Common *ABCA4* mutations in South Africans: frequencies, pathogenicity and genotype-phenotype correlations

By

Christel Nossek B.Sc (Med) Hons

NSSCHR003

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfilment of the requirements for the degree
MSc (Med) in Human Genetics

August 2010

Supervisor: Miss Lisa Roberts

Co-supervisors: Professor Rajkumar Ramesar and Professor Jacquie Greenberg

Division of Human Genetics, University of Cape Town

Table of contents

List of figures

List of tables

Abbreviations

°C	degrees Celsius
%	percentage
A	adenine
AAR(s)	ABCA4-associated retinopathy(ies)
ABCA4	retina-specific ATP-binding cassette transporter 4
ABC transporters	ATP-binding cassette transporters
adSTGD	autosomal dominant Stargardt disease
A2E	di-retinoid-pyridium- phosphatidylethanolamine
AMD	age-related macular degeneration
A2PE	di-retinoid-pyridium-ethanolamine
APEX	arrayed primer extension
AOO	age of onset
arRP	autosomal recessive retinitis pigmentosa
arSTGD	autosomal recessive Stargardt disease
AS-PCR	allele-specific polymerase chain reaction
ATP	adenosine triphosphate
BDGP	Berkeley Drosophila Genome project
BLAST	basic local alignment search tool
bp	base pair
C	cytosine
CFTR	cystic fibrosis transmembrane conductance regulator
CRALBP	cellular retinaldehyde-binding protein dH_20 distilled water
CRD	cone-rod dystrophy
dHPLC	denaturing high performance liquid chromatography
DNA	deoxyribonucleic acid
DMSO	dimethyl sulfoxide
dNTP	dinucleotide triphosphate
ddNTP(s)	dideoxynucleotide triphosphate(s)
ECD	exocytoplasmic domain

ESE(s)	exonic splicing enhancer(s)
EtBr	ethidium bromide
Exo I	exonuclease I
FFM	fundus flavimaculatus
g	grams
G	guanine
H_A	alternative hypothesis
H_O	null hypothesis
HPLC	high performance liquid chromatography
HMM	hidden markov model
kDa	kilodalton(s)
LRAT	lecithin retinol acyltransferase
L	liter
LNA	locked nucleic acid
$MgCl_2$	magnesium chloride
T_M	melting temperature
µl	microlitre
µM	micromolar
M	molecular weight marker
MD	macular degeneration
min	minute(s)
mM	millimolar
MSD	membrane-spanning domain
NBD(s)	nucleotide-binding domain(s)
NBF	nucleotide-binding folds
NCBI	National Center for Biotechnology Information
ng	nanogram
N-retinylidene-PE	*N*-retinylidene-phosphatidylethanolamine
PCR	polymerase chain reaction
PE	phosphatidylethanolamine
pH	log H^+ ions
pmol	pico mole(s)

PolyPhen	polymorphism phenotyping
RDD	retinal degenerative disorders
11*cis*RDH	11-*cis* retinol dehydrogenase
RDH8	all-*trans*-retinol dehydrogenase
REC	research ethics committee
RP	retinitis pigmentosa
RPE	retinal pigment epithelium
RPE65	retinal pigment epithelium-specific 65kDa protein
rpm	revolutions per minute
RT-PCR	reverse transcription-polymerase chain reaction
SA	South Africa
SAP	shrimp alkaline phosphatase
sec	second(s)
SIFT	sorting intolerant from tolerant
SNP(s)	single nucleotide polymorphism(s)
STGD	Stargardt disease
SR proteins	Ser/Arg-rich proteins
sub-PSEC	substitution position-specific evolutionary conservation
SUR 1	sulfonylurea receptor
T	thymine
T_a	annealing temperature
TBE	tris, boric acid, and EDTA buffer
TMDs	transmembrane domains
Tris	2-amino-2(hydroxymethyl)-1,3-propanediol
U	unit(s)
UCSC	University of California Santa Cruz
UCT	University of Cape Town
UK	United Kingdom
URC	University Research Council
USA	United States of America
V	volt(s)
WC	water control

Abstract

Stargardt disease (STGD), a juvenile-onset form of macular dystrophy resulting in a severe reduction of central vision, may be inherited in either an autosomal recessive or autosomal dominant manner. To date the only gene found to be involved with the autosomal recessive form is *ABCA4*. Mutations in this gene are associated not only with STGD, but with other autosomal recessive retinal diseases. Due to the numerous mutations detected in *ABCA4* and their associated phenotypic heterogeneity, a genotype-phenotype model has been proposed based on the amount of ABCA4 protein activity. Research in the Division of Human Genetics at the University of Cape Town (UCT) has suggested possible *ABCA4* founder mutations underlying STGD in the South African Caucasian Afrikaner population and has identified seven (C1490Y, R602W, IVS38-10T>C, L2027F, V256V, G863A, and R152X) common mutations. In a cohort of patients affected with an *ABCA4*-associated retinopathy (AAR) a total of 36% were identified as having various bi-allelic combinations of the seven mutations.

In the current study, SNaPshot PCR, allele-specific PCR (AS-PCR) and denaturing high performance liquid chromatography (dHPLC) analysis were used to screen for the seven mutations in a patient cohort and a control cohort. A high detection rate of bi-allelic disease-causing mutations in total of 28/72 patients (i.e. 38.89% were fully characterised) confirmed the designed assay to be a viable screening tool, which could be employed in a diagnostic setting. The detection of 12 heterozygotes in the Caucasian control samples (n = 269; 169 of which were specifically Afrikaner) resulted in an estimated background frequency of 4.46 per 100 individuals. This could be used by counsellors to discuss carrier risks with patients and their family members. Bioinformatic tools (PolyPhen, SIFT, PMUT, PANTHER PSEC, ESEfinder, and the BDGP Splice Site Prediction programme) revealed the predicted pathogenicity of the seven mutations to be as follows (in order of decreasing pathogenicity): C1490Y, R602W, V256V, R152X, G863A, L2027F, and IVS38-10T>C. Statistical analysis (using the Kruskal-Wallis test and the Wilcoxon Rank Sum test) showed no significant

effect of mutation combination on phenotype (i.e. AOO/severity as a measure of clinical outcome).

To improve the understanding of the genotype-phenotype correlation a larger cohort of South African STGD patients with the same common mutations in various combinations and the availability of sufficient clinical data, is required. Further investigations into the genotype-phenotype correlation, combined with the information on the pathogenicity of the mutations, could result in increased understanding regarding the impact of each mutation, thus enhancing the clinical utility of identifying *ABCA4* mutations.

Chapter 1: Introduction

1.1 The structure of the human eye

The human eye consists of three layers: 1) the outer fibrous coat, 2) the middle vascular layer, and 3) the inner coat (Smit and van Dijk 1970; Grogan and Suter 1999). The structures comprising these three layers are discussed below (Fig. 1.1).

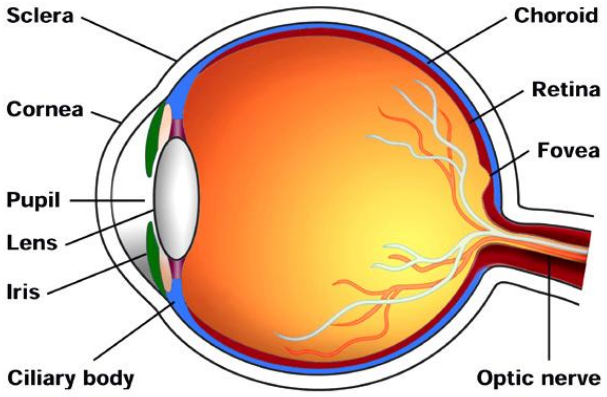


Fig. 1.1 An illustration of a cross section of the human eye, showing important structural components.

Reproduced from the website of the Steve Tatler Illustration Portfolio: http://illustration.stevetatler.co.uk/illustration6.htm.

1) The **outer fibrous coat** consists of two parts, the sclera (the posterior part of the eye) and the cornea (the anterior part of the eye).

a) The sclera is a white inflexible membrane made up of connective tissue. It functions to:

i) protect the inner parts of the eye

ii) provide for the attachment of the eye muscles, and

iii) help maintain the shape of the eye.

b) The cornea is a transparent layer and is a continuation of the sclera, but is more convex in structure. It has no blood vessels, but does consist of many

nerves. The transparent nature of the cornea allows for light rays to pass through and reach the light-sensitive cells.

2) The **middle vascular layer** consists of three parts, namely the choroid, the ciliary body, and the iris.

a) The choroid lines the sclera and is made up of loose connective tissue which is rich in blood vessels. The choroid contains dark pigments that play a role in light absorption. These pigments prevent the reflection of excessive rays of light within the eye, thus preventing blurred vision.

b) The ciliary body connects the iris to the choroid. It consists of circular ciliary muscles, which function in controlling the curvature of the lens.

c) The iris forms the anterior part of the middle vascular layer, and consists of a thin, contractile curtain-like structure with a circular opening in the centre called the pupil. The iris is comprised of involuntary circular and radial muscle fibres, which control the size of the pupil, thereby regulating the amount of light entering the eye.

3) The **inner coat** of the eye is comprised of the retina. The retina consists of an outer pigmented layer and an inner neuronal layer (Fig. 1.2).

a) The inner neuronal layer consists of five major cell types, namely:

i) the photoreceptor cell layer

ii) the horizontal cell layer

iii) the bipolar neuron layer

iv) the amacrine cell layer, and

v) the ganglion cell layer.

b) The outer pigmented layer or retinal pigment epithelium (RPE) consists of a single layer of cuboidal cells that contain dark pigment granules, which help to absorb light. The RPE is situated between the outer segments of the photoreceptor cells and the choroid (Coles 1989).

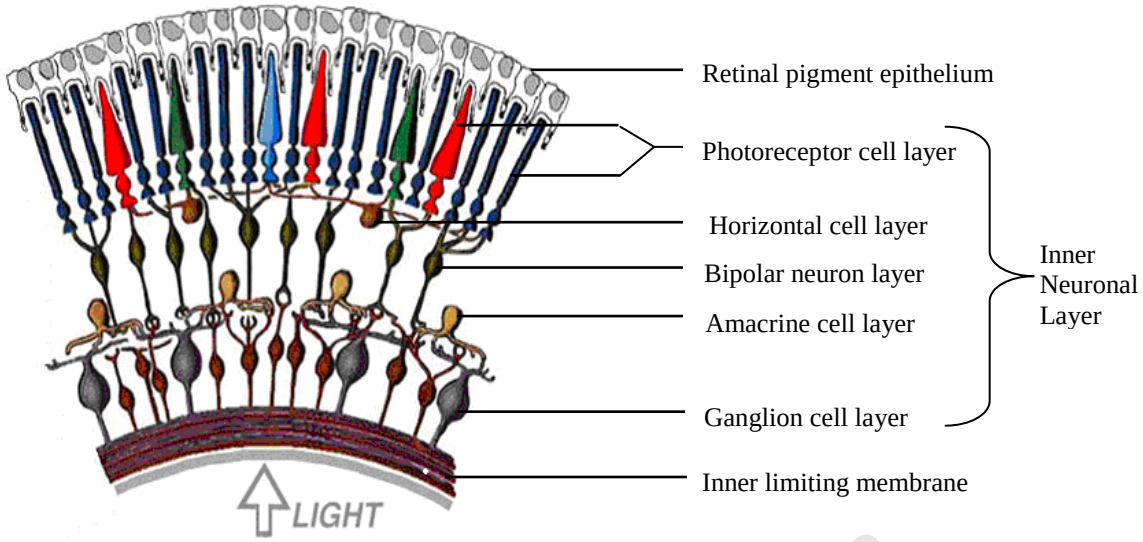


Fig. 1.2 An illustration of the retina showing the retinal pigment epithelium and five cell types of the neuronal layer.

Reproduced and adapted from the website of the Med Gadget Internet Journal of emerging medical technologies: http://medgadget.com/archives/2006/04/nonlightsensiti_1.html.

1.1.1 The retina

The retina consists of an organised layer of cells that are involved in the early stages of the visual cascade (Fig. 1.2) (Molday 2007). The photoreceptor cell layer contains the receptors required for vision, namely the rods and the cones. The horizontal cell layer connects clusters of photoreceptor cells. The visual impulse is transmitted to the bipolar cells, which in turn synapse with the amacrine cells. The amacrine cells carry the impulse to the ganglion cells. The axons of the ganglion cells join to form the optic nerve. The optic nerve conducts the impulses to the brain, where they are interpreted, resulting in vision. The retinal region where the optic nerve leaves the eye is known as the blind spot as it contains no photoreceptors, hence there is no light sensitivity in this area (Smit and van Dijk 1970; Coles 1989; Grogan and Suter 1999).

1.1.1.1 The photoreceptor cells of the retina

Photoreceptor cells are modified neurons that convert light rays into nerve impulses (Smit and van Dijk 1970; Grogan and Suter 1999). These cells consist of an outer segment and an inner segment, which are joined by a thin cilium (Fig. 1.3). The photoreceptor outer segments contain discs and are involved in the detection of light and in the conversion of light into electrical impulses in a process known as phototransduction (Molday 2007). The inner segment contains the mitochondria, endoplasmic reticulum, and other organelles essential for cell metabolism. The cell body is situated just below the inner segment. The cell body, containing the nucleus, is connected to the synaptic region. It is from the synaptic region of a photoreceptor cell that a visual impulse is transmitted to the subsequent secondary neurons in the inner neuronal layer of the retina (Molday 2007; Kawamura and Tachibanaki 2008).

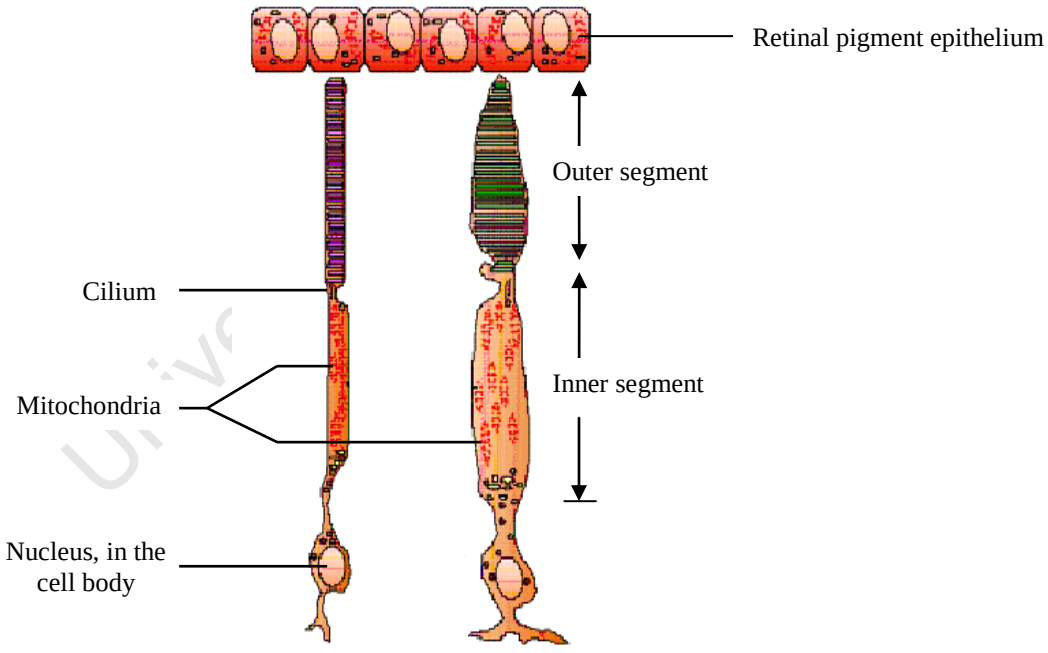


Fig. 1.3 A diagrammatic representation of the photoreceptor cells, showing the gross structure of the rod and cone cells.

Reproduced and adapted from Ganfyd:
http://www.ganfyd.org/index.php?title=Image:SchematicPhotoreceptorsRetina.png.

There are two types of photoreceptor cells (Grogan and Suter 1999; Smit and van Dijk 1970):

> Rod cells are thin and elongated and are highly concentrated in the peripheral region of the retina. The eye contains approximately 140 million rod cells. The outer segment of these cells contains the visual pigment, rhodopsin.
>
> Cone cells are broader and cone-shaped, and are found to be more abundant in the central region of the retina. The eye contains approximately seven million cone cells. The outer segment of these cells contains the visual pigment, iodopsin.

Both the rod and cone cells are sensitive to light, with the rods having a greater sensitivity. Rod cells are specialised for dim light or night vision, whereas cone cells are specialised for bright light or day vision, and are involved in visual acuity and colour perception (Smit and van Dijk 1970; Molday 1998; Grogan and Suter 1999).

The outer segments of rod cells consist of more than 1000 flattened discs that are stacked together and are surrounded by a plasma membrane (Molday 2007). In comparison to the rod cells, the plasma membrane of the outer segment of a cone cell forms a continuous folded structure with the discs.

The photoreceptor cells form a functional unit with the previously mentioned RPE. The RPE functions to: a) absorb stray light rays, b) provide transport of substrates produced in the visual cycle to the photoreceptor cells, c) create a blood-retina barrier, d) renew the visual pigments in the photoreceptor cells, and e) phagocytose the shed outer segment photoreceptor cell discs (Coles 1989). The surface of the RPE contains many processes that expand into the spaces between the rod and cone cell outer segments. These processes may be involved in phagocytosis of the photoreceptor outer segments (Essner et al. 1978; Coles 1989).

In order for photoreceptor cells to carry out their crucial function in the process of vision, the structure and components (visual pigments) of these cells needs to be maintained. The outer segments of photoreceptor cells undergo a continuous process of

phagocytosis and renewal. In rod cells, new disks are formed at the base of the outer segment in the cilium, while older disks are shed from the tips. The shed disks are released into the inter-photoreceptor space and phagocytosed by the RPE (Essner et al. 1978; Fein and Szuts 1982; Coles 1989). A similar process exists for cone cell outer segments (Anderson et al. 1978). The entire outer segments of rod cells are replaced every eight to 14 days (Coles 1989). In cone cells, disc renewal is much slower, with the process taking nine months to one year. The discs from rod cells are shed in light, mainly after extended periods of darkness, in comparison to cone cells where the discs are shed during the night.

1.1.1.2 The macula

The retina also consists of a region called the *macula lutea*, more commonly known as the macula. The macula is an oval region situated near the centre of the retina, adjacent to the previously mentioned blind spot. The macula has a darker appearance in comparison to the retina as a whole, and consists mainly of cone cells. The macula contains a central pit, which is called the fovea. This central fovea is very thin and contains no rod cells, and is the region of highest visual acuity in the retina (Fig 1.4) (Smit and van Dijk 1970; Coles 1989; Grogan and Suter 1999). Numerous hereditary diseases affecting the macula have been identified and are characterised by loss of macula functioning (Weber 1998). Due to the high concentration of cone cells in this region of the retina, both colour and central vision is affected (Fig 1.5). These eye disorders are collectively known as macular dystrophy or degeneration (MD) and can be inherited in an autosomal dominant, autosomal recessive or X-linked manner (Michaelides et al. 2003).

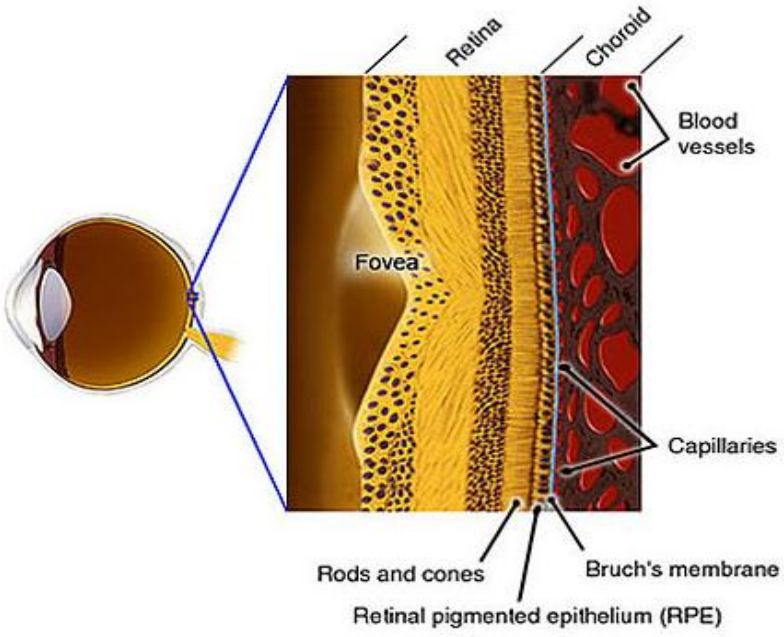


Fig. 1.4 A diagrammatic representation of the macula, showing the central fovea region that consists of mainly cone cells.

Reproduced from Macular Degeneration Research - A Program of the American Health Assistance Foundation: http://www.ahaf.org/macular/about/understanding/normal-macula.html.

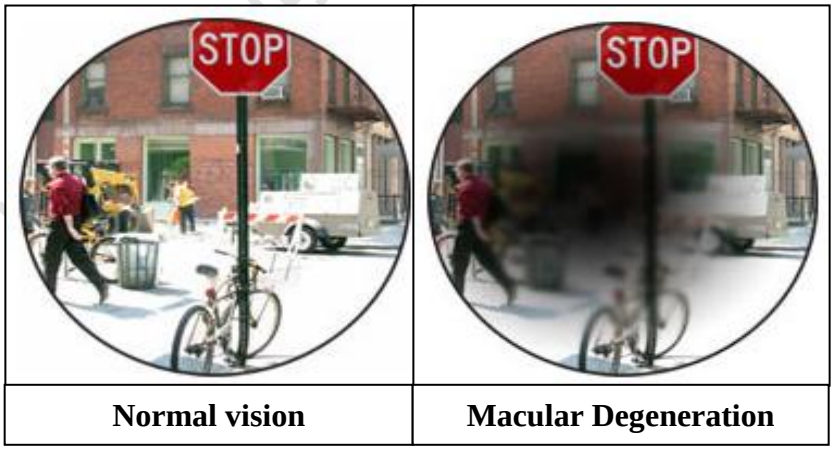


Fig. 1.5 An illustration showing the loss of vision in individuals with macular degeneration.

Reproduced from Healthy Fellow – Your Natural Health Critic Foundation: http://www.healthyfellow.com/150/vitamins-for-macular-degeneration/.

1.2 The visual cycle

The structures that form the inner lining of the human eye (the retina) play essential roles in the visual transduction pathway and process of vision. The photoreceptors make use of a chromophore, 11-*cis* retinal, for the absorption and processing of light via the visual transduction pathway (Thompson and Gal 2003). This vitamin A analog is highly photosensitive and is responsible for initiating the signal transduction that results in vision. Upon light activation the chromophore is converted to all-*trans* retinal. Sustained phototransduction, which is dependent on the replenishment of 11-*cis* retinal, can be accomplished by maintaining the cycling process of vitamin A analogs through the visual cycle (Fig. 1.6). The visual cycle involves the metabolism of vitamin A and the circulation of vitamin A analogs between the RPE and the photoreceptor cells. The visual cycle involves numerous specialised proteins and enzymes and begins with the absorption of light by the visual pigments, rhodopsin and iodopsin, in the rod and cone photoreceptor cells, respectively. The visual pigments are membrane proteins located in the outer segments of the photoreceptor cells, and contain the chromophore, 11-*cis* retinal (Thompson and Gal 2003).

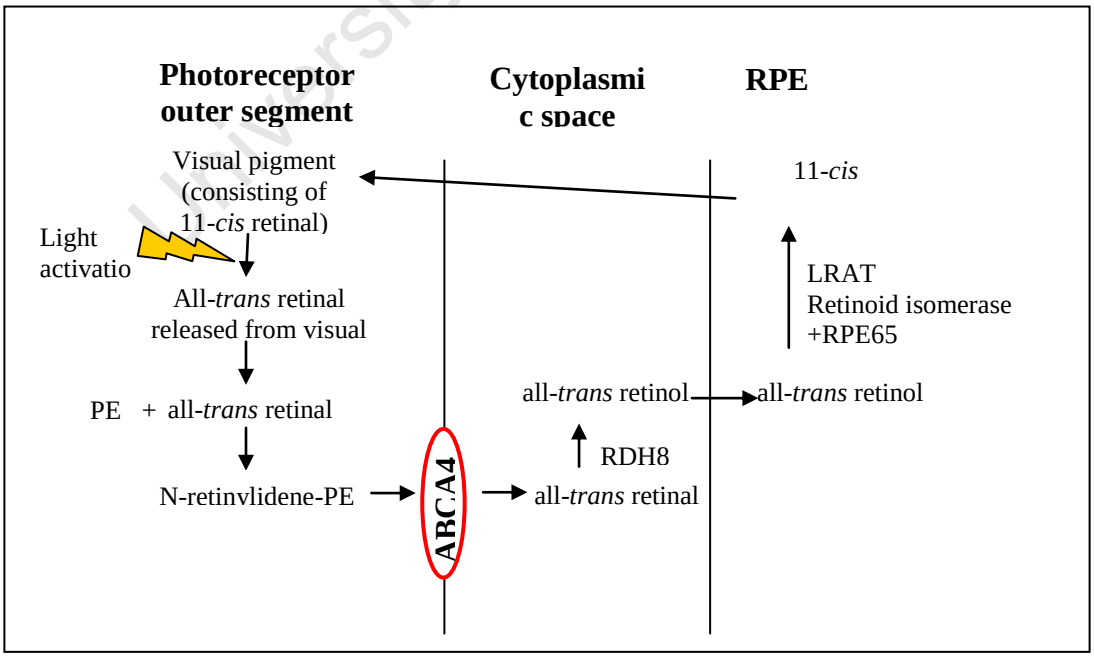


Fig. 1.6 A simplified diagrammatic representation of the visual cycle.

Following light activation, all-*trans* retinal is released from the visual pigments. In the rod photoreceptor cells, all-*trans* retinal is converted into a product called *N*-retinylidene-phosphatidylethanolamine (N-retinylidene-PE) by reacting with phosphatidylethanolamine (PE). The N-retinylidene-PE is transported to the surface of the cytoplasmic plane of the disc membrane by the retina-specific adenosine triphosphate (ATP)-binding cassette transporter (ABCA4), formerly known as ABCR. This leads to the release of all-*trans* retinal into the cytoplasmic space (Thompson and Gal 2003; Molday et al. 2009).

Once in the cytoplasm, all-*trans* retinal is converted to all-*trans* retinol by an enzyme, all-*trans* retinol dehydrogenase (*RDH8*), and is transported to the RPE. On arrival at the RPE, all-*trans* retinol is esterified in the lipid bilayer to form all-*trans* retinyl esters. The enzyme lecithin retinol acyltransferase (*LRAT*) is involved in this process. This is followed by the generation of an 11-*cis* double bond in the side chain by the enzyme, retinoid isomerase. An RPE-specific protein, RPE65, is also needed for this reaction. The final aldehyde product is produced by the enzyme 11-*cis* retinol dehydrogenase (11*cis*RDH). This reaction is accelerated by the cellular retinaldehyde-binding protein (*CRALBP*). The 11-*cis* retinal leaves the RPE, passes through the subretinal region, and enters the outer segments of the photoreceptors. The 11-cis retinal attaches to the opsin proteins, and the visual pigments are regenerated (Thompson and Gal 2003; Miyazono et al. 2008).

