deep-Intronic

Mutation in CHM Underlying Choroideremia. *Adv Exp Med Biol.* 2018;1074:83–9.

183. Tantri A, Vrabec TR, Cu-Unjieng A, Frost A, Annesley WH, Donoso LA. X-linked retinoschisis: a clinical and molecular genetic review. *Surv Ophthalmol.* 2004 Apr;49(2):214–30.

184. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet.* 2014 Jan;133(1):1–9.

185. Bush M, Setiaputra D, Yip CK, Molday RS. Cog-Wheel Octameric Structure of RS1, the Discoidin Domain Containing Retinal Protein Associated with X-Linked Retinoschisis. *PLOS ONE.* 2016 Jan 26;11(1):e0147653.

186. Safety and Efficacy of rAAV-hRS1 in Patients With X-linked Retinoschisis (XLRS) - Full Text View - ClinicalTrials.gov [Internet]. [cited 2018 Aug 25]. Available from: <https://clinicaltrials.gov/ct2/show/NCT02416622>

187. Study of RS1 Ocular Gene Transfer for X-linked Retinoschisis - Full Text View - ClinicalTrials.gov [Internet]. [cited 2018 Aug 25]. Available from: <https://clinicaltrials.gov/ct2/show/NCT02317887>

188. Molday RS, Kellner U, Weber BHF. X-linked juvenile retinoschisis: clinical diagnosis, genetic analysis, and molecular mechanisms. *Prog Retin Eye Res.* 2012 May;31(3):195–212.

189. Yang HS, Lee JB, Yoon YH, Lee JY. Correlation Between Spectral-Domain OCT Findings and Visual Acuity in X-Linked Retinoschisis. *Invest Ophthalmol Vis Sci.* 2014 May 1;55(5):3029–36.

190. Nittala MG, Laxmi G, Raman R, Rani PK, Bhargava A, Pal SS, et al. Spectral-domain OCT and microperimeter characterization of morphological and functional changes in X-linked retinoschisis. *Ophthalmic Surg Lasers Imaging Off J Int Soc Imaging Eye.* 2009 Feb;40(1):71–4.

191. Bergen AA, van Schooneveld MJ, Orth U, Bleeker-Wagemakers EM, Gal A. Multipoint linkage analysis in X-linked juvenile retinoschisis. *Clin Genet.* 1993 Mar;43(3):113–6.

192. Sauer CG, Gehrig A, Warneke-Wittstock R, Marquardt A, Ewing CC, Gibson A, et al. Positional cloning of the gene associated with X-linked juvenile retinoschisis. *Nat Genet.* 1997 Oct;17(2):164–70.

193. Plössl K, Weber BHF, Friedrich U. The X-linked juvenile retinoschisis protein retinoschisin is a novel regulator of mitogen-activated protein kinase signalling and apoptosis in the retina. *J Cell Mol Med*. 2017 Apr;21(4):768–80.
194. Plössl K, Royer M, Bernklau S, Tavraz NN, Friedrich T, Wild J, et al. Retinoschisin is linked to retinal Na/K-ATPase signaling and localization. *Mol Biol Cell*. 2017 Aug 1;28(16):2178–89.
195. Tolun G, Vijayasarathy C, Huang R, Zeng Y, Li Y, Steven AC, et al. Paired octamer rings of retinoschisin suggest a junctional model for cell-cell adhesion in the retina. *Proc Natl Acad Sci U S A*. 2016 May 10;113(19):5287–92.
196. Ramsay EP, Collins RF, Owens TW, Siebert CA, Jones RPO, Wang T, et al. Structural analysis of X-linked retinoschisis mutations reveals distinct classes which differentially effect retinoschisin function. *Hum Mol Genet*. 2016 15;25(24):5311–20.
197. Lai Y-H, Huang S-P, Chen S-P, Hu P-S, Lin S-F, Sheu M-M, et al. A novel gene mutation in a family with X-linked retinoschisis. *J Formos Med Assoc Taiwan Yi Zhi*. 2015 Sep;114(9):872–80.
198. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 2015 Jan;12(1):7–8.
199. RCSB PDB: Homepage [Internet]. [cited 2018 Aug 25]. Available from: <http://www.rcsb.org/>
200. Quan L, Lv Q, Zhang Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics*. 2016 Oct 1;32(19):2936–46.
201. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011 Oct 11;7:539.
202. Tsui I. Fundus Autofluorescence in X-Linked Retinoschisis. In: Noemi L, editor. *Fundus Autofluorescence*. Lippincott Williams & Wilkins; 2009.
203. Jarvik GP, Browning BL. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *Am J Hum Genet*. 2016 Jun 2;98(6):1077–81.
204. Yanoff M, Kertesz Rahn E, Zimmerman LE. Histopathology of juvenile retinoschisis. *Arch Ophthalmol Chic Ill* 1960. 1968 Jan;79(1):49–53.

205. Gregori NZ, Berrocal AM, Gregori G, Murray TG, Knighton RW, Flynn HW, et al. Macular spectral-domain optical coherence tomography in patients with X linked retinoschisis. *Br J Ophthalmol*. 2009 Mar;93(3):373–8.
206. Bowles K, Cukras C, Turriff A, Sergeev Y, Vitale S, Bush RA, et al. X-Linked Retinoschisis: RS1 Mutation Severity and Age Affect the ERG Phenotype in a Cohort of 68 Affected Male Subjects. *Invest Ophthalmol Vis Sci*. 2011 Nov;52(12):9250–6.
207. Bennett LD, Wang Y-Z, Klein M, Pennesi ME, Jayasundera T, Birch DG. Structure/Psychophysical Relationships in X-Linked Retinoschisis. *Invest Ophthalmol Vis Sci*. 2016 Feb;57(2):332–7.
208. Andreoli MT, Lim JI. Optical coherence tomography retinal thickness and volume measurements in X-linked retinoschisis. *Am J Ophthalmol*. 2014 Sep;158(3):567-573.e2.
209. Eksandh LC, Ponjavic V, Ayyagari R, Bingham EL, Hiriyantha KT, Andréasson S, et al. Phenotypic Expression of Juvenile X-linked Retinoschisis in Swedish Families With Different Mutations in the XLR1 Gene. *Arch Ophthalmol*. 2000 Aug 1;118(8):1098–104.
210. Deutman AF, Pinckers AJ, Aand de Kerk AL. Dominantly inherited cystoid macular edema. *Am J Ophthalmol*. 1976 Oct;82(4):540–8.
211. Wang T, Waters CT, Rothman AMK, Jakins TJ, Römisch K, Trump D. Intracellular retention of mutant retinoschisin is the pathological mechanism underlying X-linked retinoschisis. *Hum Mol Genet*. 2002 Nov 15;11(24):3097–105.
212. Sergeev YV, Caruso RC, Meltzer MR, Smaoui N, MacDonald IM, Sieving PA. Molecular modeling of retinoschisin with functional analysis of pathogenic mutations from human X-linked retinoschisis. *Hum Mol Genet*. 2010 Apr 1;19(7):1302–13.
213. Akeo K, Kameya S, Gocho K, Kubota D, Yamaki K, Takahashi H. Detailed Morphological Changes of Foveoschisis in Patient with X-Linked Retinoschisis Detected by SD-OCT and Adaptive Optics Fundus Camera [Internet]. *Case Reports in Ophthalmological Medicine*. 2015 [cited 2018 Aug 25]. Available from: <https://www.hindawi.com/journals/crionpm/2015/432782/>
214. Duncan JL, Ratnam K, Birch DG, Sundquist SM, Lucero AS, Zhang Y, et al. Abnormal Cone Structure in Foveal Schisis Cavities in X-Linked

Retinoschisis from Mutations in Exon 6 of the RS1 Gene. *Invest Ophthalmol Vis Sci*. 2011 Dec;52(13):9614–23.

215. Locus Specific Database list [Internet]. [cited 2018 Aug 26]. Available from: http://grenada.lumc.nl/LSDB_list/lstdbs/RS1

216. Johnson BA, Cole BS, Geisert EE, Ikeda S, Ikeda A. Tyrosinase is the modifier of retinoschisis in mice. *Genetics*. 2010 Dec;186(4):1337–44.

217. Farrar GJ, McWilliam P, Bradley DG, Kenna P, Lawler M, Sharp EM, et al. Autosomal dominant retinitis pigmentosa: linkage to rhodopsin and evidence for genetic heterogeneity. *Genomics*. 1990 Sep;8(1):35–40.

218. Dryja TP, McGee TL, Reichel E, Hahn LB, Cowley GS, Yandell DW, et al. A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature*. 1990 Jan 25;343(6256):364–6.

219. Dryja TP, McGee TL, Hahn LB, Cowley GS, Olsson JE, Reichel E, et al. Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. *N Engl J Med*. 1990 Nov 8;323(19):1302–7.

220. Farrar GJ, Kenna P, Redmond R, McWilliam P, Bradley DG, Humphries MM, et al. Autosomal dominant retinitis pigmentosa: absence of the rhodopsin proline----histidine substitution (codon 23) in pedigrees from Europe. *Am J Hum Genet*. 1990 Dec;47(6):941–5.

221. Inglehearn CF, Lester DH, Bashir R, Atif U, Keen TJ, Sertedaki A, et al. Recombination between rhodopsin and locus D3S47 (C17) in rhodopsin retinitis pigmentosa families. *Am J Hum Genet*. 1992 Mar;50(3):590–7.

222. Horn M, Humphries P, Kunisch M, Marchese C, Apfelstedt-Sylla E, Fugl L, et al. Deletions in exon 5 of the human rhodopsin gene causing a shift in the reading frame and autosomal dominant retinitis pigmentosa. *Hum Genet*. 1992 Nov;90(3):255–7.

223. Stenkamp RE, Teller DC, Palczewski K. Rhodopsin: a structural primer for G-protein coupled receptors. *Arch Pharm (Weinheim)*. 2005 Jun;338(5–6):209–16.

224. Li J, Edwards PC, Burghammer M, Villa C, Schertler GFX. Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol*. 2004 Nov 5;343(5):1409–38.

225. Chadderton N, Millington-Ward S, Palfi A, O'Reilly M, Tuohy G, Humphries MM, et al. Improved Retinal Function in a Mouse Model of

- Dominant Retinitis Pigmentosa Following AAV-delivered Gene Therapy. *Mol Ther J Am Soc Gene Ther*. 2009 Apr;17(4):593–9.
226. O'Reilly M, Palfi A, Chadderton N, Millington-Ward S, Ader M, Cronin T, et al. RNA interference-mediated suppression and replacement of human rhodopsin in vivo. *Am J Hum Genet*. 2007 Jul;81(1):127–35.
227. Cideciyan AV, Sudharsan R, Dufour VL, Massengill MT, Iwabe S, Swider M, et al. Mutation-independent rhodopsin gene therapy by knockdown and replacement with a single AAV vector. *Proc Natl Acad Sci*. 2018 Sep 4;115(36):E8547–56.
228. Keel SB, Doty RT, Yang Z, Quigley JG, Chen J, Knoblaugh S, et al. A heme export protein is required for red blood cell differentiation and iron homeostasis. *Science*. 2008 Feb 8;319(5864):825–8.
229. Tailor CS, Willett BJ, Kabat D. A putative cell surface receptor for anemia-inducing feline leukemia virus subgroup C is a member of a transporter superfamily. *J Virol*. 1999 Aug;73(8):6500–5.
230. Quigley JG, Yang Z, Worthington MT, Phillips JD, Sabo KM, Sabath DE, et al. Identification of a human heme exporter that is essential for erythropoiesis. *Cell*. 2004 Sep 17;118(6):757–66.
231. Chiabrando D, Castori M, di Rocco M, Ungelenk M, Gießelmann S, Di Capua M, et al. Mutations in the Heme Exporter FLVCR1 Cause Sensory Neurodegeneration with Loss of Pain Perception. *PLoS Genet* [Internet]. 2016 Dec 6 [cited 2018 Jun 25];12(12). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5140052/>
232. Castori M, Morlino S, Ungelenk M, Pareyson D, Salsano E, Grammatico P, et al. Posterior column ataxia with retinitis pigmentosa coexisting with sensory-autonomic neuropathy and leukemia due to the homozygous p.Pro221Ser FLVCR1 mutation. *Am J Med Genet B Neuropsychiatr Genet*. 174(7):732–9.
233. Ishiura H, Fukuda Y, Mitsui J, Nakahara Y, Ahsan B, Takahashi Y, et al. Posterior column ataxia with retinitis pigmentosa in a Japanese family with a novel mutation in FLVCR1. *neurogenetics*. 2011 May 1;12(2):117–21.
234. Glöckle N, Kohl S, Mohr J, Scheurenbrand T, Sprecher A, Weisschuh N, et al. Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *Eur J Hum Genet*. 2014 Jan;22(1):99–104.

235. Yusuf IH, Shanks ME, Clouston P, MacLaren RE. A splice-site variant in *FLVCR1* produces retinitis pigmentosa without posterior column ataxia. *Ophthalmic Genet.* 2018 Mar 4;39(2):263–7.
236. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D865–76.
237. Choudhury SR, Fitzpatrick Z, Harris AF, Maitland SA, Ferreira JS, Zhang Y, et al. In Vivo Selection Yields AAV-B1 Capsid for Central Nervous System and Muscle Gene Therapy. *Mol Ther J Am Soc Gene Ther.* 2016;24(7):1247–57.
238. Herrmann A-K, Grosse S, Börner K, Krämer C, Wiedtke E, Gunkel M, et al. Impact of the assembly-activating protein (AAP) on molecular evolution of synthetic Adeno-associated virus (AAV) capsids. *Hum Gene Ther.* 2018 Jul 6;
239. Tordo J, O’Leary C, Antunes ASLM, Palomar N, Aldrin-Kirk P, Basche M, et al. A novel adeno-associated virus capsid with enhanced neurotropism corrects a lysosomal transmembrane enzyme deficiency. *Brain J Neurol.* 2018 Jul 1;141(7):2014–31.
240. Smith JK, Agbandje-McKenna M. Creating an arsenal of Adeno-associated virus (AAV) gene delivery stealth vehicles. *PLoS Pathog.* 2018;14(5):e1006929.
241. Paulk NK, Pekrun K, Zhu E, Nygaard S, Li B, Xu J, et al. Bioengineered AAV Capsids with Combined High Human Liver Transduction In Vivo and Unique Humoral Seroreactivity. *Mol Ther J Am Soc Gene Ther.* 2018 Jan 3;26(1):289–303.
242. Bull’s-eye maculopathy due to hydroxychloroquine toxicity [Internet]. [cited 2018 Sep 19]. Available from: <http://webeye.ophth.uiowa.edu/eyeforum/atlas/pages/hydroxychloroquine-toxicity/index.htm>
243. Scott IU, Flynn HW, Smiddy WE. Bull’s-eye Maculopathy Associated With Chronic Macular Hole. *Arch Ophthalmol.* 1998 Aug 1;116(8):1116–7.
244. Bull’s eye maculopathy [Internet]. Macular Society. 2017 [cited 2018 Sep 19]. Available from: <https://www.macularsociety.org/bull%E2%80%99s-eye-maculopathy>

245. Nõupuu K, Lee W, Zernant J, Greenstein VC, Tsang S, Allikmets R. Recessive Stargardt disease phenocopying hydroxychloroquine retinopathy. *Graefes Arch Clin Exp Ophthalmol Albrecht Von Graefes Arch Klin Exp Ophthalmol*. 2016 May;254(5):865–72.
246. Hydroxychloroquine (Plaquenil) Toxicity and Recommendations for Screening [Internet]. [cited 2018 Sep 20]. Available from: <https://webeye.ophth.uiowa.edu/eyeforum/cases/139-plaquenil-toxicity.htm>
247. Melles RB, Marmor MF. The risk of toxic retinopathy in patients on long-term hydroxychloroquine therapy. *JAMA Ophthalmol*. 2014 Dec;132(12):1453–60.
248. Mahon GJ, Anderson HR, Gardiner TA, McFarlane S, Archer DB, Stitt AW. Chloroquine causes lysosomal dysfunction in neural retina and RPE: implications for retinopathy. *Curr Eye Res*. 2004 Apr;28(4):277–84.
249. Nanda A, Salvetti AP, Martinez-Fernandez de la Camara C, MacLaren RE. Misdiagnosis of X-linked retinitis pigmentosa in a choroideremia patient with heavily pigmented fundi. *Ophthalmic Genet*. 2018 Jun;39(3):380–3.
250. Lee S-Y, Yu W-K, Lin P-K. Large Gene Deletion and Changes in Corneal Endothelial Cells in a Family With Choroideremia. *Invest Ophthalmol Vis Sci*. 2015 Mar 1;56(3):1887–93.
251. Mura M, Sereda C, Jablonski MM, MacDonald IM, Iannaccone A. Clinical and Functional Findings in Choroideremia Due to Complete Deletion of the CHM Gene. *Arch Ophthalmol*. 2007 Aug 1;125(8):1107–13.
252. Vervoort R, Lennon A, Bird AC, Tulloch B, Axton R, Miano MG, et al. Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nat Genet*. 2000 Aug;25(4):462–6.
253. Fischer MD, McClements ME, Martinez-Fernandez de la Camara C, Bellingrath J-S, Dauletbekov D, Ramsden SC, et al. Codon-Optimized RPGR Improves Stability and Efficacy of AAV8 Gene Therapy in Two Mouse Models of X-Linked Retinitis Pigmentosa. *Mol Ther*. 2017 Aug 2;25(8):1854–65.
254. Shu X, Fry AM, Tulloch B, Manson FDC, Crabb JW, Khanna H, et al. RPGR ORF15 isoform co-localizes with RPGRIP1 at centrioles and basal bodies and interacts with nucleophosmin. *Hum Mol Genet*. 2005 May 1;14(9):1183–97.
255. Mavlyutov TA, Zhao H, Ferreira PA. Species-specific subcellular localization of RPGR and RPGRIP isoforms: implications for the phenotypic

variability of congenital retinopathies among species. *Hum Mol Genet.* 2002 Aug 1;11(16):1899–907.

256. Rao KN, Anand M, Khanna H. The carboxyl terminal mutational hotspot of the ciliary disease protein RPGRORF15 (retinitis pigmentosa GTPase regulator) is glutamylated in vivo. *Biol Open.* 2016;5(4):424–8.

257. Sun X, Park JH, Gumerson J, Wu Z, Swaroop A, Qian H, et al. Loss of RPGR glutamylation underlies the pathogenic mechanism of retinal dystrophy caused by TTLL5 mutations. *Proc Natl Acad Sci U S A.* 2016;113(21):E2925–34.

258. Beltran WA, Cideciyan AV, Lewin AS, Iwabe S, Khanna H, Sumaroka A, et al. Gene therapy rescues photoreceptor blindness in dogs and paves the way for treating human X-linked retinitis pigmentosa. *Proc Natl Acad Sci U S A.* 2012;109(6):2132–7.

259. Beltran WA, Cideciyan AV, Boye SE, Ye G-J, Iwabe S, Dufour VL, et al. Optimization of Retinal Gene Therapy for X-Linked Retinitis Pigmentosa Due to RPGR Mutations. *Mol Ther.* 2017 Aug 2;25(8):1866–80.

260. Hong D-H, Pawlyk BS, Adamian M, Sandberg MA, Li T. A single, abbreviated RPGR-ORF15 variant reconstitutes RPGR function in vivo. *Invest Ophthalmol Vis Sci.* 2005 Feb;46(2):435–41.

261. Search of: rpgr - List Results - ClinicalTrials.gov [Internet]. [cited 2018 Sep 27]. Available from: <https://clinicaltrials.gov/ct2/results?cond=&term=rpgr&cntry=&state=&city=&dist=>

262. Marlhens F, Bareil C, Griffoin J-M, Zrenner E, Amalric P, Eliaou C, et al. Mutations in RPE65 cause leber's congenital amaurosis. *Nat Genet.* 1997;17(2):139–41.

263. Marlhens F, Griffoin J-M, Bareil C, Arnaud B, Claustres M, Hamel CP. Autosomal recessive retinal dystrophy associated with two novel mutations in the RPE65 gene. *Eur J Hum Genet.* 1998;6(5):527–31.

264. Hamel CP, Griffoin J-M, Lasquelléc L, Bazalgette C, Arnaud B. Retinal dystrophies caused by mutations in RPE65: Assessment of visual functions. *Br J Ophthalmol.* 2001;85(4):424–7.

265. CDMNY. Luxturna [Internet]. [cited 2018 Sep 23]. Available from: <http://luxturna.com/image>

266. Bowne SJ, Humphries MM, Sullivan LS, Kenna PF, Tam LCS, Kiang AS, et al. A dominant mutation in *RPE65* identified by whole-exome sequencing causes retinitis pigmentosa with choroidal involvement. *Eur J Hum Genet*. 2011 Oct;19(10):1074–81.
267. Nikolaeva O, Takahashi Y, Moiseyev G, Ma J. Purified RPE65 shows isomerohydrolase activity after reassociation with a phospholipid membrane. *FEBS J*. 2009 Jun 1;276(11):3020–30.
268. Hull S, Mukherjee R, Holder GE, Moore AT, Webster AR. The clinical features of retinal disease due to a dominant mutation in RPE65. *Mol Vis*. 2016;22:626–35.
269. Jauregui R, Park KS, Tsang SH. Two-year progression analysis of RPE65 autosomal dominant retinitis pigmentosa. *Ophthalmic Genet*. 2018 Aug;39(4):544–9.
270. Choi EH, Suh S, Sander CL, Hernandez CJO, Bulman ER, Khadka N, et al. Insights into the pathogenesis of dominant retinitis pigmentosa associated with a D477G mutation in RPE65. *Hum Mol Genet*. 2018 Jul 1;27(13):2225–43.
271. Li S, Izumi T, Hu J, Jin HH, Siddiqui A-AA, Jacobson SG, et al. Rescue of enzymatic function for disease-associated RPE65 proteins containing various missense mutations in non-active sites. *J Biol Chem*. 2014 Jul 4;289(27):18943–56.
272. Ameri H. Prospect of retinal gene therapy following commercialization of voretigene neparvovec-rzyl for retinal dystrophy mediated by RPE65 mutation. *J Curr Ophthalmol*. 2018 Feb 16;30(1):1–2.
273. Last10K.com. QLT INCBC 10-K Annual Report Thu Feb 26 2015 [Internet]. [cited 2018 Sep 23]. Available from: <https://www.last10k.com/sec-filings/qlti>
274. Oral QLT091001 in Retinitis Pigmentosa (RP) Subjects With an Autosomal Dominant Mutation in Retinal Pigment Epithelial 65 Protein (RPE65) - Full Text View - ClinicalTrials.gov [Internet]. [cited 2018 Sep 23]. Available from: <https://clinicaltrials.gov/ct2/show/NCT01543906>
275. Safety and Efficacy Study in Patients With Retinitis Pigmentosa Due to Mutations in PDE6B Gene - Full Text View - ClinicalTrials.gov [Internet]. [cited 2018 Sep 2]. Available from: <https://clinicaltrials.gov/ct2/show/NCT03328130>

276. Beranova-Giorgianni S, Giorgianni F. Proteomics of Human Retinal Pigment Epithelium (RPE) Cells. *Proteomes* [Internet]. 2018 May 15 [cited 2018 Sep 14];6(2). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6027133/>
277. Hanlon KS, Chadderton N, Palfi A, Blanco Fernandez A, Humphries P, Kenna PF, et al. A Novel Retinal Ganglion Cell Promoter for Utility in AAV Vectors. *Front Neurosci*. 2017;11:521.
278. Zinn E, Pacouret S, Khaychuk V, Turunen HT, Carvalho LS, Andres-Mateos E, et al. In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. *Cell Rep*. 2015 Aug 11;12(6):1056–68.
279. Klimczak RR, Koerber JT, Dalkara D, Flannery JG, Schaffer DV. A Novel Adeno-Associated Viral Variant for Efficient and Selective Intravitreal Transduction of Rat Müller Cells. *PLOS ONE*. 2009 Oct 14;4(10):e7467.
280. van Lieshout LP, Domm JM, Rindler TN, Frost KL, Sorensen DL, Medina SJ, et al. A Novel Triple-Mutant AAV6 Capsid Induces Rapid and Potent Transgene Expression in the Muscle and Respiratory Tract of Mice. *Mol Ther Methods Clin Dev*. 2018 Jun 15;9:323–9.
281. Duncan GA, Kim N, Colon-Cortes Y, Rodriguez J, Mazur M, Birket SE, et al. An Adeno-Associated Viral Vector Capable of Penetrating the Mucus Barrier to Inhaled Gene Therapy. *Mol Ther Methods Clin Dev*. 2018 Jun 15;9:296–304.
282. Wu Z, Asokan A, Samulski RJ. Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol Ther J Am Soc Gene Ther*. 2006 Sep;14(3):316–27.
283. Song L, Llanga T, Conatser LM, Zaric V, Gilger BC, Hirsch ML. Serotype survey of AAV gene delivery via subconjunctival injection in mice. *Gene Ther*. 2018 Aug 2;
284. Cukras C, Wiley HE, Jeffrey BG, Sen HN, Turriff A, Zeng Y, et al. Retinal AAV8-RS1 Gene Therapy for X-Linked Retinoschisis: Initial Findings from a Phase I/IIa Trial by Intravitreal Delivery. *Mol Ther J Am Soc Gene Ther*. 2018 Sep 5;26(9):2282–94.
285. Buch PK, Bainbridge JW, Ali RR. AAV-mediated gene therapy for retinal disorders: from mouse to man. *Gene Ther*. 2008 Jun;15(11):849–57.
286. Phase I Dose Escalation Safety Study of RetinoStat in Advanced Age-Related Macular Degeneration (AMD) - Full Text View - ClinicalTrials.gov

- [Internet]. [cited 2018 Sep 28]. Available from: <https://clinicaltrials.gov/ct2/show/NCT01301443>
287. A Follow-up Study to Evaluate the Safety of RetinoStat® in Patients With Age-Related Macular Degeneration - Full Text View - ClinicalTrials.gov [Internet]. [cited 2018 Sep 28]. Available from: <https://clinicaltrials.gov/ct2/show/NCT01678872>
288. Verrier JD, Madorsky I, Coggin WE, Geesey M, Hochman M, Walling E, et al. Bicistronic Lentiviruses Containing a Viral 2A Cleavage Sequence Reliably Co-Express Two Proteins and Restore Vision to an Animal Model of LCA1. *PLOS ONE*. 2011 May 27;6(5):e20553.
289. Vandenberghe LH, Auricchio A. Novel adeno-associated viral vectors for retinal gene therapy. *Gene Ther*. 2012 Feb;19(2):162–8.
290. Wooley DP, Sharma P, Weinstein JR, Kotha Lakshmi Narayan P, Schaffer DV, Excoffon KJDA. A directed evolution approach to select for novel Adeno-associated virus capsids on an HIV-1 producer T cell line. *J Virol Methods*. 2017;250:47–54.
291. Ojala DS, Sun S, Santiago-Ortiz JL, Shapiro MG, Romero PA, Schaffer DV. In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ. *Mol Ther J Am Soc Gene Ther*. 2018 Jan 3;26(1):304–19.
292. Jane Farrar G, Millington-Ward S, Chadderton N, Mansergh FC, Palfi A. Gene therapies for inherited retinal disorders. *Vis Neurosci*. 2014 Sep;31(4–5):289–307.
293. Bennett J, Wellman J, Marshall KA, McCague S, Ashtari M, DiStefano-Pappas J, et al. Safety and durability of effect of contralateral-eye administration of AAV2 gene therapy in patients with childhood-onset blindness caused by RPE65 mutations: a follow-on phase 1 trial. *Lancet Lond Engl*. 2016 Aug 13;388(10045):661–72.
294. Safety and Tolerability Study of AAV2-sFLT01 in Patients With Neovascular Age-Related Macular Degeneration (AMD) - Full Text View - ClinicalTrials.gov [Internet]. [cited 2018 Sep 28]. Available from: <https://clinicaltrials.gov/ct2/show/NCT01024998>
295. Safety and Dose Escalation Study of AAV2-hCHM in Subjects With CHM (Choroideremia) Gene Mutations - Full Text View - ClinicalTrials.gov

- [Internet]. [cited 2018 Sep 28]. Available from: <https://clinicaltrials.gov/ct2/show/NCT02341807>
296. Safety Study of RPE65 Gene Therapy to Treat Leber Congenital Amaurosis - Full Text View - ClinicalTrials.gov [Internet]. [cited 2018 Sep 28]. Available from: <https://clinicaltrials.gov/ct2/show/NCT00643747>
297. Allocca M, Doria M, Petrillo M, Colella P, Garcia-Hoyos M, Gibbs D, et al. Serotype-dependent packaging of large genes in adeno-associated viral vectors results in effective gene delivery in mice. *J Clin Invest*. 2008 May 1;118(5):1955–64.
298. Wu Z, Yang H, Colosi P. Effect of Genome Size on AAV Vector Packaging. *Mol Ther*. 2010 Jan;18(1):80–6.
299. Grieger JC, Samulski RJ. Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps. *J Virol*. 2005 Aug;79(15):9933–44.
300. Grose WE, Clark KR, Griffin D, Malik V, Shontz KM, Montgomery CL, et al. Homologous Recombination Mediates Functional Recovery of Dysferlin Deficiency following AAV5 Gene Transfer. *PLoS ONE* [Internet]. 2012 Jun 15 [cited 2018 Sep 2];7(6). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3376115/>
301. Lai Y, Yue Y, Duan D. Evidence for the failure of adeno-associated virus serotype 5 to package a viral genome ≥ 8.2 kb. *Mol Ther J Am Soc Gene Ther*. 2010 Jan;18(1):75–9.
302. McClements ME, Charbel Issa P, Blouin V, MacLaren RE. A fragmented adeno-associated viral dual vector strategy for treatment of diseases caused by mutations in large genes leads to expression of hybrid transcripts. *J Genet Syndr Gene Ther*. 2016 Nov 14;7(5).
303. McClements ME, MacLaren RE. Adeno-associated Virus (AAV) Dual Vector Strategies for Gene Therapy Encoding Large Transgenes. *Yale J Biol Med*. 2017 Dec 19;90(4):611–23.
304. Pryadkina M, Lostal W, Bourg N, Charton K, Roudaut C, Hirsch ML, et al. A comparison of AAV strategies distinguishes overlapping vectors for efficient systemic delivery of the 6.2 kb Dysferlin coding sequence. *Mol Ther Methods Clin Dev*. 2015;2:15009.

305. Odom GL, Gregorevic P, Allen JM, Chamberlain JS. Gene therapy of mdx mice with large truncated dystrophins generated by recombination using rAAV6. *Mol Ther J Am Soc Gene Ther*. 2011 Jan;19(1):36–45.
306. Chen Z-Y, Yant SR, He C-Y, Meuse L, Shen S, Kay MA. Linear DNAs Concatemerize in Vivo and Result in Sustained Transgene Expression in Mouse Liver. *Mol Ther*. 2001 Mar 1;3(3):403–10.
307. Schnepf BC, Jensen RL, Chen C-L, Johnson PR, Clark KR. Characterization of adeno-associated virus genomes isolated from human tissues. *J Virol*. 2005 Dec;79(23):14793–803.
308. Lai Y, Yue Y, Liu M, Ghosh A, Engelhardt JF, Chamberlain JS, et al. Efficient in vivo gene expression by trans-splicing adeno-associated viral vectors. *Nat Biotechnol*. 2005 Nov;23(11):1435–9.
309. Ghosh A, Yue Y, Lai Y, Duan D. A hybrid vector system expands adeno-associated viral vector packaging capacity in a transgene-independent manner. *Mol Ther J Am Soc Gene Ther*. 2008 Jan;16(1):124–30.
310. Grieger JC, Soltys SM, Samulski RJ. Production of Recombinant Adeno-associated Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector. *Mol Ther J Am Soc Gene Ther*. 2016 Feb;24(2):287–97.
311. Lock M, Alvira MR, Wilson JM. Analysis of particle content of recombinant adeno-associated virus serotype 8 vectors by ion-exchange chromatography. *Hum Gene Ther Methods*. 2012 Feb;23(1):56–64.
312. Kohlbrenner E, Henckaerts E, Rapti K, Gordon RE, Linden RM, Hajjar RJ, et al. Quantification of AAV particle titers by infrared fluorescence scanning of coomassie-stained sodium dodecyl sulfate-polyacrylamide gels. *Hum Gene Ther Methods*. 2012 Jun;23(3):198–203.
313. Wright JF. Manufacturing and characterizing AAV-based vectors for use in clinical studies. *Gene Ther*. 2008 Jun;15(11):840–8.
314. Qu G, Bahr-Davidson J, Prado J, Tai A, Cataniag F, McDonnell J, et al. Separation of adeno-associated virus type 2 empty particles from genome containing vectors by anion-exchange column chromatography. *J Virol Methods*. 2007 Mar;140(1–2):183–92.
315. Urabe M, Xin K-Q, Obara Y, Nakakura T, Mizukami H, Kume A, et al. Removal of empty capsids from type 1 adeno-associated virus vector stocks

by anion-exchange chromatography potentiates transgene expression. *Mol Ther J Am Soc Gene Ther*. 2006 Apr;13(4):823–8.

316. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinforma Oxf Engl*. 2011 Nov 1;27(21):2987–93.

317. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013 Mar 1;14(2):178–92.

318. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al*. 2013 Oct 15;11(1110):11.10.1–11.10.33.

319. Sommer JM, Smith PH, Parthasarathy S, Isaacs J, Vijay S, Kieran J, et al. Quantification of adeno-associated virus particles and empty capsids by optical density measurement. *Mol Ther J Am Soc Gene Ther*. 2003 Jan;7(1):122–8.

320. Allay JA, Sleep S, Long S, Tillman DM, Clark R, Carney G, et al. Good manufacturing practice production of self-complementary serotype 8 adeno-associated viral vector for a hemophilia B clinical trial. *Hum Gene Ther*. 2011 May;22(5):595–604.

321. Zeltner N, Kohlbrenner E, Clément N, Weber T, Linden RM. Near-perfect infectivity of wild-type AAV as benchmark for infectivity of recombinant AAV vectors. *Gene Ther*. 2010 Jul;17(7):872–9.

322. Manno CS, Pierce GF, Arruda VR, Glader B, Ragni M, Rasko JJ, et al. Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat Med*. 2006 Mar;12(3):342–7.

323. Nathwani AC, Reiss UM, Tuddenham EGD, Rosales C, Chowdary P, McIntosh J, et al. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N Engl J Med*. 2014 Nov 20;371(21):1994–2004.

324. Gao K, Li M, Zhong L, Su Q, Li J, Li S, et al. Empty Virions In AAV8 Vector Preparations Reduce Transduction Efficiency And May Cause Total Viral Particle Dose-Limiting Side-Effects. *Mol Ther Methods Clin Dev*. 2014;1(9):20139.

325. Tai PWL, Xie J, Fong K, Seetin M, Heiner C, Su Q, et al. Adeno-associated Virus Genome Population Sequencing Achieves Full Vector Genome

Resolution and Reveals Human-Vector Chimeras. *Mol Ther Methods Clin Dev*. 2018 Feb 13;9:130–41.

326. Ling C, Wang Y, Lu Y, Wang L, Jayandharan GR, Aslanidi GV, et al. The Adeno-Associated Virus Genome Packaging Puzzle. *J Mol Genet Med Int J Biomed Res* [Internet]. 2015 Aug [cited 2018 Sep 16];9(3). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4778740/>

327. Kapranov P, Chen L, Dederich D, Dong B, He J, Steinmann KE, et al. Native molecular state of adeno-associated viral vectors revealed by single-molecule sequencing. *Hum Gene Ther*. 2012 Jan;23(1):46–55.

328. Hirsch ML, Wolf SJ, Samulski RJ. Delivering Transgenic DNA Exceeding the Carrying Capacity of AAV Vectors. *Methods Mol Biol Clifton NJ*. 2016;1382:21–39.

329. Xie J, Mao Q, Tai PWL, He R, Ai J, Su Q, et al. Short DNA Hairpins Compromise Recombinant Adeno-Associated Virus Genome Homogeneity. *Mol Ther J Am Soc Gene Ther*. 2017 07;25(6):1363–74.

330. Labib K, Hodgson B. Replication fork barriers: pausing for a break or stalling for time? *EMBO Rep*. 2007 Apr;8(4):346–53.

331. Day TP, Byrne LC, Schaffer DV, Flannery JG. Advances in AAV Vector Development for Gene Therapy in the Retina. *Adv Exp Med Biol*. 2014;801:687–93.

332. Minei R, Hoshina R, Ogura A. De novo assembly of middle-sized genome using MinION and Illumina sequencers. *BMC Genomics*. 2018 Sep 24;19(1):700.

333. Search of: aav | Retinal - List Results - ClinicalTrials.gov [Internet]. [cited 2018 Sep 27]. Available from: <https://clinicaltrials.gov/ct2/results?cond=Retinal&term=aav&cntry=&state=&city=&dist=>

334. Jamshidi F, Place EM, Mehrotra S, Navarro-Gomez D, Maher M, Branham KE, et al. Contribution of noncoding pathogenic variants to RPGRIP1-mediated inherited retinal degeneration. *Genet Med Off J Am Coll Med Genet*. 2018 Aug 3;

335. Huang X-F, Mao J-Y, Huang Z-Q, Rao F-Q, Cheng F-F, Li F-F, et al. Genome-Wide Detection of Copy Number Variations in Unsolved Inherited Retinal Disease. *Invest Ophthalmol Vis Sci*. 2017 01;58(1):424–9.

336. Khateb S, Hanany M, Khalaileh A, Beryozkin A, Meyer S, Abu-Diab A, et al. Identification of genomic deletions causing inherited retinal degenerations by coverage analysis of whole exome sequencing data. *J Med Genet*. 2016;53(9):600–7.
337. Van Cauwenbergh C, Van Schil K, Cannoodt R, Bauwens M, Van Laethem T, De Jaegere S, et al. arrEYE: a customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs. *Genet Med*. 2017 Apr;19(4):457–66.
338. Zhang X, Wang B, Zhang L, You G, Palais RA, Zhou L, et al. Accurate diagnosis of spinal muscular atrophy and 22q11.2 deletion syndrome using limited deoxynucleotide triphosphates and high-resolution melting. *BMC Genomics*. 2018 Jun 20;19(1):485.
339. Zubakov D, Chamier-Ciemińska J, Kokmeijer I, Maciejewska A, Martínez P, Pawłowski R, et al. Introducing novel type of human DNA markers for forensic tissue identification: DNA copy number variation allows the detection of blood and semen. *Forensic Sci Int Genet*. 2018 Sep;36:112–8.
340. Sangermano R, Khan M, Cornelis SS, Richelle V, Albert S, Garanto A, et al. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res*. 2018;28(1):100–10.
341. Liquori A, Vaché C, Baux D, Blanchet C, Hamel C, Malcolm S, et al. Whole USH2A Gene Sequencing Identifies Several New Deep Intronic Mutations. *Hum Mutat*. 2016 Feb;37(2):184–93.
342. Mayer AK, Rohrschneider K, Strom TM, Glöckle N, Kohl S, Wissinger B, et al. Homozygosity mapping and whole-genome sequencing reveals a deep intronic PROM1 mutation causing cone-rod dystrophy by pseudoexon activation. *Eur J Hum Genet*. 2016 Mar;24(3):459–62.
343. Conte I, Hadfield KD, Barbato S, Carrella S, Pizzo M, Bhat RS, et al. MiR-204 is responsible for inherited retinal dystrophy associated with ocular coloboma. *Proc Natl Acad Sci U S A*. 2015 Jun 23;112(25):E3236–45.
344. Coppieters F, Todeschini AL, Fujimaki T, Baert A, De Bruyne M, Van Cauwenbergh C, et al. Hidden Genetic Variation in LCA9-Associated Congenital Blindness Explained by 5'UTR Mutations and Copy-Number Variations of NMNAT1. *Hum Mutat*. 2015 Dec;36(12):1188–96.

345. Radziwon A, Arno G, K Wheaton D, McDonagh EM, Baple EL, Webb-Jones K, et al. Single-base substitutions in the CHM promoter as a cause of choroideremia. *Hum Mutat.* 2017;38(6):704–15.
346. Dalkara D, Goureau O, Marazova K, Sahel J-A. Let There Be Light: Gene and Cell Therapy for Blindness. *Hum Gene Ther.* 2016 Feb 1;27(2):134–47.
347. Scholl HPN, Strauss RW, Singh MS, Dalkara D, Roska B, Picaud S, et al. Emerging therapies for inherited retinal degeneration. *Sci Transl Med.* 2016 07;8(368):368rv6.
348. MacLaren RE, Bennett J, Schwartz SD. Gene Therapy and Stem Cell Transplantation in Retinal Disease: The New Frontier. *Ophthalmology.* 2016 Oct;123(10 Suppl):S98–106.
349. Bennett J. Taking Stock of Retinal Gene Therapy: Looking Back and Moving Forward. *Mol Ther.* 2017 May 3;25(5):1076–94.
350. Cideciyan AV, Jacobson SG, Beltran WA, Sumaroka A, Swider M, Iwabe S, et al. Human retinal gene therapy for Leber congenital amaurosis shows advancing retinal degeneration despite enduring visual improvement. *Proc Natl Acad Sci U S A.* 2013 Feb 5;110(6):E517–25.
351. Millington-Ward S, Chadderton N, O'Reilly M, Palfi A, Goldmann T, Kilty C, et al. Suppression and Replacement Gene Therapy for Autosomal Dominant Disease in a Murine Model of Dominant Retinitis Pigmentosa. *Mol Ther.* 2011 Apr;19(4):642–9.
352. Wert KJ, Davis RJ, Sancho-Pelluz J, Nishina PM, Tsang SH. Gene therapy provides long-term visual function in a pre-clinical model of retinitis pigmentosa. *Hum Mol Genet.* 2013 Feb 1;22(3):558–67.
353. Chadderton N, Palfi A, Millington-Ward S, Gobbo O, Overlack N, Carrigan M, et al. Intravitreal delivery of AAV-NDI1 provides functional benefit in a murine model of Leber hereditary optic neuropathy. *Eur J Hum Genet.* 2013 Jan;21(1):62–8.
354. Mookherjee S, Hiriyan S, Kaneshiro K, Li L, Li Y, Li W, et al. Long-term rescue of cone photoreceptor degeneration in retinitis pigmentosa 2 (RP2)-knockout mice by gene replacement therapy. *Hum Mol Genet.* 2015 Nov 15;24(22):6446–58.
355. LaVail MM, Yasumura D, Matthes MT, Yang H, Hauswirth WW, Deng W-T, et al. Gene Therapy for MERTK-Associated Retinal Degenerations. *Adv Exp Med Biol.* 2016;854:487–93.

356. Latella MC, Di Salvo MT, Cocchiarella F, Benati D, Grisendi G, Comitato A, et al. In vivo Editing of the Human Mutant Rhodopsin Gene by Electroporation of Plasmid-based CRISPR/Cas9 in the Mouse Retina. *Mol Ther Nucleic Acids*. 2016 Nov;5(11):e389.
357. Dalkara D, Byrne LC, Klimczak RR, Visel M, Yin L, Merigan WH, et al. In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. *Sci Transl Med*. 2013 Jun 12;5(189):189ra76.
358. Castle MJ, Turunen HT, Vandenberghe LH, Wolfe JH. Controlling AAV Tropism in the Nervous System with Natural and Engineered Capsids. *Methods Mol Biol Clifton NJ*. 2016;1382:133–49.
359. Sengupta A, Chaffiol A, Macé E, Caplette R, Desrosiers M, Lampič M, et al. Red-shifted channelrhodopsin stimulation restores light responses in blind mice, macaque retina, and human retina. *EMBO Mol Med*. 2016 Nov;8(11):1248–64.
360. Gaub BM, Berry MH, Holt AE, Isacoff EY, Flannery JG. Optogenetic Vision Restoration Using Rhodopsin for Enhanced Sensitivity. *Mol Ther*. 2015 Oct;23(10):1562–71.
361. Lopes VS, Williams DS. Gene Therapy for the Retinal Degeneration of Usher Syndrome Caused by Mutations in MYO7A. *Cold Spring Harb Perspect Med* [Internet]. 2015 Jun [cited 2018 Sep 17];5(6). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4448706/>
362. Zulliger R, Conley SM, Naash MI. Non-viral therapeutic approaches to ocular diseases: an overview and future directions. *J Control Release Off J Control Release Soc*. 2015 Dec 10;219:471–87.
363. Campbell M, Humphries MM, Kiang A-S, Nguyen ATH, Gobbo OL, Tam LCS, et al. Systemic low-molecular weight drug delivery to pre-selected neuronal regions. *EMBO Mol Med*. 2011 Apr;3(4):235–45.
364. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. 2017 Feb;19(2):249–55.
365. National Rare Disease Plan for Ireland 2014-2018 | Department of Health [Internet]. [cited 2018 Aug 6]. Available from:

<https://health.gov.ie/blog/publications/national-rare-disease-plan-for-ireland-2014-2018/>

366. The 100,000 Genomes Project by numbers [Internet]. Genomics England. 2018 [cited 2018 Aug 6]. Available from: <https://www.genomicsengland.co.uk/the-100000-genomes-project-by-numbers/>

367. Smit AK, Newson AJ, Best M, Badcock C-A, Butow PN, Kirk J, et al. Distress, uncertainty, and positive experiences associated with receiving information on personal genomic risk of melanoma. *Eur J Hum Genet EJHG*. 2018 Aug;26(8):1094–100.

368. Muthane U. Predictive genetic testing in Huntington's disease. *Ann Indian Acad Neurol*. 2011 Jul;14(Suppl1):S29–30.

369. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet Lond Engl*. 2012 Nov 17;380(9855):1778–86.

370. Beau A-B, Lynge E, Njor SH, Vejborg I, Lophaven SN. Benefit-to-harm ratio of the Danish breast cancer screening programme. *Int J Cancer*. 2017 01;141(3):512–8.

371. Green ED, Guyer MS, National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. *Nature*. 2011 Feb 10;470(7333):204–13.

372. Yohe S, Thyagarajan B. Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med*. 2017 Aug 7;141(11):1544–57.

373. Read about QR-421a for Usher syndrome type 2 [Internet]. ProQR Therapeutics. [cited 2018 Aug 6]. Available from: <http://www.proqr.com/qr-421a-for-usher-syndrome-type-2/>

374. Fetal DNA sequencing potentially could reduce need for invasive prenatal diagnostic procedures [Internet]. National Institutes of Health (NIH). 2018 [cited 2018 Aug 6]. Available from: <https://www.nih.gov/news-events/news-releases/fetal-dna-sequencing-potentially-could-reduce-need-invasive-prenatal-diagnostic-procedures>

375. Bunkar N, Bhargava A, Chaudhury K, Sharma RS, Lohiya NK, Mishra PK. Fetal nucleic acids in maternal plasma: from biology to clinical translation. *Front Biosci Landmark Ed*. 2018 Jan 1;23:397–431.

376. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. 2016 Jul 1;353(6294):78–82.
377. Navarro JF, Sjöstrand J, Salmén F, Lundeberg J, Ståhl PL. ST Pipeline: an automated pipeline for spatial mapping of unique transcripts. *Bioinforma Oxf Engl*. 2017 Aug 15;33(16):2591–3.
378. with great power comes great responsibility - YouTube [Internet]. [cited 2018 Sep 27]. Available from: <https://www.youtube.com/watch?v=b23wrRfy7SM>

7 Appendices

Article

Target 5000: Target Capture Sequencing for Inherited Retinal Degenerations

Adrian Dockery ^{1,*}, Kirk Stephenson ², David Keegan ², Niamh Wynne ³, Giuliana Silvestri ^{4,5}, Peter Humphries ¹, Paul F. Kenna ^{1,3}, Matthew Carrigan ^{1,†} and G. Jane Farrar ^{1,†}

¹ The School of Genetics & Microbiology, Trinity College Dublin, Dublin 2, Ireland; peter.humphries@tcd.ie (P.H.); paul.kenna@tcd.ie (P.F.K.); carrigma@tcd.ie (M.C.); jane.farrar@tcd.ie (G.J.F.)

² The Mater Misericordiae University Hospital, Dublin 7, Ireland; kirkstephenson@hotmail.com (K.S.); dkeegan@mater.ie (D.K.)

³ The Research Foundation, Royal Victoria Eye and Ear Hospital, Dublin 2, Ireland; wynnenc@tcd.ie

⁴ Department of Ophthalmology, The Royal Victoria Hospital, Belfast BT12 6BA, Northern Ireland, UK; julie.silvestri@belfasttrust.hscni.net

⁵ Centre for Experimental Medicine, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK

* Correspondence: dockerya@tcd.ie

† Authors contributed equally.

Academic Editor: Frans P. M. Cremers

Received: 8 September 2017; Accepted: 27 October 2017; Published: 3 November 2017

Abstract: There are an estimated 5000 people in Ireland who currently have an inherited retinal degeneration (IRD). It is the goal of this study, through genetic diagnosis, to better enable these 5000 individuals to obtain a clearer understanding of their condition and improved access to potentially applicable therapies. Here we show the current findings of a target capture next-generation sequencing study of over 750 patients from over 520 pedigrees currently situated in Ireland. We also demonstrate how processes can be implemented to retrospectively analyse patient datasets for the detection of structural variants in previously obtained sequencing reads. Pathogenic or likely pathogenic mutations were detected in 68% of pedigrees tested. We report nearly 30 novel mutations including three large structural variants. The population statistics related to our findings are presented by condition and credited to their respective candidate gene mutations. Rediagnosis rates of clinical phenotypes after genotyping are discussed. Possible causes of failure to detect a candidate mutation are evaluated. Future elements of this project, with a specific emphasis on structural variants and non-coding pathogenic variants, are expected to increase detection rates further and thereby produce an even more comprehensive representation of the genetic landscape of IRDs in Ireland.

Keywords: retina; genetics; ophthalmology; retinitis pigmentosa; genomics

1. Introduction

Inherited retinal degenerations (IRDs) represent the most frequent cause of vision loss in people of working age. As a result, these conditions have a highly significant impact on quality of life and health-related costs and loss of income. IRDs are an extremely heterogeneous set of conditions associated with the loss of retinal function, and as a group, represent one of the most genetically diverse hereditary conditions. Over 260 genes to date have been implicated in the syndromic and non-syndromic IRDs [1], with a wide range of clinical presentations and rates of progression. As this is a diverse set of conditions with frequently overlapping presentations, it is typically divided into large sub-categories, primarily by specific regions or cell types affected, such as rod photoreceptors, cone photoreceptors or, for example, peripheral versus macular regions of the retina. Retinitis Pigmentosa (RP) is the most common form of IRD, is extremely genetically heterogeneous and affects

as many as 1 in 3000 individuals [2,3]. The disease is typically characterized by progressive loss of rod photoreceptor cells, followed by the gradual death of cone photoreceptors and generally involves characteristic features such as pigmentary deposits in the peripheral retina and attenuation of retinal vessels. In contrast, some other forms of IRD can be extremely rare and have a single gene aetiology; gyrate atrophy, for example, is estimated to affect roughly one in a million people [4].

IRDs are currently thought to affect approximately 2.5 million people globally. The vast majority of these individuals have received a diagnosis based on clinical phenotype alone, rather than a genetic diagnosis, if they have been formally diagnosed at all. Clinical trials are in progress for a number of IRDs, however most such trials require patients to have a known causative mutation to participate. Here we present data from Target 5000, an ongoing next generation sequencing (NGS)-based study, which aims to genetically characterise a large national cohort of IRD patients.

The most common method chosen for IRD genetic screening is targeted NGS. Although whole-exome sequencing offers the potential to locate disease-causing mutations in novel genes, in practice diagnosis rates in whole-exome and targeted-sequencing studies are similar [5], suggesting that the coding regions responsible for the majority of IRDs have been located. Although whole genome analysis has the potential to discover non-coding disease-causing mutations, the difficulty involved with data interpretation and cost associated with the study increase dramatically.

During the course of this study, over 750 individuals from over 520 pedigrees have been sequenced with a targeted NGS panel, focused on exons of 254 IRD-associated genes, in addition to a small number of introns previously reported to harbour splice-altering mutations. Here we present novel mutations primarily from over 200 patients involved in recent recruitment but also resulting from retrospective analysis based on previously recruited patient cohorts [6]. Candidate mutations were detected in over 68% of our analysed pedigrees. This figure includes previously reported pathogenic mutations and numerous novel likely pathogenic variants. Novel variants include large structural variants, point mutations with high predicted pathogenicity, frameshift mutations and splice site mutations. A single pathogenic or likely pathogenic variant was observed in an additional 8% of pedigrees in which the gene in question is known to cause a recessive retinopathy.

2. Materials and Methods

2.1. Patient Identification and Recruitment

Probands and other family members were primarily assessed at the Research Foundation of the Royal Victoria Eye and Ear Hospital (Dublin, Ireland) and the Mater Misericordiae University Hospital (Dublin, Ireland). With informed consent, best-corrected visual acuity was assessed using revised 2000 Early Treatment Diabetic Retinopathy Study (ETDRS) charts (Precision Vision, La Salle, IL, USA). Colour vision was examined using the Lanthony desaturated D-15 panel (Gulden Ophthalmics, Elkins Park, PA, USA) under standardised lighting conditions. Goldmann perimetry was used to assess the peripheral visual fields to the IV4e, I4e and 04e targets. Full-field electroretinograms were performed according to the International Society for Clinical Electrophysiology of Vision (ISCEV) standards [7] using a Roland Consult RETI-port retiscan (Brandenburg an der Havel, Germany). Fundus colour and autofluorescence photography was performed using a Topcon CRC50DX (Topcon Great Britain Ltd., Berkshire, England) or Optos Daytona (Optos plc, Dunfermline, Scotland). Spectral domain optical coherence tomography was performed using a Cirrus HD-OCT (Carl Zeiss Meditec, Berlin, Germany).

2.2. DNA Isolation and Next Generation Sequencing

With informed consent, blood samples were taken from patients after a thorough clinical assessment. DNA was isolated from 2 mL of blood and fragmented for targeted sequencing to an average fragment size of 200–250 base pairs (bp).

Sequencing libraries were generated and target capture was performed with the Nimblegen SeqCap EZ kit (Roche Ireland Ltd., Dublin, Ireland), incorporating the exonic regions of 254 genes implicated in retinopathies (Supplementary Data, Table S1: Full list of genes captured in NGS panel). Captures were executed as per the manufacturer's instructions. Capture regions also included intronic regions in *CEP290*, *ABCA4* and *USH2A* that are known to potentially contain pathogenic mutations [8–10]. The capture panel also included a small number of genes implicated in retinal development and regulation such as *CTBP2*, which encodes the RIBEYE protein, which is essential to the formation of retinal presynaptic ribbons [11]. The total size of the captured region was approximately 750 kb.

Captured patient DNA was initially multiplexed into 24-sample pools using NimbleGen Adapters (Roche Ireland Ltd., Dublin, Ireland) and sequenced using an Illumina MiSeq (Illumina Inc., San Diego, CA, USA). In more recent captures, samples were multiplexed into 96-sample pools using dual-indexed adapters from IDT (Integrated DNA technologies, <https://www.idtdna.com/>) designed to be compatible with Illumina systems, to enable up to 96-plex pooling for higher-throughput instruments. Confirmatory single-read sequencing was also performed to verify the presence of candidate mutations.

2.3. Polymerase Chain Reaction and Sanger Sequencing

In order to validate candidate mutations found in NGS experiments, amplicons containing the mutations were generated by polymerase chain reaction (PCR) and analysed by direct sequencing. Oligonucleotides were purchased from Sigma-Aldrich (Gillingham, England). The target DNA products were amplified using Q5 High-Fidelity 2× Master Mix (New England Biolabs Inc., Ipswich, MA, USA). The annealing temperature for reactions were optimised for each mutation; all other details were executed as per the supplier's recommendations. Sanger sequencing was performed by Eurofins Genomics (Ebersberg, Germany).

2.4. Gyrate Atrophy Inversion Confirmation

Four primers were designed to cover the two breakpoints of the homozygous inversion in the *OAT* gene, located at NC_000010.11:g.124405527 and NC_000010.11:g.124422152. Primers OAT-1 (forward: 5'-GGTAACCTGGATCCGGAACA-3') and OAT-2 (reverse: 5'-CTTCTGGGGTAGGACGTTGT-3') covered the breakpoint in exon 5 of *OAT* and primers OAT-3 (forward: 5'-GCGAGGGGTTTCACATCATC-3') and OAT-4 (reverse: 5'-GTTGGTGTCTCTGGCCTG-3') targeted the breakpoint upstream of the *OAT* gene. In unaffected *OAT* genes the pairing of primers OAT-1 and OAT-2 would form a product, as would the pairings of primers OAT-3 and OAT-4. However, in the specific case of the homozygous inversion reported here, no product would result from these pairings. Products in this case could only be formed when primers OAT-1 and OAT-3 were paired and likewise with primers OAT-2 and OAT-4, which would not generate any product when unaffected patient DNA was used as a template. This is shown in the Supplementary Materials diagrammatically (Figure S1: An illustration of the PCR strategy designed to detect a large homozygous inversion) and experimentally (Figure S2: Confirmation of an *OAT* inversion using strategic PCR design).

2.5. Data Analysis

NGS data was demultiplexed and mapped to the human genome (hg38) using BWA version 0.7.15 [12]. Duplicate reads were flagged using Picard version 2.5.0 [13] and downstream analysis and variant calling performed using Freebayes version 1.1.0 [14].

Patient data from earlier stages of this study [6] that has been previously analysed with different software tools and aligned to previous versions of the human reference genome was retrospectively reanalysed using the current updated methods and reference databases for the purpose of expanding mutation detection and consistent reporting. The American College of Medical Genetics and Genomics (ACMG) criteria for classifying pathogenic variants was utilised [15].

Variants were annotated using SnpEff [16], dbNSFP [17], MetaLR [18] and M-CAP [19] for the purposes of identifying rare, pathogenic coding changes. Annotations from the SPIDEX database were used to identify variants with predicted splicing impacts. Common variants, as measured by either frequency within our sample pool or frequency in external population databases, were filtered out of the analysis. Synonymous variants were also filtered out of downstream analysis, with the exception of analysis of potential splice site alterations. 15× coverage was required at a site in order to call variants. This was calculated at the individual base level rather than the exon level. For almost all samples, more than 98% of targeted sites were covered to this depth (interquartile range: 98.3–99.1%).

In addition, structural variants were called by a separate pipeline, based on the tools LUMPY [20], which was used to identify structural variants via split reads or unusual paired read alignments, and CoNIFER [21], which was used to identify structural variants based on aberrant read depths. Results from these tools were combined and filtered to output a final, high-quality list of putative structural variants and associated confidence scores for each sample.

2.6. Ethical Approval

Ethical approval for this study was awarded by the Research and Medical Ethics committee of the Royal Victoria Eye and Ear Hospital (13-06-2011: HRA-POR201097) and by the Institutional Review Board of the Mater Misericordiae University Hospital and Mater Private Hospital (MMUH IRB 1/378/1358), Dublin, Ireland prior to commencement. All work was carried out in accordance with the approved guidelines. All patients have given written informed consent before recruitment to the study. No patients under 18 years of age were included in the study.

3. Results

3.1. Presentation and Solve Rates

Probands, in addition to additional family members where available, were sampled from 523 pedigrees in Ireland. The criteria for inclusion in the study included being over 18 years of age and a full clinical examination that implicated an IRD as a likely cause of visual impairment. A chart depicting the frequency of each condition at clinical presentation is represented in Figure 1. A current list of genes included as part of the targeted sequencing has been provided in the Supplementary Materials (Table S1: Full list of genes captured in NGS panel). A pathogenic or likely pathogenic variant has been detected in over 68% of pedigrees analysed. In addition, in 8% of cases one single candidate mutation could be detected and the gene in question has previously been associated with a recessively inherited IRD; it is likely that at least some of these patients will carry a second mutation in this gene that was not detected, such as a deep intronic or structural variant [22]. This information is further depicted in Figure 2. A large number of likely pathogenic and novel mutations have also been found as part of this study (Table 1).

3.2. Stargardt Disease

Stargardt disease (STGD1; OMIM #248200) is the most common inherited macular dystrophy with an estimated incidence ranging from one in 8000 to one in 10,000 [23]. Typical clinical features include bilateral central vision loss with macular atrophy and flecks in the retinal pigmentary epithelium. The age of onset and severity is dependent on the intrinsic pathogenicity of the causative mutations however most disease appears evident within the first two decades of life [24].

Stargardt disease is the second most common IRD observed in our study. *ABCA4* was unsurprisingly the most frequently observed gene associated with STGD1 or its phenotypically similar counterpart, fundus flavimaculatus. *ABCA4* was deemed the candidate gene in a total of 80 pedigrees across these two conditions. Several novel likely pathogenic mutations were observed in this study relating to STGD1, NM_000350.2: c.5917delG, p.Val1973fs; c.735T>G, p.Tyr245*; c.4320delT,

p.Phe1440fs. These novel *ABCA4* mutations were all observed in separate Stargardt pedigrees, segregating with the condition in each family and in *trans* with known pathogenic *ABCA4* mutations.

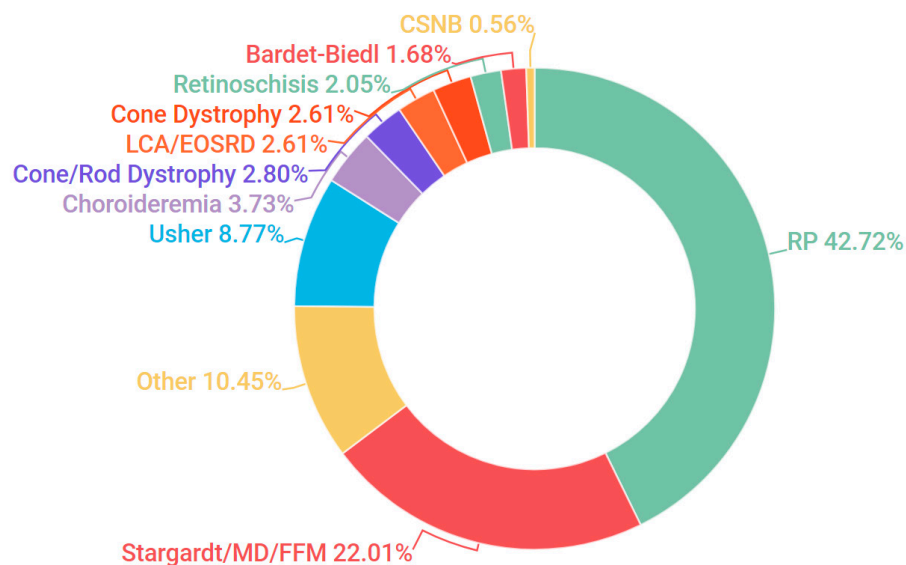


Figure 1. Clinical presentation of all inherited retinal degeneration (IRD) pedigrees included in the study. Abbreviations are listed clockwise as they appear above. RP: retinitis pigmentosa; Stargardt: Stargardt disease; MD: macular dystrophy; FFM: fundus flavimaculatus; Usher: Usher Syndrome; LCA: Leber congenital amaurosis; EOSRD: early-onset severe retinal dystrophy; Bardet-Biedl: Bardet-Biedl syndrome; CSNB: congenital stationary night blindness.

