Genetic Analysis of Inherited Retinal Diseases in Indigenous Southern African Populations

Lisa Jane Roberts MSc(Medicine)
RBRLIS002

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
in fulfilment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
in the Division of Human Genetics, Department of Pathology,
Faculty of Health Sciences
UNIVERSITY OF CAPE TOWN

March 2017

Supervisor: Prof. R. Ramesar, MSc MBA PhD
Co-Supervisor: A. Swaroop, MSc PhD

Declaration

I, Lisa Jane Roberts, hereby declare that the work on which this dissertation is based is my original research work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature:

Signature removed

Date: 8 March 2017

Ex Africa semper aliquid novi.

Out of Africa, there is always something new.

Pliny the Elder, ~A.D. 70

Dedication

To my family, for their love, strength, encouragement and support.

To the babies, for the pure joy they have brought into my world.

And in loving memory of those no longer with us.

Acknowledgements

- I thank my supervisor, Raj Ramesar, and co-supervisor, Anand Swaroop, for the opportunities they have afforded me; to undertake this PhD, and to be able to visit Anand's lab at the National Eye Institute, to engage with the whole exome sequencing technology. I have learned a tremendous amount during my PhD journey.
- I am indebted to George Rebello, my friend and mentor, for his enthusiasm over my project, for being my 'sounding board' for ideas, and for creating the perl scripts which literally saved me from doing the analysis manually. I also thank his family for their support and for putting up with the hours-long phone calls.
- I am so grateful to Rene Goliath, for going above and beyond, by volunteering to read the first draft of my thesis. Your input was invaluable.
- I thank the patients with inherited retinal diseases who choose to participate in the research project. Their altruism allows this work to continue.
- Thanks to Claudette, Gail, Yonela, Thea, Siphiwe and their teams from the various Retina South Africa branches, for following up on patient information. Your efforts were hugely appreciated. Similarly, thanks to all the clinicians who refer patients to the research programme, in particular Drs Gratia-Marie Fischer, Ismail Mayet, Robyn Rautenbach and Olivia Read, for their clinical input on the patients diagnosed during this study.
- To Jacquie Greenberg, for always listening and being supportive. Working with you for the past 17 years has been a blessing. You and Adri Winckler both knew I could do this, before I did, and I thank you both for your faith in me.
- Thanks to Gameda Benefeld, who generates the maps for the retinal research programme and kindly created additional ones for me to present in my thesis.
- To the living legend Prof. Peter Beighton. You are an inspiration, and I am so appreciative you took the time to read my abstract and make suggestions.
- I thank all colleagues at the NEI and UCT who have been generous with their assistance and advice.
- This research was supported by Retina South Africa, the South African Medical Research Council, the National Research Foundation of South Africa, the University of Cape Town, and the Intramural Research Program (EY000546) of the National Eye Institute, National Institutes of Health.
- Last, but not least, I am forever grateful to my parents, family, and friends for their support, kindness, generosity, patience, and encouragement.

Table of Contents

List of Figures

List of Tables

Abbreviations

°C	degree Celcius
µL	microlitre
µM	micromolar
AAA	adenosine triphosphatases associated with diverse cellular activities
ad	autosomal dominant
adRP	autosomal dominant retinitis pigmentosa
AGVP	African Genome Variation Project
Ala	alanine
APEX	arrayed primer extension
ar	autosomal recessive
Arg	arginine
ARMD	age-related macular degeneration
arRP	autosomal recessive retinitis pigmentosa
Asn	asparagine
Asp	aspartic acid
ATP	adenosine triphosphate
BBS	Bardet Biedl syndrome
blast	basic local alignment search tool
bp	basepairs
cDNA	complementary DNA
CNV	copy number variant
COD	confirmation of diagnosis
CSNB	congenital stationary night blindness
CT	computed tomography
Cys	cysteine
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
DOA	dominant optic atrophy
EDTA	ethylenediaminetetra-acetic acid
EEG	electroencephalogram
ERDC	European Retinal Disease Consortium
ERG	electroretinography
ExAC	Exome Aggregation Consortium
F	forward primer
g	gram
Gln	glutamine
Glu	glutamic acid
Gly	glycine
HGMD	Human Gene Mutation Database
His	histidine
IDH	isocitrate dehydrogenase
Ile	isoleucine
IRD	inherited retinal disease
IS	inner segment
Kb	kilobase
kcal/mole	kilocalorie per mole
LCA	Leber congenital amaurosis
Leu	leucine
LOVD	Leiden Open variation database
Lys	lysine
M	matrix (score) output from pathogenicity tool

Abbreviations continued

MAF	minor allele frequency
MD	macular dystrophy
Met	methionine
mL	millilitre
mM	millimolar
mRNA	messenger RNA
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NCBI	National Centre for Biotechnology Information
ncRNA	non-coding RNA
ng	nanogram
NGS	next generation sequencing
NHLBI	National Heart, Lung, and Blood Institute
OCT	optical coherence tomography
ORF15	exon open reading frame 15
OS	outer segment
PCR	polymerase chain reaction
Phe	phenylalanine
P	prediction (decision) output from pathogenicity tool
pmols	picomoles
Pro	proline
QC	quality control
R	reverse primer
RNA	ribonucleic acid
RP	retinitis pigmentosa
RPE	retinal pigment epithelium
SA	South Africa
SAP/Exo	Shrimp Alkaline Phosphatase and Exonuclease I
SCA28	spinocerebellar ataxia type 28
Ser	serine
SNP	single nucleotide polymorphism
SNP ID	SNP identifier
STGD	Stargardt macular dystrophy
TBE	tris borate EDTA buffer
TCA	tricarboxylic acid cycle
Thr	threonine
Tris	tris[hydroxymethyl]aminomethane
Trp	tryptophan
Tyr	tyrosine
U	units
UCT	University of Cape Town
UK	United Kingdom
USA	United States of America
USH	Usher syndrome
UTR	untranslated region
Val	valine
vcf	variant call format
VMD	vitelliform macular dystrophy
WES	whole exome sequencing
xl	X-linked
xlRP	X-linked retinitis pigmentosa

Abstract

Background: Inherited retinal diseases (IRDs) constitute a group of clinically and genetically heterogeneous conditions which cause degeneration of retinal photoreceptor cells and result in visual impairment. Characterisation of the genetic basis of IRD is not only beneficial for the affected families, but also contributes towards understanding of the disease pathobiology. Investigations into the molecular basis of IRDs have been ongoing in South Africa (SA) for over 30 years, however the evaluation of reported genetic mutations has yielded low returns in certain populations. Indigenous southern Africans comprise a unique population group with distinct genetic diversity, providing a valuable resource for genetic discoveries; nonetheless, this population remains largely underrepresented in genomic studies. The aim of this investigation was to characterise the underlying genetic mutations in a cohort of indigenous African IRD patients.

Methods: The IRD registry in the Division of Human Genetics (University of Cape Town) was reviewed for causative mutations. Subsequently, upon identifying a mutation underlying Usher Syndrome in two indigenous African patients, an assay was designed to screen for this mutation in probands with different IRDs (n=170) and controls (n=51), and haplotype analysis was performed on mutation-positive individuals. The registry review additionally served to identify a suitable cohort for the application of next generation sequencing (NGS) technology. Whole exome sequencing (WES) was performed on genomic DNA samples from 56 individuals from 16 families. The WES data analysis strategy involved prioritisation of variants in reported and candidate IRD genes. Rare, co-segregating, pathogenic, exonic or splice variants were validated by Sanger sequencing. Custom TaqMan assays were designed to screen seven mutations, identified by WES, in 193 unrelated indigenous African probands with IRDs.

Results: A homozygous founder mutation, c.6377delC in *MYO7A,* was identified in 43% of the indigenous African patients with Usher syndrome, which is the most common cause of deaf-blindness. Targeted WES data analysis of all known IRD genes resulted in identification of the underlying genetic defects in six distinct genes (*RHO*, *PRPF3*, *PRPF31*, *ABCA4*, *CERKL*, and *PDE6B*) in six families. Taqman screening revealed four additional probands with identical homozygous mutations in *CERKL* and *PDE6B*. An X-linked gene (*RP2*) mutation was subsequently identified in an affected family with semi-dominant retinitis pigmentosa. Supplementary analysis of the X-linked *RPGR* ORF15 mutation hotspot (not adequately covered by WES) identified two mutations in three families. A novel IRD gene, *IDH3A*, was

found in one family by analysis of 22 putative candidate genes. The large number of variants in the remainder of the indigenous African exomes presented considerable challenges for identification of additional novel genes.

Discussion: The results of this project have important implications for IRD molecular diagnostic services in SA. Using WES, a genetic diagnosis was obtained for ~73% of the indigenous African cohort, and ~70% of the causative mutations identified were novel. This outcome emphasises the superiority of NGS-based approaches over genotyping-based microarrays which screen for IRD mutations previously reported in other (mainly European-derived) populations. The unexpected identification of mutations in known X-linked genes in four families highlighted key considerations for IRD WES analysis. Cascade screening of mutations identified in this study, across larger cohorts of unrelated probands, revealed the genetic cause of IRD in additional cases and the number of indigenous African families in the registry with a genetic diagnosis was effectively doubled. Members of these families can now opt for diagnostic, carrier, or predictive testing of familial mutations. Finally, the information obtained from this research contributes towards a better understanding of the genetic architecture of IRDs in SA.

Preface

This thesis is comprised of six chapters. Two original publications have been included with the permission of the Doctoral Degrees Board of the University of Cape Town (Appendix 1).

Chapter 1 is a general introduction of the topic, and presents the rationale and aims of this research project. Chapters 2–5 each contain brief introductions, methods, results and discussions.

Chapter 2 describes the review of the inherited retinal diseases registry at the Division of Human Genetics, which was performed to identify (a) common mutations for further investigation, and (b) a suitable cohort for whole exome sequencing.

Chapter 3 is a publication describing a founder mutation identified in (a) above.

The whole exome sequencing portion of the research project is covered in Chapters 4 and 5. Chapter 4 focuses on the analysis of variants in known inherited retinal disease genes, and is divided into three sections: 4.1 is a publication of the primary analysis; 4.2 and 4.3 present the findings from re-analysis of known genes. In Chapter 5, the variants in the remainder of the exome were interrogated, initially by prioritising a specific list of candidate genes (section 5.1). The additional strategies employed to identify causative mutations in the cohort, including supplementary screening, are described in section 5.2.

Chapter 6 provides a general overview of the research findings, which are examined and discussed in terms of their implications.

The American Journal of Human Genetics convention has been used for the reference list of this thesis. Mutation nomenclature is in accordance with the latest guidelines (Version 15.11) from the Human Genome Variation Society. However, the manuscript in Chapter 3 was published prior to version 2.120831 of these guidelines, which incorporated parentheses into protein variant descriptions lacking experimental evidence (i.e. based on DNA level data).

Chapter 1. Introduction

An estimated 161–259 million people worldwide suffer from visual impairment[1]. The causes of visual impairment are diverse; however, several heritable forms are known and make up a significant proportion of the disease burden. Inherited retinal diseases (IRDs), as a group, are one of the most clinically and genetically heterogeneous disorders, which complicates both patient diagnosis and molecular genetic research endeavours.

Notwithstanding the challenges, determining the molecular genetic basis of the various forms of IRD has been the focus of numerous investigations around the globe, with the ultimate goal being the development of therapeutic interventions. Consequently, the molecular basis of several inherited forms of visual impairment have been elucidated.

Whilst investigations into IRDs have been ongoing globally, including in South Africa (SA) for more than 30 years, the genetic basis of this group of disorders in indigenous Africans remains unclear. This is partly due to underrepresentation of this population group in the patient sample archive, and to earlier screening approaches being based on results obtained in the rest of the world, mainly in Europeans and North Americans. Characterisation of the molecular basis of disease in indigenous Africans, who comprise a unique population group with distinct genetic diversity, would therefore improve the understanding of these disorders and provide novel insights to this field of study.

In this introductory chapter the retinal architecture which underlies vision, the visual transduction process, the complex genetic landscape of IRDs (globally and in the context of SA), the research approaches used, as well as the evolution of African genomes is reviewed, to provide perspective for the research strategies taken in this dissertation.

1.1 The retina and vision

The retina is an intricate, laminated tissue lining the inside of the eye, which is essential for vision. The retina is considered to be part of the central nervous system (it develops from the embryonic forebrain[2]), and houses a complex neural network comprising 50–60 cell types[3] including six major neuron types: horizontal, bipolar, amacrine and ganglion cells, and rod and cone photoreceptor cells[4]. The function of the retina is to convert light energy (photons) into electrochemical signals that are conveyed via the optic nerve to the brain, resulting in the process of vision.

Light passes through the eye into the layers of the retina and excites the sensory photoreceptor cells, which are situated in the innermost layer of the retinal architecture (Figure 1.1). The phototransduction signalling cascade is initiated in these photoreceptors when light induces bleaching of the photopigment, a complex containing a chromophore derived from Vitamin A (11-*cis*-retinal) covalently bound to the protein opsin[5]. Photoexcitation results in isomerisation of 11-*cis*-retinal to all-*trans*-retinal, which then dissociates from opsin and triggers a series of biochemical events, causing closure of ion channels and ultimately resulting in hyperpolarisation of the cell[2,6]. Photoreceptors release the neurotransmitter glutamate in the dark; however, hyperpolarisation in response to light prevents this release[2]. The glutamate signal is processed by the bipolar, horizontal, and amacrine cells, and transmitted to the brain via the ganglion cells, the axons of which form the optic nerve[4,7].

Photoreceptors occur as two subtypes, the rods and the cones (described by their shapes). Rods mediate black-and-white vision and vision in low light conditions, whilst cones mediate perception of colour and high-resolution central vision (visual acuity). Rods outnumber cones 20:1 in the human retina[4]. While rods are evenly spread throughout the retina towards its periphery, cones are concentrated in the macula, an oval-shaped area at the centre of the retina. Human rods have the photopigment rhodopsin, whilst each cone expresses one of three cone opsins; S-opsin, M-opsin and L-opsin, which are blue-, green- and red- sensitive, respectively[4]. The photoreceptors therefore allow visual perception in all light conditions and trichromatic colour vision. Photoreceptor cells have an outer segment (OS) densely packed with membranous discs containing opsin, an inner segment (IS) which contains cellular metabolic components, and a cell body containing the nucleus[4,7] (Figure 1.1). Each IS and OS is joined by a narrow connecting cilium.

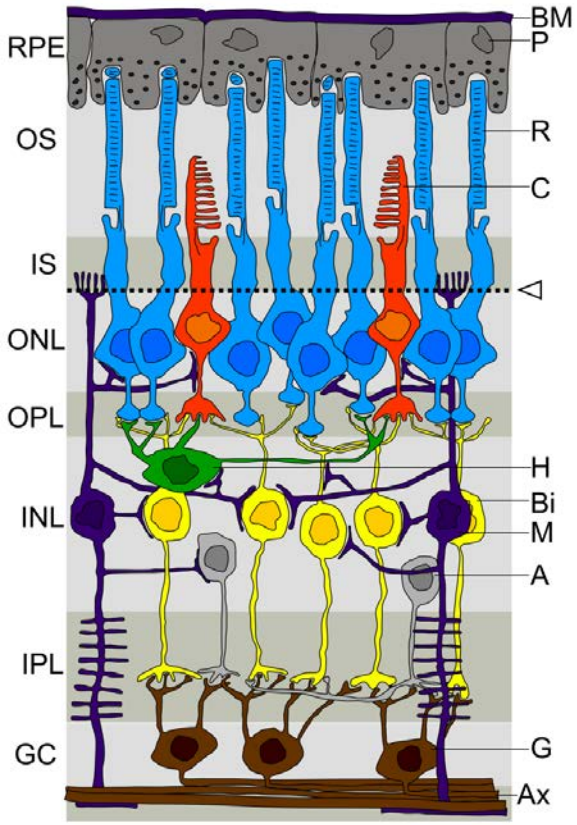


Figure 1.1 An illustration showing the layers of the retina involved in vision. Rod (R) and cone (C) photoreceptor outer segments (OS) are close to the retinal pigment epithelium (RPE). Light travels through the retinal layers (from the bottom of the image) to the OS where visual transduction is initiated in the membranous discs. Photoreceptor inner segments (IS) connect to the cell bodies and nuclei, which are located in the outer nuclear layer (ONL). Photoreceptor axons terminate in the outer plexiform layer (OPL), and synapse with the neurons in the inner nuclear layer (INL), namely the horizontal (H), bipolar (Bi), and amacrine (A) cells. The Muller (M) glial cells provide support and protection of the neurons. The neurotransmitter signal is relayed through the synapses of the inner plexiform layer (IPL) to the ganglion (G) cells in the ganglion cell layer (GC), the axons of which (Ax) form the optic nerve. The RPE contains melanin pigment and is separated from the choroidal capillaries of the eye by Bruch's membrane (BM). *Image by Peter Hartmann at de.wikipedia, edited by Marc Gabriel Schmid. Creating SVG version by Юкатан - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=27237061*

Photoreceptors are supported and maintained in four important ways by the retinal pigment epithelium (RPE), a monolayer of cells containing melanin pigment granules. Firstly, the RPE forms the outer blood-retinal barrier, suppling the

photoreceptors with nutrients and oxygen[8]. Secondly, for the retina to remain receptive to light, the 11-*cis*-retinal that is isomerised as the first step of visual transduction (as described above), must constantly be regenerated and recycled. The recycling of these Vitamin A derivatives occurs through a series of reactions in the photoreceptors and the RPE, called the "visual cycle" or the "retinoid cycle"[9]. All-*trans*-retinal is reduced to all-*trans*-retinol, which diffuses from the photoreceptors into the adjacent RPE. Esterification of the retinol with fatty acids occurs and (inside the RPE) the all-*trans*-retinal esters are converted, firstly into 11-*cis*-retinol and then 11-*cis*-retinal. The 11-*cis*-retinal diffuses back to the photoreceptor OS, and binds to opsin, regenerating the photopigment[9]. Thirdly, the RPE allows renewal of the OS content, since photoreceptors are terminally differentiated post-mitotic cells and cannot divide or regenerate. The OS tips are shed and phagocytosed by the RPE in a circadian manner (whereas rod shedding peaks in the morning, cone shedding peaks at night). Approximately 10% of the rod membranous discs are replaced in this manner, daily[9]. Finally, the melanin granules of the RPE absorb stray photons of light, preventing blurred vision and light damage to the retina[2].

The RPE, photoreceptors and other neurons of the retina function interdependently, in a delicately balanced homeostasis with high metabolic demand. Meta-analysis of retinal datasets[10] has yielded at least ~15,500 retinal and RPE genes (referred to as the '15K retinome'), of which ~13,000 have been confirmed in more than one study (i.e. the '13K retinome'). Whilst substantial overlap exists in gene expression between retina and other tissues, at least 5,000 genes were reportedly retina-specific[10]. Retinal tissue has been shown to have elevated splicing activity, expressing ~7-fold more major small nuclear RNAs (snRNA) and ~2-fold more minor snRNA, compared with other tissues (brain, heart, testis and skeletal muscle)[11]. The retina produced the greatest amount of constitutive, spliced/processed messenger RNA (mRNA) when 2,255 housekeeping gene transcripts were quantified across 31 human tissues[11]. RNA sequencing from three human retinas has revealed unprecedented transcript diversity[12], detecting over 80% of all exons in the human reference transcriptome, corresponding to 160,000 unique transcripts. This study implies that novel splicing events are enriched in the retina, with almost 80,000 novel splicing events identified (including novel exons, alternative splice sites and exon skipping events), resulting in novel exons in genes known to cause retinal diseases, as well as 116 putative novel genes. Additional RNA sequencing from temporal, macular and nasal regions of the retina has shown added complexity, with differences in gene expression levels between the peripheral

retina and macula regions[13]. This is perhaps to be expected given the distribution of different cells, namely the rod and cone photoreceptors, across the retina, as previously described. More recently (in the past year), expression analysis of a larger dataset of 50 human retinal samples has allowed a more precise estimation of the retinal transcriptome[14]. Pinelli et al confirmed 65% of known protein coding genes are expressed in the retina, and detected transcripts from ~13,000 to ~24,000 genes[14]. Notably, mitochondrial genes are amongst the most highly expressed genes in the retina[12,14].

The retina is thus a sophisticated, metabolically active tissue displaying cellular diversity and vast gene expression diversity, and as such is particularly susceptible to dysfunction.

1.2 Inherited retinal diseases

IRDs comprise of a large number of conditions causing progressive degeneration of the light-sensitive photoreceptor cells of the retina, which in turn leads to visual impairment (and can result in total blindness). IRDs are one of the most clinically and genetically heterogeneous groups of human disease.

IRD phenotypes can be broadly categorised on the basis of the primary photoreceptor type affected. Patients with primary rod cell loss experience nightblindness and reduction of the visual fields i.e. the loss of peripheral vision, resulting in tunnel-like vision[15–17]. These diseases can be congenital or have a later onset, and are either stationary (e.g. Congenital stationary night blindness, CSNB), or progressive (e.g. Leber congenital amaurosis, LCA, or retinitis pigmentosa, RP). Conversely, patients with ‘macular dystrophy’ (MD), or a primary loss of cone cells, experience loss of their central vision which can occur at an early age (e.g. the juvenile-onset Stargardt macular dystrophy, STGD) or later (e.g. age-related macular degeneration, ARMD)[15–17]. Initial loss of one photoreceptor subtype may or may not lead to secondary loss of the other subtype, worsening the individual’s prognosis. On the other hand, reduced penetrance can occur resulting in intra-familial disease variability[16,18–20]. Furthermore, whilst many IRDs are non-syndromic and limited to the loss of vision, they can also occur as part of a syndrome with multi-systemic dysfunction[15–17,21]. The phenotypic overlap between IRDs can impede a clinical diagnosis, and providing a patient prognosis can often be challenging. This is exacerbated by the fact that as the disease progresses the

diagnosis can change, which can be frustrating and confusing for patients. A definitive molecular genetic diagnosis therefore can verify and clarify a clinical diagnosis[22].

IRDs can be inherited in an autosomal dominant (*ad*), autosomal recessive (*ar*) or X-linked (*xl*) manner, although sporadic or isolated cases are common[15–17]. Rarer, atypical inheritance patterns include mitochondrial, digenic, tri-allelic and uniparental disomy cases[16,17]. Although most IRDs are monogenic, more than 200 different genes have been identified[23] (RetNet, https://sph.uth.edu/retnet/) and a significant overlap exists between the genes causing different IRDs (i.e. mutations in a single gene can result in different phenotypes)[15]. IRDs therefore demonstrate: (i) locus heterogeneity (a single phenotype, e.g. RP, can be caused by mutations in genes at different chromosomal loci, including dominant, recessive or X-linked loci), (ii) allelic heterogeneity (many different mutations can occur in each gene), (iii) genetic heterogeneity (the combination of locus and allelic heterogeneity), and (iv) clinical heterogeneity (a single gene or even a single mutation can result in different phenotypes in different individuals). A substantial amount of missing heritability further complicates the genetic diagnosis of IRDs. It is estimated that only 50–70% of cases (depending on geographical regions or populations) can be attributed to known genes[24], indicating that a considerable number of genes remain to be identified.

Identification of the specific mutation(s) causing an IRD affords several benefits to the patient. As described previously, due to the overlapping phenotypes of these diseases, a clear clinical (ophthalmological) diagnosis/prognosis may not always be possible. A genetic diagnosis on the other hand is of utility value and unequivocal, as diagnostic, predictive and carrier testing can be offered to family members. A genetic diagnosis, supporting a clinical one, may also influence the management of the disease/patient. Finally, improved molecular diagnosis in patients is important, given the number of clinical trials and treatments currently under investigation[25] for this group of disorders (http://www.clinicaltrials.gov/).

The degree of clinical and genetic heterogeneity of IRDs complicates not only the molecular diagnosis of patients, but also investigations of the pathogenic mechanisms of these disorders. For this reason, it is proposed that a candidate-free approach is required in order to better understand IRDs and characterise genes that are not yet identified. This is supported by the fact that, whilst many IRD genes are involved in phototransduction, the visual cycle, retinal metabolism or maintenance of retinal cells[26], some were not obvious candidates (in terms of expression or

function) at their time of discovery, instead encoding transcription factors[4] or ubiquitously expressed components of the spliceosome[11].

1.3 Next generation sequencing approaches in retinal research

Next generation sequencing (NGS) technologies are tools which allow the high-throughput capture of vast amounts of sequence data to study the genome, transcriptome and epigenome[27]. NGS-based genome analysis is highly applicable in scientific investigations into heterogeneous Mendelian diseases such as IRDs[28], as most (if not all) genes can be interrogated simultaneously. This hypothesis-free approach provides exceptional opportunity for gene discovery and identification of IRD mutations.

To this end, genomic NGS approaches have been employed increasingly in IRD research since 2010, with many of the earlier studies using 'targeted capture' methodologies to investigate panels of reported IRD genes. For example, ultra-high throughput sequencing of a single gene has been performed to characterise the mutation spectrum and allelic heterogeneity when: (a) a single gene is thought to be responsible for the majority of cases of a certain disorder[29]; or, (b) a newly-identified gene is investigated to determine mutation allele frequencies and contribution to the disease burden[30]. Moreover, extensive NGS strategies have allowed molecular diagnosis through the screening of panels of relevant genes. This approach has involved, in increasing order of scale, the targeted enrichment and subsequent sequencing of either, (a) exons of the genes known to cause a specific form of IRD e.g. RP[31–33] or LCA[34] or, (b) exons of many identified IRD genes in panel-based tests[23,35–43].

Subsequently, the improved capacity of whole exome sequencing (WES), with the strategic analysis of known IRD genes has become more commonplace. With WES, the entire coding portion of the genome is sequenced, together with predetermined desirable 5' and/or 3' ends of genes, and introns. The targeted analysis of IRD genes in WES data resulted in detection of causative mutations in as much as 83% of European families interrogated[44], with 50% of identified mutations being novel. Other population groups investigated in a comparable manner include Saudi Arabian[45], Chinese[46], Thai[47], and Israeli[48,49] with detection rates ranging from 49% to 83%, and the number of analysed IRD genes ranging from 60 to 226. Whilst it has

recently been suggested that panel-based diagnostic testing in IRDs is more effective than WES, due to better gene coverage[50] and cost considerations[51], the benefit of WES is that unresolved families can be re-analysed as novel IRD genes are reported, without re-designing gene panels and performing additional tests[52]. Moreover, WES has re-enforced the vast genetic (allelic) and clinical heterogeneity observed, as mutations in a gene previously thought to cause one specific IRD, are now reported for some cases of an entirely different IRD[53–55], thereby expanding the phenotypic spectrum associated with certain genes[56–58] and occasionally prompting clinical re-evaluation[59,60].

In addition to diagnostic applications, WES has facilitated huge strides in IRD research (unlike the initial targeted capture NGS panels), as *a priori* knowledge is not necessary to detect novel causative mutations. The approach has thus led to numerous gene discoveries and the elucidation of the roles these genes play in retinal functioning. More than 60 IRD genes have been identified via WES[27]. Many of these genes are involved in intraflagellar transport across photoreceptor cilia[61] or have assorted ciliary functions[62–68]; however, additional roles include photoreceptor differentiation and maintenance[69,70], fatty-acid transport in the retina[71], chaperone activity[72], ubiquitination[73], and protein glycosylation in neural development[58].

As anticipated, these large-scale NGS approaches generate “big data”, and the greatest challenge encountered with WES is therefore data analysis. The vast numbers of gene sequence variants obtained are thus interrogated via a series of prioritisation filters in order to distinguish pathogenic mutations from benign sequence alterations. The challenge of “big data” analysis is anticipated to be intensified when generated from older populations such as from Africa.

Indigenous Africans are generally underrepresented in genomic studies, and those with IRDs have not been interrogated using NGS platforms. However, WES of RP families in the United States of America (USA) has yielded a greater number of novel variants (both single nucleotide variants and small indels) in families of African ancestry compared with families of European ancestry[74]. In this study, the number of variants novel to the National Centre for Biotechnology Information (NCBI) Short Genetic Variations database (dbSNP) was reportedly >6-fold larger in a family of African American descent (n >2,500) than in Caucasian USA families (n ~400). Given that genome wide ancestry estimates show an average proportion of only ~73% African ancestry in African Americans[75] (who show substantial ancestry from west Africa[76]) , the exomes of indigenous Africans are expected to yield even more novel variants. Several thousand single nucleotide polymorphisms (SNPs) are

expected per individual, from which the causative pathogenic mutation must be identified. This is confounded by the fact that, on average, each individual carries 10 to 20 heterozygous recessive alleles for Mendelian disorders, and that the carrier frequency for recessive IRD mutations may be as high as 1 in 2[24], indicating that many heterozygous mutations identified by NGS may occur by chance and not be relevant to the disease observed. Indeed, novel candidate genes have been proposed[77] but later suggested to be false positives, prompting the recommendation of variant frequency thresholds in WES analysis for IRDs[78].

1.4 Inherited retinal diseases in South Africa

The reported prevalence of IRDs is approximately 1 in 3,500[24] in populations where some level of epidemiologic data is available. No data exists on the prevalence of this group of conditions anywhere in Africa. Nonetheless, using Statistics South Africa's 2011 population census i.e. ~51 million people in SA (http://www.statssa.gov.za/), one may extrapolate that approximately 14,500 individuals might be suffering from IRD-related visual impairment/blindness in SA, of which (taking population demographics into account) ~11,600 are estimated in the indigenous African population.

Research into IRDs in SA was initiated in 1985, when a questionnaire-based survey was utilised to determine the scope of RP in the country[79]. This was a joint endeavour between the Division of Human Genetics at the University of Cape Town (UCT) and the patient support group (then called the 'Retinitis Pigmentosa Foundation of South Africa'), and was prompted by the patients. It led to the identification of 63 families in SA containing 130 affected individuals of which ~36% were syndromic, whilst ~14%, 10% and 6% exhibited autosomal dominant (adRP), autosomal recessive (arRP) and X-linked (xlRP) inheritance patterns respectively. A large proportion (~27%) of cases were isolated, and the remaining ~8% had indeterminate inheritance patterns. It was determined that local families could benefit from the advances in genetic screening at the time, and that molecular diagnostic and carrier testing may be of some utility. The biological samples collected from these patients and their family members were thus stored in a biorepository with an associated electronic database or registry, generated in 1990–1991[80]. Within ten years, the registry had expanded to include other IRDs in addition to RP, namely MD (including STGD), LCA[81] and syndromic forms of IRD, for example Usher syndrome (USH, characterised by RP and hearing loss).

After the first ten years of study, an underrepresentation of IRD patients from indigenous African ethnolinguistic groups[82] (n=182) was noted and thought to be mainly due to a lack of access to resources (particularly in rural areas). Recruitment efforts in the last decade have therefore aimed to address the historical ascertainment bias so that the registry eventually reflects the population demographics of SA, however, this remains challenging. Patients from throughout SA are referred to the Division of Human Genetics via the support group, currently known as Retina South Africa, which also provides funding for research. Referrals are received from ophthalmologists and genetic counsellors, however these are in the minority despite attempts to raise awareness amongst professionals about the benefits of genetic testing for IRDs.

The ultimate goal of the research in SA has been to identify the causative genetic mutation in each of the families registered in the UCT IRD registry, hence the research programme has a strong translational and service component[83–88]. Traditionally, mutation analysis of candidate genes in South African patients with IRDs was performed, whereby patients were selected from the registry according to their clinical diagnosis, and an appropriate candidate gene was selected for screening based on international reports of mutation frequencies. This approach was arduous and time consuming, and the majority of South African patients remained lacking a clear molecular diagnosis until technologies advanced[89]. With the advent of microarray technology, commercially available mutation microarrays from Asper Ophthalmics in Tartu, Estonia (http://www.asperbio.com/asper-ophthalmics) have been employed since 2006. Currently, as part of the translational research programme, participants can opt for self-funded genetic screening with these microarrays which make use of arrayed primer extension (APEX[90]) technology to test specifically for previously reported mutations in candidate genes associated with specific IRDs[91–93]. The specific array used to screen each sample is selected based on the pedigree data and clinical diagnosis of that individual. Additionally, since November 2012, a limited number of samples have been subjected to NGS of a panel of 105 retinal candidate genes, by the Manchester Centre for Genomic Medicine, Central Manchester University Hospitals NHS Foundation Trust, in Manchester, in the United Kingdom (UK)[36,39]. It should be highlighted that both of the technologies, i.e. the microarray mutation panels and the NGS candidate IRD gene panels, are based on genetic findings from patients mainly in Europe and the USA.

Previous research has shown that the most prevalent genetic defects causing IRDs in the USA, the UK and Europe, are present in the South African patient cohort at an almost insignificant incidence, possibly indicating a novel gene pool for this group of disorders locally[94–97]. This is the case even in the Europe- and UK- derived Caucasian immigrant populations screened thus far in SA; identification of two novel loci for dominant IRDs[98–100], and the subsequent identification of the previously uncharacterised IRD-causing genes *PRPF8*[101] and *CA4*[102], in the South African Caucasian population, supported this notion. Furthermore, investigation of the indigenous African sub-cohort for the most common genes and mutations produced remarkably low returns[103]. The use of the microarrays emphasised that the mutation spectrum for IRDs in South Africans is dissimilar to that elsewhere, with many samples having either no mutations or partial mutation complements identified (unpublished data). This implied that South African patients either harbour novel mutations in previously identified genes, which are not being tested for with the arrays, or that their causative genes are novel. This concept is strengthened by the fact that indigenous Africans in SA exhibit vast genomic diversity[104,105].

1.5 Indigenous southern African genomes

Being the most ancient of all populations, Africans display vast genetic diversity[104,105] as a result of historical migration, population admixture, response to environmental change, and/or exposure to a plethora of infectious agents[106,107]. The 'Great Expansion' of humans out of Africa 45,000–60,000 years ago was accompanied by a loss of genetic diversity with linear correlation to the geographic distance from the origin in Africa[108].

The majority of sub-Saharan Africans speak ''Bantu'' languages, which are generally accepted to have originated from a core region in the north west of the African continent, specifically southern Nigeria and north-western Cameroon[109]. The term ''Bantu expansion'' refers to the movement of people approximately 5,600 years ago, across (west to east) and down (north to south) the continent. Evidence suggests migration accelerated around 2,000 years ago in south-central Africa (Malawi, Mozambique, Zambia and Zimbabwe), however a static frontier occurred upon arrival in SA[110] approximately 1,500 years ago[109]. The slower migration allowed interactions with (admixture with- and assimilation of-) the Khoisan hunter-gatherers already present in the region[110] who themselves are amongst the most genetically diverse of all humans[111]. Mitochondrial DNA evidence shows female

hunter-gatherers migrated into the Bantu communities at a greater rate than their male counterparts[110]. Eventually, southern African Bantu-speakers have diverged further into separate cultural and ethnolinguistic groups such as Sotho-Tswana, Xhosa and Zulu[110].

Today, there are two main Bantu-speaking groups in southern Africa, the Southwestern (subgroups R and K) and Southeastern (subgroup S) linguistic groups[109]. The R and K subgroups occupied the areas north of Namibia, whilst the S subgroup generally occupied the eastern part of present-day SA, Zimbabwe and Botswana, and are therefore most relevant to this study. The S subgroup of languages comprises the following ethnolinguistic groups: Sotho-Tswana, Venda and Nguni (which includes Xhosa and Zulu)[105]. Several languages in SA include the click sounds of Khoisan origin, which are absent in Bantu languages spoken nearby in Botswana, Mozambique and Zimbabwe[110]. Furthermore, the click sounds are restricted to certain ethnolinguistic groups within SA, namely Nguni and south-Sotho, indicating differing strengths of interactions between certain Bantu speakers and hunter-gatherers.

Two populations, the Luhya from east Africa (Kenya) and Yoruba from west Africa (Nigeria), were until recently the major African populations represented in the 1000 Genomes project[112]. It has been shown that the Luhya and Yoruba population groups are genetically diverse from the Bantu-speaking South Africans[105] and that the vast genetic diversity of African populations generally prevents them being used as proxies for one another in genetic studies[113]. This underscores the importance of the Southern African Human Genome Programme[114] and the African Genome Variation Project (AGVP)[115]. Nonetheless, a recent study on the genomic structure of indigenous southern African populations shows that the divergence of the Sotho-Tswana, Zulu, and Xhosa populations was relatively recent[116], suggesting that these people may be grouped together within genomic studies, to a greater extent than with other African populations.

Black South African individuals (referred to collectively hereafter as ‘indigenous Africans’), represent a derivation of the original Bantu expansion, and are the focus of this study as they provide a valuable resource to study genetic contributions to disease as well as for gene discovery[107,117].

1.6 Rationale

Placing the focus on indigenous Africans with IRDs was a novel approach in the study of this group of diseases. The description of IRD mutations in this population had the potential to identify ancient founder lineages, allow the development of novel diagnostic strategies, and impact on the clinical management of individuals suffering from previously unattributed IRDs.

Ultimately, the identification of novel IRD-associated genes, and the examination of the roles these genes play in IRDs, may render a better understanding of this group of diseases, potentially leading to novel approaches to disease intervention.

1.7 Aims and objectives

The aims of this project were to: (1) determine the genetic causes of IRDs in a cohort of indigenous Africans in SA, (2) translate these findings into diagnostic assays for the families concerned, and (3) contribute to the understanding of the biology of this group of conditions.

In order to achieve these aims, the objectives included:

i) Reviewing the UCT registry for previous screening results of indigenous Africans with IRDs in order to identify: a) potential causative mutations that had not previously been investigated, and b) an appropriate discovery cohort for WES

ii) Performing WES on indigenous African samples lacking reported mutations in genes known to be associated with their specific form of IRD, in order to identify the causative mutations.

iii) Investigating mutations identified either through prior screening or WES, which may account for large proportions of certain sub-cohorts (defined by clinical phenotype). These investigations included *in silico* pathogenicity investigations, assay design, family co-segregation analysis, screening in appropriate controls, haplotype analysis to ascertain an ancient founder lineage, and genotype-phenotype correlations. This information would determine whether specific assays should be developed as local diagnostic tools, for mutations having clear clinical utility.

iv) Researching putative novel genes, should they be identified, including: characterisation of the function of the gene(s), investigation of the biological pathway linking the gene to the IRD observed, and description of the associated phenotype (genotype/phenotype correlation).

Any diagnostic results achieved through this research were confirmed according to an established protocol, and translated to patients and clinicians through meaningful reports via professional genetic counsellors.

1.8 Ethics approval

Informed consent was obtained for all participants in this project, according to the Declaration of Helsinki (2008). Ethics approval for these genetic studies as well as the patient registry has been granted by the UCT Faculty of Health Sciences Human Research Ethics Committee (HREC REF. 226/2010, 768/2013 and 312/2014). The approval encompassed the Molecular Request Forms, Informed Consent forms and Confirmation of Diagnosis forms used in the IRD project recruitment process.

Chapter 2. Review of the registry: elucidating IRD trends in indigenous populations

At the start of this research project, the baseline status of genetic findings in the indigenous Africans with IRDs was established. This was achieved by reviewing the Division of Human Genetics IRD registry with regard to prior screening that had been performed in these individuals, and the results thereof. A list of mutations previously identified in this population group, that had not been fully investigated but which may represent common causes of disease, was compiled. The registry review also served to identify a cohort for WES, enriched for novel discovery.

2.1 Infrastructure of the registry

The biorepository and the associated electronic database or registry[80] archives biological samples collected from IRD patients and their family members from throughout SA over the past 30 years.

Biological specimens (venous blood or saliva, and/or genomic DNA) are accompanied by Molecular Request Forms including informed consent documentation (Appendix 2). The forms contain data such as date of birth, gender, ethnic origin/s, family history of disease, diagnosis and contact information. Ideally Confirmation of Diagnosis (COD) forms (Appendix 2) are completed by an ophthalmologist to provide detailed clinical information, and pedigrees are provided by skilled individuals, e.g. genetic counsellors or trained individuals from patient support group, Retina South Africa. Samples are de-identified using a unique code generated by the electronic database. The code is assigned as follows: the IRD abbreviation (listed below), space, family number (1 being the family number for the first proband recruited), period/full stop, individual number (assigned sequentially as new family members participate; 1 being the proband), and the first three letters of the person's first name. An example would be RPL 171.3LOU: IRD type LCA; 171st family recruited; 3rd person in this family; LOU from the person's first name. All members of a single family therefore have the same IRD code and family number,

but unique individual numbers and letters. Multiple biological sample aliquots from the same individual are coded alphabetically.

The IRD abbreviations assigned in the database are as follows (these are the product of coding for this group of disorders starting more than 30 years ago, and might be different if initiated today):

- RP (isolated or indeterminate inheritance): RP
- adRP: RPD
- arRP: RPR
- xlRP: RPX
- USH: RPU
- Other syndromic IRD, e.g. Bardet Biedl syndrome: RPO
- LCA: RPL
- MD, e.g. Sorsby fundus dystrophy, Best vitelliform macular dystrophy (VMD): RPM
- STGD: RPS
- ARMD (recruited post 2013, with a specific COD form): RPA

A map generated from the database (Figure 2.1) shows the widespread distribution across SA, of families with IRDs participating in the UCT research project.

The database is comprised of several linked Microsoft Access tables (Appendix 3) connected to each other by different reference numbers (i.e. the record number, family code and individual code). Cross-tab queries can be designed to extract relevant information across multiple tables. The data can then be exported and sorted manually in Microsoft Excel spreadsheets. All available data are captured on the registry in relevant coded fields, including clinical information, genetic screening history, if mutations have been identified, and whether genetic counselling and delivery of molecular results has occurred. Hardcopy documents, including (genetic) pedigrees, are stored in patient files.

When the causative mutation(s) are identified in a family, and validated according to a specific protocol[86], the family mode field in the registry is set as “diagnostic”. If suspected causative mutations have been identified, but this has not been confirmed, the family mode is “potential diagnostic”. Prior to this, all families are in “research” mode.

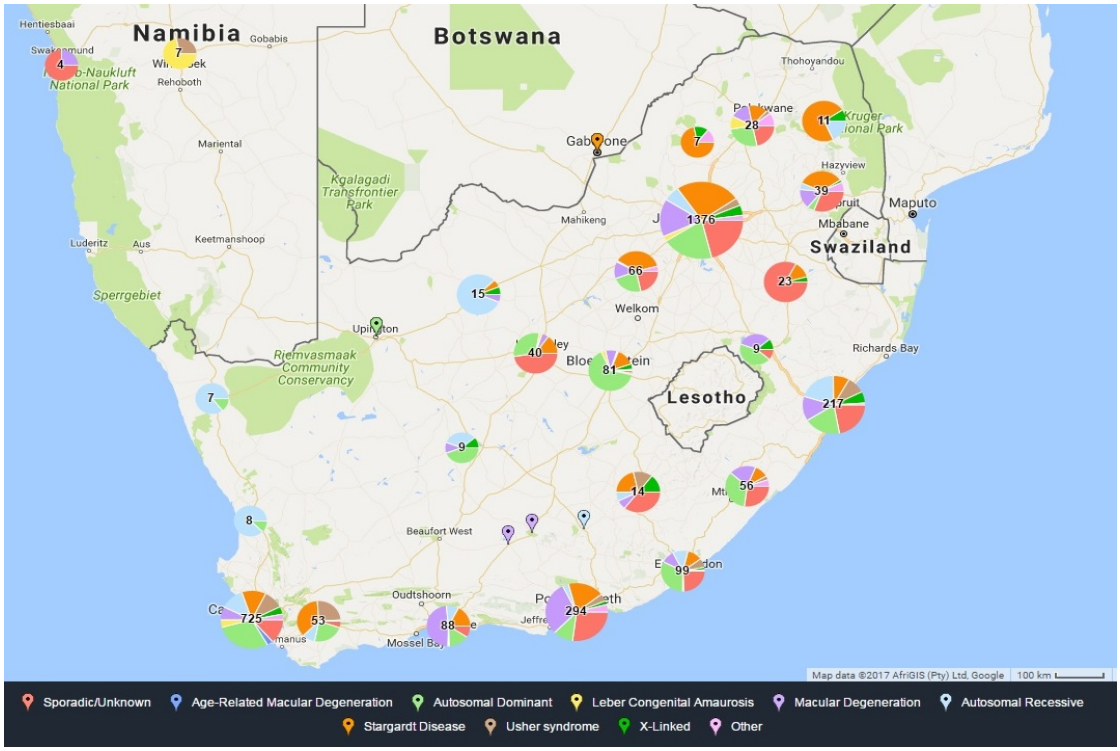


Figure 2.1 A map of South Africa illustrating the scope of individuals participating in the IRD research programme at UCT in 2017. Different IRDs are highlighted in different colours, as indicated on the bottom of the map. The autosomal dominant, autosomal recessive and X-linked labels refer to cases of retinitis pigmentosa. *Credit: Sr. G. Benefeld, Division of Human Genetics, University of Cape Town.*

An important registry field, in the context of this project, was whether samples had been sent overseas for testing by a specialist or commercial laboratory. Both overseas testing options offered in the translational research programme[86,87], namely participant-funded genetic screening using APEX[90] microarrays[91–93] or targeted capture NGS of retinal candidate genes[36,39], allow for rapid, more comprehensive tests than can be offered locally, as traditional screening of candidate genes is expensive, lengthy and arduous. The overseas testing field provided valuable information, both in cases where mutations are attributed (for translation into a diagnostic service and identification of founder mutations for future targeted screening) and where mutations were not identified (for highlighting those families with novel mutations and potential novel IRD candidate genes).

2.2 Analysis of potential common causative mutations in indigenous Africans

2.2.1 Methods

A cross-tab query was run to extract the following fields on the database (accessed 10 July 2014): family code, surname, ethnic group, family gene and mutation information, overseas testing history and family mode of analysis. This information was exported into the Microsoft Excel software program and filtered, in order to:

A) identify the indigenous African families with potential causative mutations (regardless of methodology used to identify the mutations)
B) analyse these causative mutations in the indigenous African families, with respect to frequency, predicted pathogenicity and whether they had previously been screened in significant numbers of relevant cases.
C) retrospectively evaluate the success rate of the Asper microarrays. For this, the filtering focused on all probands (regardless of ethnic group) screened using this particular methodology. All microarray testing performed from September 2005 to June 2014 was assessed to obtain:
    i) the number of families screened (counted as unrelated probands) using each of the different Asper panels (e.g. the adRP microarray, arRP microarray, etc.), and
    ii) the number of families subsequently diagnosed, using each microarray panel. Probands were considered diagnosed if the families were listed as being in ‘diagnostic’ or ‘potential diagnostic’ mode.

2.2.2 Results

This interrogation of the IRD database revealed that 33 distinct variants, in 16 different genes, had been identified in 45 unrelated indigenous African families (Table 2.1).

Table 2.1 List of the gene variants identified in indigenous African families by prior screening. These variants were considered for further investigation of potential founder effects. Distinct variants are in bold font when listed for the first time.

Family	Gene	Variant cDNA; Protein	Comment
RPS 1327	*ABCA4*	c.1804C>T; **p.(Arg602Trp)** (Homozygous)	Homozygous mutation identified in a single family.
RPS 1024	*ABCA4*	c.1927G>A; **p.(Val643Met)** & c.3602T>G; **p.(Leu1201Arg)**	p.(Val643Met) identified in two families. p.(Leu1201Arg) identified in seven families
RPS 263	*ABCA4*	c.2791G>A; **p.(Val931Met)** & c.3899G>A; **p.(Arg1300Gln)**	p.(Val931Met) identified in a single family. p.(Arg1300Gln) variant of unknown pathogenicity identified in four families.
RP 1143	*ABCA4*	c.2971G>C; **p.(Gly991Arg)** & c.3899G>A; p.(Arg1300Gln)	p.(Gly991Arg) mutation is identified in a single family. p.(Arg1300Gln) is a variant of unknown pathogenicity identified in four families.
RPS 1340	*ABCA4*	c.3602T>G; p.(Leu1201Arg) & c.2546T>C; **p.(Val849Ala)**	p.(Leu1201Arg) mutation identified in seven families. p.(Val849Ala) is a variant of unknown pathogenicity, identified in a single family.
RPS 145	*ABCA4*	c.4537dup; **p.(Gln1513Profs*42)** & c.6320G>A; **p.(Arg2107His)**	p.(Gln1513Profs*42) identified in a single family. p.(Arg2107His) identified in three families.
RPM 398	*ABCA4*	c.618C>G; **p.(Ser206Arg)** & c.3602T>G; p.(Leu1201Arg)	p.(Ser206Arg) identified in three families. p.(Leu1201Arg) identified in seven families
RPD 493 RPM 1171	*ABCA4*	c.618C>G; p.(Ser206Arg) & c.3899G>A; p.(Arg1300Gln)	p.(Ser206Arg) is identified in three families. p.(Arg1300Gln) is homozygous in RPD493, and heterozygous in RPM1171 (identified in four families in total, but is of unknown pathogenicity).
RPM 1004 RPM 1166 RPM 1168 RPS 1230	*ABCA4*	c.3602T>G; p.(Leu1201Arg). Other change unknown.	p.(Leu1201Arg) identified in seven families
RPM 200 RPS 377	*ABCA4*	c.6320G>A; p.(Arg2107His). Other change unknown.	p.(Arg2107His) identified in three families.
RPM 1075	*ABCA4*	c.1927G>A; p.(Val643Met). Other change unknown.	p.(Val643Met) mutation identified in two families.

Table 2.1 (continued)

Family	Gene	Variant cDNA; Protein	Comment
RPL 825	*AIPL1*	c.1126C>T; **p.(Pro376Ser)** & c.341C>T; **p.(Thr114Ile)**	Both are variants of unclear pathogenicity, identified in a single family
RP 1244 RPO 1000 RPO 1009 RPO 1341 RPO 879	*BBS10*	c.728_731delAAGA; **p.(Lys243Ilefs*15)** (Homozygous)	Homozygous mutation identified in five families
RPU 951	*CDH23*	c.4504C>T; **p.(Arg1502*)** (Homozygous)	Homozygous mutation identified in a single family.
RPU 968	*CDH23*	c.3625A>G; **p.(Thr1209Ala)**	Homozygous variant of unclear pathogenicity, identified in a single family.
RP 1157	*CRB1*	c.2234C>T; **p.(Thr745Met)** (Homozygous)	Homozygous mutation identified in a single family.
RPR 1113	*CRB1*	c.2506C>A; **p.(Pro836Thr).** Other change unknown.	Partial result. Heterozygous mutation identified in a single family
RPD 989 RPD 147	*CRX*	c.472G>A; **p.(Ala158Thr)**	Variant of unknown pathogenicity identified in two families
RP 1333	*EYS*	**c.3443+1G>T**. Other change unknown.	Partial result. Mutation identified in a single family
RPU 1338 RPU 564	*MYO7A*	c.6377delC; **p.(Pro2126Leufs*5)** (Homozygous)	Homozygous novel mutation, in two families.
RP 1235	*NR2E3*	c.1095C>T; **p.(Pro365=)** (Splice?)	Homozygous variant of unclear pathogenicity, identified in a single family.
RPS 1252	*OTX2*	c.664G>A; **p.(Gly222Arg)**	Variant of unclear pathogenicity
RPU 80 RPU 868	*PCDH15*	c.5601_5603delAAC; **p.(Thr1867del)**	Homozygous mutation in RPU80 and heterozygous in RPU868 (other mutation unknown).
RPD 914	*RHO*	c.329G>A; **p.(Cys110Tyr)**	Mutation identified in a single family.
RPD 1018	*RHO*	c.541G>A; **p.(Glu181Lys)**	Mutation identified in a single family.
RPD 45	*RHO*	c.1040C>T; **p.(Pro347Leu)**	Mutation identified in a single family.
RP 1251	*RP2*	c.358C>T; **p.(Arg120*)**	Mutation identified in a single family.
RPR 397	*RPE65*	c.1301C>T; **p.(Ala434Val)**. Other change unknown	Partial result. Mutation identified in a single family

Table 2.1 (continued)

Family	Gene	Variant cDNA; Protein	Comment
RPD 41	*TOPORS*	c.2556_2557delGA; **p.(Glu852Aspfs*20)**	Mutation identified in a single family.
RP 1049 RPU 765	*USH2A*	c.2137G>C; **p.(Gly713Arg)** & c.12445T>C; **p.(Trp4149Arg)**	Both are variants of unclear pathogenicity
RP 1154 (also coded as RPR 1103)	*USH2A*	c.10712C>T; **p.(Thr3571Met)** (Homozgyous)	Homozygous Mutation identified in single family.
RPR 917	*USH2A*	c.5932C>T; **p.(Pro1978Ser)**. Other change unknown.	Partial result. Mutation identified in a single family

Variants were not considered high priority for screening in additional probands, if the pathogenicity was unclear, or if they were identified in a single family. Furthermore, certain variants had already been screened in a large proportion of the relevant indigenous African sub-cohort, therefore further screening was not warranted. For example, to date *ABCA4* is the only gene known to cause *ar* STGD (RetNet https://sph.uth.edu/retnet/, accessed 27 January 2016), and 40 indigenous African probands had already been screened using the *ABCA4* array, including 16 of the 21 patients clinically diagnosed with STGD. The *ABCA4* mutations occurring in multiple families, namely p.(Val643Met) (n=2), p.(Leu1201Arg) (n=7), p.(Arg2107His) (n=3) and p.(Ser206Arg) (n=3), would thus have already been screened for in the majority (~76%) of indigenous Africans with STGD, via the process of array testing. Similarly, all samples from the IRD registry with a suspected diagnosis of Bardet Biedl syndrome (BBS, an *ar* syndrome with RP or rod-cone retinal dystrophy and other associated features such as obesity, polydactyly and/or developmental delay) are referred to the National Health Laboratory Service (NHLS) at Groote Schuur Hospital in Cape Town for diagnostic testing of the *BBS10* founder p.(Lys243Ilefs*15) mutation, which is common in indigenous Africans with this clinical diagnosis[118].

Three mutations were identified in genes known to be associated with USH in multiple families, namely p.(Arg1502*) in *CDH23*, p.(Thr1867del) in *PCDH15* and p.(Pro2126Leufs*5) in *MYO7A*. The *CDH23* and *PCDH15* mutations were included on the USH array since its initial use in the translational programme in 2010, and hence 17/20 (~85%) indigenous African USH samples had already been tested for

those mutations using the USH array. However, the *MYO7A* mutation was only added to the USH array in 2013, and all but two of the 17 samples had already been tested using the array by that time. The *MYO7A* mutation was identified in family RPU 1338 using the array, and in family RPU 564 using targeted capture NGS, in 2013. A sample from family RPU 564 had been tested with the array in 2010, however the mutation was not detected as it was not included in the panel at that stage. The p.(Pro2126Leufs*5) mutation in *MYO7A* was thus determined to be the only variant warranting further investigation as a high priority (Chapter 3).