In cone cells the transport of vitamin A analogs resulting in the initiation of vision is similar. The rate of reduction of all-*trans* retinal to all-*trans* retinol is 10-40 times greater in cones compared to the reduction in rod cells (Miyazono et al. 2008). Studies conducted on cone cells suggest a faster supply of 11-*cis* retinal occurs for regeneration of the visual pigments. Numerous reports propose that cone cell visual pigments can also be regenerated independently from the RPE in a cone-specific pathway in Müller cells utilising a retinal-retinol redox (ALOL)-coupling reaction. Müller cells are radial glial

cells present in the inner retina that help maintain the normal functioning and survival of the neurons in the retina (Franze et al. 2007).

1.3 ATP-binding cassette transporter family of proteins

The ATP-binding cassette transporters (ABC transporters) comprise a superfamily of proteins that represent the largest group of membrane transporters. ABC proteins are found in both prokaryotic and eukaryotic organisms. These transmembrane proteins bind ATP, using the energy to transport a wide range of substrates, including amino acids, metabolites, vitamins, and steroids across cell membranes to various regions in the cell (Allikmets et al. 1997a; Molday 2008; Loo and Clarke 2008). Some of the ABC protein members transport a variety of different compounds with diverse structures, whereas other ABC members are specific to one substrate. Not all ABC proteins function as transporters however, for example, the cystic fibrosis transmembrane conductance regulator (CFTR) which acts as a chloride channel, and the sulfonylurea receptor (SUR 1) which is a potassium channel regulator. To date, 49 ABC proteins have been identified, 19 of which have been linked to diseases (Loo and Clarke 2008).

1.3.1 The basic structure of an ABC transporter

ABC transporters consist of two nucleotide-binding domains (NBDs) and two transmembrane domains (TMDs). The NBDs or ATP-binding domains (also referred to as nucleotide-binding folds or NBF) are situated in the cytoplasmic region of the cell, and function to bind and hydrolyze ATP, thus providing the energy needed to transport the substrate through the membrane. Both NBDs contain sequences that are conserved for ATP-binding; these include the Walker A and Walker B motifs, Q-loop, LSGGQ signature sequence (C motif), D-loop, and H-loop. The Walker A and B motifs are a common feature of all ATP-requiring proteins. The C motif or linker peptide is characteristic of the members of the ABC superfamily. The TMDs usually consist of six transmembrane segments that are connected together by intra- and extra-cellular loops.

The TMDs provide a pathway for the translocation of a compound across the membrane (Saurin et al. 1999; Dean et al. 2001; Molday 2007; Loo and Clarke 2008). The structure or composition of the TMD determines the specificity for the type of molecule(s) transported (Dean and Allikmets 1995). Little homology exists with respect to the amino acid composition of the TMDs among the various ABC transporters (Loo and Clarke 2008).

Amino acid sequence alignments of the ATP-binding domains and the organisation of the NBDs of the various ABC proteins, has resulted in these transporters being classified into subfamilies. To date seven subfamilies exist, which are labelled alphabetically from ABCA to ABCG (Dean and Allikmets 2001, Molday 2007). For the purpose of this investigation, the members belonging to the subfamily, ABCA, will be described in more detail.

1.3.2 The ABC transporter subfamily, ABCA

Members of the ABC transporter protein subfamily ABCA show a high degree of similarity in both sequence and structural organisation (Molday et al. 2009). This subfamily consists of twelve ABCA genes, divided into two groups based on gene structure and homology. Group one consists of the genes *ABCA5*, *ABCA6*, *ABCA8*, *ABCA9* and *ABCA10*. Group two consists of the genes *ABCA1*, *ABCA2*, *ABCA3*, *ABCA4*, *ABCA7*, *ABCA12*, and *ABCA13*. The genes present in group one each have 38 introns, whereas the genes in group two have 50 to 51 introns. The second group of genes also have longer amino terminal portions when compared to the genes in the first group (Dean et al. 2001; Dean and Allikmets 2001; Molday 2007; Molday et al. 2009). Proteins encoded by ABCA genes have been implicated in the transport of a variety of lipids across cell membranes. The functions of all the proteins in this subfamily have not been elucidated (Dean et al. 2001). However, four ABCA proteins (ABCA1, ABCA3, ABCA4, and ABCA12) are known to be related with different genetic diseases that are caused by defects in lipid transport. Lipid transport is essential as it ensures the structure, function and survival of cells (Molday et al. 2009). A list of the twelve ABCA

genes, their chromosomal locations, and associated diseases (if known) are listed in Table 1.1.

Table 1.1 The location and associated diseases of the twelve genes that encode the ABCA protein subfamily.

Adapted from Kaplan et al. 1993, Dean et al. 2001 and Allikmets 2007.

ABCA Gene Symbol	Chromosomal Location	Associated Disease
ABCA1	9q31.1	Tangier disease (Rust et al. 1999)
ABCA2	9q34	
ABCA3	16p13.3	Respiratory distress syndrome (Cheong et al. 2006)
ABCA4	1p13-21	Autosomal recessive Stargardt disease (Allikmets et al. 1997a), Autosomal recessive retinitis pigmentosa (Rozet et al. 1999), Cone-rod dystrophy (Briggs et al. 2001)
ABCA5	17q24	
ABCA6	17q24	
ABCA7	19p13.3	
ABCA8	17q24	
ABCA9	17q24	
ABCA10	17q24	
ABCA12	2q34	Harlequin ichthyosis (Akiyama 2006)
ABCA13	7p11-q11	

1.3.3 The ABCA4 gene and protein

The *ABCA4* gene is located on chromosome 1, at position 1p13-21 (Kaplan et al. 1993). Initially it was thought that the gene comprised 21 exons, but further research revealed the presence of 50 exons in total (Allikmets et al. 1997a; Gerber et al. 1998). The exon

sizes range from 33 to 266 base pairs (bp), with the gene having an open reading frame greater than 6 800 bp (Allikmets et al. 1998).

Studies utilising immunofluorescent labels have revealed the expression of ABCA4 in the outer segment of both the cone and rod photoreceptor cells (Allikmets et al. 1998; Molday et al. 2000, Molday 2007). Various reports have described ABCA4 to be an inward-directed retinoid flipase which functions in the removal of vitamin A analogs from the photoreceptor outer segments (Sullivan 2009). As mentioned previously, ABCA4 functions in the cycling of vitamin A analogs between the photoreceptor cells and the RPE surface in the retina, by transporting the compound, N-retinylidene-PE, to the cytoplasmic plane of the disc membrane of the photoreceptor cells (Thompson and Gal 2003, Molday et al. 2009). The function of ABCA4 thus appears to be to ensure the complete removal of all-*trans* retinal and N-retinylidene-PE from the outer segment disc membranes once light activation has occurred, leading to the continuation of the visual cycle (Molday et al. 2009).

1.3.3.1 The structure of the ABCA4 protein

The protein is comprised of 2 273 amino acids and is structurally arranged into two halves. Computer algorithms have been used to generate topological models of the protein (Bungert et al. 2001; Molday 2007). The models indicate that each half of the protein contains:

- a single transmembrane segment with hydrophobic properties (designated H1 in the N-terminal and H7 in C-terminal half of protein)
- a large exocytoplasmic domain (ECD)
- a membrane-spanning domain (MSD), and
- a nucleotide binding domain (NBD).

(Fig. 1.7)

The N-terminal segment of the protein consists of 24 amino acids, many of which are positively charged lysine and arginine residues. Due to the nature of this region it is predicted to be situated on the cytoplasmic side of the disk membrane (Molday 2007).

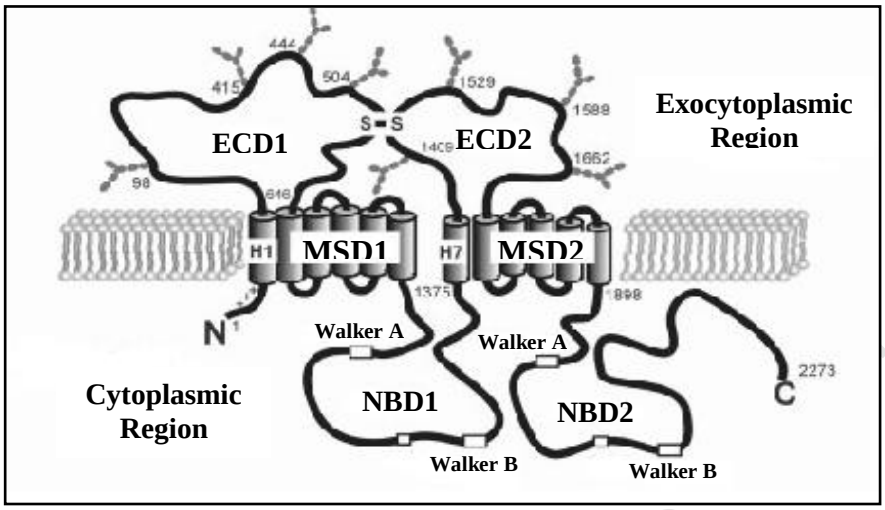


Fig. 1.7 An illustration of the proposed structure of the ABCA4 protein, showing the two exocytoplasmic domains (ECDs), two membrane-spanning domains (MSDs), and two nucleotide binding domains (NBDs).

Reproduced and adapted from Bungert et al. 2001.

The C-terminal segment of the protein, comprised of 170 amino acids, is located downstream from NBD2. This C-terminal region consists of a conserved motif containing six amino acids, Valine-Phenylalanine-Valine-Asparagine-Phenylalanine-Alanine (VFVNFA). Research has indicated that this conserved region is important in the correct folding of the ABCA4 protein (Molday 2007; Zhong et al. 2008; Molday et al. 2009).

Studies have revealed the presence of four *N*-linked glycosylation sites in each half of the ABCA4 protein (Bungert et al. 2001; Molday 2007). The presence of these N-linked glycoslylation sites on both ECD1 (602 amino acids) and ECD2 (289 amino acids) has lead to the prediction that these domains are located on the lumen side of the membrane. Intra-chain disulfide bonds are also predicted due to the presence of many conserved cysteine amino acids in both ECDs. The high degree of sequence conservation observed

amongst ECDs of various vertebrate ABCA4 proteins, has lead to investigations into the effect of mutations occurring in these regions (Biswas-Fiss et al. 2010).

Both NBDs have approximately 140 amino acid residues, and share 35% sequence identity. The NBDs contain Walker A and B motifs as well as an active signature motif (C-loop). As mentioned previously, the Walker A and B motifs are a hallmark of ATP-requiring proteins and C motifs are characteristic of ABC proteins (Bungert et al. 2001; Molday 2007).

The number and identity of the transmembrane segments forming the MSDs has yet to be defined. Computer programs and hydropathy plots have predicted five or six transmembrane segments in each MSD. These membrane-spanning segments, along with the single, hydrophobic transmembrane segment that is situated just before each ECD, presumably form the binding site for substrates (Bungert et al. 2001; Molday 2007; Molday et al. 2009).

1.4 ABCA4-associated retinopathies

1.4.1 Stargardt disease

Stargardt disease (STGD) (OMIM: 248200) was first described by Karl Stargardt in 1909 (Kaplan et al. 1993). In 1997 the link between STGD and the *ABCA4* gene was published (Allikmets et al. 1997a). STGD is a juvenile-onset form of MD (Allikmets et al. 1997a; Lewis et al. 1999; Lorenz and Preising 2005). As mentioned previously, MD involves the macula which is situated in the central region of the retina and contains a high concentration of cone photoreceptor cells.

STGD usually presents in the first or second decade of life and has an autosomal recessive mode of inheritance. It is one of the most frequent causes of macular degeneration in childhood, accounting for approximately 7% of all retinal diseases.

STGD has an estimated incidence of 1 in 10 000 individuals, and is characterised by loss of cone cells which results in a severe reduction of central vision (Kaplan et al. 1993; Lewis et al. 1999; Lorenz and Preising 2005; Allikmets 2007).

Clinical features of the disease include progressive bilateral degeneration of the macula and the RPE. The appearance of orange-yellow flecks located around the macula and/or midretinal periphery is a distinctive feature of STGD (Fig. 1.8) (Allikmets et al. 1997a; Sun and Nathans 1997). Fluorescein angiography - a test that uses fluorescein dye to study the blood vessels in the retina, iris, and choroid - has shown the choroid to be dark in many patients, which is indicative of the accumulation of lipofuscin (Allikmets 2007). Lipofuscin consists of a combination of different fluorescent by-products of the visual cycle. The exact molecules that comprise lipofuscin are not known, however a di-retinoid-pyridinium-PE (A2E) fluorophore has been shown to be the major component. This fluorophore appears to have toxic effects on the RPE (Boon et al. 2008).

STGD can also occur as an autosomal dominant disease (adSTGD). Three dominant loci have been identified on chromosomes 4, 6 and 13, respectively (Stone et al. 1994; Zhang et al. 1994; Kniazeva et al. 1999). The dominant form of STGD does not fall within the scope of this present study, and will not be discussed further.

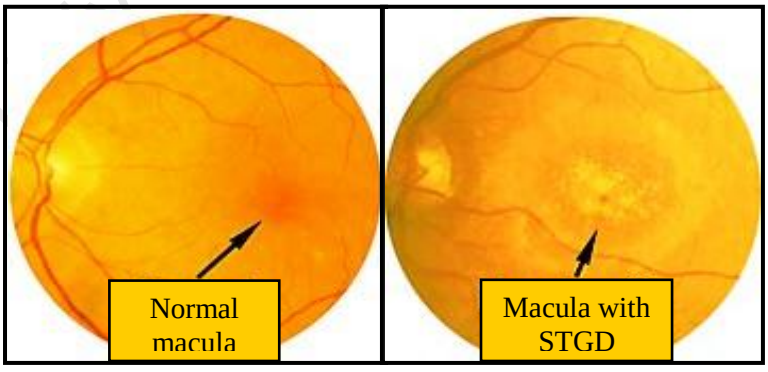


Fig. 1.8 A fundus photograph of the posterior of the eye (by focusing light through the front of the eye) showing the differences between a normal macula and a macula with STGD.

Reproduced and adapted from St Luke's Cataract & Laser Institute: http://www.stlukeseye.com/Conditions/Stargardts.asp.

1.4.2 Mechanism of ABCA4 pathogenesis

Protein misfolding, decreased functioning and mislocalisation due to mutations in the *ABCA4* gene underlie ABCA4-associated retinopathies (AARs). The proposed mechanism of *ABCA4*-related pathogenesis is that retinoid products, such as all-*trans* retinal and N-retinylidene-PE, are not completely removed from the photoreceptor outer segment disc membrane (Fig 1.9). These two compounds react to form diretinal derivatives, such as di-retinoid-pyridinium-ethanolamine (A2PE). During the process of photoreceptor disc shedding, A2PE is ingested by the RPE in phagosomes, along with other retinoid compounds. These compounds fuse, resulting in the formation of phagolysosomes which house enzymes that are capable of degrading the photoreceptor outer segments. A2PE is converted to A2E, which is unable to be metabolised further. The inability of A2E to undergo any additional metabolism results in the accumulation of A2E and other retinal compounds, collectively known as lipofuscin, in the RPE (Molday et al. 2009). It appears that A2E has the potential to cause RPE dysfunction through at least two mechanisms: light-independent inhibition of lysosomal function and phototoxicity (Schütt et al. 2000). The buildup of toxic products can cause a loss of RPE function and ultimately a loss of photoreceptor cells, which leads to loss of vision (Molday et al. 2009). Photoreceptors and vision are lost secondary to the dysfunction and/ or death of the RPE (Cideciyan et al. 2004).

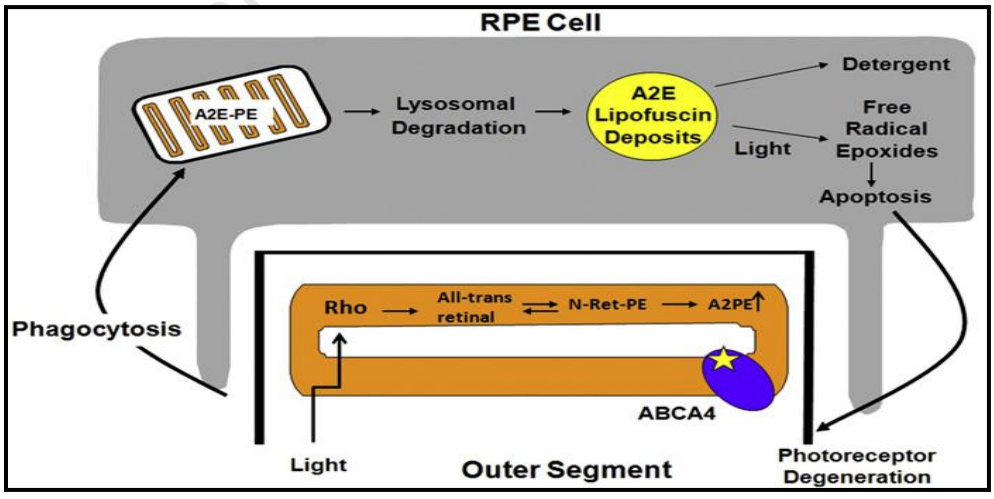


Fig. 1.9 A proposed model for ABCA4-mediated retinal degeneration.

Reproduced from Molday et al. (2009).

1.4.3 Clinical heterogeneity of ABCA4-associated retinopathies

To date *ABCA4* has been the only gene found to be involved with autosomal recessive STGD (arSTGD) (Allikmets 2007). Despite the apparent locus homogeneity, there is considerable allelic and phenotypic heterogeneity. More than 500 *ABCA4* mutations have been identified to date including single-base substitutions, duplications and deletions (Lewis et al. 1999; Briggs et al. 2001; Aguirre-Lamban et al. 2008). Mutations in the *ABCA4* gene have been found to be associated not only with STGD, but with other RDDs including fundus flavimaculatus (FFM), cone-rod dystrophy (CRD), and autosomal recessive retinitis pigmentosa (arRP) (Allikmets et al. 1997a; Allikmets et al. 1997b; Cremers et al. 1998). Research has also been conducted on the role that the *ABCA4* gene might play in age-related macular degeneration (AMD) which is the most common cause of visual loss in the elderly. AMD is similar to STGD, in that it affects the macula. Investigations linking *ABCA4* with AMD have been controversial, with some researchers showing no significant association (Allikmets et al. 1997b; Pennisi 1997; Stone et al. 1998, Patel et al. 2008). Environmental risk factors, including diet and smoking, have also been found to be associated with AMD.

FFM is clinically similar to STGD, however, it often has a later age of onset, slower disease progression, and a more widely spread distribution of the characteristic orange-yellow flecks on the retina (Maugeri et al. 1999; Allikmets 2007). The terms FFM and STGD are often used interchangeably.

CRD is characterised by a decrease in central vision, and granular changes in the macular RPE. In the initial stages of CRD there can be a reduction or absence of cone cell responses with only a slight decrease in rod cells functioning, or both photoreceptor cells types may be equally affected (Klevering et al. 2004).

Retinitis pigmentosa (RP) comprises a group of retinal diseases characterised by the loss of rod photoreceptor cells. Clinical features of RP include progressive peripheral vision loss and night blindness (Cremers et al. 1998). RP is both clinically and genetically heterogeneous and can be inherited as an autosomal dominant, autosomal recessive, or

X-linked disease. There have been more than 20 genes and loci associated with RP, but these account for less than 50% of cases (Martinez-Mir et al. 1998). arRP is the most severe retinal disease associated with mutations in the *ABCA4* gene. It is thought to be due to the complete loss of ABCA4 protein activity. Both truncating and abnormal splicing mutations have been found in families affected with ABCA4-associated arRP (Rozet et al. 1999; Wiszniewski et al. 2005).

1.4.4 Genotype-phenotype models

The vast array of *ABCA4* mutations, along with the nature of the specific mutations or combination of mutations is ultimately responsible for the broad range of the phenotype. Different mutations, leading to an altered amino acid sequence, affect protein function in different ways. These mutations could result in a decrease in ABCA4 activity, or create a misfolded protein which could give rise to a toxic effect on a) the cell(s) involved, or b) the cell(s) in the vicinity. This may be a result of either the type of mutation (i.e. missense, nonsense, frameshift, etc) or position of mutation (Sullivan 2009).

The large amount of allelic and phenotypic heterogeneity observed with *ABCA4* mutations has led to the proposal of a genotype-phenotype model (Fig. 1.10). Various research groups have attempted to determine the relationship between different *ABCA4* mutations and the severity of AARs. Van Driel et al. (1998) stated that most *ABCA4* mutations could be placed into different classes of severity, and that the phenotype can vary depending on the remaining total activity of the ABCA4 protein. The same model was later discussed by Shroyer et al. (1999).

The proposed genotype-phenotype model states that both truncated proteins (resulting from the presence of null alleles) and extremely misfolded proteins are associated with RP and CRD. Patients who retain partial ABCA4 activity due to the presence of so-called “milder” mutations (some nonsense and missense mutations) will develop arSTGD and FFM. It is hypothesised that the development of AMD in patients harbouring an *ABCA4* mutation might be the result of the gradual build up of toxic

byproducts (lipofuscin) leading to eventual photoreceptor cell loss (van Driel et al. 1998; Shroyer et al. 1999).

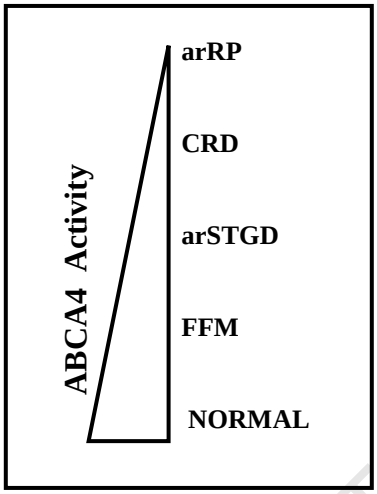


Fig. 1.10 An illustration showing the proposed genotype-phenotype correlation model for ABCA4.

Adapted and reproduced from: Shroyer et al. (1999). On the left ABCA4 activity is represented by the triangle. ABCA4 protein activity is the greatest at the bottom of the triangle, decreasing towards the top of the triangle. It has been proposed that arRP is a result of very low ABCA4 activity, while arSTGD has been proposed to result from mild ABCA4 activity. An individual with the normal amount of ABCA4 activity does not present with any AAR.

The biochemical function of the ABCA4 protein and the examination of the functional consequences of *ABCA4* mutations (i.e. the pathogenic contribution) are important factors to consider in order to fully understand the proposed genotype-phenotype correlation (Biswas-Fiss et al. 2010; Schindler et al. 2010). In 2006, Valverde et al. hypothesised that both the type of *ABCA4* mutation and its position, in conjunction with the combination of the mutant alleles, play a role in the age of onset (AOO) of visual impairment or deterioration (Valverde et al. 2006). Similarly, Lewis et al. (1999) used the AOO of visual impairment to assess the clinical severity of a pair of mutant *ABCA4* alleles. In the latter study Lewis et al. observed more frame-shift and early nonsense mutations in compound heterozygote individuals with an AOO of less than 10 years.

1.5 ABCA4-associated retinopathy research in South Africa

The implications of molecular genetic studies in diagnosing retinal diseases was realised in the 1980's with the identification of a restriction fragment length polymorphism (RFLP) on the X-chromosome found to be linked to RP (Oswald et el. 1985). Research into the genetics of retinal degenerative disorders (RDDs) was initiated in the Division of Human Genetics at the University of Cape Town (UCT) in 1990, with the banking of DNA samples for both molecular genetic studies and diagnostic purposes. To date DNA has been obtained from 2941 individuals in 1210 families. The RDD families are categorised according to their clinical diagnosis or pattern of inheritance. Table 1.2 shows the division of RDD families into STGD-affected families, families affected with arRP, and families affected with a macular-associated RDD, with specific focus on the Caucasian families for the purpose of the current study (as at May 2010).

Table 1.2 A summary of families collected for the RDD project according to their clinical diagnosis.

RDD	Number of Families	Number of Caucasian Families	Number of Affected Caucasian Individuals
STGD-affected	228	197	252
arRP	89	53	73
Macular-associated	323	290	315
Other RDD	570	331	518
Total	1210	871	1158

Research in the Division of Human Genetics at UCT has suggested that there are several founder effects underlying STGD in the South African Afrikaner population (September et al. 2004). This research revealed that five common mutations, namely c.4469G>C,

c.1885C>T, c.6079C>T, c.768G>T, and c.454C>T were observed in affected South Africans at frequencies noticeably different to those found in European populations. Four of the mutations (c.4469G>C, c.1885C>T, c.768G>T, and c.454C>T) were found to be rare in European populations (Lewis et al. 1999; Maugeri et al. 1999; Maugeri et al. 2002; Rosenberg et al. 2007). It should be noted that more than 90% of individuals in the September et al. (2004) study were of Afrikaner descent. This is probably due to the fact that individuals affected with STGD that have been recruited into the RDD research program at UCT are mainly of Caucasian Afrikaner ancestry. This scenario could have arisen from a historical ascertainment bias, due to referrals from special schools who deal with visually impaired individuals or be a reflection of the high incidence in the population group. These schools were previously restricted to individuals of certain ethnic groups (September et al. 2004).

The Caucasian population of South Africa originated mainly from the immigration of groups of individuals over the past few centuries from various European countries to South Africa. The Afrikaner population or Afrikaans-speaking Caucasians originated mainly from three countries, namely, Holland, Germany, and France (Ramesar et al. 2001; September et al. 2004). The large number of Afrikaners affected with STGD may be due to a founder effect or effects resulting from the immigration of the above mentioned groups of individuals into the Cape Province more than 330 years ago and their subsequent migration between the years 1835 to 1843 further into the interior of Southern Africa. The movement is known as “Die Groot Trek” (Botha and Beighton 1983; September et al. 2004). STGD is not the only heritable disease associated with the Afrikaner population; a very high prevalence of other genetic diseases has also been found in this group, including familial hypercholesterolemia and variegate porphyria (Gevers et al. 1987).

Traditional mutation screening methods have been expensive and laborious as the *ABCA4* gene is large in size (50 exons). Thus, in order to facilitate rapid, cost effective screening of large numbers of DNA samples from subjects with suspected AARs, Asper Ophthalmics in Estonia developed the ABCR400 microarray chip (Jaakson et al. 2003; Allikmets 2007). This microarray chip currently tests for 558 known variants in the

ABCA4 gene (http://www.asperbio.com). Investigations into the efficiency of the ABCR400 array revealed it to be more than 98% effective in identifying variants that were already present on the chip (Jaakson et al. 2003). Of all possible alleles associated with the *ABCA4* gene, the microarray identified 54-60% in patient screening studies. It should be noted that since the chip carries out direct sequencing at each of these queried positions, it does allow for the detection of novel variants at all positions included on the microarray. The ABCR400 microarray chip is constantly being expanded, with novel changes or mutations being added regularly (Jaakson et al. 2003).

The ABCR400 microarray was designed by using the arrayed primer extention (APEX) technology. This technology is essentially a direct sequencing reaction using a solid base for support. Briefly, this method involves the use of sequence-specific oligonucleotides which have been modified at their 5' end. These oligonucleotides are arrayed onto a glass slide and are usually designed in such a way that their 3' end flanks the variable site of interest. Specific target nucleic acids are prepared by the polymerase chain reaction (PCR) and annealed to the oligonucleotides on the glass slide. Subsequently, the addition of a DNA polymerase enzyme allows the sequence-specific extension of the 3' ends of primers with fluorescently-labelled dideoxynucleotide triphosphates (ddNTPs) (Jaakson et al. 2003).