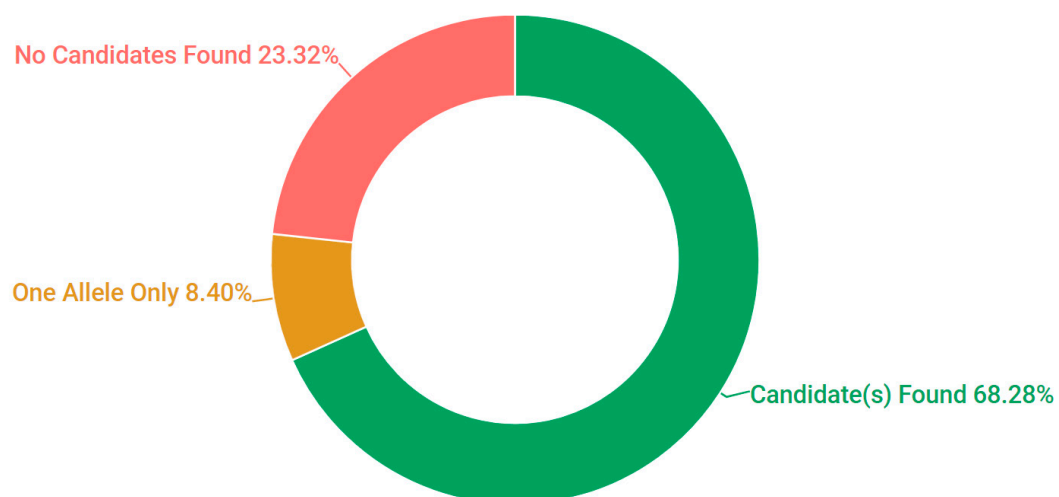


Figure 2. Pedigree solve rates employing target capture next generation sequencing for 254 IRD genes.

Despite the relative genetic homogeneity of STGD1, *ABCA4* still represents a challenge in terms of diagnostics due to the suspected prevalence of many deep intronic pathogenic mutations [25]. However, as many of these IRD diagnostic studies progress it is likely that the increasing body of data will make it possible to more accurately estimate the pathogenicity of variants discovered during sequencing. One example of this is the reclassification of the hypomorphic *ABCA4* variant p.Asn1868Ile (c.5603A>T). This variant was previously dismissed as benign due to its background population frequency of 7%. This p.Asn1868Ile variant was observed in 94 sequenced patients, including 4 patients who were homozygous for the mutation [26]. The inclusion of this single variant

in the analysis has aided in the genetic diagnosis of an additional 19 patients in our IRD patient cohort. The majority of these 19 patients were clinically diagnosed with a milder form of STGD1 or a late-onset macular degeneration but only a single pathogenic or likely pathogenic *ABCA4* mutation could previously be detected. However subsequent to modification of variant calling filters to allow for the inclusion of the p.Asn1868Ile variant, these 19 cases could be solved. The observation in the current study that the p.Asn1868Ile variant is associated with a clinically distinguishable milder form of STGD1 is consistent with the associated phenotypes outlined by Zernant and colleagues [26].

ELOVL4 and *PROM1* also contributed to a small number of pedigrees that presented as Stargardt-like dominant macular dystrophies. Novel likely pathogenic mutations were detected in *ELOVL4* during cohort analysis (Table 1). However, the most common candidate mutations associated with dominant maculopathy in our participants were found in the *BEST1* gene (Figure 3).

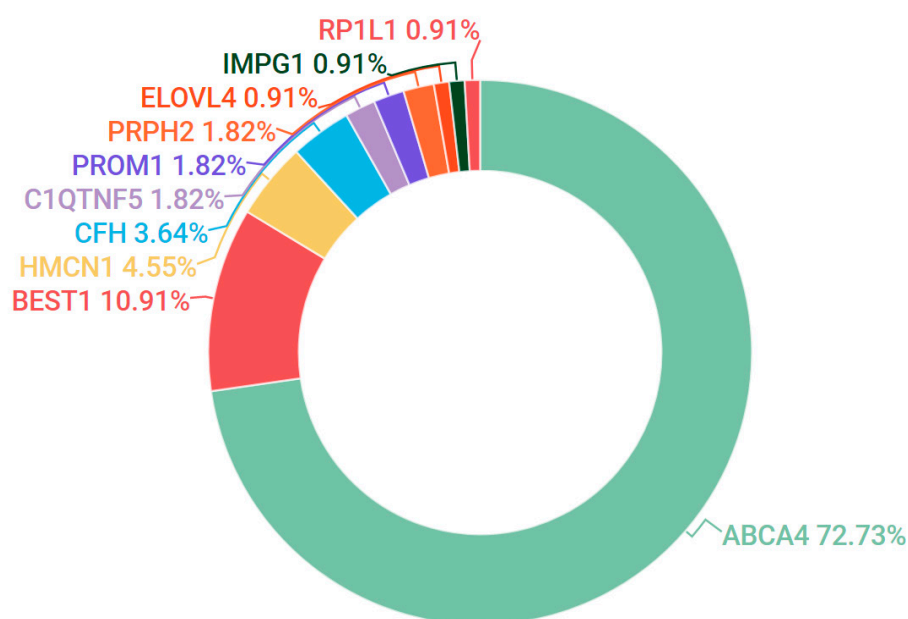


Figure 3. Gene candidates (and percentages) observed in the IRD grouping encompassing Stargardt disease, fundus flavimaculatus (FFM) and macular dystrophy (MD).

3.3. Retinitis Pigmentosa

Retinitis Pigmentosa (RP; OMIM #268000) was the most common clinical diagnosis for participants in the current study, accounting for nearly 40% of total pedigrees sequenced. This figure encompasses the various modes of inheritance (autosomal dominant, autosomal recessive and X-linked) and phenotypes (typical, inverse and paravenous among others) associated with RP.

The most common gene candidate found for dominant RP in the study was *RHO* (OMIM #180380). This accounts for over 12% of all RP pedigrees involved in the study and almost 30% of pedigrees diagnosed with dominant RP. This figure is an under representation of the prevalence of *RHO*-linked RP in this IRD population, as several pedigrees involved in prior single-gene studies on *RHO* were excluded from this study, as causative mutations had already been established [27–29]. The predominant pathogenic mutations observed are NM_000539.3:c.533A>G, p.Tyr178Cys; c.620T>G, p.Met207Arg, however, numerous novel likely pathogenic mutations have also been detected. Those variants found in sufficiently large pedigrees to verify segregation are shown in Table 1. In prior single-gene studies, the *RHO* mutations p.Tyr178Cys, p.Met207Arg and p.Thr94Ile were also commonly observed in the Irish IRD population.

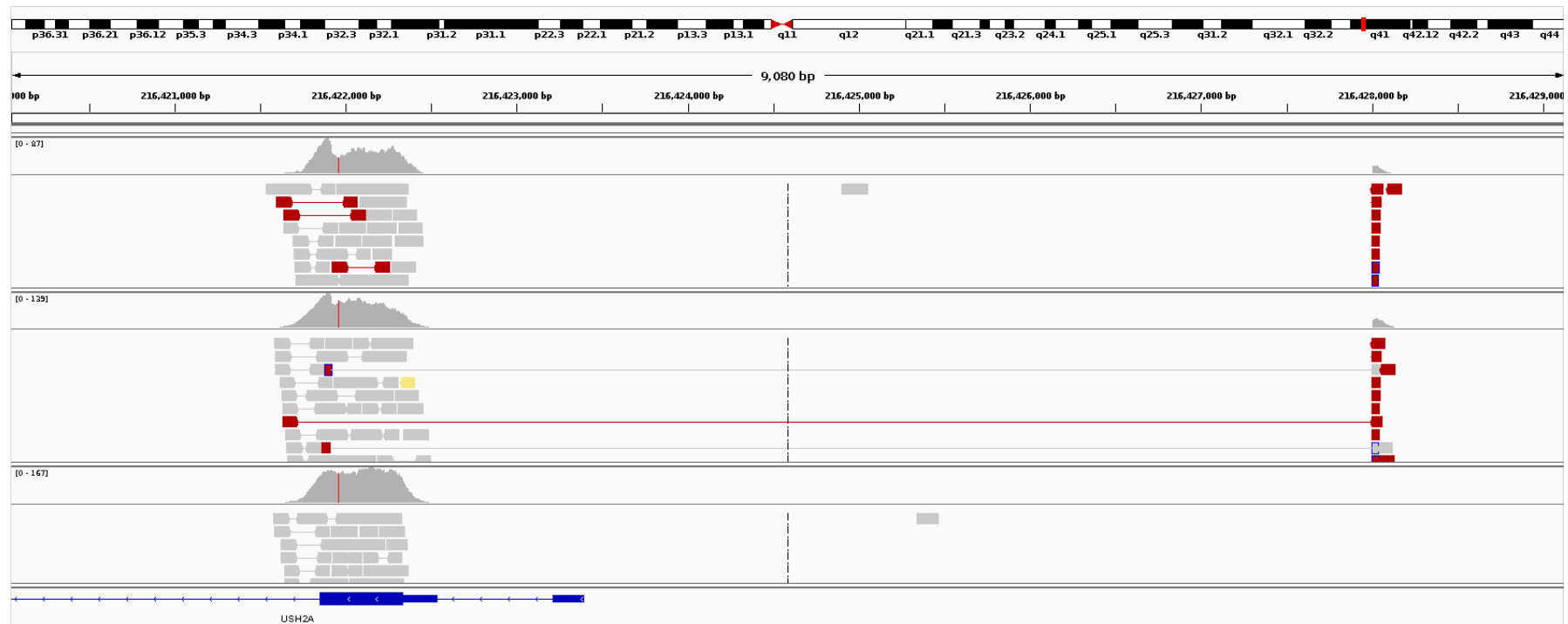


Figure 4. Sequence data from retinitis Pigmentosa (RP) patients with large *USH2A* deletions. The top two panels of reads are from the two affected patients who presented with RP. The bottom panel of reads is from an unaffected unrelated patient. Depicted above is the evidence of split reads (red) and a drop in read depth in the affected patients relative to a control.

USH2A (OMIM #608400) was the most frequently observed gene candidate for recessive RP. Pathogenic or novel mutations in this gene were observed in over 30% of recessive RP resolved cases. Although many *USH2A* pathogenic mutations are associated with Usher syndrome, none of the pedigrees which were initially clinically diagnosed with recessive RP showed any other syndromic traits. This finding of non-syndromic retinal degenerations associated with mutations in *USH2A* is consistent with observations based on other *USH2A* cohorts [30]. In one Irish IRD pedigree with two affected individuals diagnosed with recessive RP, a large heterozygous deletion was detected in *USH2A* (Figure 4) employing the structural variants analysis pipeline described in the Methods section. This deletion was approximately 9 kb in size and spans the first 150 amino acids of the coding region, genomic coordinates (hg38): g.chr1:216421919-216428002. This deletion encompasses a significant part of *USH2A*, including the start codon, transcriptional start site and part of the first exon. As such, it is highly likely to cause complete loss of function. The 9 kb mutation in *USH2A* was detected alongside a reported pathogenic mutation further downstream in the gene in both affected patients in this pedigree, NM_206933.2:c.2276G>T, p.Cys759Phe. An unaffected sibling from this pedigree was directly sequenced for both mutations and was found to be negative for the large deletion (Supplementary Data, Figure S3: Gel confirmation of *USH2A* deletion) and positive for the pathogenic single base mutation, p.Cys759Phe (Supplementary Data, Figure S4: Sanger sequencing trace of *USH2A* mutation, p.Cys759Phe) confirming that the two *USH2A* mutations observed in this pedigree are found on different alleles. The region surrounding the breakpoint in exon 2 was a long pyrimidine run, encoding a series of serine and proline residues. The region surrounding the breakpoint upstream of *USH2A* was very similar, although one purine nucleotide was present in the surrounding 25 bp. It is plausible that these sequence similarities facilitated the formation of the deletion.

RP1 (OMIM #603937) mutations have been found in both dominantly and recessively inherited forms of RP. *RP1*-RP was first described as a dominantly inherited condition [31] and then subsequently found to also be associated with a recessive mode of inheritance several years later [32]. Although there are quite a number of missense pathogenic mutations associated with *RP1*-RP [33], only a single missense mutation was observed in the current IRD cohort, p.Thr373Ile. This *RP1* mutation was found heterozygously in 34 different IRD patients from our study. As this mutation has only been observed to be causative of an IRD when homozygous or as a compound heterozygote [32], it could not be deemed causative in any of the cases in this current study as a second pathogenic mutation was not detected in the *RP1* exonic sequence. All other candidate recessive or dominant mutations in *RP1*-RP deemed to be likely causative of the observed phenotype were frameshift mutations, including a novel recessive mutation, NM_006269.1:c.160delG, p.Val54fs (see Table 1).

Of the X-linked RP cases in which a candidate variant was detected, 80% carried known or likely pathogenic variants in the *RPGR* gene (OMIM #312610). The majority of these causative *RPGR* mutations were frameshift mutations, including one novel mutation, NM_001034853.1:c.2236_2237delGA, p.Glu746fs. Given that the repetitive region of *RPGR*'s orf15 tends to negatively impact NGS efficacy [34], pathogenic mutations were rarely called in this region unless both read quality and depth were of a high standard. A small number of pedigrees also tested positive for candidate mutations in *RP2* (OMIM #300757), including a novel frameshift mutation, NM_006915.2:c.425delA, p.Asn142fs. All candidate genes found in RP pedigrees as part of this study are depicted in Figure 5.

3.4. Bardet-Biedl Syndrome

Although only 8 pedigrees in this study were initially clinically diagnosed with Bardet-Biedl Syndrome (BBS; OMIM #209900), an additional 19 pedigrees were rediagnosed as BBS, the majority of which presented clinically as simplex RP as no other syndromic signs characteristic of BBS were obvious to the clinical staff initially. This rediagnosis was prompted by genetic findings and where possible, confirmed by the ophthalmologists associated with the project. A number of these cases were subsequently found to have had polydactyly, which had typically been surgically removed at a

young age. Additionally, a smaller number of patients were seen to have more prominent intellectual disabilities and indications of obesity upon annual clinical check-ups. The most frequent mutation detected in the patient population was *BBS1* (OMIM #209901) NM_024649.4:c.1169T>G, p.Met390Arg. This mutation was observed homozygously in 17 pedigrees, and heterozygously in another 4 pedigrees. It was also the most likely BBS mutation to present as simplex RP. The findings are consistent with other studies carried out with Northern European IRD patient cohorts [35,36]. Five additional BBS pedigrees revealed candidate mutations across four other known BBS genes: *BBS4* (OMIM #600374), *BBS9/PTHB1* (OMIM #607968), *BBS10* (OMIM #610148) and *BBS16/SDCCAG8* (OMIM #613524).

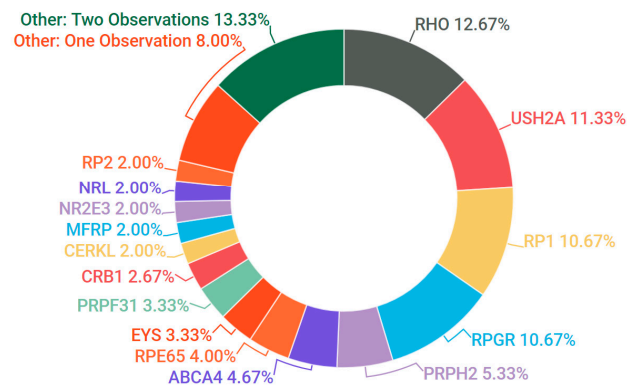


Figure 5. Gene candidates (and percentages) observed in cases of RP. Genes with two observations: *CRX*, *IFT140*, *KLHL7*, *PDE6A*, *PDE6B*, *PROM1*, *PRPF8*, *ROM1*, *SLC24A1*, *SNRNP200*. Genes with single observations: *AHI1*, *C2orf71*, *CNGA1*, *CNGB1*, *GNAT1*, *GUCA1B*, *HK1*, *IMPDH1*, *OFD1*, *PRPF6*, *RDH12*, *TULP1*.

3.5. Usher Syndrome

Usher Syndrome (USH) is diagnosed by the concurrent incidence of auditory impairment and RP. Three subtypes of USH were detected in our patient cohort, Type 1 (OMIM #276900), Type 2 (OMIM #276901) and Type 3 (OMIM #276902). These subtypes can be broadly identified by severity of hearing loss: congenital deafness, loss of hearing and progressive loss of hearing respectively. 47 pedigrees in the study were clinically diagnosed with some form of USH and candidate mutations were found in 44 pedigrees (94%). Pathogenic or likely pathogenic candidate mutations in *USH2A* were detected in 50% of all USH cases, which accounts for the majority of USH Type 2 cases. The most frequent candidate gene for USH Type 1 was *MYO7A* (OMIM #276903). Together, *USH2A* and *MYO7A* accounted for over 70% of the pathogenic mutations detected across all USH subtypes (Figure 6).

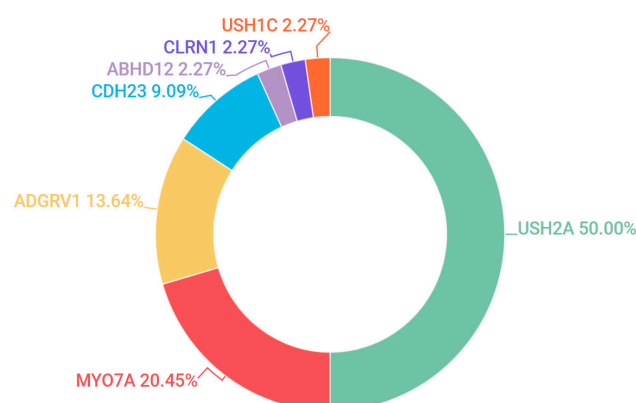


Figure 6. Gene candidates (and percentages) found across all subtypes of Usher Syndrome.

A large structural variant was also detected in an Usher pedigree. In this instance, a patient presented with Usher Type 1 (Figure 7) and no variants in Usher associated genes were detected. However, once a detailed analysis was undertaken with our structural variants pipeline (Methods: Structural Variants), it became apparent that a large homozygous deletion spanning the first four exons of *USH1C* was present, (Figure 8). These findings were verified by PCR experiments designed to amplify the first four exons of *USH1C* (Supplementary Data, Figure S5: Analysis of *USH1C* deletion by use of PCR products).

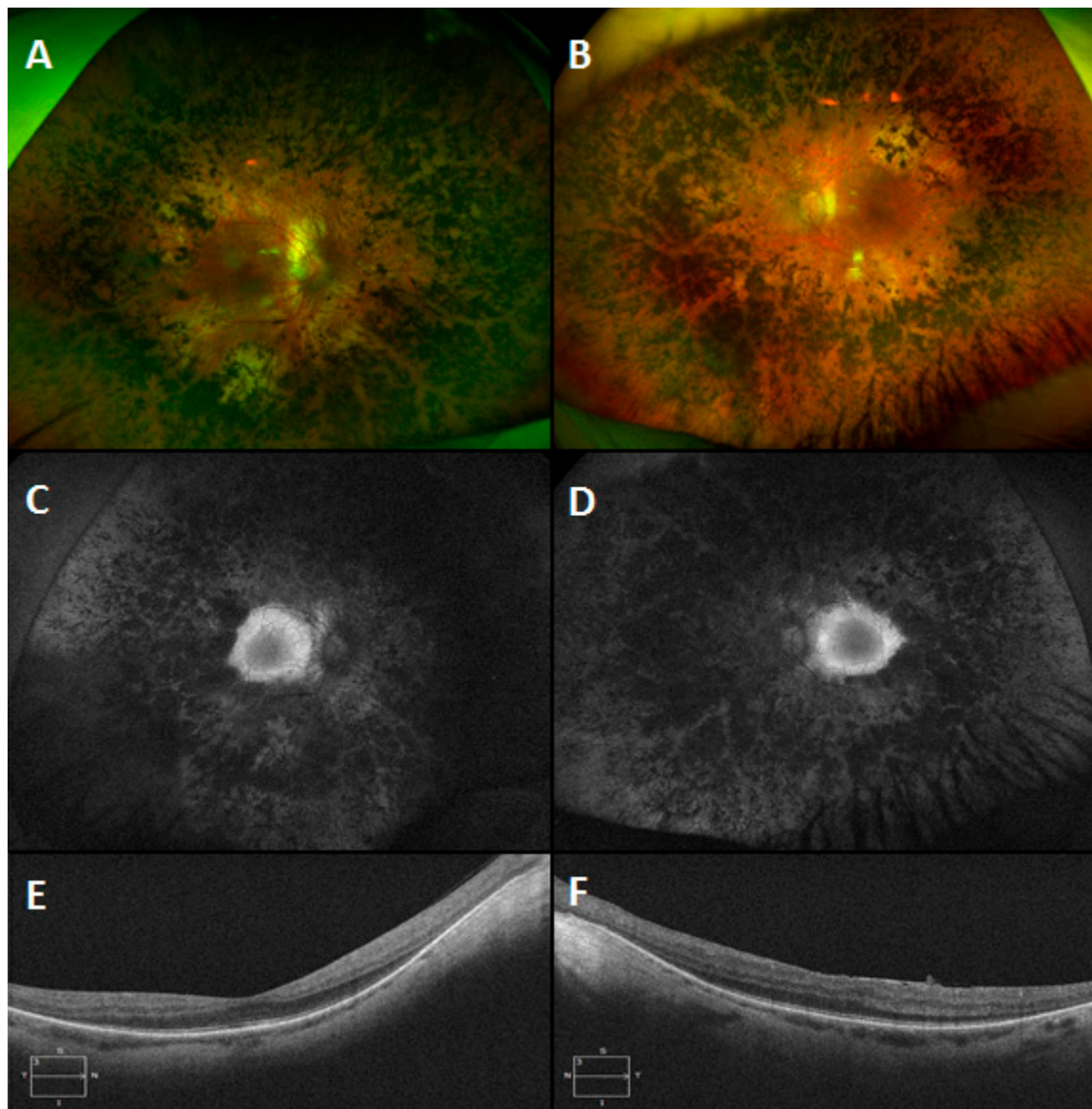


Figure 7. Montage of clinical images from an Usher syndrome patient with a large homozygous deletion in *USH1C*. Colour images of the right (A) and left (B) eyes of a 44 year old male patient with Usher syndrome type 1 which show symmetrical circumferential mid-peripheral bone spicule pigmentation and RPE atrophy with relative preservation of the maculae. The corresponding autofluorescence images of the right (C) and left (D) eyes demonstrate peripheral hypoautofluorescence and preserved macular autofluorescence. Note the symmetrical hyperautofluorescent rings, which correspond to the extent of the preserved macular photoreceptors on the Optical Coherence Tomography (OCT) scans ((E): right eye and (F): left eye).

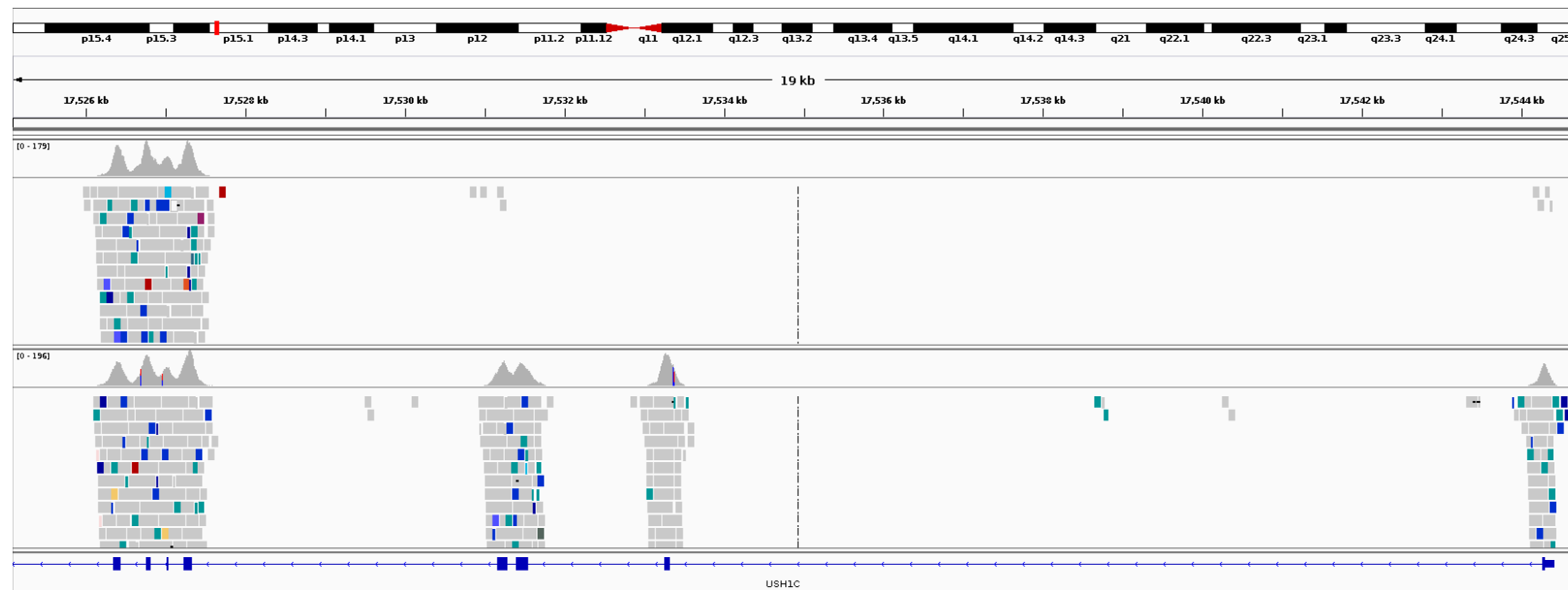


Figure 8. Large homozygous deletion spanning the first four exons of *USH1C*. In the top panel of reads from the patient sample, it is clear that there was no sequenceable template in this region compared to the control (bottom panel of reads).

3.6. Gyrate Atrophy

Gyrate atrophy/gyrate atrophy of choroid and retina (GA/GACR; OMIM #258870) is an extremely rare condition linked with a single gene aetiology, *OAT*. It is estimated that the global incidence of GA is less than one in a million, with the exception of Finland which has an estimated incidence of one in 50,000 [5]. In some regions of the world, such as Australia, the condition is sufficiently rare that the first case of GA has only been reported in the last few years [37]. The *OAT* gene encodes a mitochondrial matrix enzyme, ornithine aminotransferase, which is involved in many biologically fundamental pathways such as the urea cycle, biosynthesis of creatine and proline metabolism [38]. GA has a range of consequential phenotypes associated with it, primarily cystoid macular edema. It is for this reason that an accurate diagnosis is essential as these conditions are becoming increasingly treatable, both through dietary restrictions and medical intervention [39].

Here we report a novel homozygous inversion of the *OAT* gene. To the best of our knowledge, this is the first large inversion to be reported in this gene. The patient presented in the clinic with an obvious GA phenotype (Figure 9) and following analysis with the in-house structural variant calling pipeline (Methods: Structural Variants) it was apparent from the breakpoints that a 16.6 kb region had been inverted. The region starts approximately 3 kb upstream of the gene and ends in the middle of exon 5 of the gene (Figures 10 and 11). These findings were subsequently confirmed by alternating primer pairs designed to target the breakpoint regions of this inversion with unaffected patient DNA used as controls (Supplementary Data: Figure S2: Confirmation of an *OAT* inversion using strategic PCR design).

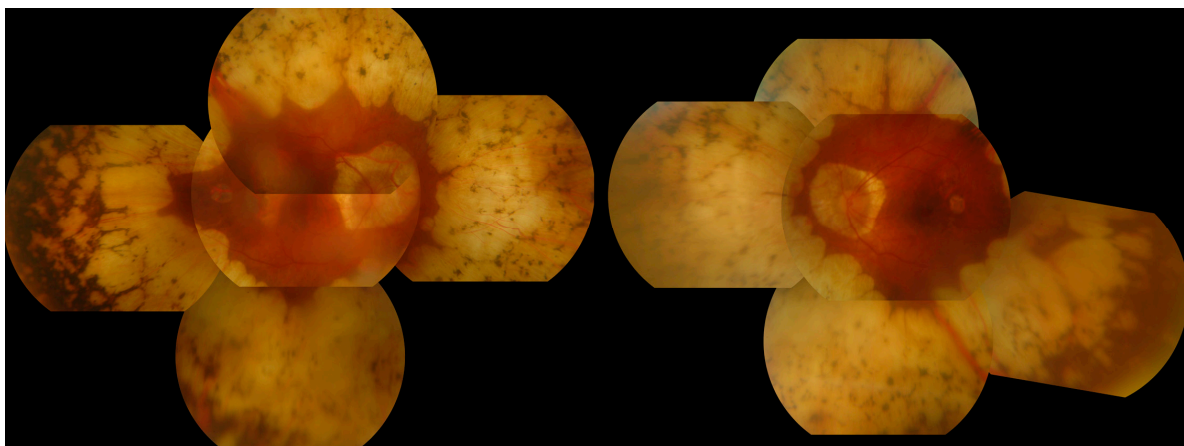


Figure 9. Montage of fundus photography of an affected gyrate atrophy (GA) patient with an *OAT* inversion. Note the marked retinal and choroidal atrophy affecting the mid-peripheral fundus of each eye with relative sparing of the central macular area. The preserved macular areas show the scalloped edges which are characteristic of the condition.

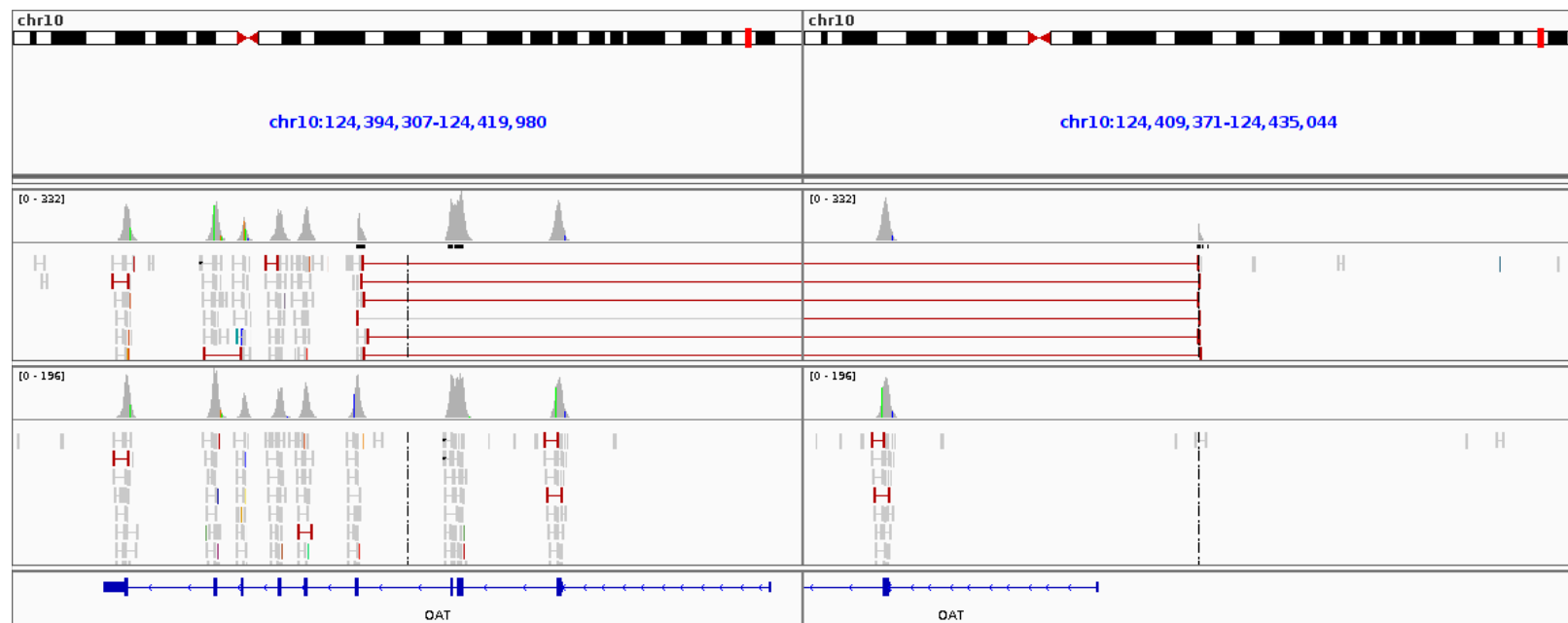


Figure 10. Sequencing results from the gyrate atrophy patient aligned to the reference genome. The top collection of sequence reads are from the affected GA patient and the bottom section of reads are from an unaffected patient sequenced in parallel. Split reads (red) from the affected patient can be seen to be spanning a 16.6 kb region of the *OAT* gene as viewed in IGV (Integrative Genomics Viewer).