The efficacy of microarray technology in identifying causative IRD mutations was evaluated and compared between different ethnic populations (Table 2.2). Samples from Caucasians outnumbered those from any other ethnic group, and different arrays were used a range of times (and with different success rates) in the various population groups, therefore no conclusions can be drawn from these data. However, by simply comparing the Caucasian vs. the indigenous African samples as a whole, 115 of 280 Caucasian patients (41.1%) were diagnosed through the use of microarray technology, compared to 14 of 109 Indigenous African patients (i.e. 12.8%).

Table 2.2 The success rate of various Asper microarrays in the diagnosis of IRD patients from different population groups. The percentage of probands diagnosed using the microarrays are presented, together with the total number of families (i.e. unrelated probands) tested.

Array	Caucasian	Indigenous African	Mixed ancestry	Indian	Unknown ethnicity
ABCA4	57.38% (n=175)	9.76% (n=41)	28.57% (n=7)	33.33% (n=6)	0% (n=2)
arRP	7.14% (n=56)	13.04% (n=23)	0% (n=1)	0% (n=7)	33.33% (n=3)
LCA	25.93% (n=27)	0% (n=2)	20% (n=5)	0% (n=2)	0% (n=2)
USH	0% (n=11)	17.64% (n=17)	0% (n=1)	0% (n=1)	0% (n=1)
Best VMD	66% (n=5)	-	-	-	-
xlRP	50% (n=2)	33.33% (n=3)	-	-	-
adRP	0% (n=4)	13.04% (n=23)	-	-	-

2.3 Selection of a discovery cohort for whole exome sequencing

2.3.1 Methods

A cross-tab query of the database (accessed 4 June 2014) was run to extract the individual code, ethnic group, diagnosis code (i.e. affected, at risk, carrier, parent, spouse, unaffected or query), family code, family mutation, overseas testing history, and family mode of analysis. The data was exported into the Microsoft Excel software program. Duplicate family numbers were removed, so that a single individual was included per family. Individuals were then stratified based on ethnic group (Caucasian, Indigenous African, Indian, Mixed Ancestry and Other), and subsequent systematic selections were performed exclusively on the indigenous African sub-cohort.

Only samples with comprehensive prior screening and no mutation identified were considered for WES, in order to increase the likelihood of novel findings. Comprehensive screening was considered to be either: i) microarray analysis, or ii) targeted capture NGS. The indigenous African cohort was therefore examined regarding the type of screening performed and the current analysis mode (i.e. 'research', 'potential diagnostic' or 'diagnostic'). Only families in research mode were selected. Those families were then examined (using the database and paper-based patient records) to determine: 1) how many individuals per family had submitted biological material for analysis, 2) the location of these individuals in the structure of the family pedigree, and (3) any supporting clinical information.

DNA samples from all family members were retrieved from the biorepository, and sample integrity was confirmed by agarose gel electrophoresis. DNA was visualised using 1–2% weight/volume (w/v) agarose gels containing 0.5–1g agarose (SeaKem® LE, Lonza, Switzerland), 50mL of 1X Tris Borate EDTA (TBE) buffer and 5µL SYBR® Safe DNA Gel Stain (ThermoFisher Scientific, MA, USA). A GeneRuler™ 100bp DNA Ladder Plus (ThermoFisher Scientific, MA, USA) was included on each gel for comparison purposes.

DNA extractions were performed on additional archived blood samples if necessary, using a salting-out method[119]. Intact DNA was quantified using the Nanodrop ND-1000Spectrophotometer (ThermoFisher Scientific, MA, USA).

Families were only included in the discovery cohort for WES if sufficient DNA was available for at least three family members. It was anticipated that a large number of sequence variants would be identified through WES; hence, careful selection of the individuals designated for WES within each family would improve the data analysis and elimination of benign variants. Variants co-segregating with disease in a family could be prioritised for validation, therefore, where possible, the most distantly-related individuals (who would share fewer variants) within each family were selected, in order to facilitate the prioritisation of candidate variants. A minimum of three samples per family were thus selected for WES, and remaining samples from each family would be used for validation and for co-segregation analysis of putative mutations.

2.3.2 Results

At the time this research project was initiated, a total of 3,215 individuals in 1,416 families had agreed to participate in the IRD research programme. The cohort for WES was selected following the stratification of 271 indigenous (Black) South African families who had biological material as well as demographic and clinical data stored in the registry (Figure 2.2). Four probands were excluded from the cohort as their visual problems were not due to a definitive IRD, but possibly due to another cause, for example physical damage or infection. Of the 267 probands remaining, 226 (85%) were still in research mode, 28 (10%) were in diagnostic mode and 13 (5%) were in potential diagnostic mode. The probands in diagnostic mode and potential diagnostic mode were excluded from the WES cohort, as their causative mutations were already known, or at least suspected.

The majority (76%, n=172) of the 226 probands in research mode had a diagnosis of RP; 128 were simplex cases, 25 had adRP, 15 had arRP and four had xlRP. A further 42 probands (19%) had some form of MD: 27 had a diagnosis simply of MD, 14 had STGD, and one had age-related MD. A small proportion (4%, n=10) of the cohort had USH. Two probands (1%) had LCA.

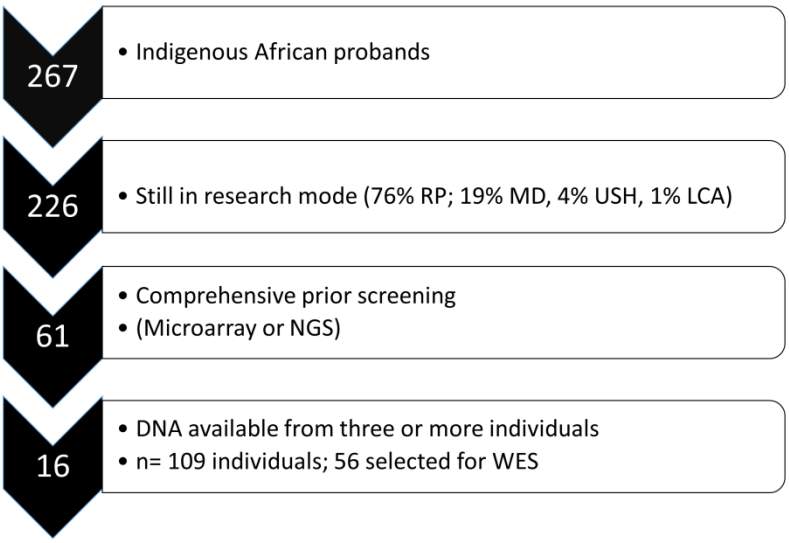


Figure 2.2 A diagram showing the criteria applied for selection of the discovery cohort, and the numbers of indigenous African IRD families at each phase of stratification.

The prior screening performed on samples from the 226 probands in research mode was assessed. In total, 65 comprehensive screens had been performed on 61 probands (four samples had more than one type of screen performed, due to a change of clinical diagnosis with progression of the disease and/or unclear Mendelian pattern of disease inheritance). The vast majority of the screening tests (97%; n=63) were microarrays, with only two samples screened by NGS.

The 61 probands remaining were evaluated regarding the biological material available from additional family members, both affected with IRD and unaffected. Genomic DNA samples, extracted from blood or saliva, were ultimately available from 109 members of 16 indigenous African families with a range of IRDs (Table 2.3).

Table 2.3 Summary of the information pertaining to the 16 families selected for WES.

Family	No affected/ No samples total	Micro-array used	Clinical Information	Pedigree inheritance pattern
RPD 55 (ZULU)	10 affected/ 18 total	adRP	RP. Age of onset 16–40 yrs.	*ad*
RPD 94 (ZULU)	9 affected/ 12 total	adRP	RP. Age of onset 10–52 years. Some relatives have slow disease progression, others rapid.	*ad*
RPD 334/ RPX 54/ RPX82 (XHOSA)	8 affected/ 16 total	adRP & xlRP	RP. Age of onset 2–8 years.	*ad*, or *xl* with occasional carrier manifestation (more males affected)
RPD 391 (TSWANA)	3 affected/ 4 total	adRP	Diffuse RP. Severe progression and profound vision loss in different relatives.	*ad*
RPD 401 (TSWANA)	2 affected/ 3 total	adRP	Signs of RP.	*ad* with reduced penetrance.
RPD 799 (UNKNOWN)	3 affected/ 4 total	adRP	RP. Age of onset 15–37 years.	*ad* with reduced penetrance.
RPD 1001 (XHOSA)	3 affected/ 4 total	adRP	RP. Age of onset 14 years in one person, rest unknown.	*ad*
RPD 1010 (XHOSA)	7 affected/ 9 total	adRP	Diffuse RP, Cataracts a common feature. Age of onset 13–40 years.	*ad*
RPD 1339 (ZULU)	2 affected/ 3 total	adRP	Diffuse RP, Age of onset 2–5 years.	*ad*
RPM 1167 (TSONGA / NDEBELE)	2 affected/ 3 total	LCA & ABCA4	2007 diagnosis: MD, 2012 diagnosis: cone-rod dystrophy, possible LCA. Age of onset 7–11 years.	*ar*
RPR 397 (SHANGAAN)	2 affected/ 3 total	arRP	Diffuse RP. Very early age of onset (birth–1year).	*ar*
RPR 624 (UNKNOWN)	2 affected/ 3 total	arRP & ABCA4	RP. One relative had age of onset 6 years, rapid progression till 18 years, then disease remained stable. Rest unknown.	*xl*

Table 2.3 (continued)

Family	No affected/ No samples total	Micro-array used	Clinical Information	Pedigree inheritance pattern
RPR 917 (XHOSA)	3 or 4 affected/ 4 total	arRP	RP. Age of onset 14–36 years.	*ar* or *xl*
RPD 1005 (XHOSA / SOTHO)	4 affected/ 6 total	arRP	RP. Age of onset 40 years in one person, rest unknown.	*ar*
RP 583 (XHOSA)	12 affected/18 total	xlRP	RP. Age of onset 3–59 years. Some have diffuse RP. One relative has possible macular involvement.	*ad*
RPM 537 (VENDA)	2 affected/ 3 total	ABCA4	STGD, Age of onset 14–18 years.	*ar*

The ethnolinguistic breakdown of these families was as follows: five Xhosa, three Zulu, two Tswana, one Shangaan, one Venda, one Tsonga/Ndebele, one Xhosa/Sotho, and two unknown. Two of the 16 families had been clinically diagnosed with *ar* MD (one had a subsequent diagnosis of cone-rod dystrophy or LCA, the other family had STGD) and the remaining 14 families had RP. Of the 109 individual samples available, 56 were selected for WES, with the remainder to be used for subsequent validation of the WES results.

2.4 Discussion

Review of the registry confirmed the premise of this research project; indigenous Africans were underrepresented in the South African IRD cohort and remained largely undiagnosed. Although a total of 1,416 families had agreed to participate in the UCT IRD research programme at the start of this investigation, only ~19% of them (n=271 families) were of indigenous African ethnicity. In 2003, a review following the first 10 years of the programme reported 182 indigenous African families[82], and despite an effort to make the programme more inclusive, the subsequent decade of recruitment resulted in only 89 additional families joining the IRD project at UCT by 2014. This indicated that many South African IRD patients do not benefit from a molecular diagnosis, which could be advantageous to their healthcare. Furthermore, the broader implication is that the true burden of disease and epidemiology of IRDs in this county remains unknown, given that the population

of SA is ~80% indigenous African (http://www.statssa.gov.za/). Awareness must continue to be raised among healthcare professionals, particularly those in rural areas and urban medical centres which serve a large proportion of the local indigenous population. The special schools for the visually impaired could provide an additional source of families, however several ethical considerations exist for this approach (mainly surrounding the recruitment of children). Nonetheless, it should be noted that in the two years since this particular investigation was initiated (from June 2014 to June 2016), 68 new cases joined the research programme of which 40 (58.8%) were indigenous African, showing progress in the effort to reduce bias.

The first goal in reviewing the registry was to ascertain whether any potential founder effects existed for IRD in the group of interest. A founder effect is the loss of genetic variation occurring when a new population is established by a small number of individuals (i.e. the 'founding individuals') from a larger population[120]. SA has local and immigrant populations, some of which exhibit significant admixture[121–123] and others which remain largely endogamous[124–126]. It can be postulated that founder mutations arose in indigenous black South Africans during the Bantu expansion[109,116]. Founder effects have been reported in indigenous South Africans with other disorders[127,128], including BBS[118], so it was prudent to investigate this population group with respect to founder mutations causing IRD.

The registry review indicated that 33 distinct variants in 16 different genes had been identified in the indigenous African sub-cohort. However, the majority of variants occurred in single families, had uncertain pathogenicity, or had already been screened in the majority of relevant cases and, thus, did not warrant further investigation in the context of this project. Although *ad* founder mutations have been reported[129], including for IRDs[130], none of the mutations identified in genes causing *ad* IRD (namely *RHO* and *TOPORS*) occurred in more than a single indigenous African family, each. These mutations could be screened in a larger ethnically-matched cohort in the future, to ascertain allele frequency, however this would be unlikely to deliver large returns, given the vast genetic diversity reported in this population group[104,105,116]. A more successful approach may be screening of the homozygous mutations identified in single families. Of these, the *CRB1* p.(Thr745Met) mutation would be of greater interest than the homozygous mutations in the USH and BBS genes as these have already been interrogated appropriately. However, ultimately a single variant, p.(Pro2126Leufs*5) in *MYO7A*, was identified in the homozygous state in two unrelated families but had not been screened in the local USH cohort. It was therefore thought to potentially be a

founder mutation in this population group and warranted further investigation as a high priority (Chapter 3).

The efficacy of microarray technology was problematic to evaluate, as the majority of samples in the biorepository (and therefore screened using the arrays) are from Caucasian patients. Furthermore, different panels have been used to different extents, and with different success rates, in the various population groups. For example, the *ABCA4* panel is the most frequently used array; driven by the translational diagnostic service aspect of the programme[87,88,131]. No definitive conclusions can thus be drawn from these data. However, since the microarrays test for reported mutations identified predominantly in patients of European/Caucasian origin, it was perhaps unsurprising that a trend was observed that indigenous African patients are not diagnosed as frequently as Caucasian South African patients using this methodology.

The second goal in reviewing the registry was to select an appropriate cohort for WES. Sixteen families met the established criteria, namely no mutation had been identified following comprehensive screening, and sufficient intact DNA samples were available from at least three family members. This cohort was thus selected to be enriched for novel variant discovery in reported IRD genes (Chapter 4), and was utilised to identify putative novel IRD genes (Chapter 5).

Chapter 3. A founder mutation in *MYO7A* underlies a significant proportion of Usher syndrome in indigenous South Africans: implications for the African diaspora

Original Publication

Roberts L, George S, Greenberg J, Ramesar RS. A founder mutation in *MYO7A* underlies a significant proportion of Usher syndrome in indigenous South Africans: implications for the African diaspora. Invest Ophthalmol Vis Sci. 2015;56:6671–6678. DOI:10.1167/iovs.15-17028

NCBI Pubmed link: http://www.ncbi.nlm.nih.gov/pubmed/26469752?dopt=Abstract

Article/journal link: http://iovs.arvojournals.org/article.aspx?articleid=2463759

Nature of the publication: Original full article

Journal/Publisher: Investigative Ophthalmology & Visual Science, an Association for Research in Vision and Ophthalmology (ARVO) journal. ISSN: 1552-5783

Candidate contribution: Conceived the project and strategy, designed the experiments, performed mutation screening, performed statistical analysis of haplotypes with respect to the control data, analysed data and drafted the manuscript.

Co-authors contributions:

SG: Performed mutation screening, genotyping, manual haplotype construction, and edited the manuscript.

JG: Assisted with patient follow up, delivery of results, confirmation of clinical and pedigree data and edited the manuscript.

RR: Supervised the research and edited the manuscript.

A Founder Mutation in *MYO7A* Underlies a Significant Proportion of Usher Syndrome in Indigenous South Africans: Implications for the African Diaspora

Lisa Roberts, Siddiqah George, Jacquie Greenberg, and Raj S. Ramesar

UCT/MRC Human Genetics Research Unit, Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

Correspondence: Lisa Roberts, Room N3.17, Level 3, Wernher and Beit North Building, Institute of Infectious Disease and Molecular Medicine, University of Cape Town Faculty of Health Sciences, Anzio Road, Observatory, 7925, Cape Town, Western Cape, South Africa; lisa.roberts@uct.ac.za.

Submitted: April 6, 2015
Accepted: August 30, 2015

Citation: Roberts L, George S, Greenberg J, Ramesar RS. A founder mutation in *MYO7A* underlies a significant proportion of Usher syndrome in indigenous South Africans: implications for the African diaspora. *Invest Ophthalmol Vis Sci*. 2015;56:6671–6678. DOI:10.1167/iovs.15-17028

Purpose. Research over the past 25 years at the University of Cape Town has led to the identification of causative mutations in 17% of the 1416 families in the Retinal Degenerative Diseases (RDD) biorepository in South Africa. A low rate of mutation detection has been observed in patients of indigenous African origin, hinting at novel genes and mutations in this population. Recently, however, data from our translational research program showed two unrelated indigenous African families with Usher syndrome (USH), with the same homozygous *MYO7A* mutation. Therefore, the extent to which this mutation contributes toward the disease burden in South Africa was investigated.

Methods. Cohorts of unrelated indigenous South African probands with different RDD diagnoses were tested for the *MYO7A* c.6377delC mutation. Familial cosegregation analysis was performed for homozygous probands, clinical data were evaluated, and SNP haplotypes were analyzed.

Results. This homozygous *MYO7A* mutation underlies a remarkable 43% of indigenous African USH cases investigated in this study, the majority of which (60%) were diagnosed clinically with Type 2 USH. All homozygotes shared a common haplotype. This mutation does not appear to cause nonsyndromic vision loss.

Conclusions. Of interest is the origin of this common mutation relevant to the Bantu population migration into southern Africa. Further investigation of the phenotype may elucidate the disease biology, and perhaps reveal a larger cohort with the same mutation, with which to assess the impact of environmental and genetic modifiers and evaluate therapeutic trials.

Keywords: Usher syndrome, African, founder mutation, *MYO7A*

Usher syndrome (USH) is characterized by vision and hearing loss, and is the most common cause of deaf-blindness.[1,2] It is an autosomal recessively inherited group of disorders, divided into three major clinical subtypes that are differentiated by the severity of hearing loss and the presence of vestibular dysfunction. Vision loss due to retinitis pigmentosa (RP) is a hallmark of all three USH subtypes. In addition to clinical heterogeneity, Usher syndrome displays genetic heterogeneity, with 12 causative genes identified to date.[3]

Usher syndrome type 1 is the most severe form, exhibiting profound congenital hearing loss and vestibular dysfunction, and prepubertal onset of progressive RP. To date, six genes have been associated with USH type 1, namely *CDH23*, *CIB2*, *MYO7A*, *PCDH15*, *USH1C*, and *USH1G*.[4] Usher syndrome type 2 is less severe, characterized by congenital hearing loss that is moderate to severe, with normal vestibular functioning and a later RP onset. Mutations in three genes, namely *DFNB31*, *GRP98*, and *USH2A*, cause type 2 USH.[1,2] Type 3 USH is characterized by variable onset of RP and hearing loss, as well as varying degree of vestibular dysfunction. Two genes are associated with USH type 3, namely *CLRN1*[2] and *HARS*,[5] with a third gene (*ABHD12*) being associated with a variant of this subtype.[6]

A number of USH protein interactions (or interactomes) have been reported,[4,7] which function in the development and maintenance of stereocilia hair bundles of the inner ear and which also colocalize in the synaptic layer, connecting cilium and the calyceal processes of the photoreceptors of the retina. The exact function of these protein interactomes is not known, as mouse models (of mutated USH genes) have little or no retinal phenotype, but they may have a role in protein trafficking between the inner and outer segments of the photoreceptors, as well as synaptic function of these sensory neurons.[4]

Due to the genetic and clinical heterogeneity, the large size of the genes, and multiple isoforms underlying the syndrome, identifying the molecular basis of USH in affected South African families using traditional candidate gene screening methods has been challenging. Technologic advances, such as the development of microarrays and next generation sequencing, have significantly improved the turnaround time and success rates of genetic mutation screening for inherited retinal degenerative diseases. As part of our translational research program in South Africa.[8,9] families can opt for genetic screening using microarrays (Asper Biotech Ltd., Tartu, Estonia)[10] or for whole exome sequencing of 105 retinal candidate genes (through the

Manchester Centre for Genomic Medicine, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK).[11] A review of data obtained from the Asper USH array and whole exome analyses revealed two unrelated indigenous (Black) South African USH probands with a homozygous c.6377delC (p.Pro2126Leufs*5) mutation in *MYO7A*, which subsequently was confirmed by cycle sequencing. This mutation has been reported previously only recently (to the best of our knowledge) in the compound heterozygous state, together with p.Arg1240Trp, in a single Caucasian individual from the United Kingdom with USH type 1.[12] The mutation subsequently was added to the panel of mutations on the Asper Usher Microarray; however, many South African samples had been tested by that time and, therefore, it was deemed necessary to rescreen appropriate samples for this mutation. The p.Pro2126Leufs*5 mutation is predicted to truncate the 2215 amino acid MYO7A protein by 86 amino acids (3.88%).

The gene *MYO7A* was the first USH gene identified[13] and has since been recognized as the most frequent cause of USH type 1, which is the most severe form of USH.[14,15] The protein MYO7A is an unconventional myosin expressed in multiple epithelial cell types,[16] including the RPE, where it functions in the light-dependent localization of the visual cycle enzyme, RPE65.[17] It is also expressed in the photoreceptor calyceal processes and cilia, and the stereocilia, together with other USH proteins.

The identification of two unrelated indigenous South African patients (and their respective families) with the same homozygous *MYO7A* mutation warranted further investigation in this population, particularly with respect to a potential founder effect and the clinical manifestation of c.6377delC. Phenotypic variation has been reported previously in USH, with mutations in *USH2A* causing nonsyndromic RP[18,19] and *MYO7A* mutations causing nonsyndromic[20] deafness. Founder mutations have been reported previously in indigenous South Africans,[21,22] and determining whether this USH mutation occurs on the same haplotype in all affected individuals would show a probable founder effect, having diagnostic implications.

Methods

Cohort

Affected individuals and their family members were recruited from throughout South Africa as part of a Retinal Degenerative Disorders (RDD) research project, established in the Division of Human Genetics at the University of Cape Town in 1990. Biological material (genomic DNA extracted from venous blood) is archived in the RDD registry, together with demographic and clinical information. Informed consent is obtained from all RDD research participants according to the tenets of the Declaration of Helsinki (2013), and ethics approval for this specific study was obtained from the institutional Human Research Ethics committee (HREC/REF: 312/2014).

A cohort of 12 unrelated indigenous South African probands with confirmed clinical diagnoses of USH (regardless of the clinical subtype), but no genetic diagnoses, were tested for the c.6377delC mutation. A further six indigenous South African probands with RDD and some hearing loss, but no clinical confirmation of USH, also were tested and considered to be a "query USH cohort." Testing also was performed on samples from three probands of Mixed Ancestry with confirmed clinical diagnoses of USH (who likely share similar ancestry with the indigenous South African individuals, as Bantu-speaking Africans are a major ancestral contributor to this admixed population[23]).

Familial cosegregation analysis was performed where possible for probands carrying the homozygous c.6377delC mutation, to ensure that the mutation cosegregated with disease within the families. A total of 51 indigenous South African population controls, who had not specifically been assessed for the absence of RDD, also was screened for the mutation. No detailed population data were available for these controls, other than that they are of indigenous South African origin. To investigate possible phenotypic variation, a cohort of 10 indigenous South African probands with nonsyndromic autosomal recessive RP (arRP) and 107 indigenous South African simplex RP cases also were tested.

c.6377del C Assay Design

The *MYO7A* Transcript variant 1 NCBI (NM_000260, accessed November 2013) and Ensembl (ENSG00000137474, ENST00000409709, accessed November 2013) sequences were used to compare the region containing exon 47 and to ensure that the entire exon was included in the assay. Transcript variant 1 was selected as it encodes the longest isoform. Primers were designed to span exon 47 and at least 50 bp of intronic sequence flanking the exon. The primers used were Forward 5′ GCAACAGGAGAGGCTGACTTTATC 3′, Reverse 5′ GTGGCTAGGAGGGCTTGTG 3′.

The primers were used to amplify a 287 bp fragment under standard 25-μL PCR conditions: 100 ng genomic DNA, 0.4 μM forward and reverse primer, 200 μM dNTPS, 1X Colourless GoTaq Reaction Buffer, and 0.5U GoTaq DNA Polymerase (Promega, Madison, WI, USA). Cycling conditions were as follows: 95°C – 5 minutes, 30 cycles of (94°C – 30 seconds, 61°C – 30 seconds, 72°C – 40 seconds), 72°C – 7 minutes.

The c.6377delC mutation creates an *Hpy*188III restriction enzyme recognition site 131 bp into the amplicon, resulting in 156 bp and 131 bp fragments in samples with the homozygous mutation, and 287, 156, and 131 bp fragments in samples with the heterozygous mutation. Restriction digests with *Hpy*188III were performed as follows: a standard 20-μL reaction containing 10 μL PCR product, 1X NEB CutSmart Buffer, and 5U NEB *Hpy*188III enzyme (New England Biolabs, Ipswich, MA, USA) was incubated at 37°C for 3 hours. The digest products were subjected to electrophoresis on a 3% agarose gel containing SYBR Safe DNA Gel Stain (Applied Biosystems by Life Technologies, Woolston, Warrington, UK) and visualized under UV light using a UVIpro Gold transilluminator (UVItec, Cambridge, UK). A heterozygous and homozygous sample (confirmed by cycle sequencing) were included as controls.

Haplotype Analysis

To examine the possibility of a founder effect in the indigenous South African population (which would have diagnostic implications), it was necessary to determine whether the mutation exists on a common haplotype in the cases. Additionally, haplotyping of African controls could determine the origin of the mutation in South Africa, and the potential contribution to USH disease burden in other African countries. The mutation of interest is located at position 76924019 on chromosome 11 (human genome build GRCh37.p13/hg19). More than 1 Mb of the sequence flanking the mutation (584 709bp 5′ and 502 265bp 3′) was interrogated for informative SNPs with minor allele frequencies > 20% in the Luhya and Yoruba 1000 Genomes dataset.[24] Three single nucleotide polymorphisms (SNPs) were selected for haplotyping analysis in mutation-positive families: rs6592706, rs948972, and

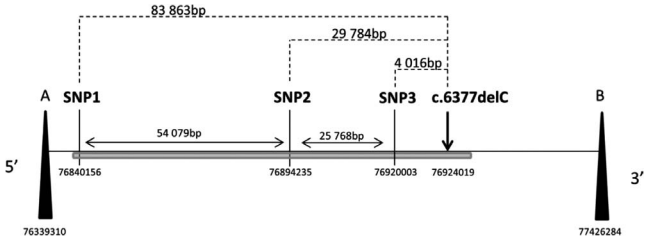


FIGURE 1. Diagrammatic representation of the >1 Mb region of chromosome 11 (between [**A**] and [**B**]) interrogated for informative SNPS, and the three SNPs selected for genotyping (SNP1 = rs6592706, SNP2 = rs948972, and SNP3 = rs11237122). The mutation location is shown as a *bold vertical arrow*. The numbers below the *horizontal solid line* indicate the position (bp) on the chromosome, the numbers above the *horizontal arrows* show the distance between adjacent SNPS, and numbers above the *dashed line* show the distance between each SNP and c.6377delC. The *MYO7A* gene is represented as a *gray box* (located from 76839310–76926284 bp).

rs11237122. All selected SNPS are 5′ of the mutation; no suitable SNPs were identified 3′ of the mutation (Fig. 1).

Primers were designed to amplify a 250, 210, and 247 bp fragment spanning rs6592706, rs948972, and rs11237122, respectively. Polymerase chain reactions (PCR) were performed in a 25-μL volume under standard PCR conditions: 100 ng genomic DNA, 0.4 μM forward and reverse primer, 200 μM dNTPS, 1X Colourless GoTaq Reaction Buffer and 0.5U GoTaq DNA Polymerase (Promega). Cycling conditions were as follows: 95°C – 5 minutes, 30 cycles of (94°C – 30 seconds, annealing temperatures (Ta) – 30 seconds, 72°C – 40 seconds), 72°C – 7 minutes. Primer sequences, annealing temperatures, and assay information are listed in Table 1.

The G allele of rs6592706 creates a *Bsr*BI restriction enzyme recognition site, generating fragments of the following sizes for the various genotypes: A/A, 250 bp; A/G, 250, 159, and 91 bp; G/G, 159 and 91 bp. Restriction digests were performed as follows: a standard 20 μL reaction containing 10 μL PCR product, 1X Tango Buffer, and 2U *Bsr*BI enzyme (ThermoScientific, Waltham, MA, USA) was incubated at 37°C for 3 hours. The digest products were subjected to electrophoresis on a 3% agarose gel containing SYBR Safe DNA Gel Stain (Applied Biosystems by Life Technologies) and visualized under UV light.

The C and G alleles of rs948972 were distinguished by cycle sequencing following purification of PCR products; 8.9 μL PCR products were purified by 1U FastAP Shrimp Alkaline Phosphatase (ThermoScientific) and 2U Exonuclease I (ThermoScientific) in a 10-μL reaction that was incubated as follows: 37°C – 60 minutes, 75°C – 15 minutes, 95°C – 5 minutes. Cycle sequencing was performed in a 20-μL reaction containing the 10 μL purified PCR products, 1 μM reverse primer, 1X Sequencing Buffer, and 1X BigDye Terminator v3.1 Reaction Mix (Applied Biosystems by Life Technologies). Cycling conditions were: 96°C – 5 minutes, 30 cycles of (96°C – 30 seconds, 50°C – 15 seconds, 60°C – 4 minutes). Sequencing products were purified by ethanol precipitation and resuspended in 10 μL Sabax water (Adcock Ingram, Johannesburg, South Africa), after which 5 μL sequencing reaction was loaded, together with 8 μL Hi-Di Formamide (Applied Biosystems by Life Technologies), onto a 3130xl Genetic Analyser (Applied Biosystems by Life Technologies).

The T allele of rs11237122 creates a *Pfl*MI restriction enzyme recognition site, generating fragments of the following sizes for the various genotypes: C/C, 247 bp; C/T, 247, 138, and 109 bp; T/T, 138 and 109 bp. Restriction digest were performed as follows: a standard 20-μL reaction containing 10 μL PCR product, 1X Buffer R and 2U *Pfl*MI enzyme (ThermoScientific) was incubated at 37°C for 3 hours. The digest products were subjected to electrophoresis on a 3% agarose gel containing SYBR Safe DNA Gel Stain (Applied Biosystems by Life Technologies) and visualized under UV light.

Haplotypes of affected individuals with the homozygous c.6377delC mutation were constructed manually based on segregation within the families, and then compared between families, and compared to control data; individual genotypes for the 3 SNPs were obtained from 97 Luhya and 88 Yoruba individuals in the 1000 Genomes dataset.[24] Linkage disequilibrium testing and haplotype analysis was performed subsequently, and χ^2 and Pearson's *P* values were calculated, using the SHEsis online program (available in the public domain at http://analysis.bio-x.cn/myAnalysis.php).[25,26]

RESULTS

Frequency of c.6377del C

After the initial identification of two unrelated indigenous USH probands homozygous for c.6377delC, another four unrelated homozygotes were identified by screening 12 additional probands. Thus, a total of six homozygotes was identified in the total cohort of 14 confirmed indigenous South African USH cases (42.86%).

No mutation-positive individuals were identified in the additional cohorts screened (Table 2). Furthermore, the mutation was not present in the Luhya or Yoruba individuals in the 1000 Genomes dataset,[24] nor was it present in 200 chromosomes from Zulu individuals sequenced as part of the African Genome Variation Project (AGVP).[27]

USH Mutation-Positive Families: Cosegregation and Haplotyping Analysis

The six identified families came from the following ethnolinguistic groups: one Sotho, two Zulu, two Xhosa, and one unknown indigenous South Africans. The families live in two large, geographically distinct provinces in South Africa, and each family lives in a different town, indicating that consanguinity was unlikely. Familial DNA was available for three of the six probands, and familial cosegregation analysis confirmed that the mutation cosegregated with disease within these families (Fig. 2). Genotyping of 3 SNPs (>83kB from c.6377delC) in all available familial DNA samples also showed the total of 10 homozygotes all shared a common haplotype (Fig. 2). The clinical and demographic information pertaining to the affected individuals is presented in Table 3.

TABLE 1. Primer Sequences, PCR Annealing Temperatures (Ta), and Assay Information for the Genotyping of Three SNPs in Families Carrying the c.6377delC Mutation

SNP	Fwd Primer, 5′→3′	Rev Primer, 5′→3′	Ta, °C	Assay
rs6592706	cttgaaggtggtctagttctca	atgtggattcaacagggcca	60	+ *Bsr*BI
rs948972	agtccaagctcacagaggag	acactcctgtctgcctgatc	60	Cycle sequencing
rs11237122	tgctgtactttggccctgaa	gcagaatctcgaagtcagagg	58	+ *Pfl*MI

Table 2. Results of Mutation Screening in South African Case and Control Cohorts

Cohort Screened	No. Individuals	No. Chromosomes	No. Chromosomes With c.6377delC
Indigenous South African USH	14	28	12
Indigenous South African query USH	6	12	0
Mixed Ancestry USH	3	6	0
Indigenous South African arRP	10	20	0
Indigenous South African simplex RP	107	214	0
Indigenous South African controls	51	102	0
South African Zulu control AGVP data*	100	200	0
Luhya control 1000 Genomes data*	97	194	0
Yoruba control 1000 Genomes data*	88	176	0

* Control data from publically available datasets.

The SHEsis Linkage Disequilibrium test was used to calculate Lewontin's D' and r^2 between each pair of the 3 SNPs in the 1000 Genomes control data from 97 Luhya and 88 Yoruba individuals. The results indicated that there is no linkage disequilibrium among the three SNPs in the separate control populations or when they are combined, and, therefore, this is not a block of low haplotype diversity in these African population groups.

We performed χ^2 calculations for genotype frequencies (Table 4) of the 3 SNPs in the 185 African controls and 6 unrelated, c.6377delC homozygous probands affected with USH. These statistical tests showed that the cases, separate control groups, and combined controls showed no deviation from Hardy Weinberg Equilibrium. The χ^2 analysis showed a significant difference between the Luyha and Yoruba controls in the frequency of rs6592706 (Pearson's $P = 0.017293$); however, this was no longer significant when the critical value was set to $P < 0.01666$ (Bonferroni correction for testing of 3 SNPs). There was no significant difference in the frequencies of rs948972 and rs11237122 between the Luhya and Yoruba controls (Pearson's $P = 0.122203$ and 0.527811, respectively) and, therefore, these control groups were combined for subsequent comparison with the cases. The results showed that for rs6592706 and rs11237122 there are significant differences in the genotype frequencies between the cases and controls, which remained significant after Bonferonni correction ($P < 0.01666$).

Subsequently, haplotypes were reconstructed and haplotype frequencies compared between the 6 unrelated cases and 185 controls (Table 5). The SHEsis analysis of 370 control and 12 case haplotypes gave a Global χ^2 of 77.993309, while $df = 6$, and a Pearson's P value was 9.27e-015. Haplotypes occurring with a frequency less than 0.05 were excluded. The results indicated a significant difference in the frequency of the ACT haplotype in the African controls compared to the six probands with Usher syndrome.

Discussion

Our experience has shown that testing for known candidate genes and mutations for RDDs (as configured through the Asper Biotech Ltd. Arrays[28]) has a good yield with our subcohort of Caucasian subjects, with 115 of 280 patients (41.1%) having their mutation(s) identified through the use of various Asper arrays, but a rather low return in our indigenous African patients, in whom only 14 of 109 patients (i.e., 12.8%) have a genetic diagnosis after microarray screening (results not shown). This is understandable, since most of the testing arrays are based on mutations generally identified in cohorts of patients of European/Caucasian origin. Our identification of a homozygous *MYO7A* mutation in two USH patients of indigenous African origin was initially surprising, but its emergence as the cause of a large proportion (42.86%) of

Table 3. Clinical and Demographic Data From 10 Affected c.6377delC Homozygous Individuals in 6 Families

DNA Code	Ethnic Group	Age of Onset, y	Clinical Diagnosis and Information
RPU 318.1UNA	Unknown	1	USH Type 2
RPU 340.3JEA*	Xhosa	Congenital	USH Type 2; VA: L = 6/60; R = 6/12 at age 43. Congenital hearing loss.
RPU 340.4PAT*	Xhosa	Congenital	USH Type 2; VA: L = 6/48; R = 6/48 at age 40. Congenital hearing loss.
RPU 340.5PET*	Xhosa	Congenital	USH Type 2; VA: L = 6/24; R = 6/60 at age 37. Congenital hearing loss.
RPU 340.8SIG*	Xhosa	Congenital	Congenital hearing loss, developed RP.
RPU 564.1FRA	Zulu	11	USH Type 2
RPU 954.1ELI	Xhosa	6/7	USH Type 2
RPU 1136.1MAN*	Zulu	Congenital	USH Type 1. Congenital hearing loss, RP developed at 6–8 y, progressive loss of day and night vision.
RPU 1136.2LIN*	Zulu	Congenital	USH Type 1. Congenital hearing loss, RP developed at 6–8 y, progressive loss of particularly night vision.
RPU 1338.1ISI	Sotho	Congenital	USH Type 1. VA: L = 6/18; R = 6/12 at age 30. Night blindness onset at 21 y. Slow disease progression, pigment, attenuation, pale discs.

VA, visual acuity; L, left eye; R, right eye.
* Multiple members of Family RPU 340 and Family RPU 1136.

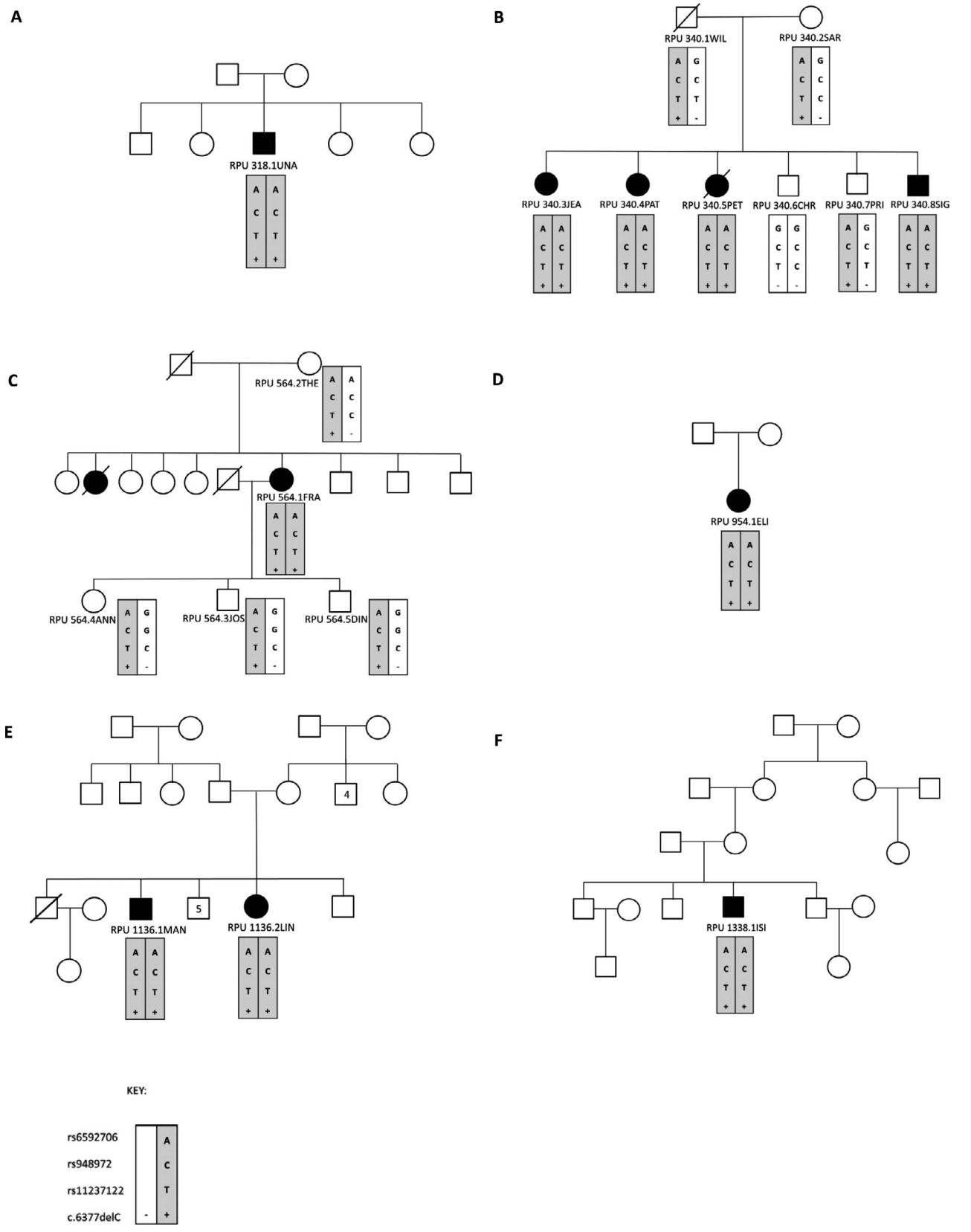


FIGURE 2. (A–F) Pedigrees of six c.6377delC mutation–positive individuals and their family members. *Squares* represent males and *circles* represent females, with *shaded symbols* indicating individuals affected by USH. All individuals from whom DNA was available are indicated by a DNA code. Key: The 3 SNPs and c.6377delC mutation are indicated in the order shown, with the founder haplotype *shaded*.

Table 4. Genotype Frequency Comparison for the 3 SNPs Between 185 African Controls and 6 Homozygous c.6377delC Probands

SNP	rs6592706			rs948972			rs11237122		
Genotype Count, Frequency	A/A (freq)	A/G (freq)	G/G (freq)	C/C (freq)	C/G (freq)	G/G (freq)	C/C (freq)	C/T (freq)	T/T (freq)
Case, $n = 6$	6 (1.000)	0 (0.000)	0 (0.000)	6 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	6 (1.000)
Luyha control, $n = 97$	14 (0.144)	45 (0.464)	38 (0.392)	46 (0.474)	44 (0.454)	7 (0.072)	28 (0.289)	49 (0.505)	20 (0.206)
Yoruba control, $n = 88$	23 (0.261)	46 (0.523)	19 (0.216)	53 (0.602)	27 (0.307)	8 (0.091)	20 (0.227)	45 (0.511)	23 (0.261)
Total control, $n = 185$	37 (0.200)	91 (0.492)	57 (0.308)	99 (0.535)	71 (0.384)	15 (0.081)	48 (0.259)	94 (0.508)	43 (0.232)
χ^2/df		21.320929/2			5.073668/2			17.951681/ $df = 2$	
Pearson's P		2.35e-005*			0.079116			0.000126*	

* $P \leq 0.001$.

unselected USH cases from this population group was remarkable. Several possibilities could explain the existence of this relatively frequent mutation underlying USH in this part of Africa, including a mutational hotspot at this nucleotide position of the *MYO7A* gene, evolutionary advantage conferred, or genetic drift. The c.6377delC mutation has been reported only once previously (in the compound heterozygous state, in a Caucasian individual from the United Kingdom with type I USH),[12] and the mutation has not been detected in the 1000 Genomes dataset, indicating that this codon is not particularly susceptible to mutagenesis. It is unlikely that the mutation confers an advantage to carriers, as heterozygous mutations in *MYO7A* can cause autosomal dominant hearing loss.[29,30] Furthermore, in the present study, USH patients from different ethnolinguistic subgroups, namely Xhosa, Zulu, and Sotho, were identified with this homozygous mutation, negating genetic drift as a cause of the mutation.

The majority of sub-Saharan Africans speak "Bantu" languages, which are believed to originate from a core region in the north west of the African continent, specifically Nigeria and western Cameroon.[31] The "Bantu expansion" refers to the movement of people approximately 5600 years ago, across and down Africa. Bantu speakers arrived in South Africa approximately 1500 years ago, where they diverged further. Today, there are two main Bantu-speaking groups in South Africa, the Southeastern (subgroup S) and Southwestern (subgroups R and K) groups. The S subgroup of languages comprises the following ethnolinguistic groups: Sotho-Tswana, Venda and Nguni (which includes Xhosa and Zulu).[32] The different ethnolinguistic groups described in this affected cohort, therefore, represent a derivation of the original Bantu expansion. All mutation-positive patients in the present study, whether Xhosa, Zulu, or Sotho, shared a common haplotype spanning >83kB of sequence which is 5′ of the mutation. This denotes that c.6377delC is a founder mutation that arose in speakers of the S-subgroup of Bantu languages before their divergence.

Table 5. Haplotype Reconstruction and Frequencies for rs6592706, rs948972, rs11237122 in 6 USH Cases and 185 African Controls

Haplotype	Case (freq)	Control (freq)	χ^2	Pearson's P
A C C	0.00 (0.000)	70.59 (0.191)	2.947	0.086060
A C T	12.00 (1.000)	38.47 (0.104)	77.993	1.09e-018*†
A G C	0.00 (0.000)	35.20 (0.095)	1.313	0.251864
A G T	0.00 (0.000)	20.74 (0.056)	0.742	0.389170
G C C	0.00 (0.000)	53.38 (0.144)	2.106	0.146678
G C T	0.00 (0.000)	106.56 (0.288)	5.060	0.024500
G G C	0.00 (0.000)	30.83 (0.083)	1.135	0.286700

* The haplotype associated with c.6377delC.
† $P \leq 0.001$.

The haplotype is imputed to be present in the 1000 Genomes data at a frequency of 10%, although the mutation is not present on this haplotype in the African populations in east Africa, that is, the Luhya of Kenya and in west Africa, that is, the Yoruba of Nigeria. Furthermore, there is no linkage disequilibrium of the three SNPs in these two populations, implying that the c.6377delC mutation arose on the haplotype after the Bantu speakers expanded southwards in Africa. A limitation of this study is the sole use of the Yoruba and Luhya data as proxy control populations, with which to investigate the haplotype frequency. It has been shown that these populations are genetically diverse from the Bantu-speaking South Africans,[32] and that proxy populations may not be applicable due to the vast genetic diversity of African populations.[33] This study highlights the paucity of genetic data from indigenous South Africans, as no local population frequency data were available for the SNPs of interest in this study in the SNP dataset recently made publically available by Ramsay et al.[32] This underscores the importance of the Southern African Human Genome Programme (SAHGP)[34] and the Africa Genome Variation Project.[27] Nevertheless, the use of the Yoruba and Luyha datasets was valuable in showing that this founder mutation arose as a more recent event: post-Bantu expansion but predivergence into the different ethnolinguistic groups of South Africa.

The 51 indigenous South African population controls used in this study to establish the frequency of the mutation comprises individuals speaking the S-group of Bantu languages (including Xhosa, Zulu, and Sotho), although a complete and defined ethnolinguistic breakdown of these samples is unavailable. Nonetheless, a recent study on the genomic structure of indigenous southern African populations shows the relatively recent divergence of the Sotho-Tswana, Zulu, and Xhosa populations, suggesting that these may serve as proxies for one another, to a greater extent than the Luhya and/or Yoruba.[35] Thus, this cohort was appropriate to compare the frequency of the mutation between cases and controls, especially when supplemented by the 200 Zulu chromosomes of the AGVP data.

Interestingly, 6 of the 10 homozygous mutation–positive patients had been clinically diagnosed with type 2 USH, whereas *MYO7A* mutations previously generally have been associated with the more severe type 1 USH. This is not the first report of *MYO7A* mutations causing type 2 USH,[14] but the observation is rare. The milder phenotype diagnosed could be due to the fact that the mutation affects the C-terminal FERM domain, and less than 4% of the protein is predicted to be truncated. A mouse study of a different *MYO7A* mutation (albeit a splice variant), affecting the same C-terminal FERM domain, showed tissue-dependent mRNA instability[36]; truncated mRNA in the ear appears to be degraded by nonsense-mediated decay, whereas mRNA expressed in the retina is not. The majority of our

homozygous cohort reported congenital onset of USH, yet the clinical diagnosis of USH type 2 indicated no vestibular dysfunction is present and RP onset is later.

Mutations in *MYO7A* are associated with nonsyndromic hearing loss[20] and it would be interesting to evaluate whether this particular founder mutation contributes to the burden of hearing loss in indigenous South Africans and other African populations. Our screening indicates that this mutation is not associated with nonsyndromic RP, which is not surprising given the lack of prior reports correlating *MYO7A* mutations with RP, and the tissue-specific protein effects reported.[36]

Providing a genetic diagnosis to a family means that individuals within that family can elect to have diagnostic, carrier, or predictive testing. Genetic testing, therefore, provides patients and their relatives with more accurate risks of developing disease, upon which they can base their informed life decisions and reproductive choices. The identification of this founder mutation will allow targeted genetic testing, based on clinical diagnosis and patient ethnicity, which will reduce the costs of genetic testing and facilitate a rapid test turnaround time. Although the c.6377delC mutation was not present in 51 indigenous South African controls, screening larger numbers of unaffected controls from Xhosa, Zulu, and Sotho populations, and combining these results with data from the AGVP[27] and SAHGP,[34] will provide mutation carrier frequency information that could be useful for genetic counseling purposes and risk calculations. Further screening in larger cohorts of Mixed Ancestry USH patients and nonsyndromic deafness patients is warranted as these patients likely share some ancestry with indigenous South Africans. Furthermore, there is potential for identifying a large number of individuals with the same pathogenic mutation, which could facilitate detailed genotype-phenotype investigations[37] and studies of phenotypic modifiers. Finally, identification of these indigenous South African patients with a *MYO7A* mutation is important given the development of UshStat, the *MYO7A* gene replacement therapy[38] currently in trials (https://clinicaltrials.gov/ identifiers NCT01505062 and NCT02065011).

Acknowledgments

Tommy Carstensen, Deepti Gurdasani, and Manjinder Sandhu, from the Wellcome Trust Sanger Institute and the University of Cambridge, analyzed the African Genome Variation Project sequencing data for the presence of the c.6377delC mutation. Alvera Vorster, from the University of Cape Town, provided valuable input and advice regarding the SNP genotyping and subsequent haplotyping analysis, Collet Dandara gave statistical input, Gameda Benefeld assisted in confirming the clinical, demographic and pedigree data, and the clinical genetics team delivered results to the patients. The authors thank the patients from throughout South Africa with inherited retinal degenerative disorders for participating in the research project.

Supported by the University of Cape Town, the Medical Research Council of South Africa, and the lay support society Retina South Africa.

Disclosure: **L. Roberts**, None; **S. George**, None; **J. Greenberg**, None; **R.S. Ramesar**, None

References

1. Millán JM, Aller E, Jaijo T, Blanco-Kelly F, Gimenez-Pardo A, Ayuso C. An update on the genetics of usher syndrome. *J Ophthalmol*. 2011;2011:417217.
2. Yan D, Liu XZ. Genetics and pathological mechanisms of Usher syndrome. *J Hum Genet*. 2010;55:327–335.
3. RetNet, the Retinal Information Network: Summaries of genes and loci causing retinal diseases. Available at: https://sph.uth.edu/Retnet/sum-dis.htm. Accessed September 8, 2014.
4. Cosgrove D, Zallocchi M. Usher protein functions in hair cells and photoreceptors. *Int J Biochem Cell Biol*. 2014;46:80–89.
5. Puffenberger EG, Jinks RN, Sougnez C, et al. Genetic mapping and exome sequencing identify variants associated with five novel diseases. *PLoS One*. 2012;7:e28936.
6. Eisenberger T, Slim R, Mansour A, et al. Targeted next-generation sequencing identifies a homozygous nonsense mutation in ABHD12, the gene underlying PHARC, in a family clinically diagnosed with Usher syndrome type 3. *Orphanet J Rare Dis*. 2012;7:59.
7. Kremer H, van Wijk E, Märker T, Wolfrum U, Roepman R. Usher syndrome: molecular links of pathogenesis, proteins and pathways. *Hum Mol Genet*. 2006;15(suppl 2):R262–R270.
8. Greenberg J, Roberts L, Bruwer Z, Schoeman M, Loggenberg K, Loubser F. Delivery of an ophthalmic genetic service in South Africa. *SA Ophthalmol J*. 2010;5:14–19.
9. Roberts LJ, Ramesar RS, Greenberg J. Clinical utility of the ABCR400 microarray: basing a genetic service on a commercial gene chip. *Arch Ophthalmol*. 2009;127:549–554.
10. Cremers FPM, Kimberling WJ, Külm M, et al. Development of a genotyping microarray for Usher syndrome. *J Med Genet*. 2007;44:153–160.
11. O'Sullivan J, Mullaney BG, Bhaskar SS, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J Med Genet*. 2012;49:322–326.
12. Le Quesne Stabej P, Saihan Z, Rangesh N, et al. Comprehensive sequence analysis of nine Usher syndrome genes in the UK National Collaborative Usher Study. *J Med Genet*. 2012;49:27–36.
13. Weil D, Blanchard S, Kaplan J, et al. Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature*. 1995;374:60–61.
14. Rong W, Chen X, Zhao K, et al. Novel and recurrent MYO7A mutations in Usher syndrome type 1 and type 2. *PLoS One*. 2014;9:e97808.
15. Jaijo T, Aller E, Oltra S, et al. Mutation profile of the MYO7A gene in Spanish patients with Usher syndrome type I. *Hum Mutat*. 2006;27:290–291.
16. Wasfy MM, Matsui JI, Miller J, Dowling JE, Perkins BD. Myosin 7Aa(-/-) mutant zebrafish show mild photoreceptor degeneration and reduced electroretinographic responses. *Exp Eye Res*. 2014;122:65–76.
17. Lopes VS, Gibbs D, Libby RT, et al. The Usher 1B protein, MYO7A, is required for normal localization and function of the visual retinoid cycle enzyme, RPE65. *Hum Mol Genet*. 2011;20:2560–2570.
18. Kaiserman N, Obolensky A, Banin E, Sharon D. Novel USH2A mutations in Israeli patients with retinitis pigmentosa and Usher syndrome type 2. *Arch Ophthalmol*. 2007;125:219–224.
19. 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;66:1975–1978.
20. Liu XZ, Walsh J, Mburu P, et al. Mutations in the myosin VIIA gene cause non-syndromic recessive deafness. *Nat Genet*. 1997;16:188–190.
21. Morgan NV, Essop F, Demuth I, et al. A common Fanconi anemia mutation in black populations of sub-Saharan Africa. *Blood*. 2005;105:3542–3544.
22. Greenberg J, Solomon GAE, Vorster AA, Heckmann J, Bryer A. Origin of the SCA7 gene mutation in South Africa: implications for molecular diagnostics. *Clin Genet*. 2006;70:415–417.