ABCR400 genotyping in a South African cohort of RDD-affected individuals conducted by the Division of Human Genetics at UCT led to the detection of causative mutations in over 60% of the cohort (Roberts et al. 2009). The cohort consisted of 132 DNA samples from patients with a clinical diagnosis of STGD, FFM, or CRD. The majority of the cohort consisted of individuals diagnosed with STGD. The DNA samples of individuals were sent to Asper Ophthalmics in Estonia for genotyping on the microarray in order to: 1) validate partial results already obtained in the Division of Human Genetics at UCT (with the objective of obtaining both causative mutations), and 2) identify the two causative mutations in samples that had not been previously investigated. The screening results revealed that of the 132 samples tested, 85 were found to carry two mutations, 30 were found to carry one mutation, and 17 were found to carry no mutations. However, in order to assess the clinical utility of the microarray screening, the results for samples

where two mutations were identified were verified using locally developed diagnostic assays. These assays revealed only 80 individuals with bi-allelic causative mutations, which thus emphasises the importance of validating the microarray results. To date a total of 181 DNA samples from individuals affected with an AAR have been sent for ABCR400 microarray screening; of which 100 samples have been completely characterised (i.e. bi-allelic mutations identified). In a total of 66 of these samples, both mutations were part of the seven most prevalent mutations (unpublished data). ABCR400 genotyping has resulted in a total of 67 different mutations being identified in the South African cohort (unpublished data).

In addition to the results obtained by September et al. (2004), the results acquired from the microarray screening identified an additional two mutations (c.5461-10T>C and c.2588G>C) that were found to also occur at a high frequency in affected individuals. Again, the majority of these mutations have been found in Caucasian individuals of Afrikaner descent. Seven of the 67 mutations were found to account for approximately 70% of all *ABCA4* mutations detected (unpublished data). These seven *ABCA4* mutations are listed in the Table 1.3 below.

Table 1.3 The seven common *ABCA4* mutations in South African patients with STGD.

Nucleotide change	Amino Acid change	Exon/Intron	Mutation Type	Reference
c.4469G>A	C1490Y	Exon 30	missense	Zhang et al. (1999)
c.1885C>T	R602W	Exon 13	missense	Webster et al. (2001)
c.6079C>T	L2027F	Exon 44	missense	Webster et al. (2001)
c.768G>T	V256V	Exon 6	splice site	Webster et al. (2001)
c.5461-10T>C	IVS38-10T>C	Intron 38	unknown	Klevering et al. (2005)
c.2588G>C	G863A	Exon 17	missense	Maugeri et al. (1999)
c.454C>T	R152X	Exon 5	termination	Zhang et al. (1999)

For the purpose of the current study, the seven mutations investigated will be referred to according to their amino acid change.

1.6 Aims and objectives of this project

Previous attempts at defining the genotype-phenotype correlation have been restricted by (i) limited number of families with the same mutations, (ii) limited number of families with the same combination of mutations for comparison, and (iii) inadequate phenotypic/ophthalmologic data. The presence of a significant cohort of South African STGD patients with the same common mutations in various combinations could allow the improved understanding of the relationship between mutation combination and patient prognosis. The investigation of the genotype-phenotype correlation, together with pathogenicity predictions of the mutations involved, could lead to increased knowledge regarding the impact of each mutation, thus enhancing the clinical utility of identifying *ABCA4* mutations.

The aim of this project is to investigate the seven common *ABCA4* mutations in South African individuals. The objectives will be to 1) determine the carrier frequencies of the mutations in a control cohort of Caucasian individuals, particularly those of Afrikaner descent, 2) assess the pathogenicity of the mutations using bioinformatic tools, and 3) gain more insight into the genotype-phenotype correlation of *ABCA4* and manifestations of STGD disease.

The proposed experimental plan is as follows:

1. To create reproducible, sensitive assays to detect the seven most common *ABCA4* mutations, using a multiplex SNaPshot reaction and allele-specific PCR (AS-PCR) assays.
2. To use the designed assays to screen for the seven *ABCA4* mutations in 72 unrelated individuals with STGD who have not been previously analysed with the ABCR400 microarray. The majority of these individuals are of Caucasian ancestry.

3. To use the designed assays to screen 269 Caucasian controls (169 of which are of Afrikaner ancestry) for the seven *ABCA4* mutations, to determine the frequency at which these mutations occur in the control population.
4. To use bioinformatic tools to investigate the functionality of each of the seven *ABCA4* mutations individually, in order to determine the predicted pathogenicity of each mutation.
5. To combine the results obtained from this screening with results from previous South African *ABCA4* mutation screening studies and to identify individuals with two of the same (homozygous) or two different (compound heterozygous) common mutations. The mutation combination will be correlated with the clinical data available for each affected individual. The clinical information will include: date of birth, gender, ethnicity, AOO, age at examination (progression), visual acuity, doctor's diagnosis, and any other clinical details available. Statistical analysis will be used to investigate the influence of the *ABCA4* mutations on the clinical phenotype of the STGD individuals (genotype-phenotype correlation).

The assays designed for this project will utilise the existence of common mutations in South Africans and may be employed in the future to target or direct testing for both research and diagnostic purposes. This could potentially reduce the testing costs to the research laboratory and patient, respectively. Determining the carrier frequencies for the seven common *ABCA4* mutations in the South African population may be useful in genetic counselling when risk prediction is discussed with patients and their family members.

Chapter 2: Materials and Methods

2.1 Cohort selection

The individuals selected for this project were divided into two groups, the patient/subject cohort and control cohort.

The patient cohort comprised of subjects affected with AARs namely STGD, FFM, CRD, and arRP. This cohort was further sub-divided into:

i) 86 subjects from 66 families who, through previous analysis using the ABCR400 microarray chip (Asper Ophthalmics in Estonia), had been shown to carry two of the seven common *ABCA4* mutations. A total of 73 individuals were diagnosed with STGD, three with FFM, eight with CRD, and two with arRP. All subjects were of Caucasian ancestry.

ii) 72 subjects who were unrelated and who had not been previously screened using the ABCR400 microarray chip. A total of 70 individuals were diagnosed with STGD and two individuals with CRD. A total of 60 individuals were Caucasian, eight were of Mixed Ancestry, three were Indian, and one was indigenous Black African.

The available demographic and clinical information of each patient included gender, ethnicity, date of birth, AOO, the ophthalmologic diagnosis, age at examination, and full details of visual acuity. This information was collated in order to be used for the latter part of this study (genotype-phenotype correlation).

Screening for the seven common *ABCA4* mutations was performed on patient DNA samples where patients had been recruited already for the broader RDD project. Informed consent was obtained from patients according to the tenets of the Declaration of Helsinki (2008). The DNA consent forms (Appendix 1) and patient information sheets used in the RDD project had previously been approved by the UCT Research

Ethics Committee (UCT URC REC REF 168/99 amended 2003 and 2005). Ethics approval for this current study was also obtained, and subsequently revised and renewed until 15 March 2011 (UCT URC REC REF 186/2009).

The control cohort comprised of 269 individuals of Caucasian ethnicity who 1) had not been specifically clinically assessed for the presence of RDDs, and 2) had not been questioned regarding a family history of RDDs (but were not part of the RDD research cohort). A total of 169 individuals were specifically of Afrikaner descent. Individuals were defined as being Afrikaner on the basis of their family name and stated ancestry upon recruitment. Furthermore, for 119 of these 169 individuals, family pedigrees indicated that this sub-cohort has been in South Africa for two or more generations. The pedigrees verified Afrikaner descent, showing the individuals to be of European origin, particularly North Western, with distinct input from the Dutch, German, and French population groups.

2.2 Determination of DNA integrity

Due to the fact that they can deteriorate over time, the quality and quantity of the DNA samples used for this project was assessed prior to molecular analysis.

2.2.1 Spectrophotometry

The concentration of each DNA sample was measured using a Nanodrop 1000 Spectrophotometer (Nanodrop Technologies Inc., Wilmington, USA). The samples were then diluted to a working concentration of 100ng/µl using SABAX distilled H_2O (dH_2O) (Adcock Ingram, Johannesburg, SA).

2.2.2 Agarose gel electrophoresis

Appendix 2 contains all recipes used for agarose gel electrophoresis.

In order to assess whether any DNA samples had compromised integrity, 200ng of diluted DNA samples together with 3µl of loading dye were loaded onto 1% agarose gels that contained ethidium bromide (EtBr) (0.6µg/ml) (Appendix 2). Agarose gel electrophoresis was conducted at a voltage of 160 volts (V) for 20 minutes (min).

The DNA was visualised by ultraviolet transillumination using the Uvipro Gold transilluminator (UVITec, Cambridge, UK). A DNA sample that appears as a smear is degraded, while a distinct band of high molecular weight indicates good integrity (Fig. 2.1).

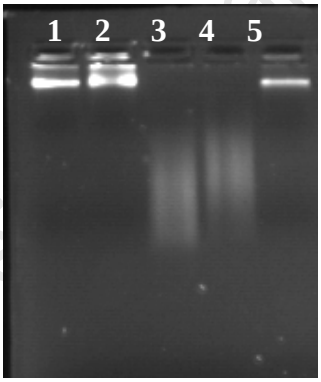


Fig. 2.1 A digital photograph of a 1% agarose gel, showing an example of DNA sample integrity, visualised with EtBr.

Lanes 1, 2, and 5 contain DNA samples with a tight, bright band near the top of the agarose gel, indicating intact DNA. By comparison, lanes 3 and 4 contain DNA samples appearing as a smear down the agarose gel, revealing degraded DNA.

2.3 Primer design

2.3.1 Gene annotation

The genomic sequence of *ABCA4* was obtained from two websites: Ensembl (http://www.ensembl.org/index.html) and National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). The reference number for the *ABCA4* genomic sequence is ENSG00000198691 for the Ensembl website and NG_009073.1 for the NCBI website. The gene sequence was annotated using the Annot v9 programme designed by Dr G. Rebello (2006). This gene annotation programme indicates the locations of the intron-exon boundaries as well as the single nucleotide polymorphisms (SNPs) in the genomic sequence. The positions of the mutations of interest were identified manually. Appendix 3 contains sequence annotations for the seven exons interrogated in the current study, indicating the locations of the seven mutations.

2.3.2 Primer information

The primers were synthesized using the Oligo 1000M DNA Synthesizer v4.20 (Beckman Instruments Inc., Department of Molecular and Cell Biology, UCT). Working stocks of 20μM (10μM for the internal SNaPshot primers) were diluted from the primer stock concentrations using dH_2O (Adcock Ingram, Johannesburg, SA). The diluted primers were stored at 4°C or -20°C. In the case of the internal SNaPshot primers, following their synthesis each primer was purified using high performance liquid chromatography (HPLC) (Department of Molecular and Cell Biology, UCT).

2.3.3 External primer design

Seven forward and reverse primer sets had been designed to amplify the exons containing the mutations of interest in the *ABCA4* gene in a PCR (Table 2.1). These external primers had been previously designed by other researchers and stored in the Division of Human Genetics at UCT. Standardly, primers were designed flanking exons, at least 20bp from intron-exon boundaries, having a length of approximately 18 to 22 nucleotides. In addition, primers were designed to have a melting temperature (T_M) of between 50°C and 60°C with both forward and reverse primer pairs having a similar T_M, to possess a maximum of three nucleotide repeats (max poly X), and to have a molar percentage of guanine (G) and cytosine (C) nucleotides (GC content) approaching 50%.

The programmes used to design the external PCR primers were:

i. Primer 3 (Whitehead Institute for Biomedical Research, 2004)
(http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)
This web-based programme was used to select adequate primers from an input sequence.

ii. IDT (integrated DNA technologies) SciTools OligoAanalyzer 3.0
(http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx)
This web-based programme was used to analyse the designed primers for the formation of any possible secondary structures such as hairpins, homodimers, and heterodimers.

iii. NCBI BLAST (Bethesda, USA)
(http://www.ncbi.nlm.nih.gov/)
This web-based programme was used to identify any genomic regions found to be similar to the sequence of the designed primers. Primers showing a great amount of non-specificity were excluded.

Table 2.1 Information pertaining to the primers used to amplify the exons containing the seven *ABCA4* mutations.

Exon	Nucleotide change	Amino acid change (Mutation)	Primer Name (Sense/Antisense)	Primer Length (bp)	Primer Sequence
5	c.454C>T	R152X	Sense	19	5' -gacccatttccccttcaac- 3'
			Antisense	18	5' -aggctgggtgcttccctc- 3'
6	c.768G>T	V256V	Sense	19	5' -ggtgtctttcctaccacag- 3'
			Antisense	20	5' -aggaatcaccttgcaattgg- 3'
13	c.1885C>T	R602W	Sense	19	5' -agctatccaagcccgttcc- 3'
			Antisense	19	5' -ccattagcgtgtcatggag- 3'
17	c.2588G>C	G863A	Sense	21	5' -ctgcggtaaggtaggataggg- 3'
			Antisense	21	5' -cacaccgtttacatagagggc- 3'
30	c.4469G>A	C1490Y	Sense	19	5' -gtcagcaactttgaggctg- 3'
			Antisense	18	5' -tccctctgtggcaggcag- 3'
39	c.5461-10T>C	IVS38-10T>C	Sense	18	5' -gccccacctgctgaagag- 3'
			Antisense	18	5' -tcccagctttggacccag- 3'
44	c.6079C>T	L2027F	Sense	20	5' -gaagcttctccagccctagc- 3'
			Antisense	20	5' -tgcactctcatgaaacaggc- 3'

2.3.4 SNaPshot primer design

SNaPshot primers were designed to bind one base pair away from the mutation of interest following the primary PCR using external primers (ABI PRISM® SNaPshot™ Multiplex Kit Protocol, Applied Biosystems). Only one primer is required to be designed per mutation for a SNaPshot reaction. The primer can either be designed to the sense or antisense strand of the DNA sequence, depending on i) whether a single nucleotide polymorphism (SNP) is present in the nucleotide sequence to which the designed primer would bind and ii) which strand shows better results based on the web-based primer analysis programmes. A SNaPshot primer should be on average 25bp in length. However, in order to design primers to be used in a multiplex SNaPshot reaction, the SNaPshot primers need to differ by at least four to six nucleotides to allow distinction between the final SNaPshot products. Modification of the length of a primer can be achieved by the addition of nucleotide tails to the 5' end of the sequence using

random nucleotides that will not interfere with primer binding and specificity (Table 2.2).

The programmes used to design the internal SNaPshot primers were:

i. IDT SciTools OligoAnalyzer 3.0 (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx)
This web-based programme was used in the same manner as in the design of the external primers for a regular PCR reaction. This programme analyses the primer sequences for possible secondary structure formation, such as hairpins, homodimers, and heterodimers.

ii. Local BLAST application in BioEdit Sequence Alignment Editor 7.0.5.2 (Hall 1999)
This analysis programme was used to align the chosen SNaPshot primer sequence with the resulting amplified region from the PCR reaction using external primers. This was performed in order to exclude primers showing non-specific homology with other regions in the primary amplicon. In order to assess primers to be used in a multiplex SNaPshot reaction, a local database, which contains the nucleotide sequences of each primary amplicon, is created in BioEdit. Each designed SNaPshot primer is then aligned to each amplicon in the local database, ensuring that there is no non-specific binding of the primers to the other amplicons in the multiplex SNaPshot reaction.

iii. AutoDimer 1.0. (Vallone and Butler 2004) (http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm)
For primer design for a multiplex SNaPshot reaction, AutoDimer was used to align each designed internal SNaPshot primer against one another. An input file containing the sequences of all the SNaPshot primers was created and this was examined in the AutoDimer programme with a minimum score requirement of three or four. The minimum score requirement is used to establish which primer

sequence interactions are stored to the output file. Thus, an interaction between two primers that has a score less than the specified minimum score requirement will not be displayed in the programme. Ideally, a SNaPshot primer should not bind to the other SNaPshot primers at more than four consecutive nucleotide positions. Specifically, a SNaPshot primer should not bind to another SNaPshot primer on their 3' end, in order to avoid non-specific single base extention of the primer. In the case of the addition of random nucleotide tails, AutoDimer was used to investigate the acceptability of the designed SNaPshot primers both with their additional random nucleotide tails and without the random nucleotide tails.

Table 2.2. Information pertaining to the SNaPshot primers designed to be used for the SNaPshot multiplex reaction.

Exon	Nucleotide change	Amino acid change (Mutation)	Primer Name (Sense/Antisense)	Primer Length (bp) -with tail	Primer Sequence (with tail)*
6	c.768G>T	V256V	Sense	50	5' -attcgtactcatgccattccatgcca **ggacttcttcaagctcttccgtgt**- 3'
			Antisense	56	5' -tatcgtactcatgccattccaca **gagcagccaaaccctcccttac**- 3'
13	c.1885C>T	R602W	Antisense	34	5' -tgttccagtgccacg **aacccgccccagatgtacc**- 3'
30	c.4469G>A	C1490Y	Antisense	32	5' -cttcgtggttac **tgagcttctccctggtgctg**- 3'
39	c.5461-10T>C	IVS38-10T>C	Sense	37	5' -ccgatgtagttga **ccccgtttccaacagtcctacttc**- 3'
44	c.6079C>T	L2027F	Antisense	41	5' -tactctggatcttagta **ggtaaagatgttctcgtcctgtga**- 3'

*****The nucleotide sequence written in bold is the sequence designed to bind complementary to the genomic DNA sequence. The nucleotide sequence not written in bold is the random nucleotide tail added to the primer sequence.

2.3.5 Allele-specific primer design

2.3.5.1 General allele-specific primer design

The allele-specific primers had been previously designed by other researchers and stored in the Division of Human Genetics at UCT. Allele-specific primer design involves creating one mutation-specific and one wild-type specific primer (Table 2.3).

Table 2.3. Information pertaining to the primers designed to be used for AS-PCR for R152X and G863A mutation detection.

Exon	Nucleotide change	Amino acid change (Mutation)	Primer Name	Primer Length (bp)	Primer Sequence*
5	c.454C>T	R152X	R152X-Rc	24	5'-**atctttcaagatatcccttattcg**- 3'
			R152X-Rt	45	5' –ttaaaaaacgctctgtcatac **atctttcaagatatcccttattca**- 3'
17	c.768G>T	G863A	G863A-Rc	26	5' -ttttt**gaagtggggttccatagtcag**- 3'
			G863A-Rg	36	5' –gcgtgcttggggtat **gaagtggggttccatagtcac**- 3'

R152X-Rc: The AS-PCR primer designed specifically to bind to the nucleotide, cytosine (C), when exon 5 does not contain the R152X mutation, and is therefore known as the wild type primer. **R152X-Rt**: The AS-PCR primer designed specifically to bind to the nucleotide, thymine (T), when exon 5 contains the R152X mutation, and is therefore known as the mutation-specific primer. **G863A-Rc**: The AS-PCR primer designed specifically to bind to the nucleotide, cytosine (C), when exon 17 does not contain the G863A mutation, and is therefore known as the wild type primer. **G863A-Rg**: The AS-PCR primer designed specifically to bind to the nucleotide, guanine (G), when exon 17 contains the G863A mutation, and is therefore known as the mutation-specific primer.

*****The nucleotide sequence written in bold is the sequence designed to bind complementary to the genomic DNA sequence. The nucleotide sequence not written in bold is the random nucleotide tail added to the primer sequence.

Both the mutation-specific and wild-type primers were designed with their most 3' nucleotide being complementary to the mutant nucleotide or wild type nucleotide, respectively (Hoti et al. 2009). Random nucleotides shown not to interfere with primer binding and specificity were added to one or both primers to allow for size-based differentiation between alleles with and without a mutation. Both mutation-specific and wild-type primers are designed to the same strand of the DNA sequence. These primers can either be designed to the sense or antisense strand of the DNA sequence, depending on:

i) the mutation involved, since certain 3' mismatches affect PCR differently. G:T mismatches more tolerable in comparison to other mismatches, and therefore should not be used in allele-specific PCR (AS-PCR) (Kwok et al. 1994).

ii) whether a SNP is present in the nucleotide sequence to which the designed primer would bind.

iii) which strand shows better results based on the web-based primer analysis programmes.

Once the allele-specific primers have been designed, the external primer is chosen. The external primer binds on the opposite strand to the one on which the mutation-specific and wild-type primers bind in order to obtain the appropriate allele-specific amplification. The suitability of the three primers (mutation-specific primer, wild type primer, and external primer) for use in an AS-PCR is determined using some of the programmes used for external primer design and SNaPshot primer design.

The programmes used to design primers for AS-PCR were:

i. NCBI Blast. (http://www.ncbi.nlm.nih.gov/), and
ii. AutoDimer 1.0.

NCBI Blast was used in the same way as for external primer design, identifying any genomic regions found to be similar to the sequence of the primers. If a great amount of non-specificity was identified, the primers were excluded. Autodimer was used in the same way as for internal SNaPshot primer design. In AutoDimer, the primers designed to be used for the AS-PCR were aligned against one another. This was conducted for the primers with and without their random nucleotide tails.

2.3.5.2 Allele-specific primer design using locked nucleic acids

In cases when allele-specific PCR (AS-PCR) was not specific, allele-specific primers containing a locked nucleic acid (LNA) was used. A LNA is a DNA analog, possessing similar chemical properties to DNA, thus allowing for primer annealing using standard

AS-PCR (Mouritzen et al. 2003). However, DNA sequences that contain a LNA nucleotide display a much higher affinity and specificity to bind to its complementary DNA sequence in comparison to a DNA sequence with no LNA residue. This appears to be due to a change in the interaction between the enzyme, DNA polymerase, and the LNA nucleotide which has a much more rigid structure (Latorra et al 2003a). In allele-specific primers a LNA is incorporated on the 3' end of the primer sequence, which replaces both the wild type and mutant-specific nucleotides. These primers were designed using the same procedure as described for general allele-specific primer design.

2.4 Detection of the seven *ABCA4* mutations

2.4.1 Polymerase chain reaction

PCR was used in order to amplify the regions surrounding the seven mutations of interest. These reactions were performed on the GeneAmp® PCR System 9700 (Applied Biosystems, California, USA), PXE 0.2 Thermocycler (Applied Biosystems, California, USA) or PX2 Thermocycler (Thermohybaid, Middlesex, UK).

The optimisation of amplification of the seven genomic regions, each one containing one mutation, included:

(i) a temperature gradient - conducted on PX2 Thermocycler (Thermohybaid, Middlesex, UK), and

(ii) a magnesium titration.

i. Temperature gradient.

This involved setting up a PCR with primer annealing temperatures (T_a) that ranged from 55˚C to 62˚C. The temperature gradient was used to determine the

annealing temperature at which the designed external primers would optimally anneal to the genomic DNA, in order to obtain the best product yield. A higher annealing temperature, leads to increased primer specificity.

ii. Magnesium titration.

This involved the use of different magnesium chloride ($MgCl_2$) buffers from the PCR optimization kit (Roche Diagnostics GmbH, Mannheim, Germany). A $MgCl_2$ titration was used to determine the magnesium concentration at which the primers anneal best to the genomic DNA. Generally, a decreased concentration of magnesium leads to the increase in PCR specificity, thus helping to decrease non-specific primer annealing. A magnesium titration was conducted in order to optimise a multiplex PCR for the amplification of four exons of the *ABCA4* gene, namely 13, 30, 39, and 44. The expected product size for each exon was 267bp, 286bp, 245bp, and 287bp, respectively (Fig. 2.2).

Although a $MgCl_2$ titration was performed in order to optimise a multiplex PCR for the amplification of four exons of the *ABCA4* gene, in most cases only three PCR products instead of the expected four were visualised by agarose gel electrophoresis. The magnesium titration in this case was not entirely successful for the optimisation of the multiplex PCR. This resulted in the attempted single multiplex PCR being separated into two PCR reactions, and is discussed later on.

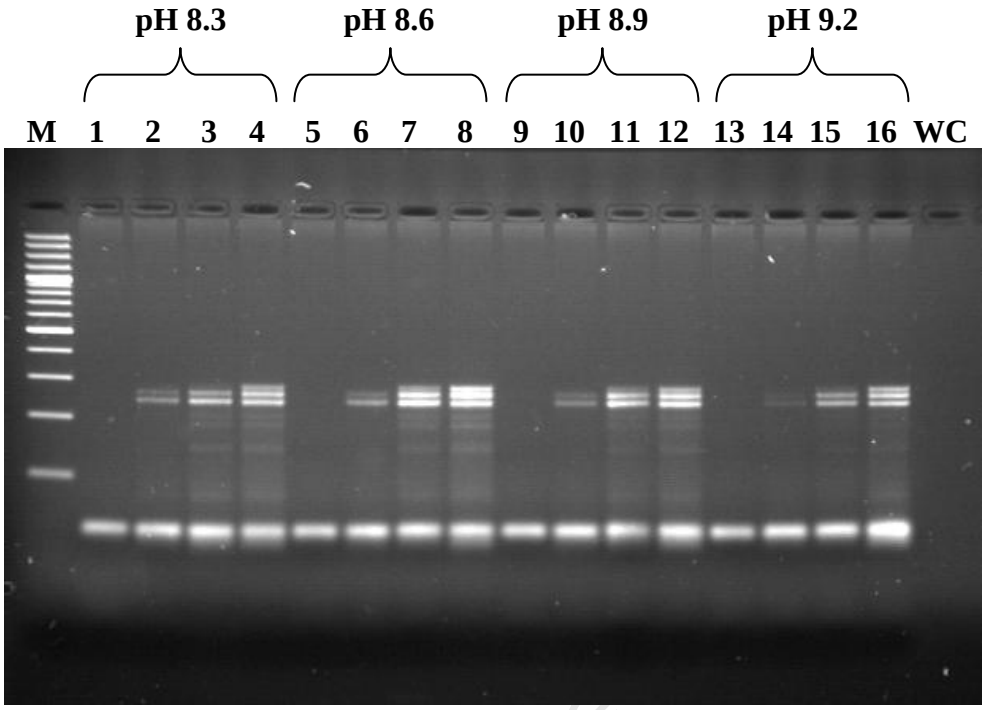


Fig. 2.2 A digital photograph of a 2% agarose gel showing an example of a MgCl2 titration conducted for optimisation of a multiplex PCR including exons 13, 30, 39, and 44, visualised with EtBr.

Lane M - GeneRuler™ 100bp Plus DNA ladder (Fermentas Life Sciences, Hanover, USA); lane 1 - 1.0mM; lane 2 - 1.5mM; lane 3 - 2.0mM; lane 4 – 2.5mM; lane 5 - 1.0mM; lane 6 - 1.5mM; lane 7 - 2.0mM; lane 8 - 2.5mM; lane 9 - 1.0mM; lane 10 - 1.5mM; lane 11 - 2.0mM; lane 12 - 2.5mM; lane 13 - 1.0mM; lane 14 – 1.5mM; lane 15 – 2.0mM; and lane 16 - 2.5mM; lane WC – Water control. Lanes 1, 5, 9, and 13, containing a MgCl2 concentration of 1.0mM at all four pH values, yielded no PCR product. Lanes 4, 8, 12, and 16, containing a MgCl2 concentration of 2.5mM all four pH values, yielded three PCR products. The best PCR products were seen in lane 16, and this buffer was chosen to be used for optimisation of the multiplex PCR for the amplification of the four exons.