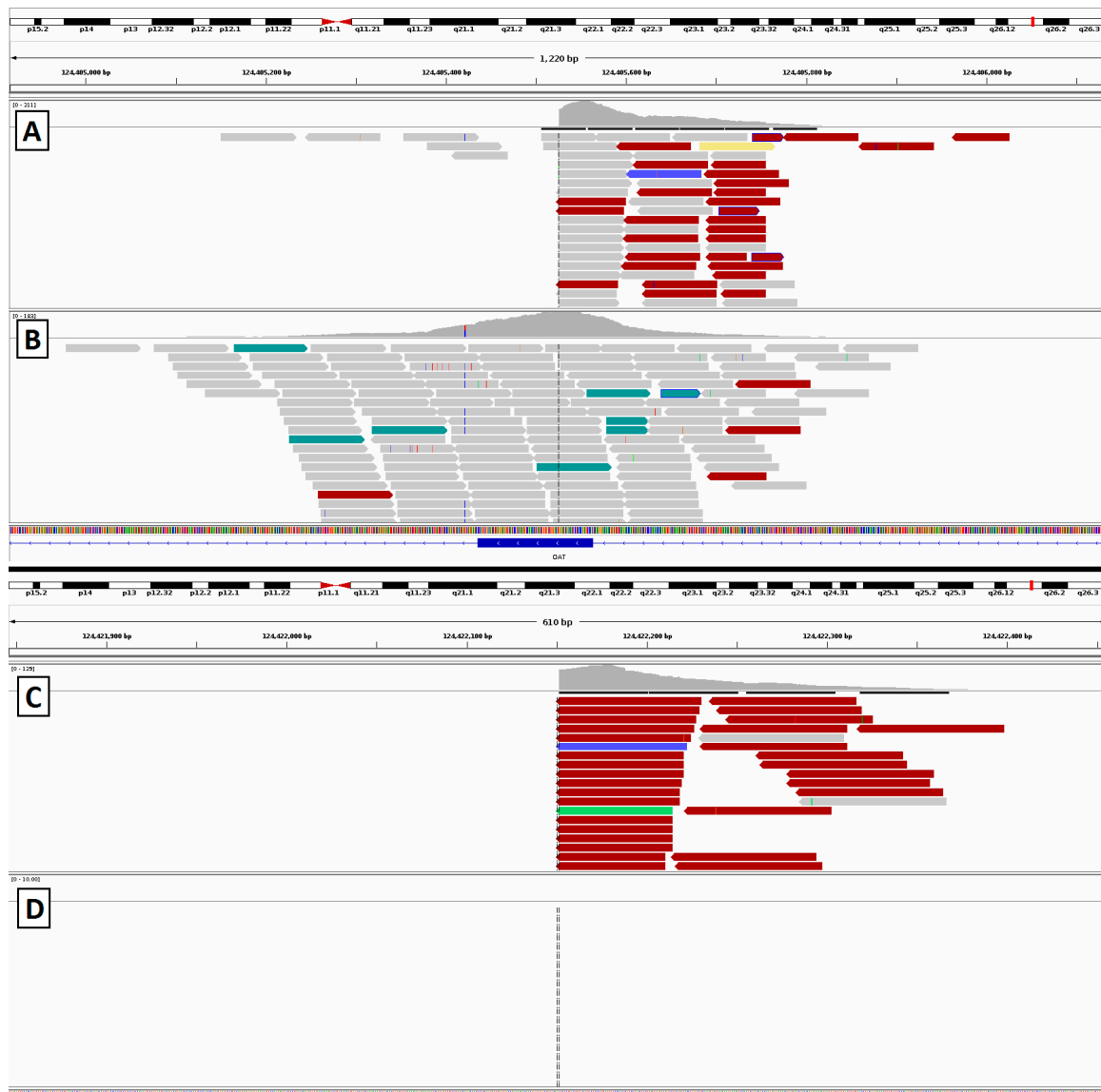


Figure 11. Sequencing results from the gyrate atrophy patient depicting alignment of reads at each breakpoint. (A,B) show reads surrounding the exonic breakpoint in the affected patient (A) and an unaffected control (B). (C,D) show the distal breakpoint. This region is not covered by our capture panel, but in the affected patient sample (C) alignments can be seen from library fragments that span the inversion breakpoint. In the control patient (D), no aligned reads are seen, as expected.

3.7. Novel Variants

Several novel likely pathogenic variants were detected in the course of this study across all conditions. Pathogenicity of novel missense mutations was predicted using the various bioinformatics tools outlined in the Methods section including MetaLR, M-CAP and SPIDEX. Segregation analysis was undertaken for pedigrees sufficiently large to do so where available. A number of the variants reported here (Table 1) had already been allocated a dbSNP ID (CNGA3 p.Arg410Trp, rs137852608; RS1 c.326+1G>A, rs281865346) as they had been detected in previous sequencing studies, most likely as part of population studies not specifically focused on the detection of IRDs. Furthermore, those variants that had been previously identified contained no record of known pathogenicity or of the phenotype of the individual in which they were identified.

Table 1. Table of Novel Likely Pathogenic Variants. Missense mutations had to show segregation in pedigrees of at least 3 members, two of whom had to be affected. This table does not include the novel large structural variants that are reported in previous sections. The mutations in this table have not been previously associated with IRDs to the best of our knowledge.

Gene	Condition	Transcript ID	Genomic Location	Nucleotide Change	Protein Change	MetaLR Score	Spidex Z-Core	Observed with:
<i>ABCA4</i>	Stargardt Disease	NM_000350.2	g.1:94007721	c.5917delG	p.Val1973fs	None	None	p.Arg290Trp
<i>ABCA4</i>	Stargardt Disease	NM_000350.2	g.1:94098827	c.735T>G	p.Tyr245 *	None	−3.542	p.Asn1868Ile
<i>ABCA4</i>	Stargardt Disease	NM_000350.2	g.1:94030459	c.4320delT	p.Phe1440fs	None	None	p.Asn1868Ile
<i>ADGRV1</i>	USH Type II	NM_032119.3	g.1:91153391	c.18795delA	p.Leu6265fs	None	None	p.Arg5772 *
<i>BBS1</i>	Bardet-Biedl	NM_024649.4	g.11:66531658	c.1614delC	p.Leu539fs	None	None	p.Met390Arg
<i>CNGA3</i>	Cone Dystrophy	NM_001298.2	g.2:98396398	c.1228C>T	p.Arg410Trp	0.9441	None	p.Arg410Trp
<i>ELOVL4</i>	Stargardt Disease	NM_022726.3	g.6:79916758	c.789delTAACTTinsAACT	p.Phe265fs_Asn264fs	None	None	/
<i>MYO7A</i>	USH Type I	NM_000260.3	g.11:77202351	c.5095C>T	p.Gln1699 *	None	−3.017	p.Ala2009fs
<i>MYO7A</i>	USH Type I	NM_000260.3	g.11:77208775	c.6025delG	p.Ala2009fs	None	None	p.Gln1699 *
<i>NR2E3</i>	Retinitis Pigmentosa (Recessive)	NM_014249.3	g.15:71813635	c.994G>T	p.Glu332 *	None	None	p.119-2A>C
<i>NYX</i>	Congenital Stationary Night Blindness	NM_022567.2	g.X:41474475	c.1022A>C	p.Asp341Ala	0.7224	None	/
<i>OPA1</i>	Optic Atrophy	NM_130837.2	g.3:193614740	c.53_62delTGAAACACAG	p.Val18fs	None	None	/
<i>PRPF31</i>	Retinitis Pigmentosa (Dominant)	NM_015629.3	g.19:54123748	c.528-1G>T	/	None	−2.131	/
<i>PRPF8</i>	Retinitis Pigmentosa (Dominant)	NM_006445.3.2	g.17:1653571	c.6337_6339delAAG	p.Lys2113del	None	None	/
<i>RHO</i>	Retinitis Pigmentosa (Dominant)	NM_000539.3	g.3:139531026	c.512C>A	p.Pro171Gln	0.4368	0.875	/
<i>RP1</i>	Retinitis Pigmentosa (Recessive)	NM_006269.1	g.8:54621124	c.160delG	p.Val54fs	None	None	p.Val54fs
<i>RP2</i>	Retinitis Pigmentosa (X-Linked)	NM_006915.2	g.X:46853795	c.425delA	p.Asn142fs	None	None	/
<i>RPGR</i>	Retinitis Pigmentosa (X-Linked)	NM_001034853.1	g.X:38286761	c.2236_2237delGA	p.Glu746fs	None	None	/
<i>RPGRIP1</i>	Leber Congenital Amaurosis	NM_020366.3	g.14:21328427	c.2899C>T	p.Gln967 *	None	−3.491	p.Gln967 *
<i>RS1</i>	Retinoschisis	NM_000330.3	g.X:18644535	c.416delA	p.Gln139fs	None	None	/
<i>RS1</i>	Retinoschisis	NM_000330.3	g.X:18647190	c.326+1G>A	/	None	−2.194	/
<i>USH2A</i>	Retinitis Pigmentosa (Recessive)	NM_206933.2	g.1:215647520	c.14791+2T>A	/	None	−2.921	p.Pro4090Thr
<i>USH2A</i>	USH Type II	NM_206933.2	g.2015680269	c.12172_12172delCTGinsTAAA	p.Leu4058fs	None	None	p.Glu767fs

4. Discussion

To date, as part of Target 5000, over 15% of the Irish IRD patient population has been sequenced, providing the first national-scale overview of the IRD landscape. The study offers not only a chance to discover new pathogenic variants in known IRD genes, but represents a vital initial step in the genetic characterisation of patients to provide them with information regarding the underlying genetic pathogenesis of their disease. Previously, we have reported the identification of over 40 novel variants in a smaller cohort of IRD patients [14,40] and here we describe an additional 23 novel mutations and 3 novel structural variants, totaling nearly 70 novel IRD mutations discovered as part of this study. Several mutations that have been previously reported such as *RHO*, p.Met207Arg [28], have presented in multiple pedigrees in this study. It is likely that some or all of these pedigrees are distantly related and current analysis is ongoing to verify this.

Significantly, the genetic pathogenesis of some previously ambiguous disease phenotypes has also been resolved, most notably the milder, late-onset phenotype of Stargardt disease that is associated with the p.Asn1868Ile *ABCA4* mutation. Also, NGS-based genetic diagnoses of IRD patients in this cohort prompted a clinical re-evaluation for many patients, predominantly from simplex RP to BBS, caused by the p.Met390Arg mutation in the *BBS1* gene, patients often presenting with subtle additional phenotypes due to, for example, early intervention for polydactyly.

Many challenges still remain for the application of NGS technologies in diagnostic medicine. Ambiguous disease phenotypes and the presence of disease genes that may be associated with multiple IRDs and different modes of inheritance can make achieving a robust diagnosis particularly difficult. The presence of stretches of repetitive sequence in some IRD genes can also make it difficult to confidently call variants in relevant portions of the genome, which we anticipate may mask some disease-causing variants from analysis. For example, an approximately 800 bp region in the centre of *RPGR* ORF15 shows a sharp drop in mapped reads due to the repetitive nature of the sequence [34]. Given that ORF15 has been implicated in cases of X-linked RP in the past [41], we anticipate that some undiagnosed patients with X-linked retinitis pigmentosa are likely to harbour mutations in this region of the gene, and efforts are underway to augment the protocol we employ to improve coverage of this region; in future sequencing panels, we hope to incorporate an augmented protocol that enhances the success of sequencing in this region. Previous studies have shown that processes can be implemented prior to sample preparation for sequencing [42] or as a parallel investigation [43] to aid sequencing this region.

The RP-associated *RP1* gene also presents some challenges in relation to genetic diagnostics. The pathogenicity and mode of inheritance of a novel mutation in *RP1* is difficult to determine as mutations in this gene have previously been associated with both dominant and recessively inherited disease. This is an issue that has been discussed in a number of other *RP1* studies [44–46]. In a recent study, a meta-analysis of previously reported *RP1* pathogenic mutations was undertaken to link the impact of each variant to the functional region of the protein. Difficulties remain in identifying new *RP1* mutations as dominant- or recessive-acting, however, as regions of the gene predominately associated with dominant RP were found to also harbour mutations associated with recessive rod-cone dystrophies [47].

Methods for NGS data analysis are undoubtedly evolving quite rapidly. Research into the effects of splice site mutations and their respective functional impacts is already providing significant insights into the effect(s) of previous potentially overlooked variants in IRD datasets [48–50]. Additionally, it is becoming increasingly commonplace for NGS studies of IRD populations to incorporate some form of detection or analysis of copy number variants (CNVs) [51,52]. It has been shown in these studies that close to 20% of previously unsolved IRD pedigrees can be resolved with the detection of pathogenic CNVs.

In the current study we describe the bioinformatics methodologies employed to retrospectively analyse datasets to detect CNVs and sequence breakpoints that are present within the captured exonic regions assessed by target capture NGS. Adopting these methodologies, we have identified three IRD

pedigrees carrying three separate large structural variants; a heterozygous large deletion in the *USH2A* gene, a homozygous large deletion in the *USH1C* gene and a homozygous large inversion in the *OAT* gene. The structural variants observed using this approach were identified in genes as diverse as the conditions themselves involving gyrate atrophy (*OAT*), retinitis pigmentosa (*USH2A*) and Usher syndrome (*USH1C*). These findings serve to emphasise the importance of implementing analysis systems that enable detection of large scale deletions and inversions in all IRD patients, as currently, we have observed that 100% of these rare SV events correlate with an IRD gene-associated pathogenic phenotype. Although split-read and read-depth analysis of short-read capture data, as performed in our study, is less sensitive to structural variants than similar methods applied to whole genome sequencing (WGS) data, it has the great advantage that the data it requires is already generated as part of standard sequencing pipelines.

It is highly likely that other structural variants may be present in our cohort but remained undetected as their breakpoints lay outside the exonic regions targeted by the capture panel. This was partly solved by the use of read depth analysis instead, as was successfully applied in the case of the *USH1C* deletion, which had no exonic breakpoints, but this method struggles to detect structural variants that do not span several exons. Despite these limitations, we have demonstrated the utility of this approach for IRD diagnostics by generating clinically actionable results even from past datasets, and we recommend that it be used as a 'stopgap' measure to improve diagnosis rates in similar projects before more comprehensive studies can be performed.

Thus far, during the course of the study, genetic analysis of IRD patients has identified candidate mutations in approximately 68% of cases. The diagnostics rates obtained is in line with other NGS studies [4,25,53–55]. The growing body of data from NGS studies of IRDs similar to this one should facilitate the formation of better correlations between genotype and phenotype. As research in parallel studies such as natural histories of IRDs [56,57] and functional analysis of modifier loci [49] continue, this information in conjunction with NGS data will undoubtedly contribute to improvements in detecting pathogenic genetic variants responsible for IRDs, as well as providing insights regarding prognoses for some IRDs and importantly may also facilitate the future delivery of gene-specific treatments to the applicable patient populations.

Non-coding variants such as splice-affecting variants, either proximal or distal to canonical splice sites, are also likely to represent a significant fraction of the unobserved disease-causing variants [22,50]. Previous studies have identified deep intronic variants that lead to intronic sequence being incorrectly retained in the mature mRNA as relevant to IRDs [58]. These variants are highly likely to be missed by current studies, as very few capture panels target introns, and the interpretation of deep intronic variants is complicated as these regions are less constrained by purifying selection, leading to large numbers of observed variants. Despite the few direct observations, strong indirect evidence of unobserved disease-causing variants in known IRD genes exists. Whole-exome studies have very similar detection rates to studies focused merely on IRD-associated genes [25], implying that coding mutations in unsequenced genes do not represent a large fraction of unobserved disease-causing mutations. Furthermore, recessive pedigrees that could not be solved in this study with a panel of 254 genes were significantly enriched for single mutations in disease-relevant genes, strongly suggesting the presence of second, as yet unobserved intronic mutations.

Despite these sources of as yet 'missed' variation causative of IRDs, the results of this study so far highlight the vast levels of genetic heterogeneity inherent in IRDs in the Irish population and the significant value of a target capture NGS-based genetic evaluation for diagnostic purposes. This has been clearly exemplified by the clinical re-categorisation of the disease pathology for several patients (for example, RP as BBS), the value of detecting pathogenic large structural variants and the continued reanalysis of patient datasets for emerging, previously undetected common pathogenic variants (*ABCA4*, p.Asn1868Ile) all of which were driven by NGS-based genetic data analysis. Future and ongoing studies, with a particular focus on structural variants and non-coding disease-causing variants,

are likely to increase mutation detection rates further and yield an even more complete picture of the genetic architecture of IRDs in Ireland.

Supplementary Materials: The following are available online at www.mdpi.com/2073-4425/8/11/304/s1, Figure S1: An illustration of the PCR strategy designed to detect a large homozygous inversion; Figure S2: Confirmation of an OAT inversion using strategic PCR design; Figure S3: Gel confirmation of USH2A deletion; Figure S4: Sanger sequencing trace of USH2A mutation, p.Cys759Phe; Figure S5. Analysis of USH1C deletion by use of PCR products, Table S1: Full list of genes captured in NGS panel.

Acknowledgments: This research was supported by grant awards from Fighting Blindness Ireland (FB Irl), The Health Research Board of Ireland (HRB), The Medical Research Charities Group (MRCG), the Irish Research Council (IRC) and Science Foundation Ireland (SFI). We would also like to acknowledge the assistance of Elaine Kenny at the Trinity Genome Sequencing Laboratory, Dublin.

Author Contributions: A.D., M.C. and G.J.F. conceived and designed the experiments; A.D. performed the experiments and generated the figures; A.D. and M.C. analysed the data; P.F.K., N.W., K.S. and D.K. recruited and clinically assessed patients involved in the project and collected patient samples; M.C. tested and implemented additional bioinformatics tools; P.F.K., K.S. and D.K. contributed fundus photography to the paper; G.S. and P.H. assisted in the formation of the study; A.D., M.C. and G.J.F. wrote the paper; G.J.F. is the PI of the laboratory in which sequence analysis and sample preparation were carried out.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. RetNet. Available online: <http://www.sph.uth.tmc.edu/RetNet/> (accessed on 1 November 2017).
2. Ferrari, S.; Di Iorio, E.; Barbaro, V.; Ponzin, D.; Sorrentino, F.S.; Parmeggiani, F. Retinitis pigmentosa: Genes and disease mechanisms. *Curr. Genomics* **2011**, *12*, 238–249. [[CrossRef](#)] [[PubMed](#)]
3. Ali, M.U.; Rahman, M.S.U.; Cao, J.; Yuan, P.X. Genetic characterization and disease mechanism of retinitis pigmentosa; current scenario. *3 Biotech* **2017**, *7*, 251. [[CrossRef](#)] [[PubMed](#)]
4. Bangal, S.; Bhandari, A.; Dhaytadak, P.; Gogri, P. Gyrate atrophy of choroid and retina with myopia, cataract and systemic proximal myopathy: A rare case report from rural India. *Australas. Med. J.* **2012**, *5*, 639–642. [[CrossRef](#)] [[PubMed](#)]
5. Farrar, G.J.; Carrigan, M.; Dockery, A.; Millington Ward, S.; Palfi, A.; Chadderton, N.; Humphries, M.; Kiang, A.S.; Kenna, P.F.; Humphries, P. Toward an elucidation of the molecular genetics of inherited retinal degenerations. *Hum. Mol. Genet.* **2017**. [[CrossRef](#)] [[PubMed](#)]
6. Carrigan, M.; Duignan, E.; Malone, C.P.G.; Stephenson, K.; Saad, T.; McDermott, C.; Green, A.; Keegan, D.; Humphries, P.; Kenna, P.F.; et al. Panel-Based Population Next-Generation Sequencing for Inherited Retinal Degenerations. *Sci. Rep.* **2016**, *6*, 33248. [[CrossRef](#)] [[PubMed](#)]
7. Marmor, M.F.; Fulton, A.B.; Holder, G.E.; Miyake, Y.; Brigell, M.; Bach, M. International Society for Clinical Electrophysiology of Vision (ISCEV) Standard for full-field clinical electroretinography (2008 update). *Doc. Ophthalmol.* **2009**, *118*, 69–77. [[CrossRef](#)] [[PubMed](#)]
8. Den Hollander, A.I.; Koeneke, R.K.; Yzer, S.; Lopez, I.; Arends, M.L.; Voesenek, K.E.J.; Zonneveld, M.N.; Strom, T.M.; Meitinger, T.; Brunner, H.G.; et al. Mutations in the *CEP290* (*NPHP6*) gene are a frequent cause of Leber congenital amaurosis. *Am. J. Hum. Genet.* **2006**, *79*, 556–561. [[CrossRef](#)] [[PubMed](#)]
9. Bauwens, M.; De Zaeytijd, J.; Weisschuh, N.; Kohl, S.; Meire, F.; Dahan, K.; Depasse, F.; De Jaegere, S.; De Ravel, T.; De Rademaeker, M.; et al. An augmented *ABCA4* screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum. Mutat.* **2015**, *36*, 39–42. [[CrossRef](#)] [[PubMed](#)]
10. Steele-Stallard, H.B.; Le Quesne Stabej, P.; Lenassi, E.; Luxon, L.M.; Claustres, M.; Roux, A.-F.; Webster, A.R.; Bitner-Glindzicz, M. Screening for duplications, deletions and a common intronic mutation detects 35% of second mutations in patients with *USH2A* monoallelic mutations on Sanger sequencing. *Orphanet J. Rare Dis.* **2013**, *8*, 122. [[CrossRef](#)] [[PubMed](#)]
11. Maxeiner, S.; Luo, F.; Tan, A.; Schmitz, F.; Südhof, T.C. How to make a synaptic ribbon: RIBEYE deletion abolishes ribbons in retinal synapses and disrupts neurotransmitter release. *EMBO J.* **2016**, *35*, 1098–1114. [[CrossRef](#)] [[PubMed](#)]
12. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]

13. Picard. Available online: <http://broadinstitute.github.io/picard/> (accessed on 1 November 2017).
14. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**, arXiv:1207.3907.
15. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. ACMG Laboratory Quality Assurance Committee Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–424. [[CrossRef](#)] [[PubMed](#)]
16. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)] [[PubMed](#)]
17. Liu, X.; Wu, C.; Li, C.; Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.* **2016**, *37*, 235–241. [[CrossRef](#)] [[PubMed](#)]
18. Dong, C.; Wei, P.; Jian, X.; Gibbs, R.; Boerwinkle, E.; Wang, K.; Liu, X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **2015**, *24*, 2125–2137. [[CrossRef](#)] [[PubMed](#)]
19. Jagadeesh, K.A.; Wenger, A.M.; Berger, M.J.; Guturu, H.; Stenson, P.D.; Cooper, D.N.; Bernstein, J.A.; Bejerano, G. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **2016**, *48*, 1581–1586. [[CrossRef](#)] [[PubMed](#)]
20. Layer, R.M.; Chiang, C.; Quinlan, A.R.; Hall, I.M. LUMPY: A probabilistic framework for structural variant discovery. *Genome Biol.* **2014**, *15*, R84. [[CrossRef](#)] [[PubMed](#)]
21. Krumm, N.; Sudmant, P.H.; Ko, A.; O’Roak, B.J.; Malig, M.; Coe, B.P.; NHLBI Exome Sequencing Project; Quinlan, A.R.; Nickerson, D.A.; Eichler, E.E. Copy number variation detection and genotyping from exome sequence data. *Genome Res.* **2012**, *22*, 1525–1532. [[CrossRef](#)] [[PubMed](#)]
22. Schulz, H.L.; Grassmann, F.; Kellner, U.; Spital, G.; Rütther, K.; Jägle, H.; Hufendiek, K.; Rating, P.; Huchzermeyer, C.; Baier, M.J.; et al. 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. *Investig. Ophthalmol. Vis. Sci.* **2017**, *58*, 394–403. [[CrossRef](#)] [[PubMed](#)]
23. Strauss, R.W.; Ho, A.; Muñoz, B.; Cideciyan, A.V.; Sahel, J.-A.; Sunness, J.S.; Birch, D.G.; Bernstein, P.S.; Michaelides, M.; Traboulsi, E.I.; et al. The Natural History of the Progression of Atrophy Secondary to Stargardt Disease (ProgStar) Studies: Design and Baseline Characteristics: ProgStar Report No. 1. *Ophthalmology* **2016**, *123*, 817–828. [[CrossRef](#)] [[PubMed](#)]
24. Michaelides, M.; Hunt, D.M.; Moore, A.T. The genetics of inherited macular dystrophies. *J. Med. Genet.* **2003**, *40*, 641–650. [[CrossRef](#)] [[PubMed](#)]
25. Haer-Wigman, L.; van Zelst-Stams, W.A.; Pfundt, R.; van den Born, L.I.; Klaver, C.C.; Verheij, J.B.; Hoyng, C.B.; Breuning, M.H.; Boon, C.J.; Kievit, A.J.; et al. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur. J. Hum. Genet.* **2017**, *25*, 591–599. [[CrossRef](#)] [[PubMed](#)]
26. Zernant, J.; Lee, W.; Collison, F.T.; Fishman, G.A.; Sergeev, Y.V.; Schuerch, K.; Sparrow, J.R.; Tsang, S.H.; Allikmets, R. Frequent hypomorphic alleles account for a significant fraction of *ABCA4* disease and distinguish it from age-related macular degeneration. *J. Med. Genet.* **2017**, *54*, 404–412. [[CrossRef](#)] [[PubMed](#)]
27. Farrar, G.J.; Kenna, P.; Redmond, R.; Shiels, D.; McWilliam, P.; Humphries, M.M.; Sharp, E.M.; Jordan, S.; Kumar-Singh, R.; Humphries, P. Autosomal dominant retinitis pigmentosa: A mutation in codon 178 of the rhodopsin gene in two families of Celtic origin. *Genomics* **1991**, *11*, 1170–1171. [[CrossRef](#)]
28. Farrar, G.J.; Findlay, J.B.; Kumar-Singh, R.; Kenna, P.; Humphries, M.M.; Sharpe, E.; Humphries, P. Autosomal dominant retinitis pigmentosa: A novel mutation in the rhodopsin gene in the original 3q linked family. *Hum. Mol. Genet.* **1992**, *1*, 769–771. [[CrossRef](#)] [[PubMed](#)]
29. Al-Jandal, N.; Farrar, G.J.; Kiang, A.S.; Humphries, M.M.; Bannon, N.; Findlay, J.B.; Humphries, P.; Kenna, P.F. A novel mutation within the rhodopsin gene (Thr-94-Ile) causing autosomal dominant congenital stationary night blindness. *Hum. Mutat.* **1999**, *13*, 75–81. [[CrossRef](#)]

30. Lenassi, E.; Vincent, A.; Li, Z.; Saihan, Z.; Coffey, A.J.; Steele-Stallard, H.B.; Moore, A.T.; Steel, K.P.; Luxon, L.M.; Héon, E.; et al. A detailed clinical and molecular survey of subjects with nonsyndromic *USH2A* retinopathy reveals an allelic hierarchy of disease-causing variants. *Eur. J. Hum. Genet.* **2015**, *23*, 1318–1327. [[CrossRef](#)] [[PubMed](#)]
31. Pierce, E.A.; Quinn, T.; Meehan, T.; McGee, T.L.; Berson, E.L.; Dryja, T.P. Mutations in a gene encoding a new oxygen-regulated photoreceptor protein cause dominant retinitis pigmentosa. *Nat. Genet.* **1999**, *22*, 248–254. [[CrossRef](#)] [[PubMed](#)]
32. Khaliq, S.; Abid, A.; Ismail, M.; Hameed, A.; Mohyuddin, A.; Lall, P.; Aziz, A.; Anwar, K.; Mehdi, S.Q. Novel association of *RP1* gene mutations with autosomal recessive retinitis pigmentosa. *J. Med. Genet.* **2005**, *42*, 436–438. [[CrossRef](#)] [[PubMed](#)]
33. Zhang, X.; Chen, L.J.; Law, J.P.; Lai, T.Y.Y.; Chiang, S.W.Y.; Tam, P.O.S.; Chu, K.Y.; Wang, N.; Zhang, M.; Pang, C.P. Differential pattern of *RP1* mutations in retinitis pigmentosa. *Mol. Vis.* **2010**, *16*, 1353–1360. [[PubMed](#)]
34. Audo, I.; Bujakowska, K.M.; Léveillard, T.; Mohand-Saïd, S.; Lancelot, M.-E.; Germain, A.; Antonio, A.; Michiels, C.; Saraiva, J.-P.; Letexier, M.; et al. Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J. Rare Dis.* **2012**, *7*, 8. [[CrossRef](#)] [[PubMed](#)]
35. Castro-Sánchez, S.; Álvarez-Satta, M.; Cortón, M.; Guillén, E.; Ayuso, C.; Valverde, D. Exploring genotype-phenotype relationships in Bardet-Biedl syndrome families. *J. Med. Genet.* **2015**, *52*, 503–513. [[CrossRef](#)] [[PubMed](#)]
36. Cox, K.F.; Kerr, N.C.; Kedrov, M.; Nishimura, D.; Jennings, B.J.; Stone, E.M.; Sheffield, V.C.; Iannaccone, A. Phenotypic expression of Bardet-Biedl syndrome in patients homozygous for the common M390R mutation in the *BBS1* gene. *Vis. Res.* **2012**, *75*, 77–87. [[CrossRef](#)] [[PubMed](#)]
37. Moloney, T.P.; O'Hagan, S.; Lee, L. Ultrawide-field fundus photography of the first reported case of gyrate atrophy from Australia. *Clin. Ophthalmol.* **2014**, *8*, 1561–1563. [[CrossRef](#)] [[PubMed](#)]
38. Ramesh, V.; McClatchey, A.I.; Ramesh, N.; Benoit, L.A.; Berson, E.L.; Shih, V.E.; Gusella, J.F. Molecular basis of ornithine aminotransferase deficiency in B-6-responsive and -nonresponsive forms of gyrate atrophy. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 3777–3780. [[CrossRef](#)] [[PubMed](#)]
39. Heller, D.; Weiner, C.; Nasie, I.; Anikster, Y.; Landau, Y.; Koren, T.; Pokroy, R.; Abulafia, A.; Pras, E. Reversal of cystoid macular edema in gyrate atrophy patients. *Ophthalmic Genet.* **2017**, 1–6. [[CrossRef](#)] [[PubMed](#)]
40. Carrigan, M.; Duignan, E.; Humphries, P.; Palfi, A.; Kenna, P.F.; Farrar, G.J. A novel homozygous truncating *GNAT1* mutation implicated in retinal degeneration. *Br. J. Ophthalmol.* **2016**, *100*, 495–500. [[CrossRef](#)] [[PubMed](#)]
41. Megaw, R.D.; Soares, D.C.; Wright, A.F. RPGR: Its role in photoreceptor physiology, human disease, and future therapies. *Exp. Eye Res.* **2015**, *138*, 32–41. [[CrossRef](#)] [[PubMed](#)]
42. Li, J.; Tang, J.; Feng, Y.; Xu, M.; Chen, R.; Zou, X.; Sui, R.; Chang, E.Y.; Lewis, R.A.; Zhang, V.W.; et al. Improved Diagnosis of Inherited Retinal Dystrophies by High-Fidelity PCR of ORF15 followed by Next-Generation Sequencing. *J. Mol. Diagn.* **2016**, *18*, 817–824. [[CrossRef](#)] [[PubMed](#)]
43. Huang, X.-F.; Wu, J.; Lv, J.-N.; Zhang, X.; Jin, Z.-B. Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: A complementary approach to clinical genetic diagnostic testing. *Genet. Med.* **2015**, *17*, 307–311. [[CrossRef](#)] [[PubMed](#)]
44. Kabir, F.; Ullah, I.; Ali, S.; Gottsch, A.D.H.; Naeem, M.A.; Assir, M.Z.; Khan, S.N.; Akram, J.; Riazuddin, S.; Ayyagari, R.; et al. Loss of function mutations in *RP1* are responsible for retinitis pigmentosa in consanguineous familial cases. *Mol. Vis.* **2016**, *22*, 610–625. [[PubMed](#)]
45. Méndez-Vidal, C.; Bravo-Gil, N.; González-Del Pozo, M.; Vela-Boza, A.; Dopazo, J.; Borrego, S.; Antiñolo, G. Novel *RP1* mutations and a recurrent *BBS1* variant explain the co-existence of two distinct retinal phenotypes in the same pedigree. *BMC Genet.* **2014**, *15*, 143. [[CrossRef](#)] [[PubMed](#)]
46. Al-Rashed, M.; Abu Safieh, L.; Alkuraya, H.; Aldahmesh, M.A.; Alzahrani, J.; Diya, M.; Hashem, M.; Hardcastle, A.J.; Al-Hazzaa, S.A.F.; Alkuraya, F.S. *RP1* and retinitis pigmentosa: Report of novel mutations and insight into mutational mechanism. *Br. J. Ophthalmol.* **2012**, *96*, 1018–1022. [[CrossRef](#)] [[PubMed](#)]