23. De Wit E, Delport W, Rugamika CE, et al. Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Hum Genet*. 2010;128:145–153.
24. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65.
25. Shi YY, He L. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell Res*. 2005;15:97–98.
26. Li Z, Zhang Z, He Z, et al. A partition-ligation-combination-subdivision EM algorithm for haplotype inference with multiallelic markers: update of the SHEsis (http://analysis.bio-x.cn). *Cell Res*. 2009;19:519–523.
27. Gurdasani D, Carstensen T, Tekola-Ayele F, et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature*. 2015;517:327–332.
28. Tõnisson N, Kurg A, Kaasik K, Lõhmussaar E, Metspalu A. Unravelling genetic data by arrayed primer extension. *Clin Chem Lab Med*. 2000;38:165–170.
29. Liu XZ, Walsh J, Tamagawa Y, et al. Autosomal dominant non-syndromic deafness caused by a mutation in the myosin VIIA gene. *Nat Genet*. 1997;17:268–269.
30. Bolz H, Bolz S-S, Schade G, et al. Impaired calmodulin binding of myosin-7A causes autosomal dominant hearing loss (DFNA11). *Hum Mutat*. 2004;24:274–275.
31. Li S, Schlebusch C, Jakobsson M. Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proc Biol Sci*. 2014;281.
32. May A, Hazelhurst S, Li Y, et al. Genetic diversity in black South Africans from Soweto. *BMC Genomics*. 2013;14:644.
33. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet*. 2002;3:611–621.
34. Pepper MS. Launch of the Southern African Human Genome Programme. *S Afr Med J*. 2011;101:287–288.
35. Chimusa ER, Meintjies A, Tchanga M, et al. A genomic portrait of haplotype diversity and signatures of selection in indigenous southern African populations. *PLoS Genet*. 2015;11: e1005052.
36. Schwander M, Lopes V, Sczaniecka A, et al. A novel allele of myosin VIIa reveals a critical function for the C-terminal FERM domain for melanosome transport in retinal pigment epithelial cells. *J Neurosci*. 2009;29:15810–15818.
37. Heathfield L, Lacerda M, Nossek C, Roberts L, Ramesar RS. Stargardt disease: towards developing a model to predict phenotype. *Eur J Hum Genet*. 2013;21:1173–1176.
38. Zallocchi M, Binley K, Lad Y, et al. EIAV-based retinal gene therapy in the shaker1 mouse model for usher syndrome type 1B: Development of UshStat. *PLoS One*. 2014;9:e94272.

Investigative Ophthalmology & Visual Science

Chapter 4. Whole exome sequencing with the strategic analysis of known retinal disease genes

4.1 Molecular diagnosis of inherited retinal diseases in indigenous African populations by whole-exome sequencing

Original Publication

Roberts L*, Ratnapriya R*, du Plessis M, Chaitankar V, Ramesar RS, Swaroop A. Molecular diagnosis of inherited retinal diseases in indigenous African populations by whole exome sequencing. Invest Ophthalmol Vis Sci. 2016;57:6374–6381. DOI:10.1167/iovs.16-19785

NCBI Pubmed link: https://www.ncbi.nlm.nih.gov/pubmed/27898983

Article/journal link: http://iovs.arvojournals.org/article.aspx?articleid=2589779

Nature of the publication: Original full article

Journal/Publisher: Investigative Ophthalmology & Visual Science, an Association for Research in Vision and Ophthalmology (ARVO) journal. ISSN: 1552-5783

Candidate contribution: Conceived the project and strategy, selected the cohort and gathered the clinical data, performed WES experiments and analysis, analysed Taqman® results, validated both WES and Taqman® results by Sanger sequencing, and performed familial cosegregation analysis by sequencing. The candidate drafted the manuscript and participated in all revisions.

Co-authors contributions:

RRP: Performed WES experiments and bioinformatic analysis, and edited the manuscript.

MdP: Assisted with bioinformatics support for WES analysis and edited the manuscript.

VC: Assisted with bioinformatics support for WES analysis and edited the manuscript.

RR: Assisted with project strategy, supervised the research and edited the manuscript.

AS: Assisted with project strategy, supervised the research and edited the manuscript.

Molecular Diagnosis of Inherited Retinal Diseases in Indigenous African Populations by Whole-Exome Sequencing

Lisa Roberts,[1] Rinki Ratnapriya,[2] Morné du Plessis,[1] Vijender Chaitankar,[2] Raj S. Ramesar,[1] and Anand Swaroop[2]

[1]University of Cape Town/MRC Human Genetics Research Unit, Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
[2]Neurobiology, Neurodegeneration & Repair Laboratory, National Eye Institute, National Institutes of Health, Bethesda, Maryland, United States

Correspondence: Anand Swaroop, NNRL, National Eye Institute, MSC0610, 6 Center Street, Bethesda, MD 20892, USA; swaroopa@nei.nih.gov.
Lisa Roberts, Room N3.14, Level 3, Wernher and Beit North Building, Institute of Infectious Disease and Molecular Medicine, University of Cape Town Faculty of Health Sciences, Anzio Road, Observatory, 7925, Cape Town, Western Cape, South Africa; lisa.roberts@uct.ac.za.

LR and RR contributed equally to the work presented here and should therefore be regarded as equivalent authors.

Submitted: April 21, 2016
Accepted: October 18, 2016

Citation: Roberts L, Ratnapriya R, du Plessis M, Chaitankar V, Ramesar RS, Swaroop A. Molecular diagnosis of inherited retinal diseases in indigenous African populations by whole-exome sequencing. *Invest Ophthalmol Vis Sci.* 2016;57:6374–6381. DOI: 10.1167/iovs.16-19785

Purpose. A majority of genes associated with inherited retinal diseases (IRDs) have been identified in patients of European origin. Indigenous African populations exhibit rich genomic diversity, and evaluation of reported genetic mutations has yielded low returns so far. Our goal was to perform whole-exome sequencing (WES) to examine variants in known IRD genes in underrepresented African cohorts.

Methods. Whole-exome sequencing was performed on 56 samples from 16 families with diverse IRD phenotypes that had remained undiagnosed after screening for known mutations using genotyping-based microarrays (Asper Ophthalmics). Variants in reported IRD genes were identified using WES and validated by Sanger sequencing. Custom TaqMan assays were used to screen for identified mutations in 193 unrelated indigenous Africans with IRDs.

Results. A total of 3494 variants were identified in 217 known IRD genes, leading to the identification of seven different mutations (including six novel) in six genes (*RHO*, *PRPF3*, *PRPF31*, *ABCA4*, *CERKL*, and *PDE6B*) in six distinct families. TaqMan screening in additional probands revealed identical homozygous *CERKL* and *PDE6B* variants in four more patients.

Conclusions. This is the first report of WES of patients with IRDs in indigenous African populations. Our study identified genetic defects in almost 40% of the families analyzed, significantly enhancing the molecular diagnosis of IRD in South Africa. Thus, WES of understudied cohorts seems to present an effective strategy for determining novel mutations in heterogeneous retinal diseases.

Keywords: next generation sequencing, genetic testing, photoreceptor dysfunction, South Africa, vision loss, inherited blindness, retinal degeneration, clinical genetics

Indigenous African populations are underrepresented in international genetic/genomic studies. The African continent includes 55 countries (https://africacheck.org/reports/how-many-countries-in-africa-how-hard-can-the-question-be/), with over 2000 distinct ethnolinguistic groups.[1] Being the most ancient of all populations, Africans display vast genetic diversity[2,3] as a result of historical migration, population admixture, response to environmental change, and/or exposure to a plethora of infectious agents.[1] Indigenous Bantu language–speaking individuals arrived in South Africa approximately 1500 years ago as a result of the movement of people, known as the "Bantu expansion," across (west to east) and down (north to south) Africa.[4,5] Subsequent divergence of Bantu speakers in South Africa occurred relatively recently into separate ethnolinguistic groups such as Sotho-Tswana, Xhosa, and Zulu. These black South African individuals, referred to collectively hereafter as indigenous Africans, are the focus of this study as they provide a valuable resource to detect genetic defects in heterogeneous Mendelian diseases including inherited retinal diseases (IRDs).

Inherited retinal diseases encompass a genetically and clinically heterogeneous group of blinding diseases, with a common phenotype of dysfunction and/or degeneration of the light-sensitive photoreceptor cells (rods and cones) in the retina.[6,7] Patients with gene defects causing a primary disease of rod photoreceptors, for example, retinitis pigmentosa (RP), initially experience night blindness and loss of peripheral vision. In contrast, IRDs showing initially the loss of cone photoreceptors, for example, macular degeneration (MD) and Stargardt disease (STGD), manifest with a loss of central vision. Inherited retinal diseases can exhibit autosomal dominant, autosomal recessive, or an X-linked pattern of inheritance and demonstrate progressive or stationary and syndromic or non-syndromic clinical phenotypes.[6,7] Over 240 genes have been

identified for IRDs (https://sph.uth.edu/Retnet/sum-dis.htm; in the public domain). Recent studies using animal models have finally begun to uncover some of the underlying disease mechanisms and pathways that affect photoreceptor development or function.[7–9] Furthermore, it is estimated that only 50% to 70% of the cases with RP (depending on geographical regions or populations) can be attributed to the known genes,[10–12] indicating that a considerable number of as yet unknown mutations and genes remain to be identified. Such a vast clinical and genetic heterogeneity displayed by IRDs confounds molecular diagnosis and investigation of the pathogenic mechanisms.

Identification of the specific genetic defect in a patient with IRD affords several potential benefits. First, overlapping phenotypes and clinical variability of IRDs do not always permit a clear clinical (ophthalmologic) diagnosis/prognosis. Genetic analysis is unequivocal and provides clinical utility as diagnostic, predictive, and carrier testing can be offered to family members. Second, genetic tests may also influence the clinical management of the disease. The IRD research program in South Africa (SA), initiated in 1990,[13] has a strong translational and service component.[14] Lastly, the knowledge of precise genetic defects can allow development of gene-based therapies for treatment of IRDs.[15]

The reported prevalence of IRDs is approximately 1 in 3500[11] in populations where epidemiologic data are available. No data exist on the prevalence of IRDs in Africa. Nonetheless, using SA's 2011 population census (http://www.statssa.gov.za/; in the public domain), one may extrapolate that approximately 14,500 individuals suffer from IRD-related visual impairment/blindness in SA; of these (taking population demographics into account), as many as 11,600 are expected in the indigenous African population. However, a high frequency of unaffected carriers of IRD gene mutations could exist because of local founder effects and further elevate the potential burden of disease.[11]

Demographic information, biological material, clinical details, and diagnoses have been archived for 3237 individuals in 1430 SA families with distinct IRDs in the University of Cape Town (UCT) registry, which contains information and biological material primarily from individuals of Caucasian origin; indigenous Africans currently comprise only 19% of the collection ($n = 275$ families). Understandably, this does not reflect the population demographics of SA and is due to ascertainment bias and the lack of resources in rural areas where a large proportion of the indigenous populations reside. To date, 249 families (249/1430 = 17%), mostly Caucasian ($n = 204/249$; 82%), are in diagnostic mode, with clear pathogenic mutations having been identified using a variety of methods.[16] The most prevalent reported genetic defects in IRDs exhibit an almost insignificant incidence in the SA patient cohort.[17–20] Investigation of the indigenous African subcohort for reported mutations through the use of Asper Ophthalmics microarrays (http://www.asperbio.com/asper-ophthalmics; in the public domain) has produced lower returns in the indigenous African IRD subcohort than in the Caucasian subcohort. Approximately 41.2% of Caucasian samples ($n = 279$) have been diagnosed by microarray screening as opposed to only 12.8% of indigenous African samples ($n = 109$) because each Asper Ophthalmics microarray specifically tests for reported mutations that have been identified predominantly in patients of European/Caucasian origin. Novel mutations are detected only if they occur at a nucleotide position(s) where a mutation has already been reported, as only select nucleotides are assayed. Thus, either SA indigenous IRD patients harbor novel mutations in known genes that are not included in the Asper arrays or causative genes are novel.

The advent of next-generation sequencing technologies has revolutionized the speed and cost at which disease mutations can be identified. An increased number of mutations are now being identified in different populations using high-throughput methods such as whole-exome sequencing (WES).[21,22] Improved molecular diagnosis in patients is important, given the number of clinical trials and treatments currently under investigation for this group of disorders.[23] We therefore resorted to a comprehensive WES approach, followed by targeted analysis of all reported IRD genes, toward understanding the genetic architecture of IRD in the indigenous SA population.

Materials and Methods

Patient Cohort

Informed consent was obtained according to the 2008 Declaration of Helsinki for all members from whom samples have been archived in the UCT IRD registry. Ethics approval was granted by the Human Research Ethics Committee of the UCT Faculty of Health Sciences (Rec Ref. 226/2010 and 768/2013). Samples from indigenous African families were selected from the registry if DNA was available from at least three family members and if a proband had been screened using the appropriate microarray but no molecular diagnosis had been obtained. A total of 16 families met the selection criteria, comprising 109 individuals; of these, 56 were chosen for WES. The selected 16 families originated from diverse, self-identified, indigenous African ethnolinguistic groups: 5 Xhosa, 3 Zulu, 2 Tswana, 1 Shangaan, 1 Venda, 1 Tsonga/Ndebele, 1 Xhosa/Sotho, and 2 Unknown. Two of the 16 families had been clinically diagnosed with autosomal recessive MD (one of whom had a subsequent diagnosis of Leber congenital amaurosis) and 14 with RP.

Whole-Exome Sequencing

Genomic DNA samples were quantified using the QuantiFluor dsDNA system (Promega, Madison, WI, USA), according to manufacturer's instructions. Whole-exome capture was performed on 50 ng DNA using the Nextera Rapid Capture Expanded Exome kit (Illumina, San Diego, CA, USA), and 125-bp paired-end sequences were obtained on a HiSeq2500 platform (Illumina), according to manufacturer's instructions. Details of WES analysis are described elsewhere.[24] FastQC (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; in the public domain) was used to confirm quality of sequencing, after which adapter indexes were removed using Trimmomatic.[25] Reads were mapped to the human reference sequence (hg19, GRCH37) using BWA,[26] and GATK[27,28] was used for variant calling, local realignment, base quality recalibration, and variant recalibration. Annotation of variants was performed with ANNOVAR.[29]

Variant Prioritization and Validation

Sequence variants present in genes (Supplementary Table S1) listed on the RetNet database (https://sph.uth.edu/Retnet/sum-dis.htm; in the public domain; accessed 12 November 2014) were extracted for further analysis. Variants with a minor allele frequency (MAF) of <0.1 in the 1000 Genomes Project[30] (October 2014 annotation) were prioritized, as were exonic or splicing variants. The variants were subsequently selected based on cosegregation with the disease phenotype within each family. For nonsynonymous variants, a minimum threshold of three pathogenic predictions was applied to the dbNSFP[31] annotation of ANNOVAR, for either of the following

Table 1. Candidate Variants in Each of the 16 Families After Prioritization Filters

Family ID	No. of IRD Variants	<0.1 MAF	Exonic/Splicing	Cosegregating Within Family	Pathogenic, >3 Predictions	Candidate Gene, Rare and Cosegregating
RPD 55	1351	749	280	17	7	0
RP 583	1431	796	302	8	8	0
RPD 94	1181	599	198	10	1	0
RP 391	1224	607	209	25	13	*PRPF3*
RPD 401	1183	619	234	30	11	0
RPD 799	1309	686	259	15	5	0
RPD 1001	1416	805	316	8	4	0
RPD 1005	1285	679	223	5	3	0
RPD 1010	1217	628	234	5	3	*RHO*
RPD 1339	1153	579	194	21	10	*PRPF31*
RPM 537	1130	550	191	9	5	*ABCA4* (x2)
RPM 1167	1086	552	198	2	0	0
RPR 397	1063	525	199	19	3	*PDE6B*
RPR 624	1200	620	217	3	0	0
RPR 917	1154	574	203	4	1	*CERKL*
RPX 54	1432	760	259	1	1	0

predictor subsets: (SIFT, PolyPhen2-HDIV, PolyPhen2-HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, MetaSVM, and MetaLR), or (VEST3 CADD-raw, CADD-phred, GERP++, phyloP46way-placental, phyloP100way-vertebrate and SiPhy-29way-logOdds). Variants were then assessed for their presence in the remainder of the cohort. High-priority candidate variants were finally evaluated by examining RetNet and Ensembl release 8[32] with particular emphasis on population data for 1000 Genomes African subpopulations and NHLBI Exome Sequencing data (http://evs.gs.washington.edu/EVS/; in the public domain) in African Americans, as well as reported phenotypes associated with the genes.

Wherever possible, additional familial samples not subjected to WES were included for validation of candidate variants by Sanger Sequencing on a 3130xl Genetic Analyzer (Applied Biosystems, ThermoFisher Scientific, Waltham, MA, USA). Finally, validated variants were checked for MAF in the African Genomes Variation Project[33] data, which include low-coverage whole-genome sequences from 100 Bagandan of Uganda, 100 Zulu of SA, and 120 Ethiopian individuals.

Screening of SA Cohort

Custom TaqMan assays (primer and reporter sequences in Supplementary Table S2) were designed to determine the allele frequency of seven variants identified by WES, in a larger cohort of 193 unrelated indigenous African probands with IRDs but no known causative mutation. In order to determine the optimal template concentration, two control samples were screened for each assay (including a positive control for each), at 8, 6, 4, 3, 2, and 1.5 ng/μL. It was empirically determined that 2 ng/μL was optimal, allowing for effective allele discrimination for each assay.

The final volume in each assay reaction was 5 μL, composed of 2.5 μL TaqMan GT mastermix (2×) (Applied Biosystems), 0.25 μL assay mix (20×), 2.25 μL DNA (at 2 ng/μL, that is, total input of 4.5-ng template). Each assay included at least two no-template controls and two positive controls. Thermal cycling was performed using the ABI 7900HT instrument (Applied Biosystems) and the following conditions: 95°C, 10 minutes; (95°C, 15 seconds; 60°C, 1 minute) × 40 cycles. If fluorescence values dictated after this cycling, a second cycling of 10× (95°C, 15 seconds; 60°C, 1 minute) cycles and subsequent postread analysis were performed. Sanger sequencing was used to validate all candidate variants.

Results

Whole-exome sequencing was performed for 56 samples that included at least three individuals from each of the 16 families. On average, 92% of the exome was captured at 25× coverage, and a total of 1,816,031 variants were identified. We excluded intergenic (n = 759,459), intronic (n = 710,303), and synonymous (n = 59,723) variants from further analysis and identified 3494 candidate variants in 217 reported IRD genes. We then filtered out variants that were present upstream or downstream (n = 298) of the coding exons, in the 5′ or 3′ untranslated region (n = 1813), or in the noncoding RNA (ncRNA) regions (n = 96). Of the remaining IRD variants (1266 exonic and 21 splice site), 561 variants were potentially pathogenic (Supplementary Table S3). At least three prediction algorithms identified 498 variants as pathogenic, and 63 variants were deletions, insertions, gain/loss of stop codons, or variants of unknown effect. The candidate variants remaining after each filtering step are shown in Table 1.

We identified seven different likely mutations in six IRD families; of these, six had not been reported previously (Table 2; Fig. 1). Four of the variants are missense, one is predicted to affect splicing, and two are predicted to result in frameshift and protein truncation. None of the variants has been reported in the whole-genome sequence data of 100 Zulu, 100 Bagandan, or 120 Ethiopian individuals in the AGVP study.[33] Additionally, these variants are not detected in 97 Luyha or 88 Yoruba individuals in the 1000 Genomes data.[30] Therefore, the seven variants identified in IRD families are not present in 505 control African individuals (1010 chromosomes), providing additional evidence in support of their pathogenicity. The previously reported autosomal recessive RP (arRP) mutation p.(His620GlnfsTer23) in *PDE6B* was present only once in 4266 alleles in the NHLBI WES dataset (ESP) of African Americans (rs769671323, as of 27 October 2015); this frameshift mutation is predicted to generate a truncated protein lacking over 200 C-terminal amino acids.[34] The second frameshift mutation identified in *ABCA4* is predicted to truncate the protein by 612 C-terminal amino acids. The c.698-1G>A variant in the acceptor splice site of exon 8 of *PRPF31*, interrogated by Human Splicing Finder 3.0,[35] is predicted to activate an intronic cryptic acceptor site while simultaneously disrupting an exon splicing silencer site and creating an exon splicing enhancer site. Therefore, all seven variants were computationally predicted to be pathogenic, cosegregated with disease in

Table 2. Potential Causative Mutations in Indigenous African Families With IRDs

Family	Disorder	Ethnicity	Gene	Variant: cDNA; Protein	Comment	Pathogenicity, ACMG Category[36]	Reported/Novel
RP 391	adRP	Tswana	*PRPF3*	c.1480A>G; p.(Thr494Ala)	Heterozygous, 9 pathogenic predictions	Likely pathogenic	Novel
RPD 1010	adRP	Xhosa	*RHO*	c.154T>G; p.(Phe52Val)	Heterozygous, 4 pathogenic predictions	Likely pathogenic	Novel
RPD 1339	adRP	Zulu	*PRPF31*	c.698-1G>A; p.(?)	Heterozygous	Likely pathogenic	Novel
RPM 537	arSTGD	Venda	*ABCA4*	c.4832delC; p.(Thr1611MetfsTer51) c.1043T>G; p.(Phe348Cys)	Compound heterozygous, 9 pathogenic predictions for nonsynonymous p.(Phe348Cys)	Pathogenic (frameshift truncation), likely pathogenic (missense)	Both novel
RPR 397	arRP	Shangaan	*PDE6B*	c.1860delC; p.(His620GlnfsTer23)	Homozygous	Pathogenic	Reported, Danciger et al.[34]
RPR 917	arRP	Xhosa	*CERKL*	c.365T>G; p.(Leu122Arg)	Homozygous, 4 pathogenic predictions	Likely pathogenic	Novel

adRP, autosomal dominant RP; arRP, autosomal recessive RP; arSTGD, autosomal recessive STGD.

the respective families, verified by Sanger sequencing, and exhibited conservation across vertebrates (Fig. 2). According to ACMG guidelines for the interpretation of sequence variants,[36] the frameshift truncations identified in this study have sufficient evidence to classify them as "pathogenic," while each of the splice site or missense variants meets the criteria of "likely pathogenic" variants in the absence of functional studies.

We then performed TaqMan assays for these seven pathogenic or likely pathogenic variants, identified here, in an additional 193 indigenous Africans with IRDs. Five of these variants were not detected in this cohort. The *PDE6B* c.1860delC mutation was identified in a homozygous state in one additional individual (diagnosed with arRP, from infancy) and in a heterozygous state in four individuals (two sporadic RP, one arRP, and one with an apparent dominant family history). In addition, we identified the homozygous *CERKL* c.365T>G variant in three patients with different IRD phenotypes: one each of sporadic RP, sporadic STGD, and arRP. This c.365T>G variant was also identified in the heterozygous state in one RP proband.

Discussion

The use of indigenous SA populations, combined with next-generation sequencing platforms, provides an enriched resource for discovering novel IRD genes and mutations. Due to the vast clinical and genetic heterogeneity, traditional candidate gene-based approaches have been less effective for the molecular diagnosis of IRDs. Targeted capture of specific IRD genes, associated with particular retinal phenotypes, is a strategy being used for molecular diagnosis with increasing frequency.[37–41] Both targeted capture and WES allow for the detection of novel mutations in genes (in contrast to microarrays). Recently targeted capture of known IRD genes in panel-based testing was reportedly more successful than WES,[42] probably due to better coverage of the genes of interest. We believe that panel-based testing is especially not suitable for the research on understudied populations, like the indigenous Africans, where WES with targeted bioinformatic analysis could enhance molecular diagnosis and even lead to novel gene discovery. Collaborative and combined analysis of WES data from different groups can yield genetic evidence for novel IRD genes. In addition, WES data from unresolved families can be reanalyzed when novel IRD genes are reported without redesigning diagnostic gene panels and performing new experiments. The latter is an important consideration when providing a molecular diagnosis for patients in resource-limited settings.

Our targeted analysis approach was successful in assigning molecular diagnosis in 38% of the indigenous African families, a clear improvement on the 13% detection rate using the commercially available arrays that test for specific reported variants. Six of seven (85%) variants discovered were novel, supporting the high genetic heterogeneity in IRDs as well as genetic diversity among indigenous Africans. Analysis of a larger cohort of unrelated indigenous African probands revealed that five out of seven variants were rare and detected in a single family each, further advocating the use of WES-based diagnosis instead of the genotyping-based microarrays used previously to screen this population group. Nonetheless, the detection rate is still much lower than the reported 83% of European families interrogated using a similar approach.[43] Other population groups investigated in a comparable manner include Saudi Arabian,[44] Chinese,[45] Thai,[46] and Israeli,[22,47] with detection rates ranging from 49% to 83% and the number of analyzed genes ranging from 60 to 226.

The relatively low detection of causal mutations in the SA cohort of IRD families can be attributed to multiple factors. Whole-exome sequencing is a capture-based method with genomic regions of low coverage and poor detection of large genomic alterations. Additionally, WES will not detect less obvious pathogenic variants, such as ncRNA or regulatory variants and those present in the untranslated regions or introns. The clinical complexity of IRDs, that is, nonpenetrance, frequent manifestation in carriers of X-linked disorders, and variable expressivity within families, could result in an incorrect inheritance pattern being assumed and hence incorrect variant filtering during cosegregation analysis. This problem is exacerbated in SA, where frequently sparse clinical information accompanies the samples particularly from the more rural areas of the country, and where language barriers can often result in misinformation. However, it is also plausible that causative mutations in many families reside in an as yet

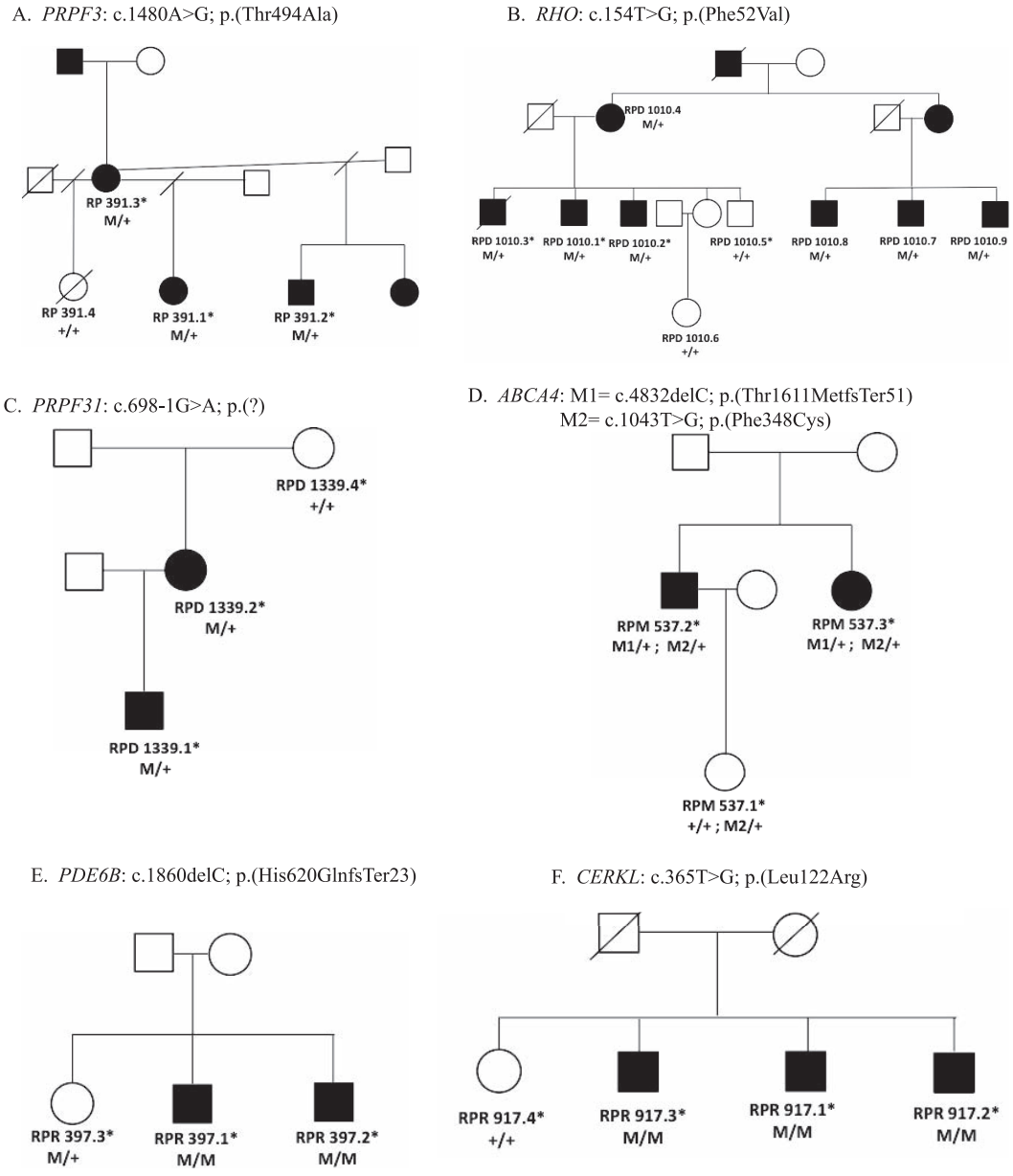


Figure 1. Pedigrees of IRD families showing cosegregation of the variants identified by WES. *Squares* represent males, and *circles*, females. *Shaded symbols* indicate individuals with IRD. Identifier codes show individuals from whom biological material is available, and those selected for whole-exome sequencing are noted with an *asterisk*. Segregation of mutation(s) in the families is indicated as +/+, homozygous for wild-type allele; M/+, heterozygous; M/M, homozygous for mutation. Clinical information is presented in Supplementary Table S4.

unreported IRD gene. We believe that the use of previously understudied populations is a sensible approach for ascertaining missing heritability in genetically heterogeneous diseases such as IRDs.

PDE6B mutations have been associated with autosomal dominant congenital stationary night blindness (adCSNB) and arRP.[48] In our patient samples, two probands with arRP carried the homozygous c.1860delC mutation of *PDE6B*. In addition, we identified four IRD patients (two sporadic RP, one arRP, and one with an apparent autosomal dominant [adRP] family history) with a heterozygous *PDE6B* c.1860delC allele. The relatively high frequency of this allele (1.9%; $n = 8/418$ alleles) in the SA IRD cohort could imply compound heterozygosity for *PDE6B*, digenic inheritance, or enhanced genetic burden. The individual RPR 397.1 (in the WES cohort) had been tested previously by the arRP microarray; however, this array platform was designed to detect the c.1857_1858delC *PDE6B* variant and not c.1860. We also noted the relatively frequent occurrence of the *CERKL* c.365T>G variant in SA IRD patients ($n = 9/418$ alleles; 2.2%). The four homozygous cases with this mutation displayed varying phenotypes: two arRP, one sporadic RP, and one sporadic STGD. *CERKL* mutations are shown to result in autosomal recessive forms of cone dystrophy, cone–rod dystrophy and RP (RetNet). In our study, an identical *CERKL* mutation is associated with distinct IRD phenotypes, implying the existence of modifier variants or the impact of vastly different environmental and epigenetic landscape in this genetically diverse cohort compared to the reported Caucasian patients. Given the existence of the large number of sequence variants in native Africans,[2,3] it would be prudent to perform WES on carriers of *PDE6B* and *CERKL* variants to identify causal IRD mutation(s).

***PRPF3*: c.1480A>G; p.(Thr494Ala)**

Human (Homo sapiens)	GAAGCTGTTCAAGACCCC**A**CGAAGGTAGAAGCCCA
Chimpanzee (Pan troglodytes)	GAAGCTGTTCAAGACCCC**A**CGAAGGTAGAAGCCCA
Cow (Bos taurus)	GAAGCTGTTCAAGACCCC**A**CGAAGGTAGAAGCCCA
Dog (Canis lupus familiaris)	GAAGCTGTTCAAGACCCC**A**CGAAGGTAGAAGCCCA
Zebrafish (Danio rerio)	GAAGCTGTGCAGGACCCC**A**CTAAAGTGGAGGCCCA
Chicken (Gallus gallus)	GAAGCTGTTCAGGACCCA**A**CAAAAGTTGAAGCTCA
Mouse (Mus musculus)	GAAGCAGTTCAAGACCCT**A**CAAAGGTAGAAGCCCA

***RHO*: c.154T>G; p.(Phe52Val)**

Human (Homo sapiens)	CTGCTGATCGTGCTGGGC**T**TCCCCATCAACTTCCTC
Chimpanzee (Pan troglodytes)	CTGCTGATCGTGCTGGGC**T**TCCCCATCAACTTCCTC
Cow (Bos taurus)	CTGCTGATCATGCTTGGC**T**TCCCCATCAACTTCCTC
Dog (Canis lupus familiaris)	CTGCTGATCGTGCTCGGC**T**TCCCCATCAACTTCCTC
Zebrafish (Danio rerio)	TTCCTCATCATCACCGGC**T**TCCCCGTCAACTTCCTC
Chicken (Gallus gallus)	ATGCTGATCCTGCTCGGC**T**TCCCCGTCAACTTCCTC
Mouse (Mus musculus)	CTGCTCATCGTGCTGGGC**T**TCCCCATCAACTTCCTC

***PRPF31*: c.698-1G>A; p.(?)**

Human (Homo sapiens)	CTCCCTCCCACCGCA**G**GTGGGCCGGCGGCCTGACC
Chimpanzee (Pan troglodytes)	CTCCCTCCCACCGCA**G**GTGTGGCCGGCGGCCTGACG
Cow (Bos taurus)	----------CA**G**GGGTGGCCGGCGGCCTGACC
Dog (Canis lupus familiaris)	CTCCTTCCTGTTGCA**G**GGGTGGCTGGGGGCCTGACC
Zebrafish (Danio rerio)	TTTCCTTTTATTTCA**G**GTGTTGCTGGTGGTCTGACT
Chicken (Gallus gallus)	no homologue
Mouse (Mus musculus)	----------GCA**G**GTGTGGCTGGAGGGCTCACC

***ABCA4*: c.4832delC; p.(Thr1611MetfsTer51)**

Human (Homo sapiens)	AAACATCTAG AAA**C**TGAAGA CAACATTAAG GTACTT
Chimpanzee (Pan troglodytes)	AAACATCTAG AAA**C**TGAAGA CAACATTAAG GTACTT
Cow (Bos taurus)	AAACAACTAG AAA**C**TGAAGA CAATATTAAG GTATTG
Dog (Canis lupus familiaris)	AAACATCTAG AAA**C**TGAAGA CAACATTAAG GTATTG
Zebrafish (Danio rerio)	no homologue
Chicken (Gallus gallus)	no homologue
Mouse (Mus musculus)	AAACATCTTG AAA**C**CACAGA CAACATTAAG GTAC--

***ABCA4*: c.1043T>G; p.(Phe348Cys)**

Human (Homo sapiens)	TATAAGGCCT**T**TCTGGGGAT TGACTCCACA AGGAAG
Chimpanzee (Pan troglodytes)	TATAAGGCCT**T**TCTGGGGAT TGACTCCACA AGGAAG
Cow (Bos taurus)	TATAAGGCCT**T**CTAGGGAT CGACTCCACA AGGAAA
Dog (Canis lupus familiaris)	TATAAGGCCT**T**CTTAGGGAT TGACTCCACA AGGAAG
Zebrafish (Danio rerio)	no homologue
Chicken (Gallus gallus)	TACAAAGCTT**T**CCTGGGGAT TGATTCCACA AAGAAA
Mouse (Mus musculus)	TATAAAGCCT**T**CCTGGGGAT TGATTCCACA AGGAAA

***PDE6B*: c.1860delC; p.(His620GlnfsTer23)**

Human (Homo sapiens)	TGGCTAAGCTCCA**C**GGCTCC TCGATTTTGG AGCGGC
Chimpanzee (Pan troglodytes)	no homologue
Cow (Bos taurus)	TAGCCAAGCTCCA**C**GGCTCC TCGATTTTGG AGCGAC
Dog (Canis lupus familiaris)	TGGCCAAGCTCCA**C**GGCTCC TCCATCCTGG AGCGGC
Zebrafish (Danio rerio)	no homologue
Chicken (Gallus gallus)	no homologue
Mouse (Mus musculus)	TAGCCAAGTTACATGGCTCCTCAATTCTGGAAAGGC

***CERKL*: c.365T>G; p.(Leu122Arg)**

Human (Homo sapiens)	TTTATTAGGTATCACAC**T**CTTCATCTGCTTGAAAAA
Chimpanzee (Pan troglodytes)	TTTATTAGGTATCACAC**T**CTTCATCTGCTTGAAAAA
Cow (Bos taurus)	TTTATTAGGAATCACTC**T**CTTCATATGTTTGAAAAC
Dog (Canis lupus familiaris)	TTTATTGGGTATCACAC**T**CTTCATTTGTTTGAAAAA
Zebrafish (Danio rerio)	no homologue
Chicken (Gallus gallus)	no homologue
Mouse (Mus musculus)	CTGCTGGGCATAACGC**T**TTTCATCTGCTTGAAGGA

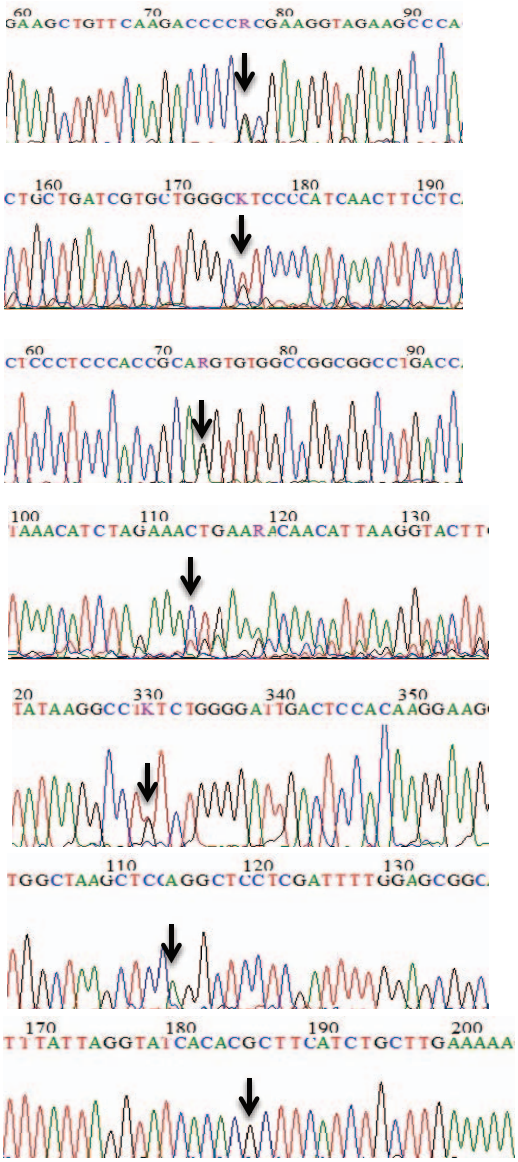


Figure 2. Sequence alignments across vertebrate species demonstrating nucleotide conservation of each identified variant (highlighted in *red*). *Right*: the corresponding Sanger sequencing electropherogram. *Arrow* indicates the position of the mutation.

Our study shows that genetic investigations of the SA indigenous population present considerable challenges and unique opportunities in human disease gene discovery.[49] Africans have smaller haplotype blocks and low levels of linkage disequilibrium compared to non-African populations, as well as evidence of genetic admixture, leading to unique diversity.[3,4] Whole-exome sequencing of RP families in the United States has yielded a greater number of novel variants (both single nucleotide variants and small indels) in the families of African ancestry compared with families of European ancestry.[50] In this study, the number of variants novel to the National Center for Biotechnology Information Short Genetic Variations database (dbSNP) was reportedly >6-fold larger in a family of African American descent ($n > 2500$) than in Caucasian U.S. families ($n \sim 400$). Given that genome-wide ancestry estimates show an average proportion of only ~73% African ancestry in African Americans,[51] the exomes of indigenous Africans are expected to yield even more novel variants. Therefore, inclusion of African populations in genomics research should facilitate the discovery of genetic defects associated with human disease.[52]

This study employs the first next generation sequencing (NGS)-based approach in an indigenous SA cohort as an opportunity for improved understanding of the genetic architecture of IRDs. We have shown that success of diagnosis is enhanced considerably using WES, and have identified important genes and novel variants for genetic counseling for IRD patients. Our study provides valuable insight into the etiology of IRD in SA, and contributes toward more comprehensive understanding of this heterogeneous group of disorders by cataloguing novel causative variants.

Acknowledgments

The authors thank patients and family members for participation and Linn Gieser and Ash-Police Reddy for technical assistance with WES. This study made use of data generated by the African Genomes Variation Project, for the African Partnership for Chronic

Disease Research. A full list of the investigators and funders who contributed to the generation or collation of the data is available from www.apcdr.org.

Research in South Africa was funded by Retina South Africa and the Medical Research Council of South Africa. WES and data analysis were supported by the Intramural Research Program (EY000546) of the National Eye Institute and utilized computational resources of the National Institutes of Health High-Performance Computing Biowulf cluster (https://hpc.nih.gov).

Disclosure: **L. Roberts**, None; **R. Ratnapriya**, None; **M. du Plessis**, None; **V. Chaitankar**, None; **R.S. Ramesar**, None; **A. Swaroop**, None

References

1. Campbell MC, Tishkoff SA. The evolution of human genetic and phenotypic variation in Africa. *Curr Biol*. 2010;20:R166–R173.
2. May A, Hazelhurst S, Li Y, et al. Genetic diversity in black South Africans from Soweto. *BMC Genomics*. 2013;14:644.
3. Schuster SC, Miller W, Ratan A, et al. Complete Khoisan and Bantu genomes from southern Africa. *Nature*. 2010;463:943–947.
4. Chimusa ER, Meintjies A, Tchanga M, et al. A genomic portrait of haplotype diversity and signatures of selection in indigenous southern African populations. *PLoS Genet*. 2015;11:e1005052.
5. Li S, Schlebusch C, Jakobsson M. Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proc Biol Sci*. 2014;281.
6. Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res*. 2010;29:335–375.
7. Wright AF, Chakarova CF, Abd El-Aziz MM, Bhattacharya SS. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat Rev Genet*. 2010;11:273–284.
8. Swaroop A, Kim D, Forrest D. Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nat Rev Neurosci*. 2010;11:563–576.
9. Veleri S, Lazar CH, Chang B, Sieving PA, Banin E, Swaroop A. Biology and therapy of inherited retinal degenerative disease: insights from mouse models. *Dis Model Mech*. 2015;8:109–129.
10. Anasagasti A, Irigoyen C, Barandika O, de Munain AL, Ruiz-Ederra J. Current mutation discovery approaches in retinitis pigmentosa. *Vis Res*. 2012;75:117–129.
11. Nishiguchi KM, Rivolta C. Genes associated with retinitis pigmentosa and allied diseases are frequently mutated in the general population. *PLoS One*. 2012;7:e41902.
12. Daiger SP, Bowne SJ, Sullivan LS. Genes and mutations causing autosomal dominant retinitis pigmentosa. *Cold Spring Harb Perspect Med*. 2015;5:a017129.
13. Greenberg J, Bartmann L, Ramesar R, Beighton P. Retinitis pigmentosa in southern Africa. *Clin Genet*. 1993;44:232–235.
14. Greenberg J, Roberts L, Bruwer Z, Schoeman M, Loggenberg K, Loubser F. Delivery of an ophthalmic genetic service in South Africa. *S A Ophthalmol J*. 2010;5:14–19.
15. Dalkara D, Goureau O, Marazova K, Sahel JA. Let there be light: gene and cell therapy for blindness. *Hum Gene Ther*. 2016;27:134–147.
16. Roberts L, Goliath R, Rebello G, et al. Inherited retinal disorders in South Africa and the clinical impact of evolving technologies. *S Afr Med J*. 2016;106:10988.
17. Roberts L, Bartmann L, Ramesar R, Greenberg J. Novel variants in the hotspot region of RP1 in South African patients with retinitis pigmentosa. *Mol Vis*. 2006;12:177–183.
18. Roberts L, Ramesar R, Greenberg J. Low frequency of rhodopsin mutations in South African patients with autosomal dominant retinitis pigmentosa. *Clin Genet*. 2000;58:77–78.
19. Roberts L, Rebello G, Greenberg J, Ramesar R. Great expectations: RPE65 mutations in South Africa. In: Baert M, Peeters C, eds. *Retinitis Pigmentosa: Causes Diagnosis and Treatment*. New York, NY: Nova Science Publishers; 2010:89–110.
20. Greenberg J, Roberts L, Ramesar R. Unusual frequencies of rhodopsin mutations and polymorphisms in South African patients with retinitis tigmentosa. In: Anderson RE, LaVail MM, Hollyfield JG, eds. *New Insights into Retinal Degenerative Diseases. A Book on the Proceedings of the IXth International Symposium on Retinal Degeneration*. New York: Kluwer Academic/Plenum Publishers; 2002:329–331.
21. Ratnapriya R, Swaroop A. Genetic architecture of retinal and macular degenerative diseases: the promise and challenges of next-generation sequencing. *Genome Med*. 2013;5:84.
22. Lazar CH, Mutsuddi M, Kimchi A, et al. Whole exome sequencing reveals GUCY2D as a major gene associated with cone and cone-rod dystrophy in Israel. *Invest Ophthalmol Vis Sci*. 2015;56:420–430.
23. Thompson DA, Ali RR, Banin E, et al. Advancing therapeutic strategies for inherited retinal degeneration: recommendations from the Monaciano Symposium. *Invest Ophthalmol Vis Sci*. 2015;56:918–931.
24. Chaitankar V, Karakulah G, Ratnapriya R, Giuste FO, Brooks MJ, Swaroop A. Next generation sequencing technology and genomewide data analysis: perspectives for retinal research [published online ahead of print June 10, 2016]. *Prog Retin Eye Res*. doi: 10.1016/j.preteyeres.2016.06.001.
25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–2120.
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760.
27. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–498.
28. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303.
29. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
30. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.
31. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat*. 2013;34:E2393–E2402.
32. Cunningham F, Amode MR, Barrell D, et al. Ensembl 2015. *Nucleic Acids Res*. 2015;43:D662–D669.
33. Gurdasani D, Carstensen T, Tekola-Ayele F, et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature*. 2015;517:327–332.
34. Danciger M, Blaney J, Gao YQ, et al. Mutations in the PDE6B gene in autosomal recessive retinitis pigmentosa. *Genomics*. 1995;30:1–7.
35. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37:e67.
36. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics

Investigative Ophthalmology & Visual Science

and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–424.

37. Audo I, Bujakowska KM, Leveillard T, 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.
38. Booij JC, Bakker A, Kulumbetova J, et al. Simultaneous mutation detection in 90 retinal disease genes in multiple patients using a custom-designed 300-kb retinal resequencing chip. *Ophthalmology*. 2011;118:160–167.
39. Ge Z, Bowles K, Goetz K, et al. NGS-based Molecular diagnosis of 105 eyeGENE((R)) probands with Retinitis Pigmentosa. *Sci Rep*. 2015;5:18287.
40. O'Sullivan J, Mullaney BG, Bhaskar SS, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J Med Genet*. 2012;49:322–326.
41. Wang X, Wang H, Sun V, et al. Comprehensive molecular diagnosis of 179 Leber congenital amaurosis and juvenile retinitis pigmentosa patients by targeted next generation sequencing. *J Med Genet*. 2013;50:674–688.
42. Consugar MB, Navarro-Gomez D, Place EM, 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*. 2015;17:253–261.
43. Corton M, Nishiguchi KM, Avila-Fernandez A, et al. Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis. *PLoS One*. 2013;8:e65574.
44. Abu-Safieh L, Alrashed M, Anazi S, et al. Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Res*. 2013;23:236–247.
45. Xu Y, Guan L, Shen T, et al. Mutations of 60 known causative genes in 157 families with retinitis pigmentosa based on exome sequencing. *Hum Genet*. 2014;133:1255–1271.
46. Jinda W, Taylor TD, Suzuki Y, et al. Whole exome sequencing in Thai patients with retinitis pigmentosa reveals novel mutations in six genes. *Invest Ophthalmol Vis Sci*. 2014;55:2259–2268.
47. Beryozkin A, Shevah E, Kimchi A, et al. Whole exome sequencing reveals mutations in known retinal disease genes in 33 out of 68 Israeli families with inherited retinopathies. *Sci Rep*. 2015;5:13187.
48. Manes G, Cheguru P, Majumder A, et al. A truncated form of rod photoreceptor PDE6 beta-subunit causes autosomal dominant congenital stationary night blindness by interfering with the inhibitory activity of the gamma-subunit. *PLoS One*. 2014;9:e95768.
49. Ramsay M, Tiemessen CT, Choudhury A, Soodyall H. Africa: the next frontier for human disease gene discovery? *Hum Mol Genet*. 2011;20:R214–R220.
50. Koboldt DC, Larson DE, Sullivan LS, et al. Exome-based mapping and variant prioritization for inherited Mendelian disorders. *Am J Hum Genet*. 2014;94:373–384.
51. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet*. 2015;96:37–53.
52. Sirugo G, Hennig BJ, Adeyemo AA, et al. Genetic studies of African populations: an overview on disease susceptibility and response to vaccines and therapeutics. *Hum Genet*. 2008;123:557–598.

Investigative Ophthalmology & Visual Science

4.2 Extended WES analysis of known IRD genes and the identification of a novel X-linked mutation

The analysis of all known candidate genes gave a relatively low yield of causal mutations (Chapter 4.1). Given the pace at which IRD genes are currently being identified, it was considered prudent to re-analyse the WES data with an updated list of IRD genes. Furthermore, accurate clinical information such as individual disease status, phenotypes and mode of inheritance within a family remain crucial for WES, as per traditional genetic approaches, with errors in clinical data impeding downstream analysis and variant prioritisation for mutation identification. As described previously, it is impossible to predict the causative IRD gene based solely on phenotype and inheritance pattern. This is particularly true for RP, and specifically xlRP. Consequently, an incorrect assumption of inheritance modes in the previous analysis required consideration.

RP is the most common IRD, with an estimated prevalence of about 1 in 3,500–4,000 individuals[15,17] . Approximately 15–25% of cases are adRP, 35–50% arRP, and 7–15% xlRP[17]. Simplex or isolated cases are also common[17,132]. RP is usually monogenic, yet an expansive list of genes has been associated with the disease. There are currently 22, 36 and two genes identified as causing adRP, arRP and xlRP respectively (RetNet https://sph.uth.edu/retnet/, accessed 18 May 2016). However, several genes have been associated with multiple inheritance patterns, e.g. *RHO*, *NRL*, and *RP1* have each been associated with both adRP and arRP[26,133].

Although phenotypic variability occurs, it can be generalised that xlRP cases display a severe RP phenotype, with early age of onset (childhood to late teen years) and rapid disease progression[132,134]. The two causative genes, *RP2* and *RPGR,* account for 10–20% and 50–90% of xlRP cases, respectively[132,134,135]. In contrast with many other *xl* recessive Mendelian disorders where males are exclusively affected, female carriers of mutations in the xlRP genes can manifest with disease, resulting in a pedigree which displays an apparent *ad* trait[17,136]. It has been reported that ~9% of families with a provisional diagnosis of adRP actually have xlRP[137], and 15% of simplex males with RP have a mutation in *RP2* or *RPGR*[132].

There is a spectrum of disease severity in female carriers of xlRP[134,138–141]; they often display some degree of fundus changes despite having normal visual acuity, however 10–30% of carriers show no fundus abnormalities[142]. Moreover, families with semi-dominant xlRP have been reported in whom several heterozygous females have a highly penetrant phenotype[136,143–145] whilst remaining more mildly affected than hemizygous males. Furthermore, intrafamilial clinical heterogeneity has been reported amongst obligate carriers[143]. This variable disease expression in xlRP carrier females is often attributed to lyonisation (random inactivation of most genes on either the maternal or paternal X chromosome, occurring early in female embryo development). If the wild type allele is inactivated in retinal tissues, and insufficient cells express the functional gene product, disease manifestation can occur.

4.2.1 Methods

4.2.1.1 Patient cohort and WES

The 10 families (of the original cohort of 16 families) in whom no causative mutations had been detected through the preceding WES analysis were interrogated further. Pedigrees with an absence of male-to-male transmission of disease were identified as possible cases of xlRP.

WES methodology has been described (Chapter 4.1). Briefly, exome capture was performed using the Nextera Rapid Capture Expanded Exome kit (Illumina, San Diego, CA, USA) on 50ng DNA, and sequencing was performed on a HiSeq2500 platform (Illumina, San Diego, CA, USA), according to manufacturer's instructions. Reads were mapped to the human reference sequence (hg19, GRCH37) and software versions for exome analysis were as follows: Fastqc v.0.11.4 (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) , Trimmomatic v.0.33[146], BWA v.0.7.12[147], GATK v.3.4-46[148,149], Samtools v.1.2[150], and ANNOVAR 2014Nov12[151]. This WES analysis utilised computational resources of the high performance computing Biowulf cluster (https://hpc.nih.gov) at the National Institutes of Health, USA.

4.2.1.2 Exclusion of intergenic, intronic and synonymous WES variants

Intergenic, intronic and synonymous variants were excluded sequentially using an inverse (-v) 'grep' command on a PC running the UBUNTU LINUX operating system. Quality control (QC) was performed by viewing the outputs of the various grep commands (using the 'whole' flag -w), with 'intergenic', 'intronic' and 'synonymous' keywords, and ensuring the variant totals in the output files added up correctly (Appendix 4). Examples of the commands used are below:

> grep –w "intergenic" DATA.csv > intergenicDATA.csv

> grep –v –w "synonymous" NOintergenic_or_intronicDATA.csv > NOintergenic_intronic_OR_synonymousDATA .csv

4.2.1.3 Extraction of IRD variants from WES data

An updated list of 714 gene symbols was obtained from the RetNet database (https://sph.uth.edu/retnet/, accessed 27 January 2016), which included the original 217 analysed genes (Chapter 4.1). The 714 gene names comprised NCBI gene symbols for genes associated with all IRD phenotypes, including non-Mendelian diseases and syndromic IRDs, as well as gene aliases, therefore this count has redundancy. It also included four genes not yet listed on the database (personal communication with Prof. S. Daiger, The University of Texas HSC at Houston). Nonetheless, the revised list now contained 240 identified IRD genes, including 22 adRP, 34 arRP and two xlRP genes (Table 4.2.1).