2.4.1.1 PCR conditions

PCR was generally conducted in a 25µl reaction volume using 200µM deoxynucleotide triphosphates (dNTPs) (Bioline, Michigan, USA), 1X reaction buffer containing 1.5mM $MgCl_2$ (Colourless GoTaq™ Reaction Buffer, Promega, Madison, USA), 0.5 units (U) GoTaq® DNA polymerase (Promega, Madison, USA), 10pmol of each external primer (sense and antisense), and 100ng or 200ng of genomic DNA. The remaining volume was made up by using SABAX dH_2O (Adcock Ingram, Johannesburg, SA).

The general PCR thermal cycling conditions included an initial denaturation step at 94°C for 3 min, followed by 30 cycles at 94°C for 30 seconds, T_a for 30 seconds, and 72°C for 40 seconds, followed by a final elongation step at 72°C for 5 min.

The alterations made to either the PCR components or cycling conditions in order to obtain optimal PCR products are listed below:

1. Amplification of the exons containing four of the seven mutations of interest were conducted in two multiplex PCR reactions. One multiplex reaction amplified exons 30 (containing the C1490Y mutation site) and 39 (containing the IVS38-10T>C mutation site). The second multiplex reaction amplified exons 13 (containing the R602W mutation site) and 44 (containing the L2027F mutation site) (Fig. 2.3). The PCR components and cycling conditions for both multiplex reactions were the same. PCR was performed using 2.5mM $MgCl_2$ (pH 9.2) (Roche Diagnostics GmbH, Mannheim, Germany), 3pmol of each external primer (sense and antisense for each mutation in the multiplex), and 200ng genomic DNA. The cycling conditions included an initial denaturation step at 94°C for 3 min, followed by 28 cycles at 94°C for 30 seconds, 63°C for 30 seconds, and 72°C for 40 seconds, followed by a final elongation step at 72°C for 5 min.
2. PCR was performed using 200ng genomic DNA for the amplification of exon 6 (containing the V256V mutation site). The cycling conditions included an initial denaturation step at 94°C for 3 min, followed by 30 cycles at 94°C for 30 seconds, 57.9°C for 30 seconds, and 72°C for 40 seconds, followed by a final elongation step at 72°C for 5 min.

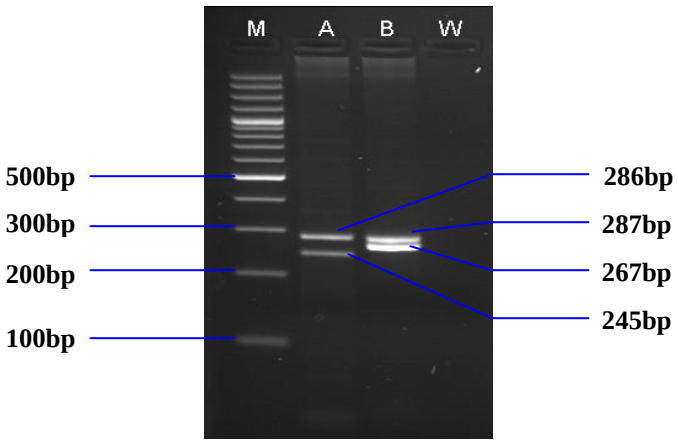


Fig. 2.3 A digital photograph of a 2% agarose gel showing the products of two multiplex PCR's, visualised with EtBr.

Lane M represents the GeneRuler™ 100bp Plus DNA ladder (Fermentas Life Sciences, Hanover, USA), Lane A represents the first of two multiplex PCRs, Lane B represents the second of the two multiplex PCRs, and Lane W represents the water control. The 287bp, 286bp, 267bp, and 245bp products are representative of the amplified exons 44 (L2027F), 30 (C1490Y), 13 (R602W), and 39 (IVS38-10T>C), respectively.

3. The detection of the G863A mutation in exon 17 was performed using an AS-PCR. Optimised conditions included 10pmol forward external primer, 15pmol G863A-Rc primer, 5pmol G863A-Rg primer, and 200ng genomic DNA. The cycling conditions included an initial denaturation step at 94°C for 3 min, followed by 35 cycles at 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 40 seconds, a final elongation step at 72°C for 5 min, followed by an additional denaturation step 95°C for 5 min. On completion of the PCR, the products were placed directly on ice for 5 min. These two additional steps (additional 5 min denaturation and snap-freeze step) were added to reduce the non-specific binding initially visualised by agarose gel electrophoresis.

4. Initially, numerous attempts were made to optimise an AS-PCR to detect the R152X mutation in exon 5, however they all proved to be unsuccessful. Optimisation attempts included the use of temperature gradients, magnesium gradients, DMSO gradients, glycerol gradients, and the replacement of the 3' nucleotide in the mutation-specific and wild-type primer sequences with a LNA residue. As mentioned previously, the incorporation of an LNA residue in place

of a normal DNA residue into the primer appears to result in a much higher binding affinity and specificity of the primer to its complementary DNA sequence (Latorra et al 2003b). As a result of unsuccessful AS-PCR optimisation, basic PCR (with no LA residue) was performed to amplify exon 5 using 100ng genomic DNA. The cycling conditions involved an initial denaturation step at 94°C for 3 min, followed by 28 cycles at 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 40 seconds, followed by a final elongation step at 72°C for 5 min.

The PCR products obtained from the two multiplex PCR reactions were electrophoresed on a 2% agarose gel. The PCR products for exons 5 and 6 were electrophoresed on a 1.5% agarose gel, and visualised with EtBr (Results not shown). A GeneRuler™ 100bp Plus DNA ladder (Fermentas Life Sciences, Hanover, USA) was used to ensure that the correct PCR product size was obtained for each sample (Appendix 2). The PCR products obtained from the AS-PCR for exon 17 were not electrophoresed on an agarose gel as their product size difference was only 10bp. Instead, the products obtained from the AS-PCR for the detection of the G863A mutation in exon 17 were resolved by denaturing HPLC (dHPLC) using size-based separation.

2.4.2 SNaPshot reaction

SNaPshot PCR is used as a form of SNP genotyping and is essentially an adaptation of the basic PCR procedure, requiring a template, primers, Taq polymerase, and reaction buffer. The SNaPshot method involves an initial primary PCR to amplify the region that contains the SNP or mutation of interest. The SNaPshot reaction uses PCR product as the template, in comparison to DNA in a normal PCR, and a single primer (forward or reverse internal primer) instead of a pair of primers (forward and reverse external primers). Instead of incorporating dNTPs, SNaPshot utilises the incorporation of fluorescently labelled ddNTPs, allowing the single base extension on the 3' end of the internal SNaPshot primer (ABI PRISM® SNaPshot™ Multiplex Kit Protocol, Applied Biosystems).

The results can be viewed using the Applied Biosystems (ABI) 3100 Genetic Analyser (Applied Biosystems, California, USA) (ABI PRISM® SNaPshot™ Multiplex Kit Protocol, Applied Biosystems). Each of the four ddNTPs that are incorporated in the SNaPshot reaction are labelled with different fluorescent dye (Table 2.4). Therefore, depending on which ddNTP is incorporated, peak colours specific for each ddNTP will be used in the data analysis.

Table 2.4 A table containing a list of the fluorescently labelled dyes for each ddNTP, as well as the colour observed during data analysis.

Incorporated ddNTP	Name of Fluorescent Dye Label	Colour of the Analysed Data
Adenine (A)	dR6G	Green
Cytosine (C)	dTAMRA™	Black
Guanine (G)	dR110	Blue
Thymine (T)	dROX™	Red

2.4.2.1 Primary PCR purification

As described previously, two multiplex primary PCR's were conducted in order to amplify four exons, each of which contained a mutation site of interest. One multiplex reaction amplified exons 30 (containing the C1490Y mutation site) and 39 (containing the IVS38-10T>C mutation site). The second multiplex reaction amplified exons 13 (containing the R602W mutation site) and 44 (containing the L2027F mutation site). The PCR products obtained from the two multiplex PCR reactions were pooled using equal volumes of each prior to purification treatment.

Following the primary PCR, the obtained amplified products were purified using a shrimp alkaline phosphatase (SAP) (Promega, Madison, USA) and Exonuclease I (*E. Coli*) *Exo I* (New England Biolabs Inc., USA) treatment. This purification method is

used to remove the remaining dNTPs and primers that were not depleted during the primary PCR step. The SAP (Promega, Madison, USA) functions in the dephosphorylation of phosphates on the 5' end of DNA sequences (http://www.promega.com/catalog/catalogproducts.aspx?categoryname=productleaf_1478). *Exo I* (New England Biolabs Inc., USA) functions in a 3' to 5' direction and is involved in removing nucleotides from the remaining single-stranded DNA sequences, therefore enhancing the degradation of the external primers in the obtained in the PCR mixture (http://www.neb.com/nebecomm/products/productm0293.asp).

This post-PCR purification was conducted in a 10µl and 20µl reaction volume for multiplex and single-plex reactions, respectively. Single-plex reactions were set-up when the multiplex reaction was unsuccessful. A 10µl Multiplex reaction volume consisted of 2.5U of SAP (Promega, Madison, USA) and 2U of *Exo I* (New England Biolabs Inc., USA), with the remaining volume being the pooled multiplex PCR products. A 20µl single-plex reaction volume consisted of 1.5U of SAP (Promega, Madison, USA), 1U of *Exo I* (New England Biolabs Inc., USA), 5µl PCR product, with the remaining volume being made up using SABAX dH_2O (Adcock Ingram, Johannesburg, SA).

The primary PCR purification procedure involved thermal cycling conditions at 37°C for one hour, followed by an enzyme inactivation step at 75°C for 15 min.

2.4.2.2 SNaPshot PCR conditions

Following primary PCR purification, a SNaPshot mutiplex reaction was conducted using all four SNaPshot primers, for the detection of C1490Y, IVS38-10T>C, R602W, and L2027F mutations. The four SNaPshot primers were pooled in an empirically determined ratio of 0.2:2:2:0.5 for R602W, C1490Y, IVS38-10T>C, and L2027F, respectively. The optimised amount of each primer included in the SNaPshot reaction was determined by performing various primer titrations.

Multiplex SNaPshot PCR was conducted in a reaction volume of 10µl, consisting of 5µl pooled, purified PCR product, 1µl pooled SNaPshot primers (using a working stock of 10µM), and 1µl SNaPshot Multiplex Ready Reaction Mix (Applied Biosystems, Warringon, UK). The remaining volume was made up by using SABAX dH_2O (Adcock Ingram, Johannesburg, SA). A negative control consisting of a reaction in SABAX dH_2O (Adcock Ingram, Johannesburg, SA) and no pooled, purified PCR product was included in each multiplex SNaPshot PCR set-up. Single-plex SNaPshot reactions were conducted for four samples known to be heterozygous for each of the four mutations, respectively. These single-plex reactions were performed each time a batch of samples were screened for the four mutations using the multiplex assay and were used as positive controls. When repeated attempts on a sample failed to show all four mutation results, due to less amplification of a certain exon, subsequent single-plex SNaPshot reactions were used. Single-plex SNaPshot PCR was also conducted in a 10µl reaction volume, consisting of 1µl PCR product, 1µl single SNaPshot primers specific for mutation of interest (using a working stock of 10µM), and 1µl SNaPshot® Multiplex Ready Reaction Mix (Applied Biosystems, Warringon, UK). The remaining volume was made up by using SABAX dH_2O (Adcock Ingram, Johannesburg, SA).

The cycling conditions for both the multiplex and single-plex SNaPshot PCRs were as follows: 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 30 seconds. This was conducted on the PX2 Thermocycler (Thermohybaid, Middlesex, UK). Samples were stored at 4°C.

2.4.2.3 Post-SNaPshot purification

Following SNaPshot PCR reaction, the SNaPshot products were purified using SAP (Promega, Madison, USA). An amount of 1U of SAP (Promega, Madison, USA) was added to the entire SNaPshot reaction, in order to remove unincorporated ddNTPs. The post-SNaPshot purification procedure involved thermal cycling conditions at 37°C for one hour, followed by an enzyme inactivation step at 75°C for 15 min. Once the reaction

was completed, the products were stored in the dark, in order to prevent a reduction in the amount of fluorescence of each product due to degradation.

2.4.2.4 Analysis of SNaPshot results using capillary electrophoresis on the ABI 3100 Genetic Analyser

Once the post-PCR purification was completed, the products were resolved on the ABI 3100 Genetic Analyser (Applied Biosystems, California, USA). A volume of 0.5µl of purified SNaPshot product, together with 9µl of Hi-Di™ Formamide (Applied Biosystems, Warrington, UK) and 0.1µl of GeneScan™ -120LIZ™ size standard (Applied Biosystems, Warrington, UK) was pipetted into a MicroAmp 96-well reaction plate (Applied Biosystems, Warrington, UK) (Appendix 2). The function of the Hi-Di™ Formamide (Applied Biosystems, Warrington, UK) is to maintain the SNaPshot products in a single-stranded conformation, while the GeneScan™ -120LIZ™ size standard (Applied Biosystems, Warrington, UK) is used to estimate the size of the alleles obtained from the SNaPshot reaction. The mixture consisting of the sample, Hi-Di™ Formamide (Applied Biosystems, Warrington, UK) and GeneScan™ -120LIZ™ size standard (Applied Biosystems, Warrington, UK) in the reaction plate were denatured at 95 ºC for 5 min, and cooled on ice for a further 5 min. The reaction plate was loaded onto the ABI 3100 Genetic Analyser (Applied Biosystems, California, USA). The settings for data capture were as follows: the dye set used for the analysis was E5, the analysis method used was GeneScan-120 liz, and the run module was SNP36_POP4. The SNaPshot data were analysed using the Genemapper 3.0 Genescan Software (Applied Biosystems, California, USA).

Although the expected internal primer sizes for the detection of the R602W, C1490Y, IVS38-10T>C and L2027F mutations was 32, 34, 37, and 41bp respectively, the analysed SNaPshot results were not visualised at those exact locations. Instead they were visualised at 36/37bp, 38/39bp, 40/41.5bp, and 45/46bp, respectively. A possible reason

for these observed shifts could be due the specific ddNTP that was incorporated, as each of the four ddNTPs contained a different fluorescent dye.

2.4.2.5 Trouble-shooting: incorporation of V256V

Attempts were conducted to incorporate the detection of the V256V mutation into the multiplex SNaPshot reaction along with the C1490Y, R602W, L2027F, and IVS28-10T>C mutations. Initially, a SNaPshot primer specific for the detection of the V256V mutation was designed to bind to the antisense strand of exon 6. The V256V mutation is a splice mutation caused by a G>T nucleotide. Although numerous optimisations were attempted, the SNaPshot results for both the multiplex and single-plex reactions for a sample known to be heterozygous for V256V produced oscillating black and green peaks, representative of a C and A allele, respectively (Fig. 2.4). When optimisation was attempted using a known wild type sample (no V256V mutation), oscillating black peaks were visible (results not shown). A SNaPshot primer specific for the detection of the V256V mutation was designed to bind to the sense strand of exon 6, but SNaPshot results for both the multiplex and single-plex reactions for a sample known to be heterozygous for V256V produced oscillating blue and red peaks, representative of a G and T allele, respectively (Fig. 2.5). When optimisation was attempted using a known wild type sample (no V256V mutation), oscillating blue peaks were visible (results not shown). Optimisation attempts included i) altering the annealing temperature used during the SNaPshot reaction, which was attempted by setting up a temperature gradient ii) using a different post-PCR purification procedure (Appendix 2), and iii) excluding a post-PCR purification procedure step, in order to determine whether or not this was affecting the subsequent SNaPshot reaction.

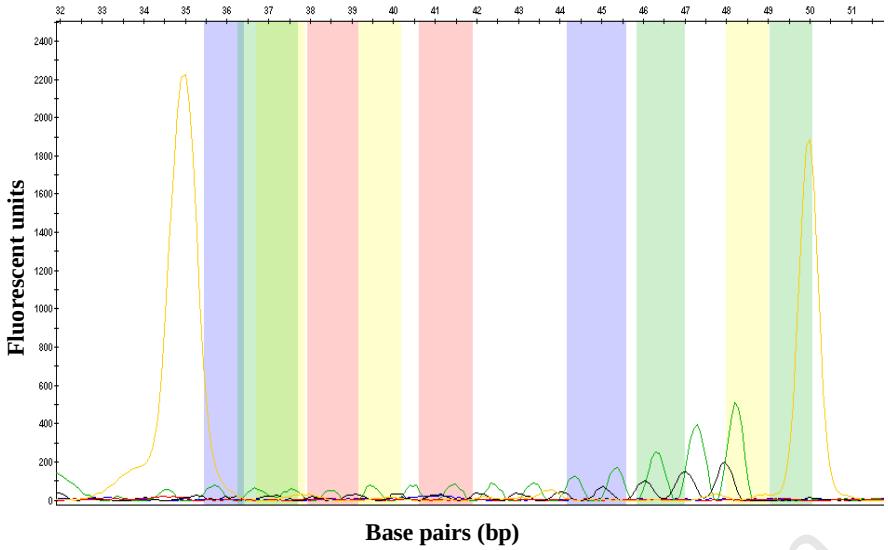


Fig. 2.4 An electropherogram of the results obtained for a single-plex SNaPshot reaction to detect the V256V mutation using a SNaPshot primer specific for the antisense strand of exon 6 in the *ABCA4* gene.

The product size in bp is shown on the x-axis. The level of fluorescence is shown on the y-axis. The expected SNaPshot size for the known heterozygous sample was 46bp with a single green and black peak. However, the results revealed oscillating green and black peaks, across the region between 34bp and 49bp instead.

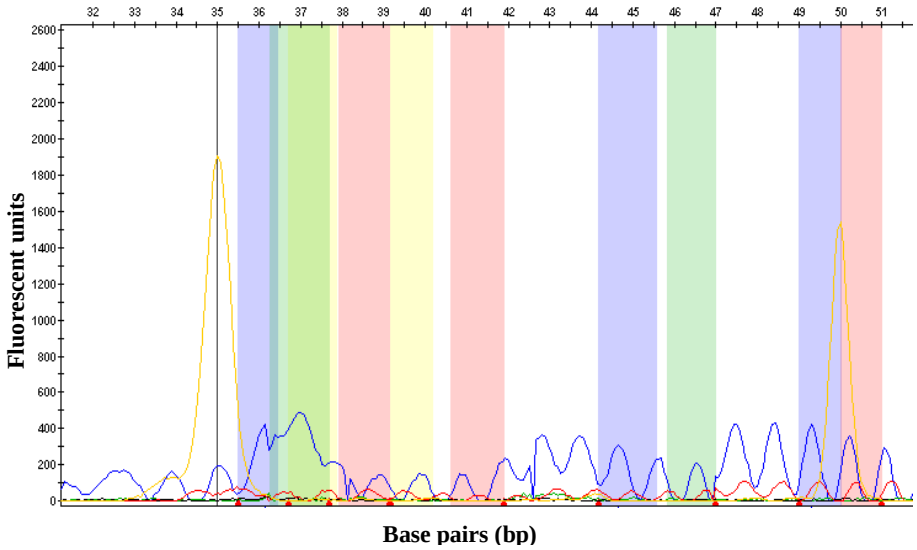


Fig. 2.5 An electropherogram of the results obtained for a single-plex SNaPshot reaction to detect the V256V mutation using a SNaPshot primer specific for the sense strand of exon 6 in the *ABCA4* gene.

The product size in bp is shown on the x-axis. The level of fluorescence is shown on the y-axis. The expected SNaPshot size for the known heterozygous sample was 50bp, with a single blue and red peak. However, the results revealed oscillating blue and red peaks, across the region between 34bp and 51bp instead.

Incorporation of a fifth SNaPshot primer for the detection of the V256V proved unsuccessful. SNaPshot primers designed to bind to both the sense and antisense strand did not yield the desired specificity for V256V mutation detection. After numerous attempts, the V256V mutation was optimised to be detected using mutation detection by dHPLC analysis.

2.4.3 Allele-specific PCR

AS-PCR is used to differentiate between a wild type and mutant allele. The technique is based on the 3' positioning of a nucleotide in one primer sequence that is complementary to the wild type allele, and the 3' positioning of a nucleotide in the other primer sequence that is complementary to the mutant allele. In theory, only the primer consisting of the correctly matched 3' nucleotide should bind to the appropriate allele and extend (Latorra et al. 2003b). DNA Taq polymerase does not have 3' to 5' exonuclease activity, and 3' end mismatched nucleotides are extended less often than that of matched 3' end nucleotides (Kwok et al. 1994).

As mentioned in section 2.3.5, three primers are required for an AS-PCR, an external primer, a mutation-specific primer and a wild type-specific primer. The mutation-specific and wild-type primers bind to one strand, and the external primer binds to the opposite strand.

General AS-PCR was used to detect for the G863A mutation, using size-based separation, due to the addition of random nucleotide tails at the 5' end of the wild type and mutation-specific primer sequences. The difference in size between the wild type- and mutation-specific PCR products was 10bp. For this reason, screening for the G863A mutation was detected using non-denaturing size-based separation conditions on the WAVE® Nucleic Acid Fragment Analysis System (Transgenomic Inc, USA) (see section 2.4.4).

General AS-PCR was attempted to detect the R152X mutation, with an expected PCR product of 84bp for a wild type allele, and 105bp for an allele containing the R152X mutation. This technique, however, was found to result in non-specific binding of the mutation-specific primer, resulting in the formation of two PCR products for samples known to be wild type, the amplification of which should therefore only produce a single band (Fig.2.6). As mentioned previously, allele-specific primers can be designed by placing a LNA residue on the 3' end of the primer. The replacement of a normal 3' end DNA nucleotide with a LNA residue at the mutation site has been shown to improve the discrimination between wild type and mutant alleles (Latorra et al. 2003b; Mouritzen et al. 2003). An AS-PCR using a 3' LNA nucleotide in the sequences for both the wild type and mutation-specific primers was thus attempted. Optimisation procedures for the AS-PCR with the 3' LNA residue, however, also proved unsuccessful.

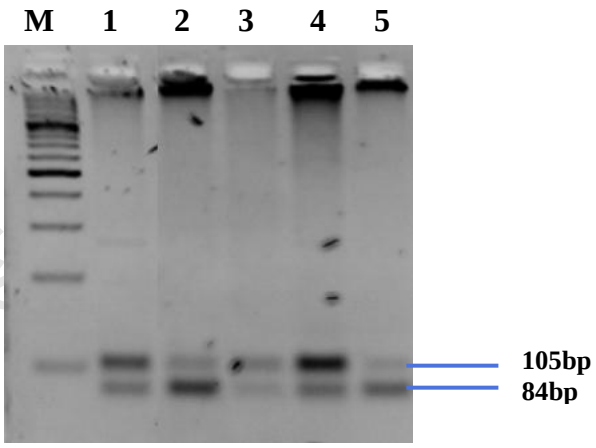


Fig. 2.6 An inverted digital photograph of a 1.5% agarose gel showing the products of from an AS-PCR to specifically detect the R152X mutation in exon 5, visualised with EtBr.

Lane M represents the GeneRuler™ 100bp Plus DNA ladder (Fermentas Life Sciences, Hanover, USA), Lanes 1, 3, and 4 represent AS-PCR results from known heterozygous samples for the R152X mutation, whereas Lanes 2 and 5 represent AS-PCR results from known wild type samples that do not contain an R152X mutation. A slight non-specific mutant allele at 105bp was visualised in both wild type samples in lanes 2 and 5.

2.4.4. Denaturing high performance liquid chromatography

dHPLC was used to detect sequence differences between two alleles or chromosomes, in particular the presence or absence of some of the mutations of interest investigated in this study. This technique was performed using the WAVE® Nucleic Acid Fragment Analysis System (Transgenomic Inc, USA) and can detect single base pair substitutions, and small deletions and insertions (Xiao and Oefner 2001).

The dHPLC technique involves the binding of DNA molecules to a DNASep™ column - which is known as the stationary phase - and the subsequent separation of the molecules from the column. DNA molecules separate from the column, due to increasing concentrations of an organic solvent, called acetonitrile. This is known as the mobile phase (Xiao and Oefner 2001).

There are three modes in which the WAVE® system can operate, namely fully-denaturing, partially-denaturing, and non-denaturing. Fully-denaturing conditions are used to estimate the size of single stranded DNA molecules, such as primers, and will not be discussed further as it did not fall into the scope of this project. Partially-denaturing conditions are used for mutation detection and SNP discovery, with the basis of strand separation being dependent on i) the size of the fragment, ii) the nucleotide composition of the fragment, and ii) the temperature(s) at which the fragment will optimally denature or melt. Non-denaturing conditions are used for the sizing of double-stranded DNA molecules; strand separation is thus dependent on size but not sequence (WAVE® System Operator's Guide).

2.4.4.1 Mutation detection (partially-denaturing)

Prior to using partially denaturing conditions for mutation detection, the sequence of the specific PCR product being investigated is analysed using WAVEMAKER® software. The software is used to predict the temperature(s) at which a DNA sequence will

partially denature, thus resulting in the partial separation of the two DNA strands in the helical structure. This is known as the helical fraction. Ideally, it is optimal to melt the sequence at a helical fraction of between 70 and 90 percent. At the specified temperature, heteroduplex PCR products elute first, due to their lower affinity for the column as a result of the mis-match in their sequence (WAVE® System Operator's Guide). Before the commencement of mutation detection, samples are heated to 95°C for 5 min, followed by slow cooling for approximately an hour to allow for the re-annealing of the alleles. Samples that are heterozygous for a mutation have a ratio of one-to-one for wild type-to-mutant allele; therefore heating will result in a combination of homoduplexes and heteroduplexes. Samples that are wild type, containing no mutation, will consist of only homoduplexes. Samples that are homozygous for a mutation will show the same profile as a wild type sample, since both alleles consist of the same DNA sequence. Thus, samples showing a wild type profile need to be mixed with a known wild type control sample and heated and cooled again in order to reveal whether it is actually wild type or not (WAVE® System Operator's Guide). This is known as the "WAVE mixing experiment".

Screening for the V256V and R152X mutations was conducted using mutation detection under partially-denaturing conditions. Since the positions of the mutations were known, temperatures specific to the exact position of the mutation were used to melt the PCR product. Various parameters were optimised for both the V256V and R152X mutation detection methods in order to adequately visualise the profile difference between a sequence with and without the mutation. The optimised parameters included: i) the ideal temperature for discriminating between a sequence containing a mutation and a wild type sequence (temperature), ii) the rate at which the sample should be eluted from the column (flow rate), iii) the time period during which the sample should be eluted from the column (gradient duration), and iv) the amount of sample injected onto the column (volume of PCR product injected) (Table 2.5).

Table 2.5 Optimised parameters for mutation detection using dHPLC for mutations V256V and R152X.

Mutation	PCR Product Size (bp)	Temperature (°C)	Flow Rate (ml/min)	Gradient Duration (min)	Volume of PCR Product injected (µl)
V256V	294	61.6	1.5	3	5
R152X	230	56.5	1.5	2	5

A wild type control and mutation positive control sample (heterozygous for the mutation) were included for every batch of test samples that was run by mutation detection using dHPLC for comparison.

The samples which showed altered profiles compared to the wild type control after mutation detection using dHPLC, were selected for sequence analysis. This is essential in order to confirm the presence of a mutation at the specific position rather than a SNP or mutation in the nearby vicinity.