47. El Shamieh, S.; Boulanger-Scemama, E.; Lancelot, M.-E.; Antonio, A.; Démontant, V.; Condroyer, C.; Letexier, M.; Saraiva, J.-P.; Mohand-Saïd, S.; Sahel, J.-A.; et al. Targeted next generation sequencing identifies novel mutations in RP1 as a relatively common cause of autosomal recessive rod-cone dystrophy. *Biomed. Res. Int.* **2015**, *2015*, 485624. [[CrossRef](#)] [[PubMed](#)]
48. Bax, N.M.; Sangermano, R.; Roosing, S.; Thiadens, A.A.H.J.; Hoefsloot, L.H.; van den Born, L.I.; Phan, M.; Klevering, B.J.; Westeneng-van Haaften, C.; Braun, T.A.; et al. Heterozygous deep-intronic variants and deletions in *ABCA4* in persons with retinal dystrophies and one exonic *ABCA4* variant. *Hum. Mutat.* **2015**, *36*, 43–47. [[CrossRef](#)] [[PubMed](#)]
49. Sangermano, R.; Bax, N.M.; Bauwens, M.; van den Born, L.I.; De Baere, E.; Garanto, A.; Collin, R.W.J.; Goercharn-Ramlal, A.S.A.; den Engelsman-van Dijk, A.H.A.; Rohrschneider, K.; et al. Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the *ABCA4* c.5461-10T→C Mutation in Stargardt Disease. *Ophthalmology* **2016**, *123*, 1375–1385. [[CrossRef](#)] [[PubMed](#)]
50. Cornelis, S.S.; Bax, N.M.; Zernant, J.; Allikmets, R.; Fritsche, L.G.; den Dunnen, J.T.; Ajmal, M.; Hoyng, C.B.; Cremers, F.P.M. In Silico Functional Meta-Analysis of 5962 *ABCA4* Variants in 3928 Retinal Dystrophy Cases. *Hum. Mutat.* **2017**, *38*, 400–408. [[CrossRef](#)] [[PubMed](#)]
51. De Castro-Miró, M.; Tonda, R.; Escudero-Ferruz, P.; Andrés, R.; Mayor-Lorenzo, A.; Castro, J.; Ciccioli, M.; Hidalgo, D.A.; Rodríguez-Ezcurra, J.J.; Farrando, J.; et al. Novel Candidate Genes and a Wide Spectrum of Structural and Point Mutations Responsible for Inherited Retinal Dystrophies Revealed by Exome Sequencing. *PLoS ONE* **2016**, *11*, e0168966. [[CrossRef](#)] [[PubMed](#)]
52. Bujakowska, K.M.; Fernandez-Godino, R.; Place, E.; Consugar, M.; Navarro-Gomez, D.; White, J.; Bedoukian, E.C.; Zhu, X.; Xie, H.M.; Gai, X.; et al. Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. *Genet. Med.* **2017**, *19*, 643–651. [[CrossRef](#)] [[PubMed](#)]
53. Bernardis, I.; Chiesi, L.; Tenedini, E.; Artuso, L.; Percesepe, A.; Artusi, V.; Simone, M.L.; Manfredini, R.; Camparini, M.; Rinaldi, C.; et al. Unravelling the Complexity of Inherited Retinal Dystrophies Molecular Testing: Added Value of Targeted Next-Generation Sequencing. *Biomed. Res. Int.* **2016**, *2016*, 6341870. [[CrossRef](#)] [[PubMed](#)]
54. Riera, M.; Navarro, R.; Ruiz-Nogales, S.; Méndez, P.; Burés-Jelstrup, A.; Corcóstequi, B.; Pomares, E. Whole exome sequencing using Ion Proton system enables reliable genetic diagnosis of inherited retinal dystrophies. *Sci. Rep.* **2017**, *7*, 42078. [[CrossRef](#)] [[PubMed](#)]
55. Tiwari, A.; Bahr, A.; Bähr, L.; Fleischhauer, J.; Zinkernagel, M.S.; Winkler, N.; Barthelmes, D.; Berger, L.; Gerth-Kahlert, C.; Neidhardt, J.; et al. Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. *Sci. Rep.* **2016**, *6*, 28755. [[CrossRef](#)] [[PubMed](#)]
56. Savoie, B.T.; Ferrone, P.J. Complicated Congenital Retinoschisis. *Retinal Cases Brief Rep.* **2017**, *11* (Suppl. 1), S202–S210. [[CrossRef](#)] [[PubMed](#)]
57. Dimopoulos, I.S.; Freund, P.R.; Knowles, J.A.; MacDonald, I.M. The Natural History of Full-Field Stimulus Threshold Decline in Choroideremia. *Retina* **2017**. [[CrossRef](#)] [[PubMed](#)]
58. Littink, K.W.; Pott, J.-W.R.; Collin, R.W.J.; Kroes, H.Y.; Verheij, J.B.G.M.; Blokland, E.A.W.; de Castro Miró, M.; Hoyng, C.B.; Klaver, C.C.W.; Koenekoop, R.K.; et al. A novel nonsense mutation in *CEP290* induces exon skipping and leads to a relatively mild retinal phenotype. *Investig. Ophthalmol. Vis. Sci.* **2010**, *51*, 3646–3652. [[CrossRef](#)] [[PubMed](#)]



Supplementary Materials

Target 5000: Target Capture Sequencing for Inherited Retinal Degenerations

Adrian Dockery ^{1,*}, Kirk Stephenson ², David Keegan ², Niamh Wynne ³, Giuliana Silvestri ^{4,5}, Peter Humphries ¹, Paul F. Kenna ^{1,3}, Matthew Carrigan ^{1,†} and G. Jane Farrar ^{1,†}

Table S1. Full list of genes captured in NGS panel.

A	<i>BBS9</i>	<i>CNGB3</i>	<i>GNAT1</i>	K	<i>NEK2</i>	<i>PEX1</i>	<i>RDH12</i>	T	<i>WFS1</i>
<i>ADIPOR1</i>	<i>BEST1</i>	<i>CNNM4</i>	<i>GNAT2</i>	<i>KCNJ13</i>	<i>NEUROD1</i>	<i>PEX2</i>	<i>RDH5</i>	<i>TEAD1</i>	<i>WFS1</i>
<i>ABCA4</i>	C	<i>COL11A1</i>	<i>GNB3</i>	<i>KCNV2</i>	<i>NMNAT1</i>	<i>PEX7</i>	<i>RGR</i>	<i>TIMM8A</i>	<i>WHRN</i>
<i>ABCC6</i>	<i>C12orf65</i>	<i>COL2A1</i>	<i>GNPTG</i>	<i>KIAA1549</i>	<i>NPHP1</i>	<i>PGK1</i>	<i>RGS9</i>	<i>TIMP3</i>	Z
<i>ABHD12</i>	<i>C1QTNF5</i>	<i>COL9A1</i>	<i>GPR179</i>	<i>KIF11</i>	<i>NPHP3</i>	<i>PHYH</i>	<i>RGS9BP</i>	<i>TMEM126A</i>	<i>ZNF408</i>
<i>ACBD5</i>	<i>C21orf2</i>	<i>CRB1</i>	<i>GRK1</i>	<i>KIZ</i>	<i>NPHP4</i>	<i>PITPNM3</i>	<i>RHO</i>	<i>TMEM216</i>	<i>ZNF423</i>
<i>ADAM9</i>	<i>C2orf71</i>	<i>CRX</i>	<i>GRM6</i>	<i>KLHL7</i>	<i>NR2E3</i>	<i>PLA2G5</i>	<i>RIMS1</i>	<i>TMEM237</i>	<i>ZNF513</i>
<i>ADAMTS18</i>	<i>C8orf37</i>	<i>CSPP1</i>	<i>GUCA1A</i>	L	<i>NR2F1</i>	<i>PLK4</i>	<i>RLBP1</i>	<i>TOPORS</i>	
<i>ADGRA3</i>	<i>CA4</i>	<i>CTBP2</i>	<i>GUCA1B</i>	<i>LAMA1</i>	<i>NRL</i>	<i>PNPLA6</i>	<i>ROM1</i>	<i>TREX1</i>	
<i>ADGRV1</i>	<i>CABP4</i>	<i>CTNNA1</i>	<i>GUCY2D</i>	<i>LCA5</i>	<i>NYX</i>	<i>POC1B</i>	<i>RP1</i>	<i>TRIM32</i>	
<i>ADIPOR1</i>	<i>CACNA1F</i>	<i>CYP4V2</i>	H	<i>LRAT</i>	O	<i>POMGNT1</i>	<i>RP1L1</i>	<i>TRNT1</i>	
<i>AGBL5</i>	<i>CACNA2D4</i>	D	<i>HARS</i>	<i>LRIT3</i>	<i>OAT</i>	<i>PRCD</i>	<i>RP2</i>	<i>TRPM1</i>	
<i>AHI1</i>	<i>CAPN5</i>	<i>DHDDS</i>	<i>HGSNAT</i>	<i>LRP5</i>	<i>OFD1</i>	<i>PRDM13</i>	<i>RP9</i>	<i>TSPAN12</i>	
<i>AIPL1</i>	<i>CC2D2A</i>	<i>DHX38</i>	<i>HK1</i>	<i>LZTFL1</i>	<i>OPA1</i>	<i>PROM1</i>	<i>RPE65</i>	<i>TTC8</i>	
<i>ALMS1</i>	<i>CDH23</i>	<i>DMD</i>	<i>HMCN1</i>	M	<i>OPA3</i>	<i>PRPF3</i>	<i>RPGR</i>	<i>TTL5</i>	
<i>ARL2BP</i>	<i>CDH3</i>	<i>DRAM2</i>	<i>HMX1</i>	<i>MAK</i>	<i>OPN1LW</i>	<i>PRPF31</i>	<i>RPGR</i>	<i>TTPA</i>	
<i>ARL3</i>	<i>CDHR1</i>	<i>DTHD1</i>	I	<i>MAPKAPK3</i>	<i>OPN1MW</i>	<i>PRPF4</i>	<i>RPGRIP1</i>	<i>TUB</i>	
<i>ARL6</i>	<i>CEP164</i>	E	<i>IDH3B</i>	<i>MERTK</i>	<i>OPN1SW</i>	<i>PRPF6</i>	<i>RPGRIP1L</i>	<i>TUBGCP4</i>	
<i>ASRGL1</i>	<i>CEP250</i>	<i>EFEMP1</i>	<i>IFT140</i>	<i>MFN2</i>	<i>OTX2</i>	<i>PRPF8</i>	<i>RS1</i>	<i>TUBGCP6</i>	
<i>ATF6</i>	<i>CEP290</i>	<i>ELOVL4</i>	<i>IFT172</i>	<i>MFRP</i>	P	<i>PRPH2</i>	<i>RTN4IP1</i>	<i>TULP1</i>	
<i>ATXN7</i>	<i>CERKL</i>	<i>EMC1</i>	<i>IFT27</i>	<i>MFS2D8</i>	<i>PANK2</i>	<i>PRPS1</i>	S	U	
B	<i>CFH</i>	<i>EXOSC2</i>	<i>IMPDH1</i>	<i>MIR204</i>	<i>PAX2</i>	R	<i>SAG</i>	<i>UNC119</i>	
<i>BBIP1</i>	<i>CHM</i>	<i>EYS</i>	<i>IMPG1</i>	<i>MKKS</i>	<i>PCDH15</i>	<i>RAB28</i>	<i>SDCCAG8</i>	<i>USH1C</i>	
<i>BBS1</i>	<i>CIB2</i>	F	<i>IMPG2</i>	<i>MKS1</i>	<i>PCYT1A</i>	<i>RAX2</i>	<i>SEMA4A</i>	<i>USH1G</i>	
<i>BBS10</i>	<i>CLN3</i>	<i>FAM161A</i>	<i>INPP5E</i>	<i>MTTP</i>	<i>PDE6A</i>	<i>RB1</i>	<i>SLC24A1</i>	<i>USH2A</i>	
<i>BBS12</i>	<i>CLRN1</i>	<i>FLVCR1</i>	<i>INVS</i>	<i>MVK</i>	<i>PDE6B</i>	<i>RBP3</i>	<i>SLC25A46</i>	V	
<i>BBS2</i>	<i>CLUAP1</i>	<i>FSCN2</i>	<i>IQCB1</i>	<i>MYO7A</i>	<i>PDE6C</i>	<i>RBP4</i>	<i>SLC7A14</i>	<i>VCAN</i>	
<i>BBS4</i>	<i>CNGA1</i>	<i>FZD4</i>	<i>ITM2B</i>	N	<i>PDE6G</i>	<i>RCBTB1</i>	<i>SNRNP200</i>	W	
<i>BBS5</i>	<i>CNGA3</i>	G	J	<i>NBAS</i>	<i>PDE6H</i>	<i>RD3</i>	<i>SPATA7</i>	<i>WDPCP</i>	
<i>BBS7</i>	<i>CNGB1</i>	<i>GDF6</i>	<i>JAG1</i>	<i>NDP</i>	<i>PDZD7</i>	<i>RDH11</i>	<i>SPP2</i>	<i>WDR19</i>	

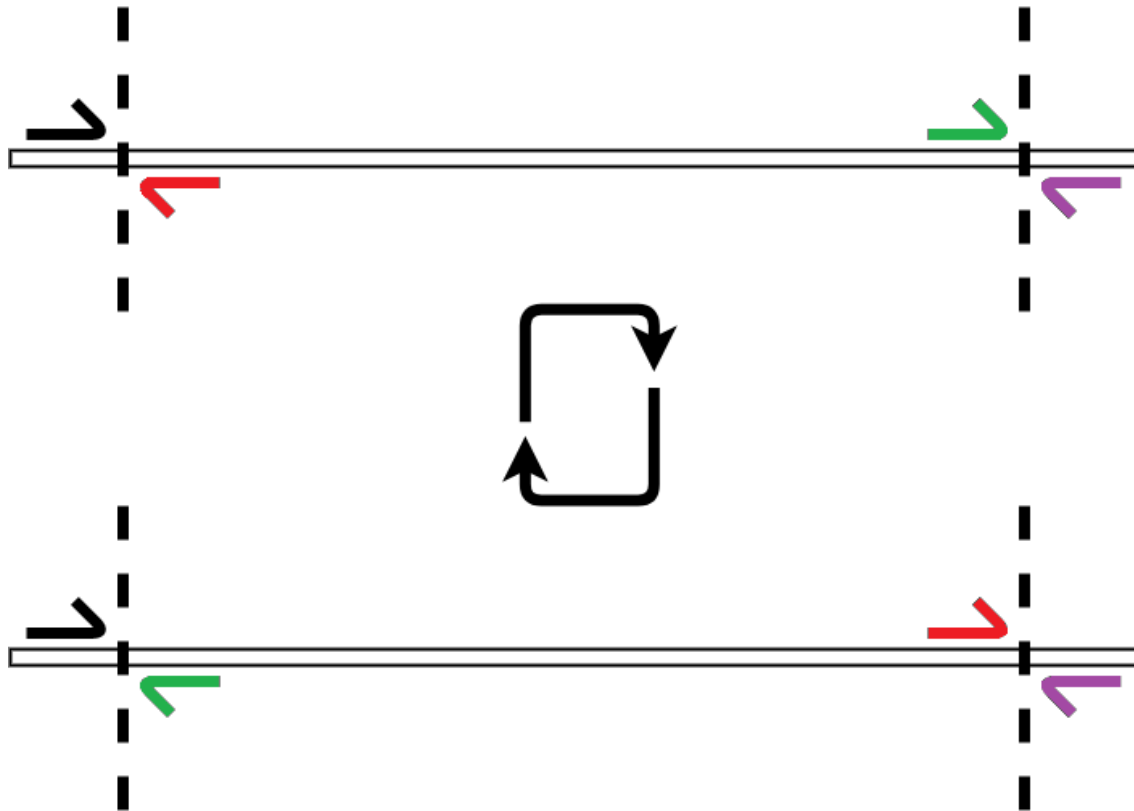


Figure S1. An illustration of the PCR strategy designed to detect a large homozygous inversion. Primers are indicated by colour; OAT-1 (**black**), OAT-2 (**red**), OAT-3 (**green**) and OAT-4 (**purple**). Broken lines indicate genomic breakpoints. Circular arrows indicate an inversion event. The top illustration depicts how primers would anneal around the breakpoints in a control sample. The bottom illustration shows how the internal primers (OAT-2 and OAT-3) are shuffled in an inversion event to form new primer pairings, detectable by PCR analysis.

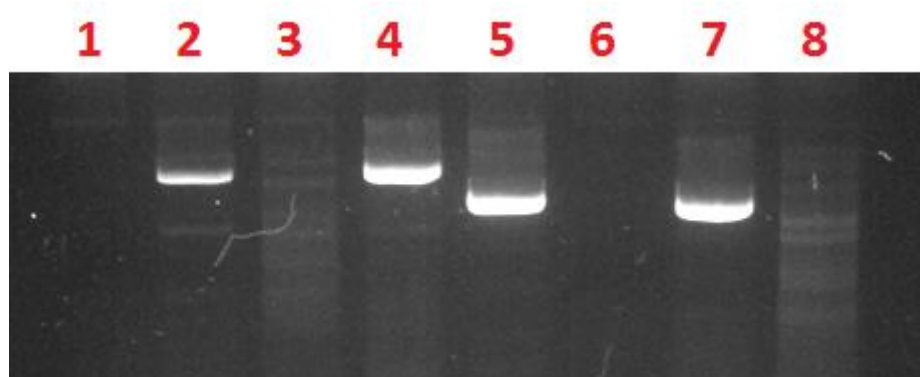


Figure S2. Confirmation of an *OAT* inversion using strategic PCR design. If the wildtype *OAT* sequence is present, the primer sets 1 + 2 and 3 + 4 can be used to generate a PCR product. If the inversion has occurred primer sets 1 + 3 and 2 + 4 can be used to generate of a PCR product. Unaffected patients are positive for products of a wildtype sequence. Patient H58 is positive only for inversion-specific products. Lane 1: Patient H58 Primers 1 and 2; Lane 2: Patient H58 Primers 3 and 4; Lane 3: Patient H58 Primers 1 and 3; Lane 4: Patient H58 Primers 2 and 4; Lane 5: Unaffected Patient Primers 1 and 2; Lane 6: Unaffected Patient Primers 3 and 4; Lane 7: Unaffected Patient Primers 1 and 3; Lane 8: Unaffected Patient Primers 2 and 4.

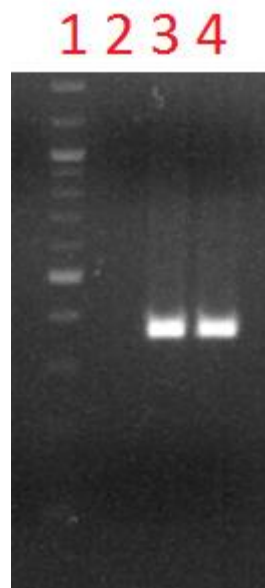


Figure S3. Gel confirmation of *USH2A* deletion. Primers were designed to target the region of deletion in affected patients C41 and E15. A product of the desired size (~300bp) would only be formed if the deletion was present. The sample from the unaffected sibling in the same pedigree, P36, did not form a product. A product was formed from the samples of C41 and E15. Lane 1: 100 bp Ladder; Lane 2: Unaffected Patient P36; Lane 3: Affected Patient C41; Lane 4: Affected Patient E15.

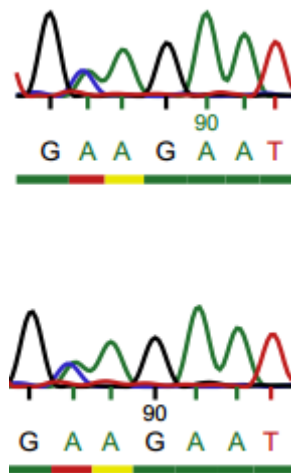


Figure S4. Sanger sequencing trace of *USH2A* mutation, p.Cys759Phe. The top trace is unaffected patient P36 and the bottom trace is affected patient E15. It is clear that both patients share the same heterozygous point mutation.

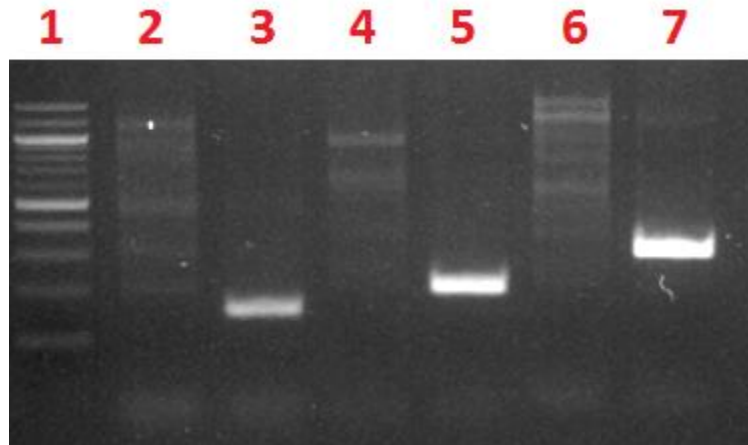


Figure S5. Analysis of *USH1C* deletion by use of PCR products. Primers were designed to amplify each exon believed to be deleted in Patient 1363. If the sequence for the exon was present a product would be formed. Exons 3 and 4 were designed as a single amplicon due to their proximity. The unaffected individual is positive for all of the *USH1C* exons. Patient 1363 is negative for all *USH1C* exons. Lane 1: 100 bp Ladder; Lane 2: Patient 1363 with primers for exon 1; Lane 3: Unaffected Patient with primers for exon 1; Lane 4: Patient 1363 with primers for exon 2; Lane 5: Unaffected Patient with primers for exon 2; Lane 6: Patient 1363 with primers for exons 3 + 4; Lane 7: Unaffected Patient with primers for exons 3 + 4.

INVITED REVIEW

Toward an elucidation of the molecular genetics of inherited retinal degenerations

G. Jane Farrar^{1,*}, Matthew Carrigan¹, Adrian Dockery¹,
Sophia Millington-Ward¹, Arpad Palfi¹, Naomi Chadderton¹,
Marian Humphries¹, Anna Sophia Kiang¹, Paul F. Kenna² and
Pete Humphries^{1,*}

¹Institute of Genetics, School of Genetics and Microbiology, University of Dublin, Trinity College, Dublin 2, Ireland and ²Research Foundation, Royal Victoria Eye and Ear Hospital, Dublin 2, Ireland

*To whom correspondence should be addressed at: The Ocular Genetics Unit, Institute of Genetics, Trinity College, University of Dublin, Dublin 2, Ireland.
Tel: +353 18962482; Email: gjfarrar@tcd.ie (G.J.F.); Email: pete.humphries@tcd.ie (P.H.)

Abstract

While individually classed as rare diseases, hereditary retinal degenerations (IRDs) are the major cause of registered visual handicap in the developed world. Given their hereditary nature, some degree of intergenic heterogeneity was expected, with genes segregating in autosomal dominant, recessive, X-linked recessive, and more rarely in digenic or mitochondrial modes. Today, it is recognized that IRDs, as a group, represent one of the most genetically diverse of hereditary conditions - at least 260 genes having been implicated, with 70 genes identified in the most common IRD, retinitis pigmentosa (RP). However, targeted sequencing studies of exons from known IRD genes have resulted in the identification of candidate mutations in only approximately 60% of IRD cases. Given recent advances in the development of gene-based medicines, characterization of IRD patient cohorts for known IRD genes and elucidation of the molecular pathologies of disease in those remaining unresolved cases has become an endeavor of the highest priority. Here, we provide an outline of progress in this area.

Genetic Heterogeneity in IRDs

Inherited retinal degenerations (IRDs) represent the most frequent cause of visual dysfunction in those of working age, such conditions therefore having a highly significant impact on quality of life and health economics. The more common IRDs include Retinitis Pigmentosa (RP), Choroideremia, Leber congenital amaurosis (LCA), Usher syndrome, Congenital stationary night blindness (CSNB), Vitelliform macular dystrophy, Stargardt Macular Dystrophy, Best disease, Retinoschisis, cone-rod dystrophy and myocilin-based hereditary forms of open angle glaucoma (see www.sph.uth.tmc.edu/retnet/ for details of disorders, causative genes and mutations). Typically, IRDs

result in gradual photoreceptor loss and compromised vision, often leading to registered blindness. Patterns of inheritance include autosomal recessive, autosomal dominant and X-linked, however, more rare mitochondrial and digenic forms of retinal dystrophies have been characterized.

Since the localization, through genetic linkage analysis, of the first IRD genes involved in X-linked recessive and autosomal dominant forms of RP (1–3), relentless progress has been made over three decades in defining an extensive portfolio of disease-causing genes, revealing an extreme level of genetic heterogeneity (4,5). Thus far, approximately 260 genes have been characterized, with more still to be identified (6–9; Tables 1–3). DNA

Received: April 27, 2017. Revised: May 5, 2017. Accepted: May 8, 2017

© The Author 2017. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Representative NGS studies of IRD patient populations

IRD patient numbers	Patient origin (predominant)	NGS technology	Pathogenic/likely pathogenic mutations (%)	Reference
100	Dutch	Target capture 111 genes	~ 50%	Neveling et al. 2012 (10)
20	France	Target capture 254 genes	57%	Audo et al. 2012 (11)
50	UK	Target capture 105 genes	50-50%	O'Sullivan et al. 2012 (12)
50	UK	Target capture 73 genes	53% (early onset IRD; 25% overall)	Shanks et al. 2013 (13)
192	US	Target capture (GEDI) 214 genes	51%	Consugar et al. 2015 (14)
105	eyeGENE® database	Target capture 195 genes	49.5%	Ge et al. 2015 (15)
82	N Ireland	Target capture 186 genes	60%	Zhao et al. 2015 (16)
88	Italy	Target capture 72 genes	59.1%	Bernardis et al. 2016 (17)
309	Ireland	Target capture 216 genes	57%	Carrigan et al. 2016 (18)
537	Worldwide	Target capture 105 genes	51%	Ellingford et al. 2016 (19)
47	Spain	Target capture 75 genes	57.4%	Perez-Carro et al. 2016 (20)
88	European	Target capture & WES	61%	Weisschuh et al. 2016 (21)
58	Swiss	WES	64%	Tiwari et al. 2016 (22)
33	Mixed	WES	55%	de Castro-Miró et al. 2016 (23)
59	Spain	WES	71.8%	Riera et al. 2017 (24)
266	Dutch	Target panel & WES	52%	Haer-Wigman et al. 2017 (25)

sequencing, and more recently, next-generation sequencing (NGS) technologies, in particular, now provide an opportunity to characterize the majority of the remaining genes (10–25).

IRD genes and encoded proteins are involved in a wide array of functions. It is of note that mutations in essentially all genes encoding components of the visual transduction and retinoid cycles have been implicated in disease etiology, as well as major structural components of photoreceptors (for extensive reviews of gene function see 26, 27; Tables 1–3). RP is the most common form of IRD, affecting approximately 1 in 3,000 people (4,5), the disease being characterized by progressive loss of rod photoreceptor function and their demise, followed by the death of the cone photoreceptors, initial symptoms of nyctalopia (night blindness) being followed usually by extensive loss of daytime vision as cones degenerate. Mutations within around 70 disease-causing genes have so far been implicated in RP (www.sph.uth.tmc.edu/retnet/; date last accessed April 25, 2017), while other genes still remain to be identified, indicating that this condition may, perhaps, be better regarded as a cluster of related conditions with similar clinical presentations.

Next-Generation Sequencing of IRD Populations

Future exome and whole genome sequencing, in conjunction with emerging methods to define the function of candidate mutations in regulatory and intronic regions and to better characterize copy number variations (CNVs), all of which have been implicated as causative of some forms of IRD, will undoubtedly progress our understanding of the molecular genetics of IRDs to a new level. The situation for IRDs is further complicated by the diversity of clinical presentations that can be caused by mutations even within a single IRD gene, as well as overlapping clinical phenotypes which may be the result of mutations in entirely distinct genes, making it impossible for a diagnosis to be given in the majority of cases on the basis of disease phenotype alone (17,28–30). By way of example, in a recent NGS study of an Irish IRD patient cohort, new disease phenotypes were associated

with the GNAT1 and SLC24A1 genes (28). In both cases a severe homozygous mutation in a known congenital stationary night blindness (CSNB) gene caused a late-onset form of RP involving photoreceptor loss (Fig. 1). Furthermore IRD phenotypes can be influenced by modifier loci (31–33). Given the limitations associated with the clinical diagnosis of specific forms of IRD, it is no surprise that the majority of patients with IRDs do not know the mutation(s) responsible for their condition. Recent advances in the development of gene-specific therapies for IRDs has made this issue all the more pressing, since this offers, for the first time, the prospect of modulating, halting or reversing the degeneration associated with some of these conditions (34), however, such therapies cannot, obviously, be administered without a precise knowledge of the underlying mutation to be treated.