The list of 714 gene symbols was extracted using the file 'grep' command, and QC was performed again by viewing the output file and ensuring the variant totals added up correctly. Examples of commands used are:

> grep –w –f Full714IRDlist20160127.csv DATA.csv > ALL714RETNETcandidatesDATA.csv

> grep –w –f Full714IRDlist20160127.csv NOintergenic_intronic_OR_synonymousDATA.csv > Filtered2016RetnetDATA.csv

Table 4.2.1 Summary of all Mendelian IRD phenotypes, with the number of identified loci and genes for each. Reproduced from the RetNet database (27/01/2016)

Disease Category	Total no. of Genes and Loci	No. of Identified Genes
Bardet-Biedl syndrome, autosomal recessive	17	17
Chorioretinal atrophy or degeneration, autosomal dominant	1	1
Cone or cone-rod dystrophy, autosomal dominant	9	5
Cone or cone-rod dystrophy, autosomal recessive	15	14
Cone or cone-rod dystrophy, X-linked	1	0
Congenital stationary night blindness, autosomal dominant	1	1
Congenital stationary night blindness, autosomal recessive	9	9
Congenital stationary night blindness, X-linked	2	2
Leber congenital amaurosis, autosomal dominant	1	1
Leber congenital amaurosis, autosomal recessive	11	11
Macular degeneration, autosomal dominant	14	9
Macular degeneration, autosomal recessive	3	3
Ocular-retinal developmental disease, autosomal dominant	1	1
Optic atrophy, autosomal dominant	6	3
Optic atrophy, autosomal recessive	2	1
Optic atrophy, X-linked	1	0
Retinitis pigmentosa, autosomal dominant	23	22
Retinitis pigmentosa, autosomal recessive	37	34
Retinitis pigmentosa, X-linked	5	2
Syndromic/systemic diseases with retinopathy, autosomal dominant	9	8
Syndromic/systemic diseases with retinopathy, autosomal recessive	49	45
Syndromic/systemic diseases with retinopathy, X-linked	3	2
Usher syndrome, autosomal recessive	16	13
Other retinopathy, autosomal dominant	13	9
Other retinopathy, autosomal recessive	15	13
Other retinopathy, mitochondrial	7	7
Other retinopathy, X-linked	8	7
TOTALS	**279**	**240**

The data at this stage represented the sum of variants across all 56 individuals, and not necessarily those occurring in each distinct family. Thus, data were separated into the ten families of interest in Microsoft Excel, and all subsequent analyses were executed separately for each family.

A perl script, dotzero.pl (written by Dr. G. Rebello) was used to detect variants without alternative genotypes, i.e. a wild type genotype (0/0) or gaps (./.). The script detected variants if their occurrence within a family was only with a 0/0 genotype or gaps, or a combination of the two, across all individuals in that family. The option "keep" would extract these rows, whilst the option "throw" would extract the inverse, for example:

> perl dotzero.pl Fam55ALL2016RETNETDATA.csv keep > FAM55dotzeros.csv

> perl dotzero.pl Fam55ALL2016RETNETDATA.csv throw > FAM55nodotzeros.csv

QC was performed by manually viewing the "keep" output file to ensure the correct rows were extracted, and also ensuring the variant totals added up correctly between the "keep" and "throw" output files.

4.2.1.4 Variant prioritisation

Variants with a minor allele frequency (MAF) of <0.1, in the 1000 Genome Project[112] (Oct 2014 annotation) were prioritized, as were exonic or splicing variants. These variants were extracted in Microsoft Excel, by filtering for variants with no MAF information and MAF <0.1. Subsequently, downstream, upstream, 3' untranslated region (UTR), 5' UTR, and non-coding RNA (ncRNA) variants were removed by filtering in Microsoft Excel.

Family pedigrees were assessed for male-to-male transmission, and apparent mode of inheritance. A perl script, coseg.pl, was written in collaboration with Dr. G. Rebello, to apply rules for Mendelian inheritance (*ad*, *ar* and *xl* patterns) to each family. A text file, cosegfamily.txt which included identifier, column number, affection status and gender information for each sample in the cohort, was employed by the script. The coseg.pl script executed allele co-segregation analysis of each family, retaining variants which segregated appropriately within the family based on a specified inheritance model, e.g.

> perl cosegv03.pl FAM55ExonicSplice.csv cosegfamily.txt D > FAM55cosegD.csv

> perl cosegv03.pl FAM55ExonicSplice.csv cosegfamily.txt X > FAM55cosegX.csv

The above commands would retain variants co-segregating with disease in a family, based on an *ad* and *xl* inheritance model, respectively. The script was used to analyse each family for each potential mode of inheritance, hence some families were analysed for two different modes if male-to-male transmission was absent. All segregating variants per family were then incorporated into a single input file for downstream analyses.

During coseg.pl script design, QC was performed by ensuring that output variants matched those that would have been correctly selected manually. This confirmed that conservative results were obtained using the script which could be refined manually. This was preferable to a less conservative script incorrectly filtering out variants that should be retained.

Another perl script, scorep.pl (written by Dr. G. Rebello), was used to select pathogenic variants from the segregating variants extracted above. This script requires a text document, svalues.txt, containing user-established thresholds for the different prediction tools present in the Database for nonsynonymous SNPs' functional predictions (dbNSFP)[152,153] , annotated by ANNOVAR[151]. Two types of pathogenicity predictions, termed Predictions (P, where the prediction tool gives a decision) and Matrix (M, where the prediction tool gives a score but no decision) were used (Table 4.2.2). Low- and high- thresholds were established, based on the literature[154]. Scores below the low threshold, between the low threshold and high threshold, and above the high threshold, were scored 0, 1 and 2 respectively. Alternatively, a binary decision was used (i.e. 0 or 2) if the prediction gave only two outcomes.

Table 4.2.2 Summary of 16 pathogenicity prediction tools and the score thresholds applied in this study to prioritise pathogenic variants.

Pathogenicity prediction tool	Prediction basis	P / M	Score threshold
SIFT_pred	Protein sequence conservation among homologs[155].	P	T = 0, D = 2
Polyphen2_HDIV _pred	Several sequence, phylogenetic and structure based features, (default HumDiv-trained predictor)[156].	P	B = 0, P = 1, D = 2
Polyphen2_HVAR _pred	Several sequence, phylogenetic and structure based features, (HVAR distinguishes drastic effects and is better for Mendelian disease diagnostics)[156].	P	B = 0, P = 1, D = 2
LRT_pred	DNA sequence homology, evolutionary model using 32 vertebrate species (≥ 10 eutherian mammals)[152,154,157].	P	N/U = 0, D = 2
MutationTaster _pred	Functional prediction using DNA sequence conservation, splice site prediction, mRNA stability prediction, and protein feature annotations[154,158].	P	N/P = 0, A/D = 2
MutationAssessor _pred	Sequence homology of protein families within and between species[154,159].	P	L/N = 0, M = 1, H = 2
FATHMM_pred	Sequence conservation[160].	P	T = 0, D = 2
RadialSVM_pred	Ensembl score to aggregate individual prediction scores and MAF (Support Vector Machine approach)[154].	P	T = 0, D = 2
LR_pred	Ensembl score to aggregate individual prediction scores and MAF (Logistic regression approach)[154].	P	T = 0, D = 2
VEST3_score	Identifies causal variants based on enrichment for functional changes in across disease exomes[161].	M	<0.5 = 0, ≥0.5 and <0.75 = 1, ≥0.75 = 2
CADD_raw	Contrasts variants surviving natural selection (comparison between human and chimpanzee) with simulated mutations[162]. The raw score is effectively arbitrary.	M	<5 = 0, ≥5 = 2
CADD_phred	Contrasts variants surviving natural selection (comparison between human and chimpanzee) with simulated mutations[162]. The phred score is a scaled-C score that ranks variants relative to the most deleterious substitutions.	M	<10 = 0, ≥10 and <20 = 1, ≥20 = 2
GERP++_RS	Identifies slowly evolving regions (constrained elements) and measures functional constraint[163–165].	M	<4.4 = 0, ≥4.4 = 2
phyloP46way _placental	Measures conservation in terms of faster/slower evolution than expected[166]. The dataset used is 46 placental mammals.	M	<1.6 = 0, ≥1.6 = 2

Table 4.2.2 (continued)

Pathogenicity prediction tool	Prediction basis	P / M	Score threshold
phyloP100way _vertebrate	Measures conservation in terms of faster/slower evolution than expected[166]. The dataset used is 100 vertebrates.	M	<1.6 = 0, ≥1.6 = 2
SiPhy_29way _logOdds	Estimates likely substitution patterns, as well as accelerated/decelerated mutation rates of bases, by examining patterns of base subsitutions[167].	M	<12.7 = 0, ≥12.7 = 2

The scorep.pl script scored each prediction as 0, 1 or 2 and totalled the sum of the 16 prediction scores. QC was performed by manually confirming scoring and the expected sum totals in the output file. Since frameshifts and truncation mutations would be excluded from analysis by many of the prediction tools (i.e. no prediction would be assigned), the script also totalled the number of missing predictions. QC was performed by manually confirming the total number of missing predictions. The command could be flagged to output all variants, or only variants scoring above a minimum baseline for pathogenicity as follows:

> perl scorep7.pl FAM55tracking.csv svalues.txt P,0,M,0,A,0,AND > Fam55P0M0A0ANDout.csv

> perl scorep7.pl FAM55tracking.csv svalues.txt P,6,M,6,A,0,OR > Fam55P6M6A0ORout.csv

The first command would output all variants with their overall score and number of gaps. The second command would output only those variants with a pathogenic total score of 6 for either the P or M predictors. A minimum baseline of 6 was selected as it would include either: a) 3 predictions with a score of 2 each (3X2); or b) 4 predictions of which half had a score of 2, and half had a score of 1.

Both commands were run for each family, thereafter the output was combined to include all variants with an overall pathogenicity sum total of 6 or more and all variants without pathogenicity predictions due to the nature of the variants. Heterozygous variants in recessive families were discarded at this stage if a second mutation was absent in the same gene.

Remaining variants were then analysed using the scanfexome.pl script (written by Dr. G. Rebello) to ascertain the frequency of each variant within the cohort of 56 individuals subjected to WES. This script would search for each variant (by amino acid effect and gene information) in the larger variant call format (vcf) file for the whole cohort, and extract the genotype occurrences in the larger dataset. A text file,

fexomefamline.txt, was used by the script, and contained relevant pedigree and disease information for each individual, to be presented with the output. In order for the script to function, the column containing the amino acid effect and gene information (referred to as 'column 9') required indexing to ensure no duplicates, and 'fixing' to replace missing column 9 data with unique chromosome coordinates. Both correctional steps were performed for the larger cohort vcf file and the individual family input files:

> perl fixcolumn9.pl FAM583pathogenic.csv comma > FAM583pathFIXED.csv

> perl indexcolumn9v3.pl FAM583pathFIXED.csv comma

> perl "scanfexome1(3).pl" FAM583pathFIXED.csv fexomefamline.txt > FAM583OUT.csv

Manual confirmation was used as QC for this script. The output of scanfexome.pl was manually interrogated to remove variants which occurred frequently and randomly (i.e. without appropriate co-segregation with disease) in other individuals and families in the larger cohort. Variants with a single spurious occurrence were, however, retained in a conservative approach to allow for a potential false positive variant call.

Variants remaining after the analysis pipeline were considered as potential candidates for manual appraisal. Reads and coverage were viewed using Integrative Genomics Viewer (https://www.broadinstitute.org/igv/). Each variant was evaluated with respect to the gene, phenotype and mode of inheritance listed in RetNet (https://sph.uth.edu/retnet/). Variants were then checked in the Leiden Open variation database (LOVD www.lovd.nl/)[168] and the Human Gene Mutation Database (HGMD) (www.hgmd.cf.ac.uk/). Variant MAF's were examined using Ensembl release 8 (http://www.ensembl.org/)[169] with particular emphasis on African subpopulations in the 1000 Genomes Data[112], and African American exome sequencing data from the National Heart, Lung, and Blood Institute (NHLBI)[170]. Furthermore, the MAF's were compared in >60,000 exomes of the Exome Aggregation Consortium (ExAC) Browser[171], and 640 control African chromosomes in the AGVP[115]. The AGVP data includes low coverage whole genome sequences from 100 Bagandan, 100 Zulu and 120 Ethiopian individuals (from Uganda, SA and Ethiopia, respectively). It has recently been suggested that the frequency-centred threshold should not exceed 1:100 000 for an *ad* IRD gene[78], which was considered when evaluating high priority candidate variants.

4.2.1.5 Validation by cycle sequencing

In order to verify candidate variants, primers were designed flanking the mutation(s) of interest, using Primer3 Plus (available at http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and Oligoanalyzer 3.1 (available at https://www.idtdna.com/calc/analyzer). General rules for primer design were: (1) the primer pair should amplify the entire exon if possible, including 10–20bp of intronic flanking sequence, however the amplicon product should not exceed 600bp; (2) forward and reverse primers should have similar melting temperature, in the range of 55–65°C and similar GC content, of approximately 45–50% (3) Hairpin structures should not involve the 3' end of the oligonucleotide, and should melt at temperatures lower than 40°C; (4) Homo- and hetero- dimers ideally have a maximum delta G of -5 kcal/mole, involve no more than 3bp and should not include the 3' end of the oligonucleotides; (5) Primer specificity should be confirmed using the NCBI Primer-blast tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against the Refseq representative genomes database for Homo Sapiens, and the UCSC In-Silico Polymerase Chain Reaction (PCR) tool (http://genome.ucsc.edu/) against the GRCh37/hg19 reference genome assembly.

Primers designed for validation in this particular study amplified a 359bp portion of exon 2 [F: 5' GATGCAGGGCTAAGTATCTTC 3'; R: 5' AATGGAAGTACTCCAGAGTGG 3'] of the *RP2* mRNA NCBI Reference Sequence, NM_006915. PCR was performed under standard conditions in a 25µL reaction containing 200ng DNA, 10pmols each primer, 200µM dNTPs (Bioline, MA, USA), 1X GoTaq® Colourless buffer (Promega, WI, USA) and 0.5U GoTaq® G2 DNA Polymerase (Promega, WI, USA). The final volume was made up with SABAX distilled water (Adcock Ingram, Johannesburg, SA). The cycling conditions were as follows: 94°C – 3 minutes, 30 cycles of {94°C – 30 seconds, 58°C – 30 seconds, 72°C – 40 seconds}, 72°C – 7 minutes, on a Multigene thermocycler (Labnet International Inc., NJ, USA). A volume of 5µL PCR products was visualised using 2% weight/volume (w/v) agarose gels containing 1g agarose (SeaKem® LE, Lonza, Switzerland), 50mL of 1X TBE buffer and 5µL SYBR® Safe DNA Gel Stain (ThermoFisher Scientific, MA, USA). Size estimation of products was performed by comparison against a GeneRuler™ 100bp DNA Ladder Plus (ThermoFisher Scientific, MA, USA).

PCR products were subjected to enzymatic Shrimp Alkaline Phosphatase and Exonuclease I (SAP/Exo) purification as follows: 8.9µL PCR products, 1U FastAP™ Thermosensitive Alkaline Phosphatase (ThermoFisher Scientific, MA, USA) and 2U

Exonuclease I (ThermoFisher Scientific, MA, USA) were incubated in a 10µL final volume on a 2720 thermal cycler (Applied Biosystems by ThermoFisher Scientific, MA, USA) as follows: 37°C – 60 minutes, followed by 75°C – 15 minutes, and a final hold of 95°C – 5 minutes.

Cycle sequencing was performed in a 20µL reaction containing 4–7µL purified PCR product, 20pmols reverse primer, 1X BigDye® Terminator v3.1 Ready Reaction Mix (Applied Biosystems by ThermoFisher Scientific, MA, USA), 1X BigDye® sequencing buffer (Applied Biosystems by ThermoFisher Scientific, MA, USA), and the final volume was made up with SABAX distilled water (Adcock Ingram, Johannesburg, SA). Sequencing reaction cycling was performed on a 2720 thermal cycler (Applied Biosystems by ThermoFisher Scientific, MA, USA) as follows: 96°C – 5 minutes, {96°C – 30 seconds, 50°C – 15 seconds, 60°C – 4 minutes} X 30 cycles.

Cycle sequencing products were purified by ethanol precipitation: samples were mixed with 50µL absolute ethanol (Merck Chemicals, Gauteng, SA) and 2µL sodium acetate (3mM, pH5.5), and incubated overnight at -20°C. Samples were centrifuged at 10,000 revolutions per minute (rpm) for 10 minutes using the Microcentrifuge 5415D (Eppendorf, NY, USA). The supernatant was removed, after which samples were washed in 50µL of 70% ethanol. A repeat centrifugation at 10,000 rpm for 10 minutes was followed by the removal of supernatant, upon which samples were air dried for 60 minutes. A volume of 5µL purified sequencing reaction was added to 8µL Hi-Di™ Formamide (Applied Biosystems by ThermoFisher Scientific, MA, USA) and capillary electrophoresis was performed on an ABI Prism 3130xl Genetic Analyser (Applied Biosystems by ThermoFisher Scientific, MA, USA). Sequences were aligned to the reference sequence (NM_006915) using Bioedit Sequence Alignment editor v.7.1.3.0[172].

4.2.1.6 Clinical examination

In order to review the phenotype of a semi-dominant xlRP family diagnosed in this investigation, and to compare it with published findings, some participants were re-evaluated. Dilated retinal examinations were performed by the same ophthalmologist on multiple family members (n=11) and their visual acuity was tested. Other clinical information was available from a previous clinical examination (n=1), or family history taken during counselling sessions and recruitment. Electroretinography (ERG) and Optical Coherence Tomography (OCT) was

unavailable in the Eastern Cape Province where the family resides, at the time of the clinical examination.

4.2.2 Results

Approximately 12 million reads (n=12,610,197) were captured per sample, and ~1.8 million variants identified in the 16 indigenous African families. However, 34,579 variants were identified in the updated list of RetNet genes, and 4,866 of these were retained after eliminating intergenic, intronic and synonymous variants (Appendix 4). Four of the ten families showed male-to-male transmission of the disease. The remaining six families comprised four apparent adRP and two arRP pedigrees, each thus had potential for actually having xlRP, and were analysed accordingly.

On average ~2,140 IRD variants were identified in each of the 10 families (Range: 1,881–2,329), of which ~1,355 had a low MAF (Range: 1,168–1,502) and ~288 (Range: 232–359) resided in exonic or splicing locations. On average, ~12 variants per family co-segregated appropriately with disease (Range: 2–36), and ~5 of these were predicted to be pathogenic (Range: 0–13), however most variants were subsequently eliminated based on their frequency in the larger cohort and/or publically available control data.

Ultimately all but two variants were excluded via the analysis pipeline and subsequent manual appraisal, namely c.1144G>A; p.(Val382Ile) in exon 9 of *AFG3L2* (Chapter 4.3), and an *xl* mutation in a pedigree with an apparent *ad* inheritance pattern of disease. This c.704C>A; p.(S235*) mutation in exon 2 of *RP2* (NM_006915) was identified in a large Xhosa family with an initial diagnosis of adRP. This mutation was predicted (using Mutalyzer v.2.0.21[173]) to truncate 116 C-terminal amino acids from the wild type 350 amino acid protein (Figure 4.2.1). Whilst the X chromosome was not included in the AGVP data, the c.704C>A variant was not listed in Ensembl, nor was it reported in the ExAC Browser, LOVD or HGMD database (accessed 24 February 2016), indicating it is novel and exceedingly rare.

```
  1   MGCFFSKRRK ADKESRPENE EERPKQYSWD QREKVDPKDY MFSGLKDETV GRLPGTVAGQ
 61   QFLIQDCENC NIYIFDHSAT VTIDDCTNCI IFLGPVKGSV FFRNCRDCKC TLACQQFRVR
121   DCRKLEVFLC CATQPIIESS SNIKFGCFQW YYPELAFQFK DAGLSIFNNT WSNIHDFTPV
181   SGELNWSLLP EDAVVQDYVP IPTTEELKAV RVSTEANRSI VPISRGQRQK SSDESCLVVL
241   FAGDYTIANA RKLIDEMVGK GFFLVQTKEV SMKAEDAQRV FREKAPDFLP LLNKGPVIAL
301  EFNGDGAVEV CQLIVNEIFN GTKMFVSESK ETASGDVDSF YNFADIQMGI *
```

Figure 4.2.1 Predicted truncation of the RP2 protein sequence by the c.704C>A mutation. The sequence highlighted in red is absent when the protein is truncated, resulting in a predicted loss of 116 amino acids.

The presence of the mutation was confirmed by cycle sequencing in the four relatives originally selected for WES. Furthermore, cycle sequencing of samples from all family members in the extended pedigree (n=18) established that the variant co-segregated with disease (Figure 4.2.2, Figure 4.2.3). According to recently published guidelines for the interpretation of sequence variants[174] this variant meets the specified combination of criteria that classify it as pathogenic, namely it is a nonsense mutation (very strong evidence) that is absent from controls (moderate evidence) and co-segregates with disease (supporting evidence).

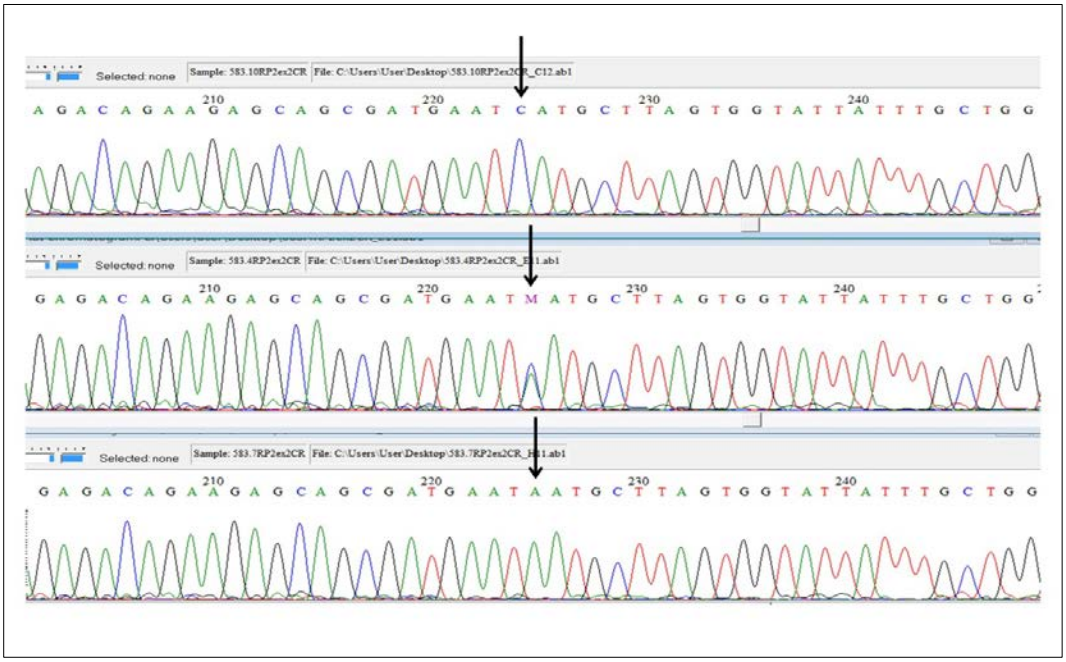


Figure 4.2.2 Sequencing electropherograms showing the wild type *RP2* sequence (top), and heterozygous (middle) and hemizygous (bottom) c.704C>A mutation in members of family RP 583.

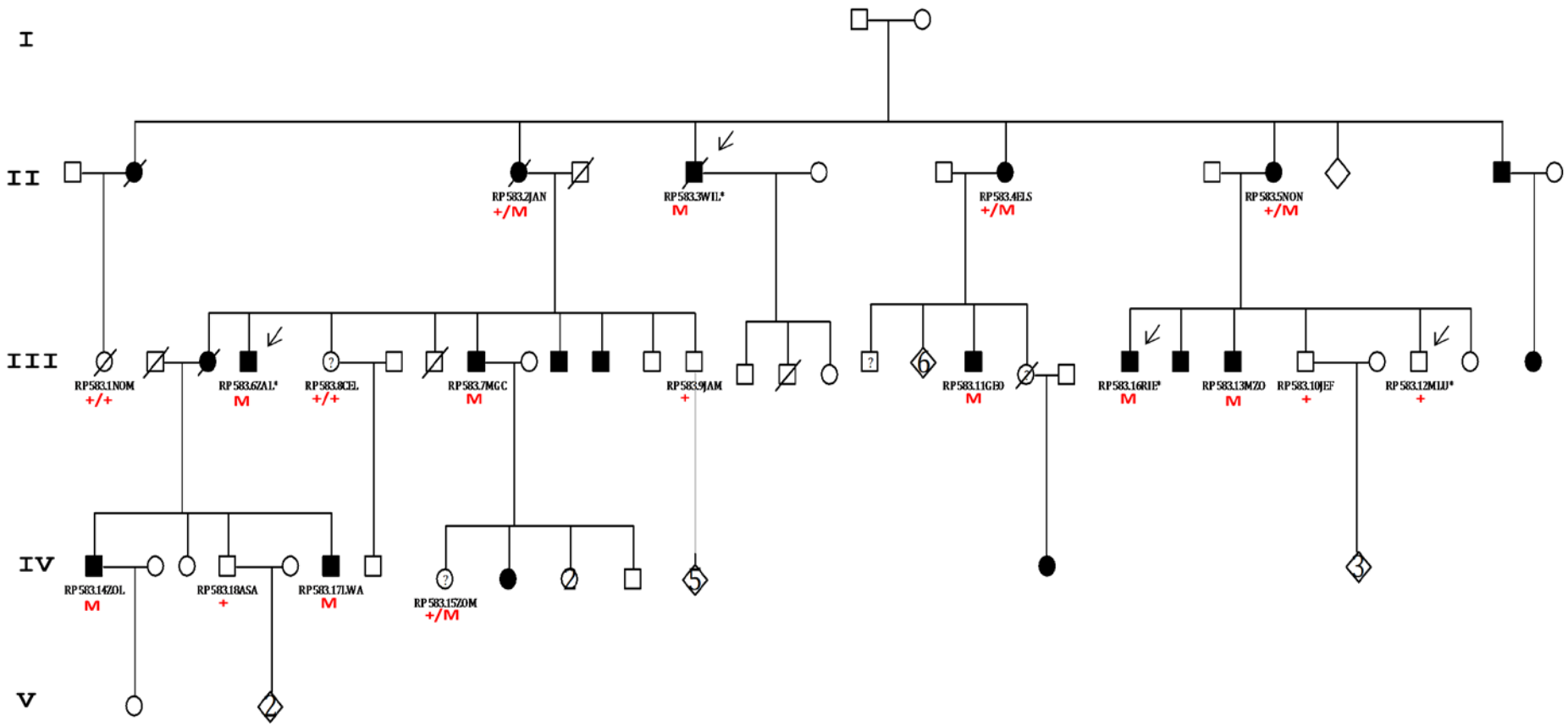


Figure 4.2.3 Pedigree of family RP 583, demonstrating co-segregation of the *RP2* c.704C>A mutation with disease.

Squares represent males, and circles, females. Shaded symbols indicate individuals affected with IRD. Identifier codes show individuals from whom biological material was available, and those selected for whole exome sequencing are noted with an arrow. Segregation of the mutation in the families is indicated as: M (hemizygous mutation); +/M (heterozygous mutation); +/+ and + are wild type alleles in females and males, respectively.

Clinical information of family members is presented in Table 4.2.3.

Table 4.2.3 Clinical information for the members of family RP 583. Females manifesting with IRD are indicated in bold.

Individual number in the pedigree	DNA identifier	Clinical description
II 5, (M)	RP 583.3WIL	AOO 16 years
II 8, (F)	**RP 583.4ELS**	**AOO 18 years, no typical RP features, atypical macular changes, bilateral pigment clumping and atrophy in macula, bilateral tessellated fundus, slowly progressive, bilateral cataracts at age 82.***
II 10, (F)	**RP 583.5NON**	**AOO 15 years, diffuse RP (grade 3 with macular involvement), slowly progressive, cataracts.***
II 12, (M)	N/A	Onset childhood, no typical RP features, atypical macular dystrophy, bilateral geographic macular atrophy, slow progression.*
III 4, (M)	RP 583.6ZAL	Onset early childhood, diffuse RP with macular involvement, slow progression.*
III 5, (F)	RP 583.8CEL	No nyctalopia, no visual complaints, no reduction to visual fields, however both retinas have a diffuse mottled, granular appearance. No refractive error.*
III 8, (M)	RP 583.7MGC	AOO 16 years
III 10 (M)	N/A	AOO 20 years, Diffuse RP, slowly progressive *
III 19 (M)	RP 583.11GEO	AOO 6 years, original diagnosis was STGD, diffuse RP with macular involvement, slow progression, early cataracts, high myopia.*
III 22 (M)	RP 583.16RIE	AOO 10 years, diffuse RP with macular involvement, slow progression, cataracts.*
III 24 (M)	RP 583.13MZO	AOO 12 years
III 29, (F)	**N/A**	**AOO 30 years, both retinas have a diffuse mottled, granular appearance, but no bone spicule pigmentation. Slowly progressive.***
IV 1, (M)	RP 583.14ZOL	AOO 11 years, diffuse RP with macular involvement, slow progression, epiretinal membranes, Pseudopapilloedema.*
IV 6, (M)	RP 583.17LWA	AOO 3 years.
IV 7, (M)	N/A	Normal retinal examination*
IV 13, (F)	**N/A**	**Onset early childhood, diffuse mottling/granularity of the RPE and macular area, severe refractive error with anisometropia.***

(M) male, (F) female, “AOO” Age of Onset, * Clinically confirmed.

4.2.3 Discussion

Missing heritability is an issue in the investigations of IRD, with known disease genes accounting for only 50% of the cases[17]. It has been estimated that 20–30% of arRP cases, 60–70% of adRP cases and 80–85% of xlRP cases can be diagnosed using a combination of Sanger Sequencing and targeted capture NGS[175]. WES with targeted analysis of candidate genes for IRD detects 49–83% cases in different population groups[44–49] and ~38% of the indigenous South African cohort was diagnosed using this approach (Chapter 4.1). However one large Xhosa family was subsequently identified as carrying a causative pathogenic mutation in an xlRP gene, that had been overlooked previously when the family was analysed using an *ad* model. Incorrect assumption of the Mendelian inheritance pattern had produced a false negative result, emphasising that all pedigrees lacking male-to-male transmission of IRD should be considered for mutations in xlRP genes. It has recently been suggested that targeted capture of the X chromosome, or 'X-exome sequencing' may be useful in identifying causal variants[176], yet this only applies in cases where the *xl* inheritance pattern is obvious. Additionally, whilst it is clear that approaches such as WES and targeted capture NGS are preferable to traditional screening methods, it must be considered that the mutation hotspot in exon ORF15 of *RPGR* is repetitive and purine rich[177] and may not be captured adequately by these technologies (Chapter 5.2.2).

A majority of mutations in *RPGR* and *RP2* are predicted to result in a truncated protein, implying the loss of gene function causes xlRP[132,135,137,178]. N-terminal RP2 has homology to the tubulin-specific chaperone cofactor C and forms a complex with the small G-protein Arl3, suggesting that RP2 is a GTP'ase activating protein for Arl3[179]. In photoreceptor cells, Arl3 is localised in the microtubule-rich connecting cilium (which forms a bridge between the IS and OS), and RP2 is localised to the photoreceptor plasma membrane, suggesting they function in the ciliary transport and distribution of proteins like rhodopsin. Photoreceptors are neurons which are exceptionally metabolically active, demanding accurate protein synthesis, sorting and trafficking.

The C-terminus of RP2 (predicted to be abolished by c.704C>A), has similar structure to nucleoside diphosphate (NDP) kinase, which catalyses the phosphorylation of nucleoside diphosphates to triphosphates, but there are significant differences around the active site[180] between these two proteins, hence the function of this domain in RP2 is unclear. Nevertheless the p.(Ser235*) mutation

identified in the present study falls within the ferredoxin-like fold of the C-terminal α/β domain and should cause a loss of function similar to other reported *RP2* truncation mutations.

The clinical expression of the c.704C>A mutation in this large Xhosa family is that of diffuse RP with an early age of onset in males ranging from 3 to 16 years. The disease is characterised by slow progression in many individuals. Interestingly, four individuals had cataracts. Eight of the 15 affected individuals with clinical information had macular involvement, including one with an initial diagnosis of STGD. This is in accordance with the recognisable RP2 phenotype (early onset macular atrophy, high myopia and reduced visual acuity)[181], although only one individual was reported with myopia in this family. Of the five affected females examined, one (RP 593.5NON) was classified as Grade 3, according to the Grover et al classification for manifesting female carriers[140]. The remaining females did not match the classification system, as they had reduced vision with either: (1) Pigment clumping in the macula but without typical RP features (RP 584.4ELS); or (2) a diffuse mottled, granular appearance of the retina but without bone-spicule pigmentation (III 29 and IV 13).

Individual RP 583.8CEL presented an interesting case, as she did not carry the mutation but had abnormal fundus photographs indicating some regions with pigment clumping and the diffuse mottled granular appearance referred to as 'salt-and-pepper fundus'. However, this patient, unlike the other females examined, did not experience nyctalopia, visual complaints or visual field reduction. Multiple archived blood samples were tested for this case, eliminating the possibility of a sample substitution error. There are thus two explanations for this scenario, either: (1) the observed clinical abnormalities are not due to the mutation, but instead are as a result of congenital infection or chronic medication; or (2) genetic mosaicism has occurred, and although the mutation is not present in the blood it may be present in her other tissues, including the retina. Intriguingly, a germ-line mosaic for an *RP2* mutation has already been reported in SA previously, albeit for a different mutation and in a Caucasian family[182].

There is intrafamilial clinical heterogeneity observed amongst the females, still the large number of women manifesting with visual complaints is interesting. Whilst many semi-dominant families with *RPGR* mutations have been reported, to our knowledge this is only the second such family to be described with an *RP2* mutation. In 2009, Pomares et al[143] described a large Spanish pedigree with a wide phenotypic range in female carriers due to an intronic *RP2* mutation, and reporting

that '*RP2* joins *RPGR* as the cause for semi-dominant xlRP'. The c.1073-9T>A mutation reported in that study resulted in exon skipping and protein truncation, however the aberrant protein escaped nonsense-mediated decay, making it difficult to establish whether the mutation caused a gain or loss of function. In that family, affected males expressed 95% of the aberrant and 5% of the wild type product respectively, whereas carrier females expressed a range (8–90%) of aberrant transcript levels, leading the authors to suggest skewed X-inactivation as the cause of the phenotypic range observed in females.

X-inactivation status, however, cannot be measured directly in the retina[139], and indirect measurements (in lymphocytes) show random X-inactivation in the majority of female carriers regardless of their affection status[144,183]. Recently, adaptive optics with scanning laser ophthalmoscopy has allowed non-invasive study of individual cone photoreceptors in obligate xlRP carriers[142]. Although only five patients were examined in that study, each showed arrangements of healthy, unaffected cones and regions with abnormal cones distributed in irregular patterns. This prompted the authors to suggest that healthy cones expressed predominantly the wild type X-allele and abnormal cones principally expressed the mutant allele i.e. skewed X-inactivation occurred in a mosaic pattern across the retina in all five carriers with different mutations. Skewed X-inactivation is possibly the cause of the occasional affected carriers within xlRP families. However, this is unlikely to be the reason for semi-dominant families, unless the chromosome inactivation is consistently skewed in retinal cells by a genetic modifier in a chromosomal region co-segregating with the primary xlRP mutation[136,144].

Banin et al[144] examined several potential explanations for the phenomenon of semi-dominant xlRP pedigrees, and eliminated the following factors in phenotype severity in carriers: 1) it is not gene-specific as both *RPGR* and *RP2* have been associated with both forms of xlRP (i.e. both dominantly- and recessively- manifesting); 2) it is not mutation-specific as the same mutation, on different chromosomal backgrounds, can cause both forms. The authors concluded that an additional modifier allele linked to *RPGR*, either in the gene promoter or in the chromosomal segment segregating with the causative mutation, may affect penetrance in female carriers. Additional evidence for genetic modifiers that affect disease severity is the report of discordant phenotypes observed between non-identical twin males with the same *RPGR* mutation[20].

The identification of the novel *RP2* mutation in this large African family holds potential for prospective analysis of the genetic variants modifying disease

expression in xlRP carrier females. Improved understanding of genetic modifiers is key to providing more accurate prognosis for patients and measuring efficacy of potential treatments[184]. Of more immediate benefit is the molecular diagnosis of this family (mistakenly assumed to have adRP), finally allowing accurate recurrence risk prediction and genetic counselling for family members.

4.3 Expanding the disease spectrum associated with *AFG3L2* mutations

The second significant finding which emerged from re-analysis of the ten unresolved families, utilising the updated set of IRD genes and multiple inheritance models, was the identification of a putative mutation in *AFG3L2* in a Sotho-Tswana family.

4.3.1 Methods

4.3.1.1 WES analysis pipeline

The 714 gene symbols included four genes not yet recorded on the RetNet database (personal communication with Prof. S. Daiger, The University of Texas HSC at Houston). The reasons these particular genes (*RTN41P1*, *SLC19A2A*, *CO2* and, importantly, *AFG3L2*) were being considered for RetNet inclusion were not disclosed.

Candidate variants were prioritised using the scripts, criteria and analysis pipeline defined previously. However in order to improve the disease-risk prediction of the primary candidate, an additional, larger dataset namely the Beacon Network[185] was accessed. This is a federated system (https://beacon-network.org/#/), whereby a meta-database is created through virtual connections between multiple autonomous databases. The Beacon Network can be queried simply for the presence or absence of a specific allele, using chromosome coordinates. The tiered-access approach circumvents patient/data privacy concerns by providing additional details only upon authorisation.

4.3.1.2 Validation by cycle sequencing

Primers were designed to confirm the mutation of interest, according to the criteria described (Chapter 4.2). Primers for validation in this particular study amplified a 559bp portion of exon 9 [F: 5' AGCCCAAGTCACCCTTTACA 3'; R: 5' CCAGGGACTGGTGAGAGGT 3'] of the *AFG3L2* mRNA NCBI Reference Sequence, NM_006796.2. PCR was performed under standard conditions

described in Chapter 4.2. The cycling conditions were as follows: 94°C – 3 minutes, 30 cycles of {94°C – 30 seconds, 60°C – 30 seconds, 72°C – 40 seconds}, 72°C – 7 minutes, on a Multigene thermocycler (Labnet International Inc., NJ, USA). A volume of 5µL PCR products was visualised using 2% weight/volume (w/v) agarose gels as described in Chapter 4.2.

PCR products were subjected to enzymatic SAP/Exo purification and cycle sequencing was performed in a 20µL reaction as described (Chapter 4.2). Sequences were aligned to the reference sequence (NM_006796.2) using Bioedit Sequence Alignment editor v.7.1.3.0[172].

4.3.1.3 Clinical examination

Members of family RPD 401 were ophthalmologically re-examined as the clinical information available was sparse at the outset of this research project, only noting signs of RP in several family members, with reduced penetrance in some. Furthermore, the reported phenotype associated with the primary candidate gene identified in this investigation was unusual. Thus, retinal examinations were performed by two ophthalmologists on multiple family members (n=5). Additional clinical information was collected from previous medical evaluations and the anecdotal family history taken during counselling sessions for recruitment. The previous medical examinations included computed tomography (CT) brain scan, and electroencephalogram (EEG), of two relatives respectively. ERG and OCT were not performed due to cost considerations.

4.3.2 Results

A variant of interest, c.1144G>A; p.(Val382Ile), was identified in exon 9 of *AFG3L2* in the Sotho-Tswana family RPD 401.The pedigree displayed male-to-male disease transmission, therefore the family had been analysed using an *ad* inheritance model. The number of variants in this family, after each filtering step, are listed in Table 4.3.1, beginning with 2,002 IRD gene variants that were neither intergenic, intronic nor synonymous. The details pertaining to the five candidate variants that were manually evaluated as potential causative mutations are presented in Table 4.3.2.

Table 4.3.1 Number of candidate variants in family RPD 401 after prioritization filters.

No. of IRD variants	<0.1 MAF	Exonic or splicing	Co-segregating within family	Pathogenic	Candidates (rare and co-segregating in cohort)
2,002	1,262	270	36	13	5

Table 4.3.2 Summary of the five candidate variants in family RPD 401 evaluated as putative causative mutations.

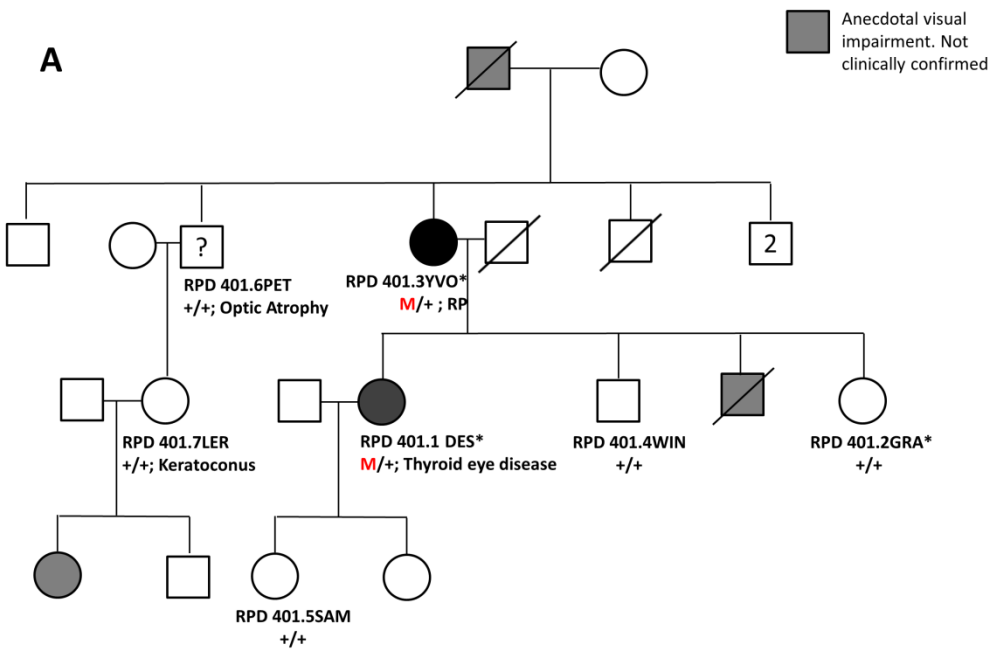
Variant	ExAC MAF	1000G MAF	SNP ID	Pathogenicity Score			Missing Pathogenic Scores (gaps)		
				P	M	All	P	M	All
CNGB1: NM_001135639 c.376G>C; p.(Ala126Pro)	6.38E-04	0.0014	rs200332871	9	1	10	1	0	1
EYS: NM_001142800 c.4543C>T; p.(Arg1515Trp)	0.096	0.094	rs62415827	7	1	8	1	0	1
PLK4: NM_001190799 c.1765G>T; p.(Val589Leu)	.	.	.	2	6	8	0	0	0
AFG3L2: NM_006796 c.1144G>A; p.(Val382Ile)	.	.	.	16	12	28	0	0	0
KIZ: chr20 (position 21106804 C>T); unknown	1.41E-03	0.0018	rs150243168	0	0	0	9	7	16

The *CNGB1*, *EYS*, and *KIZ* variants were excluded, as allele frequencies for these variants in the ExAC Browser[171] (accessed 24 February 2016) were high, at 79/120,638 alleles, 2,042/19,786 alleles, and 61/9,602 alleles respectively. These variants each vastly exceed the likely frequency threshold of true dominant, pathogenic RP mutations (1:100,000 individuals[78]).

The *PLK4* variant was not listed in the ExAC browser. However, this nucleotide position was multiallelic, and a G>A variant at the same genomic coordinate (Chromosome 4:128812277) resulted in a Valine to Methionine substitution with the

reference SNP ID rs773122147. This alternative allele (G>A) had a frequency of 1/120,830 alleles in the ExAC browser[171] (accessed 24 February 2016). *PLK4* is associated with *ar* microcephaly, dwarfism, severe intellectual disability, and retinopathy[186]. This mode of inheritance conflicted with that observed in family 401. Furthermore, the *PLK4* variant had a much lower predicted pathogenicity score than the *AFG3L2* variant.

The strongest candidate was c.1144G>A; p.(Val382Ile) in exon 9 of *AFG3L2*. This variant had the highest predicted pathogenicity score of the five putative mutations, with 14 of the 16 prediction tools reporting it as pathogenic, and an overall score of 28. The c.1144G>A variant was not recorded in any of the datasets examined (accessed 24 February 2016), namely LOVD (www.lovd.nl/)[168], HGMD (www.hgmd.cf.ac.uk/), 1000 Genomes Data[112] or African American NHLBI exome sequencing data[170] in Ensembl release 8 (http://www.ensembl.org/)[169], the ExAC Browser[171], or AGVP data[115]. Furthermore, the only representation of this allele (chromosome 18: 12356713 C>T) in the Beacon Network[185] (accessed 14 September 2016) occurred in an *in silico* prediction database containing almost every possible single-nucleotide variant i.e. the variant was not present in any patient sample (personal communication with David Caplan of SolveBio Genomic Intelligence, https://www.solvebio.com/). This indicates that this novel variant is very rare, as expected for an *ad* IRD mutation. The variant was not detected in the remaining cohort of 56 individuals, and was in fact the only exonic *AFG3L2* variant identified in the WES data (four 3' UTR and three upstream variants were excluded). The heterozygous variant was subsequently confirmed via cycle sequencing (Figure 4.3.1) and initially appeared to track with disease in the family, occurring in individuals RPD 401.1DES and 401.3YVO, and absent in the five other family members tested. However, upon clinical evaluation, individual RPD 401.1DES, who was included for WES under the assumption that she had RP, was subsequently diagnosed with thyroid eye disease (Table 4.3.3).



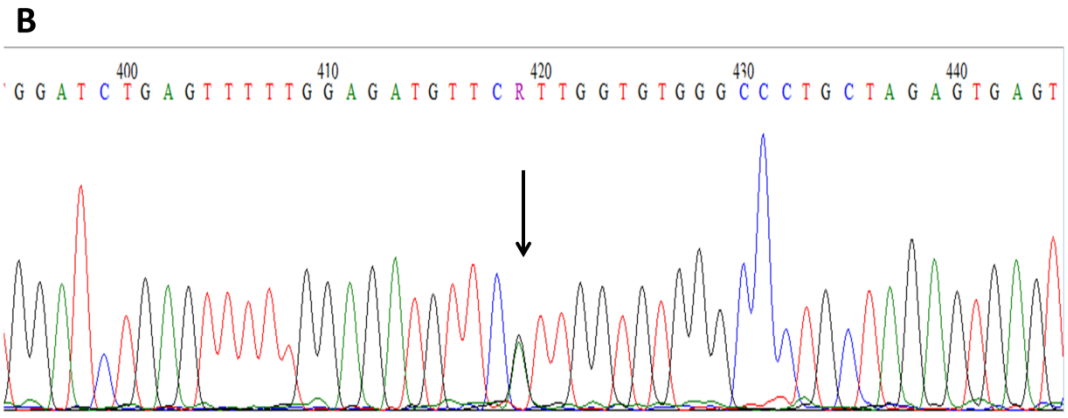


Figure 4.3.1 (A) Pedigree of family RPD 401, showing the presence of the novel c.1144G>A variant in *AFG3L2* in two family members Squares represent males, and circles, females. Shaded symbols indicate individuals affected with IRD: black represents RP, dark grey represents thyroid eye disease and light grey represents individuals with visual impairment that has not been clinically confirmed. Confirmed clinical diagnoses are listed. Identifier codes represent the individuals from whom biological material was available, and those selected for whole exome sequencing are noted with an asterisk. Segregation of the variant in the families is indicated as: M/+ (heterozygous variant) and +/+ (wild type alleles). **(B) Sequencing electropherogram showing the heterozygous *AFG3L2* c.1144G>A variant (indicated by an arrow).**

Table 4.3.3 Clinical information for the members of family 401 with ocular phenotypes

DNA identifier	Clinical description
RPD 401.1DES	Thyroid eye disease. No IRD or optic atrophy. CT scan performed in 2015 showed residual proptosis after bilateral retraction repairs. Optic nerve and visualised brain appear normal.
RPD 401.3 YVO	AOO 30 years. Diffuse RP. Hearing loss. Hypertension. EEG performed to determine whether she had a seizure in 2013, was normal (no epileptiform activity).
RPD 401.6PET	AOO 18 years. Cavernous optic atrophy, due to low tension glaucoma or dominant optic atrophy. Visual acuity 6/60. Normal intra-ocular pressures and nerve fibre assessment. No maculopathy or RP-like features. Slight dysarthria, but no obvious ataxia or cerebellar dysfunction.
RPD 401.7LER	Keratoconus. No optic nerve or retinal pathology.

"AOO" Age of Onset

4.3.3 Discussion

Given the benefit of hindsight due to the clinical re-evaluation, family RPD 401 would perhaps not have been included in the WES cohort as there is no clear (single) IRD phenotype. Nevertheless, *AFG3L2* is an interesting potential candidate gene for further investigation in this family, particularly as no causative mutation was identified in any other known IRD genes. *AFG3L2* is ubiquitously expressed[187] including in the retina[10,13]. It is a pleiotropic gene in humans (Figure 4.3.2), with mutations causing a range of effects such as *ad* spinocerebellar ataxia type 28 (SCA28)[188,189] and dominant optic atrophy (DOA)[56].

The AFG3L2 protein is a mitochondrial metalloprotease. It contains a conserved adenosine triphosphate (ATP) binding site in a AAA domain[187] ('adenosine triphosphatases associated with diverse cellular activities'). AFG3L2 forms homo- (with itself) and hetero- (with paraplegin, encoded by *SPG7*) oligomeric proteases, known as *m*-AAA proteases in the inner mitochondrial membrane. These proteases assert protein quality control via chaperone activity, facilitating the correct assembly of the respiratory chain complexes[190].

Two AFG3L2 mouse models, a null mutant and a missense mutation carrier, both display severe neuromuscular defects due to impaired axonal development[190]. Respiratory chain complex enzymatic activity is impaired in the models, affecting mitochondrial metabolism and morphology of neuronal tissue (but without affecting non-neuronal tissue). These data indicate a key role for AFG3L2 in neurons.

The SCA28 phenotype in humans is characterised by slow disease progression, imbalance, gait and limb ataxia, dysarthria (slurred speech), eye movement abnormalities (nystagmus), ptosis (drooping eyelids) and ophthalmoparesis (weak extraocular muscles)[188,191]. It has been established that about 78% of SCA28 patients have at least one of these ocular signs[192]. As is the case with mouse models, the mitochondrial abnormalities are restricted to neuronal tissue in humans[189]. The vast majority of SCA28 mutations occur in exons 15 and 16, affecting the proteolytic domain[189,191,192], however one mutation has also been reported in exon 10, affecting the AAA domain[189].

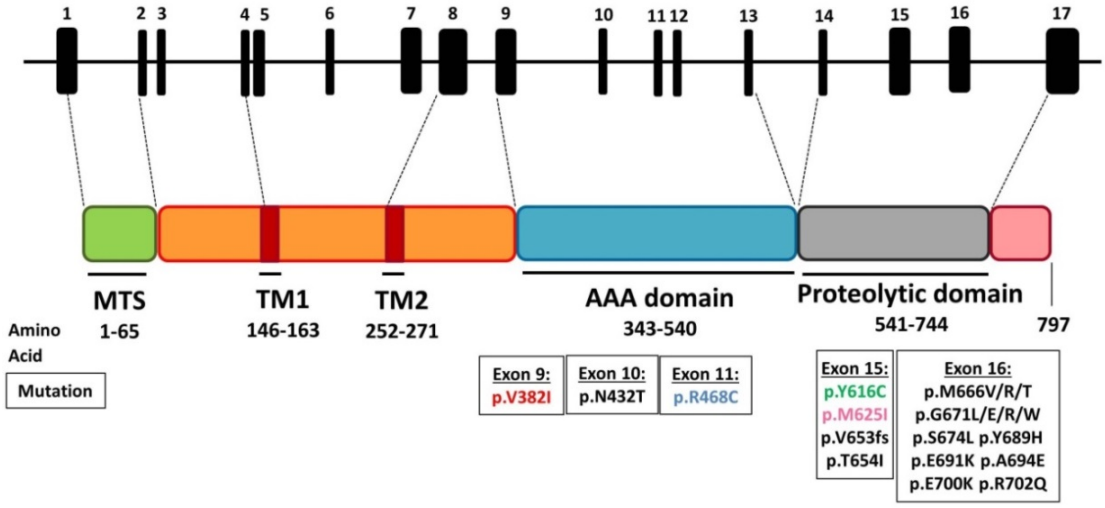


Figure 4.3.2 Illustration showing the *AFG3L2* exons above the protein domains they encode. The amino acid numbers involved in each domain are presented, and various mutations causing different phenotypes are given in different coloured text: spinocerebellar ataxia type 28 (black); progressive myoclonus epilepsy (pink); spastic ataxia neuropathy (green); dominant optic atrophy (blue); retinitis pigmentosa and thyroid eye disease in this study (red). Legend: MTS, mitochondrial targeting sequence; TM1/TM2, transmembrane domain 1 and 2. *Image adapted from references*[56,189]

Two classes of AFG3L2 mutations have been proposed, which explain the dominant and recessive phenotypes observed: (1) dominant negative mutations which affect both homo- and hetero- complexes and are thus fully penetrant, and (2) haploinsufficient mutations, which are alleviated when oligomers are formed with wild type AFG3L2 or paraplegin thus demonstrating reduced penetrance. The mutations do not appear to affect protein stability, but rather interfere with

specialised functions of the m-AAA proteases such as substrate recognition and/or chaperone function[189].

A homozygous mutation (p.Tyr616Cys) in exon 15 has been previously identified in siblings of a consanguineous family with a spastic ataxia neuropathy syndrome[193]. The siblings had an early onset phenotype with spastic gait, dystonia (abnormal muscle tone), dysarthria and progressive myoclonic epilepsy. The ocular signs included oculomotor apraxia (absence of controlled, purposeful eye movement) and ptosis. Cerebellar atrophy was also observed. This mutation, again, resides in the proteolytic domain. However, this mutation was functionally proven to be hypomorphic and did not result in complete loss-of-function (both proteolytic function and ATP'ase activity was retained)[193]. Moreover, in yeast studies the mutation affected the homo-oligomeric proteases and, to an even greater extent, the hetero-oligomeric complexes. The isoenzymes can function synergistically, partially substituting for each other. However with both complexes affected, functional m-AAA protease levels were low. This resulted in a recessive phenotype, combining features of both AFG3L2 and paraplegin deficiencies. Two additional cases of homozygous mutations, proximate to p.Tyr616Cys, have been reported[194]. In these two apparently unrelated individuals, a homozygous p.(Met625Ile) mutation was identified as causing progressive myoclonic epilepsy[194].