2.4.4.2 Size-based separation (non-denaturing)

Generally, the WAVE® system is able to distinguish between fragments that differ by one percent of the size of the PCR product, for example, a PCR product of 200bp can be distinguished from a PCR product of 202bp (WAVE® System Operator's Guide). PCR products that are smaller in size elute first in comparison to larger ones, which is a result of less phosphates being present in the smaller sized fragment. Detection of the G863A mutation was completed by AS-PCR. The PCR products obtained from the AS-PCR were visualised by size-based separation under non-denaturing conditions at 50°C. The primers designed for AS-PCR for G863A mutation detection contained a random nucleotide tail, which lead to a size difference of 10bp between the wild type and mutant alleles, which was visible using the WAVEMAKER® system.

A wild type control and mutation positive control sample (heterozygous for the mutation) were included for every batch of test samples that was run by size-based separation using dHPLC for comparison.

2.4.5 Cycle sequencing

Cycle sequencing was used to determine the sequence changes causing altered dHPLC chromatogram profiles, in order to determine whether or not a mutation was indeed present in a sample.

Cycle sequencing was conducted directly on the products obtained from PCR. No purification of the products obtained from the PCR's was required as no primer dimers or non-specific products were visible in the PCR product when they were electrophoresed on an agarose gel.

Cycle sequencing was performed using the BigDye® terminator cycle sequencing kit version 3.1 (Applied Biosystems, California, USA). Sequencing was conducted in a 20µl reaction volume using 3-5µl PCR product, 2µl termination mix (Applied Biosystems, California, USA), 4µl BigDye termination buffer (Applied Biosystems, California, USA), and 20pmol of the external primer (sense or antisense). The sense primer was used for the V256V mutation detection, whilst the antisense primer was used for the R152X mutation detection.

The sequencing thermal cycling conditions included an initial denaturation step at 96°C for 5 min, followed by 30 cycles of 95°C for 30 seconds, 50°C for 30 seconds, and 60°C for 4 min. The protocol for the purification of the products following cycle sequencing is described in Appendix 2.

Sequencing resolution on the ABI 3100 Genetic Analyser (Applied Biosystems, California, USA) involved loading 5µl purified sequencing product, together with 8µl Hi-Di^{TM} Formamide (Applied Biosystems, Warrington, UK) into a 96-well reaction plate. The reaction plate was loaded onto the ABI 3100 Genetic Analyser (Applied Biosystems, California, USA). The settings for data capture were as follows: the dye set used for the analysis was Z, the analysis method used was BC-3100POP4UR_seqOffFtOff.saz, and the run module was ultraseq36_POP4 def. Analysis was conducted using the ABI Prism™ DNA Sequencing Analysis Software.

2.5 Family co-segregation analysis

Following screening of the patient cohort for the seven common mutations, family studies were conducted (wherever possible, depending on the availability of samples) for the patients in whom two *ABCA4* mutations were found. Only the two mutations identified in the affected probands were screened for in specific family members for whom a DNA sample was available. Where one or two mutations were identified in the affected individual using the multiplex SNaPshot assay, single-plex SNaPshot reactions were performed in the family studies for the specific mutation(s) detected. DNA samples from a patient's family (usually from the mother, father, or siblings) are collected to screen for the patient's mutations, once they have been identified in order to determine whether or not: (i) the mutations were co-segregating with disease within the family and (ii) the two mutations were transmitted in *cis*/*trans*. If an *ABCA4* allele was found to carry more than one mutation (i.e. in *cis*) it was designated as a complex allele and not classified as disease-causing on its own. The 86 individuals, for whom two disease-causing mutations were identified with the ABCR400 microarray, had been verified and family co-segregation studies were conducted prior to the initiation of this project.

2.6 Bioinformatics

A significant aspect of the current study involved attempting to determine whether there was any genotype-phenotype correlation in individuals found to carry two disease-associated *ABCA4* mutations. In this regard, an attempt was made to assess the pathogenicity of each of the seven *ABCA4* mutations in this study using several bioinformatic tools/resources.

Amino acid substitutions resulting from missense mutations (i.e. C1490Y, R602W, L2027F, and G863A) were investigated using four bioinformatic tools: a polymorphism phenotyping programme, PolyPhen (www.bork.embl-hiedelberg.de/PolyPhen); Sorting Intolerant From Tolerant (SIFT) (www.blocks.fhcrc.org/sift/SIFT.html); PANTHER PSEC (Thomas et al. 2003); and PMUT (http://mmb2.pcb.ub.es:8080/PMut/). Mutations thought to affect splicing were investigated using different programmes compared to missense mutations. The V256V mutation, which is located in exon 6, was investigated using the programme ESEfinder (http://exon.cshl.edu/ESE/), which is specific for exonic mutations. The IVS38-10T>C mutation, which is located in intron 38, was investigated using the Berkeley Drosophila Genome Project (BDGP) Splice Site Prediction programme (http://www.fruitfly.org/seq_tools/splice.html).

2.6.1 A polymorphism phenotyping programme, PolyPhen

The required input data for the PolyPhen (www.bork.embl-hiedelberg.de/PolyPhen) server is the amino acid sequence of the protein in conjunction with the position of the mutation in the sequence and the two amino acid variants involved (Ramensky et al. 2002). PolyPhen uses homologues of the input amino acid sequence via a blast search to determine whether the amino acid replacement is compatible with the range of substitutions that have been previously observed at that particular position. Reference sequences that have an identity of between 30-94% of the input sequence are used by

software to calculate the "profile matrix". The index is measured by the position-specific independent counts (PSIC). Profile scores of the matrix represent logarithmic ratios of the likelihood that an amino acid will occur at a specific site to the likelihood that this same amino acid will occur at any other site in the peptide sequence. PolyPhen calculates the value of the difference between the profile scores of both variants occurring at the polymorphic position. The output reveals the proportion of sequences that were aligned to the query protein at the specific position, which could be used to determine the reliability of the resulting profile score (Ramensky et al. 2002).

2.6.2 Sorting intolerant from tolerant (SIFT)

SIFT (www.blocks.fhcrc.org/sift/SIFT.html) predicts whether an amino acid change will alter protein function, based on both the position and type of the change (Ng and Henikoff 2003). This programme presumes that amino acids that are conserved within protein families are important; therefore if a change is detected at a well-conserved position it tends to be predicted as being damaging to the protein. This presumption also pertains to amino acid changes from, for example, a polar amino acid to a hydrophobic one. The programme uses the query protein sequence to choose homologous proteins, creating an alignment of these proteins together with the query protein. By making use of the amino acids that appear at each site in the sequence, SIFT calculates the probability of an amino acid replacement being tolerated, and this is determined in part by the most frequent amino acid governing the specific site. SIFT provides predictions for amino acid substitutions at every position in the sequence for all 20 amino acids. Conservation values for amino acids at certain positions in the protein vary from $\log_2 20$ (which equals a value of 4.32), when only one amino acid is seen at a specific position, to zero, when all possible amino acids are seen at a specific position. SIFT uses a median conservation value of 3.0. The median conservation value is an indication of the confidence of the prediction given for a certain amino acid substitution. The predictions of the function of an amino acid alteration on a protein with a higher median conservation score are less varied and therefore resulting in a greater false positive error. The output from SIFT is a score which represents how well an amino acid substitution is

predicted to be tolerated. An amino acid substitution is predicted as being harmful to the protein if the value is less than 0.05 (Ng and Henikoff 2003).

2.6.3 PANTHER PSEC

PANTHER PSEC is a programme that uses a collection of protein families and subfamilies to determine whether or not a specific amino acid can occur at a particular position in protein families that are evolutionarily related amongst various species. This programme is very similar to SIFT, however, PANTHER PSEC utilises a different model or algorithm. The likelihood that a specific variant will alter protein function is determined by the substitution position-specific evolutionary conservation (sub-PSEC) score. The sub-PSEC score is obtained from the likelihood of observing different amino acids using a PANTHER hidden Markov model (HMM). The HMM profile has a different amino acid change at each site in the profile, which is based on the array of amino acids observed in an alignment of multiple homologous sequences (Thomas et al. 2003; Thomas and Kejariwal 2004; Brunham et al. 2005).

2.6.4 PMUT

The PMUT (http://mmb2.pcb.ub.es:8080/PMut/) software functions at two levels: i) it gathers information from a local database containing mutation hotspots, and ii) it examines a query SNP in a particular protein. The input data required by PMUT is a protein sequence together with the position of the DNA variant. It can be specified to either analyse only one variant or to conduct a complete scan of all the possible variants that are found at the specific position. PMUT uses two neural networks that serve as predictor engines for mutational data (Ferrer-Costa et al. 2005). PMUT data outputs consist of:

- a pathogenicity index that ranges zero to one. Mutations are noted as pathological when the index is greater than 0.5.

- a confidence index that ranges from zero to nine. A value of zero is seen as low, whilst a value of 9 is seen as high.

2.6.5 Splicing mutations occurring in exonic regions

Mutations occurring in the exons of the *ABCA4* gene that appear to have an effect on splicing were investigated using a programme called ESEfinder (http://exon.cshl.edu/ESE/). This web-based programme scans for presumed exonic splicing enhancers (ESEs) in the given query sequence. Matrices which correspond to the motifs of four different Ser/Arg-rich (SR) proteins are used in this programme. SR proteins are composed of a family of splicing factors that are conserved and play various roles in the splicing pathway. ESEs act as binding sites for SR proteins. The human SR proteins are SF2/ASF, SC35, SRp40 and SRp55. The input data required for ESEfinder is the query nucleotide sequence. The output is given as a series of scores that have been designed in increments of one nucleotide. A score is considered to be a high value if it is larger than the threshold value set at the input stage. The default threshold values are as follows: SF2/ASF = 1.956, SC35 = 2.383, SRp40 = 2.670 and SRp55 = 2.676. For interpretation enhancement and standardisation, the results are displayed graphically, with the query sequence on the x-axis. Colour-coded bars represent the presence of a motif with a high score. The higher the motif score, the higher the bar. The width of the bar indicates the length and positioning of the motif (Cartegni et al. 2003).

2.6.6 Splicing mutations occurring in intronic regions

Mutations occurring in the introns of the *ABCA4* gene that appear to have an effect on splicing were investigated using a programme called the BDGP Splice Site Prediction programme (http://www.fruitfly.org/cgi-bin/seq_tools/splice.pl). The programme uses a neural network to screen for both 5' and 3' splice sites. Subsequently, intronic mutations in the *ABCA4* gene that didn't appear to alter splicing were investigated further by looking at the conservation of the specific nucleotide across different species. This was

perfomed by using the University of California Santa Cruz (UCSC) Genome browser (http://genome.ucsc.edu/index.html?org=Human&db=hg19&hgsid=154378218). This browser compares the input sequence to the genomes of different species. In order to determine whether or not the nucleotide was conserved amongst species, the tool "Blat" was used to map a portion of the sequence containing the mutation site to the specified Human genome (Human Mar. 2006, NCBI36/hg18, Assembly).

2.7 Statistical analysis

An attempt was made to correlate genotype status of the seven common *ABCA4* mutations with the phenotype of the subjects affected with an AAR. Statistical analysis, using the programme StatSoft Statistica (Release 9), was conducted in order to investigate this genotype-phenotype correlation. The analysis was performed using the genotyping results obtained from the mutation screening of the seven *ABCA4* mutations conducted in the current study as well as the results obtained from the ABCR400 microarray chip screening (collectively referred to as the total patient cohort). The combined results were analysed against phenotype data available for each patient. The clinical data available for the majority of the individuals was the AOO of the disease. AOO is the difference in years between birth date and the self-reported start of AAR symptoms (for example poor visual acuity and/or blurred vision and/or night blindness). Only patients who were identified to carry two of the seven mutations and who had a self-reported AOO were used in this analysis.

Non-parametric analysis using the Kruskal-Wallis test was conducted to analyse the data. Non-parametric tests may be used as an alternative to parametric tests when the assumption of normality is violated and therefore the data does not cluster around a mean or expected value. A histogram depicts whether the data is normally distributed. Normally distributed data are represented by a symmetrical, bell-shaped curve, whereas data that is not normally distributed is asymmetrical showing skewing to either the left or right of the graph. The resulting histogram generated from the data revealed non-

normal distribution and showed that the age distribution of the cohort was significantly different to what was expected, with a significant p-value of less than 0.01 (Fig. 2.7). This could be due to the fact that only a few individuals were identified with high AOO, since the AOO of an AAR is generally within the first two decades of life.

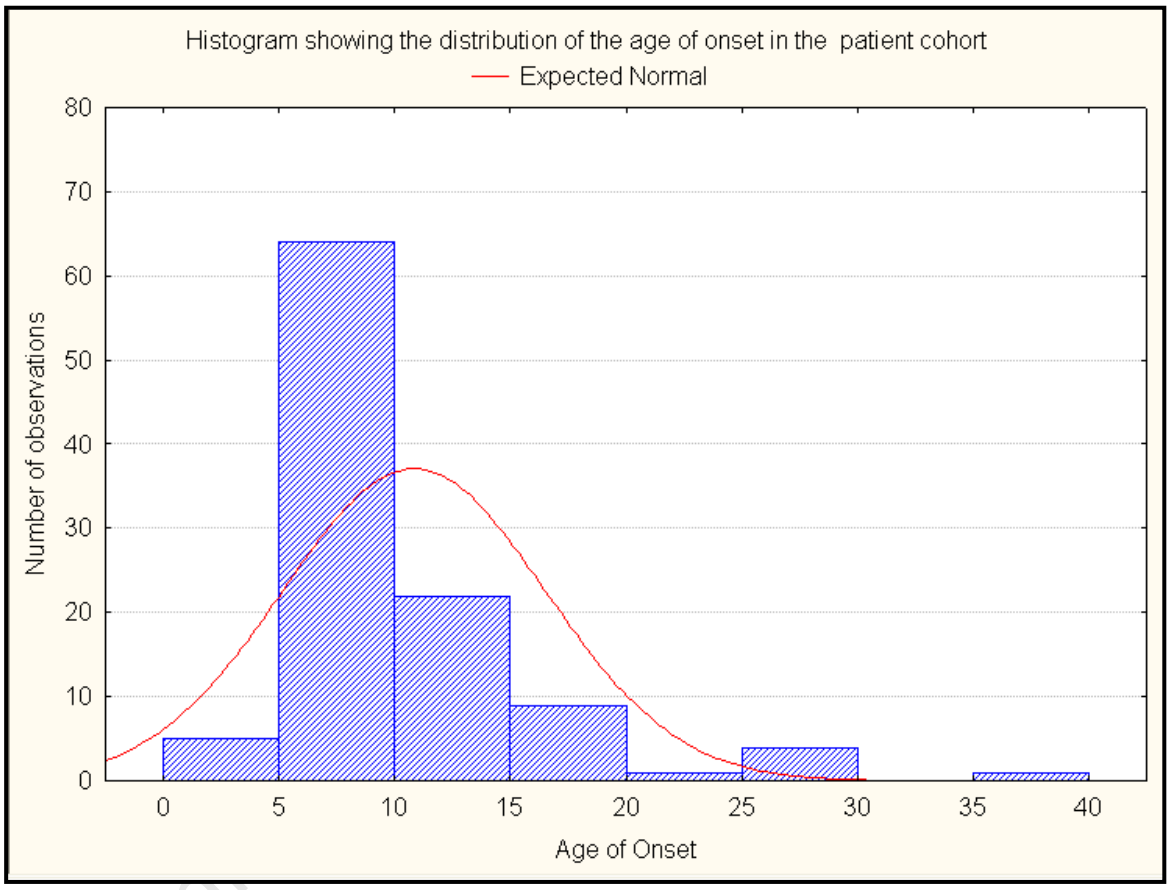


Fig. 2.7. A histogram showing the non-normal age distribution of the patient cohort.

The Kruskal-Wallis test can be used to test whether differences exist between the collection of paired mutation combinations of *ABCA4* variations. In order for the sampling distribution of the test statistic to be well-approximated by the chi-squared distribution, sample sizes per paired combination of disease-causing variations in *ABCA4* greater than or equal to five were included (Keller and Warrack 2000). AOO was used as the dependent variable (plotted on the y-axis in Fig. 2.7) in the analysis. The

various paired mutation combinations that the patients were found to carry (genotypes) were used as the independent variable (plotted on the x-axis in Fig. 2.7). Individuals with the same combination of two of the seven *ABCA4* mutations were grouped together and analysed against AOO.

The proposed null hypothesis (H_O) for the Kruskal-Wallis test was that in a South African cohort with a recessive AAR there is no significant difference between the median AOO for all paired mutation combinations in *ABCA4*. The proposed alternative hypothesis (H_A) was that in a South African cohort with a recessive AAR a significant difference between the median AOO exists between at least one pair of paired mutation combinations in *ABCA4*.

In order to reject the H_O, a *p*-value of 0.05 was used. In the case of this analysis a *p*-value of less than 0.05 would mean that the H_O would be rejected, and that H_A is assumed true (Keller and Warrack 2000). The Kruskal-Wallis test can only be used to determine whether a difference between the collection of paired mutation combinations of *ABCA4* variations (genotypes) exists. Therefore, if the Kruskal-Wallis test does produce a significant *p*-value, a subsequent set of tests need to be performed in order to determine exactly which pairs differ with respect to median AOO.

The Wilcoxon Rank Sum test for independent samples was used to conduct a pairwise comparison between the various paired mutation combinations (genotypes) against one another, in order to determine if there was a significant difference between each of the paired mutation combinations. Like the Kruskal Wallis testing, this test is conducted on non-normally distributed data, and is thus also a type of nonparametric test (Samuels and Witmer 1999; Keller and Warrack 2000).

The proposed null hypothesis (H_O) for the Wilcoxon Rank Sum test was that in a South African cohort with a recessive AAR there is no significant difference between the median AOO for the paired mutation combination **A** in *ABCA4* compared to the paired mutation combination **B** in *ABCA4*. The proposed alternative hypothesis (H_A) was that

in a South African cohort with a recessive AAR there is a significant difference between the median AOO of one paired mutation combination **A** in *ABCA4* compared to the paired mutation combination **B** in *ABCA4*. These test were conducted and compared all possible paired mutation combinations in *ABCA4*. Variation **A** represents one paired mutation combination and variation **B** represents a different paired mutation combination.

In the Wilcoxon Rank Sum test, paired mutation combinations in *ABCA4* consisting of three or more individuals were included in the analysis. This was conducted in order to determine if paired mutation combinations in *ABCA4* held any significance. To account for the amount of statistical comparisons performed on the data, adjustments for multiple tests needed to be conducted. This was performed using the Bonferroni correction, where the *p*-value of each pairwise comparison was adjusted for the total number of tests conducted (Green and Britten 1998).

Chapter 3: Results

3.1 Detection of the seven *ABCA4* mutations

Detection of C1490Y (exon 30), R602W (exon 13), L2027F (exon 44), and IVS38-10T>C (exon 39) mutations were successfully incorporated into a multiplex SNaPshot reaction. The V256V (exon 6) and R152X (exon 5) mutations were detected by partially-denaturing conditions using dHPLC analysis. The G863A mutation (exon 17) was screened for using AS-PCR, and visualised by size-based separation using dHPLC analysis.

Molecular testing for mutations in the *ABCA4* gene was conducted on 1) the test cohort of 72 patients affected with an AAR, and 2) a control cohort of 269 individuals (100 Caucasians and 169 specifically self-identified Afrikaners). Both cohorts were screened for the seven disease-causing *ABCA4* mutations (i.e. C1490Y, R602W, L2027F, IVS38-10T>C, V256V, G863A, and R152X). A summary of the results obtained from the patient and control screening is shown at the end of this section in Table 3.1.

3.1.1 C1490Y mutation detection

A sample was designated as heterozygous C1490Y by the presence of both a C and T allele, which were represented on the electropherogram by a black and red peak, respectively (Fig. 3.1). A sample was designated as homozygous C1490Y by the presence of a T allele, which was represented on the electropherogram by a red peak (no example shown). A sample was designated as wild type (i.e. not carrying a C1490Y mutation) by the presence of a C allele, which was represented on the electropherogram by a black peak (Fig. 3.2 and 3.3).

The C1490Y mutation was detected in the heterozygous state in 12/72 patient samples (16.67%) i.e. 12/144 alleles (8.33%). Homozygous C1490Y alleles were also detected in 2/72 individuals (2.77%) in the patient cohort, i.e. 4/144 alleles (2.77%).

The C1490Y mutation was detected in the heterozygous state in 2/269 (0.74%) individuals in the control cohort, i.e. 2/538 alleles (0.37%). The two heterozygotes (1.18%) were detected in the Afrikaner sub-cohort. This variant was only observed in the heterozygous state in the control cohort.

3.1.2 R602W mutation detection

A sample was designated as heterozygous R602W by the presence of both a G and A allele, which were represented on the electropherogram by a blue and green peak, respectively (Fig. 3.1). A sample was designated as homozygous R602W by the presence of an A allele, which was represented on the electropherogram by a green peak (no example shown). A sample was designated as wild type (i.e. not carrying a R602W mutation) by the presence of a G allele, which was represented on the electropherogram by a blue peak (Fig. 3.2 and 3.3).

The R602W mutation was detected in the heterozygous state in 10/72 patient samples (13.88%), i.e. 10/144 alleles (6.94%). This variant was only observed in the heterozygous state in the patient cohort.

The R602W mutation was detected in 2/269 individuals (0.74%) in the control cohort, i.e. 2/538 alleles (0.37%). The two heterozygotes (1.18%) were detected in the Afrikaner sub-cohort. This variant was only observed in the heterozygous state in the control cohort.

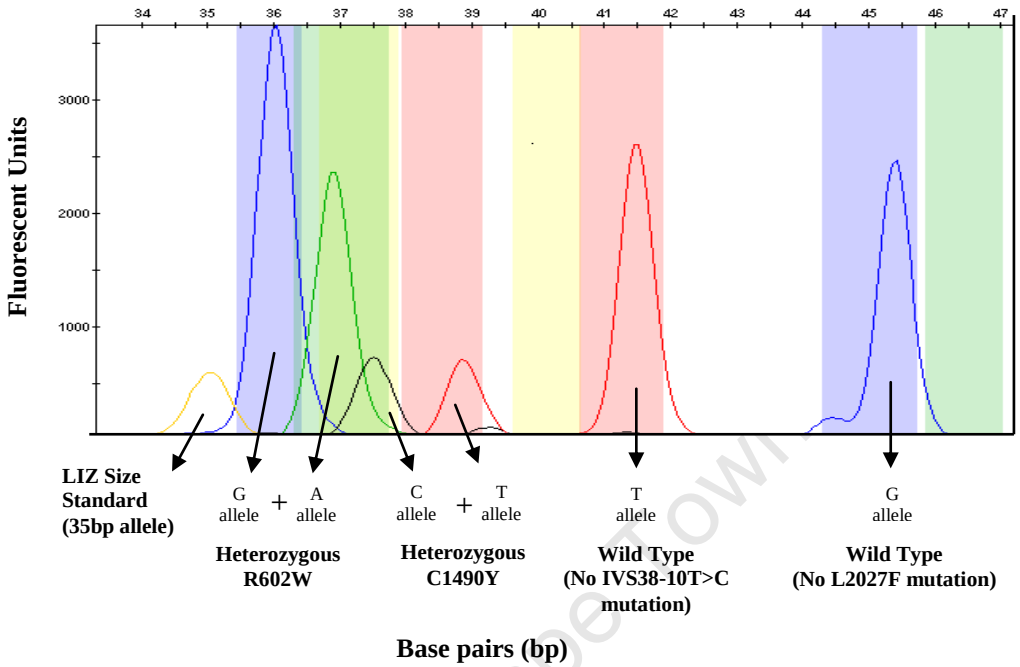


Fig. 3.1 An electropherogram of the multiplex SNaPshot reaction, showing the results obtained for a sample that is heterozygous for both the R602W and C1490Y mutations.

The R602W alleles are indicated by blue and green peaks at 36bp and 37 bp, respectively. The C1490Y alleles are indicated by black and red peaks at 38bp and 39bp, respectively. The IVS38-10T>C wild type allele is represented by a red peak at 41.5bp. The L2027F wild type allele is represented by a blue peak at 45bp. The LIZ Size Standard can be seen at 35bp.

3.1.3 L2027F mutation detection

A sample was designated as heterozygous L2027F by the presence of both a G and A allele, which were represented on the electropherogram by a blue and green peak, respectively (Fig. 3.2). A sample was designated as homozygous L2027F by the presence of an A allele, which was represented on the electropherogram by a green peak (no example shown). A sample was designated as wild type (i.e. not carrying a L2027F mutation) by the presence of a G allele, which was represented on the electropherogram by a blue peak (Fig. 3.1 and 3.3).

The L2027F mutation was detected in the heterozygous state in 12/72 patient samples (16.67%), i.e. 12/144 alleles (8.33%). This variant was only observed in a heterozygous state in the patient cohort.

The L2027F mutation was detected in the heterozygous state in 2/269 individuals (0.74%) in the control cohort, i.e. 2/538 alleles (0.37%). One heterozygote (0.59%) was detected in the Afrikaner sub-cohort and the other (1%) the Caucasian sub-cohort. This variant was only observed in a heterozygous state in the control cohort.

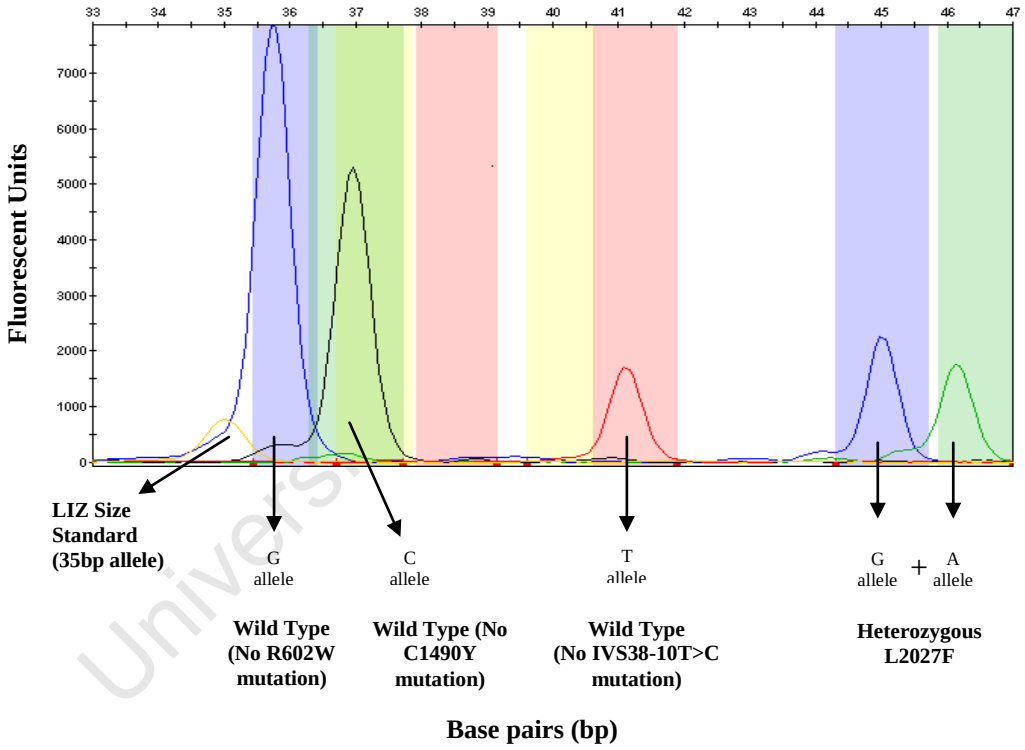


Fig. 3.2 An electropherogram of the multiplex SNaPshot reaction, showing the results obtained for a sample that is heterozygous for the L2027F mutation.