High throughput NGS technologies offer an opportunity to rapidly characterize causative mutations in the growing number of IRD genes that have now been identified. These technologies have greatly accelerated DNA sequencing by providing a means of simultaneously sequencing many small fragments of DNA from different regions of the genome in a single reaction, and moreover by barcoding patient DNA to enable sequencing of samples from multiple patients in parallel. The short sequence reads from such fragments are then reconstructed by comparing the sequence to a reference genome; hg38 being currently the reference genome of choice for many such studies (hg38; GCA_000001405.15). NGS can be undertaken on the whole genome, or limited to expressed (coding) sequences (whole exome), or targeted to particular regions of the genome (employing target capture panel NGS); where genes of interest are, in essence, captured and sequenced. This form of sequencing has advanced at an exponential pace, with the cost per megabase of sequence halving regularly (35) and set to reduce further with the introduction of the new NovaSeq series from Illumina, among other innovations. IRDs have an ideal profile for NGS studies (14,18,36), representing a group of Mendelian conditions where many (to date approx. 260 genes), but not all of the causative genes, have been characterized. Given this scenario, NGS studies enable identification of known, or novel mutations in known genes and the stratification of IRD patient cohorts into

Table 2. Recessive Retinitis Pigmentosa genes, localizations and functions

Gene	Location	Function
ABCA4	1p22.1	ATP binding cassette subfamily A member 4; Photoreceptor-specific ATP-binding cassette (ABC) transporter with N-retinylidene-PE as a substrate.
AGBL5	2p23.3	ATP/GTP binding protein like 5; Metalloprotease that mediates protein dephosphorylation; can remove glutamate residues from both carboxyl termini and side chains of protein substrates.
ARL6	3q11.2	ADP ribosylation factor like GTPase 6; Involved in membrane protein trafficking at the base of the cilium.
ARL2BP	16q13.3	ADP ribosylation factor like GTPase 2 binding protein; Binds to ARL2.GTP with high affinity. Role in nuclear translocation, retention and transcriptional activity of STAT3.
BBS1	11q13.2	Bardet-Biedl syndrome 1 protein; Required for BBSome complex assembly, its ciliary localization and ciliogenesis.
BBS2	16q13	Bardet-Biedl syndrome 2 protein; Required for BBSome complex assembly, its ciliary localization and ciliogenesis.
BEST1	11q12.3	Bestrophin 1; Forms calcium-sensitive chloride channels; highly permeable to bicarbonate.
C2orf71	2p23.2	Chromosome 2 open reading frame 71; Photoreceptor cilia protein connecting inner and outer segments.
C8orf37	8q22.1	Chromosome 8 Open Reading Frame 37; Co-localizes with polyglutamylated tubulin at the base of the primary cilium in RPE cells.
CERKL	2q31.3	Ceramide kinase-like protein; has no detectable ceramide-kinase activity; protects cells from apoptosis in oxidative stress.
CLRN1	3q25.1	Clarín 1; A role in excitatory ribbon synapse junctions between hair cells and cochlear ganglion cells and possibly in analogous synapses in retina.
CNGA1	4p12	Cyclic nucleotide gated channel alpha 1; Rod cGMP-gated channel alpha subunit; Forms cGMP-gated cation channel for depolarization of rod photoreceptors.
CNGB1	16q21	Cyclic nucleotide gated channel beta 1; Rod cGMP-gated channel beta subunit; Forms cGMP-gated cation channel for depolarization of rod photoreceptors.
CRB1	1q31.3	Crumbs 1; Plays a role in photoreceptor morphogenesis in the retina. May maintain cell polarization and adhesion.
CYP4V2	4q35.1	Cytochrome P450 family 4 subfamily V member 2; Involved in fatty acid and steroid metabolism in the eye.
DHDDS	1p36.11	Dehydrodolichyl Diphosphate Synthase Subunit; forms the dehydrodolichyl diphosphate synthase (DDS) complex, an essential component of the dolichol monophosphate (Dol-P) biosynthetic machinery
DHX38	16q22.2	DEAH-Box Helicase 38; Unknown function; probable ATP-binding RNA helicase involved in pre-mRNA splicing.
EMC1	1p36.13	ER membrane protein complex subunit 1; Unknown function.
EYS	6q12	Eyes Shut Homolog (Drosophila); An extracellular matrix protein, one of the largest human genes at > 2mb. Possibly involved in stability of photoreceptor ciliary axonemes.
FAM161A	2p15	Family with sequence similarity 161 member A; Ciliary protein and a member of the Golgi-centrosomal interactome; involved in development of retinal progenitors.
GNAT1	3p21.31	G Protein Subunit Alpha Transducin 1; α subunit of the heterotrimeric G-protein transducing; stimulates the coupling of rhodopsin and cGMP-phosphodiesterase.
GPR125/ADGRA3	4p15.2	Adhesion G protein-coupled receptor A3/G protein-coupled receptor 125; Involved in signal transduction.
HGSNAT	8p11.21	Heparan-alpha-glucosaminide N-acetyltransferase; Lysosomal acetyltransferase that catalyzes transmembrane acetylation of the terminal glucosamine residues of heparan sulfate, prior to hydrolysis by α -N-acetylglucosaminidase.
IDH3B	20p13	Isocitrate Dehydrogenase 3 (NAD(+)) Beta; catalyzes the oxidation of isocitrate to alpha-ketoglutarate in the citric acid cycle.
IFT140	16p13.3	Intraflagellar transport 140; Involved in ciliogenesis, cilia maintenance and retrograde ciliary transport.
IFT172	2p23.3	Intraflagellar transport 172; Involved in formation and maintenance of cilia.
IMPG2	3q12.3	Interphotoreceptor matrix proteoglycan 2; Involved in organization of interphotoreceptor matrix.
KIAA1549	7q34	KIAA1549 protein; Unknown function.
KIZ	20p11.23	Kizuna centrosomal protein; Localizes to centrosomes, strengthening and stabilizing the pericentriolar region prior to spindle formation.
LRAT	4q32.1	Lecithin retinol acyltransferase; Transfers the acyl group from phosphatidylcholine to all-trans retinol, producing all-trans retinyl esters.
MAK	6p24.2	Male germ-cell associated kinase; Unknown function.
MERTK	2q13	C-mer proto-oncogene receptor tyrosine kinase; A regulator of rod outer segment fragment phagocytosis.
MVK	12q24.11	Mevalonate kinase; Involved in the synthesis of isopentenyl diphosphate from (R)-mevalonate.
NEK2	1q32.3	NIMA (never in mitosis gene A)-related kinase 2; Involved in regulation of centrosome disjunction.

Continued

Table 2. Continued

Gene	Location	Function
NEUROD1	2q31.3	Neuronal differentiation protein 1; Acts as a transcriptional activator.
NR2E3	15q23	Nuclear receptor subfamily 2 group E3; Transcription factor, that is an activator of rod development and repressor of cone development.
NRL	14q11.2	Neural retina leucine zipper; Transcription factor which regulates expression of rod-specific genes, involved in photoreceptor cell development and function
OFD1	Xp22.2	Oral-facial-digital syndrome 1 protein; Involved in biogenesis of the cilium and controlling centriole length during cell division.
PDE6A	5q32	Alpha subunit of cGMP phosphodiesterase; involved in visual transduction cycle.
PDE6B	4p16.3	Beta subunit of cGMP phosphodiesterase; involved in visual transduction cycle.
PDE6G	17q25.3	Gamma subunit of cGMP phosphodiesterase; involved in visual transduction cycle.
POMGNT1	1p34.1	Protein O-linked acetylglucosaminyltransferase 1 (beta 1,2-); Participates in O-mannosyl glycosylation.
PRCD	17q25.1	Progressive rod-cone degeneration protein; Unknown function.
PRIM1	12q13.3	DNA primase small subunit; Polymerase that synthesizes small RNA primers for Okazaki fragments made during discontinuous DNA replication.
RBP3	10q11.22	Retinol-binding protein 3 (IRBP); Shuttles 11-cis and all trans retinoids between the retinol isomerase in RPE and visual pigments in photoreceptor cells.
RGR	10q23.1	RPE-retinal G protein-coupled receptor; may be involved in regulating recycling of retinal.
RHO	3q22.1	Rhodopsin; Light-induced isomerization of 11-cis to all-trans retinal leading to G-protein activation and release of all-trans retinal.
RLBP1	15q26.1	Retinaldehyde-binding protein 1; Soluble retinoid carrier essential for function of photoreceptors.
RP1	8q11.23-q12.1	Oxygen-regulated protein 1; Involved in organization of photoreceptor outer segments.
RP1L1	8p23.1	Retinitis pigmentosa 1-like 1 protein; Required for differentiation of photoreceptors and organization of photoreceptor outer segments.
RP2	Xp11.3	Retinitis pigmentosa 2 (X-linked); Involved in trafficking between Golgi and the ciliary membrane.
RPE65	1p31.2	Retinoid isomerohydrolase; Involved in production of 11-cis retinal and visual pigment regeneration.
RPGR	Xp11.4	X-linked retinitis pigmentosa GTPase regulator; Involved in ciliogenesis
SAG	2q37.1	Rod photoreceptor-arrestin; Binds to photoactivated-phosphorylated rhodopsin, thereby apparently preventing the transducin-mediated activation of phosphodiesterase
SLC7A14	3q26.2	Solute carrier family 7 member 14; Unknown function.
SLC24A1	15q22.31	Solute carrier family 24 (sodium/potassium/calcium exchanger) member 1; Critical component of visual transduction, controlling the calcium concentration of outer segments.
SPATA7	14q31.3	Spermatogenesis associated protein 7; Required for recruitment and localization of RPGRIP1 to the photoreceptor cilium.
TRNT1	3p26.2	CCA adding tRNA nucleotidyl transferase 1; Adds and repairs CCA sequences necessary for attachment of amino acids to the 3' terminus of tRNA molecules.
TTC8	14q31.3	Tetratricopeptide repeat protein 8; Subunit of the BBSome complex required for proper cilia functions.
TULP1	6p21.31	Tubby-like protein 1; Required for normal development of photoreceptor synapses.
USH2A	1q41	Usherin; Involved in photoreceptor cell maintenance.
ZNF408	11p11.2	Zinc finger protein 408; Unknown function.
ZNF513	2p23.3	Zinc finger protein 513; Transcriptional regulator involved in retinal development and maintenance.

those whose genetic pathogenesis has been resolved, and those for whom new genes, or variants in regulatory sequences, splice variants or CNVs associated with disease genes, may be causative of disease and therefore additional analyses deploying whole exome or whole genome sequencing would be appropriate.

A variety of NGS studies, many employing a target panel-based NGS approach focused on sequencing the exons of known IRD genes, have been undertaken (10–25; Table 1). These studies support the view that IRDs represent an exemplar group of disorders for the application of panel-based NGS or WES as effective tools for detection of causative mutations. From such studies, some common patterns in findings have been observed, as have some unique findings, which at present, remain specific to individual studies. Of note, using target-capture based NGS, approximately 50–60% of IRD patients were found to carry disease causing or likely disease causing mutations in many studies. Given additional family studies indicating segregation of

the disease, and replication of the findings in a certified diagnostic laboratory, these patients would be deemed to be categorised as ‘resolved’ in terms of the genetic etiology of the retinal pathology. Among recent examples of capture-panel NGS for IRDs, is a study of 537 IRD patients in which the variants identified were deemed to account, or likely account for the disease in 51% of cases and a recent study of Italian IRD patients, in which 59% of cases were resolved, similar to the identification levels obtained in many other NGS studies (17,19; Tables 1–3). Detection rates were significantly affected by the range of conditions under study, the number of genes included in the capture panel, as well as whether the rate was corrected based on previous screening of the same population, making precise comparisons of detection rates between studies challenging. Nonetheless, there is a consistent finding that between 25–50% of cases were not solved by targeted sequencing or WES. It is of interest that studies employing WES, rather than capture-panel NGS, have resulted in the identification of the underlying

Table 3. Autosomal Dominant Retinitis Pigmentosa genes, localizations & function

Gene	Location	Function
ARL3	10q24.32	ADP-ribosylation factor-like GTPase 3; Required for cytokinesis and cilia signaling.
BEST1	11q12.3	Bestrophin-1; Forms calcium-sensitive chloride channels.
CA4	17q23.1	Carbonic anhydrase IV; Involved in acid overload removal from retina and acid release in the choriocapillaris.
CRX	19q13.33	Cone-rod homeobox protein; Transcription factor that binds and transactivates the 5'-TAATC[CA]-3' sequence found upstream of photoreceptor-specific genes, including opsin genes.
FSCN2	17q25.3	Fascin-2; Acts as an actin bundling protein.
GUCA1B	6p21.1	Guanylyl cyclase-activating protein 2; Stimulates guanylyl cyclase 1 and 2 in response to changing Ca ²⁺ concentrations.
HK1	10q22.1	Hexokinase-1; Catalyses D-hexose + ATP to D-hexose 6-phosphate + ADP.
IMPDH1	7q32.1	Inosine-5'-monophosphate dehydrogenase 1; Involved in rate-limiting step in de novo synthesis of guanine nucleotides.
KLHL7	7p15.3	Kelch-like protein 7; Substrate-specific adapter of a BCR (BTB-CUL3-RBX1) E3 ubiquitin ligase complex which mediates ubiquitination and degradation of substrate proteins.
NR2E3	15q23	Nuclear receptor subfamily 2 group E3; Transcriptional factor that is an activator of rod development and repressor of cone development.
NRL	14q11.2	Neural retina luciferase zipper; Transcription factor which regulates expression of rod-specific genes.
PRPF3	1q21.2	Pre-mRNA-processing-splicing factor 3; A small nuclear ribonucleoprotein, required for the assembly of the spliceosome U4/U5/U6 tri-snRNP complex.
PRPF4	9q32	Pre-mRNA-processing-splicing factor 4; A small nuclear ribonucleoprotein, required for the assembly of the spliceosome U4/U5/U6 tri-snRNP complex.
PRPF8	17p13.3	Pre-mRNA-processing-splicing factor 8; Acts as a scaffold aiding assembly of spliceosomal proteins.
PRPF31	19q13.42	Pre-mRNA-processing-splicing factor 31; A small nuclear ribonucleoprotein, required for the assembly of the spliceosome U4/U5/U6 tri-snRNP complex.
PRPH2	6p21.1	Peripherin-2; Involved in photoreceptor disk morphogenesis.
RDH12	14q24.1	Retinol dehydrogenase; An NADPH-dependent retinal reductase with high activity toward 9-cis and all-trans-retinol.
RHO	3q22.1	Rhodopsin; Light-induced isomerization of 11-cis to all-trans retinal leading to G-protein activation and release of all-trans retinal.
ROM1	11q12.3	Rod outer segment membrane protein; Essential for disk morphogenesis.
RP1	8q11.23	Oxygen-regulated protein 1; Required for the differentiation of photoreceptor cells, Involved in organization of rod and cone photoreceptor outer segments.
RP9	7p14.3	Retinitis pigmentosa 9 protein; possibly a target protein for the PIM1 kinase.
RPE65	1p31.3	Retinoid isomerohydrolase; Involved in production of 11-cis retinal and visual pigment regeneration.
SEMA4A	1q22	Semaphorin-4A; Promotes axon growth cone collapse.
SNRNP200	2q11.2	U5 small nuclear ribonucleoprotein 200 kDa helicase; Involved in spliceosome assembly, activation and disassembly.
SPP2	2q37.1	Secreted phosphoprotein 2; Unknown function.
TOPORS	9p21.1	E3 ubiquitin-protein ligase Topors; Functions as an E3 ubiquitin-protein ligase and as an E3 SUMO1-protein ligase.

genetic defect in similar levels of IRD patients ~60% (22), although in a recent study of Spanish IRD patients, WES resulted in successful identification of pathogenic variants in approximately 70% of patients (24). Levels of mutation detection achieved for IRDs that are less genetically heterogeneous, such as choroideremia, involving solely the CHM gene, and Stargardt disease, predominantly caused by ABCA4 mutations, were typically significantly higher (18; Fig. 2). However, even within such comparatively homogenous forms of IRD, the level of intragenic mutational heterogeneity present is still being characterised. For example, in a recent study in which 148 pathogenic or likely pathogenic mutations in the ABCA4 gene were identified, about a third of these (n = 48) represented new Stargardt associated disease alleles (37).

Deciphering the Genetic Pathogenesis of Unresolved IRDS

It is evident from this body of research that currently the genetic pathogenesis of IRDs remains unresolved in ~40% of cases

(10–25; Tables 1–3). IRD mutation detection may remain elusive for a variety of reasons, possibly due to inadequate sequence coverage, inappropriate filtering of data (and loss of the pathogenic variant), variants in intronic sequences which affect splicing (such sequences do not form part of most capture panels), structural variants such as inversions, duplications and so on which can be difficult to resolve with NGS, regulatory variants or new, as yet uncharacterized, IRD genes and which therefore were not included in capture panels, among many other causes. However from some NGS studies there are indicators that support the hypothesis that, within the known IRD genes included in many target capture panels, additional significant levels of pathological genetic variants are present but as yet remain undetected. For example, it has been found that there can be a preponderance of patients with one causative mutation in a recessive IRD gene; such over-representation of heterozygotes compared to a control population has been observed in a number of NGS studies, including our own study of the Irish IRD population, for example, for the ABCA4 gene causative of Stargardt

disease. The majority of clinically defined Irish Stargardt cases who did not have two ABCA4 mutations in the coding sequence, had a single ABCA4 mutation and this was highly significantly enriched compared to the frequency of ABCA4 mutant heterozygotes in a control Irish population (Carrigan M, unpublished data). Recent studies have focused on revealing the genetic variants underlying such observations. Various tools have been developed to detect copy number variations (CNVs) in NGS datasets and have been employed to successfully detect CNVs in IRD patient cohorts (38–40). Indeed, CNVs were identified in a recent WES study in approx. 10% of the 60 IRD patients assessed by exon coverage data analysis and confirmed by PCR (39). Alternative methods for CNV detection, that previously have been employed extensively in diagnostic laboratories for disorders other than IRDs, include quantitative PCR (qPCR), multiplex ligation-dependent probe amplification (MLPA) and comparative genomic hybridization. Using the latter, Van Cauwenbergh and colleagues developed a custom microarray (arrEYE) with coding and noncoding sequence from 166 known and candidate IRD genes and 196 noncoding RNAs for CNV detection in IRD patients (40); the CNV detection rate obtained with arrEYE from a first study of 57 IRD patients was 3.5%. In a recent study, 18% of unresolved cases ($n = 28$) were resolved by CNV mapping (38). It is clear that CNVs may represent a significant contributor to the unresolved cases of IRDs and that methods of CNV analysis, both computational and experimental, should, where possible be included in future IRD studies, which has not always been the case thus far.

In addition to CNVs, a proportion of the remaining unresolved IRD cases may be caused by variants that affect RNA splicing and thereby contribute to disease. Mutations that may have an impact on pre-mRNA splicing can be predicted using various *in silico* tools such as Human Splicing Finder (www.umd.be/HSF/; date last accessed April 25, 2017) among others, and can be confirmed by transcript analysis. RNA sequencing technologies are greatly enabling high throughput transcriptomics and undoubtedly will be applied to a greater extent to explore splice variants in IRDs in the future. A transcriptomic approach to assess the effects of IRD mutations on splicing may in principle be readily adopted for ubiquitously expressed IRD genes, and for tissue specific IRD genes, using iPS cells differentiated into appropriate cell lineages as per the example above. One approach to identify intronic variants with effects on RNA splicing in IRD patients involves deep intronic sequencing. For example, in a recent study of the USH2A gene, the most frequent cause of Usher syndrome type II (USH2) involving RP and sensorineural hearing loss, an analysis of the whole 800kb of the USH2A gene was prompted by the identification of patients with an USH2 phenotype and a single exonic mutation in the USH2A gene. Deep sequencing of intronic sequences enabled the identification of candidate splice mutations and subsequent functional validation of these mutations using reporter minigene assays leading to the resolution of 3 out of 5 USH2 patients with a single exonic mutation (41). In another study, the functional effect of a commonly observed mutation in the ABCA4 gene (the c.5461-10T→C variant) causative of Stargardt disease was analysed using patient-derived fibroblasts reprogrammed into induced pluripotent stem (iPS) cells and then differentiated into photoreceptor progenitor cells. This ABCA4 variant was found to induce skipping of exon 39 or exon 39 and 40 in the mature transcript again using a minigene assay (42). Recently homozygosity mapping and whole genome sequencing revealed a variant deep in intron 18 of the PROM1 gene causative of a recessive cone-rod dystrophy, resulting in the inclusion of a pseudo-exon in the mutant transcript, which was functionally validated

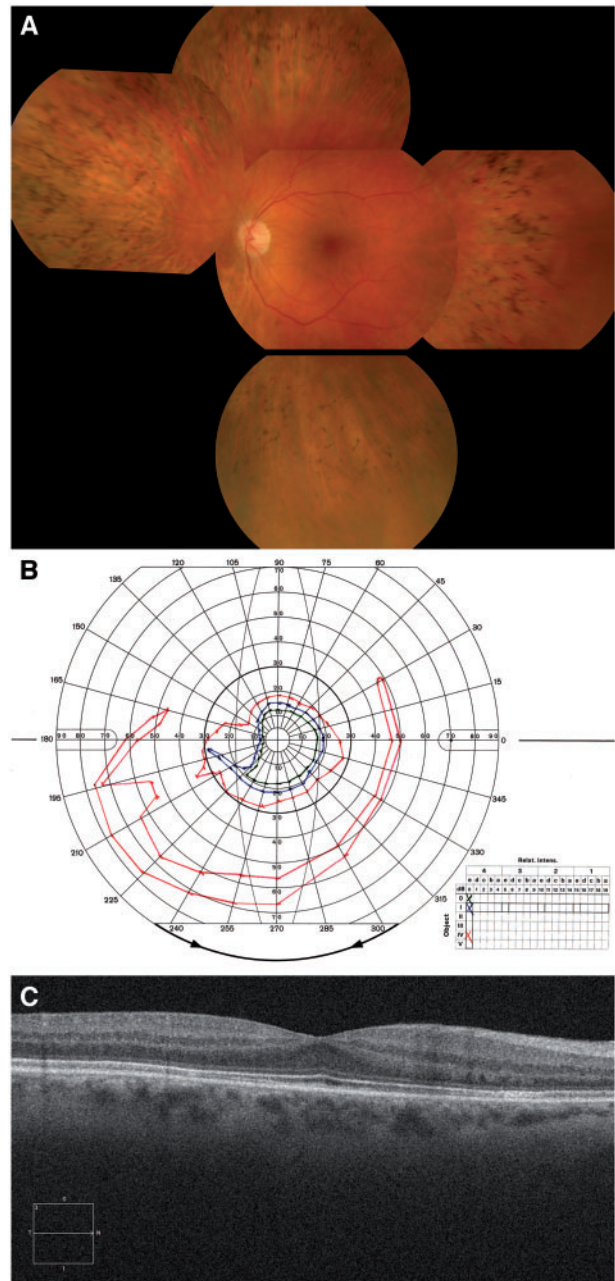


Figure 1. SLC24A1 causative of an autosomal recessive retinal degeneration. **Panel A:** Fundus photograph of the right eye of the proband with an SLC24A1 mutation indicating mild optic disc pallor, moderate arteriolar attenuation and quite extensive thinning of the retinal pigment epithelium in the retinal mid-periphery with abundant bone spicule intra-retinal pigment deposits, typically seen in patients with Retinitis Pigmentosa. **Panel B:** Goldmann kinetic perimetry of the left eye. Marked concentric constriction is evident, even to the large IV4e target (red), with preservation of an inferior island of field. This pattern of visual field loss is very typical of that seen in Retinitis Pigmentosa. **Panel C:** Optical Coherence Tomography (OCT) image of the left macula showing preservation of retinal structures, consistent with the normal appearance of the macula in Panel A and the patient's good best-corrected visual acuity of 6/6.

again using a minigene assay (43). Detection of pathogenic IRD intronic variants will elude many of the current capture panel NGS strategies for IRDs which are focused solely on exons and hence the extension of screening studies in the future to include intronic sequences will aid in elucidating what proportion of IRD cases involve aberrant splicing.

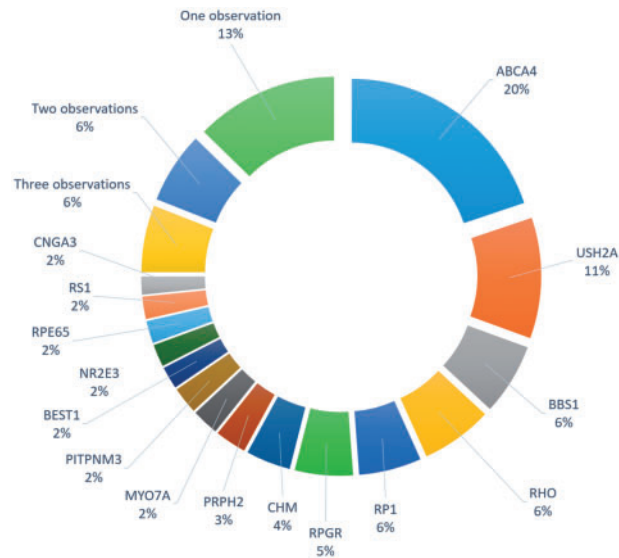


Figure 2. Frequencies with which different genes were found to harbour disease-causing mutations in the Target 5000 study in Ireland (Carrigan et al. 2016), counted as number of independent pedigrees. Genes that were observed less than four times, could not be cleanly visually represented, and so are listed below.

Three observations: ADGRV1, CERKL, CRX, EYS, KLHL7

Two observations: CNGB3, CRB1, NRL, PROM1, PRPF8, RP2, SNRNP200, TRPM1

One observation: ABCC6, ABHD12, APL1, BBS10, BBS9, C2ORF71, CDH23, CLRN1, CNGA1, ELOVL4, EMC1, FZD4, GNAT1, GUCY2D, HK1, HMCN1, IFT140, IMPDH1, IQCB1, LCA5, MFN2, MFRP, NMNAT1, NYX, OAT, PDE6A, PRPF31, RDH5, SDCCAG8, SLC24A1.

In addition to CNVs and splice variants as a source of as yet ‘unresolved’ pathogenic IRD mutations, variants in regulatory sequences, such as, promoter sequences or miRNAs, and their associated target sites, may also be implicated in some forms of IRD. Indeed a mutation in the seed region of microRNA-204 has been found to segregate in a family with autosomal dominantly inherited retinal dystrophy and bilateral coloboma (44), a 5'UTR sequence in the NMNAT1 gene has been implicated in a form of LCA (45) and recently a single base mutation in the promoter region of the CHM gene has been implicated as causative of choroideremia (46), all highlighting the potential role of regulatory mutations in some forms of IRD. The implementation of more extensive NGS strategies in the future will aid in characterising such IRD regulatory mutations.

In conclusion, while significant advances in high throughput genomic and transcriptomic NGS has greatly facilitated an elucidation of much of the genetic architecture of IRD patient cohorts, substantial levels of unresolved IRD cases still remain. Recently it has become evident that a significant proportion of these will be accounted for by CNVs, splicing defects and regulatory variants. It still remains to be established within these unresolved cases, what proportion will be caused by new, as yet uncharacterised, retinal disease genes.

A Diagnostic Imperative Driven by Developments in Gene Therapy

There is an obvious rationale for pursuing such studies to their conclusion, in that gene therapies for a growing number of ocular disorders are now in clinical trial (47–51). To date,

approximately 300 IRD patients have been treated with ocular gene therapies (www.clinicaltrials.gov). In some of these trials, a continued retinal degeneration has been observed (52) and the parameters determining this still need to be elucidated. It is of note however, that an adeno-associated virus (AAV)-RPE65 therapy (voretigene neparvovec) has successfully progressed through to Phase III clinical trial (CHOP/Spark Therapeutics; www.sparktx.com), as has an AAV-ND4 therapy for Leber hereditary optic neuropathy (LHON) (51). Data from many ocular gene therapy trials support the view that intraocular delivery of AAV is well tolerated in the human eye and hence represents a safe platform for gene delivery (48–51). A large number of gene therapies employing AAV vectors are also at the stage of preclinical evaluation in animal IRD models, in which benefit has been demonstrated (51,53–57). Thus far many gene therapies have been directed towards recessive forms of IRD, although strategies such as suppression and replacement or genome editing are being considered to address the 30% of IRDs that are dominantly inherited (53,58). The efficacy of the therapeutic approaches will be determined in part by features specific to individual IRDs. For example, the target cell type within the retina, whether photoreceptors, RPE cells or retinal ganglion cells (51,53–57), among others, will influence vector choice and route of administration. Strategies such as directed evolution are also expediting the generation of AAV serotypes with a predilection for specific retinal cell types (59,60). Additionally, the severity of disease and whether retention of a relatively intact target cell population is maintained over years, thereby providing a significant timeframe for therapeutic intervention, will also greatly influence therapeutic efficacy. For severe degenerative IRDs, where few or no photoreceptors remain, optogenetic therapies may provide an alternative option by enabling remaining cells to become photosensitive (61,62). While AAV has shown substantial promise for a number of ocular indications, preclinical and clinical studies have also been undertaken with a variety of other non-viral and viral vectors, for example, nanoparticles and lentiviral vectors amongst others (63,64). Moreover, methodologies for systemic delivery of potentially therapeutic compounds into the retina by modulation of permeability at the inner blood-retina barrier have been extensively tested in animal systems and have employed AAV vectors (65). Given the rapid pace at which therapies are being developed, it is imperative to establish the genetic architecture of IRDs in different national and ethnic groups, in principle thereby facilitating participation of patients in future human clinical trials or access in the future to marketed gene therapies. In parallel with genotyping of IRDs, it is vital that patients are clinically profiled over time, as this will greatly augment our understanding of the natural histories of these ocular disorders and genotype-phenotype correlations. Knowledge regarding the natural history of a disorder is of direct value to patients, and moreover, will in principle facilitate future patient participation in clinical trials and will aid in informing the choice of appropriate primary and secondary endpoints for clinical trial design. Therefore as such, identification of the disease-causing mutations is of immediate clinical relevance to the entire IRD patient population.

Conflict of Interest statement. NC, GJF, PH, PK, AP & SMW are shareholders in Spark Therapeutics.

Funding

Vision research at the Ocular Genetics Unit at TCD is supported by awards from Fighting Blindness Ireland (FB Irl), the Health

Research Board of Ireland (HRA_POR-2013-376; HRA_POR-2015-1140), the Medical Research Charities Group (MRCG-2012-4; MRCG-2013-8; MRCG-2016-14), the Irish Research Council (IRC), Science Foundation Ireland (SFI 11/PI/1080) and the European Research Council (ERC-2012-AdG 322656-Oculus). Funding to pay the Open Access publication charges for this article was provided by Science Foundation Ireland.