Most recently, a novel *AFG3L2* mutation was identified in a family with DOA[56]. DOA causes vision loss due to degeneration of the optic nerve, and can be associated with other symptoms like muscle weakness, ataxia and peripheral neuropathy. The most frequent associated symptom is neurosensory deafness. All identified DOA genes (*OPA1*, *OPA3* and *SPG7*) encode mitochondrial membrane proteins, yet this was the first report linking *AFG3L2* to the phenotype[56]. The causative mutation did not reside within the proteolytic domain, like most previously described mutations, but in the AAA domain (exon 11). The clinical description of the DOA family included reduced visual acuity, photophobia, colour vision impairment, and optic disc pallor. Importantly, a moderate to marked reduction in retinal nerve fibre thickness was observed. No other oculomotor or neuromuscular problems, ataxia or cerebellar atrophy were evident, although mild intellectual disability was reported[56]. The DOA gene, *OPA1*, is a substrate for AFG3L2 and it has been suggested that the p.(Arg468Cys) mutation in this family may disrupt the interaction, resulting in this particular clinical manifestation.

To the best of our knowledge, the variant identified in this Sotho-Tswana African family is the first reported in exon 9 of *AFG3L2*. From the sparse clinical information it was initially inferred that the family was affected with adRP. However, upon clinical re-evaluation the two 'mutation-positive' individuals were diagnosed with RP with hearing loss, and thyroid eye disease, respectively. Whilst functional studies are required to ascertain whether this novel c.1144G>A; p.(Val382Ile) variant is pathogenic[174], it affects an amino acid in the AAA domain[189] that is conserved amongst homologues[187]. Indeed, p.Val382 exists in the highly conserved central pore loop motif F/YVG[189]. It has been suggested that this particular motif is essential for the recognition of substrates and ATP-depending movement of polypeptides into the proteolytic chamber[189]. The p.(Val382Ile) missense mutation may therefore affect the interaction between m-AAA proteases and their substrates.

The combination of phenotypes in the p.(Val382Ile) carriers is very interesting. The ophthalmologist noted hearing loss (a commonly associated symptom of DOA), but no signs of ataxia in the patient with RP (RPD 401.3 YVO). There were no obvious neurological problems in the patient with thyroid eye disease (RPD 401.1DES), who had a CT scan showing normal brain and optic nerve.

Thyroid eye disease, also known as Graves' disease, is an autoimmune disorder characterised by proptosis (bulging eyes), eyelid retraction and oedema[195]. Graves' ophthalmopathy mainly occurs in patients with hyperthyroidism due to Graves' disease (which is systemic). There are, however, cases of Grave's ophthalmopathy without Graves' disease or hyperthyroidism[195]. Intriguingly, one of the Graves' disease risk loci, *RAC2* (an important immune response gene[196]), encodes a product which interacts with AFG3L2[197]. Thus, it would be interesting to perform additional investigations into whether genetic modifiers could influence the phenotype such that it resembles thyroid eye disease. Variants in RAC2, for example, may affect it's interaction with the p.(Val382Ile) form of AFG3L2. There are 26 *RAC2* variants in family RPD 401, seven of which occur in the individual with thyroid eye disease. Five of these variants are present in the thyroid eye disease case and absent in the individual with RP, or are homozygous in the thyroid eye disease case but heterozygous in the RP case. None of these five variants occur in coding regions, but they might affect important regulatory elements.

Individual RPD 401.6PET, exhibiting cavernous optic atrophy, did not carry the *AFG3L2* variant. The WES data was subsequently examined for variants in the DOA genes (*OPA1*, *OPA3* and *SPG7*) in RPD 401.3 YVO only. This was performed in order to exclude variable expressivity as the cause of the combination of RP and

DOA in the family. No rare exonic or splice variants of these genes were present in the RP patient. Moreover, the phenotype of individual RPD 401.6PET could not be distinguished between low tension glaucoma or DOA, suggesting his visual impairment could be coincidental. Indeed, his normal nerve fibre assessment is in contrast with the DOA family described previously in the literature[56]. Alternatively, the presence of dysarthria may imply genetic mosaicism, which may explain why the *AFG3L2* variant was not detected in this patient's buccal DNA sample.

There is currently insufficient evidence to translate this family into "diagnostic" mode. Furthermore, the lack of a single, heritable phenotype complicates additional WES data analysis for this family. However, if further investigations prove the *AFG3L2* variant to be functional it would expand the phenotypic spectrum associated with this gene to include the first case of RP. The question would then remain whether the RP phenotype is because the variant affects the m-AAA protease recognition of — or interaction with — a specific retinal substrate, or whether retinal-specific thresholds exist for certain substrates that are perhaps more sensitive than other tissues.

Chapter 5. Analysis of exome data towards identifying potential novel candidate genes

5.1 *IDH3A* identification, facilitated by analysis of the European Retinal Disease Consortium candidate genes list

The genetic heterogeneity displayed by IRDs decreases the likelihood of identifying multiple families with different mutations in a novel gene, particularly within a single, small study cohort. The use of consortia, like the European Retinal Disease Consortium (ERDC, http://www.erdc.info/), is therefore valuable for researchers working on rare, heterogeneous diseases, allowing joint analysis efforts to take place in a collaborative manner. The ERDC is a consortium, comprising 15 international research groups, which has published a list of novel IRD gene candidates on its website. For each of these genes, the likely causative variant(s) has been identified in a single family. In the present study, eight indigenous African families remained without molecular diagnosis after interrogation of all known IRD genes, and were subsequently analysed in a targeted approach for variants in the ERDC candidate genes.

5.1.1 Methods

5.1.1.1 WES analysis pipeline

The WES methodology and analysis strategy has been described in detail (Chapter 4.2). After intergenic, intronic and synonymous variants were excluded, the genes of interest were extracted from the data. Although 26 genes were reported by the ERDC (http://www.erdc.info/#!candidateirdgenes/cihc accessed 27 January 2016),

four (*ACBD5*, *IFT140*, *PRPS1*, and *ZNF513*) were not included in this enquiry, as they had already been incorporated in the preceding RetNet gene analysis (Appendix 4). The commands used to extract the genes of interest from the WES were (for n=26 and n=22 ERDC genes, respectively):

> grep –w –f ERDC20160127.csv NOintergenic_intronic_OR_synonymousDATA.csv > Filtered2016ERDCDATA.csv

> grep –w –f ERDC20160127COMPACT.csv NOintergenic_intronic_OR_synonymousDATA.csv > COMPACTFiltered2016ERDCDATA.csv

The WES data were separated into the eight families of interest using Microsoft Excel software. The prioritisation filters were applied to variants in each family, using the scripts, criteria and analysis pipeline described previously (Chapter 4.2).

5.1.1.2 Validation by cycle sequencing

Primers were designed to confirm the variant of interest, according to the criteria described in Chapter 4.2, and amplified a 330bp fragment spanning exon 5 [F: 5' TTTCACAAGGTAGCCGAGGT 3'; R: 5' GATGCACAGAAAGCAGTCCA 3'] and a 494bp fragment spanning exon 10 [F: 5' TTGTATTGCTGAGGAAAGATGG 3'; R: 5' TGTCTATACATCAGTGCTGCTTAACTT 3'] of the *IDH3A* mRNA NCBI Reference Sequence, NM_005530.2

PCR was performed under the standard conditions defined in Chapter 4.2. The cycling conditions were as follows: 94°C – 3 minutes, 30 cycles of {94°C – 30 seconds, 60°C – 30 seconds, 72°C – 40 seconds} , 72°C – 7 minutes, on a Multigene thermocycler (Labnet International Inc., NJ, USA). A volume of 5µL PCR products was visualised using 2% weight/volume (w/v) agarose gels as described in Chapter 4.2.

PCR products were subjected to enzymatic SAP/Exo purification and cycle sequencing was performed in a 20µL reaction containing 20pmols forward primer, as described (Chapter 4.2). Sequences were aligned to the reference sequence (NM_005530.2) using Bioedit Sequence Alignment editor v.7.1.3.0[172].

5.1.1.3 Longitudinal clinical evaluation

The clinical records of the family of interest were evaluated, in order to describe the phenotype associated with a putative novel gene. The proband from the family had initially been evaluated and diagnosed in 2006. Both the proband and her affected sister were subsequently re-examined in 2012 by a single ophthalmologist. After molecular diagnosis in 2016, only one of these original patients was ophthalmologically re-examined, as the proband was deceased. This latest examination was performed by the same ophthalmologist who evaluated both siblings in 2012. Fundus photography and OCT were performed.

5.1.2 Results

In total, 2,685 variants were detected in the ERDC gene list (n=26 genes). However, only 362 variants were retained after the removal of all intergenic, intronic and synonymous variants (Appendix 4). Upon further elimination of the *ACBD5*, *IFT140*, *PRPS1* and *ZNF513* genes, 314 variants remained.

The eight families of interest each had, on average, ~135 ERDC variants of which ~65 had a MAF <0.1 and ~16 were exonic or splice variants (Table 5.1.1). Although multiple possible inheritance modes were applied per family, only two families carried variants that co-segregated with disease. These variants were both predicted to be pathogenic, and were rare in the larger WES cohort. Although *SAMD11* has subsequently been confirmed as an IRD gene[69], the *SAMD11* c.122G>A; p.(Arg41Gln) variant (SNP ID rs148711625) in family RPD 1001 was excluded upon manual appraisal due to a high MAF, with 54 G/A genotypes in 661 African individuals in the 1000 Genomes data[112], 132 G/A and four A/A genotypes in 2,180 African Americans in the NHLBI exome data[170], 136 G/A and one A/A genotypes in 1,522 Africans in the ExAC data[171], as well as 32 G/A and two A/A genotypes in 320 Africans in the AGVP data[115].

Table 5.1.1 Summary of variants in the ERDC candidate genes after prioritisation filters. The inheritance models applied for each family are listed in the co-segregation column (*ad*, *ar* and/or *xl*).

Family	Variants in ERDC genes	<0.1 MAF	Exonic /splice	Co-segregating	Pathogenic	Candidate gene (rare in WES cohort)
RPX 54	159	79	19	0 (*ad* or *xl*)	0	0
RPD 55	146	63	13	0 (*ad* or *xl*)	0	0
RPD 94	116	55	11	0 (*ad*)	0	0
RPR 624	131	60	17	0 (*ar* or *xl*)	0	0
RPD 799	135	63	11	0 (*ad*)	0	0
RPD 1001	149	79	19	1 (*ad*); 0 (*xl*)	1	*SAMD11*
RPD 1005	132	67	21	0 (*ar* or *xl*)	0	0
RPM 1167	111	54	15	2 (*ar*)	2	*IDH3A*

Ultimately, a single family (RPM 1167, comprising three individuals) was identified as having putative mutations in an ERDC gene. Two heterozygous *IDH3A* variants were identified in affected sisters of Tsonga (maternal) and Ndebele (paternal) origin; a truncation mutation, c.463G>T; p.(Gly155*) in exon 5, and a missense mutation, c.946C>T; p.(Arg316Cys) in exon 10 (SNP ID rs770798851). Cycle sequencing validation in the siblings and their mother showed the mutations were *in trans* (Figure 5.1.1). No additional familial samples were available for sequencing.

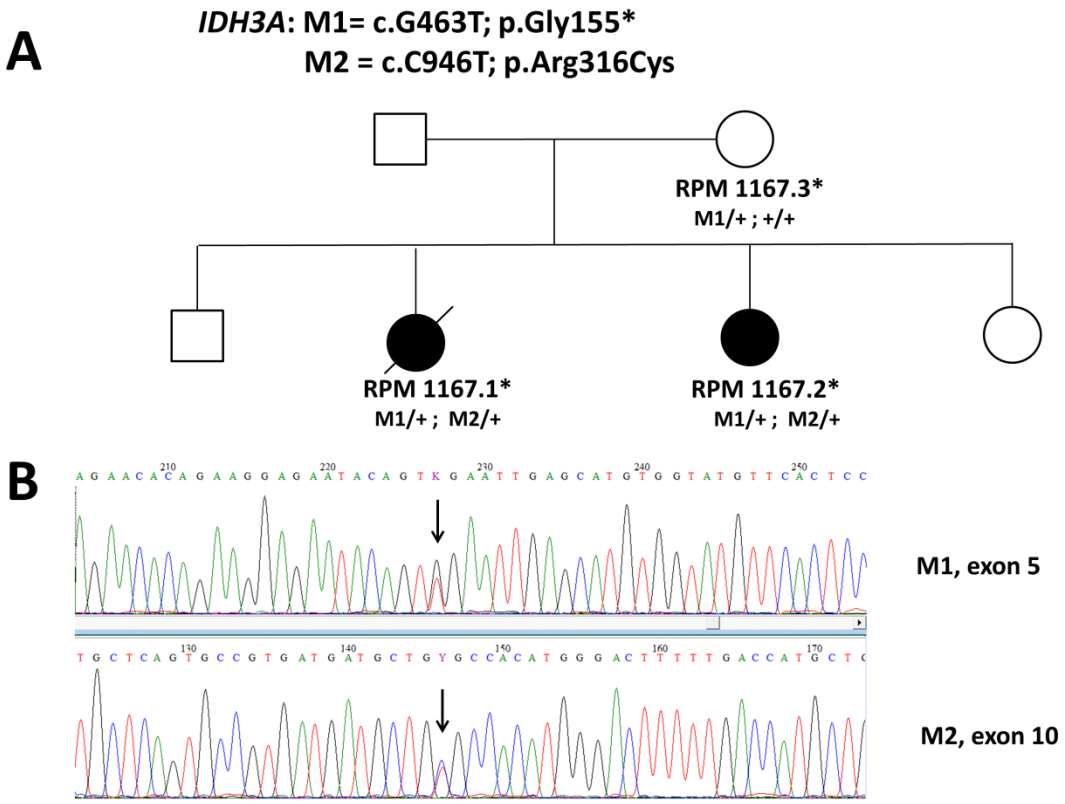


Figure 5.1.1 (A) Pedigree of family RPM 1167, showing co-segregation of two *IDH3A* mutations with disease. Squares represent males, and circles, females. Shaded symbols indicate individuals affected with IRD. Identifier codes show individuals from whom biological material was available, and those selected for whole exome sequencing are noted with an asterisk. Segregation of mutation(s) in the families is indicated as: M/+, heterozygous for mutant allele; +/+, homozygous for wild type allele. **(B) Sequencing electropherograms of *IDH3A* exon 5 and exon 10 confirming the presence of each heterozygous mutation, indicated by an arrow.**

The truncation mutation is presumably novel, as it was not listed in the databases searched (accessed 4 March 2016). This mutation is predicted (using Mutalyzer v.2.0.21[173]) to truncate the 366 amino acid protein by 212 C-terminal amino acids, effectively reducing the product by half (Figure 5.1.2). The missense mutation was predicted to be pathogenic by 12/16 ANNOVAR software tools as well as the PON-P2 online pathogenicity predictor (http://structure.bmc.lu.se/PON-P2/)[198], and is exceedingly rare, present in only one of 121,410 alleles in the ExAC data[171] (the heterozygous allele occurred in the 'European non-Finnish' population group). No further *IDH3A* exonic variants were identified in the WES data of the 16 families in this study.

```
  1  MAGPAWISKV SRLLGAFHNP KQVTRGFTGG VQTVTLIPGD GIGPEISAAV MKIFDAAKAP
 61  IQWEERNVTA IQGPGGKWMI PSEAKESMDK NKMGLKGPLK TPIAAGHPSM NLLLRKTFDL
121  YANVRPCVSI EGYKTPYTDV NIVTIRENTE GEYSGIEHVI VDGVVQSIKL ITEGASKRIA
181  EFAFEYARNN HRSNVTAVHK ANIMRMSDGL FLQKCREVAE SCKDIKFNEM YLDTVCLNMV
241  QDPSQFDVLV MPNLYGDILS DLCAGLIGGL GVTPSGNIGA NGVAIFESVH GTAPDIAGKD
301  MANPTALLLS AVMMLRHMGL FDHAARIEAA CFATIKDGKS LTKDLGGNAK CSDFTEEICR
361  RVKDLD*
```

Figure 5.1.2 Predicted truncation of the IDH3A protein sequence due to the c.463G>T mutation. The sequence highlighted in red is absent when the protein is truncated, resulting in a predicted loss of 212 amino acids.

In 2006, individual RPM 1167.1 (born in 1986) was clinically diagnosed with rapid progression of diffuse RP and macular coloboma (a hole/missing tissue in the macula). In 2007, that individual was referred for molecular testing for a diagnosis of MD. In 2010, a sample from RPM 1167.1 was tested using the *ABCA4* Microarray[91] which at that stage tested for 558 mutations in the *ABCA4* gene, however no mutation was identified. In 2012, individual RPM 1167.2 (born in 1989) was referred with a diagnosis of possible cone-rod dystrophy. Both sisters had an age of onset of ~8 years, and in 2012 their diagnosis was re-evaluated to LCA as fundus examination showed severe mid peripheral bone spicule pigmentary changes, mild arteriolar narrowing, severe macular pigmentation, coloboma-like atrophy of the macula, and optic atrophy. Neither had nystagmus or other ocular associated diseases e.g. strabismus (abnormal eye alignment or “squint”), hypermetropia (long-sightedness), keratoglobus (thinning and protrusion of the cornea) or cataract, nor any systemic associations. Their vision at that stage was limited to counting fingers at 1 meter. Therefore in 2013, a sample from RPM 1167.1 was tested using the LCA Microarray[92], which at that stage tested for 780 mutations in 15 LCA-associated genes, but again no mutation was identified.

In 2016, after WES analysis identified *IDH3A* as the causative gene, RPM 1167.2 was evaluated by the same ophthalmologist that evaluated both sisters in 2012 (RPM 1167.1 was deceased by 2016). Fundus photographs showed peripheral bone-spicule pigmentary changes (Figure 5.1.3 A), peripheral and mid-peripheral pigmentary retinopathy, macular pigmentation with severe coloboma-like atrophy, mild arteriolar narrowing and mild optic atrophy. Fundus autofluorescence (Figure 5.1.3 B) was strongly reduced within the macula, diffusely increasing in a broad ring around the macula, and fading towards the periphery. The OCT (Figure 5.1.3 C) showed macular atrophy and confirmed the coloboma-like atrophy, also showing

lamellar disorganization (lamellar structures were visible but could not be identified). The clinical picture thus represented arRP with a macular coloboma-like atrophy, resembling LCA.

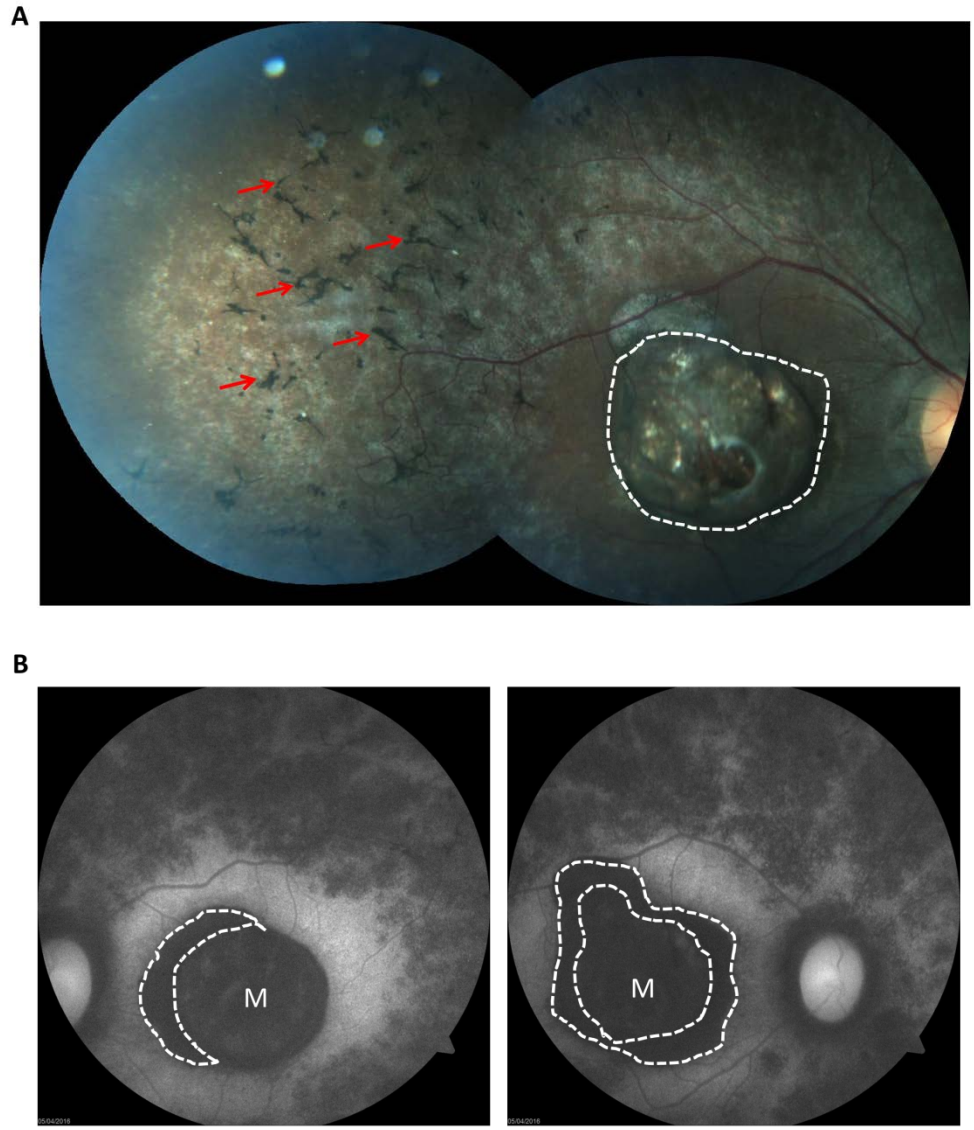


Figure 5.1.3 Clinical findings in individual RPM 1167.2, who carries compound heterozygous mutations in *IDH3A*. *Credit: Dr. G. Fischer.* (A) Fundus photograph showing the peripheral bone-spicules (red arrows) and coloboma-like atrophy (white line); (B) The broad ring of reduced fundus autofluorescence (white lines) within the macula (M); (continued next page)

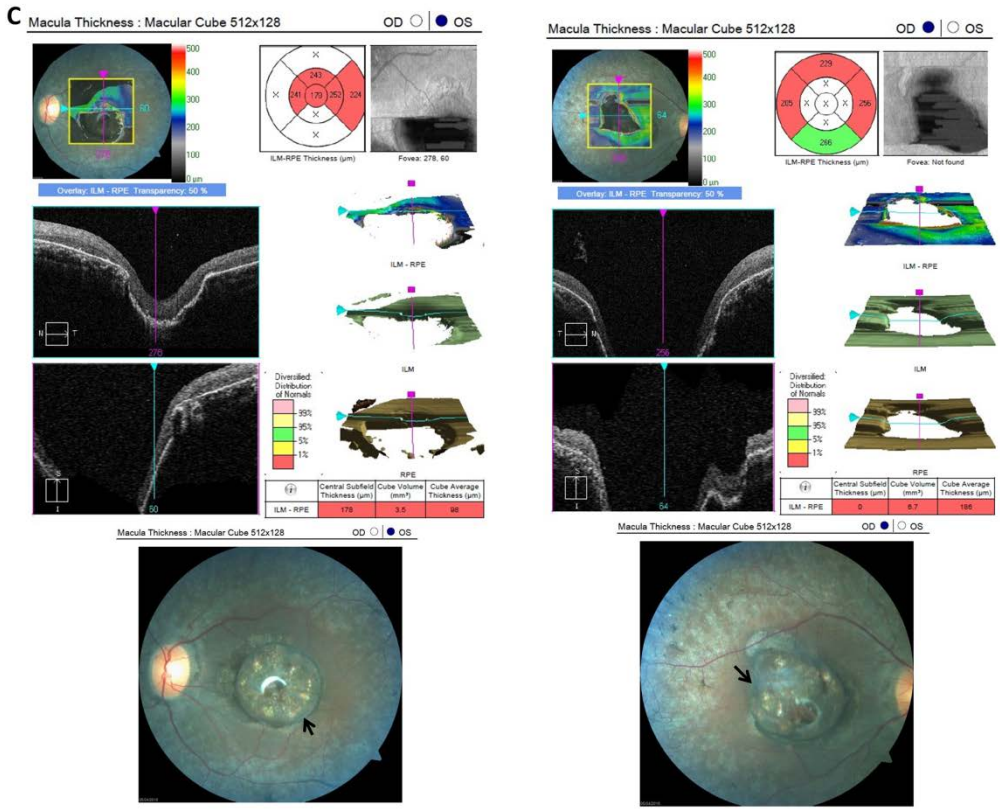


Figure 5.1.3 (continued) (C) optical coherence tomography results, confirming the coloboma-like atrophy (arrows).

5.1.3 Discussion

Having excluded all known IRD genes, eight families were analysed for mutations in 22 putative novel IRD genes proposed by the ERDC. As a result of this approach, one South African family was identified with compound heterozygous mutations in *IDH3A*.

It is likely that the genes responsible for a large proportion of IRDs have already been identified, and therefore any novel causative gene will probably be a rare source of cases[65]. The rare nature of IRDs, the complexity of these disorders (i.e. the vast genetic and allelic heterogeneity), and the fact that 50–70% of cases can be attributed to known genes, further reduces the likelihood that multiple families will be identified with the same novel causative gene. However, the demonstration of several families with mutations attributed to a putative novel gene provides stronger evidence of causality than a single family. Thus, collaborations (like this one set out by the ERDC) are immensely constructive, and the sharing of these data amongst contributors is beneficial for those researchers who participate. The ERDC

candidate gene list is an important asset, the utilisation of which was shown to be an effective approach in this investigation. This particular resource should be highlighted amongst researchers to foster global collaboration for gene discovery.

The *IDH3A* gene was first cloned in humans in 1995[199] and assigned to chromosome 15q25.1–q25.2 in 1996[200]. *IDH3A* is expressed in the retinal transcriptome[10,13], and encodes isocitrate dehydrogenase 3 (NAD(+)) alpha.

Isocitrate dehydrogenases (IDH) catalyse the oxidative decarboxylation of isocitrate to α-ketoglutarate[200] in the tricarboxylic acid cycle (TCA). The TCA, also known as the Citric acid cycle, or Krebs cycle, is a series of reactions in the mitochondrial matrix whereby organic fuel molecules, e.g. glucose, are catabolised in the presence of oxygen to generate energy in the form of ATP.

IDH enzymes belong to two subclasses, distinguished by whether they use the coenzyme nicotinamide adenine dinucleotide (NAD(+)) or nicotinamide adenine dinucleotide phosphate (NADP(+)) as an electron acceptor (NCBI Gene Entry http://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=3419). There are three NAD-IDH enzymes that localise to the mitochondrial matrix and catalyse the rate-limiting step of the TCA. Each isozyme is a tetramer, composed of two α-subunits, one β-subunit, and one γ-subunit[200]. *IDH3A* encodes the α-subunit of one isozyme of NAD-IDH.

Loss of function mutations in the NAD-IDH β-subunit (*IDH3B*) causing two unrelated cases of RP were reported in 2008[201]. Besides the IRD, no phenotype associated with mitochondrial dysfunction (e.g. reduced muscle strength or cardiac problems) was observed. This was despite the fact that the measured NAD-IDH activity in homozygous cell lysates from the patients was less than ~5% of normal control activity. It was proposed that NADP-IDH adequately substitutes for the mutant NAD-IDH in all tissues except the retina, and it was noted that retina is the only tissue where NADP-IDH expression levels do not exceed those of NAD-IDH. The authors suggested that NADP-IDH may be the major enzyme responsible for this step of the TCA in tissues other than the retina, and concluded that NAD-IDH is essential for proper retinal functioning[201].

Visual phototransduction is an energy-demanding process. This, coupled with the photoreceptor OS renewal (described in Chapter 1), results in high metabolic activity in the retina. It was previously thought that the mitochondria housed in the photoreceptor IS supplied the energy required for these processes. However it has subsequently been shown that the eight enzymes of the TCA cycle, including NAD-

IDH, are catalytically active in the OS at comparable levels to those in the retinal mitochondria[202]. Furthermore, the same group previously detected IDH3A in OS discs[203]. Panfoli et al thus suggested that the OS, which lacks mitochondria, possesses functional mitochondrial machinery in the cytosol to meet the high demand for ATP[202].

Mutations in *IDH3A* may impair the metabolic processes, or result in oxidative stress intolerable to the photoreceptor, and thus cause apoptosis and consequent visual loss. Although animal models of *IDH3A* mutations are available, including commercially-available zebrafish (https://zfin.org/ZDB-GENE-040426-1007) and mouse models (https://www.mousephenotype.org/data/genes/MGI:1915084 , http://www.informatics.jax.org/allele/summary?markerId=MGI:1915084), to the best of our knowledge, no associated phenotypes have been reported. Functional investigations of the molecular mechanisms leading to *IDH3A*-related IRD are thus planned with ERDC collaborators. Furthermore, the spectrum of *IDH3A* mutations and phenotypic data associated with seven mutation-positive individuals has been compiled with these colleagues for publication (manuscript accepted 3 March 2017, *Ophthalmology,* the journal of the American Academy of Ophthalmology). It appears that, even with this limited number of cases, some genotype-phenotype correlation exists; all subjects developed symptoms in the first decade of life. Furthermore, patients with missense mutations are diagnosed with typical arRP, while those with truncating mutations generally manifest with a more severe phenotype (including the coloboma-like lesions) although intrafamilial variation in disease expression was noted. The identification of genotype-phenotype correlations is valuable for 'personalised ophthalmology'. An accurate prediction of visual prognosis is important for patients and facilitates their decision-making[204]. Furthermore, the knowledge of genotype-phenotype correlations may be useful for clinicians, and can help inform disease management. Knowledge about disease progression is also beneficial in identifying patients for clinical trials, particularly in terms of their stage of disease and whether therapy will be beneficial, as well as evaluating treatment outcomes[25].

The identification of the compound heterozygous mutations in *IDH3A* has important implications for this local family. The clinical diagnosis of the two sisters in family RPM 1167 has been revised several times. Individual RPM 1167.1 was clinically diagnosed with diffuse RP, MD, and LCA in her lifetime. Similarly, RPM 1167.2 has been diagnosed, at different stages, with possible cone-rod dystrophy and LCA. It is possible that the phenotype in affected individuals may evolve with age, and clinical

follow-up provides an opportunity to observe the onset and evolution of the disorder with time. Nonetheless, this genetic finding now provides an unambiguous diagnosis (arRP) for the family, and permits carrier testing for relatives. Of broader significance, is the identification of a new arRP gene; *IDH3A* can now be added to IRD gene panels or interrogated using other NGS applications, for patients internationally, including in SA. Screening of patient samples is required to ascertain whether arRP is the only IRD associated with this gene. Moreover, the identification of additional *IDH3A* IRD cases will provide information regarding the incidence of mutations in this gene in various population groups, and thus the contribution of this gene to the overall disease burden.

5.2 Elimination of candidates, supplementary screening and pathway analysis

For seven families, mutations in all known IRD genes and novel candidate genes on the ERDC list had been carefully excluded. Analysis was thus extended to evaluate the remaining exome data of these families.

The WES data originally comprised a total of 286,546 exonic, non-synonymous variants as described previously (Appendix 4). Having eliminated the variants in RetNet genes (n=4,866) and ERDC genes (n=314) a total of 281,366 variants remained in the cohort of 56 samples. The seven unattributed families were extracted from this data for further analysis. A multitude of approaches were employed to characterise as many of these families as possible, therefore, in the interests of clarity, the methods are embedded with results in this section.

5.2.1 Exome analysis in seven families

The unattributed families comprised 26 individual samples that had been subjected to WES. The data (minus all genes interrogated in previous analyses, Chapters 4.2–5.1) were separated into the seven families of interest using Microsoft Excel software. On average 115,148 variants were identified per family (Table 5.2.1). Prioritisation filters were applied to variants in each family separately, using the scripts, criteria and analysis pipeline previously described (Chapter 4.2). Co-segregation analysis was performed using the apparent mode of inheritance and if male-to-male transmission was absent, an *xl* inheritance mode was used in addition. Hence five families were analysed using two different inheritance models. The scorep.pl script for scoring pathogenicity was used with the same baseline threshold of 6, corresponding to a minimum of three pathogenic predictions, in order to be less stringent so as to not exclude variants in error. As before, pathogenic variants were interrogated to remove those occurring frequently and randomly (i.e. without appropriate co-segregation with disease) in the larger cohort of 56 samples. However, an additional filtering step was included in this exome analysis to remove variants in the 2,157 genes on the "gene exclusion list"[205]. These genes (including pseudogenes and paralogues) have been identified by retrospective analysis of NGS data and shown to contain multiple variants per individual. It was presumed

that such hypervariable genes are unlikely to be IRD candidates, and could thus be excluded.

Table 5.2.1 Numbers of exomic candidate variants remaining in seven unresolved families, after each prioritisation filter. The inheritance models applied for each family are listed in the co-segregation column (*ad, ar* and/or *xl*).

Family	Variants	<0.1 MAF	Exonic /splice	Co-segregating	Pathogenic	Candidates {lacking a SNP ID}
RPX 54	126,213	67,329	17,439	895 (*ad*) + 46 (*xl*) = 941	485	5 {0}
RPD 55	116,704	61,878	16,836	495 (*ad*) + 8 (*xl*) = 503	267	30 {9}
RPD 94	108,586	54,132	13,322	568 (*ad*)	221	48 {13}
RPR 624	104,944	53,047	13,305	692 (*ar*) + 89 (*xl*) = 781	357	26 {8}
RPD 799	114,883	59,994	15,674	1,278 (*ad*)	602	61 {26}
RPD 1001	121,250	66,372	18,097	540 (*ad*) + 75 (*xl*) = 615	329	28 {11}
RPD 1005	113,457	59,069	15,904	574 (*ar*) + 11 (*xl*) = 585	252	32 {2}

In total, 230 candidate variants were detected in 190 distinct genes. Manual appraisal using extant knowledge for this large number of variants was impractical. Furthermore, prioritisation of candidates based on gene expression in the retina was unreliable, as different candidate genes were present in various retinome datasets. The 190 distinct genes were compared to published datasets using the Microsoft Excel software conditional formatting tool to highlight duplicates. It was established that 170 of the 190 candidate genes (89%) were listed in the Whitmore et al data[13], generated by sequencing RNA from temporal, macular and nasal regions of the retina. In contrast, only 63 and 73 of the candidate genes (33–38%) were listed in the 13K and 15K retinome gene lists[10], respectively. These 13K and 15K retinal/RPE gene lists were compiled by meta-analysis of retinal datasets. Retinal gene expression could therefore not be used to eliminate candidates with confidence.

In order to ascertain whether mutations in the same gene could be causative in more than one family, the Microsoft Excel software conditional formatting tool was used to highlight candidate gene duplicates between families. Four candidate genes were found to occur in two families each, however in each case the specific variant

was different (Table 5.2.2). Each variant was evaluated for allele frequency in ExAC Browser[171] (accessed 16 June 2016).

Table 5.2.2 Summary of the four candidate genes from the residual exome data, with variants in multiple families. The inheritance models applied, that resulted in selection of these genes, are listed in the gene column (*ad, ar* and/or *xl*).

Gene	Family	Variant	SNP ID	ExAC allele count/total allele number
ATPAF2 (*ad*)	RPX 54	c.722A>G; p.(Glu241Gly)	rs34607655	98/105,310
ATPAF2 (*ad*)	RPD 55	c.402_404del; p.(Arg134del)	-	-
NIN (*ad*)	RPD 55	c.673G>A; p.(Glu225Lys)	rs148473555	17/120,634
NIN (*ad*)	RPD 799	c.3083_3085del; p.(Glu1028del)	rs369382014	-
DNAH17 (*ad*)	RPD 94	c.13033T>C; p.(Trp4345Arg)	rs140378962	123/121,294
DNAH17 (*ad*)	RPD 799	c.7901C>T; p.(Ala2634Val)	rs143375889	102/113,760
'NBPF8, NBPF9' (*ar*)	RPR 624	5 variants of unknown effect (not annotated)	-	-
'NBPF8, NBPF9' (*ad*)	RPD 799	1 variant of unknown effect (not annotated)	-	-

For each variant listed on ExAC, the majority of alleles were reported in the African population group. The allele counts of each of the *ATPAF2, NIN* and *DNAH17* variants given in the ExAC browser exceeded the recommended frequency centred threshold of 1/100,000 for an *ad* IRD gene[78]. In addition, the annotations for the gene '*NBPF8,NBPF9*' were incomplete, and the chromosomal coordinates spanning these variants in families RPR 624 and RPD 799 revealed a non-coding gene in the region in Ensembl (http://www.ensembl.org/)[169]. Thus, no coding candidate gene was found to be causative in multiple families i.e. each family could have mutations in a unique IRD gene.

In the absence of alternative strategies, the family with the least candidate variants was interrogated further. Four candidate variants (identified using an *ad* inheritance model) in family RPX 54 were evaluated and ultimately eliminated based on allele frequency data in the ExAC data[171] (accessed 9 June 2016). These four variants were *ATPAF2* rs34607655 (listed in Table 5.2.2 above), *MATN2* rs34354598

(326/120,044 alleles, including five homozygous Africans), *C1orf131* rs34759016 (189/121,396 alleles, including one homozygous African), and *PDHX* rs61752925 (236/120,694 alleles, including two homozygous Africans). A fifth variant was confirmed by cycle sequencing to be an indel in *LGALS13*, c.230_240delinsATTTACACTAT; p.(Thr77_Asp79delinsAsnLeuHis). PCR and cycle sequencing were performed using primers which amplified a 390bp fragment spanning exon 3 of *LGALS13* [F: 5' GTGTGTGTCTGCGCAAGG 3'; R: 5' CCAGGGGCAGGAGTAGTTAT 3'], as described in Chapter 4.2, with the exception that the PCR annealing temperature was 60°C and 20pmols forward primer were used for the sequencing reaction. The *LGALS13* variant was predicted (using Mutalyzer v.2.0.21[173]) to affect three consecutive amino acids; Thr77, Thr78 and Asp79 (Figure 5.2.1). However, the amino acid variants namely p.(Thr77Asn); p.(Thr78Leu) and p.(Asp79His) have been previously reported to occur in a mutation hotspot in a South African study of preterm labour and pre-eclampsia (with no record of retinal disease), and were identified in 30 individuals[206]. This variant was therefore deemed unlikely to cause IRD.

```
  1 MSSLPVPYKL PVSLSVGSCV IIKGTPIHSF INDPQLQVDF YTDMDEDSDI AFRFRVHFGN
 61 HVVMNRREFG IWMLEETTDY VPFEDGKQFE LCIYVHYNEY EIKVNGIRIY GFVHRIPPSF
121 VKMVQVSRDI SLTSVCVCN*

------------------------------------------------------------------

  1 MSSLPVPYKL PVSLSVGSCV IIKGTPIHSF INDPQLQVDF YTDMDEDSDI AFRFRVHFGN
 61 HVVMNRREFG IWMLEENLHY VPFEDGKQFE LCIYVHYNEY EIKVNGIRIY GFVHRIPPSF
121 VKMVQVSRDI SLTSVCVCN*
```

Figure 5.2.1 The predicted alteration of the LGALS13 protein sequence due to the c.230_240delinsATTTACACTAT variant. The wild type sequence (top) highlighted in red is altered by the variant in family RPX 54 (bottom).

Thus, family RPX 54 concluded with no candidate variants at the end of the analysis pipeline, raising concern that the causative variant had been incorrectly eliminated, or not captured. All five eliminated variants in family RPX 54 (above) had been filtered using an *ad* inheritance model, highlighting the possibility that an *xl* gene could be associated with IRD in this family. However, no candidates passed the prioritisation filters when an *xl* inheritance model was used. This prompted consideration that a variant in the *xl* mutation hotspot, *RPGR* ORF15, had been overlooked.

5.2.2 Supplementary analysis of *RPGR* ORF15

Two challenges were faced in the analysis of residual WES data towards identifying novel genes involved in IRDs in our cohort. Firstly, one family had no candidates at the end of the analysis pipeline. Secondly, large numbers of candidates were present in the six remaining families. It was essential that mutations in the known IRD genes be excluded, prior to interrogating multiple variants to implicate (and prove) novel IRD candidates. Given the lack of male-to-male disease transmission in 5/7 unresolved families, the primary xlRP gene, *RPGR*, required further consideration.

The *RPGR* gene was identified as causing xlRP in 1996[207,208]. Four years later, an RPGR transcript with a novel 3' terminal exon, known as exon 'open reading frame 15' (ORF15), was characterised[209]. ORF15 was reported as a mutational hotspot, accounting for the majority of xlRP cases studied. Multiple isoforms of RPGR exist, and the isoform containing ORF15 is the predominant transcript expressed in the retina[209]. ORF15 includes exon 15 and a portion of intron 15, and is comprised of a highly repetitive, low complexity, purine rich sequence. The nature of this sequence has been postulated to be responsible for the region's high mutability, possibly by adopting unusual structural conformations, thereby reducing the replication fidelity[209]. Additionally, this genomic sequence generates technical challenges in molecular investigations.

It has recently been shown that *RPGR* ORF15 is difficult to capture using NGS approaches due to the highly repetitive sequence therein, exceeding 1Kb in length[210]. Huang et al showed that ORF15 was insufficiently captured during NGS, causing false negative results. They therefore proposed that a complementary approach be taken in the molecular diagnosis of IRDs, with additional sequencing of targets like *RPGR* ORF15 which evade capture-based technologies[210].

It was evident that intron 15 (which forms part of ORF15) was not captured sufficiently in our WES data, when the .bam files of affected males from unresolved families were viewed using Integrative Genomics Viewer (https://www.broadinstitute.org/igv/) (Figure 5.2.2). Subsequently, data from affected males of each of the five unresolved potential *xl* families were interrogated to ascertain whether this genomic region was shared between relatives. Affected males within a family, sharing the same X-chromosomal region around this gene,

could potentially represent a family harbouring an *RPGR* ORF15 mutation that was not captured or covered by WES.

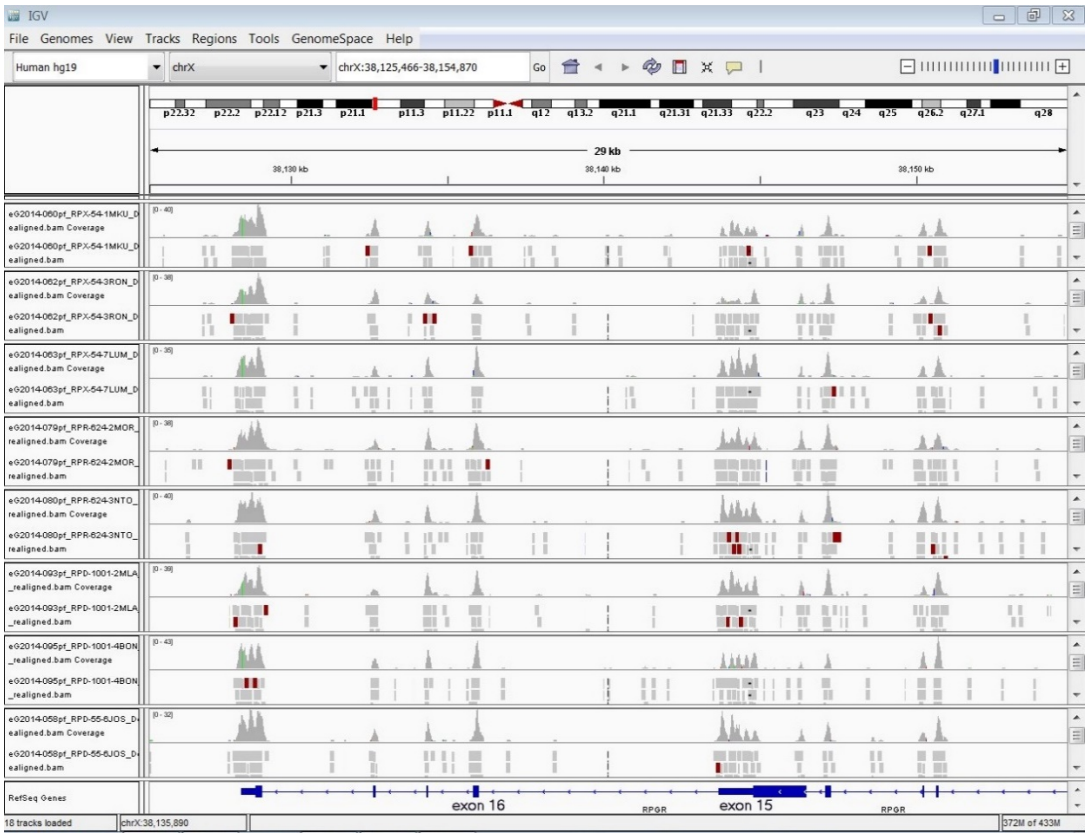


Figure 5.2.2 A screenshot of the Integrative Genome Viewer profiles of .bam files in this study, showing the coverage across the *RPGR* gene. Grey peaks and corresponding blocks beneath them indicate covered regions. *RPGR* ORF15 includes exon 15 (blue, bottom panel) and a portion of intron 15 which has not been captured.

All variants on the X chromosome were extracted from the WES data (including intergenic, intronic, synonymous, upstream and downstream, ncRNA and UTR variants), using the following command on a PC running the UBUNTU LINUX operating system:

> grep "^chrX" DATA.csv > chrX_DATA.csv

A total of 39,480 X chromosome variants were identified in the total cohort. The 129,158 bp region spanning the *RPGR* gene was then extracted, from chromosomal co-ordinates 38,080,739 to 38,209,897, and encompassed 64 variants. Affected males from the families of interest were manually inspected to determine whether the same 'haplotype' occurred in relatives, as determined by the presence of identical variants spanning a portion of the region. A single affected male from family RPD 55 had been subjected to WES, therefore no intra-familial comparison could be performed. The results of the five unresolved potential *xl* families are presented in Table 5.2.3.

Table 5.2.3 Results of analysis of the genomic region encompassing *RPGR* in unresolved families with possible X-linked disease inheritance.

Family	Number of affected males subjected to WES	Number of consecutive variants shared by related males	Size of the genomic region shared by related males (bp)
RPX 54	3	29	41,320
RPD 55	1	N/A	N/A
RPR 624	2	20	28,464
RPD 1001	2	15	16,211
RPD 1005	2	None	None

Only one family, RPD 1005, did not exhibit a shared genomic region around *RPGR*. The affected males in this family differed at 15 positions between chromosomal co-ordinates 38,128,811 and 38,156,677. The family was therefore unlikely to carry an ORF15 mutation. However, the haplotype analysis indicated that ORF15 mutations could indeed be present in three unresolved families. Interestingly, the two affected males in family RPD 1001 had a stretch of variants in common with three affected males in family RPX 54, indicating this genomic region could be identical by descent. The final family had a single affected male subjected to WES therefore the possibility of an ORF15 mutation could not be excluded.

The ORF15 region was therefore sequenced in an affected proband from each of four families (RPX 54, RPD 55, RPR 624 and RPD 1001). Due to the technical challenges of conventionally (i.e. Sanger) sequencing the region, this supplementary analysis was performed commercially (Asper Ophthalmics). Two mutations were identified in three families and no mutation was identified in the individual from family RPD 55. As suggested by the haplotype analysis, the

individuals from families RPX 54 and RPD 1001 each carried an identical mutation, namely c.2790_2791delGG; p.(Glu931Glyfs*147). The individual from family RPR 624 carried a different mutation, namely c.2964_2965delGG; p.(Glu989Glyfs*89). The individual selected from family RPX 54 for ORF15 sequencing had previously been tested using the Asper xlRP microarray. In 2010, the xlRP microarray screened for 182 specific mutations in *RP2* and *RPGR*, excluding the ORF15 region. The c.2790_2791delGG mutation was therefore not detected at that time.

Both mutations c.2790_2791delGG and c.2964_2965delGG were listed in LOVD (www.lovd.nl/)[168], with variant numbers RPGR_00175 and RPGR_00314 respectively, and each mutation had been previously associated with xlRP[211,212]. No population MAF data was presented for either mutation in Ensembl release 8 (Ensembl transcript ENST00000378505.6, http://www.ensembl.org/)[169] or the ExAC Browser[171] (both accessed 31 October 2016). Using the corresponding mRNA NCBI Reference Sequence NM_001034853.1 in Mutalyzer v.2.0.21[173], it was predicted that both deletions truncate the protein by 76 amino acids, although the frameshift length differs between them (Figure 5.2.3).

The analysis of *RPGR* ORF15 to supplement the WES data thus resulted in identification of the probable causal mutation in three additional families (Figure 5.2.4 A–C, presented on the three successive pages after Figure 5.2.3), thus leaving only four families unresolved.

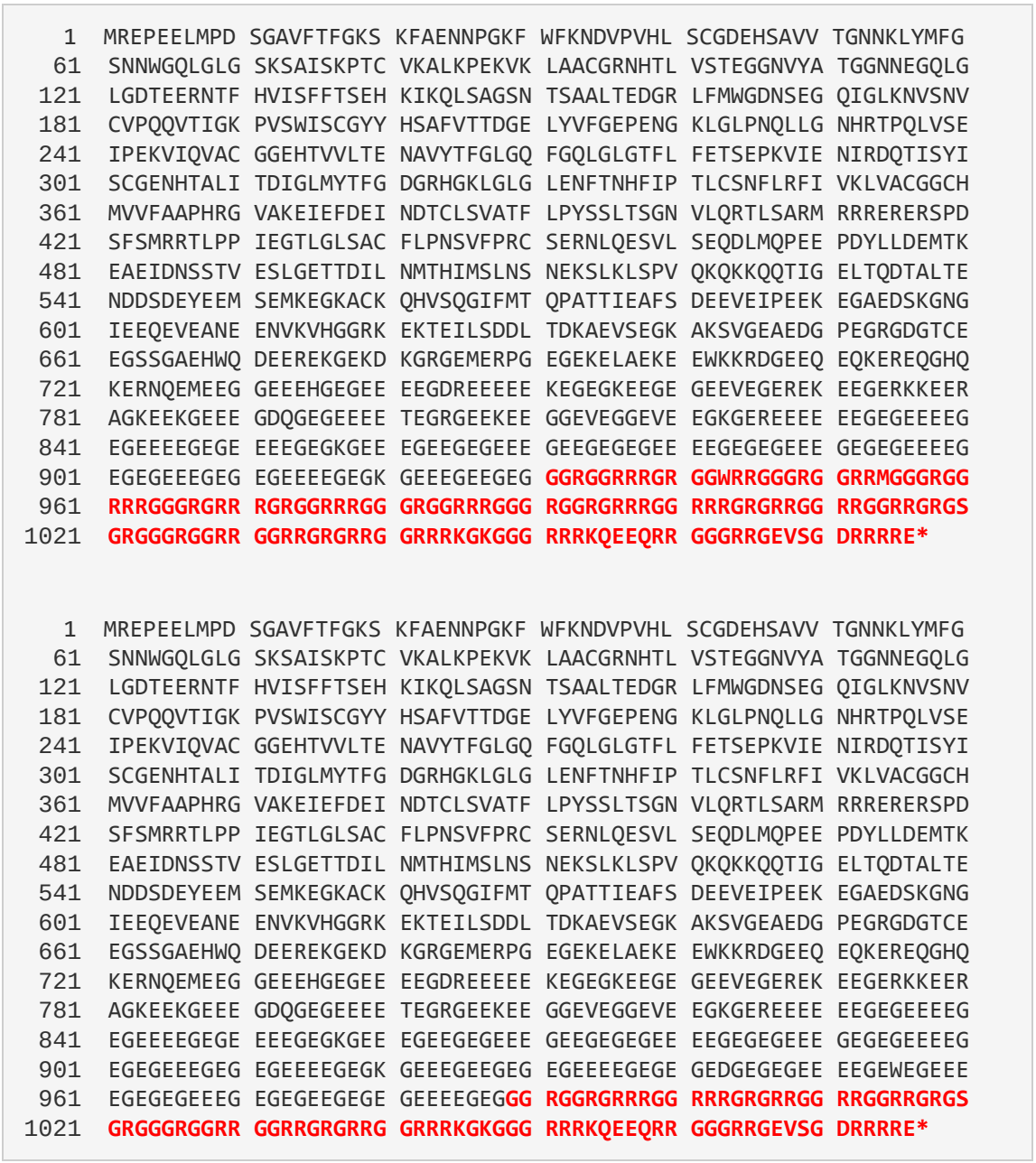
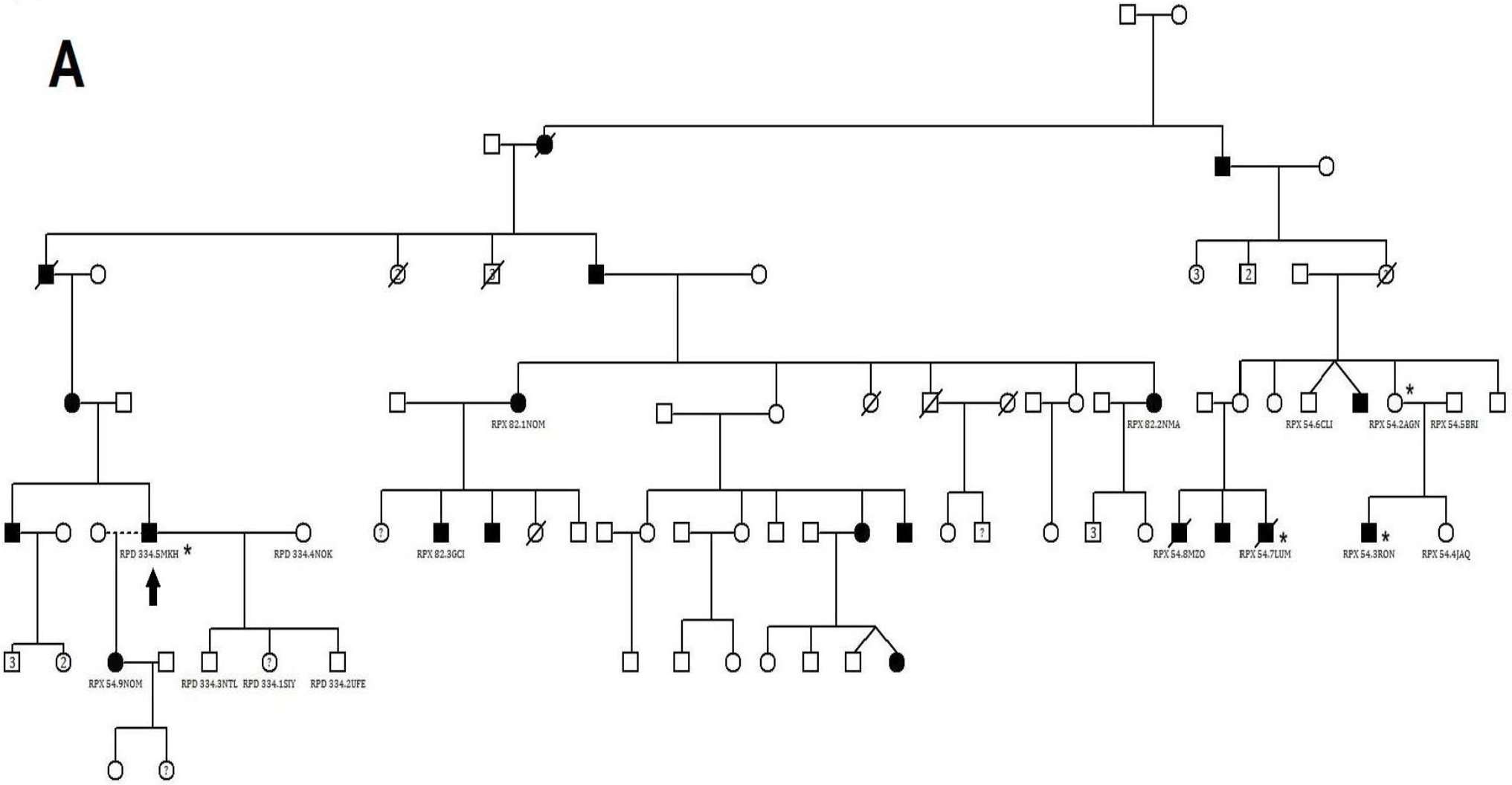
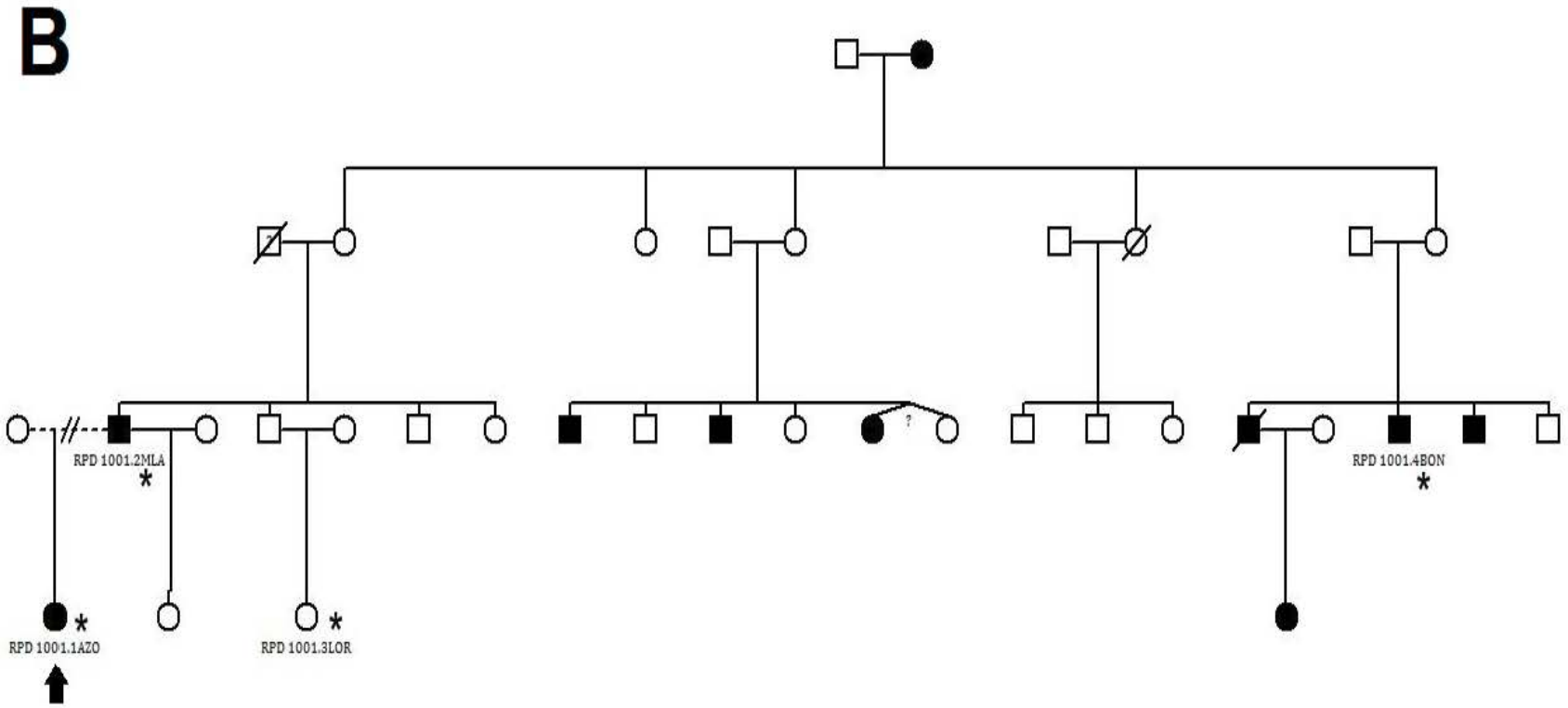


Figure 5.2.3 Mutalyzer protein predictions indicating the different amino acid frameshift lengths, but identical stop codon positions, generated by the c.2790_2791delGG; p.(Glu931Glyfs*147) (top) and c.2964_2965delGG; p.(Glu989Glyfs*89) (bottom) mutations in RPGR ORF15. The amino acid frameshifts are indicated in red, and premature stop codons are represented by an asterisk.