The R602W wild type allele is indicated by a blue peak at 36bp. The C1490Y wild type allele is indicated by a black peak at 38bp. The IVS38-10T>C wild type allele is represented by a red peak at 41.5bp. The L2027F alleles are represented by blue and green peaks at 45bp and 46bp, respectively. The LIZ Size Standard can be seen at 35bp.

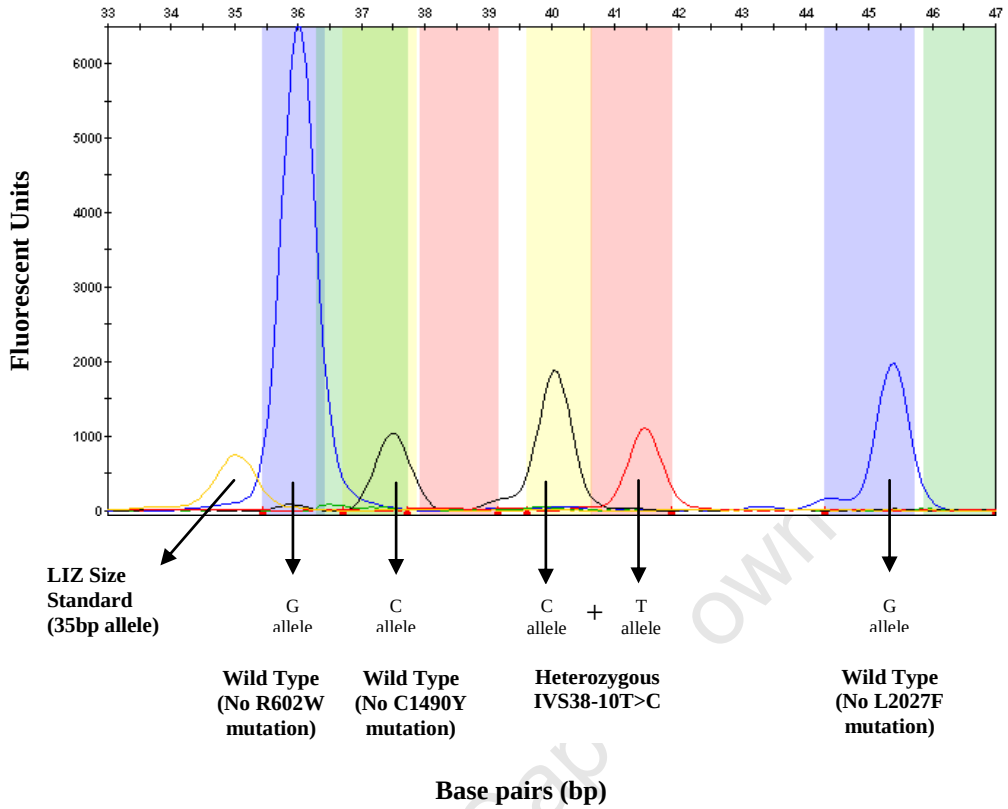


Fig. 3.3 An electropherogram of the multiplex SNaPshot reaction, showing the results obtained for a sample that is heterozygous for the IVS38-10T>C mutation.

The R602W wild type allele is indicated by a blue peak at 36bp. The C1490Y wild type allele is indicated by a black peak 38bp. The IVS38-10T>C alleles are represented by a black and red peak at 40bp and 41.5bp, respectively. The L2027F wild type allele is represented by a blue peak at 45bp. The LIZ Size Standard can be seen at 35bp.

3.1.4 IVS38-10T>C mutation detection

A sample was designated as heterozygous IVS38-10T>C by the presence of both a C and T allele, which were represented on the electropherogram by a black and red peak, respectively (Fig. 3.3). A sample was designated as homozygous IVS38-10T>C by the presence of a C allele, which was represented on the electropherogram by a black peak (no example shown). A sample was designated as wild type (i.e. not carrying an IVS38-10T>C mutation) by the presence of a T allele, which was represented on the electropherogram by a red peak (Fig. 3.1 and 3.2).

The IVS38-10T>C mutation was detected in the heterozygous state in 9/72 patient samples (12.5%), i.e. 9/144 alleles (6.25%). The mutation was detected in a homozygous state in 2/72 (2.77%) patients, i.e. 4/144 alleles (2.77%).

The IVS38-10T>C mutation was not observed in the control cohort of 269 individuals.

3.1.5 V256V mutation detection

After the "WAVE mixing experiment" (defined in section 2.4.4.1) was performed, mutation screening by dHPLC showed 11 patient samples to have the identical chromatogram profile to that of the positive control, while one sample was suspected to be homozygous V256V. Two control samples were found to have an identical chromatogram profile to that of the positive control. One control sample was of Caucasian ethnicity and the other was of Afrikaner descent. A different Afrikaner control sample (sample 2.13) was shown to have a different chromatogram profile to that of both the positive control and the negative control (Fig. 3.4).

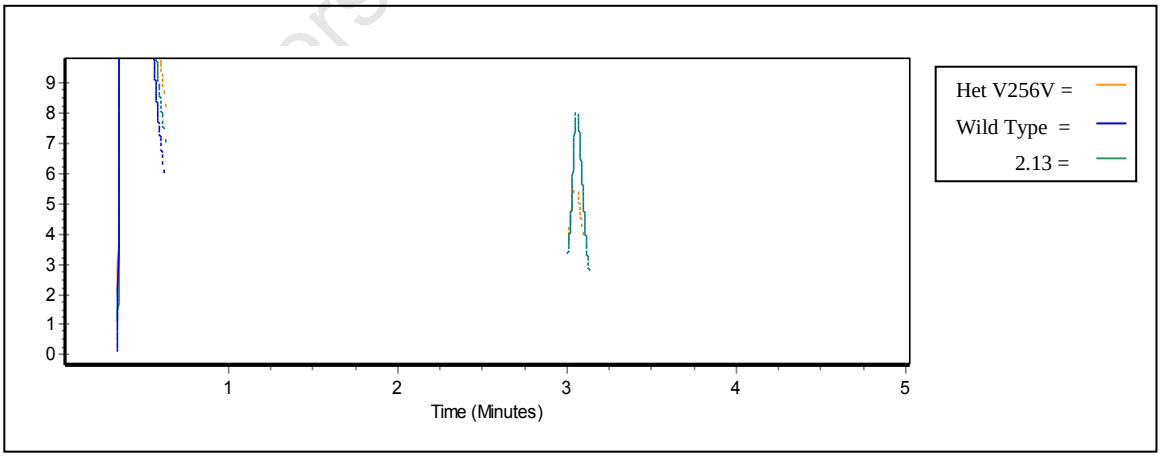


Fig. 3.4 A dHPLC chromatogram showing the heterozygous profile (Het V256V), wild type profile (Wild Type), and a unique profile (2.13) with an altered pattern compared to the profiles of both the heterozygous and wild type control samples.

The 15 samples shown to have altered chromatogram profiles in comparison to the wild type control when exon 6 was screened by dHPLC, were sequenced in order to confirm the presence of the V256V mutation (Fig. 3.5). Cycle sequencing results for the 12 patient samples confirmed the presence of 11 heterozygotes and one homozygote in the patient cohort. Cycle sequencing results furthermore confirmed the presence of a heterozygous V256V mutation in two control samples; however no V256V mutation (or any other sequence alteration) was present in exon 6 of sample 2.13.

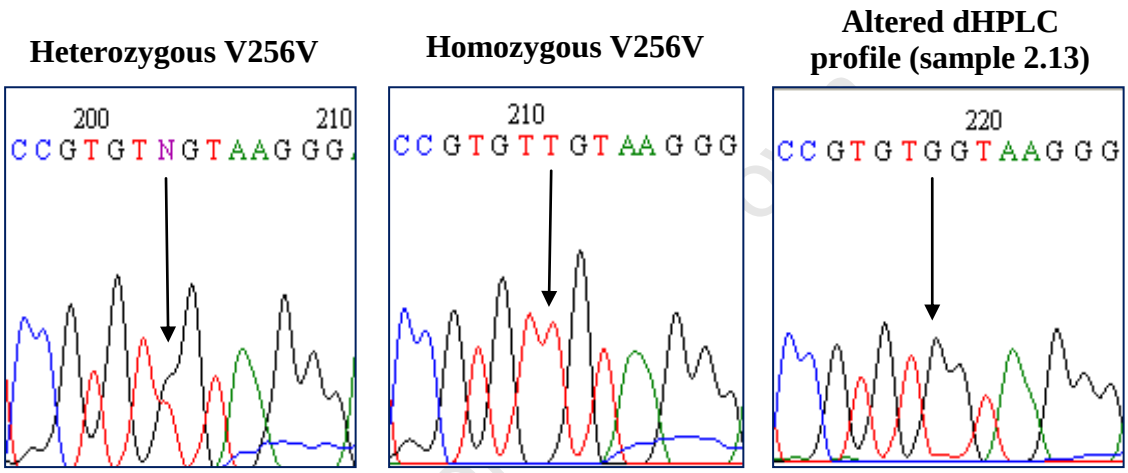


Fig. 3.5 Sequencing electropherograms showing the heterozygous c.768G>T (V256V) mutation (detected in 13 individuals), homozygous c.768G>T (V256V) mutation (detected in one individual) and the wild type sequence of sample 2.13, that had produced an aberrant dHPLC profile.

In summary, a heterozygous V256V mutation was detected in 11/72 (15.28%) patients, i.e. 11/144 alleles (7.63%). Two V256V mutations (homozygotes) were also detected in 1/72 patient samples (1.39%), i.e. 2/144 alleles (1.39%).

A V256V mutation was detected in 2/269 (0.74%) samples in the control cohort, i.e. 2/538 alleles (0.37%). One heterozygote was detected in each of the sub-cohorts of Caucasian (1%) and Afrikaner (0.59%) descent.

3.1.6 G863A mutation detection

A sample was designated as heterozygous G863A by the presence of both a C and a G allele, producing a chromatogram profile containing 3 peaks (Fig. 3.6). A sample was designated as wild type, and therefore did not carry a G863A mutation, by the presence of a C allele, producing a chromatogram profile containing a single peak (Fig. 3.6).

A heterozygous G863A mutation was detected in 7/72 patient samples (9.72%), i.e. 7/144 alleles (4.86%). No patients were found to carry two G863A alleles in the homozygous state.

A heterozygous G863A mutation was detected in 3/269 of the individuals (1.12%) in the control cohort, i.e. 3/538 alleles (0.56%). Two heterozygotes were detected in the Caucasian sub-cohort (2%) and one in the Afrikaner sub-cohort (0.59%).

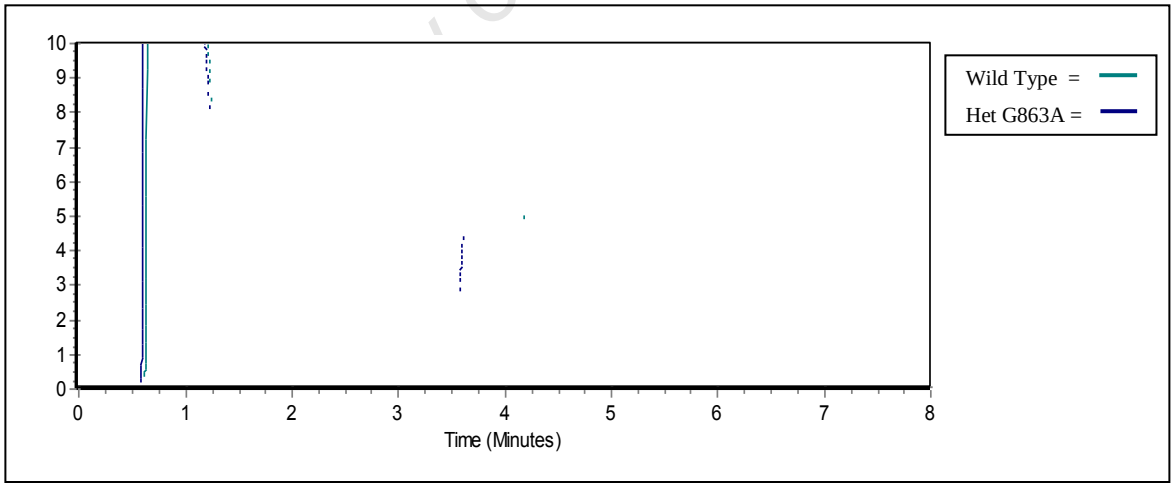


Fig. 3.6 A dHPLC chromatogram showing the wild type profile and the profile obtained when a heterozygous G863A mutation is present.

3.1.7 R152X mutation detection

The R152X mutation in exon 5 could not be detected using an AS-PCR, even when a single LNA nucleotide was incorporated at the 3' end of both the mutation-specific and wild type-specific primers. Screening for the R152X mutation in exon 5 was performed by partially-denaturing conditions using dHPLC analysis.

After the "WAVE mixing experiment" was performed, mutation screening by dHPLC showed eight patient samples to have identical chromatogram profiles to that of the positive control. Three control samples were found to have an identical chromatogram profile to that of the positive control. All three control samples were of Afrikaner descent (Fig. 3.7).

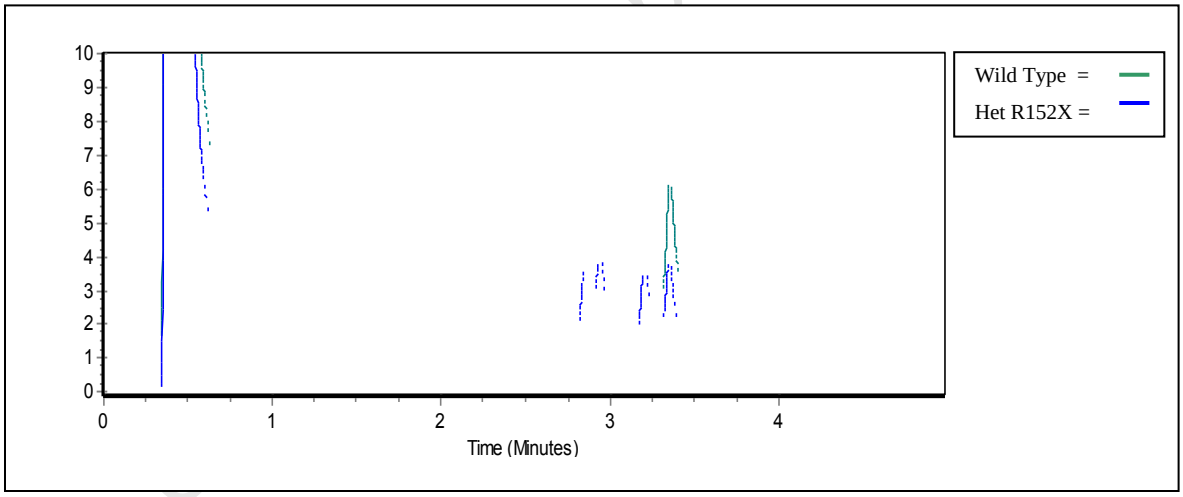


Fig. 3.7 A dHPLC chromatogram showing the wild type profile, and the profile obtained when a heterozygous R152X mutation is present.

The 11 samples shown to have altered chromatogram profiles in comparison to the wild type control when exon 5 was screened by dHPLC, were sequenced in order to confirm the presence of the R152X mutation (Fig. 3.8). Cycle sequencing using the antisense primer revealed that the R152X mutation was not present in one of the patient samples and in two of the Afrikaner control samples. Instead, these samples were shown to carry

a single c.455G>A mutation (R152Q) which was found directly 3' of the R152X mutation site (c.454C>T). Due to the fact that the R152Q mutation was situated in such close proximity to the R152X mutation site, it had resulted in the chromatogram profile of the three samples being identical to the dHPLC chromatogram profile of the positive control. Cycle sequencing results confirmed the presence of a heterozygous R152X mutation in seven patient samples and one control sample in the Afrikaner sub-cohort.

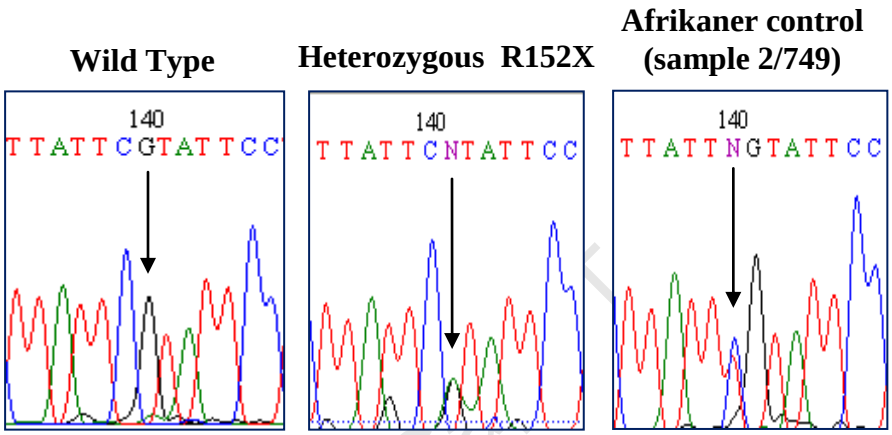


Fig. 3.8 Sequencing electropherograms showing the wild type sequence, heterozygous c.454C>T (R152X) mutation, and the Afrikaner control sample (2/749) that was shown to carry a c.455G>A (R152Q) directly flanking the R152X mutation site.

In summary, a R152X mutation was detected in 7/72 (9.72%) patient samples, i.e. 7/144 alleles (4.86%). No patients were found to carry two R152X mutations in the homozygous state.

A R152X mutation was detected in 1/269 (0.37%) individuals in the control cohort, i.e. 1/538 alleles (0.19%). This heterozygote was detected in the Afrikaner sub-cohort (0.59%).

Table 3.1 The results from screening both the patient and control cohort for seven *ABCA4* mutations, showing the number and frequency of mutant alleles detected per mutation.

The frequencies of the mutant alleles are shown in brackets.

Cohort	Mutant alleles						
	C1490Y	**R602W**	**L2027F**	**IVS38-10T>C**	**V256V**	**G863A**	**R152X**
Patient (n = 72) (144 alleles)	16 (11.11%)	10 (6.94%)	12 (8.33%)	13 (9.03%)	13 (9.03%)	7 (4.86%)	7 (4.86%)
Control (total) (n = 269) (538 alleles)	2 (0.37%)	2 (0.37%)	2 (0.37%)	0	2 (0.37%)	3 (0.56%)	1 (0.19%)
a) Caucasian (n = 100) (200 alelles)	0	0	1 (0.5%)	0	1 (0.5%)	2 (1%)	0
b) Afrikaner (n = 169) (338 alleles)	2 (0.59%)	2 (0.59%)	1 (0.296%)	0	1 (0.296%)	1 (0.296%)	1 (0.296%)

In summary, 78/144 (54.17%) pathogenic alleles were detected in the 72 patients screened. Twenty-two (30.55%) patients were found to carry one mutation, 28 (38.89%) were found to carry two mutations, and 22 (30.55%) were shown not to carry any of the seven mutations. The majority of the patients found to carry two mutations were compound heterozygotes (n = 23 individuals) (82.14%), with the remaining five being identified as homozygotes (17.86%). A total of 16 different paired combinations of pathogenic variations in *ABCA4* were detected in the 28 patients with two mutations.

Of the 269 control samples screened, 12 (4.46%) were shown to carry a heterozygous mutation. Four heterozygotes (4%) were detected in the Caucasian sub-cohort, with eight heterozygotes (4.73%) detected in the Afrikaner sub-cohort.

3.2 Familial co-segregation analysis

Where possible, family studies were conducted in order to determine whether the two mutations identified in 28 patient samples (23 compound heterozygotes and five homozygotes) were co-segregating with the disease within the family, and whether these mutations existed "in *trans*" (i.e. were bi-allelic). Extended pedigrees and biological material from family members were available for 10 of the 28 individuals, and co-segregation analysis could thus be conducted (Fig. 3.9 a to j). The results for each of the 10 families are shown below:

Family 107 (Fig. 3.9a)**:** The affected individual, RPS 107.2TAB, was found to carry two C1490Y mutations (i.e. homozygous C1490Y). Familial co-segregation analysis showed that each parents carried a heterozygous C1490Y mutation.

Family 125 (Fig. 3.9b)**:** The affected individual, RPS 125.1TRB, was found to be a compound heterozygote, carrying both a C1490Y and R152X mutation. Familial co-segregation analysis showed that the father (RPS 125.3SOR) carried a heterozygous R152X mutation, while the mother (RPS 125.2ELI) carried a heterozygous C1490Y mutation.

Family 133 (Fig. 3.9c)**:** The affected individual, RPS 133.4CHA, was found to carry two V256V mutations (i.e. homozygous V256V). Familial co-segregation analysis showed that each parent carried a heterozygous V256V mutation. The unaffected sister (RPS 133.3YOL) was shown not to carry any V256V mutation.

Family 151 (Fig. 3.9d)**:** The affected individual, RPS 151.3DAN, was found to be a compound heterozygote, carrying both a C1490Y and R602W mutation. Familial co-segregation analysis showed that the father (RPS 151.1DJF) carried a heterozygous R602W mutation, while the mother (RPS 151.2AMA) carried a heterozygous C1490Y

mutation. The unaffected brother (RPS 151.5DEW) was shown to carry a heterozygous C1490Y mutation (the maternal allele). The mutations carried by the affected younger brother (RPS 151.4FRE) could not be determined as no biological material was available for this individual.

Family 114 (Fig. 3.9e)**:** The affected individual, RPS 114.3BER, was found to be a compound heterozygote, carrying both a C1490Y and IVS38-10T>C mutation. Familial co-segregation analysis showed that the father (RPS 144.7STE) carried a heterozygous IVS38-10T>C mutation, while the mother (RPS 114.5THE) carried a heterozygous C1490Y mutation. The unaffected paternal uncle (RPS 114.1EUG) was also shown to carry a heterozygous IVS38-10T>C mutation. The unaffected maternal aunt (RPS 114.2WAY) did not carry the C1490Y mutation. The unaffected sister (RPS 114.6NIC) was shown to carry a heterozygous C1490Y mutation (maternally inherited). The younger affected brother (RPS 114.4BAR) was also shown to be a compound heterozygote, carrying both a C1490Y and IVS38-10T>C mutation.

Family 232 (Fig. 3.9f)**:** The affected individual, RPS 232.1SUR, was found to be a compound heterozygote, carrying both a R152X and V256V mutation. Familial co-segregation analysis showed that the father (RPS 232.2JOH) carried a heterozygous V256V mutation. No biological material was available for the mother of the affected individual, who, according to the pedigree had a retinal disease, but was deceased. However, since the father (RPS 232.2JOH) was shown to carry one V256V mutation, it could be inferred, that the R152X mutation was transmitted from the mother.

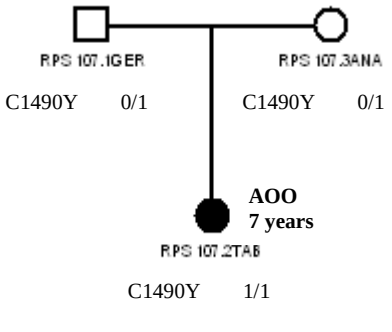
Family 226 (Fig. 3.9g)**:** The affected individual, RPS 226.1KOB, was found to be a compound heterozygote, carrying both a C1490Y and R152X mutation. Familial co-segregation analysis showed that the father (RPS 226.4ALB) carried a heterozygous C1490Y mutation, while the mother (RPS 226.2PET) carried a heterozygous R152X mutation. The unaffected sister (RPS 226.3JAC) was shown not to carry either mutation.

Family 549 (Fig. 3.9h)**:** The affected individual, RPS 549.1ZAN, was found to be a compound heterozygote, carrying both a V256V and L2027F mutation. Familial co-segregation analysis showed that the father (RPS 549.2PHI) carried a heterozygous L2027F mutation. No biological material was available for the mother. However, since the father (RPS 549.2PHI) was shown to carry one L2027F mutation, it is most likely that the V256V mutation was inherited from the mother.

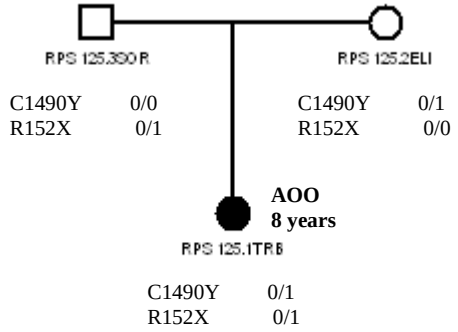
Family 464 (Fig. 3.9i)**:** The affected individual, RPS 464.1SAN, was found to be compound heterozygous, carrying both a R602W and G863A mutation. Familial co-segregation analysis showed that her unaffected offspring (RPS 464.2STE and RPS 464.3NEI) each carried a heterozygous R602W mutation. The offspring were not shown to carry the heterozygous G863A mutation. The disease status of the affected individual's husband was unknown, and no biological material was available for this individual.

Family 352 (Fig. 3.9j)**:** The affected individual, RPS 352.1DEO, was found to be a compound heterozygote, carrying both a R152X and L2027F mutation. This patient was married to RPS 352.2ARL who was also shown to be affected with a retinal disease. Although the clinical details for the proband indicated that he was affected with STGD disease, the exact retinal disease of his spouse was not known. The two disease-causing mutations detected in the proband were not detected in his spouse. Interpretation for this result could be that the spouse carries two other *ABCA4* mutations (if she is affected with an AAR) or that she is affected with a different retinopathy.