References

- Bhattacharya, S.S., Wright, A.F., Clayton, J.F., Price, W.H., Phillips, C.I., McKeown, C.M., Jay, M., Bird, A.C., Pearson, P.L., Southern, E.M. et al. (1984) Close genetic linkage between X-linked retinitis pigmentosa and a restriction fragment length polymorphism identified by recombinant DNA probe L1.28. *Nature*, **309**, 253–255.
- McWilliam, P., Farrar, G.J., Kenna, P.F., Bradley, D.G., Humphries, M.M., Sharp, E.M., McConnell, D.J., Lawler, M., Sheils, D., Ryan, C. and Humphries, P. (1989) *Genomics*, **3**, 619–622.
- Farrar, G.J., Kenna, P.F., 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*, **354**, 478–480.
- Ferrari, S., Di Iorio, E., Barbaro, V., Ponzin, D., Sorrentino, F.S. and Parmeggiani, F. (2011) Retinitis pigmentosa: genes and disease mechanisms. *Curr. Genomics*, **12**, 238–249.
- Chiang, J.P. and Trzuppek, K. (2015) The current status of molecular diagnosis of inherited retinal dystrophies. *Curr. Opin. Ophthalmol.*, **26**, 346–351.
- Arno, G., Agrawal, S.A., Eblimit, A., Bellingham, J., Xu, M., Wang, F., Chakarova, C., Parfitt, D.A., Lane, A., Burgoyne, T. et al. (2016) Mutations in REEP6 Cause Autosomal-Recessive Retinitis Pigmentosa. *Am. J. Hum. Genet.*, **99**, 1305–1315.
- Corton, M., Avila-Fernández, A., Campello, L., Sánchez, M., Benavides, B., López-Molina, M.I., Fernández-Sánchez, L., Sánchez-Alcudia, R., da Silva, L.R., Reyes, N. et al. (2016) Identification of the Photoreceptor Transcriptional Co-Repressor SAMD11 as Novel Cause of Autosomal Recessive Retinitis Pigmentosa. *Sci. Rep.*, **6**, 35370.
- Pierrache, L.H.M., Kimchi, A., Ratnapriya, R., Roberts, L., Astuti, G.D.N., Obolensky, A., Beryozkin, A., Tjon-Fo-Sang, M.J.H., Schuil, J., Klaver, C.C.W. et al. (2017) Whole-exome sequencing identifies biallelic IDH3A variants as a cause of retinitis pigmentosa accompanied by pseudocoloboma. *Ophthalmology*, doi: 10.1016/j.ophtha.2017.03.010. [Epub ahead of print]
- Xu, M., Xie, Y.A., Abouzeid, H., Gordon, C.T., Fiorentino, A., Sun, Z., Lehman, A., Osman, I.S., Dharmat, R., Riveiro-Alvarez, R. et al. (2017) Mutations in the spliceosome component CWC27 cause retinal degeneration with or without additional developmental anomalies. *Am. J. Hum. Genet.*, **100**, 592–604.
- Neveling, K., Collin, R.W., Gilissen, C., van Huet, R.A., Visser, L., Kwint, M.P., Gijzen, S.J., Zonneveld, M.N., Wiskamp, N., de Ligt, J. et al. (2012) Next-generation genetic testing for retinitis pigmentosa. *Hum. Mutat.*, **33**, 963–972.
- Audo, I., Bujakowska, K.M., Léveillard, T., Mohand-Saïd, S., Lancelot, M.E., Germain, A., Antonio, A., Michiels, C., Saraiva, J.P., Letexier, M. et al. (2012) Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet. J. Rare Dis.*, **7**, 8.
- O'Sullivan, J., Mullaney, B.G., Bhaskar, S.S., Dickerson, J.E., Hall, G., O'Grady, A., Webster, A., Ramsden, S.C. and Black, G.C. (2012) A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J. Med. Genet.*, **49**, 322–326.
- Shanks, M.E., Downes, S.M., Copley, R.R., Lise, S., Broxholme, J., Hudspith, K.A., Kwasniewska, A., Davies, W.I., Hankins, M.W., Packham, E.R. et al. (2013) Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease. *Eur. J. Hum. Genet.*, **21**, 274–280.
- 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. et al. (2015) Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet. Med.*, **17**, 253–261.
- Ge, Z., Bowles, K., Goetz, K., Scholl, H.P., Wang, F., Wang, X., Xu, S., Wang, K., Wang, H. and Chen, R. (2015) NGS-based Molecular diagnosis of 105 eyeGENE® probands with Retinitis Pigmentosa. *Sci. Rep.*, **5**, 18287.
- Zhao, L., Wang, F., Wang, H., Li, Y., Alexander, S., Wang, K., Willoughby, C.E., Zaneveld, J.E., Jiang, L., Soens, Z.T. et al. (2015) Next-generation sequencing-based molecular diagnosis of 82 retinitis pigmentosa probands from Northern Ireland. *Hum. Genet.*, **134**, 217–230.
- Bernardis, I., Chiesi, L., Tenedini, E., Artuso, L., Percesepe, A., Artusi, V., Simone, M.L., Manfredini, R., Camparini, M., Rinaldi, C. et al. (2016) Unravelling the complexity of inherited retinal dystrophies molecular testing: added value of targeted next-generation sequencing. *Biomed. Res. Int.*, **2016**, 6341870.
- Carrigan, M., Duignan, E., Malone, C., Stephenson, K., Saad, T., McDermott, C., Green, A., Keegan, D., Humphries, P., Kenna, P.F. and Farrar, G.J. (2016) Panel-based population next-generation sequencing for inherited retinal degenerations. *Sci. Rep.*, **6**, 33248.
- 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. et al. Molecular findings from 537 individuals with inherited retinal disease. *J. Med. Genet.*, doi: 10.1136/jmedgenet-2016-103837. [Epub ahead of print]
- 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. et al. (2016) Panel-based NGS reveals novel pathogenic mutations in autosomal recessive retinitis pigmentosa. *Sci. Rep.*, **6**, 19531.
- Weisschuh, N., Mayer, A.K., Strom, T.M., Kohl, S., Glöckle, N., Schubach, M., Andreasson, S., Bernd, A., Birch, D.G., Hamel, C.P. et al. (2016) Mutation detection in patients with retinal dystrophies using targeted next generation sequencing. *PLoS One*, **11**, e0145951.
- Tiwari, A., Bahr, A., Bähr, L., Fleischhauer, J., Zinkernagel, M.S., Winkler, N., Barthelmes, D., Berger, L., Gerth-Kahlert, C., Neidhardt, J. et al. (2016) Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. *Sci. Rep.*, **6**, 28755.
- de Castro-Miró, M., Tonda, R., Escudero-Ferruz, P., Andrés, R., Mayor-Lorenzo, A., Castro, J., Ciccioli, M., Hidalgo, D.A., Rodríguez-Ezcurra, J.J., Farrando, J. et al. (2016) Novel candidate genes and a wide spectrum of structural and point mutations responsible for inherited retinal dystrophies revealed by exome sequencing. *PLoS One*, **11**, e0168966.

24. Riera, M., Navarro, R., Ruiz-Nogales, S., Méndez, P., Burés-Jelstrup, A., Corcóstegui, B. and Pomares, E. (2017) Whole exome sequencing using Ion Proton system enables reliable genetic diagnosis of inherited retinal dystrophies. *Sci. Rep.*, **7**, 42078.
25. Haer-Wigman, L., van Zelst-Stams, W.A., Pfundt, R., van den Born, L.I., Klaver, C.C., Verheij, J.B., Hoyng, C.B., Breuning, M.H., Boon, C.J., Kievit, A.J. et al. (2017) Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur. J. Hum. Genet.*, **25**, 591–599.
26. Humphries, P., Humphries, M.M., Tam, L.C.S., Farrar, G.J., Kenna, P.F., Campbell, M. and Kiang, A.-S. (2012) Hereditary Retinopathies - Progress in Development of Genetic and Molecular Therapies. *Springer-Verlag New York* DOI 10.1007/978-1-4614-4499-2
27. Daiger, S.P., Sullivan, L.S. and Bowne, S.J. (2013) Genes and mutations causing retinitis pigmentosa. *Clin. Genet.*, **84**, 132–141.
28. Carrigan, M., Duignan, E., Humphries, P., Palfi, A., Kenna, P.F. and Farrar, G.J. (2016) A novel homozygous truncating GNAT1 mutation implicated in retinal degeneration. *Br. J. Ophthalmol.*, **100**, 495–500.
29. Shankar, S.P., Hughbanks-Wheaton, D.K., Birch, D.G., Sullivan, L.S., Conneely, K.N., Bowne, S.J., Stone, E.M. and Daiger, S.P. (2016) Autosomal dominant retinal dystrophies caused by a founder splice site mutation, c.828 + 3A>T, in PRPH2 and protein haplotypes in trans as modifiers. *Invest. Ophthalmol. Vis. Sci.*, **57**, 349–359.
30. Tee, J.J., Smith, A.J., Hardcastle, A.J. and Michaelides, M. (2016) RPGR-associated retinopathy: clinical features, molecular genetics, animal models and therapeutic options. *Br. J. Ophthalmol.*, **00**, 1022–1027.
31. Vollrath, D., Yasumura, D., Benchorin, G., Matthes, M.T., Feng, W., Nguyen, N.M., Sedano, C.D., Calton, M.A. and LaVail, M.M. (2015) Tyro3 modulates MerTK-associated retinal degeneration. *PLoS Genet.*, **11**, e1005723.
32. 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. *Sci. Rep.*, **6**, 19450.
33. Conley, S.M., Stuck, M.W., Watson, J.N. and Naash, M.I. (2017) Rom1 converts Y141C-Prph2-associated pattern dystrophy to retinitis pigmentosa. *Hum. Mol. Genet.*, **26**, 509–518.
34. Pierce, E.A. and Bennett, J. (2015) The status of RPE65 gene therapy trials: safety and efficacy. *Cold Spring Harb. Perspect. Med.*, **5**, a017285.
35. Lecuit, M. and Eloit, M. (2015) The potential of whole genome NGS for infectious disease diagnosis. *Expert. Rev. Mol. Diagn.*, **15**, 1517–151519.
36. Siemiatkowska, A.M., Collin, R.W., den Hollander, A.I. and Cremers, F.P. (2014) Genomic approaches for the discovery of genes mutated in inherited retinal degeneration. *Cold Spring Harb. Perspect. Med.*, **4**, pii: a017137.
37. Schulz, H.L., Grassmann, F., Kellner, U., Spital, G., Rütger, K., Jägle, H., Hufendiek, K., Rating, P., Huchzermeyer, C., Baier, M.J. et al. (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. *Invest. Ophthalmol. Vis. Sci.*, **58**, 394–403.
38. Bujakowska, K.M., Fernandez-Godino, R., Place, E., Consugar, M., Navarro-Gomez, D., White, J., Bedoukian, E.C., Zhu, X., Xie, H.M., Gai, X. et al. (2016) Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. *Genet. Med.*, 2016 [Epub ahead of print]
39. Khateb, S., Hanany, M., Khalaileh, A., Beryozkin, A., Meyer, S., Abu-Diab, A., Abu Turkey, F., Mizrahi-Meissonnier, L., Lieberman, S. et al. (2016) Identification of genomic deletions causing inherited retinal degenerations by coverage analysis of whole exome sequencing data. *J. Med. Genet.*, **53**, 600–607.
40. Van Cauwenbergh, C., Van Schil, K., Cannoodt, R., Bauwens, M., Van Laethem, T., De Jaegere, S., Steyaert, W., Sante, T., Menten, B., Leroy, B.P. et al. (2017) arrEYE: a customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs. *Genet. Med.*, **19**, 457–466.
41. Liquori, A., Vaché, C., Baux, D., Blanchet, C., Hamel, C., Malcolm, S., Koenig, M., Claustres, M. and Roux, A.F. (2016) Whole USH2A gene sequencing identifies several new deep intronic mutations. *Hum. Mutat.*, **37**, 184–193.
42. Sangermano, R., Bax, N.M., Bauwens, M., van den Born, L.I., De Baere, E., Garanto, A., Collin, R.W., Goercharn-Ramlal, A.S., den Engelsman-van Dijk, A.H., Rohrschneider, K. et al. (2016) Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c.5461-10T→C Mutation in Stargardt Disease. *Ophthalmology*, **123**, 1375–1385.
43. Mayer, A.K., Rohrschneider, K., Strom, T.M., Glöckle, N., Kohl, S., Wissinger, B. and Weisschuh, N. (2016) Homozygosity mapping and whole-genome sequencing reveals a deep intronic PROM1 mutation causing cone-rod dystrophy by pseudoexon activation. *Eur. J. Hum. Genet.*, **24**, 459–462.
44. Conte, I., Hadfield, K.D., Barbato, S., Carrella, S., Pizzo, M., Bhat, R.S., Carissimo, A., Karali, M., Porter, L.F., Urquhart, J. et al. (2015) MiR-204 is responsible for inherited retinal dystrophy associated with ocular coloboma. *Proc. Natl. Acad. Sci. U S A*, **112**, E3236–E3245.
45. Coppieters, F., Todeschini, A.L., Fujimaki, T., Baert, A., De Bruyne, M., Van Cauwenbergh, C., Verdin, H., Bauwens, M., Ongenaert, M., Kondo, M. et al. (2015) Hidden Genetic Variation in LCA9-Associated Congenital Blindness Explained by 5'UTR Mutations and Copy-Number Variations of NMNAT1. *Hum. Mutat.*, **36**, 1188–1196.
46. Radziwon, A., Arno, G.K., Wheaton, D., McDonagh, E.M., Baple, E.L., Webb-Jones, K.G., Birch, D., Webster, A.R. and MacDonald, I.M. (2017) Single-base substitutions in the CHM promoter as a cause of choroideremia. *Hum. Mutat.*, [Epub ahead of print]
47. Farrar, G.J., Millington-Ward, S., Chadderton, N., Mansergh, F.C. and Palfi, A. (2014) Gene therapies for inherited retinal disorders. *Vis. Neurosci.*, **31**, 289–307.
48. Dalkara, D., Goureau, O., Marazova, K. and Sahel, J.A. (2016) Let there be light: gene and cell therapy for blindness. *Hum. Gene Ther.*, **27**, 134–147.
49. Scholl, H.P., Strauss, R.W., Singh, M.S., Dalkara, D., Roska, B., Picaud, S. and Sahel, J.A. (2016) Emerging therapies for inherited retinal degeneration. *Sci. Transl. Med.*, **8**, 368rv6.
50. MacLaren, R.E., Bennett, J. and Schwartz, S.D. (2016) Gene therapy and stem cell transplantation in retinal disease: The New Frontier. *Ophthalmology*, **123**, S98–S106.
51. Bennett, J. (2017) Taking Stock of Retinal Gene Therapy: Looking Back and Moving Forward. *Mol. Ther.*, [Epub ahead of print]

52. Cideciyan, A.V., Jacobson, S.G., Beltran, W.A., Sumaroka, A., Swider, M., Iwabe, S., Roman, A.J., Olivares, M.B., Schwartz, S.B., Komáromy, A.M. et al. (2013) Human retinal gene therapy for Leber congenital amaurosis shows advancing retinal degeneration despite enduring visual improvement. *Proc. Natl. Acad. Sci. U S A*, **110**, 517–525.
53. Millington-Ward, S., Chadderton, N., O'Reilly, M., Palfi, A., Goldmann, T., Kilty, C., Humphries, M., Wolfrum, U., Bennett, J., Humphries, P. et al. (2011) Suppression and replacement gene therapy for autosomal dominant disease in a murine model of dominant retinitis pigmentosa. *Mol. Ther.*, **19**, 642–649.
54. Wert, K.J., Davis, R.J., Sancho-Pelluz, J., Nishina, P.M. and Tsang, S.H. (2013) Gene therapy provides long-term visual function in a pre-clinical model of retinitis pigmentosa. *Hum. Mol. Genet.*, **22**, 558–567.
55. Chadderton, N., Palfi, A., Millington-Ward, S., Gobbo, O., Overlack, N., Carrigan, M., O'Reilly, M., Campbell, M., Ehrhardt, C., Wolfrum, U. et al. (2013) Intravitreal delivery of AAV-NDI1 provides functional benefit in a murine model of Leber hereditary optic neuropathy. *Eur. J. Hum. Genet.*, **21**, 62–68.
56. Mookherjee, S., Hiriyanna, S., Kaneshiro, K., Li, L., Li, Y., Li, W., Qian, H., Li, T., Khanna, H., Colosi, P. et al. (2015) Long-term rescue of cone photoreceptor degeneration in retinitis pigmentosa 2 (RP2)-knockout mice by gene replacement therapy. *Hum. Mol. Genet.*, **24**, 6446–6458.
57. LaVail, M.M., Yasumura, D., Matthes, M.T., Yang, H., Hauswirth, W.W., Deng, W.T. and Vollrath, D. (2016) Gene therapy for MERTK-associated retinal degenerations. *Adv. Exp. Med. Biol.*, **854**, 487–493.
58. Latella, M.C., Di Salvo, M.T., Cocchiarella, F., Benati, D., Grisendi, G., Comitato, A., Marigo, V. and Recchia, A. (2016) In vivo editing of the human mutant rhodopsin gene by electroporation of plasmid-based CRISPR/Cas9 in the mouse retina. *Mol. Ther. Nucleic Acids*, **5**, e389.
59. Dalkara, D., Byrne, L.C., Klimczak, R.R., Visel, M., Yin, L., Merigan, W.H., Flannery, J.G. and Schaffer, D.V. (2013) In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. *Sci. Transl. Med.*, **5**, 189ra76.
60. Castle, M.J., Turunen, H.T., Vandenberghe, L.H. and Wolfe, J.H. (2016) Controlling AAV tropism in the nervous system with natural and engineered capsids. *Methods Mol. Biol.*, **1382**, 133–149.
61. Sengupta, A., Chaffiol, A., Macé, E., Caplette, R., Desrosiers, M., Lampič, M., Forster, V., Marre, O., Lin, J.Y. et al. (2016) Red-shifted channelrhodopsin stimulation restores light responses in blind mice, macaque retina, and human retina. *EMBO Mol. Med.*, **8**, 1248–1264.
62. Gaub, B.M., Berry, M.H., Holt, A.E., Isacoff, E.Y. and Flannery, J.G. (2015) Optogenetic vision restoration using rhodopsin for enhanced sensitivity. *Mol. Ther.*, **23**, 1562–1571.
63. Lopes, V.S. and Williams, D.S. (2015) Gene therapy for the retinal degeneration of Usher syndrome caused by mutations in MYO7A. *Cold Spring Harb. Perspect. Med.*, **5**, pii: a017319.
64. Zulliger, R., Conley, S.M. and Naash, M.I. (2015) Non-viral therapeutic approaches to ocular diseases: An overview and future directions. *J. Control Release*, **219**, 471–487.
65. Campbell, M., Humphries, M.M., Kiang, A.S., Nguyen, A.T., Gobbo, O.L., Tam, L.C., Suzuki, M., Hanrahan, F., Ozaki, E., Farrar, G.J. et al. (2011) Systemic low-molecular weight drug delivery to pre-selected neuronal regions. *EMBO Mol. Med.*, **4**, 235–245.

RESEARCH ARTICLE

Open Access



Multimodal imaging in a pedigree of X-linked Retinoschisis with a novel *RS1* variant

Kirk Stephenson^{1*}, Adrian Dockery^{2†}, Niamh Wynne³, Matthew Carrigan², Paul Kenna³, G. Jane Farrar² and David Keegan¹

Abstract

Background: To describe the clinical phenotype and genetic cause underlying the disease pathology in a pedigree (affected $n = 9$) with X-linked retinoschisis (XLR $RS1$) due to a novel *RS1* mutation and to assess suitability for novel therapies using multimodal imaging.

Methods: The Irish National Registry for Inherited Retinal Degenerations (Target 5000) is a program including clinical history and examination with multimodal retinal imaging, electrophysiology, visual field testing and genetic analysis. Nine affected patients were identified across 3 generations of an XLR $RS1$ pedigree. DNA sequencing was performed for each patient, one carrier female and one unaffected relative. Pedigree mapping revealed a further 4 affected males.

Results: All affected patients had a history of reduced visual acuity and dyschromatopsia; however, the severity of phenotype varied widely between the nine affected subjects. The stage of disease was classified as previously described. Phenotypic severity was not linearly correlated with age. A novel *RS1* (Xp22.2) mutation was detected (NM_000330: c.413C > A) resulting in a p.Thr138Asn substitution. Protein modelling demonstrated a change in higher order protein folding that is likely pathogenic.

Conclusions: This family has a novel gene mutation in *RS1* with clinical evidence of XLR $RS1$. A proportion of the older generation has developed end-stage macular atrophy; however, the severity is variable. Confirmation of genotype in the affected grandsons of this pedigree in principle may enable them to avail of upcoming gene therapies, provided there is anatomical evidence (from multimodal imaging) of potentially reversible early stage disease.

Keywords: X-linked Retinoschisis, Retinoschisin, Inherited retinal dystrophy, Inherited maculopathy

Background

Target 5000 [1, 2] (Fighting Blindness, Ireland), the Irish National Inherited Retinal Degeneration Registry, aims to phenotype and genotype all patients in Ireland with inherited retinal degenerations (IRD). A concurrent goal of the study is to facilitate the implementation of individualised management plans including, where appropriate, novel therapeutic options for those patients with modifiable disease. This paper is focused on the

application of this process for a pedigree with X-Linked Retinoschisis (OMIM: 312700, XLR $RS1$) and the clinical and genetic workup of these patients for potential new therapies and future participation in appropriate trials now emerging for this form of IRD.

XLR $RS1$ is a rare IRD (1:15000–30,000 [3]) due to variants in the *RS1* gene encoding the retinoschisin protein (OMIM: 300839, Xp22.1). XLR $RS1$ is a mutationally heterogeneous disorder with over 230 known variants [4–6], the majority of mutations falling in exons 4–6 (the structurally important discoidin domain of the retinoschisin protein [7]). Congenital poor vision in males is due to macular schisis and results in variable outer retinal atrophy with age. As novel gene therapies are in clinical trials [8, 9], to

* Correspondence: kirkstephenson@hotmail.com

[†]Kirk Stephenson and Adrian Dockery contributed equally to this work.

¹The Catherine McAuley Centre, Mater Private Hospital, Nelson Street, Dublin 7, Ireland

Full list of author information is available at the end of the article



optimise options for patients, confirmation of the presence of an *RS1* variant is essential and moreover must be supplemented with various clinical assays to assess the stage of disease and potential suitability for treatment.

With wide variability in phenotype between and within families [10–12], clinical diagnosis can be challenging. Multimodal imaging has been an excellent addition for confirming phenotype and guiding molecular genetic testing. Optical Coherence Tomography (OCT) imaging may detect intraretinal cystic spaces [13], which may appear in multiple retinal planes. Fundus autofluorescence (AF) imaging can detect subtle patterns of change in the retinal pigment epithelium (RPE), which allows staging of severity [14]. Microperimetry can delineate areas of residual function [15]. Thus, multimodal imaging can be used to demonstrate maintenance of anatomy which may determine suitability for novel treatments.

The *RS1* genomic location was first implicated as pathogenic using multipoint linkage analysis of Dutch families affected by XLRS1 [16]. Locus localization was later refined and the *RS1* gene was identified by positional cloning. Mapping and expression analysis led to the discovery of the then novel transcript (XLRS1) [17] encoding the retinoschisin protein.

Retinoschisin, a 224 amino acid protein, is secreted from photoreceptors and bipolar cells in retina and has been shown to be crucial for cell adhesion during retinal development [18]. It has been suggested that retinoschisin may also be involved in mitogen-activated protein kinase signaling and apoptosis in retina [18] and recently has been shown to influence Na/K-ATPase signaling and localization [19]. Employing cryo-electron microscopy, it has been discovered that functional retinoschisin forms a dimer of octamer rings comprising a hexadecamer. Given this intricate structure, a mutation that affects the folded structure of the *RS1* protein monomer is likely to inhibit the proper assembly of the 16-subunit oligomer [20, 21] as evidenced by the large number of reported pathogenic mutations [5]. Pathogenic *RS1* mutations are typically associated with XLRS1, although some mutations show greater phenotypic variability than others, including greater ranges of onset age and severity of condition [12].

Methods

Clinical phenotyping

A large pedigree with a clinical diagnosis of XLRS1 was invited to participate given informed consent. Nine affected males (7 adults, 2 children), one carrier female and one unaffected male attended the recruiting hospitals as specified above to take part in the IRD registry (Fig. 1).

Phenotyping included an ophthalmic and medical history, pedigree mapping, and dilated ophthalmic examination. Colour photography, autofluorescence (AF, Optos plc, Scotland) and spectral domain optical coherence tomography (SD-OCT, Cirrus HD-OCT, Carl Zeiss Meditec AG, CA, USA) images were acquired. This multimodal imaging was assessed and the macular findings were categorised into their appropriate stages. Stage was assessed by individual eye, not entire patient.

DNA isolation and next generation sequencing

Blood samples were taken from patients after clinical assessment. DNA was isolated from 2 ml of blood and fragmented for targeted sequencing to an average fragment size of 200–250 base pairs. Sequencing libraries were generated and target capture was performed with the Nimblegen SeqCap EZ kit (Roche), incorporating the exonic regions of over 200 genes implicated in IRDs. Capture regions also included intronic regions in the CEP290, ABCA4 and USH2A genes that are known to potentially contain pathogenic mutations [22–24]. The total size of the captured region was approximately 750 kb.

Captured patient DNA was multiplexed into 24-sample pools and sequenced using an Illumina MiSeq. Confirmatory single-read sequencing was also performed to verify the presence of candidate mutations.

Polymerase chain reaction and sanger sequencing

To validate the *RS1* variant identified by NGS, an amplicon for direct Sanger sequencing was designed incorporating the variant. The sequence used for primer design was Human reference transcript NM_000330.3. Forward primer: 5'- GCAGATGATCCACTGTGCTG – 3'. Reverse primer: 5' - TTTCTTGGGAGGTGGAGATG – 3'. Oligonucleotides were purchased from Sigma-Aldrich (www.sigmaaldrich.com/). The target DNA products were

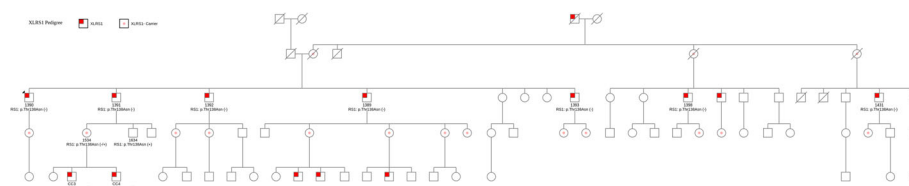


Fig. 1 Pedigree of 5 generations of an XLRS1 pedigree. Individuals marked with a red square were confirmed to be clinically affected. *RS1* genotype has been annotated within the pedigree tree for those investigated

amplified using Q5 High-Fidelity 2x Master Mix (New England Biolabs Inc). The annealing temperature for reactions was 65 °C; all other details were executed as per the supplier's recommendations. PCR products were purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific). Sanger sequencing was performed by Eurofins Genomics (www.eurofinsgenomics.eu).

Data analysis

Data obtained from NGS was subsequently demultiplexed and mapped to the human genome (hg38) using BWA version 0.7.15 [25]. Duplicate reads were flagged using Picard version 2.5.0 [26] and downstream analysis and variant calling performed using Freebayes version 1.1.0 [27].

Variants were identified and scored based on methods outlined in Carrigan et al. [1]. Synonymous variants, polymorphisms and mutations with high frequency in any population were filtered out, and the remaining list of rare variants with the potential to affect protein sequence was output for manual inspection. Output scores from the following ensemble variant pathogenicity predictor tools are shown in Table 1. The scale for each score ranges from 0 (likely benign) to 1 (likely pathogenic).

MetaLR

Meta Logistic Regression (LR) [28] incorporates pathogenicity prediction scores and maximum minor allele frequency from nine different tools for more accurate and thorough evaluation of deleterious effect of missense mutations. This allows for the more accurate assessment of variants than any of the singular methods alone.

M-CAP

Mendelian Clinically Applicable Pathogenicity (M-CAP) Score [29] was the first high sensitivity pathogenicity classifier for rare missense variants in the human genome aimed at the clinic. It combines pathogenicity scores from other tools and databases (including SIFT, Polyphen-2 and CADD) with novel features to create a more powerful model.

Revel

REVEL (rare exome variant ensemble learner) [30], is also an ensemble predictor tool and was trained with recently discovered pathogenic and rare benign missense variants, excluding those previously used to train its constituent tools. REVEL also claims to have the best

performance for categorising pathogenic from rare benign variants with allele frequencies < 0.5%.

The American College of Medical Genetics and Genomics (ACMG) criteria for classifying pathogenic variants was utilised [31]. The other rare variants detected were for autosomal recessive conditions and showed pathogenic variants in only 1 allele. This was consistent between affected male relatives in this study.

Protein modelling

3D models of single subunit wildtype RS1 (NM_000330.3) and the Thr138Asn mutant were generated using Iterative Threading ASSEmbly Refinement, I-TASSER [32]. Polymer structures of RS1 were obtained from the Protein Data Bank (PDB, ID#3JD6,) [33]. The effect of single point mutations on protein stability was measured using STRUM [34]. Protein alignments were generated using Clustal Omega [35].

Results

Mean patient age of affected adult males ($n = 7$) was 67 $\pm$ 5.38 years. The recruited carrier female (daughter) was 47 years of age. Mean age of the 2 affected grandsons was 13 $\pm$ 1.41 years. Mean visual acuity for adult affected males was LogMAR 0.78 $\pm$ 0.16 for right eyes and 0.74 $\pm$ 0.28 for left eyes. Two eyes were not assessable due to one case of dense cataract (amblyopic eye) and one enucleation (due to complications following previous retinal surgery, presumed post vitreous haemorrhage or detachment). Only one of 14 adult eyes developed either a retinal detachment or haemorrhage previously (indeterminable). Mean visual acuity of affected grandsons was LogMAR 0.6 $\pm$ 0.14 for right eyes and 0.5 for both left eyes.

Macular changes varied in severity, which showed no correlation to patient age. A severity staging was proposed by Tsui and Tsang [14], outlined in Table 2. Of 12 assessable adult eyes, 1 was in Stage 1 (8%, mean age 74y), 6 were in Stage 2 (50%, mean age 65.6 $\pm$ 5.72y) and 5 were in Stage 3 (42%, mean age 68.8 $\pm$ 4.55y). The predominantly advanced stage of disease here is likely due to advanced patient age. Corresponding images from this series are in accordance with this and are shown in Fig. 2. Both affected grandsons had bilateral Stage 1 disease. Electrophysiology data for the 2 affected grandsons is shown in Fig. 3. Confirmation of this RS1 variant's pathogenicity will

Table 1 A direct comparison of the novel mutation, p.Thr138Asn, and a known pathogenic variant in RS1. Both variants were analysed with the same bioinformatic pipeline and have been scored by the same computational tools

Variant	Nucleotide	Protein	DBSNP ID	MetaLR	M-CAP	REVEL	ClinVar Report
Novel	c.413C > A	p.Thr138Asn	None	0.9742	0.8775	0.929	None
Known	c.636G > A	p.Arg209His	rs281865362	0.9446	0.9233	0.796	Pathogenic

Table 2 The clinical staging of Retinoschisis [14]

Stage	Colour	AF	SD-OCT
1	Macular schisis, classically radial (A)	Macular pigments displaced between schitic areas (B)	≥1 intraretinal schisis plane (C)
2	RPE pigment mottling (D)	Hyperautofluorescent ring at macula (E)	Collapse of schisis plane(s) (F)
3	Macular atrophy (G)	Large area of central hypoautofluorescence with surrounding hyperautofluorescent ring (H)	Outer retinal atrophy (I)

be of most benefit for this younger generation with potentially reversible macular changes.

Figure 2 demonstrates the clinical staging of XLRS using multimodal imaging. Patient 1 (74y) demonstrates the features of Stage 1 (mild) disease with BCVA LogMAR 0.6. There are subtle pigmentary changes on colour photography (A), with displacement of macular pigment leading to radial pattern of hyper- vs hypo-autofluorescence (B). The most striking feature is on OCT with intraretinal cystic spaces (C) in the inner plexiform and inner nuclear layers.

Patient 2 (60y) shows Stage 2 (intermediate) disease with BCVA of LogMAR 0.9. Colour photography reveals a hypopigmented ring at the macula, surrounding an area of normal pigmentation (D). Autofluorescence confirms this as hypoautofluorescent

with a surrounding ring of hyperautofluorescence and a preserved iso-autofluorescent foveal area (E). OCT reveals the absence/collapse of intraretinal cystic spaces, but lacking complete outer retinal atrophy (F). Incidental vitreo-foveal traction is noted in this case.

Patient 3 (71y) has stage 3 (advanced) disease with BCVA of LogMAR 1.0. The colour photograph demonstrates macular atrophy (G), which appears as hypoautofluorescence with a hyperautofluorescent ring (H). This corresponds to outer retinal atrophy without schitic intraretinal space on OCT (I).