A
RPX 82.1NOM
RPX 82.2NMA
RPX 54.6CLI
RPX 54.2AGN
RPX 54.5BRI
RPD 334.5MKH
RPD 334.4NOK
RPX 82.3GCI
RPX 54.8MZO
RPX 54.7LUM
RPX 54.3RON
RPX 54.4JAQ
RPX 54.9NOM
RPD 334.3NTL
RPD 334.1SIY
RPD 334.2UFE


B
RPD 1001.2MLA
RPD 1001.4BON
RPD 1001.1AZO
RPD 1001.3LOR

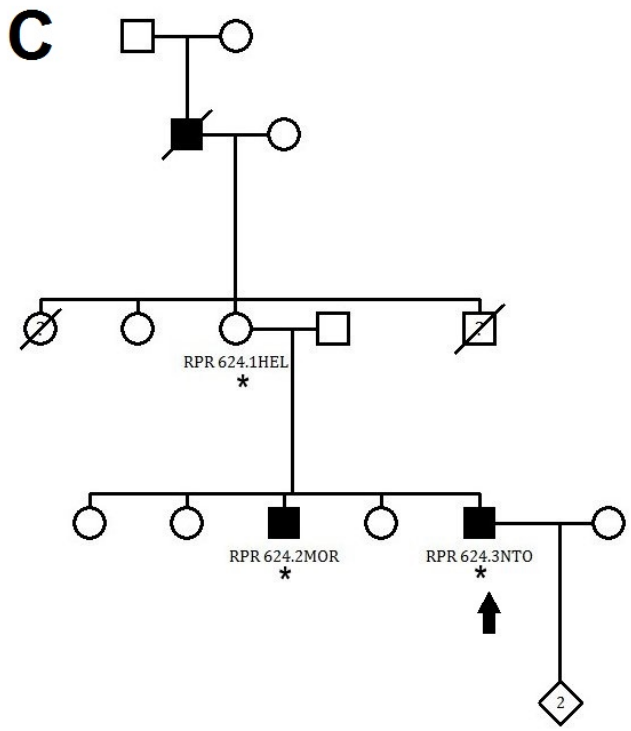


Figure 5.2.4 Pedigrees of three families with *RPGR* ORF15 mutations. Squares represent males, and circles, females. Shaded symbols indicate individuals affected with IRD, and a question mark highlights subjects for whom the clinical status is unknown. Identifier codes show individuals from whom biological material is available, and those selected for WES are noted with an asterisk. The individuals indicated with an arrow were subjected to *RPGR* ORF15 sequencing; these persons in families RPD 54 (A) and RPD 1001 (B) each carry the c.2790_2791delGG mutation, whilst the individual in family RPR 624 (C) carries the c.2964_2965delGG mutation.

Family RPD 54 is large, and comprises what were originally deposited as three Xhosa nuclear families in the registry, and subsequently found to be related (as indicated in Table 2.3). Although RPD 1001 is also a Xhosa family, it appears unrelated to RPD 54, which is supported by the fact that these two families identify with different tribes/clans within the Xhosa ethnolinguistic group (self-reported by family members). The ethnolinguistic group of family RPR 624 was unknown. Clinical information was sparse for members of the three families. Within family RPD 54, affected males RPD 334.5MKH, RPX 54.3RON, and RPX 54.7LUM had a reported age of onset of three years, two years, and eight years, respectively, whilst RPX 54.8MZO was simply diagnosed with diffuse RP. Within family RPD 1001, RPD 1001.2MLA also had diffuse RP, and RPD 1001.4BON had an age of onset of 14 years. In the third family, RPR 624.3NTO was reported to have had diffuse RP from six years of age. He experienced a rapid deterioration of his vision until the age of 18 years, after which the disease was stable. By the age of 30 years his visual acuity was 20/200, bilaterally. Clinical detail was not available for any affected females in the pedigrees.

5.2.3 Functional enrichment and pathway analysis of 144 candidate genes

Families RPD 55, RPD 94, RPD 799 and RPD 1005 remained unresolved (Appendix 5), and hence were enriched for novel gene discovery, with a total of 171 candidate variants in 144 unique genes (Table 5.2.1). Of the 171 variants, 50 had no SNP ID and could perhaps be analysed in the first instance, however SNP IDs have been assigned to pathogenic mutations in population databases[24,174,213] hence this approach could result in rejection of a true candidate.

The 144 genes were compared to retinome datasets using the Microsoft Excel software conditional formatting tool to highlight duplicates. It was established that 128 genes are expressed in the retina, according to the Whitmore et al data[13] whereas only 51 and 58 genes are listed in the 13K and 15K retinome gene lists[10], respectively. A core set of 47 genes are listed in all three datasets, however restricting analysis to these genes could also have resulted in the exclusion of a true candidate.

The most inclusive strategy for identification of novel IRD genes would be evaluation of each of the 171 candidate variants, however this remains a daunting

and inefficient approach. Therefore, in order to guide the future research, functional enrichment analysis was performed for all 144 genes using the WEB-based GEne SeT AnaLysis Toolkit (WebGestalt[214,215] http://www.webgestalt.org/, accessed 9 November 2016). WebGestalt organises and visualises gene sets by user-selected attributes, and reports categories within the input list that are significantly enriched compared to a reference gene set[214]. The organism of interest and input gene identifier types were specified as 'Homo sapiens' and 'gene symbols', respectively. The genes not initially recognised by the software were replaced with gene aliases until ultimately all but one (*GOLGA8M*) could be unambiguously mapped to Entrez genes. The reference gene set selected was the Homo sapiens genome (i.e. the entire Entrez gene list), and the remaining default selections were used, namely hypergeometric statistical method and Benjamini & Hochberg (BH) multiple test adjustment[214]. In order to provide a functional overview of the 143 input genes, the 10 most enriched categories were chosen as output, rather than specifying a significance level[215]. The minimum number of genes per category was set to 1. The attributes selected for enrichment analysis were: disease, phenotype, pathways and gene ontology.

No diseases pertinent to this study were enriched in the 143 genes. The top 10 enriched disease categories included Brenner tumour of ovary, carcinoma, breast diseases, breast neoplasms, skin diseases, Alzheimers disease, and very-long-chain acyl-coenzyme A dehydrogenase deficiency. Similarly, some of the enriched phenotypes were seemingly unrelated to this investigation, e.g. female hypogonadism and elevated alpha-fetoprotein. However some enriched phenotypes may be relevant for candidate prioritisation, namely (i) morphological abnormalities of the central nervous system, such as abnormal cortical gyration (n=4 genes) and microcephaly (n=9 genes); and (ii) abnormalities of cell physiology (n=5 genes) including carnitine metabolism and fatty-acid anion metabolism.

For pathway enrichment analysis, three collections were selected within WebGestalt, namely KEGG pathways, Pathway Commons and WikiPathways. The top 10 enriched categories for each collection are presented in Table 5.2.4.

Table 5.2.4 Summary of results of the WebGestalt pathway analysis, using the KEGG, Pathway Commons and WikiPathways collections. The top 10 enriched categories of genes are listed, together with the number of genes in each category.

KEGG pathways	Pathway Commons pathways	Wikipathways
Focal adhesion (n=5)	IL5-mediated signalling events (n=13)	MicroRNAs in cardiomyocyte hypertrophy (n=3)
Hypertrophic cardiomyopathy (n=3)	LKB1 signalling events (n=13)	Focal adhesion (n=4)
Dilated cardiomyopathy (n=3)	EGFR-dependent endothelin signalling events (n=13)	Mitochondrial LC-fatty acid beta-oxidation (n=2)
p53 signalling pathway (n=2)	mTOR signalling pathway (n=13)	Monoamine transport (n=2)
Arrhythmogenic right ventricular cardiomyopathy (n=2)	IL3-mediated signalling events (n=13)	Ovarian infertility genes (n=2)
Adherens junction (n=2)	Glypican pathway (n=14)	Synaptic vesicle pathway (n=2)
Gastric acid secretion (n=2)	Arf6 signalling events (n=13)	Thyroxine (thyroid hormone) production (n=1)
Tyrosine metabolism (n=2)	PDGFR-beta signalling pathway (n=13)	ErbB signalling pathway (n=2)
Fatty acid metabolism (n=2)	GMCSF-mediated signalling events (n=13)	Fatty acid beta oxidation (n=2)
Melanoma (n=2)	Signalling events mediated by hepatocyte growth factor receptor (c-Met) (n=13)	miRNAs involved in DNA damage response (n=1)

None of the enriched pathways are connected to IRDs in an obvious way, however the 13 identical genes within the various signalling categories (Pathway Commons) may be of interest. Again, interpretation of these data is complicated by the heterogeneity of known IRD genes. Lastly, Gene Ontology (GO) enrichment analysis was performed in WebGestalt, yielding enriched categories of biological processes, molecular functions, and cellular components. The most applicable biological processes were: calcium-dependent cell-cell adhesion (n=3 genes), localization of cell (15 genes), centrosome-templated microtubule nucleation (n=1 gene) and positive regulation of small GTPase mediated signal transduction (n=3 genes). The most relevant molecular functions were: voltage-gated cation channel activity (n=4 genes) and protein tyrosine kinase activator activity (n=2 genes). The most pertinent enriched cellular components were the axoneme (n=4 genes), and cell projection (n=20 genes). Since it is not known whether the putative four novel IRD genes are in any way related to each other, (and therefore categorised

together) these WebGestalt results should be interpreted with caution, and should not be used to eliminate candidates. Instead, enriched categories should complement future analysis, by providing additional information which may aid in the prioritisation of candidate genes.

Finally, in order to ascertain whether any of the candidates interact directly with known IRD genes, RPGeNet v1.0[216] (https://compgen.bio.ub.edu/RPGeNet/, accessed 11 November 2016) was used. RPGeNet is a searchable, web-based, RP-specific network, which integrates the gene and protein interaction data of 110 RP/LCA genes. This 'interactome database' was searched for the most direct interactions of the 144 candidate genes, i.e. the distance to other nodes was set to 1, and the interaction level to 0. These skeleton network settings were required for searching such a large number of candidate genes. Since ubiquitously expressed genes are implicated in IRD, the 'absolute' gene expression option (i.e level of expression in the retina) was selected, rather than the 'relative' option (i.e expression fold-change in retina versus the average expression in all tissues) for colour coding of results. When 144 genes were input simultaneously, only two direct interactions were identified: the candidate *ATM* interacts with the syndromic gene *PANK2*, and the candidate *MPDZ* interacts with *CRB1* (Figure 5.2.5).

In order to confirm the results, the 144 genes were input as smaller batches of 10. The same results were obtained (i.e. *ATM* and *MPDZ* were the only candidates interacting with known IRD genes) however six batches (i.e 60 genes) gave an error output. Furthermore, the same error was obtained for single genes in those batches, indicating that the candidate gene input names were not recognised, or the genes have no interactions at the specified interaction level. It should be noted that the 110 RP/LCA gene "seeds" in RPGeNet do not represent a comprehensive list of all IRD genes. Furthermore, the application does not permit easy export of the results. Nevertheless, this is a useful tool for candidate prioritisation in the field of IRD research, and will hopefully be developed further.

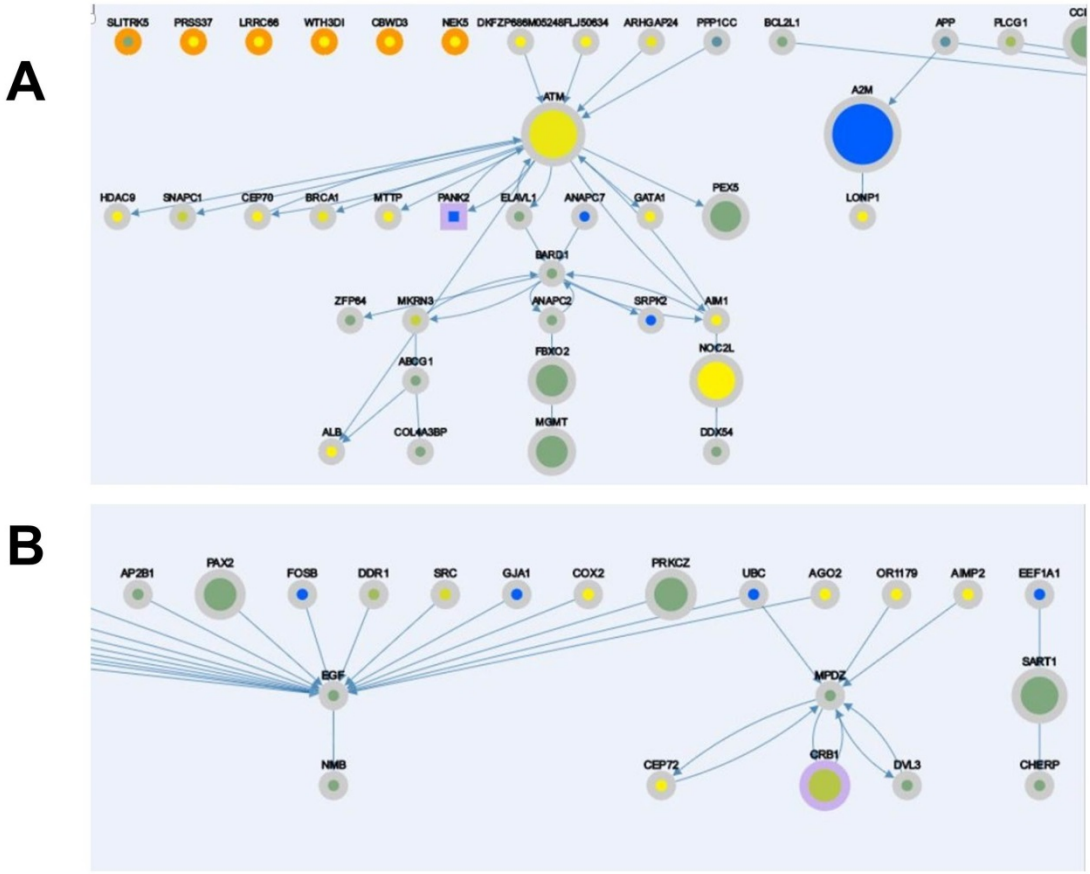


Figure 5.2.5 Screenshots of the RPGeNet analysis output, showing the interaction of two candidates, *ATM* (A) and *MPDZ* (B) with known IRD genes. The results are portrayed as follows: gene expression is represented by the colour of the inner circle, with black representing null expression, and shades of yellow to green to blue as retinal expression level increases. RP/LCA "seed" genes e.g. CRB1 and PANK2, are highlighted by a violet border (squares indicate syndromic genes); orange borders represent genes with no interaction in RPGeNet; a gray border shows the gene interactions at level 0 (skeleton network).

5.2.4 Discussion

At the end of this investigation, three families were identified with mutations in *RPGR*, and four (out of the original 16) families remain without a genetic diagnosis.

Family RPX 54 concluded with no candidate variants after the analysis pipeline, highlighting a limitation in the WES approach. *RPGR*, a major IRD gene, is not captured adequately by this methodology due to the nature of its sequence (particularly in the ORF15 region). The mutation hotpot, ORF15, was found to be omitted in our data. This originally gave rise to a false-negative result for mutations in *RPGR*, corresponding to recently published findings[210].

Supplementary sequencing of this region identified two *RPGR* ORF15 mutations in three indigenous African families, an important finding in this study. Although the WES cohort was limited, and one family should likely be discounted due to a lack of clear IRD (RPD 401), this nevertheless indicates that ORF15 mutations account for as much as 3/15 (20%) of the indigenous African IRD cohort. Furthermore, when the *RP2* mutation is included, 4/15 (~27%) of the families subjected to WES are, in fact, attributed to mutations in the xlRP genes. This relatively large fraction could be artifactually skewed, as technical challenges have largely prevented analysis of the ORF15 region, despite previous efforts in the Division of Human Genetics at UCT. However, the identification of the same c.2790_2791delGG; p.(Glu931Glyfs*147) mutation, on a shared haplotype, in two unrelated Xhosa families (albeit from different tribes/clans) implies that a founder effect may exist in SA, which could contribute to the burden of IRD disease.

Both of the identified ORF15 mutations can be classified as pathogenic according to recently published guidelines[174], as they: (i) are predicted null variants in a gene in which loss of function is a known mechanism of disease, (ii) probably result in a reduction of protein length, (iii) have been previously reported as pathogenic, (iv) occur in a known mutation hotspot, and (v) are absent from control databases. Nevertheless, variant co-segregation with disease should be confirmed in all available samples, in order to properly convert these three families into diagnostic mode in the translational programme. It is therefore recommended, for both research and diagnostic purposes, that an effort is made to overcome the technical challenges and perform *RPGR* ORF15 screening in local southern African populations.

Understanding the challenges and limitations of NGS is essential for informed analysis and interpretation of research results, and drawing accurate conclusions from the findings[217]. Awareness about which genes have been inadequately interrogated is important, particularly when negative results are obtained. In this particular instance, there was an initial misconception that *RPGR* could be excluded in the cohort, as no mutations were identified in the coding regions of this gene. The observation that ORF15 was insufficiently covered, was a crucial lesson in this study.

The targeted capture NGS diagnostic reports issued by the Manchester Centre for Genomic Medicine (received through the UCT IRD translational research programme) specify that ORF15 has not been included in that analysis. Indeed, ORF15 is explicitly excluded in many publications reporting the NGS findings for

IRDs[39,42,50,60]. However, an approach has recently been put forward which permits NGS-based analysis of this region[218]. It was suggested that high-fidelity PCR of a 2.1Kb region encompassing ORF15, followed by fragmentation and indexing of the PCR products and subsequent NGS, greatly improves the coverage of this region. The authors reported that 31% of cases which previously tested negative through various approaches (including targeted capture NGS and WES), were found with pathogenic ORF15 mutations using this technique[218]. It is therefore anticipated that the high-fidelity PCR step can be employed, and ORF15 thereby integrated into the design of IRD targeted capture NGS panels. However, should resource limitations prevent the implementation of IRD capture panels in SA, traditional sequencing technologies should be considered. Sanger sequencing of the ORF15 region requires particular PCR reagents and multiple reverse primers (personal communication with Dr SJ Bowne, The University of Texas HSC at Houston).

Ultimately, mutations were identified in all but four families in the WES study cohort. For these four families, there remains the possibility that their causative gene is novel. However, the vast number of WES variants in each family, due to the genetic diversity of the indigenous African populations, resulted in a "bottleneck" in the identification of novel IRD candidate genes. This was further exacerbated by the marked complexity of the known molecular causes of IRDs, which limited the ability to efficiently exclude genes based on their tissue expression or biological function. It would appear that the most effective strategy moving forward would be evaluation of each of the candidate genes, in each family. To this end, prioritisation software such as ToppGene[219] or ENDEAVOUR[220] could be used to rank the genes containing the 30 candidate variants in family RPD 55, 48 variants in RPD 94, 61 variants in RPD 799 and 32 variants in RPD 1005, in turn. These tools require the user to submit a training set of genes (for example, all known RP genes or IRD genes), together with the list of candidate genes to be prioritised. The candidates are ultimately ranked according to their similarity to the training genes, based on specific functional annotations/attributes which are selected by the user. In addition, supporting evidence may be gained by classifying all the candidate variants of this study, based on their pathogenicity scores. Finally, these gene- and variant- level data should be interrogated in combination with the information obtained from Webgestalt and RPGeNet in this study, in an inclusive approach towards IRD gene discovery.

Chapter 6. Discussion

The considerable clinical and genetic heterogeneity displayed by IRDs confounds our understanding of the aetiology and pathophysiology of these conditions. Compelling clinical utility accompanies a genetic diagnosis, and major advancements have been made in gene-based therapeutic strategies. However, the molecular basis of IRDs in indigenous southern Africans has remained largely unexplored, due to challenges within the SA context, leading to recruitment bias and the use of testing strategies that had been constructed upon internationally-reported molecular findings in patients of European descent.

The goal of the current investigation was to delineate the molecular basis of disease in the indigenous African population group, in a novel approach towards studying genetic contributions to IRDs. As part of this endeavour, NGS technology (namely WES) was applied to identify IRD mutations. This advanced genomic technology has significantly impacted the characterisation of Mendelian diseases, especially IRDs. It was anticipated that the examination of these heterogeneous disorders in such a genetically diverse population group (which has historically been underrepresented in genomic studies), would be challenging. Nonetheless, an additional aim of this study was to provide direct benefit to local families by developing diagnostic assays in order to facilitate their translation towards a molecular genetic service. Furthermore, the information gathered from this research would contribute towards the understanding of IRDs in SA.

6.1 Summary of main results

The general strategies and approaches employed in this research, and the results obtained, are summarised in Figure 6.1.

In this study, mutations were identified in a novel IRD gene, *IDH3A*, which had not previously been linked to human retinal disease. In addition, pathologically relevant mutations were detected in several known IRD genes, one of which was a founder mutation.

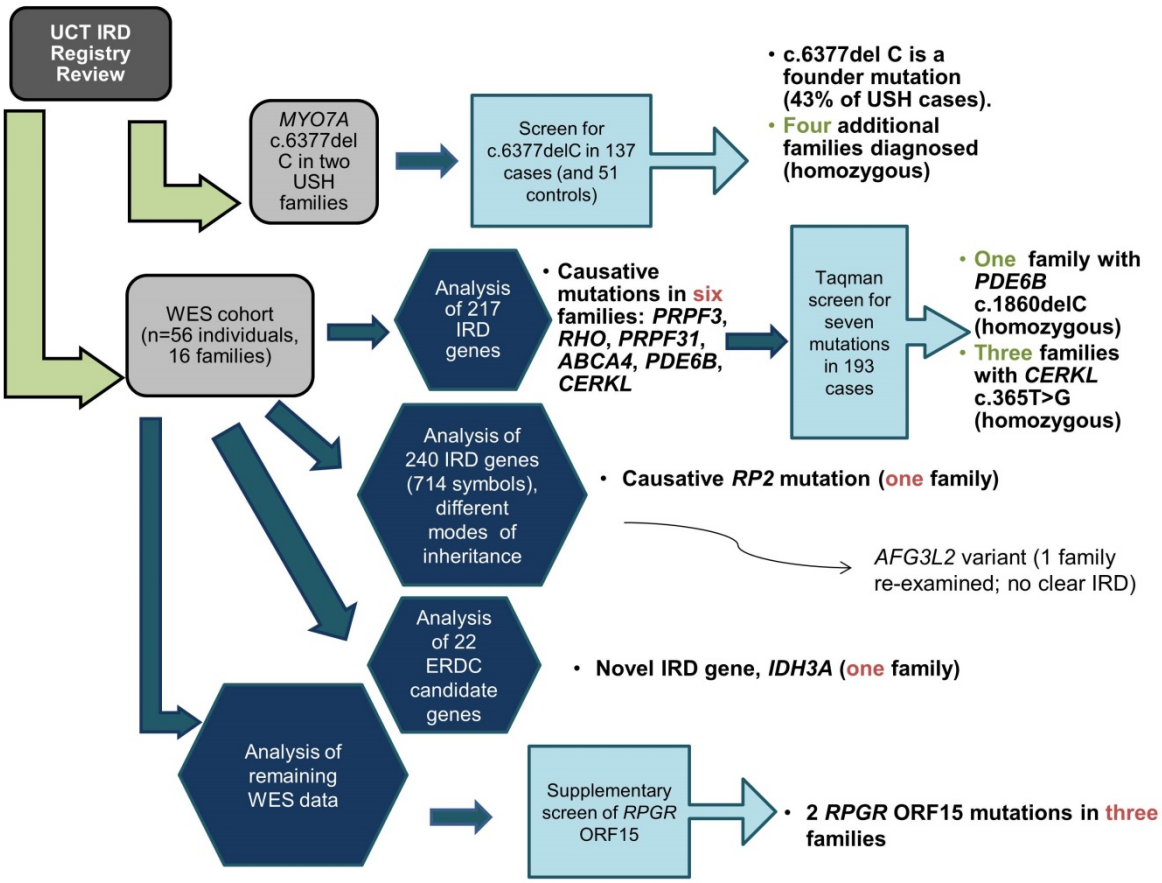


Figure 6.1 An illustration summarising the research strategy and the principal findings of this investigation. The general approaches are indicated in grey, stages of WES analysis are shown in dark blue hexagons, and screening strategies in light blue. The major findings obtained using each strategy are bulleted, with the number of resolved families highlighted in red (identified by WES) or green (identified by cascade screening).

Overall, thirteen different causative mutations were identified in ten distinct IRD genes as a result of this investigation. Single mutations (homozygous in *ar* cases and heterozygous/hemizygous in *ad/xl* cases) were identified in all genes except *ABCA4* and *IDH3A*, which each carried compound heterozygous mutations (i.e. *in trans)*, and *RPGR* ORF15 which carried two different mutations. Importantly, as a result of this research project, 48 affected individuals, from 19 distinct families, acquired a new genetic diagnosis (Figure 6.2A and B).

The strategy to use WES technology was extremely successful, as 11/15 (~73%) families in this cohort were resolved and the novel IRD gene (*IDH3A*) was identified in this portion of the project. This excludes the family in which a variant of interest, but unclear significance, was identified in *AFG3L2*. This particular family was not translated into diagnostic mode, due to the range of ophthalmological phenotypes

segregating within the family upon subsequent clinical re-evaluation. The *AFG3L2* variant is therefore not included in the calculations presented in this section.

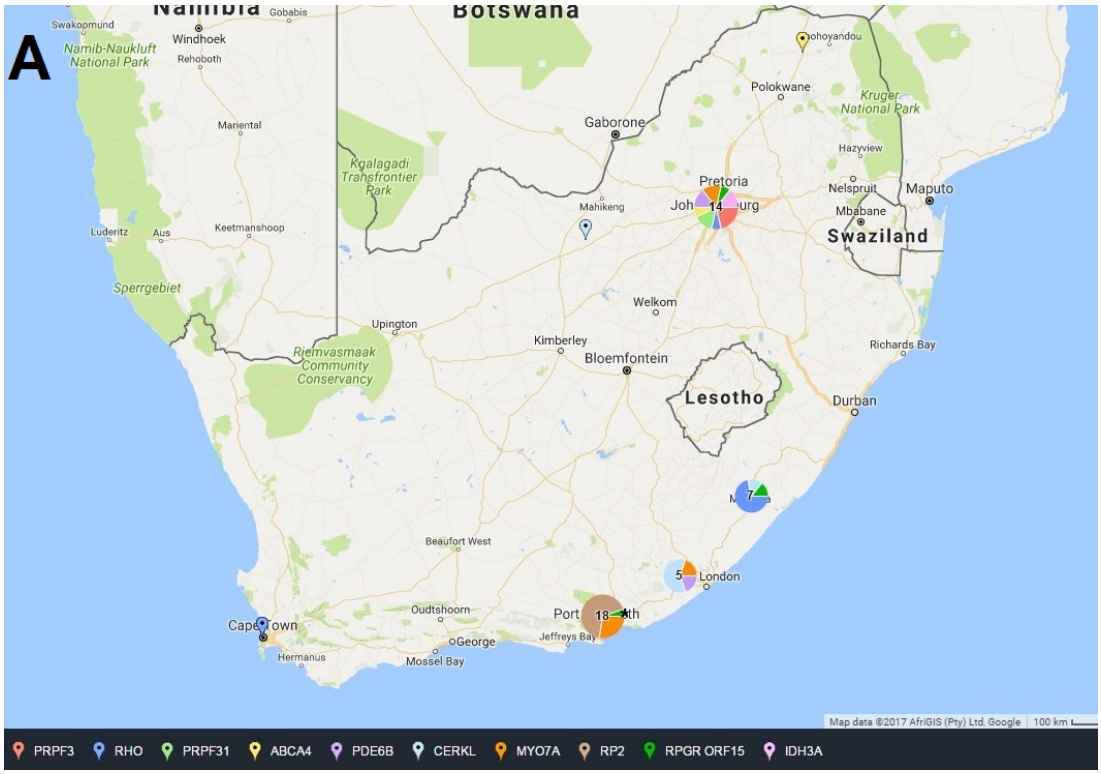


Figure 6.2(A) Map of South Africa illustrating the distribution of all affected individuals with IRD mutations identified during this study. Causative genes are represented in different colours, defined on the bottom of each map. The individual with the *RPGR* ORF15, c.2964_2965delGG mutation is differentiated from the two individuals with the *RPGR* ORF15 c.2790_2791delGG mutation, with an asterisk. *Credit: Sr. G. Benefeld, Division of Human Genetics, University of Cape Town.*

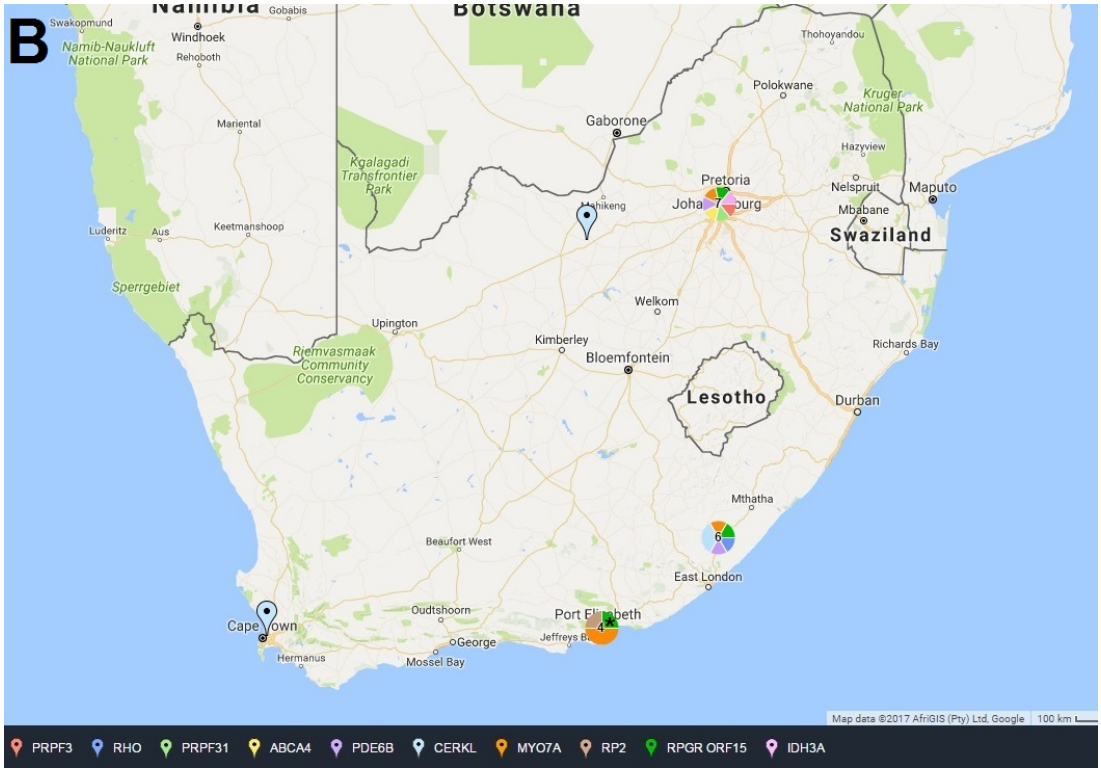


Figure 6.2(B) Map of South Africa illustrating the distribution of distinct families with IRD mutations identified during this study. Causative genes are represented in different colours, defined on the bottom of each map. A single proband from each family is shown, to highlight identical mutations (in *PDE6B*, *CERKL*, *MYO7A* and *RPGR* ORF15) identified in unrelated families. The individual with the *RPGR* ORF15, c.2964_2965delGG mutation is differentiated from the two unrelated individuals with the *RPGR* ORF15 c.2790_2791delGG mutation, with an asterisk. *Credit: Sr. G. Benefeld, Division of Human Genetics, University of Cape Town.*

Diagnostic assays were designed to: (a) facilitate rapid, sensitive and accurate testing for family members in the translational research programme, and (b) determine allele frequencies, and hence identify potential common mutations contributing to the burden of disease in SA. These assays, utilising either restriction enzyme digests or TaqMan technology, were designed to screen for eight of the mutations in larger cohorts of cases (an additional 137 or 193 unrelated, indigenous African IRD probands). Eight of the 19 families were diagnosed via this cascade screening, with their IRD attributed to the *MYO7A*, *PDE6B* or *CERKL* mutations. However, neither of the two *ABCA4* mutations, nor the *RHO*, *PRPF3* or *PRPF31* mutation was detected in any additional families. This observation raised the question of whether such cascade screening for any given mutation, simply by virtue of the fact that it has been detected in an indigenous African patient, is an

effective method (or an efficient use of resources). Thus, due to time constraints and cost considerations, and in order to channel research efforts optimally, assays were not designed to screen for the remaining five causative mutations in larger case cohorts.

Although screening assays were not specifically designed for the remaining five mutations, a majority (11/13) of mutations were validated by cycle sequencing, therefore these sequencing experiments have been optimised for any future diagnostic tests requested by relatives. The exceptions are the two *RPGR* ORF15 mutations, where the repetitive, low complexity nucleotide sequence presents significant technical challenges to any PCR-based analysis. It has been suggested that Sanger sequencing validation of NGS variants with robust quality scores may be unnecessary, because it limits the high-throughput nature of NGS technology[221]. Beck et al have questioned the utility of this practice, particularly since they found 17 NGS variants that were incorrectly called by a single round of Sanger sequencing in their study. However, cycle sequencing remains the gold-standard, serving the dual purpose of validation and the co-segregation analysis of variants, to establish pathogenicity. In this study, cycle sequencing was performed since many of the mutations detected were novel, and required confirmation of co-segregation with IRD within families.

6.2 Research perspectives and implications of findings

6.2.1 Genetic basis of IRDs in Africans

The findings of this 'pilot' study have made a significant impact towards elucidating the genetic landscape of IRD among the indigenous African populations of SA. At the time this project was initiated in 2014, 271 families (19%) in the IRD registry at UCT were of indigenous African ethic origin. Only 20 of these families (n=20/271, ~7%) were in diagnostic mode, having had their full complement of pathogenic mutations identified (Table 2.1). The causative genes in these previously diagnosed families were identified by reviewing the registry. These 10 causative genes were: *ABCA4* (n=4 families diagnosed), *BBS10* (n=5), *MYO7A* (n=2), *RHO* (n=3); *CDH23*, *CRB1*, *PCDH15*, *RP2*, *TOPORS*, and *USH2A* (n=1 family diagnosed with each gene).

The present study identified additional mutations in *ABCA4*, *RP2* and *RHO* in the local population, and implicated six IRD genes for the first time in indigenous Africans (*PRPF3*, *PRPF31*, *CERKL*, *PDE6B*, *IDH3A* and *RPGR*). This project has therefore almost doubled the number of indigenous African families with a molecular diagnosis in the UCT IRD registry, from 20 to 39 families.

Only one founder mutation had previously been associated with an IRD phenotype in this population group, *BBS10* c.728_731delAAGA, which causes BBS[118]. Founder mutations are important as: (i) the presence of numerous patients with the same causative mutation can facilitate research into the penetrance and genetic modifiers of disease, (ii) these mutations may be a major cause of disease, thus targeted testing along population lineages may be an efficient diagnostic strategy, and (iii) the presence of a founder mutation in different populations may be an indication of relatedness (i.e. the sharing of an identical segment of DNA, by descent)[120]. The research efforts of this study have revealed another founder mutation associated with syndromic IRD, namely *MYO7A* c.6377delC, which accounts for a large proportion (43%) of the USH cases in SA. Intriguingly, this USH founder mutation occurs in indigenous Africans of different ethnolinguistic groups, reflecting relatedness of these populations. This is a significant finding, as the targeted

screening of this mutation is a cost-effective enhancement to the diagnostic service in the country,

The c.6377delC *MYO7A* mutation had been reported previously[222], as had the *PDE6B* c.1860delC mutation[223] and two *RPGR* mutations, c.2790_2791delGG and c.2964_2965delGG[211,212], identified during this study. The remaining nine IRD mutations (~70%) were novel, explaining the low detection rate of Asper microarrays (testing 'Eurocentric' mutations) when applied to this group of patients. NGS approaches have detected novel IRD mutations to differing degrees in various population groups. Novel IRD mutations comprise ~30–60% of those identified in the UK and European populations[32,40,42,44,60,224], and 60% in the USA, including the Hispanic population[41,225]. In contrast, 67% and 76% of IRD mutations are novel in Chinese[46] and Thai[47] populations, respectively. Hence, the value of studying different populations (i.e. divergent from USA, UK and European populations) is clear, particularly for cataloguing novel IRD mutations and for understanding IRD in humans of all ethnic origins.

In 2016, Ellingford et al reported that, when the Manchester NGS panel of 105 IRD genes was applied to patients referred from institutions worldwide, 52% of identified mutations were not reported in databases[39]. This large, international cohort (n=537 individuals) provides a useful resource for comparison with our study, although the indigenous African WES cohort is small (n=16 families) and the majority of families in the UCT registry have not been comprehensively screened. The findings presented immediately hereunder should thus be extrapolated with caution.

The summary of the Manchester group's results revealed that, although 402 causative IRD mutations had been identified, overall the most common variant was *CERKL* c.375C>G; p.(Cys125Trp)[39]. This particular *CERKL* mutation was not identified in the African cohort, however the proximal mutation c.365T>G; p.(Leu122Arg) was identified in four unrelated homozygous probands (4/19 families, 21% of families characterised in this present study; 4/39 families, 10% of all characterised indigenous African families to date). Other findings also support the emerging picture of a dissimilar IRD mutation profile in indigenous Africans. It was reported that, within a single clinical category, the most frequent IRD-related genes identified using the Manchester panel were *EYS* for *ar* cone-rod dystrophy/RP and *RHO* for *ad* cone-rod dystrophy/RP[39]. In South Africa, it appears that (as well as the previously reported BBS mutation), mutations in *MYO7A, ABCA4* and *RHO* are the most frequent in causing USH (six families; four identified by screening in this project, and two already in the registry), MD/STGD (five families, one characterised

by this WES study and four already in the registry) and adRP (four families; one characterised by this WES study and three already in the registry), respectively. In Manchester, the most common mutations are missense variants (n=137/311, 44%), followed by nonsense variants (n=76/311, 24%) and frameshift insertion/deletion events (54/311, 17%). In contrast, in the 39 indigenous South African families resolved to date, 46% of mutations are missense (n=19/41 mutations), and a comparable proportion i.e. 41% of mutations (n=17/41 mutations) are frameshift insertion/deletion events, whilst 10% (n=4/41) mutations are nonsense variants. These figures are likely to have been impacted by the frameshift insertion/deletion mutations in *BBS10*, *MYO7A* and *PDE6B*, which each occur in several families.

It was interesting that four families in the WES cohort (n=4/15; 26%) were attributed to mutations in the known *xl* genes, *RP2* and *RPGR*. Technical challenges have largely prevented analysis of the ORF15 region in the Division of Human Genetics at UCT, therefore this high percentage may be artefactual. Nevertheless, with affected females manifesting in 3/4 families (RPD 54, RPD 1001, and RP 583), an *xl* mode of inheritance was not initially expected. Also surprising, was the identification of two unrelated Xhosa families (from different clans/tribes) with the same c.2790_2791delGG mutation, on a shared haplotype spanning *RPGR* ORF15. This suggests that this particular mutation may occur frequently in the local population therefore, in this particular case, cascade screening of a larger cohort may be warranted.

The WES data generated by this project provided additional information of general interest. A total of 1,816,031 variants were detected in 56 individuals. The description of these variants is presented in Table 6.1.

Table 6.1 Categories of ~1.8M variants identified in the exome sequencing data of 56 indigenous Africans.

Category	Number	Percentage of total (%)
Intergenic	759,459	41.82
Intronic (including ncRNA)	710,303	39.11
3’ UTR/5’ UTR	148,210	8.16
Exonic	129,631	7.14
ncRNA (excluding intronic ncRNA)	33,418	1.84
Upstream/downstream	33,660	1.85
Splicing	1,350	0.07

Approximately 7.2% of all genetic variation detected in the African exomes occurred within gene exons and splice regions. Detailed examination of the genetic variation in each of these 56 exomes, and comparison with the exomes of individuals from other population groups exceeded the scope of this study. Furthermore, to the best of our knowledge, most genomic studies in Africans involve genome-wide genotyping[105,109,111,116,226,227], NGS analysis in African populations which are different from those in our study[228–230], or very low coverage (4x) whole genome sequencing in relevant population groups[115]. The only publication reporting exome data from a "Bantu-speaking" individual, namely Archbishop Emeritus Desmond Tutu[104], which might be compared with the 56 exomes, focussed on SNPs. However, it is noted that the number of filtered (high priority) exome variants in the present research cohort exceeds that reported by Weisschuh et al in 2016[60]. In a largely European cohort, their WES analysis pipeline identified 100–150 variants per dominant family that were rare, potentially pathogenic and shared by two family members, thereby making the prioritisation of novel IRD candidate genes challenging[60]. In the present South African study, the five dominant families unresolved after the examination of IRD genes and ERDC genes, had on average 380 such variants (Table 5.2.1), more than two fold the reported figure. A minor number of *xl* variants are included in this number, however more relatives were included per family which reduced the number of variants co-segregating with disease.

Notwithstanding genomic diversity impeding the identification of further novel IRD genes, this study has several important implications for all the families which have been resolved. Members of these families can request genetic testing in order to confirm a clinical diagnosis, or determine carrier status. In addition, predictive testing and reproductive options[231] may be possible. Importantly, these patients may now be eligible for participation in gene-based therapy clinical trials (https://clinicaltrials.gov/ct2/home)[25,232], which is generally impossible in the absence of a genetic diagnosis. Of particular relevance are the lentiviral vector constructs being tested in humans to replace *MYO7A* and *ABCA4*, namely USHstat (ClinicalTrials.gov identifier NCT01505062)[233], and Stargen (ClinicalTrials.gov identifier NCT01367444)[234], respectively. Preclinical studies have also shown success for *RPGR*[177,235] and *RP2*[184] gene replacement, therefore human clinical trials are anticipated. In the aforementioned cases, haploinsufficiency is generally the cause of IRD and supplementation of the wild type gene ameliorates disease progression. However, for dominant negative mutations or toxic gain-of-function mutations, like the mutations in *RHO* which cause adRP, the mutant allele requires

silencing, and many such suppression and replacement approaches are being investigated[232,236,237]. It should be noted that there are currently many additional treatments for IRDs being investigated, which are not gene-specific, for instance stem-cell transplants[238], and optogenetic sensory replacement to confer light sensitivity to other cells in the photoreceptor-deprived retina[239].

6.2.2 Genotype – phenotype observations

Several interesting genotype-phenotype associations were noted during this investigation. For example, 60% of the *MYO7A* cases were clinically diagnosed with Type 2 USH, a less severe form of disease than the Type 1 USH which this gene is usually associated with.

The phenotypes associated with *RHO*, *PRPF3*, *PRPF31* (i.e. adRP) and *ABCA4* (i.e. MD) were as expected. However, observations were made regarding the phenotypes of the patients in our cohort with *PDE6B* and *CERKL* mutations. The mutations in each of these genes were initially identified through WES analysis, and subsequently detected in additional probands by TaqMan screening. The *CERKL* c.365T>G mutation was ultimately identified in four homozygous probands with different IRD phenotypes (i.e. two arRP, one sporadic RP, and one sporadic STGD). This particular variant was also identified in the heterozygous state in one proband with RP, whose mother apparently had vision problems (however no biological material or clinical data was available for the mother). *CERKL* is reportedly associated with *ar* forms of cone dystrophy, cone-rod dystrophy and RP (RetNet, https://sph.uth.edu/retnet/). It was thus somewhat surprising that an identical mutation resulted in distinct phenotypes in the various South African patients. The varying clinical diagnoses in the homozygous probands may be attributed to modifier effects or different stages of disease at the time of diagnosis. In the heterozygous case several possible explanations exist: (1) the *CERKL* mutation may be co-incidental with a mutation in a different (*ad*) gene; (2) a second heterozygous mutation exists in CERKL, and different compound heterozygous mutations within *CERKL* have resulted in the successive generations manifesting with disease; or (3) the observation regarding the mother's vision is misleading and she does not suffer from IRD. Overall, the *PDE6B* mutation was identified in two homozygous probands (with arRP) and four heterozygous probands (i.e. two sporadic RP cases, one arRP case, and one individual with an apparent adRP family history). *PDE6B* has previously been associated with the phenotypes of *ad*

CSNB and arRP[240]. CSNB is phenotypically similar to early stage arRP, as rod photoreceptor dysfunction causes nightblindness; however, rod apoptosis does not occur in CSNB as with arRP. The mutations resulting in arRP are thus presumed to be due to a complete loss of enzyme activity, causing elevated cGMP levels and subsequent rod photoreceptor apoptosis. Conversely, mutations producing the CSNB phenotype result in constitutive activation of PDE6 and consequently rod photoreceptor desensitisation[240]. No mutations in *PDE6B* have been associated with adRP, to date[241]. Hence, the proband with an apparent dominant family history is probably carrying the heterozygous mutation coincidentally as it appears to be relatively frequent in the affected indigenous Africans (n=8/418 alleles; 1.9%). The remaining three heterozygous mutation carriers (2 sporadic RP; 1 arRP) might harbour a mutation in the alternate allele of the gene.

It was remarkable that three of the four families with mutations in *xl* genes did not display a conventional *xl* inheritance pattern in their pedigrees, but contained affected females. This is a well-known phenomenon with *RPGR* mutations, however to the best of our knowledge only one other family with semi-dominant RP due to an *RP2* mutation has been reported[143]. Many individuals in the South African family with an *RP2* mutation had cataracts, and most of the affected females (n=4/5) who were clinically examined did not match the Grover et al classification for manifesting female carriers[140].

The phenotype associated with the novel IRD gene, *IDH3A,* appears to be correlated to genotype, in so far as patients with arRP and coloboma-like atrophy of the macula carried a null variant, whereas patients with missense variants present with classic arRP (manuscript accepted 3 March 2017, *Ophthalmology*, the journal of the American Academy of Ophthalmology). However, intrafamilial phenotypic variability has been noted. Furthermore, the first occurrence of a recessive mutation in the *IDH3A* gene was reported very recently[242], and was associated with severe encephalopathy in an infant of North-African-Jewish ancestry. The clinical manifestations of the child included retinitis pigmentosa and bilateral optic atrophy, amongst other nervous system defects. To the best of our knowledge the seven patients (four families) identified via the ERDC collaboration and this infant with encephalopathy represent the first cases of human disease linked to *IDH3A*. Identification of additional cases with recessive *IDH3A* mutations may provide more information surrounding the phenotype(s) associated with this gene.

Finally, the identification of a missense variant in *AFG3L2*, in a mother with RP and her daughter with thyroid eye disease was intriguing, despite the fact that this family

did not display segregation of a clear, single, Mendelian IRD. Functional studies are required to assess whether this variant is pathological. If shown to cause disease, this would extend the phenotypic spectrum associated with this gene. Mutations in *AFG3L2* have previously been linked with SCA28, progressive myoclonus epilepsy; spastic ataxia neuropathy and DOA, but not with RP or thyroid eye disease.

In conclusion, we hypothesise that IRD phenotypes observed in the indigenous Africans, with their vastly different genetic and environmental landscape may, in some cases, be distinct from those described in mostly Caucasian patients in the literature.

6.2.3 Genetic screening approaches in indigenous Africans

During the course of this investigation, Van Huet et al reported a low efficacy of the Asper Ophthalmics (http://www.asperbio.com/asper-ophthalmics) arRP APEX microarray in a European cohort[243], confirming the premise of our study. The microarray solved only ~9% of a large cohort of 250 European probands with arRP, corresponding to the detection rates of 7% Caucasians and 13% indigenous Africans in SA (although the local cohort was smaller, Table 2.2) using the same array. They ascribed this low efficacy to several reasons, which would equally explain the overall poor yield of this technology in indigenous African samples, namely: (1) The microarrays test only for specific sets of reported mutations, which have primarily been identified in patients of European ancestry; (2) the microarrays are not comprehensive, as they are not updated frequently and do not contain rare mutations and genes; (3) the clinical and genetic heterogeneity of IRDs mean that an inappropriate microarray might be selected for use in a given proband; and (4) older versions of microarrays contained benign variants, thereby reducing their capacity[243]. In this analysis, WES (and supplementary *RPGR* ORF screening) solved 73% of the indigenous African families with IRDs, a vast improvement over our previous success rate of 13% using arrays, thereby confirming the superiority of NGS approaches.

Prior reports indicated that known IRD genes account for only 50–70% of diagnosed cases[24]. Since the genomic portraits of global populations differ significantly from each other, the contribution of known IRD genes to the burden of disease in indigenous Africans was previously unknown, hence WES was the method of choice for this study. It was established that 10/15 (~67%) of the local cases were resolved with mutations in reported IRD genes. Four families remain unattributed

and one family was characterised with a novel IRD gene. Considering *RPGR* ORF 15 is not captured in targeted panels, it is likely that 7/15 (~47%) of the South African families would have been identified, had they been screened using a panel-based approach, with targeted enrichment of known IRD genes. This is only slightly lower than the 51–57% panel diagnostic rate published by several groups in the last year[39,50,51].

In their review of the Manchester panel, Ellingford et al reported a trend that 'newly identified genes are becoming increasingly rarer causes' of IRD, with 79% of diagnosed cases attributed to genes discovered between 1995 and 2004[39]. They suggested that more extensive analysis of known IRD genes, for example for deep-intronic or copy number variants, might further impact the diagnostic yield. The use of targeted IRD panels rather than WES has also been advocated by Consugar et al[50]. In a direct, empiric, comparison of panel vs WES for four samples, they reported that WES was ~10% less sensitive, due to less efficient coverage of key genomic positions. They thus indicated that probe design in panels was improved, compared to the probes used in commercial exome capture kits, resulting in increased and more uniform coverage. Other advantages of panels were described, namely: (1) reduced turnaround time, (2) increased probability of a meaningful result, and (3) reduced possibility of incidental findings[50]. Souzeau et al have recently presented the ethical considerations surrounding the return of incidental findings in ophthalmic genetic research[244], and in terms of practicalities, important points were raised. Notably, most research laboratories are not accredited nor equipped to report incidental findings, as "researchers often have a lack of expertise for results or conditions that are outside of their scope of research" and "the requirements for the return of incidental findings….put an unsustainable burden on the research enterprise"[244].