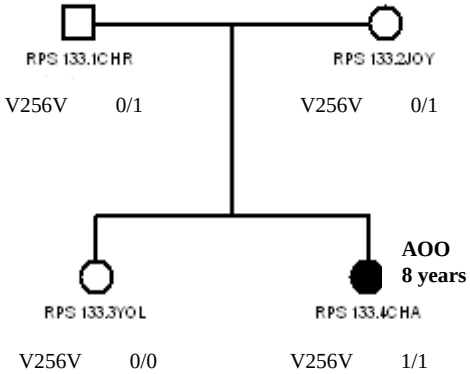
a) Family 107



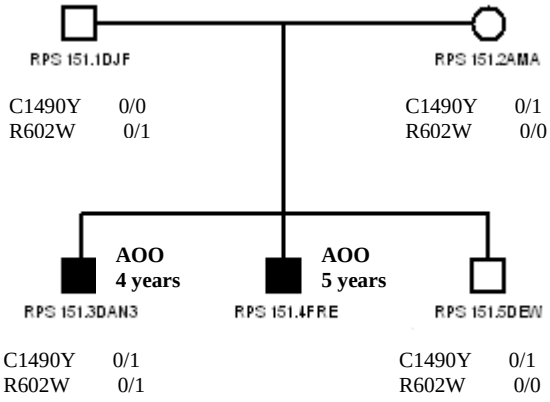
b) Family 125



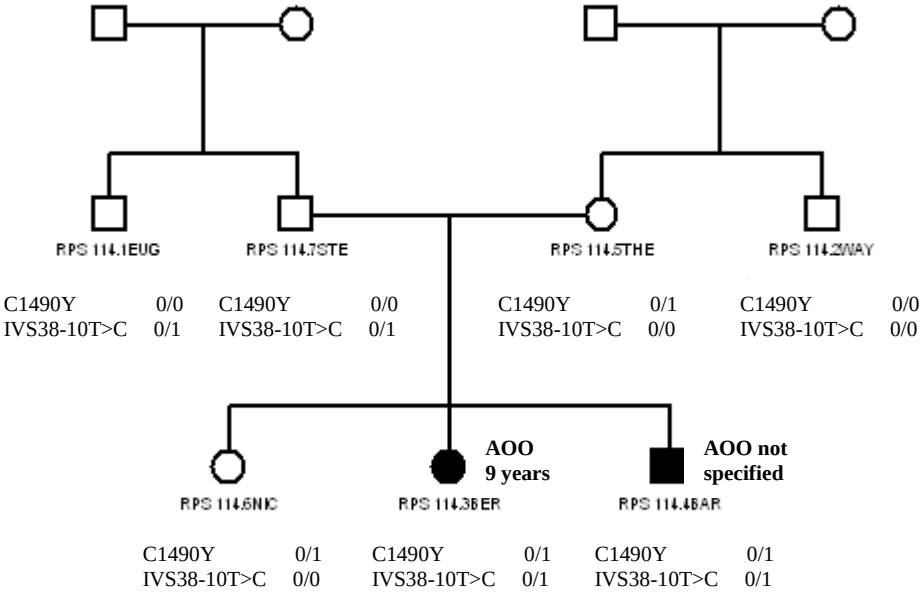
c) Family 133



d) Family 151



e) Family 114



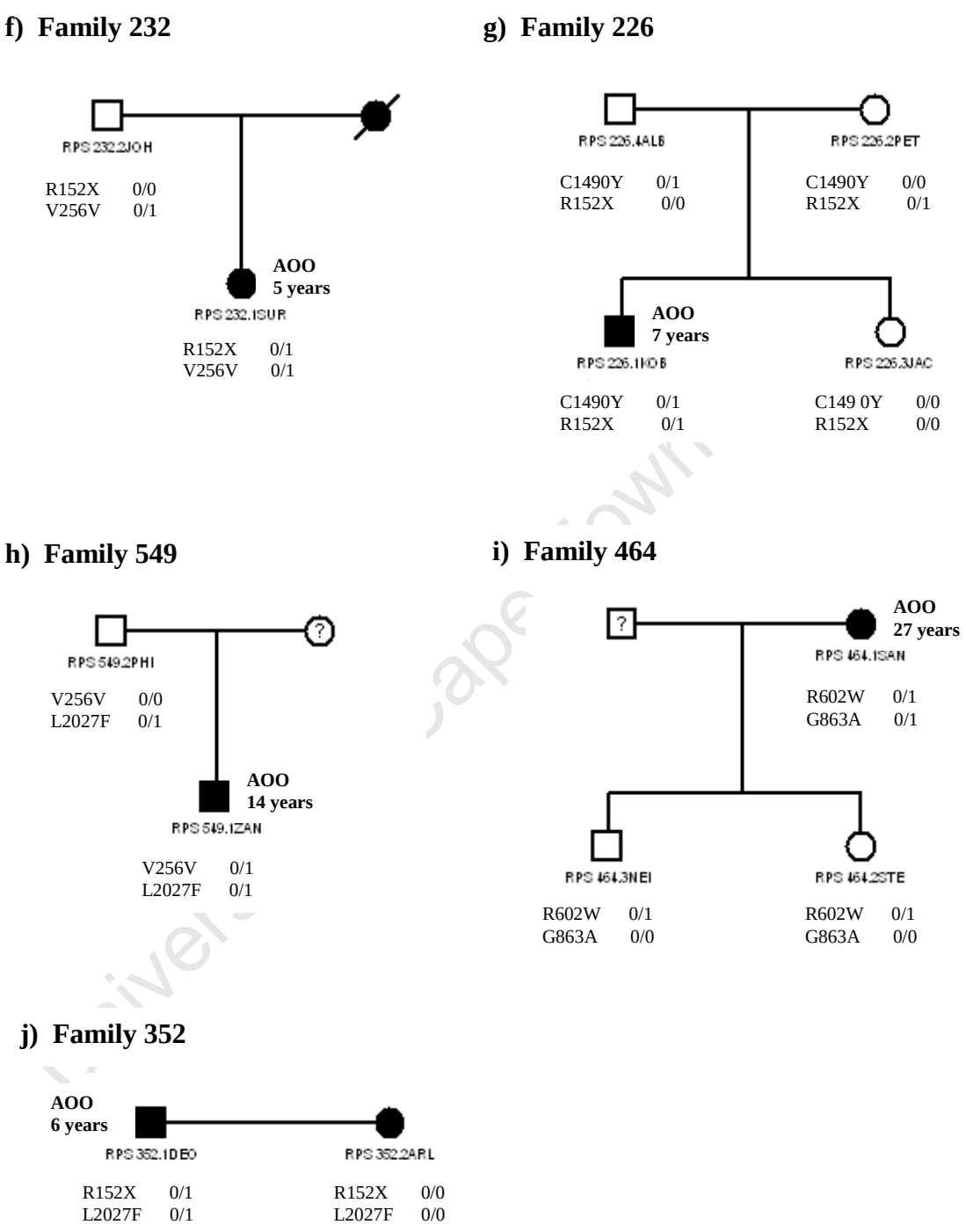


Fig. 3.9 The 10 family pedigrees available for co-segregation analysis.

The square represents a male and a circle represents a female. Shaded symbols indicate an affected individual and a symbol that has a line through it indicates that individual is deceased. AOO indicates the age of onset of the disease. An individual given the number 0/0 next to the appropriate mutation name indicates the individual is wild type and therefore does not carry that mutation. An individual given the number 0/1 next to the appropriate mutation name indicates the individual is heterozygous for the mutation, carrying only one allele with that mutation. An individual given the number 1/1 next to the appropriate mutation name indicates the individual is homozygous and therefore carries two alleles with that mutation.

3.3 Bioinformatics

3.3.1 Missense Mutations

In order to assess the pathogenicity of each of the four missense *ABCA4* mutations (C1490Y, R602W, G863A, and L2027F) investigated in this study, several bioinformatics programmes were utilised, namely PolyPhen; SIFT; PANTHER PSEC; and PMUT (Table 3.2).

Of the four missense mutations, the C1490Y mutation was predicted to have the highest pathogenic score in three out of the four software packages used. Although SIFT revealed R602W to have a slightly higher median conservation score in comparison to the G863A mutation (i.e. nucleotide change is less diverse and therefore has a higher false positive rate), this mutation was predicted as the second most pathogenic. The G863A mutation was ranked as the third most pathogenic overall. The L2027F mutation was assigned a lower subPSEC score in comparison to the G863A mutation using PANTHER PSEC, suggesting it to be more deleterious. However, when L2027F was investigated using the other three bioinformatics programmes (PolyPhen, SIFT, and PMUT) it was predicted to be the least pathogenic of the four missense mutations.

Table 3.2 The bioinformatic results obtained using PolyPhen, PMut, Panther PSEC, and SIFT for the four *ABCA4* missense mutations C1490Y, R602W, L2027F, and G863A.

Rating	Mutation	PolyPhen	PMut	Panther PSEC	SIFT
1	***C1490Y***	**Prediction:** Probably damaging ***PSIC score difference:*** 3.326 (0 = neutral, high positive number = damaging) ***Basis:*** Multiple alignment	**Prediction:** Pathological ***Reliabiliy:*** 9 (0 = low, 9 = high) ***NN output:*** 0.9853 (>0.5 = pathological)	Does not align to the HMM	**Amino acid substitution score:** 0.00 (<0.5 = deleterious) **Median sequence conservation:** 2.46
2	***R602W***	**Prediction:** Probably damaging ***PSIC score difference:*** 2.817 (0 = neutral, high positive number = damaging) ***Basis:*** Multiple alignment	**Prediction:** Pathological ***Reliabiliy:*** 5 (0 = low, 9 = high) ***NN output:*** 0.7931 (>0.5 = pathological)	***subPSEC:*** -6.41899 (0 = neutral, -10 = most likely to be deleterious) $P_{Deleterious}$**:** 0.96829	**Amino acid substitution score:** 0.01 (<0.5 = deleterious) **Median sequence conservation:** 1.96
3	***G863A***	**Prediction:** Probably damaging ***PSIC score difference:*** 2.042 (0 = neutral, high positive number = damaging) ***Basis:*** Multiple alignment	**Prediction:** Pathological ***Reliabiliy:*** 1 (0 = low, 9 = high) ***NN output:*** 0.5785 (>0.5 = pathological)	***subPSEC:*** -4.47326 (0 = neutral, -10 = most likely to be deleterious) $P_{Deleterious}$**:** 0.81355	**Amino acid substitution score:** 0.01 (<0.5 = deleterious) **Median sequence conservation:** 1.94
4	***L2027F***	**Prediction:** Possibly damaging ***PSIC score difference:*** 1.842 (0 = neutral, high positive number = damaging) ***Basis:*** Multiple alignment	**Prediction:** Neutral ***Reliabiliy:*** 1 (0 = low, 9 = high) ***NN output:*** 0.4356 (>0.5 = pathological)	***subPSEC:*** -5.23022 (0 = neutral, -10 = most likely to be deleterious) $P_{Deleterious}$**:** 0.90293	**Amino acid substitution score:** 0.04 (<0.5 = deleterious) **Median sequence conservation:** 1.90

3.3.2 Splicing mutations occurring in exonic regions

The web-based programme, ESEfinder, predicted a change in the splicing profile of the wild type sequence in comparison to the sequence containing the V256V mutation. Analysis of both sequences revealed the binding of an additional SR protein, SRp40, in the mutant V256V sequence (Fig. 3.10). The binding of SRp40 to this sequence, is the result of the generation of another ESE motif in the sequence, created by the V256V mutation. As mentioned previously, SR proteins play various roles in the splicing pathway of a nucleotide sequence. By creating a binding site for an additional SR protein, the V256V has been predicted to alter the normal splicing mechanism of the *ABCA4* gene.

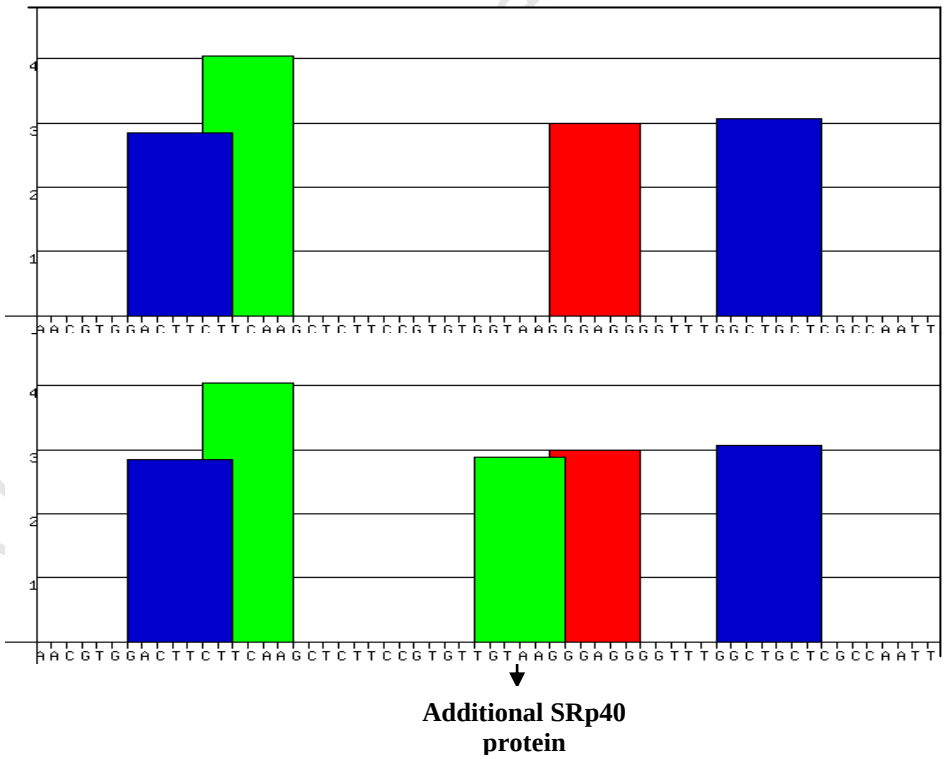


Fig. 3.10 Graph showing the results from ESEfinder analysis for the V256V mutation.

The top graph shows the results obtained for the wild type sequence, with the bottom graph showing the results from the sequence containing the V256V mutation. An additional SRp40 protein was found to bind to the sequence when the V256V mutation was present. This is shown by the additional green motif that is present on the left hand side of the red motif in the V256V mutant sequence.

3.3.3 Splicing mutations occurring in intronic regions

The BDGP Splice Site Prediction programme did not predict any significant alteration to the donor or acceptor splice site when the wild type sequence of exon 39 (including part of intron 38) was compared to the sequence containing the IVS38-10T>C mutant allele. This result was not expected since the IVS38-10T>C mutation is situated in the 5' splice acceptor site, 10bp away from the intron-exon boundary of exon 39. Investigation into the conservation of the wild type nucleotide amongst species, using the UCSC Genome browser, showed that the nucleotide was highly conserved (Fig. 3.11). Conservation of the wild type nucleotide amongst species could imply its importance in the correct functioning of the *ABCA4* gene.

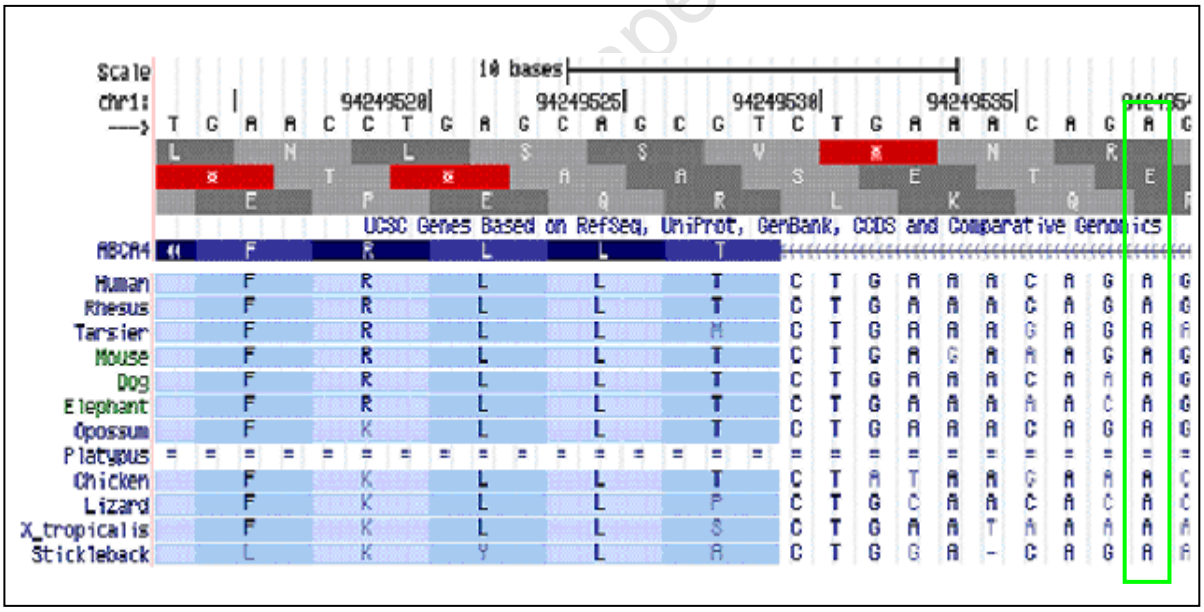


Fig. 3.11 The results obtained from the UCSC Genome browser showing the conservation of the wild type nucleotide at position 10bp 5' of exon 39.

The green rectangle highlights the conservation of the nucleotide amongst the different species.

IVS38-10T>C is listed as a mutation on the ABCR400 microarray chip (M Mällo, personal communication). In addition, an article by Klevering et al. (2005) lists IVS38-

10T>C as a pathogenic mutation. For the purpose of this study, IVS38-10T>C was thus regarded as a disease-causing variant.

3.3.4 Rating the pathogenicity of the seven common *ABCA4* mutations

Both bioinformatic tools as well as extensive literature research were utilised in order to assess the pathogenicity of all seven *ABCA4* mutations. Of the seven mutations four were missense mutations, one resulted in a stop codon, whilst the remaining two were expected to affect splicing.

Previous investigations assessing the functions of the missense mutations C1490Y and R602W, revealed a decrease in the amount of hydrolytic activity in both the mutant alleles compared to that of the wild type allele (Wisniewsk et al. 2005). It was also noted that the presence of either of these two missense mutations resulted in the expression of the ABCA4 protein in the inner segment of rod cells and not in the original site of the outer segment. These variants have therefore been designated as mislocalisation alleles. According to Wisniewsk et al. (2005), these two mutations appear to have a dual effect on the ABCA4 protein, affecting both the function and the end localisation of the protein. The results obtained from the bioinformatic analysis predicted the C1490Y and R602W mutations to be the most pathogenic. Therefore, C1490Y and R602W were ranked as the first and second most pathogenic of the seven mutations, respectively.

The V256V mutation does not result in an altered amino acid residue. The nucleotide substitution (c.768G>T) has, however, been shown to create an alternative splice site, which leads to an abnormal ABCA4 transcript (Valverde et al. 2006). The result obtained using the bioinformatic tool, ESEfinder, also indicated a possible role for V256V in alternative splicing, with the prediction of an additional ESE binding site for an SR protein, SRp40, when the mutation is present. Investigations into the functional consequence of this mutation have lead to its classification as a severe mutation

(Klevering et al. 2004). This mutation was thus rated as the third most pathogenic mutation.

The R152X mutation occurs as a result of a nucleotide substitution (c.454C>T), and leads to a premature stop codon in exon 5, thus resulting in a truncated protein. When the R152X mutation is present in the heterozygous state, it causes the inactivation of the *ABCA4* allele (Zhang et al. 1999). Unlike the C1490Y and R602W mutations, which result in mislocalisation of the ABCA4 protein, and the V256V mutation, which appears to be involved in alternative splicing, and could be more hazardous to the retina; R152X results in a non-functional allele, leading to a large decrease in total ABCA4 protein activity. This mutation has thus been designated as the fourth most pathogenic mutation of those screened for in the present study.

Previous investigations into the function of the missense mutation (c.2588G>C) resulting in the G863A amino acid change, have shown that this mutation has a mild effect on ABCA4 protein activity in individuals affected with STGD disease and is associated with a later AOO (Maugeri et al. 1999; Shroyer et al. 2001). Bioinformatic analysis of the mutation predicted it to be less pathogenic in comparison to the C1490Y and R60W mutations. The G863A mutation was designated as the fifth most pathogenic *ABCA4* mutation.

The missense mutation, c.6079C>T, results in a L2027F amino acid change. This mutation is located in the NBD2 and research has shown it has an effect on ABCA4 protein expression and ATP binding (Sun et al. 2000). In most cases bioinformatic analysis predicted L2027F to be less severe in comparison to the other three missense mutations, namely C1490Y, R602W, and G863A (listed in decreasing predicted pathogenicity according to the bioinformatic results obtained) and was therefore predicted as the least pathogenic of the four mutations. In the group of mutations as a whole, L2027F was designated as the sixth most pathogenic *ABCA4* mutation of the seven investigated in the current study.

Several researchers have investigated the function of the IVS38-10T>C mutation on ABCA4 protein function (Maugeri et al. 1999; Rivera et al. 2000; Klevering et al. 2004). Although situated in a splice acceptor site, this mutation is not predicted to alter splicing, however, other researchers still consider it pathogenic (M Mällo, personal communication; Klevering et al. 2005). This mutation was thus designated the least pathogenic of the seven *ABCA4* mutations investigated in this study.

3.4 Statistical analysis

Non-parametric analysis was conducted in order to investigate the influence of the seven common *ABCA4* mutations on the clinical phenotype of AAR-affected individuals. The analysis was performed using the results obtained from the mutation screening of the seven *ABCA4* mutations in this project as well as the results obtained from the ABCR400 microarray chip screening. A total of 28 samples in the patient cohort of this project were found to carry two of the seven mutations. Results from previous microarray screening and family studies in our laboratory, identified a total of 86 individuals from 66 families that were found to carry two of the seven mutations. This resulted in a total of 114 individuals who were found to carry two of the seven *ABCA4* mutations and a total of 23 different paired mutation combinations. Of the 114 individuals, only patients for whom a self-reported AOO was available were used in this analysis. Data was available from a total of 106 patients from 94 families (26 of the 28 patients in the current study, and 80 of the 86 patients screened by the ABCR400 microarray chip) (Table 3.3).

The AOO varied depending on the mutation combination. The four mutations predicted to be most highly pathogenic (C1490Y, R602W, V256V, and R152X) appeared to have a similar average AOO range of between six to 12 years. The C1490Y mutation in combination with either itself or another common mutation was detected in 56 of the 106 patient samples. The IVS38-10T>C mutation was identified in a total of 23 patients (including two homozygotes), most of whom had an AOO of between six to 11 years,

with an exception being seen when it was found to occur in combination with G863A, resulting in a later AOO. In fact, when the G863A mutation was present in combination with one of the other common mutations, the AOO was later. No individuals were found to carry two G863A mutations.

Table 3. 3 A summary of the 23 mutation combinations observed in 106 patients with AARs, showing the number of patients per combination and the range and average of the AOO (in years) per combination.

Mutation combination	ABCR400 Microarray	Seven Common *ABCA4* Mutation Screening	Total Patient Cohort	AOO (in years): Range; Average
C1490Y + IVS38-10T>C	9	3	12	5-12; 8.9
C1490Y + R602W	9	2	11	4-13; 8.1
C1490Y + L2027F	10	-	10	7-18; 10.5
C1490Y + V256V	9	1	10	7-12; 10.3
C1490Y (HOMO)	6	2	8	7-9; 8.1
R602W + L2027F	6	2	8	8-22; 12
L2027F + R152X	5	1	6	6-11; 7.6
C1490Y + R152X	3	2	5	7-9; 8
L2027F + V256V	2	3	5	10-14; 12
V256V + R152X	3	1	4	5-10; 8
C1490Y + G863A	3	-	3	19-20; 19.3
R602W + V256V	2	1	3	4-10; 6.7
L2027F + IVS38-10T>C	2	1	3	7-16; 11
IVS38-10T>C + R152X	3	-	3	5-8; 7
G863A + R152X	3	-	3	30-36; 32
V256V + G863A	1	2	3	12-20, 16.7
R602W + G863A	1	1	2	8-27; 17.5
IVS38-10T>C (HOMO)	-	2	2	6-7; 6.5
R602W + IVS38-10T>C	1	-	1	11;11
IVS38-10T>C + G863A	1	-	1	20;20
L2027F (HOMO)	1	-	1	28; 28
V256V (HOMO)	-	1	1	8; 8
IVS38-10T>C + V256V	-	1	1	6; 6

The Kruskal-Wallis test was performed where AOO was used as the dependent variable (plotted on the y-axis) and the various paired combinations of mutations (genotypes)

were used as the independent variable (plotted on the x-axis). Patients with the same mutation combination were grouped together and sample sizes less than five were excluded (Fig. 3.12). This resulted in a total of nine different mutation combinations being included in the analysis (n = 76 patients).

The Kruskal-Wallis analysis revealed a *p*-value of 0.003. Since the *p*-value was less than 0.05, the proposed null hypothesis (H_O), stating that in a South African cohort with a recessive AAR there is no significant difference between the median AOO for all paired combinations of disease-causing variations in *ABCA4*, was rejected. The proposed alternative hypothesis (H_A) stating that in a South African cohort with a recessive AAR a significant difference between the median AOO exists between at least one pair of disease-causing variations in *ABCA4*, was then assumed.

Having established that there were differences in AOO between the different genotypes, the Wilkoxon Rank Sum test was used to determine exactly which genotype groups differ with respect to AOO. For one particular paired combination (G863A+R152X, n = 3 individuals), a much later AOO (i.e. 30 years and older) was observed in comparison to the other mutation combinations. In order to test whether this observation held any significance, paired mutation combinations carried by three or more individuals were included for this analysis. This resulted in a total of 16 different mutation combinations being included in the analysis (n = 97 individuals). A total of 120 different pairwise comparisons (mutation combination **A** versus mutation combination **B**) were conducted and significant *p*-values were obtained for 37 different paired mutation combinations. Due to the large number of statistical comparisons being performed on the data, adjustments for the multiple pairwise comparisons was conducted. This was performed using the Bonferroni correction, where the *p*-value of each pairwise comparison was adjusted for the 120 tests. After Bonferroni corrections were made for the multiple comparisons, by multiplying each observed *p*-value obtained from the analysis by 120 (the number of tests conducted), none of the 37 previously significant *p*-values remained significant (Table 3.4).

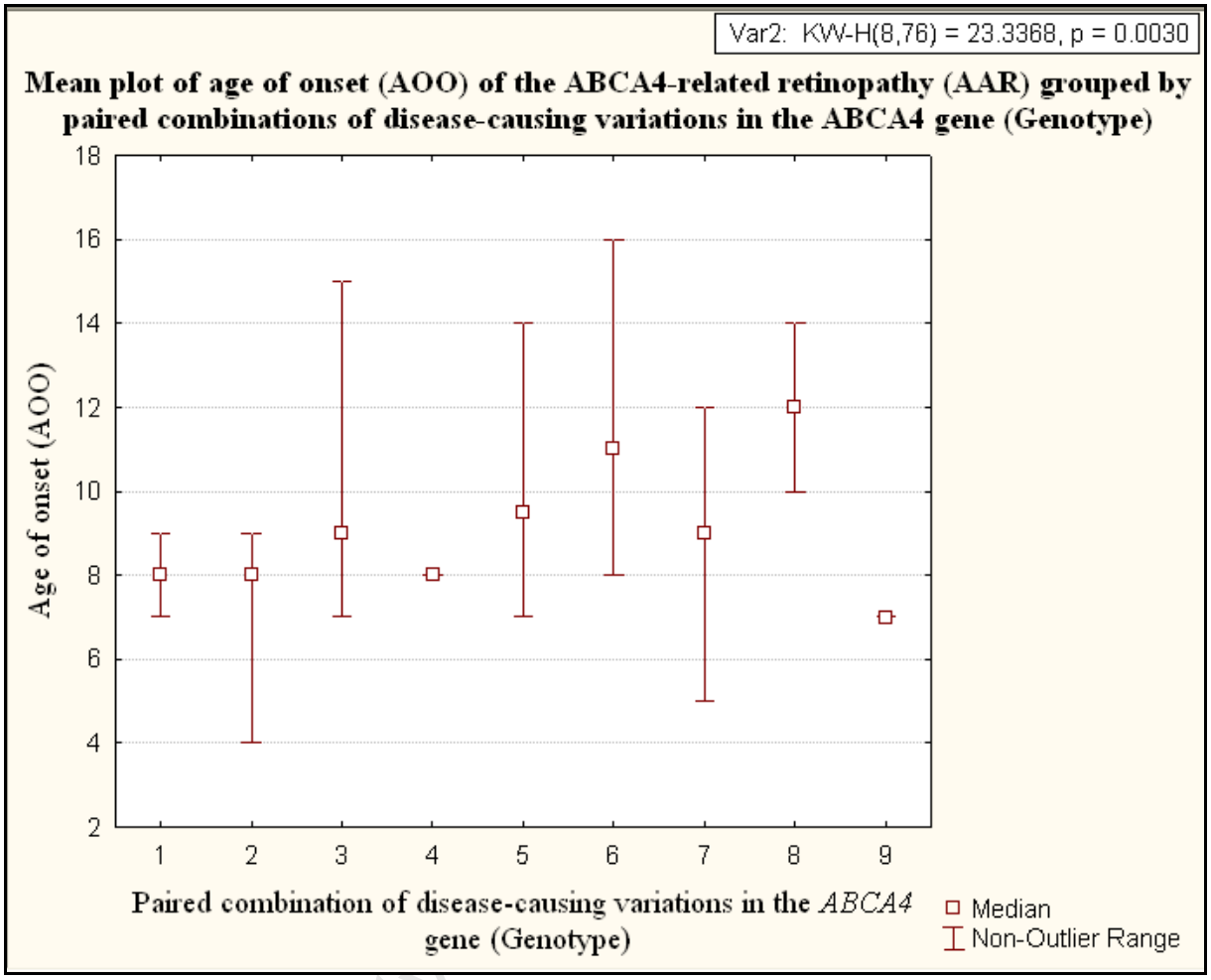


Fig. 3.12 A graph showing the results from the non-parametric Kruskal-Wallis test comparing AOO and nine genotype groups.