Next generation sequencing detected a novel missense mutation in *RS1* (c.413C>A; p.Thr138Asn [1]) falling within exon 5 of the discoidin domain [7]. This is a variant of unknown significance; however, due to its location adjacent to known pathogenic loci, it was deemed to be

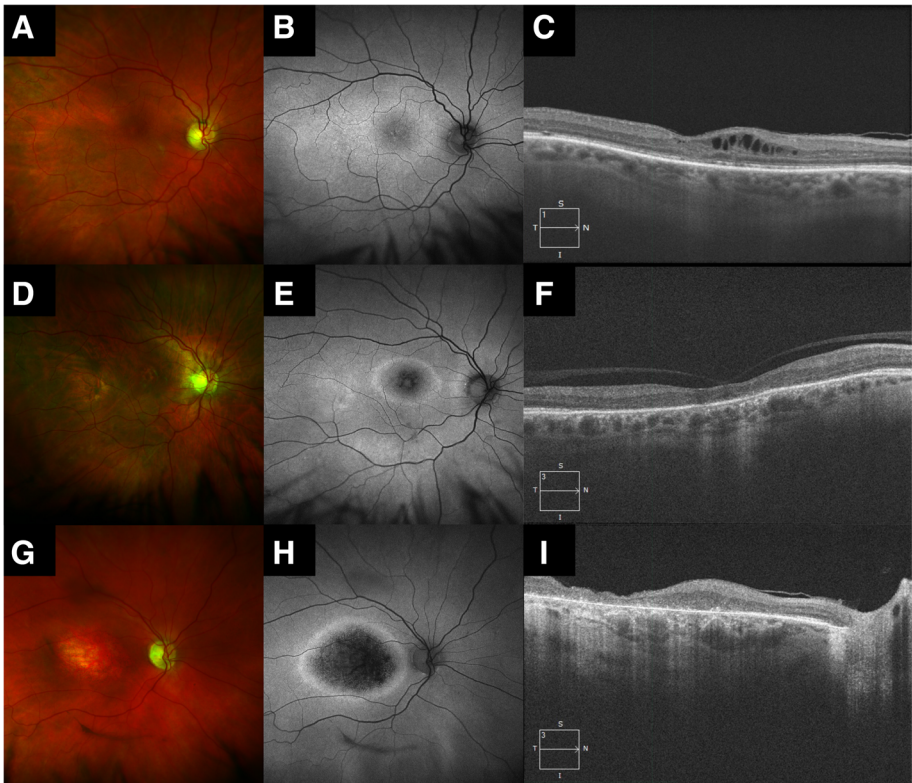


Fig. 2 Phenotypic features of selected cases from this XLRS1 pedigree. Colour photographs (left column), autofluorescence (middle column), and SD-OCT (right column) of the right eyes of 3 patients in Stage 1 (top row), Stage 2 (middle row) and Stage 3 (bottom row). Further description is available in Table 2 and the ‘Results’ section

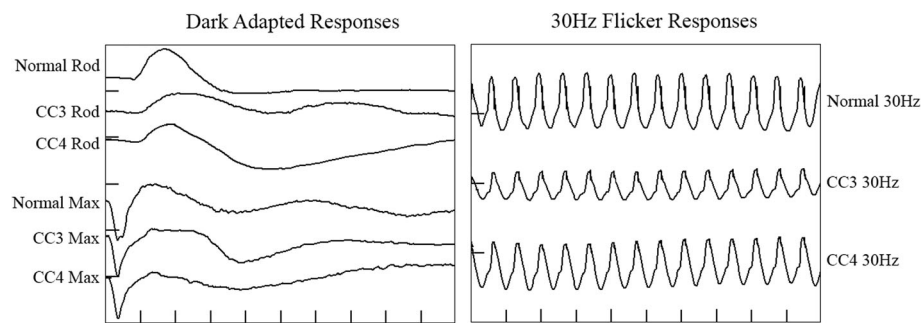


Fig. 3 ISCEV full-field electroretinogram [45] responses from affected grandsons (CC3 and CC4). Horizontal axis: 50 msec per division; Vertical axis: Dark Adapted Responses 500 μ V per division; 30 Hz Flicker Responses 250 μ V per division. The rod-isolated responses of CC3 and CC4 were slightly, but significantly reduced in amplitudes. The mixed rod and cone responses to the maximal intensity flash stimulus presented to the dark-adapted eye showed a normal a-wave amplitude in each subject, (CC3 445 μ V, CC4 441 μ V). The b-wave amplitudes were significantly reduced in each patient (CC3 515 μ V, CC4 502 μ V) and approximated the a-wave amplitudes. Ordinarily, the b-wave amplitude greatly exceeds that of the a-wave. The normal a-wave and selective reduction of the b-wave indicates preserved photo-transduction but impaired post-receptoral retinal function. This has been reported in RS1-associated X-linked retinoschisis [44]

potentially pathogenic and in silico analysis was undertaken.

Mutational analysis

The exonic regions from a panel of 218 IRD genes were sequenced by NGS in the proband from this pedigree. An additional 5 affected males were sequenced with a subsequent panel of the exonic regions of 254 IRD genes. Analysis of this NGS data led to the identification of a mutation in the *RS1* gene within the pedigree. Further analysis of each of the members of the family by PCR amplification of *RS1* and Sanger sequencing demonstrated that each affected male was found to have a hemizygous nonsynonymous mutation in exon 5 of the *RS1* gene (NM_000330.3:c.413C > A, p.Thr138Asn). In addition, the single carrier female analyzed was heterozygous for the same *RS1* mutation while the unaffected male was hemizygous for the reference base at that position (Fig. 4a, b).

The p.Thr138Asn *RS1* mutation involves a polar to polar amino acid substitution. Given that this novel as yet previously unreported mutation is located in the critical discoilin domain of the retinoschisin protein, where many pathogenic mutations have been identified previously, additional in silico analysis was undertaken to explore the potential pathogenicity of this *RS1* mutation on protein structure and conformation. The destabilizing effect of this mutation as predicted in silico is shown below in terms of a prediction of the changes to fold stability (Table 3), where a negative delta delta G ($\Delta\Delta G$) score is indicative of a destabilizing effect. This change is also represented visually as theoretical structural changes to the retinoschisin monomeric protein structure (Fig. 5a, b). Notable secondary structural changes are identified with arrows in Fig. 5a while Fig. 5b clearly visualizes the position of the specific amino acid substitution. The complexity of the

hexadecameric structure of retinoschisin is shown in Fig. 6 for reference.

Table 1 displays the comparative output scores of three highly regarded variant pathogenicity prediction tools for this novel variant and a known pathogenic variant. As seen in the respective columns of the tools, both variants have been deemed likely pathogenic by all three tools. Also, the novel variant described in this study is scored more likely to be pathogenic by two out of three methods listed.

The ACMG guidelines [31] have also been implemented to classify the novel variant. The relevant information has been outlined in Table 4 along with supporting evidence where appropriate [20, 36, 37]. The guidelines state that there is sufficient evidence to classify this variant as pathogenic as there are one strong (PS1) and three moderate (PM1, PM2 and PM5) lines of evidence.

Discussion

A key function of retinoschisin is in retinal adhesion [7]. There is some debate over which retinal plane is affected by the schisis, which has been documented on OCT at the level of the photoreceptor inner segments, the inner nuclear layer and the nerve fibre layer [13, 38]. Decreased retinoschisin protein function may, and likely does, cause multiple planes of schisis.

There is significant variability of the severity of disease phenotype between the two eyes of the same individual, between different members of this XLRS1 pedigree and, as previously documented, between unrelated individuals [10–12]. This variability appears greater with age with some patients progressing through the clinical stages at different rates, possibly due to acquired/environmental factors or other genetic modifiers (see below). Although this variability is noted, there are clinical characteristics that guide the clinician to this diagnosis (e.g. pedigree,

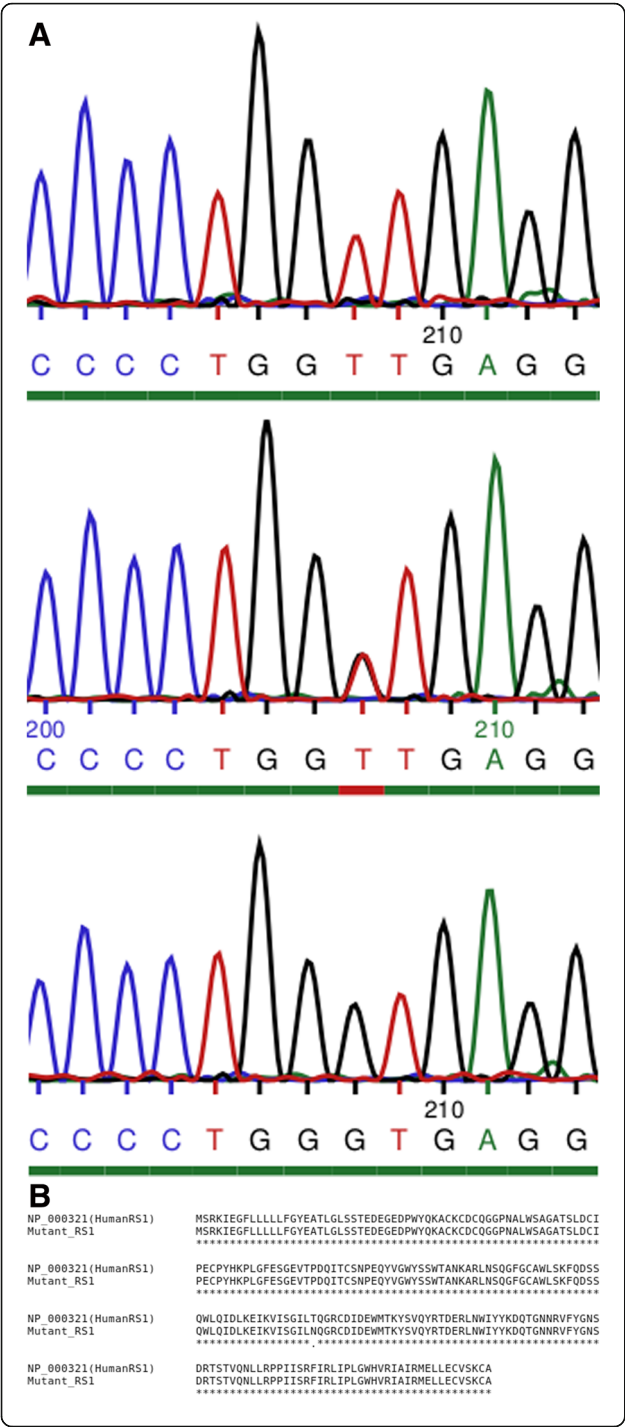


Fig. 4 a Sanger sequencing of *RS1*-exon 5 polymerase chain reaction (PCR) products from members of the pedigree. Top: Sanger sequencing result for affected male (ID#1392) showing a T nucleotide trace at sequencing position 208. Middle: Sanger sequencing result for carrier female (ID#1534) showing a heterozygous result at sequencing position 207, revealing traces from both signals of wild-type G allele and the mutant T allele. Bottom: Sanger sequencing result for unaffected male (ID#1634) showing a G nucleotide trace at sequencing trace position 208. **b** Clustal Omega results for protein alignment of reference *RS1* protein sequence against the observed mutant protein sequence. “.” denotes the position of the substitution

peripheral retinoschisis, lack of subretinal flecks or vitelliform changes) which are supported by a likely-pathogenic *RS1* variant and lack of other known pathogenic genetic variants. This supports the bimodal presentation of XLR₁. Severe cases of the disease (approximately 30%) are detected in infancy with poor fixation and/or strabismus. Milder cases (approximately 70%) are detected at school entry or in adulthood [39]. The current study highlights that the underlying genetic variant on the *RS1* gene is not predictive of phenotypic severity for this, or indeed any of the other, genetic variant(s) that have been described [12]. Thus, in assessing patients for trial/intervention suitability both a genetic and full phenotypic analysis is required. For example, gene therapy will not be beneficial if there is significant structural change (i.e. stage 3). However, the combination of a stage 1 or 2 case with confirmed genotype would suggest a possible role for genetic manipulation/replacement. As our knowledge of the impact of multiple variants increases we may better predict the molecular effects of each on retinal structure and function thus tailoring our approach to intervention further.

The clinical findings may be subtle in early stage or mild forms of disease. Classically, radial cystic maculopathy at the fovea is seen in 98–100% of cases [40] (Fig 2a-c). Ancillary tests have proven useful, particularly in subtle cases. An electronegative electroretinogram is suggestive of, but not specific, to XLR₁ [41] as this feature may be seen in acquired disease of the inner retina (e.g. central retinal artery occlusion). Microperimetry may assist in detection of subclinical disease [15]. Late stage disease is a diagnostic challenge; however, clinical examination, family history and genotyping as outlined above will aid in determining the diagnosis.

Colour photography, Autofluorescence and Optical Coherence Tomography are non-invasive and accessible

Table 3 Output results of the STRUM program. The change in stability scores between the mutant and wild type *RS1* protein is given as delta delta G (ddG)

Position	Wild Type	Mutant Type	ddG
138	T	N	−0.78

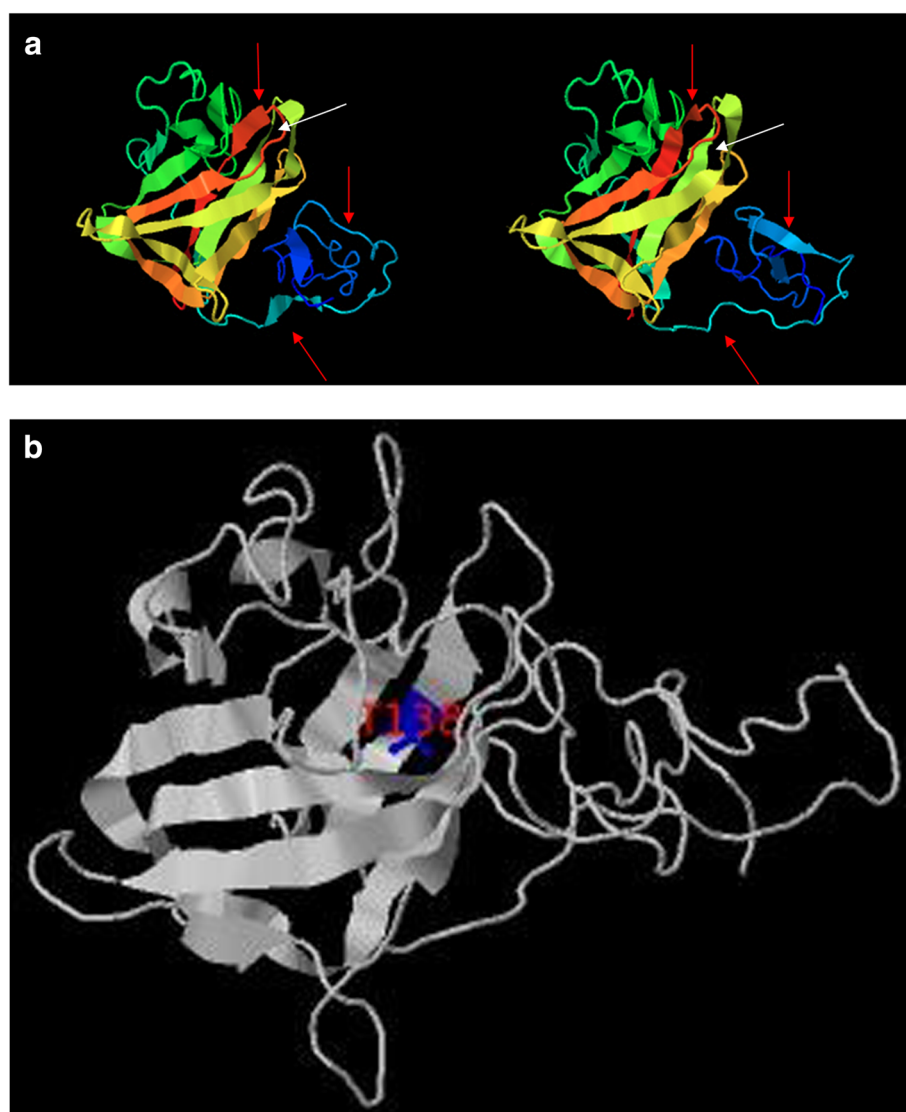


Fig. 5 **a** Protein models of reference RS1 (left) and mutant T138 N (right) proteins generated by the I-TASSER program. Red arrows denote obvious structural changes to protein structures. A white arrow highlights the position of the amino acid change at codon 138. **b** Protein model of RS1 mutant T138 N generated by STRUM. The affected amino acid residue is displayed in blue and annotated in red

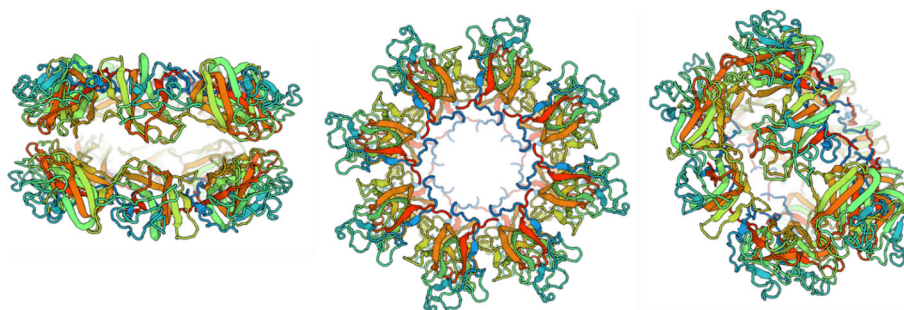


Fig. 6 Different views of the back-to-back octamer rings of wild type RS1 obtained from the Protein Data Bank. (Left) Objective view. (Middle) Plan view. (Right) Perspective view

Table 4 The ACMG guidelines applicable to the novel variant in question, p. Thr128Asn

Category Code	Qualifying Criteria	Relevance to Case
PS4	The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls.	This variant is found in a sufficient number of relatives to qualify it as strong evidence of pathogenicity [36].
PM1	Located in a mutational hot spot and/or critical and well-established functional domain (e.g. active site of an enzyme) without benign variation.	Located in well established functional domain as reported by several studies [20, 37].
PM2	Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes or ExAC.	Absent from every databased queried.
PM5	Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before.	Novel mutation at same amino acid as previously reported p.Thr138Ala (HGMD Database).
PP1	Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease.	Cosegregation in a family in a gene known to cause disease.
PP3	Multiple lines of computational evidence support a deleterious effect on the gene or gene product.	Multiple lines of computational evidence presented here (Table 1).
PP4	Patient's phenotype or family history is highly specific for a disease with a single genetic etiology.	Patient phenotype is highly specific for a disease.

tests useful in determining the clinical stage of XLR1. This is relevant, for current and upcoming gene therapy trials [8, 9], in selecting those cases most likely to benefit from intervention (i.e. early to intermediate disease [40]). Ancillary tests, such as adaptive optics (AO) imaging and microperimetry, can complement assessment of cone structure/spacing in those who are deemed early/intermediate stage [42, 43].

To our knowledge, this specific mutation is currently unreported in any clinical database available including the Leiden University's Open Variation Database for Retinosis (LOVD 3.0 [6]) and represents a novel and likely pathogenic mutation given the clear segregation in this pedigree and the significant predicted effects on protein structure.

Recent studies have outlined the importance of certain key residues in correctly assembling a functional RS1 oligomer. Several key residue positions necessary to ensure proper folding of the retinosis protein have been identified [20]. Of note, amino acids at positions 137 and 139 were highlighted in that study as fundamental to forming a beta-sheet required in the interaction between subunits essential in the formation of their functional octamer ring structure. It is possible therefore, that an amino acid substitution at position 138 may provide some steric hindrance, that in turn, would interfere with the relationship between different subunits on the same octamer ring. Such interference could in principle have deleterious consequences for the formation and stability of the oligomeric RS1 complex. This hypothesis is further supported by in silico analysis of the p.Thr138Asn mutation. Computational analysis was utilized to predict protein folding changes in the mutant protein (using I-TASSER); physiochemical properties, position specific conservation and secondary structure prediction scores were used to determine the destabilization

effect(s) of the protein substitution at position 138 (using STRUM) where significant changes in the above were predicted for the p.Thr138Asn mutation.

The prediction scores in Table 1 are useful for assessing the confidence with which a variant can be called deleterious or not. However, these are confidence scores and not pathogenicity scores, so although the scores listed would be considered to be quite high, these scores cannot indicate the detrimental effect that this mutation may have relative to a similar amino acid substitution elsewhere in the protein. Increased bioavailability of functional retinosis protein would logically correlate to the lasting preservation of the retinal layers; however, to assess this dosage effect, each variant would have to be assessed individually for function and half-life. This is a challenging task for retinosis as it has typically proven quite difficult to produce in useful quantities [20]. Currently we rely on in silico predictors of a variant's effect on protein structure. This includes protein modelling (Figs. 4a and b) where possible and delta delta g scores (Table 3), to predict the destabilising effects a mutation may have on a protein. However, when all lines of evidence are considered, this novel RS1 variant is classified as pathogenic under the guidelines outlined by the ACMG [31] (Table 4).

In XLR1, age is not an absolute factor, particularly in adulthood, when evaluating severity of disease and the likely benefit from novel gene therapies. The individuals in this XLR1 pedigree exhibited no clear correlation between patient age and 1) stage of disease, 2) visual acuity, 3) central retinal thickness or 4) age at onset despite a shared novel RS1 gene variant. Advanced cases prove a diagnostic challenge as many inherited and acquired macular diseases follow a final common pathway of outer retinal atrophy. Each case must be judged on its potential (anatomically within Stage 1 or 2 and

genetically possessing a pathogenic *RS1* variant) to respond to novel therapies. Furthermore, the variability in clinical presentations between family members with the same mutation suggests the possible involvement of genetic modifier loci and/or environmental effect(s).

While there is variability in phenotype between relatives affected by the same genetic variant, the significance is that the clinician not be discouraged from the diagnosis due to these differences but consider them within the spectrum of the 3 described clinical stages. The diagnostic value of ERG, in particular the characteristic decrease in b-wave activity in combination with suggestive clinical findings are powerful tools for diagnosing younger patients, however the diagnostic value of these tests may be less certain with advanced age [44]. Confirmation of extent of severity (clinical stages) and a pathogenic *RS1* variant open a potential avenue of treatment for those in Stage 1 or 2. This information then guides genetic testing for at risk relatives who could possibly benefit from novel disease-modifying therapies.

There is a possibility that there may be some form of modifier or digenic effect which could account for the observed phenotypic variability. With this in mind, the coding regions of known genes associated with retinal degenerations were subjected to sequence analysis. The presence of multiple pathogenic variants within a single patient is extremely rare but has been previously observed, even in our own IRD cohort. Although, as previously mentioned, this method unfortunately did not detect any likely genetic cause of phenotypic variability between the affected patients in this pedigree.

The retinoschisin protein is widely expressed throughout the layers of the retina in childhood and more specifically in rod and cone photoreceptor cells in adulthood and is seen to act as a cellular adhesive when functioning correctly. However, other proteins are also capable of fulfilling similar adhesive functions, for example adherins and cadherins in adherens junctions or claudins and occludin in tight junctions. Given that some of these other proteins belong to large protein families, it is a possibility that variation in relative composition of these cellular adhesive protein cocktails may have a more notable contribution to a variable phenotype in the absence of functional retinoschisin.

This variability creates a diagnostic challenge in advanced XLRS1 disease. The clinical clues in conjunction with adjunctive testing, confirmed by robust genetic assessment (i.e. ruling out other causes and confirming a pathogenic *RS1* variant) confirm the diagnosis, thus empowering those affected by bringing them closer to accessing novel gene therapies where appropriate and preparing younger generations for those potential therapeutic options.

Conclusions

Here we describe how multimodal imaging in XLRS1 and confirmation of a pathogenic *RS1* variant help to a) confirm diagnosis, b) monitor progression, c) assess clinical stage and thus suitability for novel therapeutic options even in patients of advanced age. Autofluorescence and OCT are non-invasive, well-tolerated and readily available in most clinical centres. Standardisation of clinical assessment and staging criteria, perhaps in conjunction with a central European reading centre, would allow detection of individuals with modifiable XLRS1 who may be suitable for novel gene therapy clinical trial participation.

Abbreviations

ACMG: American College of Medical Genetics; AF: Autofluorescence; AO: Adaptive Optics; IRD: Inherited Retinal Degeneration; I-TASSER: Iterative Threading ASSEMBly Refinement; LogMAR: Logarithm of the Minimum Angle of Resolution; LOVD: Leiden University's Open Variation Database; MMUH: Mater Misericordiae University Hospital, Dublin, Ireland; MPH: Mater Private Hospital, Dublin, Ireland; OCT: Optical Coherence Tomography; OMIM: Online Mendelian Inheritance in Man; PCR: Polymerase Chain Reaction; PDB: Protein Data Bank; RPE: Retinal Pigment Epithelium; RVEEH: Royal Victoria Eye & Ear Hospital, Dublin, Ireland; XLRS1: X-Linked Retinoschisis

Acknowledgements

Fighting Blindness Ireland; Clinical Research Centre: University College Dublin/Mater Misericordiae University Hospital, Dublin, Ireland; The Research Foundation, The Royal Victoria Eye & Ear Hospital, Dublin, Ireland; School of Genetics, Trinity College Dublin, Ireland; Retinal Eye Centre, Mater Private Hospital, Dublin, Ireland.

Funding

All patients were seen as part of the Target 5000 Irish National Inherited Retinal Degeneration Registry which is funded through the Irish charity Fighting Blindness (charity number: 20013349). There are no specific sources of industry or institutional funding supporting this publication. All publications arising as part of this project are independent endeavours of the publishing authors.

Availability of data and materials

All source notes are securely stored at the Catherine McAuley Research & Education Centre, Nelson Street, Dublin 7 or the Research Foundation, Royal Victoria Eye & Ear Hospital, Adelaide Road, Dublin 2. The 2 children were recruited via RVEEH while all adults were recruited via MMUH/MPH. Source documentation is available upon request from the corresponding author.

Authors' contributions

KS, NW, PK and DK recruited the patients, performed history and clinical examination, and wrote the clinical portions of the manuscript. AD, MC and GJF performed next generation sequencing on blood samples from the patients, performing protein modelling/pathogenicity analysis and wrote the genetic portions of the manuscript. All authors have read and approved the manuscript.

Ethics approval and consent to participate

This study was approved by the Institutional Review Boards of the Mater Misericordiae University Hospital, Mater Private Hospital and Royal Victoria Eye & Ear Hospital, Dublin, Ireland. All patients (or their legal guardians in the case of children) whose data is described in this manuscript signed informed consent before taking part in this study. This study is in accordance with the Declaration of Helsinki.

Consent for publication

All patients (or their legal guardians in the case of children) whose data is described in this manuscript signed informed consent giving permission for their data to be published anonymously.

Competing interests

None of the authors have any competing interests as laid out in BioMed Central's guidance document.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹The Catherine McAuley Centre, Mater Private Hospital, Nelson Street, Dublin 7, Ireland. ²Department of Genetics, Trinity College Dublin, Dublin, Ireland. ³The Research Foundation, The Royal Victoria Eye and Ear Hospital, Dublin, Ireland.

Received: 4 October 2017 Accepted: 29 October 2018

Published online: 12 November 2018

References

- Carrigan M, et al. Panel-based population next-generation sequencing for inherited retinal degenerations. *Sci Rep*. 2016;6:33248.
- Dockery A, et al. Target 5000: Target Capture Sequencing for Inherited Retinal Degenerations. *Genes (Basel)*. 2017;8(11). <https://doi.org/10.3390/genes8110304>.
- George ND, et al. Infantile presentation of X linked retinoschisis. *Br J Ophthalmol*. 1995;79(7):653–7.
- Functional implications of the spectrum of mutations found in 234 cases with X-linked juvenile retinoschisis. The Retinoschisis consortium. *Hum Mol Genet*. 1998;7(7):1185–92. <https://www.ncbi.nlm.nih.gov/pubmed/9618178>.
- Stenson PD, et al. The human gene mutation database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*. 2014; 133(1):1–9.
- Leiden University Medical Center, N. Leiden Open Variation Database 3.0. 2018; Available from: <http://www.lovd.nl/3.0/home>. Accessed 5 July 2018.
- Bush M, et al. Cog-wheel Octameric structure of RS1, the Discoidin domain containing retinal protein associated with X-linked Retinoschisis. *PLoS One*. 2016;11(1):e0147653.
- ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2015 Apr 15. Identifier NCT02416622, Safety and Efficacy of rAAV-hRS1 in Patients With X-linked. Retinoschisis (XLR5); [6 screens]. Available from: <https://clinicaltrials.gov/ct2/show/NCT02416622>. Accessed 5 July 2018.
- ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2014 Dec 14. Identifier NCT02317887, Study of RS1 Ocular Gene Transfer for X-linked Retinoschisis; [8 screens]. Available from: <https://clinicaltrials.gov/ct2/show/NCT02317887>. Accessed 5 July 2018.
- Eksandh LC, et al. Phenotypic expression of juvenile X-linked retinoschisis in Swedish families with different mutations in the XLR51 gene. *Arch Ophthalmol*. 2000;118(8):1098–104.
- Molday RS, Kellner U, Weber BH. X-linked juvenile retinoschisis: clinical diagnosis, genetic analysis, and molecular mechanisms. *Prog Retin Eye Res*. 2012;31(3):195–212.
- Lai YH, et al. A novel gene mutation in a family with X-linked retinoschisis. *J Formos Med Assoc*. 2015;114(9):872–80.
- Gregori NZ, et al. Macular spectral-domain optical coherence tomography in patients with X linked retinoschisis. *Br J Ophthalmol*. 2009;93(3):373–8.
- Tsui I, Tsang SH. In: Noemi Lois JVF, editor. Fundus autofluorescence in X-linked Retinoschisis, in Fundus Autofluorescence. Philadelphia: Lippincott Williams & Wilkins; 2009.
- Nittala MG, et al. Spectral-domain OCT and microperimeter characterization of morphological and functional changes in X-linked retinoschisis. *Ophthalmic Surg Lasers Imaging*. 2009;40(1):71–4.
- Bergen AA, et al. Multipoint linkage analysis in X-linked juvenile retinoschisis. *Clin Genet*. 1993;43(3):113–6.
- Sauer CG, et al. Positional cloning of the gene associated with X-linked juvenile retinoschisis. *Nat Genet*. 1997;17(2):164–70.
- Plossl K, et al. Retinoschisin is linked to retinal Na/K-ATPase signaling and localization. *Mol Biol Cell*. 2017;28(16):2178–89.
- Plossl K, Weber BH, Friedrich U. The X-linked juvenile retinoschisis protein retinoschisin is a novel regulator of mitogen-activated protein kinase signalling and apoptosis in the retina. *J Cell Mol Med*. 2017;21(4):768–80.
- Tolun G, et al. Paired octamer rings of retinoschisin suggest a junctional model for cell-cell adhesion in the retina. *Proc Natl Acad Sci U S A*. 2016; 113(19):5287–92.
- Ramsay EP, et al. Structural analysis of X-linked retinoschisis mutations reveals distinct classes which differentially effect retinoschisin function. *Hum Mol Genet*. 2016;25(24):5311–20.
- den Hollander AI, et al. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet*. 2006;79(3): 556–61.
- Bauwens M, et al. An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. *Hum Mutat*. 2015;36(1):39–42.
- Steele-Stallard HB, et al. Screening for duplications, deletions and a common intronic mutation detects 35% of second mutations in patients with USH2A monoallelic mutations on sanger sequencing. *Orphanet J Rare Dis*. 2013;8:122.
- Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
- Picard. Available from: <http://broadinstitute.github.io/picard/>. Accessed 18 Aug 2017.
- Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907 2012; Available from: <http://arxiv.org/abs/1207.3907>.
- Dong C, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24(8):2125–37.
- Jagadeesh KA, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet*. 2016;48(12): 1581–6.
- Ioannidis NM, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877–85.
- Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405–24.
- I-TASSER. Available from: <http://zhanglab.cmb.med.umich.edu/I-TASSER/>. Accessed 18 Aug 2017.
- RCSB Protein Data Bank. Available from: <https://www.rcsb.org/>. Accessed 18 Aug 2017.
- Quan L, Lv Q, Zhang Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics*. 2016;32(19): 2936–46.
- Sievers F, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol Syst Biol*. 2011;7:539.
- Jarvik GP, Browning BL. Consideration of Cosegregation in the pathogenicity classification of genomic variants. *Am J Hum Genet*. 2016; 98(6):1077–81.
- Plossl K, et al. Pathomechanism of mutated and secreted retinoschisin in X-linked juvenile retinoschisis. *Exp Eye Res*. 2018;177:23–34.
- Yanoff M, Kertesz Rahn E, Zimmerman LE. Histopathology of juvenile retinoschisis. *Arch Ophthalmol*. 1968;79(1):49–53.
- Deutman AF, Pinckers AJ, Aan de Kerk AL. Dominantly inherited cystoid macular edema. *Am J Ophthalmol*. 1976;82(4):540–8.
- Wang T, et al. Intracellular retention of mutant retinoschisin is the pathological mechanism underlying X-linked retinoschisis. *Hum Mol Genet*. 2002;11(24):3097–105.
- Sergeev YV, et al. Molecular modeling of retinoschisin with functional analysis of pathogenic mutations from human X-linked retinoschisis. *Hum Mol Genet*. 2010;19(7):1302–13.
- Akeo K, et al. Detailed morphological changes of Foveoschisis in patient with X-linked Retinoschisis detected by SD-OCT and adaptive optics fundus camera. *Case Rep Ophthalmol Med*. 2015;2015:432782.
- Duncan JL, et al. Abnormal cone structure in foveal schisis cavities in X-linked retinoschisis from mutations in exon 6 of the RS1 gene. *Invest Ophthalmol Vis Sci*. 2011;52(13):9614–23.
- Bradshaw K, et al. Mutations of the XLR51 gene cause abnormalities of photoreceptor as well as inner retinal responses of the ERG. *Doc Ophthalmol*. 1999;98(2):153–73.
- McCulloch DL, et al. ISCEV standard for full-field clinical electroretinography (2015 update). *Doc Ophthalmol*. 2015;130(1):1–12.