Although WES has significant advantages over panels in a pure research arena, and allows re-analysis of data as novel genes are reported, the filtering of large amounts of this data in order to isolate a causative mutation remains the primary challenge. This challenge in the current study in particular, was exacerbated with the indigenous African samples. Both targeted IRD panels and WES can detect novel mutations in reported genes, as well as mutations in unexpected IRD genes[245] (outside of the known genotype-phenotype relationships). Therefore, perhaps a two-tiered approach, as suggested by Carrigan et al[51] should be employed for indigenous African IRD sufferers in the translational research programme in SA. If targeted capture panels were utilised initially for these patients, it could reduce the

number of families requiring WES by as much as 50%. However, the design of such panels requires careful consideration, whether to include deep intronic mutations known to cause IRDs[39,50], and whether to include syndromic genes, as hypomorphic mutations of these genes could result in a milder phenotype of IRD without extra-ocular symptoms[41]. In addition, targeted enrichment probes would still not capture GC rich sequences with repetitive elements like *RPGR* ORF15[50], and this region, in particular, would require supplementary analysis[218] in SA.

6.3 Study limitations

The small size of the WES cohort (n=16 families) was constrained by the selection criteria that were applied. A larger cohort would have been preferable for evaluating the WES findings and extrapolating these results as putative trends in the molecular basis of indigenous African IRDs.

It is also important to note that there are several reasons true causative mutations could have been missed during WES analysis, which may be relevant to the four unresolved families that remain.

Beginning with the laboratory experiment, WES kits do not actually capture all the exons of all known genes and are biased against GC rich regions[27]. Although 92–95% of exons are captured[27], several known IRD genes have been reported with low coverage of particular exons due to the sequence GC content, for example *IMPDH1* exon 1, *PEX7* exon 1, *PITPNM3* exon 1, *FLVCR1* exon 1, *RP9* exon 1, *CHM* exon 5[42], and *RPGR* ORF15[210]. There are particular types of mutations which are not detected by WES, such as deep intronic mutations, for example *CEP290* c.2991+1655A>G, a frequent cause of LCA[246–248]. In addition, during the alignment stage of WES, if an indel occurs, the entire read is discarded as it does not align to the human reference sequence, thus large genomic alterations are not detected using this technique[249,250]. For example, a 356bp *Alu* repeat insertion in the *MAK* gene (associated with RP) is missed by variant callers and detection of this recurrent mutation requires analysis of raw sequence reads[250]. Approximately 7% of mutations listed in HGMD are large indels, repeats or complex rearrangements[250]. It is known that deletions of *PRPF31* are responsible for ~3% of adRP, and *ABCA4* and *RPGR* deletions have also been identified[175]. Furthermore, in 2016, Khateb et al performed manual coverage analysis of IRD gene exons captured by WES, and determined that 10% of IRD cases had large homozygous or hemizygous deletions[251]. The deletions ranged from ~2Kb to ~76Kb, even encompassing as

many as seven consecutive genes on the X chromosome. The finding that copy number variants (CNVs) contribute significantly to the missing heritability of IRDs was supported by Bujakowska et al, who identified large deletions of IRD genes in 18% of the cases that were not solved by NGS approaches, ~7% of IRD cases overall, and further identifying 35 known IRD genes may be prone to non-allelic homologous recombination thus representing CNV hotspots[252].

True mutations may also have been eliminated during almost any stage of the WES candidate prioritisation pipeline. In extracting certain sets of genes, e.g. known IRD genes, from WES data the possibility of gene aliases in either dataset is problematic, as relevant data may not have been retrieved. Pathogenicity prediction thresholds may also have been too stringent. The identification of low MAF variants may also introduce errors; during initiation of their bioinformatics pipeline, Crowgey et al noted that, for several alleles, the hg19 human reference genome used for alignment actually contained the minor allele[253]. This is an issue, particularly with understudied populations, as the true minor allele is filtered out because it is reported "in the 96–100% range rather than the expected 0–4%"[253]. Variant prioritisation based on cosegregation might fail if: (1) the incorrect inheritance pattern is assumed, (2) genetic mosaicism results in an allelic imbalance[249,253], or (3) digenic inheritance is the cause of disease[254]. It is therefore possible that mutations in reported IRD genes are in fact, causal in the four remaining families.

6.4 Future research

This research project has identified several avenues for future exploration. Primarily, there are four families outstanding, which should be analysed further. This may involve prioritisation tools such as TOPPGENE, as described in Chapter 5.2. Though reported IRD genes display a wide range of functions, about 15% of IRD genes (the largest category) are those involved in photoreceptor ciliary functioning[7,65,255], therefore these genes might be considered first, within the four families.

Alternatively, approaches could be taken to either screen for known indels, like the *Alu* repeat in *MAK*[250], or to attempt to leverage the WES data to detect CNVs. The latter may be challenging; several tools have been developed for CNV detection in WES data, (e.g. ExomeCNV, CoNIFER, cnvOffSeq, ExomeDepth, and CEQer), however, Khateb et al relied on a manual approach to detect CNVs, as many tools require a minimum coverage quality which might not be achieved in WES, resulting

in numerous false-positives. Furthermore, read depth aberrations from different WES batches suggests that larger numbers of (unrelated) samples, or larger batch sizes should be used for CNV analysis[256], and the 16 families of this study may be insufficient. Thus, a better approach might be to generate array-based comparative genomic hybridisation (array CGH) data for the four families, to combine with the existing WES data. Van Cauwenbergh and colleagues in Belgium, for example, have recently designed a customised array CGH platform, arrEYE, for CNV analysis of IRD genes[257].

There are several additional findings from this study which may form the basis of new research investigations. *RPGR* ORF15 warrants screening, using either traditional or NGS approaches, in IRD patients of all ethnic origins in SA and this should not be limited to families with an obvious *xl* inheritance pattern. Also related to xlRP, the WES data from the semi-dominant family with an *RP2* mutation may be interrogated to identify disease modifiers to explain the disease manifestation in females. These modifiers may segregate with the *RP2* mutation in this family, given the number of affected females observed. Hence, all variants on the X chromosome in the affected males could be analysed, extending outwards (5' and 3' directions) from the c.704C>A; p.(S235*) mutation. The modifiers may affect CpG islands, or genes interacting with *RP2*, to effect a consistently skewed X-inactivation in retinal cells.

Functional analysis of *AFG3L2* is recommended, to ascertain whether the variant identified in this study is pathogenic. This variant may affect the interaction of this m-AAA protease with a retinal substrate, or it could be a mild mutation which only meets retinal-specific thresholds for disease (i.e. is not pathological in other tissues).

Finally, the novel IRD gene, *IDH3A*, should be screened in large cohorts of patients (not restricted to SA) to determine the contribution towards disease, therefore the gene should be included in IRD targeted capture panels. Understanding the molecular mechanism of disease of *IDH3A* should be a priority.

6.5 Conclusion and recommendations

In conclusion, this study has provided new insights into the genetic profile of IRDs in the South African population. Whilst common causative mutations do occur in indigenous Africans, in general each family harbours unique mutations. It is thus recommended that APEX microarrays should no longer be used to characterise the mutations causing IRD in these patients. The majority of causative mutations in this population group are novel, and therefore are not detected using this technique. Since most mutations do occur in reported IRD genes, resources should instead be directed towards NGS approaches, either WES or targeted capture panels, which are likely to identify approximately 50% of local cases. Additional cases will almost certainly be diagnosed if supplementary *RPGR* screening is performed. Furthermore, cascade screening of specific mutations should only be implemented in indigenous African case cohorts if at least two unrelated probands are identified with an identical mutation; although founder mutations can contribute significantly to IRDs, screening for a given mutation by virtue of the fact it was detected in a single patient, is inefficient.

An important lesson learnt from this study, which is applicable to IRD researchers globally, is that with all NGS analysis the *xl* genes should be considered, regardless of the perceived mode of inheritance in a family.

The strategies employed, and outcomes of this work, have significant implications for stakeholders in Africa. Specifically, these findings are important to those initiating IRD research (or diagnostic services) in other African countries. Furthermore this approach may be useful to those interested in examining other heritable disabilities on the continent.

Finally, the importance of including African populations in genomic research cannot be overstated. Research in Africa has its' challenges, including lack of resources, logistical obstacles and language/geographic barriers. However, unless this bias is addressed, and diverse genomes are better represented in medical studies, inequalities in precision medicine and human health will remain unaddressed. Mutation- and gene- discoveries in diverse populations may lead to improved understanding of disease mechanisms and novel therapeutics. African populations, and African scientists, play an important role in this endeavour, and the body of work presented in this study makes an important contribution towards inclusivity in this regard.

References

1. Dandona, L., and Dandona, R. (2006). What is the global burden of visual impairment? BMC Med. *4*, 6–15.

2. Kolb, H. (2003). How the Retina Works. Am. Sci. *91*, 28–35.

3. Siegert, S., Scherf, B.G., Del Punta, K., Didkovsky, N., Heintz, N., and Roska, B. (2009). Genetic address book for retinal cell types. Nat. Neurosci. *12*, 1197–1204.

4. Swaroop, A., Kim, D., and Forrest, D. (2010). Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. Nat. Rev. Neurosci. *11*, 563–576.

5. Bok, D. (1990). Processing and transport of retinoids by the retinal pigment epithelium. Eye *4*, 326–332.

6. Chabre, M., and Deterre, P. (1989). Molecular mechanism of visual transduction. Eur. J. Biochem. *179*, 255–266.

7. Wright, A.F., Chakarova, C.F., Abd El-Aziz, M.M., and Bhattacharya, S.S. (2010). Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. Nat. Rev. Genet. *11*, 273–284.

8. Boulton, M., and Dayhaw-Barker, P. (2001). The role of the retinal pigment epithelium: topographical variation and ageing changes. Eye *15*, 384–389.

9. Palczewski, K. (2012). Chemistry and biology of vision. J. Biol. Chem. *287*, 1612–1619.

10. Schulz, H.L., Goetz, T., Kaschkoetoe, J., and Weber, B.H.F. (2004). The Retinome - defining a reference transcriptome of the adult mammalian retina/retinal pigment epithelium. BMC Genomics *5*, 50–60.

11. Tanackovic, G., Ransijn, A., Thibault, P., Abou Elela, S., Klinck, R., Berson, E.L., Chabot, B., and Rivolta, C. (2011). PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. Hum. Mol. Genet. *20*, 2116–2130.

12. Farkas, M.H., Grant, G.R., White, J.A., Sousa, M.E., Consugar, M.B., and Pierce, E.A. (2013). Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. BMC Genomics *14*, 486–499.

13. Whitmore, S.S., Wagner, A.H., DeLuca, A.P., Drack, A. V., Stone, E.M., Tucker, B.A., Zeng, S., Braun, T.A., Mullins, R.F., and Scheetz, T.E. (2014). Transcriptomic analysis across nasal, temporal, and macular regions of human neural retina and RPE/choroid by RNA-Seq. Exp. Eye Res. *129*, 93–106.

14. Pinelli, M., Carissimo, A., Cutillo, L., Lai, C.H., Mutarelli, M., Moretti, M.N., Singh, M.V., Karali, M., Carrella, D., Pizzo, M., et al. (2016). An atlas of gene expression and gene co-regulation in the human retina. Nucleic Acids Res. *44*, 5773–5784.

15. Berger, W., Kloeckener-Gruissem, B., and Neidhardt, J. (2010). The molecular basis of human retinal and vitreoretinal diseases. Prog. Retin. Eye Res. *29*, 335–375.

16. Rivolta, C., Sharon, D., DeAngelis, M.M., and Dryja, T.P. (2002). Retinitis pigmentosa and allied diseases: numerous diseases, genes, and inheritance patterns. Hum. Mol. Genet. *11*, 1219–1227.

17. Ayuso, C., and Millan, J.M. (2010). Retinitis pigmentosa and allied conditions today: a paradigm of translational research. Genome Med. *2*, 34–44.

18. Frio, T.R., Civic, N., Ransijn, A., Beckmann, J.S., and Rivolta, C. (2008). Two trans-acting eQTLs modulate the penetrance of PRPF31 mutations. Hum. Mol. Genet. *17*, 3154–3165.

19. Rivolta, C., McGee, T.L., Frio, T.R., Jensen, R.V, Berson, E.L., and Dryja, T.P. (2006). Variation in retinitis pigmentosa-11 (PRPF31 or RP11) gene expression between symptomatic and asymptomatic patients with dominant RP11 mutations. Hum. Mutat. *27*, 644–653.

20. Walia, S., Fishman, G.A., Swaroop, A., Branham, K.E.H., Lindeman, M., Othman, M., and Weleber, R.G. (2008). Discordant phenotypes in fraternal twins having an identical mutation in exon ORF15 of the RPGR gene. Arch. Ophthalmol. *126*, 379–384.

21. Werdich, X.Q., Place, E.M., and Pierce, E.A. (2014). Systemic diseases associated with retinal dystrophies. Semin. Ophthalmol. *29*, 319–328.

22. Gillespie, R.L., Hall, G., and Black, G.C. (2014). Genetic testing for inherited ocular disease: delivering on the promise at last? Clin. Exp. Ophthalmol. *42*, 65–77.

23. Daiger, S.P., Sullivan, L.S., Bowne, S.J., Birch, D.G., Heckenlively, J.R., Pierce, E.A., and Weinstock, G.M. (2010). Targeted high-throughput DNA sequencing for gene discovery in retinitis pigmentosa. Adv. Exp. Med. Biol. *664*, 325–331.

24. Nishiguchi, K.M., and Rivolta, C. (2012). Genes associated with retinitis pigmentosa and allied diseases are frequently mutated in the general population. PLoS One *7*, e41902.

25. Thompson, D.A., Ali, R.R., Banin, E., Branham, K.E., Flannery, J.G., Gamm, D.M., Hauswirth, W.W., Heckenlively, J.R., Iannaccone, A., Jayasundera, K.T., et al. (2015). Advancing therapeutic strategies for inherited retinal degeneration: recommendations from the Monaciano Symposium. Invest. Ophthalmol. Vis. Sci. *56*, 918–931.

26. 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.

27. Chaitankar, V., Karakülah, G., Ratnapriya, R., Giuste, F.O., Brooks, M.J., and Swaroop, A. (2016). Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research. Prog. Retin. Eye Res. *55*, 1–31.

28. Ratnapriya, R., and Swaroop, A. (2013). Genetic architecture of retinal and macular degenerative diseases: the promise and challenges of next-generation sequencing. Genome Med. *5*, 84–97.

29. Zernant, J., Schubert, C., Im, K.M., Burke, T., Brown, C.M., Fishman, G.A., Tsang, S.H., Gouras, P., Dean, M., and Allikmets, R. (2011). Analysis of the ABCA4 gene by next-generation sequencing. Invest. Ophthalmol. Vis. Sci. *52*, 8479–8487.

30. Benaglio, P., Mcgee, T.L., Capelli, L.P., Harper, S., Berson, E.L., and Rivolta, C. (2011). Next generation sequencing of pooled samples reveals new SNRNP200 mutations associated with retinitis pigmentosa. Hum. Mutat. *32*, E2246–E2258.

31. Simpson, D.A., Clark, G.R., Alexander, S., Silvestri, G., and Willoughby, C.E. (2011). Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa. J. Med. Genet. *48*, 145–151.

32. González-del Pozo, M., Borrego, S., Barragán, I., Pieras, J.I., Santoyo, J., Matamala, N., Naranjo, B., Dopazo, J., and Antiñolo, G. (2011). Mutation screening of multiple genes in Spanish patients with autosomal recessive retinitis pigmentosa by targeted resequencing. PLoS One *6*, e27894.

33. Clark, G.R., Crowe, P., Muszynska, D., O'Prey, D., O'Neill, J., Alexander, S., Willoughby, C.E., McKay, G.J., Silvestri, G., and Simpson, D.A. (2010). Development of a diagnostic genetic test for simplex and autosomal recessive retinitis pigmentosa. Ophthalmology *117*, 2169–77.e3.

34. Coppieters, F., De Wilde, B., Lefever, S., De Meester, E., De Rocker, N., Van Cauwenbergh, C., Pattyn, F., Meire, F., Leroy, B.P., Hellemans, J., et al. (2012). Massively parallel sequencing for early molecular diagnosis in Leber congenital amaurosis. Genet. Med. *14*, 576–585.

35. 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*, 8–24.

36. 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.

37. Neveling, K., Collin, R.W.J., Gilissen, C., van Huet, R.A.C., Visser, L., Kwint, M.P., Gijsen, S.J., Zonneveld, M.N., Wieskamp, N., de Ligt, J., et al. (2012). Next-generation genetic testing for retinitis pigmentosa. Hum. Mutat. *33*, 963–972.

38. Booij, J.C., Bakker, A., Kulumbetova, J., Moutaoukil, Y., Smeets, B., Verheij, J., Kroes, H.Y., Klaver, C.C.W., van Schooneveld, M., Bergen, A.A.B., et al. (2011). Simultaneous mutation detection in 90 retinal disease genes in multiple patients using a custom-designed 300-kb retinal resequencing chip. Ophthalmology *118*, 160–167.

39. 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. (2016). Molecular findings from 537 individuals with inherited retinal disease. J. Med. Genet. *53*, 761–767.

40. 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.

41. Zhang, Q., Xu, M., Verriotto, J.D., Li, Y., Wang, H., Gan, L., Lam, B.L., and Chen, R. (2016). Next-generation sequencing-based molecular diagnosis of 35 Hispanic retinitis pigmentosa probands. Sci. Rep. *6*, 32792.

42. Boulanger-Scemama, E., El Shamieh, S., Démontant, V., Condroyer, C., Antonio, A., Michiels, C., Boyard, F., Saraiva, J.P., Letexier, M., Souied, E., et al. (2015). Next-generation sequencing applied to a large French cone and cone-rod dystrophy cohort: mutation spectrum and new genotype-phenotype correlation. Orphanet J Rare Dis. *10*, 85–104.

43. Chen, X., Sheng, X., Sun, X., Zhang, Y., Jiang, C., Li, H., Ding, S., Liu, Y., Liu, W., Li, Z., et al. (2016). Next-generation Sequencing Extends the Phenotypic Spectrum for LCA5 Mutations: Novel LCA5 Mutations in Cone Dystrophy. Sci. Rep. *6*, 24357.

44. Corton, M., Nishiguchi, K.M., Avila-Fernández, A., Nikopoulos, K., Riveiro-Alvarez, R., Tatu, S.D., Ayuso, C., and Rivolta, C. (2013). Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis. PLoS One *8*, e65574.

45. Abu-Safieh, L., Alrashed, M., Anazi, S., Alkuraya, H., Khan, A.O., Al-Owain, M., Al-Zahrani, J., Al-Abdi, L., Hashem, M., Al-Tarimi, S., et al. (2013). Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. Genome Res. *23*, 236–247.

46. Xu, Y., Guan, L., Shen, T., Zhang, J., Xiao, X., Jiang, H., Li, S., Yang, J., Jia, X., Yin, Y., et al. (2014). Mutations of 60 known causative genes in 157 families with retinitis pigmentosa based on exome sequencing. Hum. Genet. *133*, 1255–1271.

47. Jinda, W., Taylor, T.D., Suzuki, Y., Thongnoppakhun, W., Limwongse, C., Lertrit, P., Suriyaphol, P., Trinavarat, A., and Atchaneeyasakul, L. (2014). Whole exome sequencing in Thai patients with retinitis pigmentosa reveals novel mutations in six genes. Invest. Ophthalmol. Vis. Sci. *55*, 2259–2268.

48. Lazar, C.H., Mutsuddi, M., Kimchi, A., Zelinger, L., Mizrahi-Meissonnier, L., Marks-Ohana, D., Boleda, A., Ratnapriya, R., Sharon, D., Swaroop, A., et al. (2015). Whole Exome Sequencing Reveals GUCY2D as a Major Gene Associated With Cone and Cone-Rod Dystrophy in Israel. Invest. Ophthalmol. Vis. Sci. *56*, 420–430.

49. Beryozkin, A., Shevah, E., Kimchi, A., Mizrahi-Meissonnier, L., Khateb, S., Ratnapriya, R., Lazar, C.H., Blumenfeld, A., Ben-Yosef, T., Hemo, Y., et al. (2015). Whole Exome Sequencing Reveals Mutations in Known Retinal Disease Genes in 33 out of 68 Israeli Families with Inherited Retinopathies. Sci. Rep. *5*, 13187.

50. 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.

51. 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. (2016). Panel-Based Population Next-Generation Sequencing for Inherited Retinal Degenerations. Sci. Rep. *6*, 33248.

52. Lee, K., Berg, J.S., Milko, L., Crooks, K., Lu, M., Bizon, C., Owen, P., Wilhelmsen, K.C., Weck, K.E., Evans, J.P., et al. (2015). High Diagnostic Yield of Whole Exome Sequencing in Participants With Retinal Dystrophies in a Clinical Ophthalmology Setting. Am. J. Ophthalmol. *160*, 354–363.e9.

53. Wang, X., Wang, H., Cao, M., Li, Z., Chen, X., Patenia, C., Gore, A., Abboud, E.B., Al-Rajhi, A.A., Lewis, R.A., et al. (2011). Whole-exome sequencing identifies ALMS1, IQCB1, CNGA3, and MYO7A mutations in patients with Leber congenital amaurosis. Hum. Mutat. *32*, 1450–1459.

54. Manes, G., Meunier, I., Avila-Fernández, A., Banfi, S., Le Meur, G., Zanlonghi, X., Corton, M., Simonelli, F., Brabet, P., Labesse, G., et al. (2013). Mutations in IMPG1 Cause Vitelliform Macular Dystrophies. Am. J. Hum. Genet. *93*, 571–578.

55. Avila-Fernandez, A., Perez-Carro, R., Corton, M., Lopez-Molina, M.I., Campello, L., Garanto, A., Fernandez-Sanchez, L., Duijkers, L., Lopez-Martinez, M.A., Riveiro-Alvarez, R., et al. (2015). Whole-exome sequencing reveals ZNF408 as a new gene associated with autosomal recessive retinitis pigmentosa with vitreal alterations. Hum. Mol. Genet. *24*, 4037–4048.

56. Charif, M., Roubertie, A., Salime, S., Mamouni, S., Goizet, C., Hamel, C.P., and Lenaers, G. (2015). A novel mutation of AFG3L2 might cause dominant optic atrophy in patients with mild intellectual disability. Front. Genet. *6*, 311.

57. Haer-Wigman, L., Newman, H., Leibu, R., Bax, N.M., Baris, H.N., Rizel, L., Banin, E., Massarweh, A., Roosing, S., Lefeber, D.J., et al. (2015). Non-syndromic retinitis pigmentosa due to mutations in the mucopolysaccharidosis type IIIC gene, heparan-alpha-glucosaminide N-acetyltransferase (HGSNAT). Hum. Mol. Genet. *24*, 3742–3751.

58. Xu, M., Yamada, T., Sun, Z., Eblimit, A., Lopez, I., Wang, F., Manya, H., Xu, S., Zhao, L., Li, Y., et al. (2016). Mutations in POMGNT1 cause non-syndromic retinitis pigmentosa. Hum. Mol. Genet. *25*, 1479–1488.

59. Almoguera, B., Li, J., Fernandez-San Jose, P., Liu, Y., March, M., Pellegrino, R., Golhar, R., Corton, M., Blanco-Kelly, F., López-Molina, M.I., et al. (2015). Application of Whole Exome Sequencing in Six Families with an Initial Diagnosis of Autosomal Dominant Retinitis Pigmentosa: Lessons Learned. PLoS One *10*, e0133624.

60. 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.

61. Xu, M., Yang, L., Wang, F., Li, H., Wang, X., Wang, W., Ge, Z., Wang, K., Zhao, L., Li, H., et al. (2015). Mutations in human IFT140 cause non-syndromic retinal degeneration. Hum. Genet. *134*, 1069–1078.

62. El Shamieh, S., Neuillé, M., Terray, A., Orhan, E., Condroyer, C., Démontant, V., Michiels, C., Antonio, A., Boyard, F., Lancelot, M.E., et al. (2014). Whole-exome sequencing identifies KIZ as a ciliary gene associated with autosomal-recessive rod-cone dystrophy. Am. J. Hum. Genet. *94*, 625–633.

63. Sergouniotis, P.I., Chakarova, C., Murphy, C., Becker, M., Lenassi, E., Arno, G., Lek, M., MacArthur, D.G., UCL-Exomes Consortium, Bhattacharya, S.S., et al. (2014). Biallelic variants in TTLL5, encoding a tubulin glutamylase, cause retinal dystrophy. Am. J. Hum. Genet. *94*, 760–769.

64. Kastner, S., Thiemann, I.J., Dekomien, G., Petrasch-Parwez, E., Schreiber, S., Akkad, D.A., Gerding, W.M., Hoffjan, S., Günes, S., Günes, S., et al. (2015). Exome Sequencing Reveals AGBL5 as Novel Candidate Gene and Additional Variants for Retinitis Pigmentosa in Five Turkish Families. Invest. Ophthalmol. Vis. Sci. *56*, 8045–8053.

65. Soens, Z.T., Li, Y., Zhao, L., Eblimit, A., Dharmat, R., Li, Y., Chen, Y., Naqeeb, M., Fajardo, N., Lopez, I., et al. (2016). Hypomorphic mutations identified in the candidate Leber congenital amaurosis gene CLUAP1. Genet. Med. *18*, 1044–1051.

66. Roosing, S., Lamers, I.J.C., de Vrieze, E., van den Born, L.I., Lambertus, S., Arts, H.H., POC1B Study Group, Peters, T.A., Hoyng, C.B., Kremer, H., et al. (2014). Disruption of the basal body protein POC1B results in autosomal-recessive cone-rod dystrophy. Am. J. Hum. Genet. *95*, 131–142.

67. Namburi, P., Ratnapriya, R., Khateb, S., Lazar, C.H., Kinarty, Y., Obolensky, A., Erdinest, I., Marks-Ohana, D., Pras, E., Ben-Yosef, T., et al. (2016). Bi-allelic Truncating Mutations in CEP78, Encoding Centrosomal Protein 78, Cause Cone-Rod Degeneration with Sensorineural Hearing Loss. Am. J. Hum. Genet. *99*, 777–784 (Erratum 1222–1223).

68. Nikopoulos, K., Farinelli, P., Giangreco, B., Tsika, C., Royer-Bertrand, B., Mbefo, M.K., Bedoni, N., Kjellström, U., El Zaoui, I., Di Gioia, S.A., et al. (2016). Mutations in CEP78 Cause Cone-Rod Dystrophy and Hearing Loss Associated with Primary-Cilia Defects. Am. J. Hum. Genet. *99*, 770–776.

69. 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.J., 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.

70. 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.

71. Xu, M., Eblimit, A., Wang, J., Li, J., Wang, F., Zhao, L., Wang, X., Xiao, N., Li, Y., Wong, L.C., et al. (2016). ADIPOR1 Is Mutated in Syndromic Retinitis Pigmentosa. Hum. Mutat. *37*, 246–249.

72. Minegishi, Y., Sheng, X., Yoshitake, K., Sergeev, Y., Iejima, D., Shibagaki, Y., Monma, N., Ikeo, K., Furuno, M., Zhuang, W., et al. (2016). CCT2 Mutations Evoke Leber Congenital Amaurosis due to Chaperone Complex Instability. Sci. Rep. *6*, 33742.

73. Coppieters, F., Ascari, G., Dannhausen, K., Nikopoulos, K., Peelman, F., Karlstetter, M., Xu, M., Brachet, C., Meunier, I., Tsilimbaris, M.K., et al. (2016). Isolated and Syndromic Retinal Dystrophy Caused by Biallelic Mutations in RCBTB1, a Gene Implicated in Ubiquitination. Am. J. Hum. Genet. *99*, 470–480.

74. Koboldt, D.C., Larson, D.E., Sullivan, L.S., Bowne, S.J., Steinberg, K.M., Churchill, J.D., Buhr, A.C., Nutter, N., Pierce, E.A., Blanton, S.H., et al. (2014). Exome-based mapping and variant prioritization for inherited mendelian disorders. Am. J. Hum. Genet. *94*, 373–384.

75. Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D., and Mountain, J.L. (2015). The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States. Am. J. Hum. Genet. *96*, 37–53.

76. Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.M., Doumbo, O., et al. (2009). The genetic structure and history of Africans and African Americans. Science *324*, 1035–1044.

77. Ma, X., Guan, L., Wu, W., Zhang, Y., Zheng, W., Gao, Y.T., Long, J., Wu, N., Wu, L., Xiang, Y., et al. (2015). Whole-exome sequencing identifies OR2W3 mutation as a cause of autosomal dominant retinitis pigmentosa. Sci. Rep. *5*, 9236.

78. Sharon, D., Kimchi, A., and Rivolta, C. (2016). OR2W3 sequence variants are unlikely to cause inherited retinal diseases. Ophthalmic Genet. *37*, 366–368.

79. Oswald, A.H., Goldblatt, J., Sampson, G., Clokie, R., and Beighton, P. (1985). Retinitis pigmentosa in South Africa. S. Afr. Med. J. *68*, 863–866.

80. Greenberg, J., Ramesar, R., and Beighton, P. (1994). Genetic mapping of retinitis pigmentosa--implications for South African patients. S. Afr. Med. J. *84*, 410–412.

81. Rebello, M.T., Greenberg, L.J., and Ramesar, R.S. (2002). A computer-based register for inherited retinal dystrophies in Southern Africa. Ophthalmic Genet. *23*, 61–65.

82. Ramesar, R.S., Roberts, L., Rebello, G., Goliath, R., Vorster, A., September, A., Ehrenreich, L., Gama, D., and Greenberg, J. (2003). Retinal degenerative disorders in Southern Africa: a molecular genetic approach. Adv. Exp. Med. Biol. *533*, 35–40.

83. Roberts, L., Rebello, G., Ramesar, R., and Greenberg, J. (2008). Management of a South African family with retinitis pigmentosa-should potential therapy influence translational research protocols? J. Ocul. Biol. Dis. Infor. *1*, 55–58.

84. Greenberg, J., and Roberts, L. (2012). The parent trap : Unexpected revelations that influenced the genetic management of a family with Stargardt disease. SA Ophthalmol. J. *7*, 18–21.

85. Rebello, G., Vorster, A., Greenberg, J., Coutts, N., Roberts, L., Ehrenreich, L., Gama, D., and Ramesar, R. (2003). Analysis of RPGR in a South African family with X-linked retinitis pigmentosa: research and diagnostic implications. Clin. Genet. *64*, 137–141.

86. Greenberg, J., Roberts, L., Bruwer, Z., Schoeman, M., Loggenberg, K., and Loubser, F. (2010). Delivery of an ophthalmic genetic service in South Africa. SA Ophthalmol. J. *5*, 14–19.

87. Roberts, L.J., Ramesar, R.S., and Greenberg, J. (2009). Clinical utility of the ABCR400 microarray: basing a genetic service on a commercial gene chip. Arch. Ophthalmol. *127*, 549–554.

88. Roberts, L.J., Hardie, S., Goolam Hoosen, T., Ramesar, R.S., and Greenberg, L.J. (2013). The value of genetic testing for inherited retinal disease caused by mutations in the ABCA4 gene in South Africans. S. Afr. Med. J. *103*, 702–703.

89. Roberts, L., Goliath, R., Rebello, G., Bardien, S., September, A.V, Bartmann, L., Loubser, F., Greenberg, L.J., and Ramesar, R.S. (2016). Inherited retinal disorders in South Africa and the clinical impact of evolving technologies. S. Afr. Med. J. *106*, S33–S37.

90. Tõnisson, N., Kurg, A., Kaasik, K., Lõhmussaar, E., and Metspalu, A. (2000). Unravelling genetic data by arrayed primer extension. Clin. Chem. Lab. Med. *38*, 165–170.

91. Jaakson, K., Zernant, J., Külm, M., Hutchinson, A., Tonisson, N., Glavac, D., Ravnik-Glavac, M., Hawlina, M., Meltzer, M.R., Caruso, R.C., et al. (2003). Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. Hum. Mutat. *22*, 395–403.

92. Zernant, J., Külm, M., Dharmaraj, S., den Hollander, A.I., Perrault, I., Preising, M.N., Lorenz, B., Kaplan, J., Cremers, F.P.M., Maumenee, I., et al. (2005). Genotyping microarray (disease chip) for Leber congenital amaurosis: detection of modifier alleles. Invest. Ophthalmol. Vis. Sci. *46*, 3052–3059.

93. Cremers, F.P.M., Kimberling, W.J., Külm, M., de Brouwer, A.P., van Wijk, E., te Brinke, H., Cremers, C.W.R.J., Hoefsloot, L.H., Banfi, S., Simonelli, F., et al. (2007). Development of a genotyping microarray for Usher syndrome. J. Med. Genet. *44*, 153–160.

94. Roberts, L., Ramesar, R., and Greenberg, J. (2000). Low frequency of rhodopsin mutations in South African patients with autosomal dominant retinitis pigmentosa. Clin. Genet. *58*, 77–78.

95. Roberts, L., Bartmann, L., Ramesar, R., and Greenberg, J. (2006). Novel variants in the hotspot region of RP1 in South African patients with retinitis pigmentosa. Mol. Vis. *12*, 177–183.

96. Roberts, L., Rebello, G., Greenberg, J., and Ramesar, R. (2010). Great Expectations : RPE65 Mutations in South Africa. In Retinitis Pigmentosa: Causes, Diagnosis and Treatment, M. Baert, and C. Peeters, eds. (Nova Science Publishers), pp. 89–110.

97. September, A.V, Vorster, A.A., Ramesar, R.S., and Greenberg, L.J. (2004). Mutation spectrum and founder chromosomes for the ABCA4 gene in South African patients with Stargardt disease. Invest. Ophthalmol. Vis. Sci. *45*, 1705–1711.

98. Greenberg, J., Goliath, R., Beighton, P., and Ramesar, R. (1994). A new locus for autosomal dominant retinitis pigmentosa on the short arm of chromosome 17. Hum. Mol. Genet. *3*, 915–918.

99. Goliath, R., Shugart, Y., Janssens, P., Weissenbach, J., Beighton, P., Ramasar, R., and Greenberg, J. (1995). Fine localization of the locus for autosomal dominant retinitis pigmentosa on chromosome 17p. Am. J. Hum. Genet. *57*, 962–965.

100. Bardien, S., Ebenezer, N., Greenberg, J., Inglehearn, C.F., Bartmann, L., Goliath, R., Beighton, P., Ramesar, R., and Bhattacharya, S.S. (1995). An eighth locus for autosomal dominant retinitis pigmentosa is linked to chromosome 17q. Hum. Mol. Genet. *4*, 1459–1462.

101. McKie, A.B., McHale, J.C., Keen, T.J., Tarttelin, E.E., Goliath, R., van Lith-Verhoeven, J.J., Greenberg, J., Ramesar, R.S., Hoyng, C.B., Cremers, F.P., et al. (2001). Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). Hum. Mol. Genet. *10*, 1555–1562.

102. Rebello, G., Ramesar, R., Vorster, A., Roberts, L., Ehrenreich, L., Oppon, E., Gama, D., Bardien, S., Greenberg, J., Bonapace, G., et al. (2004). Apoptosis-inducing signal sequence mutation in carbonic anhydrase IV identified in patients with the RP17 form of retinitis pigmentosa. Proc. Natl. Acad. Sci. U. S. A. *101*, 6617–6622.

103. Greenberg, J., Roberts, L., and Ramesar, R. (2002). Unusual frequencies of Rhodopsin mutations and polymorphisms in South African patients with Retinitis Pigmentosa. In New Insights into Retinal Degenerative Diseases. A Book on the Proceedings of the IXth International Symposium on Retinal Degeneration, R.E. Anderson, M.M. LaVail, and J.G. Hollyfield, eds. (Kluwer Academic/Plenum Publishers), pp. 329–333.

104. Schuster, S.C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, B., Kasson, L.R., Harris, R.S., Petersen, D.C., Zhao, F., Qi, J., et al. (2010). Complete Khoisan and Bantu genomes from southern Africa. Nature *463*, 943–947.

105. May, A., Hazelhurst, S., Li, Y., Norris, S.A., Govind, N., Tikly, M., Hon, C., Johnson, K.J., Hartmann, N., Staedtler, F., et al. (2013). Genetic diversity in black South Africans from Soweto. BMC Genomics *14*, 644–655.

106. Campbell, M.C., and Tishkoff, S.A. (2010). The Evolution of Human Genetic and Phenotypic Variation in Africa. Curr. Biol. *20*, R166–R173.

107. Sirugo, G., Hennig, B.J., Adeyemo, A.A., Matimba, A., Newport, M.J., Ibrahim, M.E., Ryckman, K.K., Tacconelli, A., Mariani-Costantini, R., Novelli, G., et al. (2008). Genetic studies of African populations: An overview on disease susceptibility and response to vaccines and therapeutics. Hum. Genet. *123*, 557–598.

108. Henn, B.M., Cavalli-Sforza, L.L., and Feldman, M.W. (2012). The great human expansion. Proc. Natl. Acad. Sci. U. S. A. *109*, 17758–17764.

109. Li, S., Schlebusch, C., and Jakobsson, M. (2014). Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. Proc. Biol. Sci. *281*, 20141448.

110. Marks, S.J., Montinaro, F., Levy, H., Brisighelli, F., Ferri, G., Bertoncini, S., Batini, C., Busby, G.B.J., Arthur, C., Mitchell, P., et al. (2015). Static and Moving Frontiers: The Genetic Landscape of Southern African Bantu-Speaking Populations. Mol. Biol. Evol. *32*, 29–43.

111. Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodriguez-Botigue, L., Ramachandran, S., Hon, L., Brisbin, A., et al. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci U S A *108*, 5154–5162.

112. The 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

113. Tishkoff, S.A., and Williams, S.M. (2002). Genetic analysis of African populations: human evolution and complex disease. Nat. Rev. Genet. *3*, 611–621.

114. Pepper, M.S. (2011). Launch of the Southern African Human Genome Programme. S. Afr. Med. J. *101*, 287–288.

115. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. Nature *517*, 327–332.

116. Chimusa, E.R., Meintjies, A., Tchanga, M., Mulder, N., Seoighe, C., Soodyall, H., and Ramesar, R. (2015). A Genomic Portrait of Haplotype Diversity and Signatures of Selection in Indigenous Southern African Populations. PLOS Genet. *11*, e1005052.

117. Ramsay, M., Tiemessen, C.T., Choudhury, A., and Soodyall, H. (2011). Africa: the next frontier for human disease gene discovery? Hum. Mol. Genet. *20*, R214–R220.

118. Fieggen, K., Milligan, C., Henderson, B., and Esterhuizen, A.I. (2016). Bardet Biedl syndrome in South Africa: A single founder mutation. S. Afr. Med. J. *106*, S72–S74.

119. Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. *16*, 1215.

120. Zeegers, M.P.A., van Poppel, F., Vlietinck, R., Spruijt, L., and Ostrer, H. (2004). Founder mutations among the Dutch. Eur. J. Hum. Genet. *12*, 591–600.

121. Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C., van Helden, P.D., Hoal, E.G., and Behar, D.M. (2010). Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. Am. J. Hum. Genet. *86*, 611–620.

122. de Wit, E., Delport, W., Rugamika, C.E., Meintjes, A., Möller, M., van Helden, P.D., Seoighe, C., and Hoal, E.G. (2010). Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. Hum. Genet. *128*, 145–153.

123. Schlebusch, C.M., and Soodyall, H. (2012). Extensive population structure in San, Khoe, and mixed ancestry populations from southern Africa revealed by 44 short 5-SNP haplotypes. Hum. Biol. *84*, 695–724.

124. Greeff, J.M. (2007). Deconstructing Jaco: genetic heritage of an Afrikaner. Ann. Hum. Genet. *71*, 674–688.

125. Botha, M.C., and Beighton, P. (1983). Inherited disorders in the Afrikaner population of southern Africa. Part I. Historical and demographic background, cardiovascular, neurological, metabolic and intestinal conditions. S. Afr. Med. J. *64*, 609–612.

126. Hayden, M.R., Hopkins, H.C., Macrae, M., and Beighton, P.H. (1980). The origin of Huntington's chorea in the Afrikaner population of South Africa. S. Afr. Med. J. *58*, 197–200.

127. Morgan, N.V., Essop, F., Demuth, I., de Ravel, T., Jansen, S., Tischkowitz, M., Lewis, C.M., Wainwright, L., Poole, J., Joenje, H., et al. (2005). A common Fanconi anemia mutation in black populations of sub-Saharan Africa. Blood *105*, 3542–3544.

128. Greenberg, J., Solomon, G.A.E., Vorster, A.A., Heckmann, J., and Bryer, A. (2006). Origin of the SCA7 gene mutation in South Africa: implications for molecular diagnostics. Clin. Genet. *70*, 415–417.

129. Auranen, M., Ylikallio, E., Toppila, J., Somer, M., Kiuru-Enari, S., and Tyynismaa, H. (2013). Dominant GDAP1 founder mutation is a common cause of axonal Charcot-Marie-Tooth disease in Finland. Neurogenetics *14*, 123–132.

130. 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.

131. Roberts, L.J., Nossek, C.A., Greenberg, L.J., and Ramesar, R.S. (2012). Stargardt macular dystrophy: common ABCA4 mutations in South Africa--establishment of a rapid genetic test and relating risk to patients. Mol. Vis. *18*, 280–289.

132. Branham, K., Othman, M., Brumm, M., Karoukis, A.J., Atmaca-Sonmez, P., Yashar, B.M., Schwartz, S.B., Stover, N.B., Trzupek, K., Wheaton, D., et al. (2012). Mutations in RPGR and RP2 account for 15% of males with simplex retinal degenerative disease. Invest. Ophthalmol. Vis. Sci. *53*, 8232–8237.

133. Daiger, S.P., Bowne, S.J., and Sullivan, L.S. (2014). Genes and Mutations Causing Autosomal Dominant Retinitis Pigmentosa. Cold Spring Harb. Perspect. Med. *5*, a017129.

134. Koenekoop, R.K., Loyer, M., Hand, C.K., Al Mahdi, H., Dembinska, O., Beneish, R., Racine, J., and Rouleau, G.A. (2003). Novel RPGR mutations with distinct retinitis pigmentosa phenotypes in French-Canadian families. Am. J. Ophthalmol. *136*, 678–687.

135. Breuer, D.K., Yashar, B.M., Filippova, E., Hiriyanna, S., Lyons, R.H., Mears, A.J., Asaye, B., Acar, C., Vervoort, R., Wright, A.F., et al. (2002). A comprehensive mutation analysis of RP2 and RPGR in a North American cohort of families with X-linked retinitis pigmentosa. Am. J. Hum. Genet. *70*, 1545–1554.

136. Wu, D.M., Khanna, H., Atmaca-Sonmez, P., Sieving, P.A., Branham, K., Othman, M., Swaroop, A., Daiger, S.P., and Heckenlively, J.R. (2010). Long-term follow-up of a family with dominant X-linked retinitis pigmentosa. Eye *24*, 764–774.

137. Churchill, J.D., Bowne, S.J., Sullivan, L.S., Lewis, R.A., Wheaton, D.K., Birch, D.G., Branham, K.E., Heckenlively, J.R., and Daiger, S.P. (2013). Mutations in the X-linked retinitis pigmentosa genes RPGR and RP2 found in 8.5% of families with a provisional diagnosis of autosomal dominant retinitis pigmentosa. Invest. Ophthalmol. Vis. Sci. *54*, 1411–1416.

138. Yokoyama, A., Maruiwa, F., Hayakawa, M., Kanai, A., Vervoort, R., Wright, A.F., Yamada, K., Niikawa, N., and Naōi, N. (2001). Three novel mutations of the RPGR gene exon ORF15 in three Japanese families with X-linked retinitis pigmentosa. Am. J. Med. Genet. *104*, 232–238.

139. Friedrich, U., Warburg, M., and Jørgensen, A.L. (1993). X-inactivation pattern in carriers of X-linked retinitis pigmentosa: a valuable means of prognostic evaluation? Hum. Genet. *92*, 359–363.

140. Grover, S., Fishman, G.A., Anderson, R.J., and Lindeman, M. (2000). A longitudinal study of visual function in carriers of X-linked recessive retinitis pigmentosa. Ophthalmology *107*, 386–396.

141. Comander, J., Weigel-DiFranco, C., Sandberg, M.A., and Berson, E.L. (2015). Visual Function in Carriers of X-Linked Retinitis Pigmentosa. Ophthalmology *122*, 1899–1906.

142. Pyo Park, S., Hwan Hong, I., Tsang, S.H., and Chang, S. (2013). Cellular imaging demonstrates genetic mosaicism in heterozygous carriers of an X-linked ciliopathy gene. Eur J Hum Genet *21*, 1240–1248.

143. Pomares, E., Riera, M., Castro-Navarro, J., Andrés-Gutiérrez, A., Gonzàlez-Duarte, R., and Marfany, G. (2009). Identification of an intronic single-point mutation in RP2 as the cause of semidominant X-linked retinitis pigmentosa. Invest. Ophthalmol. Vis. Sci. *50*, 5107–5114.

144. Banin, E., Mizrahi-Meissonnier, L., Neis, R., Silverstein, S., Magyar, I., Abeliovich, D., Roepman, R., Berger, W., Rosenberg, T., and Sharon, D. (2007). A non-ancestral RPGR missense mutation in families with either recessive or semi-dominant X-linked retinitis pigmentosa. Am. J. Med. Genet. A *143A*, 1150–1158.

145. Souied, E., Segues, B., Ghazi, I., Rozet, J.M., Chatelin, S., Gerber, S., Perrault, I., Michel-Awad, A., Briard, M.L., Plessis, G., et al. (1997). Severe manifestations in carrier females in X linked retinitis pigmentosa. J Med Genet *34*, 793–797.

146. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

147. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

148. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V, Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

149. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

150. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

151. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

152. Liu, X., Jian, X., and Boerwinkle, E. (2011). dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum. Mutat. *32*, 894–899.

153. Liu, X., Jian, X., and Boerwinkle, E. (2013). dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum. Mutat. *34*, E2393–E2402.

154. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. *24*, 2125–2137.

155. Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. *4*, 1073–1081.

156. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. *Chapter 7*, Unit 7.20.

157. Chun, S., and Fay, J.C. (2009). Identification of deleterious mutations within three human genomes. Genome Res. *19*, 1553–1561.

158. Schwarz, J.M., Rödelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat. Methods *7*, 575–576.

159. Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. *39*, e118.

160. Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L.A., Edwards, K.J., Day, I.N.M., and Gaunt, T.R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum. Mutat. *34*, 57–65.

161. Carter, H., Douville, C., Stenson, P.D., Cooper, D.N., and Karchin, R. (2013). Identifying Mendelian disease genes with the variant effect scoring tool. BMC Genomics *14 Suppl 3*, S3.

162. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

163. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. *6*, e1001025.

164. Cooper, G.M., Stone, E.A., Asimenos, G., NISC Comparative Sequencing Program, Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. *15*, 901–913.

165. Goode, D.L., Cooper, G.M., Schmutz, J., Dickson, M., Gonzales, E., Tsai, M., Karra, K., Davydov, E., Batzoglou, S., Myers, R.M., et al. (2010). Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. Genome Res. *20*, 301–310.

166. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121.

167. Garber, M., Guttman, M., Clamp, M., Zody, M.C., Friedman, N., and Xie, X. (2009). Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics *25*, i54–i62.

168. Cremers, F.P.M., den Dunnen, J.T., Ajmal, M., Hussain, A., Preising, M.N., Daiger, S.P., and Qamar, R. (2014). Comprehensive Registration of DNA Sequence Variants Associated with Inherited Retinal Diseases in Leiden Open Variation Databases. Hum. Mutat. *35*, 147–148.

169. Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2015). Ensembl 2015. Nucleic Acids Res. *43*, D662–D669.

170. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/).

171. Lek, M., Karczewski, K.J., Minikel, E.V, Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

172. Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. *41*, 95–98.

173. Wildeman, M., van Ophuizen, E., den Dunnen, J.T., and Taschner, P.E.M. (2008). Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. Hum. Mutat. *29*, 6–13.

174. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424.

175. Daiger, S.P., Sullivan, L.S., and Bowne, S.J. (2013). Genes and mutations causing retinitis pigmentosa. Clin. Genet. *84*, 132–141.

176. Lim, H., Park, Y.M., Lee, J.K., and Taek Lim, H. (2016). Single-Exome sequencing identified a novel RP2 mutation in a child with X-linked retinitis pigmentosa. Can. J. Ophthalmol. *51*, 326–330.

177. Wu, Z., Hiriyanna, S., Qian, H., Mookherjee, S., Campos, M.M., Gao, C., Fariss, R., Sieving, P.A., Li, T., Colosi, P., et al. (2015). A long-term efficacy study of gene replacement therapy for RPGR-associated retinal degeneration. Hum. Mol. Genet. *24*, 3956–3970.

178. Mears, A.J., Gieser, L., Yan, D., Chen, C., Fahrner, S., Hiriyanna, S., Fujita, R., Jacobson, S.G., Sieving, P.A., and Swaroop, A. (1999). Protein-truncation mutations in the RP2 gene in a North American cohort of families with X-linked retinitis pigmentosa. Am. J. Hum. Genet. *64*, 897–900.

179. Veltel, S., Gasper, R., Eisenacher, E., and Wittinghofer, A. (2008). The retinitis pigmentosa 2 gene product is a GTPase-activating protein for Arf-like 3. Nat. Struct. Mol. Biol. *15*, 373–380.

180. Kühnel, K., Veltel, S., Schlichting, I., and Wittinghofer, A. (2006). Crystal structure of the human retinitis pigmentosa 2 protein and its interaction with Arl3. Structure *14*, 367–378.

181. Jayasundera, T., Branham, K.E.H., Othman, M., Rhoades, W.R., Karoukis, A.J., Khanna, H., Swaroop, A., and Heckenlively, J.R. (2010). RP2 phenotype and pathogenetic correlations in X-linked retinitis pigmentosa. Arch. Ophthalmol. *128*, 915–923.

182. Vorster, A.A., Rebello, M.T., Coutts, N., Ehrenreich, L., Gama, A.D., Roberts, L.J., Goliath, R., Ramesar, R., and Greenberg, L.J. (2004). Arg120stop nonsense mutation in the RP2 gene: mutational hotspot and germ line mosaicism? Clin. Genet. *65*, 7–10.

183. Pelletier, V., Jambou, M., Delphin, N., Zinovieva, E., Stum, M., Gigarel, N., Dollfus, H., Hamel, C., Toutain, A., Dufier, J.L., et al. (2007). Comprehensive survey of mutations in RP2 and RPGR in patients affected with distinct retinal dystrophies: genotype–phenotype correlations and impact on genetic counseling. Hum. Mutat. *28*, 81–91.

184. 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.

185. Global Alliance for Genomics and Health (2016). A federated ecosystem for sharing genomic, clinical data. Science *352*, 1278–1280.

186. Martin, C.A., Ahmad, I., Klingseisen, A., Hussain, M.S., Bicknell, L.S., Leitch, A., Nürnberg, G., Toliat, M.R., Murray, J.E., Hunt, D., et al. (2014). Mutations in PLK4, encoding a master regulator of centriole biogenesis, cause microcephaly, growth failure and retinopathy. Nat. Genet. *46*, 1283–1292.

187. Banfi, S., Bassi, M.T., Andolfi, G., Marchitiello, A., Zanotta, S., Ballabio, A., Casari, G., and Franco, B. (1999). Identification and characterization of AFG3L2, a novel paraplegin-related gene. Genomics *59*, 51–58.

188. Mariotti, C., Brusco, A., Di Bella, D., Cagnoli, C., Seri, M., Gellera, C., Di Donato, S., and Taroni, F. (2008). Spinocerebellar ataxia type 28: A novel autosomal dominant cerebellar ataxia characterized by slow progression and ophthalmoparesis. Cerebellum *7*, 184–188.

189. Di Bella, D., Lazzaro, F., Brusco, A., Plumari, M., Battaglia, G., Pastore, A., Finardi, A., Cagnoli, C., Tempia, F., Frontali, M., et al. (2010). Mutations in the mitochondrial protease gene AFG3L2 cause dominant hereditary ataxia SCA28. Nat. Genet. *42*, 313–321.

190. Maltecca, F., Aghaie, A., Schroeder, D.G., Cassina, L., Taylor, B.A., Phillips, S.J., Malaguti, M., Previtali, S., Guenet, J.L., Quattrini, A., et al. (2008). The Mitochondrial Protease AFG3L2 Is Essential for Axonal Development. J. Neurosci. *28*, 2827–2836.

191. Edener, U., Wöllner, J., Hehr, U., Kohl, Z., Schilling, S., Kreuz, F., Bauer, P., Bernard, V., Gillessen-Kaesbach, G., and Zühlke, C. (2010). Early onset and slow progression of SCA28, a rare dominant ataxia in a large four-generation family with a novel AFG3L2 mutation. Eur. J. Hum. Genet. *18*, 965–968.

192. Cagnoli, C., Stevanin, G., Brussino, A., Barberis, M., Mancini, C., Margolis, R.L., Holmes, S.E., Nobili, M., Forlani, S., Padovan, S., et al. (2010). Missense mutations in the AFG3L2 proteolytic domain account for ~1.5% of European autosomal dominant cerebellar ataxias. Hum. Mutat. *31*, 1117–1124.

193. Pierson, T.M., Adams, D., Bonn, F., Martinelli, P., Cherukuri, P.F., Teer, J.K., Hansen, N.F., Cruz, P., Mullikin, J.C., Blakesley, R.W., et al. (2011). Whole-exome sequencing identifies homozygous AFG3L2 mutations in a spastic ataxia-neuropathy syndrome linked to mitochondrial m-AAA proteases. PLoS Genet. *7*, e1002325.

194. Muona, M., Berkovic, S.F., Dibbens, L.M., Oliver, K.L., Maljevic, S., Bayly, M.A., Joensuu, T., Canafoglia, L., Franceschetti, S., Michelucci, R., et al. (2015). A recurrent de novo mutation in KCNC1 causes progressive myoclonus epilepsy. Nat. Genet. *47*, 39–46.

195. Bahn, R.S. (2010). Graves' ophthalmopathy. N. Engl. J. Med. *362*, 726–738.

196. Zhao, S.X., Xue, L.Q., Liu, W., Gu, Z.H., Pan, C., Yang, S.Y., Zhan, M., Wang, H.N., Liang, J., Gao, G.Q., et al. (2013). Robust evidence for five new Graves' disease risk loci from a staged genome-wide association analysis. Hum. Mol. Genet. *22*, 3347–3362.

197. Vinayagam, A., Stelzl, U., Foulle, R., Plassmann, S., Zenkner, M., Timm, J., Assmus, H.E., Andrade-Navarro, M.A., and Wanker, E.E. (2011). A directed protein interaction network for investigating intracellular signal transduction. Sci. Signal. *4*, rs8.

198. Niroula, A., Urolagin, S., and Vihinen, M. (2015). PON-P2: Prediction method for fast and reliable identification of harmful variants. PLoS One *10*, e0117380.

199. Kim, Y.O., Oh, I.U., Park, H.S., Jeng, J., Song, B.J., and Huh, T.L. (1995). Characterization of a cDNA clone for human NAD(+)-specific isocitrate dehydrogenase alpha-subunit and structural comparison with its isoenzymes from different species. Biochem. J. *308*, 63–68.

200. Huh, T.L., Kim, Y.O., Oh, I.U., Song, B.J., and Inazawa, J. (1996). Assignment of the human mitochondrial NAD+ -specific isocitrate dehydrogenase alpha subunit (IDH3A) gene to 15q25.1-->q25.2 by in situ hybridization. Genomics *32*, 295–296.

201. Hartong, D.T., Dange, M., McGee, T.L., Berson, E.L., Dryja, T.P., and Colman, R.F. (2008). Insights from retinitis pigmentosa into the roles of isocitrate dehydrogenases in the Krebs cycle. Nat. Genet. *40*, 1230–1234.

202. Panfoli, I., Calzia, D., Ravera, S., Bruschi, M., Tacchetti, C., Candiani, S., Morelli, A., and Candiano, G. (2011). Extramitochondrial tricarboxylic acid cycle in retinal rod outer segments. Biochimie *93*, 1565–1575.

203. Panfoli, I., Musante, L., Bachi, A., Ravera, S., Calzia, D., Cattaneo, A., Bruschi, M., Bianchini, P., Diaspro, A., Morelli, A., et al. (2008). Proteomic analysis of the retinal rod outer segment disks. J. Proteome Res. *7*, 2654–2669.

204. Porter, L.F., and Black, G.C.M. (2014). Personalized ophthalmology. Clin. Genet. *86*, 1–11.

205. Adams, D.R., Sincan, M., Fuentes Fajardo, K., Mullikin, J.C., Pierson, T.M., Toro, C., Boerkoel, C.F., Tifft, C.J., Gahl, W.A., and Markello, T.C. (2012). Analysis of DNA sequence variants detected by high-throughput sequencing. Hum. Mutat. *33*, 599–608.

206. Gebhardt, S., Bruiners, N., and Hillermann, R. (2009). A novel exonic variant (221delT) in the LGALS13 gene encoding placental protein 13 (PP13) is associated with preterm labour in a low risk population. J. Reprod. Immunol. *82*, 166–173.

207. Meindl, A., Dry, K., Herrmann, K., Manson, E., Ciccodicola, A., Edgar, A., Carvalho, M.R.S., Achatz, H., Hellebrand, H., Lennon, A., et al. (1996). A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X–linked retinitis pigmentosa (RP3). Nat. Genet. *13*, 35–42.

208. Roepman, R., van Duijnhoven, G., Rosenberg, T., Pinckers, A.J.L.G., Bleeker-Wagemakers, L.M., Bergen, A.A.B., Post, J., Beck, A., Reinhardt, R., Ropers, H.H., et al. (1996). Positional cloning of the gene for X-linked retinitis pigmentosa 3: homology with the guanine-nucleotide-exchange factor RCC1. Hum. Mol. Genet. *5*, 1035–1041.

209. Vervoort, R., Lennon, A., Bird, A.C., Tulloch, B., Axton, R., Miano, M.G., Meindl, A., Meitinger, T., Ciccodicola, A., and Wright, A.F. (2000). Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. Nat. Genet. *25*, 462–466.

210. Huang, X.F., Wu, J., Lv, J.N., Zhang, X., and Jin, Z.B. (2015). Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: a complementary approach to clinical genetic diagnostic testing. Genet. Med. *17*, 307–311.

211. Shu, X., Black, G.C., Rice, J.M., Hart-Holden, N., Jones, A., O'Grady, A., Ramsden, S., and Wright, A.F. (2007). RPGR mutation analysis and disease: an update. Hum. Mutat. *28*, 322–328.

212. García-Hoyos, M., Garcia-Sandoval, B., Cantalapiedra, D., Riveiro, R., Lorda-Sánchez, I., Trujillo-Tiebas, M.J., Rodriguez de Alba, M., Millan, J.M., Baiget, M., Ramos, C., et al. (2006). Mutational screening of the RP2 and RPGR genes in Spanish families with X-linked retinitis pigmentosa. Invest. Ophthalmol. Vis. Sci. *47*, 3777–3782.

213. MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., et al. (2014). Guidelines for investigating causality of sequence variants in human disease. Nature *508*, 469–476.

214. Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: An integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. *33*, W741–W748.

215. Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res. *41*, W77–W83.

216. Boloc, D., Castillo-Lara, S., Marfany, G., Gonzàlez-Duarte, R., and Abril, J.F. (2015). Distilling a Visual Network of Retinitis Pigmentosa Gene-Protein Interactions to Uncover New Disease Candidates. PLoS One *10*, e0135307.

217. Facio, F.M., Lee, K., and O'Daniel, J.M. (2014). A genetic counselor's guide to using next-generation sequencing in clinical practice. J. Genet. Couns. *23*, 455–462.

218. 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. (2016). Improved Diagnosis of Inherited Retinal Dystrophies by High-Fidelity PCR of ORF15 followed by Next-Generation Sequencing. J. Mol. Diagn. *18*, 817–824.

219. Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. *37*, W305–W311.

220. Tranchevent, L.C., Barriot, R., Yu, S., Van Vooren, S., Van Loo, P., Coessens, B., De Moor, B., Aerts, S., and Moreau, Y. (2008). ENDEAVOUR update: a web resource for gene prioritization in multiple species. Nucleic Acids Res. *36*, W377–W384.

221. Beck, T.F., Mullikin, J.C., and Biesecker, L.G. (2016). Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. Clin. Chem. *62*, 647–654.

222. Le Quesne Stabej, P., Saihan, Z., Rangesh, N., Steele-Stallard, H., Ambrose, J., Coffey, A., Emmerson, J., Haralambous, E., Hughes, Y., Steel, K.P., et al. (2012). Comprehensive sequence analysis of nine Usher syndrome genes in the UK National Collaborative Usher Study. J. Med. Genet. *49*, 27–36.

223. Danciger, M., Blaney, J., Gao, Y.Q., Zhao, D.Y., Heckenlively, J.R., Jacobson, S.G., and Farber, D.B. (1995). Mutations in the PDE6B gene in autosomal recessive retinitis pigmentosa. Genomics *30*, 1–7.

224. Shanks, M.E., Downes, S.M., Copley, R.R., Lise, S., Broxholme, J., Hudspith, K.A.Z., Kwasniewska, A., Davies, W.I.L., 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.

225. Ge, Z., Bowles, K., Goetz, K., Scholl, H.P.N., 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.

226. Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., Kure, B., Wata Mpoloka, S., Nakagawa, H., Naumann, C., et al. (2012). The genetic prehistory of southern Africa. Nat. Commun. *3*, 1143.

227. González-Santos, M., Montinaro, F., Oosthuizen, O., Oosthuizen, E., Busby, G.B.J., Anagnostou, P., Destro-Bisol, G., Pascali, V., and Capelli, C. (2015). Genome-Wide SNP Analysis of Southern African Populations Provides New Insights into the Dispersal of Bantu-Speaking Groups. Genome Biol. Evol. *7*, 2560–2568.

228. Henn, B.M., Botigué, L.R., Peischl, S., Dupanloup, I., Lipatov, M., Maples, B.K., Martin, A.R., Musharoff, S., Cann, H., Snyder, M.P., et al. (2016). Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc. Natl. Acad. Sci. U. S. A. *113*, E440–E449.

229. Kidd, J.M., Sharpton, T.J., Bobo, D., Norman, P.J., Martin, A.R., Carpenter, M.L., Sikora, M., Gignoux, C.R., Nemat-Gorgani, N., Adams, A., et al. (2014). Exome capture from saliva produces high quality genomic and metagenomic data. BMC Genomics *15*, 262–278.

230. Lachance, J., Vernot, B., Elbers, C.C., Ferwerda, B., Froment, A., Bodo, J.M., Lema, G., Fu, W., Nyambo, T.B., Rebbeck, T.R., et al. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell *150*, 457–469.

231. Brezina, P.R., and Kutteh, W.H. (2014). Clinical applications of preimplantation genetic testing. BMJ *349*, g7611.

232. Boye, S.E., Boye, S.L., Lewin, A.S., and Hauswirth, W.W. (2013). A comprehensive review of retinal gene therapy. Mol. Ther. *21*, 509–519.

233. Zallocchi, M., Binley, K., Lad, Y., Ellis, S., Widdowson, P., Iqball, S., Scripps, V., Kelleher, M., Loader, J., Miskin, J., et al. (2014). EIAV-based retinal gene therapy in the shaker1 mouse model for usher syndrome type 1B: Development of UshStat. PLoS One *9*, e94272.

234. Kong, J., Kim, S.R., Binley, K., Pata, I., Doi, K., Mannik, J., Zernant-Rajang, J., Kan, O., Iqball, S., Naylor, S., et al. (2008). Correction of the disease phenotype in the mouse model of Stargardt disease by lentiviral gene therapy. Gene Ther. *15*, 1311–1320.

235. Beltran, W.A., Cideciyan, A.V, Lewin, A.S., Hauswirth, W.W., Jacobson, S.G., and Aguirre, G.D. (2014). Gene augmentation for X-linked retinitis pigmentosa caused by mutations in RPGR. Cold Spring Harb. Perspect. Med. *5*, a017392.

236. Palfi, A., Millington-Ward, S., Chadderton, N., O'Reilly, M., Goldmann, T., Humphries, M.M., Li, T., Wolfrum, U., Humphries, P., Kenna, P.F., et al. (2010). Adeno-associated virus-mediated rhodopsin replacement provides therapeutic benefit in mice with a targeted disruption of the rhodopsin gene. Hum. Gene Ther. *21*, 311–323.

237. Chadderton, N., Millington-Ward, S., Palfi, A., O'Reilly, M., Tuohy, G., Humphries, M.M., Li, T., Humphries, P., Kenna, P.F., and Farrar, G.J. (2009). Improved retinal function in a mouse model of dominant retinitis pigmentosa following AAV-delivered gene therapy. Mol. Ther. *17*, 593–599.

238. Schwartz, S.D., Hubschman, J.P., Heilwell, G., Franco-Cardenas, V., Pan, C.K., Ostrick, R.M., Mickunas, E., Gay, R., Klimanskaya, I., and Lanza, R. (2012). Embryonic stem cell trials for macular degeneration: a preliminary report. Lancet *379*, 713–720.

239. Doroudchi, M.M., Greenberg, K.P., Zorzos, A.N., Hauswirth, W.W., Fonstad, C.G.,

Horsager, A., and Boyden, E.S. (2011). Towards optogenetic sensory replacement. In 2011 Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, (IEEE), pp. 3139–3141.

240. Manes, G., Cheguru, P., Majumder, A., Bocquet, B., Sénéchal, A., Artemyev, N.O., Hamel, C.P., and Brabet, P. (2014). A truncated form of rod photoreceptor PDE6 β-subunit causes autosomal dominant congenital stationary night blindness by interfering with the inhibitory activity of the γ-subunit. PLoS One *9*, e95768.

241. Gao, Y.Q., Danciger, M., Zhao, D.Y., Blaney, J., Piriev, N.I., Shih, J., Jacobson, S.G., Heckenlively, J.H., and Farber, D.B. (1996). Screening of the PDE6B gene in patients with autosomal dominant retinitis pigmentosa. Exp. Eye Res. *62*, 149–154.

242. Fattal-Valevski, A., Eliyahu, H., Fraenkel, N.D., Elmaliach, G., Hausman-Kedem, M., Shaag, A., Mandel, D., Pines, O., and Elpeleg, O. (2017). Homozygous mutation, p.Pro304His, in IDH3A, encoding isocitrate dehydrogenase subunit is associated with severe encephalopathy in infancy. Neurogenetics *18*, 57–61.

243. van Huet, R.A.C., Pierrache, L.H.M., Meester-Smoor, M.A., Klaver, C.C.W., van den Born, L.I., Hoyng, C.B., de Wijs, I.J., Collin, R.W.J., Hoefsloot, L.H., and Klevering, B.J. (2015). The efficacy of microarray screening for autosomal recessive retinitis pigmentosa in routine clinical practice. Mol. Vis. *21*, 461–476.

244. Souzeau, E., Burdon, K.P., Mackey, D.A., Hewitt, A.W., Savarirayan, R., Otlowski, M., and Craig, J.E. (2016). Ethical Considerations for the Return of Incidental Findings in Ophthalmic Genomic Research. Trans. Vis. Sci. Tech. *5*, 3.

245. Wang, X., Feng, Y., Li, J., Zhang, W., Wang, J., Lewis, R.A., and Wong, L.J. (2016). Retinal diseases caused by mutations in genes not specifically associated with the clinical diagnosis. PLoS One *11*, e0165405.

246. den Hollander, A.I., Koenekoop, 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. (2006). Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. Am. J. Hum. Genet. *79*, 556–561.

247. Coppieters, F., Lefever, S., Leroy, B.P., and De Baere, E. (2010). CEP290, a gene with many faces: Mutation overview and presentation of CEP290base. Hum. Mutat. *31*, 1097–1108.

248. Garanto, A., Duijkers, L., and Collin, R.W.J. (2015). Species-dependent splice recognition of a cryptic exon resulting from a recurrent intronic CEP290 mutation that causes congenital blindness. Int. J. Mol. Sci. *16*, 5285–5298.

249. Beryozkin, A., Levy, G., Blumenfeld, A., Meyer, S., Namburi, P., Morad, Y., Gradstein, L., Swaroop, A., Banin, E., and Sharon, D. (2016). Genetic Analysis of the Rhodopsin Gene Identifies a Mosaic Dominant Retinitis Pigmentosa Mutation in a Healthy Individual. Invest. Ophthalmol. Vis. Sci. *57*, 940–947.

250. Bujakowska, K.M., White, J., Place, E., Consugar, M., and Comander, J. (2015). Efficient in silico identification of a common insertion in the MAK gene which causes retinitis pigmentosa. PLoS One *10*, e0142614.

251. Khateb, S., Hanany, M., Khalaileh, A., Beryozkin, A., Meyer, S., Abu-Diab, A., Abu Turky, F., Mizrahi-Meissonnier, L., Lieberman, S., Ben-Yosef, T., 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.

252. 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. doi: 10.1038/gim.2016.158.

253. Crowgey, E.L., Stabley, D.L., Chen, C., Huang, H., Robbins, K.M., Polson, S.W., Sol-Church, K., and Wu, C.H. (2015). An integrated approach for analyzing clinical genomic variant data from next-generation sequencing. J. Biomol. Tech. *26*, 19–28.

254. Liu, Y.P., Bosch, D.G.M., Siemiatkowska, A.M., Rendtorff, N.D., Boonstra, F.N., Möller, C., Tranebjærg, L., Katsanis, N., and Cremers, F.P.M. (2016). Putative digenic inheritance of heterozygous RP1L1 and C2orf71 null mutations in syndromic retinal dystrophy. Ophthalmic Genet. doi: 10.3109/13816810.2016.1151898.

255. Estrada-Cuzcano, A., Roepman, R., Cremers, F.P.M., den Hollander, A.I., and Mans, D.A. (2012). Non-syndromic retinal ciliopathies: translating gene discovery into therapy. Hum. Mol. Genet. *21*, R111–R124.

256. Fromer, M., Moran, J.L., Chambert, K., Banks, E., Bergen, S.E., Ruderfer, D.M., Handsaker, R.E., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., et al. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am. J. Hum. Genet. *91*, 597–607.

257. 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. (2016). 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. doi: 10.1038/gim.2016.119.

Appendices

Appendix 1: Permission to include publications in thesis

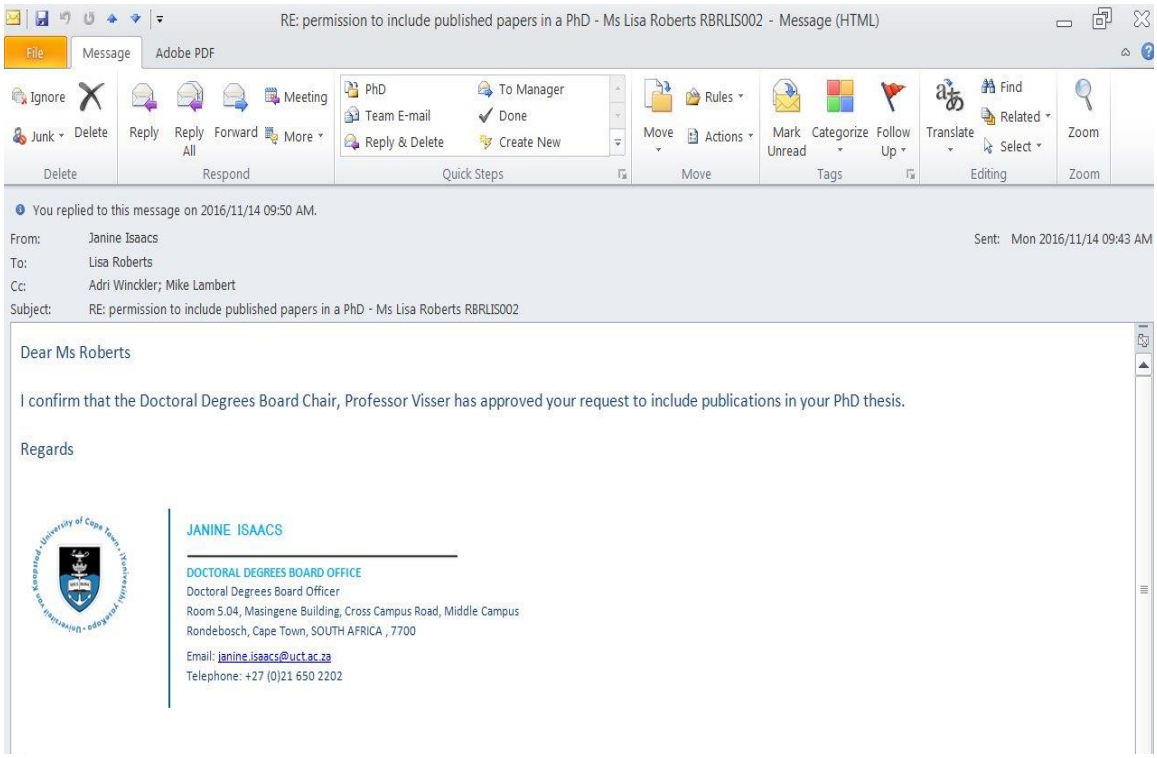
Appendix 2: Molecular request form and confirmation of diagnosis form

Appendix 3: Screenshots of the UCT Human Genetics IRD database

Appendix 4: Flow diagram summarising WES analysis

Appendix 5: Pedigrees of four unresolved families

Appendix 1. Permission to include publications in thesis



RE: permission to include published papers in a PhD - Ms Lisa Roberts RBRLIS002 - Message (HTML)

File | Message | Adobe PDF

Ignore | Junk | Delete | Reply | Reply All | Forward | Meeting | More | PhD | Team E-mail | Reply & Delete | To Manager | Done | Create New | Move | Rules | Actions | Mark Unread | Categorize | Follow Up | Translate | Find | Related | Select | Zoom

Delete | Respond | Quick Steps | Move | Tags | Editing | Zoom

You replied to this message on 2016/11/14 09:50 AM.

From: Janine Isaacs
Sent: Mon 2016/11/14 09:43 AM
To: Lisa Roberts
Cc: Adri Winckler; Mike Lambert
Subject: RE: permission to include published papers in a PhD - Ms Lisa Roberts RBRLIS002

Dear Ms Roberts

I confirm that the Doctoral Degrees Board Chair, Professor Visser has approved your request to include publications in your PhD thesis.

Regards

JANINE ISAACS

DOCTORAL DEGREES BOARD OFFICE
Doctoral Degrees Board Officer
Room 5.04, Masingene Building, Cross Campus Road, Middle Campus
Rondebosch, Cape Town, SOUTH AFRICA , 7700
Email: janine.isaacs@uct.ac.za
Telephone: +27 (0)21 650 2202

Appendix 2. Molecular request form and confirmation of diagnosis form

REQUEST FOR MOLECULAR STUDIES FORM

GENETICS OF RETINAL DEGENERATIVE DISORDERS

DIVISION OF HUMAN GENETICS, WERNHER & BEIT NORTH
FACULTY OF HEALTH SCIENCES, UNIVERSITY OF CAPE TOWN, OBSERVATORY, 7925
TEL: 021 406-6299 FAX: 021 650-2010 EMAIL: jacquie.greenberg@uct.ac.za

PATIENT DETAILS

SURNAME: _____ **NAME:** _____

DATE OF BIRTH: ____/ ____/_____ **SEX: FEMALE - ☐ MALE - ☐**

ETHNICITY:_____

NEW FAMILY: YES - ☐ NO - ☐ (If NO please fill in family name)

FAMILY NAME:_____

Number of Children: _____ **Number of affected family members:** _____

CONTACT ADDRESS:_____

_____**CODE** _____

TEL: …….. FAX: …… E-mail ……

REFERRAL SOURCE

PLEASE NOTE: A confirmation of diagnosis (COD) form is required to accompany all samples. This separate form needs to be completed by an Ophthalmologist and can be faxed separately.

NAME OF REFERRING DOCTOR: _____ REFERRING FACILITY:_____

FAX: _____ TEL:_____ E-mail: _____

ADDRESS: _____

REASON FOR REFERRAL (CLINICAL DIAGNOSIS)

AFFECTED - ☐ AT RISK - ☐ CARRIER - ☐ SPOUSE - ☐ UNAFFECTED - ☐

RETINITIS PIGMENTOSA ☐	USHER SYNDROME ☐	**DOMINANT INHERITANCE ☐**
STARGARDT DISEASE ☐	MACULAR DYSTROPHY ☐	**RECESSIVE INHERITANCE ☐**
ARMD – WET ☐	ARMD – DRY ☐	**X-LINKED INHERITANCE ☐**
OTHER DISORDER:	**AGE OF ONSET:**	**ISOLATED CASE ☐**
	DIAGNOSIS AGE:	

FAMILY HISTORY INFORMATION

ADDITIONAL FAMILY HISTORY _____

ADDITIONAL DISORDERS (APPARENT OR PREVIOUSLY TREATED):
_____

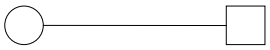
RELEVANT CLINICAL DETAILS: _____

PHYSICAL DISABILITY-☐ INTELLECTUAL DISABILITY- ☐ DEAFNESS - ☐ IMPAIRED VISION - ☐
NIGHT BLINDNESS - ☐ AGE OF ONSET: _____ OTHER: _____

PEDIGREE/ FAMILY TREE (If more extensive, please use a separate page)

Maternal Ethnicity/ Genetic Origin:_____
Paternal Ethnicity/ Genetic Origin:_____

Grandparents

Parents/Siblings 

Proband (Patient)/
Siblings

Children

Have samples from this patient been sent to a DNA lab before? Yes - ☐ No - ☐ Don't Know - ☐.

If "Yes": Where: _____

When:_____

Have samples from other family members previously been sent to a DNA lab for genetic ophthalmic disease testing? Yes - ☐ No - ☐ Unsure - ☐.

If "Yes": Name of lab:_____Mutation identified:_____

RETINA SA MEMBERSHIP

This is not a requirement, but is necessary information to assist us with our database.

Are you currently a Retina SA member? Yes - ☐ No - ☐

If "No", have you ever been a Retina SA member? Yes - ☐ No - ☐

RETINA SOUTH AFRICA

TEST REQUESTED

ABCA4 Quick 7 - ☐ "Asper Chip" - ☐ Specify_____

Known family mutation - ☐ Specify:_____ Other - ☐ Specify:_____

For Laboratory use only: DNA number: _____ Vol.Blood:_____ (ml)

Other: _____ Date Received: Year: _____ Month:____ Day:____ Computer Index No: _____

SPECIMEN TUBES REQUIRED: 2x 4ml Plastic purple top tubes (containing EDTA)

NAME AND DATE OF BIRTH TO BE ON EACH TUBE
BLOODS ARE TO BE KEPT REFRIGERATED. SPECIMENS ARE TO BE CAREFULLY PACKAGED AND TRANSPORTED IN A POLYSTYRENE COOLBOX WITH AN ICE BRICK. DO NOT FREEZE.
BLOODS ARE CODED ON ARRIVAL IN THE LABORATORY ACCORDING TO FAMILY NAME
WHEN AVAILABLE, RESULTS ARE GIVEN TO PARTICIPANTS ACCORDING TO AN ESTABLISHED PROTOCOL

CONSENT FORM REQUIRED FOR DNA ANALYSIS AND STORAGE
GENETICS OF RETINAL DEGENERATIVE DISORDERS

PLEASE DELETE DETAILS THAT ARE NOT APPLICABLE:

1. I, _____, request that an attempt be made using genetic material to assess the probability that, I / my child / my unborn child, might have inherited a disease-causing mutation in the gene for: ***(Name of Disorder):***

2. I understand that the genetic material for analysis is to be obtained from:
blood cells / skin sample / other (specify)

3. I request that no portion of the sample be stored for later use. ☐ *(MARK IF APPLICABLE)*

OR I request that a portion of the sample be stored indefinitely for:*(DELETE WHERE NOT APPLICABLE)*
(a) Possible re-analysis
(b) Analysis for the benefit of members of my immediate family
(c) Research purposes, subject to the approval of the University of Cape Town Human Research Ethics Committee, provided that any information from such research will remain confidential

4. Regarding the results of any current and future analysis carried out on this sample of stored biological material: only meaningful results that have clear diagnostic implications will be made known to me, via my doctor, in accordance with the relevant protocol, if and when available. In addition, I authorise that these results may be made known to: *(DELETE WHERE NOT APPLICABLE)*

Doctor/Family member/Friend: _____

5. I authorise / do not authorise my doctor(s) to provide relevant clinical details to the Division of Human Genetics, UCT.

6. I have been informed that:

- There are risks and benefits associated with genetic analysis and storage of biological material and these have been explained to me.
- The analysis procedure is specific to the genetic condition related to the visual impairment mentioned above and cannot determine the complete genetic makeup of an individual.
- The genetics laboratory is under an obligation to respect medical confidentiality.
- Although samples are stored in the Division of Human Genetics, collaborations with international facilities may require the samples and clinical information to be shared. In these cases, all samples will be de-identified and given a unique code, and full patient confidentiality will be maintained.
- Genetic analysis may not be informative for some families or family members.
- Even under the best conditions, current technology of this type is not perfect and could lead to incorrect results.
  Where biological material is used for research purposes, there may be no direct benefit to me.

7. I understand that I may withdraw my consent for any aspect of the above at any time without this affecting my future medical care.

ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME AND ALL THE QUESTIONS I HAD WERE ANSWERED BY: _____IN A LANGUAGE THAT I UNDERSTAND.

DATE _____ /_____ /____ **PLACE:** _____

PATIENT'S SIGNATURE: _____ **WITNESS CONSENT:** _____

Tel: +27 21 406 6297
Fax: +27 21 650 2010
Email: jacquie.greenberg@uct.ac.za

Division of Human Genetics
Level 3, Wernher and Beit North
Institute of Infectious Disease and
Molecular Medicine
Faculty of Health Sciences
University of Cape Town

CONFIRMATION OF DIAGNOSIS

To be completed by the eye specialist – PLEASE PRINT CLEARLY

Name of Patient:

Date of Birth: Y Y Y Y - M M - D D **Tel/Cell** **Fax** ( )

Address:

Gender: M F **Ethnic Group** Asian Black Coloured Indian White

In my opinion the patient has one of the following conditions:

RETINITIS PIGMENTOSA	Diffuse Form	
	Sectoral (regional) form	
USHER SYNDROME ***(RP & congenital hearing loss)***	Type I – profound deafness	
	Type II – severe deafness	
MACULAR DEGENERATION	Age-related MD - Wet	
	- Dry	
	Best Disease	
	Cone & Rod Dystrophy	
	Sorsby Fundus Dystrophy	
	Pattern Dystrophy	
	Stargardt Disease	
	Fundus Flavimaculatus	

Other retinal disorder (specify):

MODE OF INHERITANCE

Dominant **Recessive** **X-Linked** **Isolated Case** **Unknown but familial**

Age of onset: **years**

Progression of disease:

Other clinical features:

Tests performed: **ERG** **Visual Acuity** **Visual Fields** **DNA** **Fluorescein Angiogram** **Colour Fundus Photographs**

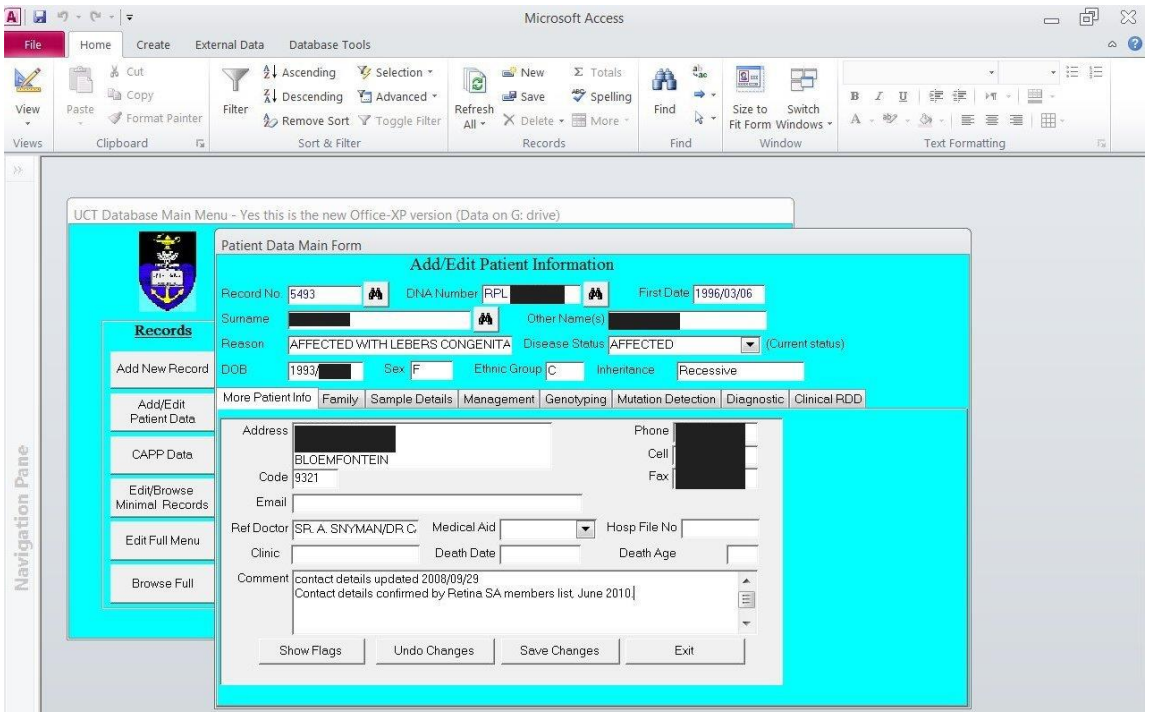
Other family members affected:

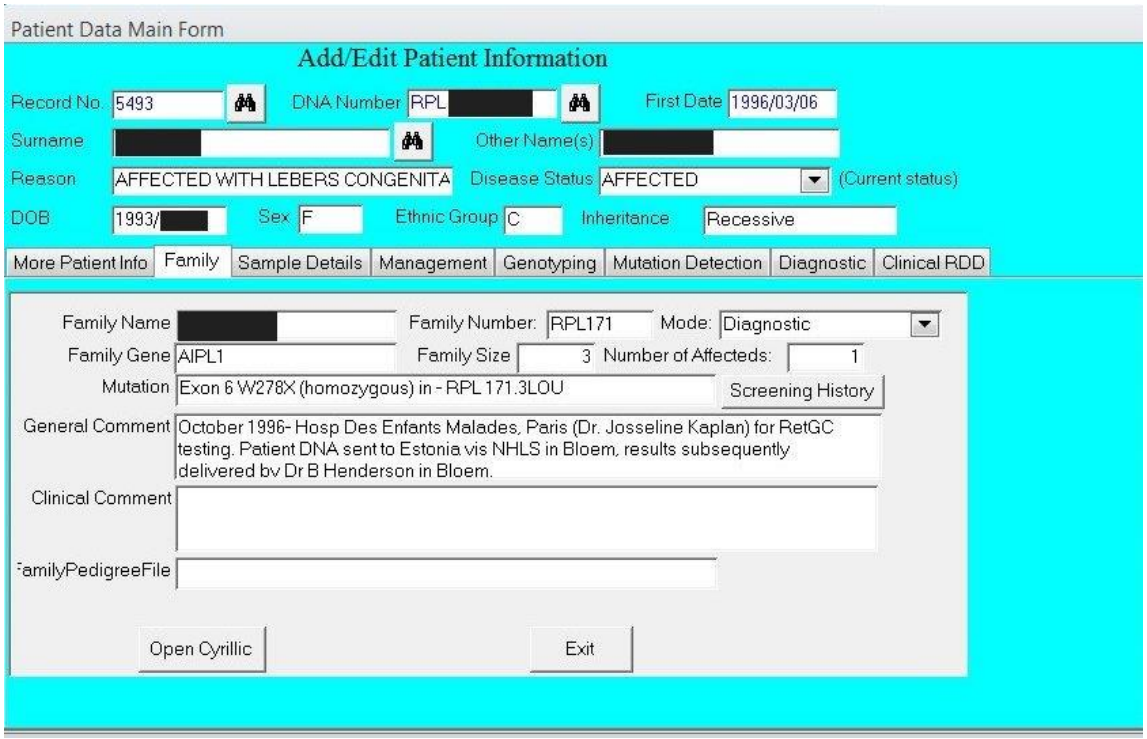
Name of Doctor: **Signature:**

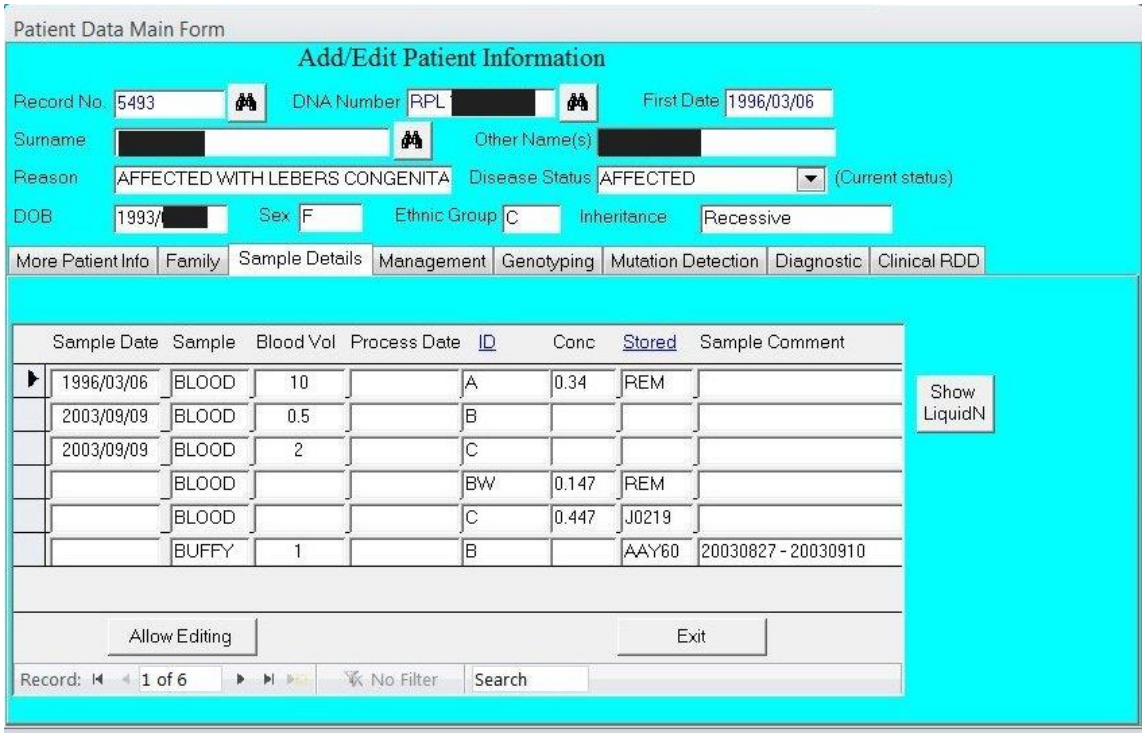
Date: Y Y Y Y – M M – D D **Tel.** ( ) **Fax** ( )

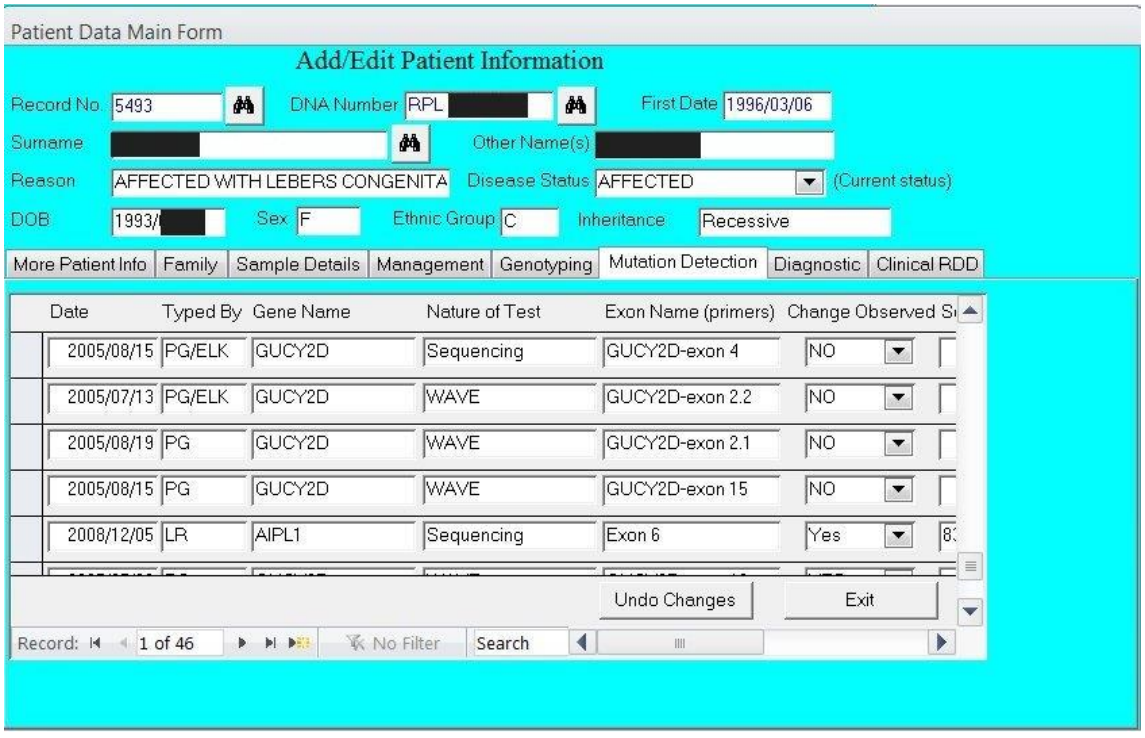
PLEASE RETURN THE COMPLETED FORM TO:
Prof Jacquie Greenberg (jacquie.greenberg@uct.ac.za / fax: 021 650-2010)

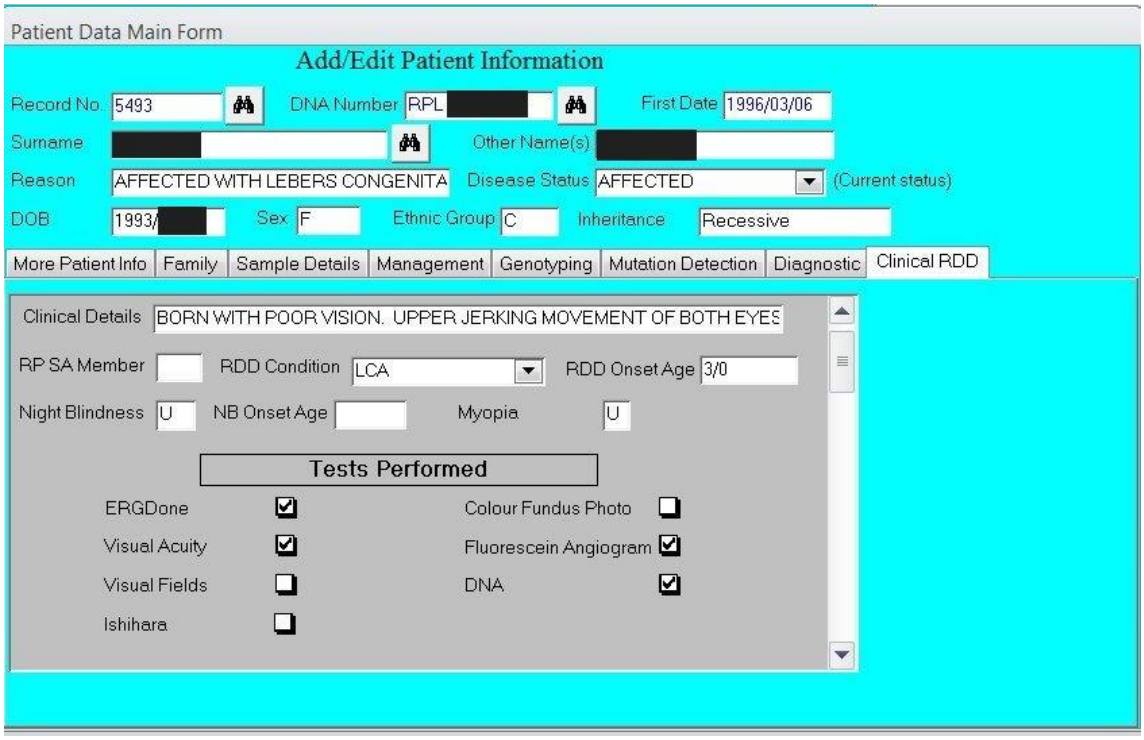
Appendix 3. Screenshots of the UCT Human Genetics IRD database






Patient Data Main Form
Add/Edit Patient Information
Record No 5493
DNA Number RPL [redacted]
First Date 1996/03/06
Surname [redacted]
Other Name(s) [redacted]
Reason AFFECTED WITH LEBERS CONGENITA
Disease Status AFFECTED
(Current status)
DOB 1993/[redacted]
Sex F
Ethnic Group C
Inheritance Recessive
More Patient Info
Family
Sample Details
Management
Genotyping
Mutation Detection
Diagnostic
Clinical RDD
Sample Date | Sample | Blood Vol | Process Date | ID | Conc | Stored | Sample Comment
1996/03/06 | BLOOD | 10 | | A | 0.34 | REM |
2003/09/09 | BLOOD | 0.5 | | B | | |
2003/09/09 | BLOOD | 2 | | C | | |
| BLOOD | | | BW | 0.147 | REM |
| BLOOD | | | C | 0.447 | J0219 |
| BUFFY | 1 | | B | | AAY60 | 20030827 - 20030910
Show LiquidN
Allow Editing
Exit
Record: 1 of 6
No Filter
Search

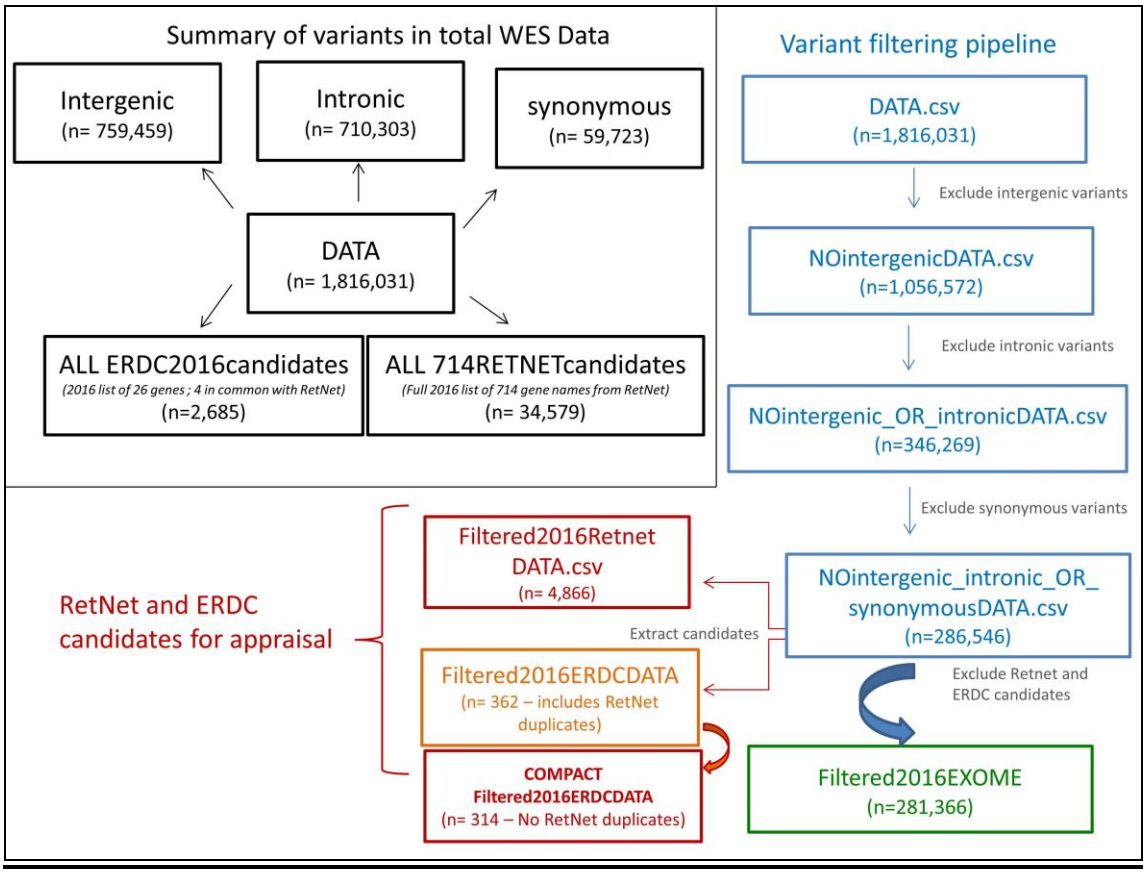

Patient Data Main Form
Add/Edit Patient Information
Record No 5493
DNA Number RPL [redacted]
First Date 1996/03/06
Surname [redacted]
Other Name(s) [redacted]
Reason AFFECTED WITH LEBERS CONGENITA
Disease Status AFFECTED
(Current status)
DOB 1993/[redacted]
Sex F
Ethnic Group C
Inheritance Recessive
More Patient Info
Family
Sample Details
Management
Genotyping
Mutation Detection
Diagnostic
Clinical RDD
Date | Typed By | Gene Name | Nature of Test | Exon Name (primers) | Change Observed
2005/08/15 | PG/ELK | GUCY2D | Sequencing | GUCY2D-exon 4 | NO
2005/07/13 | PG/ELK | GUCY2D | WAVE | GUCY2D-exon 2.2 | NO
2005/08/19 | PG | GUCY2D | WAVE | GUCY2D-exon 2.1 | NO
2005/08/15 | PG | GUCY2D | WAVE | GUCY2D-exon 15 | NO
2008/12/05 | LR | AIPL1 | Sequencing | Exon 6 | Yes
Undo Changes
Exit
Record: 1 of 46
No Filter
Search


Patient Data Main Form
Add/Edit Patient Information
Record No 5493
DNA Number RPL
First Date 1996/03/06
Surname
Other Name(s)
Reason AFFECTED WITH LEBERS CONGENITA
Disease Status AFFECTED
(Current status)
DOB 1993/
Sex F
Ethnic Group C
Inheritance Recessive
More Patient Info
Family
Sample Details
Management
Genotyping
Mutation Detection
Diagnostic
Clinical RDD
Clinical Details BORN WITH POOR VISION. UPPER JERKING MOVEMENT OF BOTH EYES
RP SA Member
RDD Condition LCA
RDD Onset Age 3/0
Night Blindness U
NB Onset Age
Myopia U
Tests Performed
ERGDone
Visual Acuity
Visual Fields
Ishihara
Colour Fundus Photo
Fluorescein Angiogram
DNA

Appendix 4. Flow diagram summarising WES analysis

The summaries of variants (black) were obtained by using the 'grep' or 'file grep' commands, using keywords or gene lists respectively, to extract particular variants from the DATA.csv file

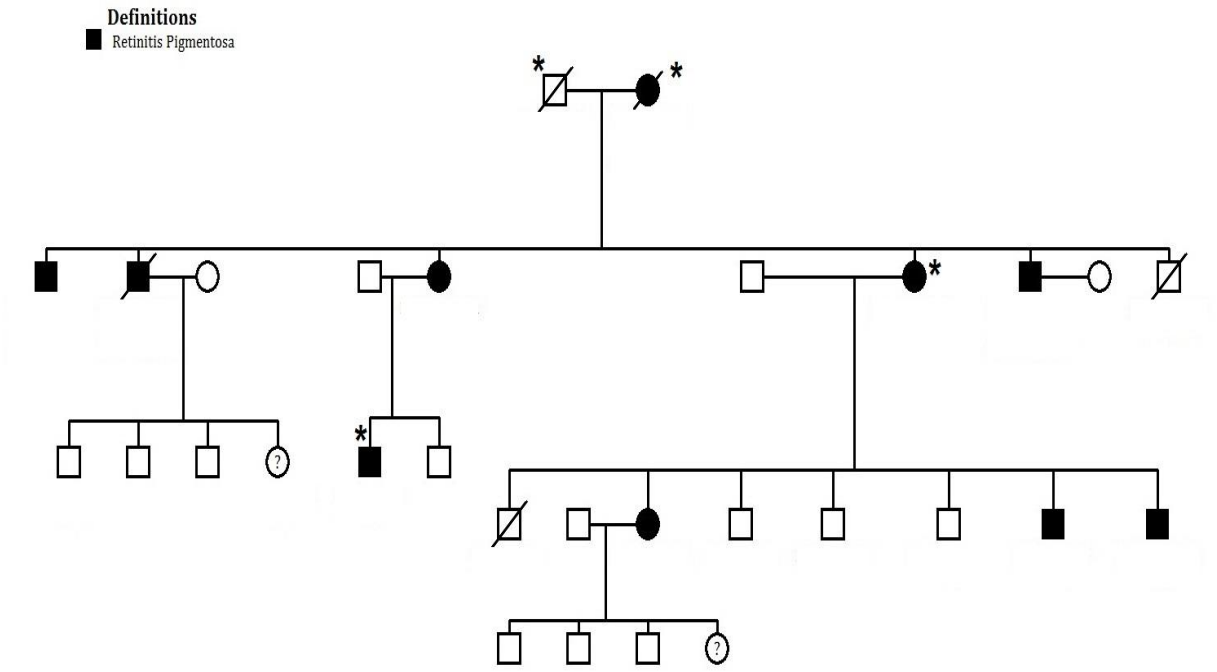
Variant filtering (blue) was done by sequentially using the 'inverse grep' command to exclude types of variants. Thereafter, the "file grep" command was used to extract RetNet IRD and ERDC genes (red and orange) for analysis. The 'inverse file grep' command was used to exclude the RetNet IRD and ERDC variants, thereby leaving the residual exome data (green).



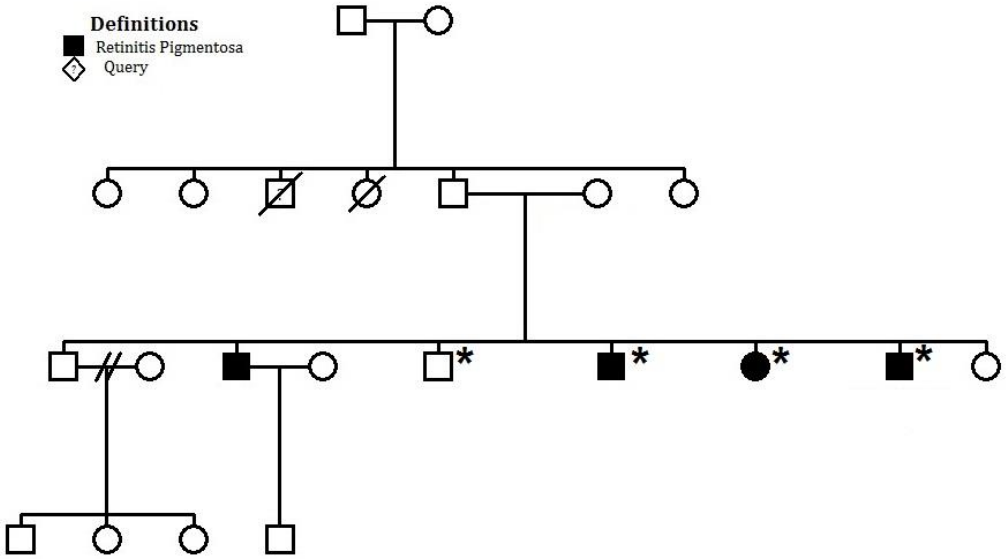
Appendix 5. Pedigrees of four unresolved families

Squares represent males, and circles, females. Shaded symbols indicate individuals affected with IRD. A question mark highlights subjects for whom the clinical status is unknown. Individuals selected for WES are noted with an asterisk.

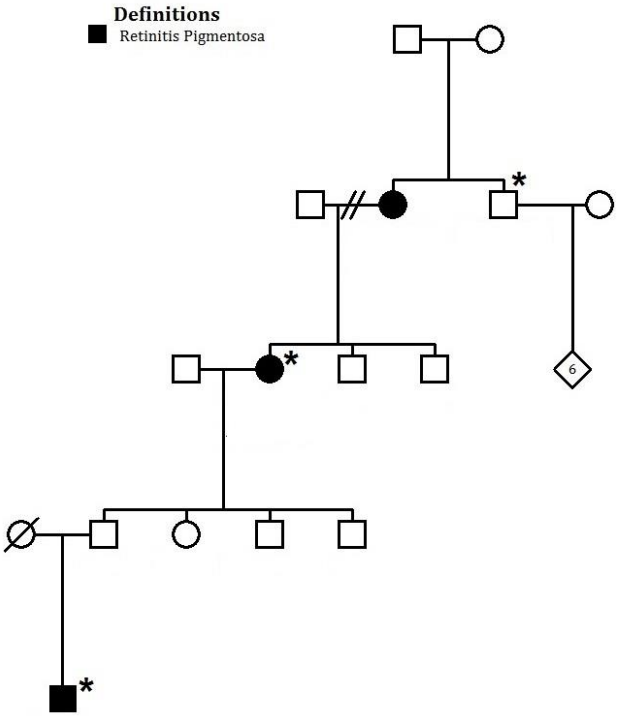
Family RPD 55



Family RPD 1005



Family RPD 799



Family RPD 94