AOO is represented on the y-axis. The various genotypes are shown on the x-axis: 1 - homozygous C1490Y genotype, 2 - Heterozygous C1490Y and R602W genotype, 3 - heterozygous C1490Y and V256V genotype, 4 - heterozygous C1490Y and R152X genotype, 5 - C1490Y and L2027F genotype, 6 - R602W and L2027F genotype, 7 - C1490Y and IVS38-10T>C genotype, 8 - L2027F and V256V genotype, and 9 - L2027F and R152X genotype.

Table 3.4 The 37 different pairwise comparisons (mutation combination A versus mutation combination B) shown to be significantly different with respect to AOO prior to Bonferroni corrections.

Paired Combination (A)	Paired Combination (B)	Sample size (A)	Sample size (B)	Median AOO (A) (years)	Median AOO (B) (years)	*p*-value before Bonferroni correction	p-value after Bonferroni correction
C1490Y + C1490Y	C1490Y + G863A	8	3	8	19	0.006061	0.727273
C1490Y + C1490Y	G863A + R152X	8	3	8	30	0.006061	0.727273
C1490Y + C1490Y	L2027F + V256V	8	5	8	12	0.00554	0.186480
C1490Y + C1490Y	V256V + G863A	8	3	8	18	0.006061	0.727273
C1490Y + G863A	C1490Y + IVS38-10T>C	3	12	19	9	0.004396	0.527473
C1490Y + G863A	C1490Y + L2027F	3	10	19	9.5	0.006993	0.839161
C1490Y + G863A	C1490Y + R602W	3	11	19	8	0.002747	0.329670
C1490Y + G863A	C1490Y + V256V	3	11	19	9	0.005495	0.659341
C1490Y + IVS38-10T>C	G863A + R152X	12	3	9	30	0.004396	0.527473
C1490Y + IVS38-10T>C	V256V + G863A	12	3	9	18	0.006593	0.791209
C1490Y + L2027F	G863A + R152X	10	3	9.5	30	0.006993	0.839161
C1490Y + R602W	G863A + R152X	11	3	8	30	0.002747	0.329670
C1490Y + R602W	L2027F + V256V	11	5	8	12	0.002289	0.274725
C1490Y + V256V	G863A + R152X	11	3	9	30	0.005495	0.659341
C1490Y + R152X	L2027F + V256V	5	5	8	12	0.007937	0.952381
G863A + R152X	L2027F + R602W	3	9	30	11	0.009091	1.000000
C1490Y + R602W	L2027F + R602W	11	9	8	11	0.010199	1.000000
C1490Y + IVS38-10T>C	L2027F + V256V	12	5	9	12	0.011797	1.000000
C1490Y + C1490Y	L2027F + R602W	8	9	8	11	0.012505	1.000000
C1490Y + G863A	C1490Y + R152X	3	5	19	8	0.017857	1.000000
C1490Y + G863A	L2027F + R152X	3	5	19	7	0.017857	1.000000
C1490Y + R602W	V256V + G863A	11	3	8	18	0.013736	1.000000
C1490Y + R152X	G863A + R152X	5	3	8	30	0.017857	1.000000
C1490Y + R152X	V256V + G863A	5	3	8	18	0.017857	1.000000

Paired Combination (A)	Paired Combination (B)	Sample size (A)	Sample size (B)	Median AOO (A) (years)	Median AOO (B) (years)	*p*-value before Bonferroni correction	p-value after Bonferroni correction
G863A + R152X	**L2027F + R152X**	3	5	30	7	0.017857	1.000000
L2027F + R602W	**L2027F + R152X**	9	5	11	7	0.015984	1.000000
L2027F + R152X	**V256V + G863A**	5	3	7	18	0.017857	1.000000
C1490Y + R152X	**L2027F + R602W**	5	9	8	11	0.020979	1.000000
L2027F + R152X	**L2027F + V256V**	5	5	7	12	0.02381	1.000000
C1490Y + V256V	**V256V + G863A**	11	3	9	18	0.024725	1.000000
C1490Y + G863A	**L2027F + V256V**	3	5	19	12	0.035714	1.000000
C1490Y + L2027F	**V256V + G863A**	10	3	9.5	18	0.034965	1.000000
G863A + R152X	**L2027F + V256V**	3	5	30	12	0.035714	1.000000
IVS38-10T>C + R152X	**L2027F + V256V**	3	5	8	12	0.035714	1.000000
IVS38-10T>C + R152X	**L2027F + R602W**	3	9	8	11	0.040909	1.000000
C1490Y + L2027F	**L2027F + R152X**	10	5	9.5	7	0.045288	1.000000
C1490Y + V256V	**L2027F + R152X**	11	5	9	7	0.049451	1.000000

Chapter 4: Discussion

Since *ABCA4* was identified as the gene responsible for causing arSTGD, its involvement with this disease and other retinal diseases has been investigated extensively (Allikmets et al. 1997a; Cremers et al. 1998; Briggs et al. 2001; Klevering et al. 2005; Rosenberg et al. 2007). Historically, studies have focused on populations in Europe and research into AARs from emerging countries, including South Africa, has been lacking.

September et al. (2004) investigated the spectrum of *ABCA4* mutations involved with arSTGD and showed six common disease-associated sequence variants underlying arSTGD in that cohort. More than 90% of the cohort was of Afrikaner descent. The study formed part of a broader investigation to determine the genes underlying many different types of RDD in the South African population (September PhD thesis, 2003).

Subsequently, material from a cohort of South Africans affected with an AAR, archived in the Division of Human Genetics at UCT, was screened for mutations in the *ABCA4* gene using the ABCR400 microarray chip, in order to assess the spectrum of mutations in the population. This screening detected a large number of causative mutations, with seven mutations found to be relatively common, namely C1490Y, R602W, IVS38-10T>C, L2027F, V256V, G863A, and R152X.

The aim of the current study was to investigate the founder effect of the seven most common *ABCA4* mutations in South African individuals in order to determine the carrier frequency in the general population and to examine the previously described genotype-phenotype correlation.

4.1 Detection of the seven *ABCA4* mutations

Reproducible, sensitive assays were created in order to detect all seven *ABCA4* mutations.

Detection of the C1490Y, R602W, IVS38-10T>C, and L2027F mutations was performed by designing a multiplex SNaPshot assay. The incorporation of an additional internal SNaPshot primer for the detection of the V256V mutation into the designed multiplex SNaPshot reaction was attempted, however, this proved problematic. Firstly, the internal SNaPshot primer was designed to bind to the antisense strand of the DNA sequence. Oscillating black and green peaks (corresponding to a C and A allele, respectively) were visualised during optimisation when using a heterozygous V256V mutation-positive control. Also, oscillating black peaks (corresponding to a C allele) were seen when a wild type control was tested. Instead of the expected SNaPshot product size of around 46bp, the oscillating peaks were observed across the region between 34bp and 49bp. Secondly, an internal SNaPshot primer was designed to bind to the sense strand of the DNA sequence, however, this produced similar results. Oscillating blue and red peaks (corresponding to a G and T allele, respectively) were visualised during optimisation using a heterozygous V256V mutation-positive control, and oscillating blue peaks (corresponding to a G allele) were seen when a wild type control sample was tested. The V256V mutation was thus detected by partially-denaturing conditions using dHPLC analysis for the purposes of this project.

Detection of the G863A mutation was attempted using an AS-PCR assay. Due to the small difference in size between the amplified wild type and mutant alleles (10bp) the products generated from the AS-PCR were visualised using size-based separation by dHPLC analysis.

Similarly, detection of the R152X mutation was also attempted using an AS-PCR assay. The designed AS-PCR primers resulted in a 24bp product difference between the wild type and mutant alleles. However, non-specific amplification of the mutant allele was

observed in this assay, even after numerous optimisation procedures were attempted, including the use of 3' LNA residues in the primers. The R152X mutation was consequently detected by partially-denaturing conditions using dHPLC analysis. However, the R152Q (c.455G>A) variant next to the R152X (c.454C>T) mutation site, created an aberrant dHPLC chromatogram profile identical to that of the positive control. As a result, all samples exhibiting aberrant dHPLC profiles of exon 5 needed to be sequenced in order to determine the true nature of the sequence variant. Literature reports revealed that previous researchers have classified the R152Q as a rare sequence variant, due to the fact that it was found to occur at similar frequencies in both a control and patient cohort (Rivera et al. 2000).

For future laboratory work involving AS-PCR primer design, the incorporation of a mismatch either one, two or three nucleotides, 5' of the mutation of interest, should be considered, as this has been found to improve allele discrimination (Kwok et al. 1994). Furthermore, in order to aid in the size-based differentiation between wild type and mutation-containing alleles, longer random nucleotide tails should be incorporated on the 5' end of either one of the allele-specific primers. In the case where random nucleotides are incorporated on both allele-specific primers, the length of the tails should vary enough between the two primers in order for adequate allele discrimination.

In designing methods for dHPLC analysis, one should be aware of the fact that alterations in the sequence that are found to be in close proximity to the mutation site of interest could interfere with result analysis. Therefore results from mutation analysis performed under partially-denaturing conditions using dHPLC analysis should be verified by cycle sequencing.

G863A and R152X AS-PCR primers were available lab resources and were thus used for this project. Ideally, the V256V, G863A and R152X mutations should be incorporated into the SNaPshot multiplex assay. Recently, subsequent to the completion of the benchwork of the current project, the detection of the V256V mutation by SNaPshot PCR using a shorter primer (with no random nucleotide tail) has been

successful. In addition, attempts are currently in progress to incorporate the detection of the G863A and R152X mutations, into the multiplex SNaPshot reaction.

4.2 *ABCA4* mutations in South Africa

4.2.1 *ABCA4* mutation screening in a patient cohort

In the present study, DNA samples from 72 unrelated patients affected with AARs, and not previously analysed with the ABCR400 microarray, were screened using the designed assay. Of the 72 individuals in the patient cohort, 70 were clinically diagnosed as having STGD and two individuals were diagnosed as having CRD. At least one of the seven *ABCA4* mutations was detected in 50 (69.44%) individuals in the patient cohort; this included 28 (38.89%) individuals in whom two mutations were identified (i.e. were completely characterised). In agreement with previous studies, the majority of these patients (n = 23) were compound heterozygotes. In order to calculate the predicted proportion of individuals expected to be fully characterised, the results obtained from the ABCR400 microarray chip screening were used as a reference: 100 of 181 (55.25%) patients tested with the microarray were found to be completely characterised, of which 66 of 100 (66%) were found with two of the seven common mutations (unpublished data). One would thus expect 36.3% (0.55 x 0.66) of the cohort screened in this study to carry two of the seven *ABCA4* mutations. The results obtained in this current study are therefore in agreement with what was expected.

C1490Y was found to be the most frequently occurring mutation, detected in 14 (19.44%) patient samples. This mutation was also found to be the most commonly occurring in the September et al. (2004) study, as well as in the ABCR400 microarray analysis (unpublished data), where the majority of the patient cohorts were affected with STGD. This is therefore the most common mutation in the South African STGD cohort.

Of the seven *ABCA4* mutations, the R152X and G863A mutations occurred least frequently; each was seen in seven (9.72%) patient samples. These mutations were also found to be the least commonly detected in the September et al. (2004) study, as well as in the ABCR400 microarray analysis (unpublished data). These two mutations are therefore the least commonly occurring mutations of the seven screened for in the South African STGD cohort.

The frequency of IVS38-10T>C in the current study (15.3%) was in accordance with results from the ABCR400 microarray screening, which detected a frequency of IVS38-10T>C mutations of almost 16%.

As discussed by September et al. (2004), the majority of the six common *ABCA4* mutations detected in their study were found to occur at a higher frequency in the South African population in comparison to populations in Europe, with the exception of the G863A mutation, which appears to have a founder effect in The Netherlands. As mentioned previously, September et al. (2004) classified IVS38-10T>C as non-disease-causing and it was only detected in 1/64 (1.56%) of patients with arSTGD. In the Rivera et al. (2000) study, IVS38-10T>C (identified as a rare sequence variant) was detected at a higher frequency, in 9/144 (6.25%) patients with arSTGD in the German population.

Investigations into the spectrum of *ABCA4* mutations in the Italian population also revealed a very low or non-existent frequency of the seven mutations, with only the V256V mutation detected in 1/71 (1.41%) patient samples (Passerini et al. 2009).

An investigation of *ABCA4* in 144 German patients affected with STGD identified three of the seven common mutations i) G863A, ii) R152X, and iii) V256V, of which G863A was detected most often - occurring in 17/144 (11.81%) patient samples (Rivera et al. 2000). The other two mutations were detected at much lower levels, i.e. 0.69%. The results obtained from the current study are thus in agreement with the conclusions drawn by September et al. (2004) that apart from G863A, these commonly occurring mutations in the South African population occur at a much lower frequency in various European populations.

4.2.2 *ABCA4* mutation screening in a control cohort

Screening of a total of 269 individuals in a control cohort using the designed assay resulted in the detection of a heterozygous mutation in 12 samples (4.46%). Four mutations (4%) were detected in a general Caucasian sub-cohort and eight mutations (4.73%) were detected in the Afrikaner-specific sub-cohort. The results obtained from the control cohort screening indicate that the carrier frequency of the C1490Y, R602W, and G863A mutations is slightly higher in comparison to the other mutations (L2027F, V256V, and R152X), with the IVS38-10T>C mutation not being detected at all. The identification of 12/269 heterozygotes in the control cohort resulted in a background frequency of carriers in this population to be 4.46 per 100 individuals (approximately 5% of South African individuals were found to be carriers of one of the seven *ABCA4* mutations screened for in the present study). This is greater than previous reports, estimating the carrier frequency to be between five and 10% for all mutations tested for on the ABCR400 microarray (Jaakson et al. 2003). In the current study, half the number of heterozygotes was detected in the Caucasian sub-cohort in comparison to the Afrikaner sub-cohort. One may therefore gauge that STGD occurs at a relatively high incidence in the Afrikaner population. However, percentage-wise this is not the case, the Caucasian cohort (n = 100) is almost half the size of the Afrikaner-specific cohort (n = 169). In addition, the Caucasian sub-cohort is likely to contain some individuals of Afrikaner descent. If the Caucasian sub-cohort had consisted only of individuals of British or English heritage, it is possible that an even higher difference in frequency between the two sub-groups could have been observed. Despite this, the results obtained could aid in genetic counselling when risk prediction is discussed with patients and their family members.

4.3 Genotype-phenotype correlation

The large amount of *ABCA4* allelic and phenotypic heterogeneity has led to investigations of a genotype-phenotype model. Numerous investigations have attempted to determine the exact relationship between the different *ABCA4* mutations and the severity of RDD. In 1998, Van Driel et al. stated that most *ABCA4* mutations could be placed into different classes of severity, and that phenotype can vary depending on the remaining total activity of the ABCA4 protein. Alleles which result in a non-functional protein have been associated with a more severe RDD phenotype, namely arRP, whereas alleles which result in partial *ABCA4* activity have been associated with milder RDD phenotypes, namely CRD and STGD. Valverde et al. (2006) hypothesised that both the type of mutation and its position, in conjunction with the paired combination of disease-causing *ABCA4* variations, play a role in the AOO of visual impairment or deterioration. Similarly, Lewis et al. (1999) used the AOO of visual impairment to assess the clinical severity of a paired combination of mutations.

4.3.1 Investigating the pathogenicity of *ABCA4* mutations

In order to assess the effect that each of the seven *ABCA4* mutations has on the ABCA4 protein function, pathogenicity predictions were assigned to each mutation using bioinformatic tools and an extensive literature search. It was thought that determination of the pathogenicity of the mutations could aid in the understanding of the genotype-phenotype correlation. Due to the fact that four (C1490Y, R602W, L2027F, and G863A) of the mutations were missense mutations, one (R152X) a nonsense mutation, and the other two (V256V and IVS38-10T>C) possible splice site mutations, different bioinformatic tools were used to investigate the possible biological consequence of the mutations. Predicted pathogenicity for each mutation was assigned according to the results obtained. Overall predictions revealed the C1490Y mutation to be the most pathogenic, whereas the IVS38-10T>C mutation was predicted to be the least pathogenic. The order of predicted pathogenicity, with decreasing severity, of the seven

ABCA4 mutations was as follows: C1490Y, R602W, V256V, R152X, G863A, L2027F, and IVS38-10T>C.

Previous research revealed that a large proportion of the patients found to carry the C1490Y mutation presented with an earlier AOO (i.e. 11 years) in comparison to patients with no C1490Y mutation (September et al. 2004). Functional analysis has revealed that the presence of a C1490Y allele decreases the amount of ATP hydrolytic activity and causes mislocalisation of the ABCA4 protein to the inner segment (and not the outer segment) of the photoreceptor (Wisznieski et al. 2005). It is therefore plausible that C1490Y could be predicted to be highly pathogenic. The R602W mutation has also been shown to interrupt ABCA4 localisation in a similar way to C1490Y. Mutations affecting protein localisation were predicted to have greater pathogenicity in comparison to both splice-site mutations and mutations resulting in protein truncation.

The BDGP Splice Site Prediction programme, used to evaluate the intronic IVS38-10T>C mutation, did not detect any change in splicing of exon 39 in *ABCA4*. Rivera et al. (2000) investigated the functional effect of the IVS38-10T>C alteration by means of an Exon Trapping System (pSPL3b), which is a plasmid that consists of β-globin coding sequences that are of rabbit origin and are separated by a fragment of the HIV-*tat* gene which is present in humans (Burn et al. 1995). The wild type and mutant clones consisting of IVS38-10T>C were transfected into COS7 cells (Rivera et al. 2000); cell lines derived from monkey kidney cells. The mRNA was isolated from the cells and analysed using reverse-transcription (RT)-PCR. The results obtained revealed no difference in the product size of exon 39 and no exon 39 skipping between the wild type and mutant clones. In an alternate assay, Rivera et al. (2000) conducted further RT-PCR analysis using RNA obtained from an Epstein-Barr virus-transformed lymphoblastoid cell line from one of the patients found to be heterozygous for the mutation. These findings were in agreement with the previous results obtained using COS7 cell lines as well as with the result obtained from the BDGP Splice Site Prediction programme in the current study. Despite the fact that IVS38-10T>C is situated at the – 10 nucleotide position in a splice-acceptor site, it does not appear to affect the splicing mechanism of

exon 39. However, investigations into whether or not the wild type nucleotide is conserved amongst various species revealed 100% conservation, which could thus indicate the value of a wild type allele with no IVS38-10T>C alteration. Even though IVS38-10T>C was not found to affect the splicing of exon 39, it is possible that it could affect splicing of another exon in *ABCA4*. It may, however, not be a mutation and instead be in linkage disequilibrium with a mutation that has yet to be identified. But, if IVS38-10T>C is indeed a mutation, it could affect splicing of another exon or it could affect the ABCA4 protein some other way, acting as an enhancer or playing a role in methylation (Dr. R. Allikmets, personal communication). A recent study, conducted by Schindler et al. (2010) investigated the pathogenic contribution of recessive *ABCA4* alleles. The researchers found that disease-causing *ABCA4* alleles carrying the IVS38-10T>C mutation were one of the most common alleles in the cohort investigated had a severe effect on patients' visual acuities and visual fields. However, due to the fact that the function of this mutation could not be definitively determined, IVS38-10T>C was given the lowest pathogenicity score in comparison with the other six *ABCA4* mutations. If and when the exact nature and function of IVS38-10T>C, and its true pathogenicity is determined, its placement relative to the other six mutations analysed here may be reconsidered.

4.3.2 Statistical analysis

Due to the presence of a relatively large cohort of individuals with two disease-causing *ABCA4* variations, statistical analysis was conducted in the current study in order to further investigate the proposed genotype-phenotype model. As mentioned earlier, in other investigations into the relationship between the different *ABCA4* mutations and the severity of RDD, AOO was used as a measure of phenotype. AOO of the disease was the only clinical data available for the majority of the individuals in the current study, and was subsequently used as the dependent variable for the statistical analysis.

The results obtained from the ABCR400 microarray chip screening and subsequent family studies identified 86 individuals from 66 families with two *ABCA4* mutations. Reasonable clinical data with a clear AOO of the disease was available for 80 of the individuals. The results obtained from screening of the seven *ABCA4* mutations in this current study identified 28 patients with two mutations. The AOO was available for 26 of these individuals. This amounted to a total cohort of 106 individuals from 94 families on whom statistical analysis could be conducted and included a total of 23 different bi-allelic mutation combinations.

Although a variety of different mutation combinations of the *ABCA4* mutations were identified in the combined cohort, a relatively small number of individuals were identified for each genotype. Due to the types of statistical tests conducted on the data - the Kruskal-Wallis test and the Wilkoxon Rank Sum test – only certain mutation combinations were included in the analysis, resulting in fewer individuals being included in the final analyses. For the Kruskal-Wallis test, mutation combinations consisting of five or more individuals were included in the analysis. This resulted in a total of 76 individuals and a total of nine mutation combinations being included in the analysis. For the Wilkoxon Rank Sum test, mutation combinations consisting of three or more individuals were included for this analysis. This resulted in a total of 97 individuals and a total of 16 different mutation combinations being included in the analysis.

Statistical analysis using the Kruskal-Wallis test revealed a significant difference between the mutation combinations (genotype) and the AOO of disease in the patient cohort. When this result was investigated further using the Wilkoxon Rank Sum test, in order to determine which two paired mutation combinations of *ABCA4* mutations (genotype) were significantly different, 37 different paired combinations were shown to be significant. However, when the Bonferroni correction for multiple testing was performed, due to the large number of tests conducted (n = 120), no significant differences remained.

4.3.3 Analysis of the genotype-phenotype model

Inspection of the average AOO observed amongst the various mutation combinations, revealed that individuals who were found to carry one of the four most highly predicted pathogenic mutations (C1490Y, R602W, V256V, and R152X - in decreasing order of predicted pathogenicity) was found to be between six and 12 years of age. Thus, due to the early AOO, these four mutations do appear to be more pathogenic in comparison to the other three mutations.

An interesting observation in the present study is that patients found to carry at least one C1490Y mutation had an AOO of 20 years and less. This is in agreement with the findings of September et al. (2004), who also found individuals carrying a C1490Y mutation had an AOO of less than 20 years. Furthermore, this is also in agreement with the pathogenicity analysis which revealed this variant to be the most pathogenic.

Although the IVS38-10T>C was predicted as the least pathogenic, due to the fact that its function is still unknown, this mutation was detected in 23 patients, two of whom were found to be homozygous. In addition to this, the average AOO was found to range from six to 11 years of age, except for one individual (AOO = 20 years) who was found to carry a single G863A mutation of her other allele. These results indicate the possibility of this mutation being incorrectly placed as the least pathogenic of the seven mutations, and this might need to be re-assessed.

In the current study, three patients within the same family (Family 9) were found to have a relatively late AOO of disease, of between 30 and 36 years of age. All three patients were found to be compound heterozygotes for the G863A and R152X mutations. The AOO in these three patients is much later in comparison to individuals carrying a C1490Y mutation. It has been noted previously that an individual with a R152X mutation on one allele has only one functional allele remaining (Zhang et al. 1999). Familial co-segregation analysis showed that the G863A and R152X mutations were

transmitted on different alleles (i.e. "*in trans*"). Therefore the G863A mutation on the functional allele may be responsible for the later AOO observed in these patients. The later AOO, however, may not be solely due to the two *ABCA4* mutations themselves, but may involve as yet unidentified genetic and/or environmental modifiers specifically occurring in this family.

Interestingly, although only a small group (n = 12), all patients who were found to carry a G863A mutation had a later AOO. No homozygotes were detected; however, the results obtained thus far are in accordance with the literature and previous investigations, which suggested G863A has a mild effect on ABCA4 protein activity. This mutation was placed as the fifth highest pathogenic mutation. In addition, in most cases, individuals with an L2027F mutation, although not as delayed, were also found to have a slightly later AOO. One homozygote was detected for this mutation, and was shown to have an AOO of 28 years. This mutation was placed as the sixth highest pathogenic mutation. Based on what has been discussed, it is possible that that the G863A mutation is actually less pathogenic in comparison to L2027F. Nevertheless, in agreement with the pathogenicity predictions, these two mutations do appear to be less pathogenic in comparison to C1490Y, R602W, V256V, and R152X.

September et al. (2004) identified a patient who was a compound heterozygote for the V256V and R152X mutations. The disease phenotype for this individual was reported to change over time from STGD to a more severe phenotype, arRP. This was not the case for the three patients identified by the ABCR400 microarray screening who were also found to be compound heterozygotes for the V256V and R152X mutations, but whose disease phenotype remained STGD. However, adequate ophthalmologic data was not available for these three patients in order to conduct further analysis.

A change in disease phenotype from CRD to the more severe arRP phenotype was, however, observed for two patients who were shown to carry two different mutation combinations. One individual, RPM 303.2EBE, was found to be a compound

heterozygote for the C1490Y and V256V mutations, whilst the other, RP 428.1IGN, was found to be a compound heterozygote for the C1490Y and R602W mutations. This observation of C1490Y, albeit in a relatively small subset of re-classified patients, does support the previous predictions of its pathogenicity.

The re-classification or progression of disease phenotype from STGD to CRD was observed in three individuals who were also identified as having two of the seven mutations. Individual RPS 165.2IZA was found to be a compound heterozygote for the C1490Y and V256V mutations (same mutation combination as individual RPM 303.2EBE, discussed above). Individual RP 442.2HUG was found to be a compound heterozygote for the C1490Y and IVS38-10T>C mutations. Individual RPM 518.2JAR was found to be compound heterozygous for the IVS38-10T>C and R152X mutations. These individuals had AOO varying from five to 15 years. It is possible that these three mutation combinations are associated with the development of an AAR at an earlier age, however, more individuals per mutation category are required in order to test this assumption more effectively. As all three individuals were shown to carry different mutation combinations, one could not assess the exact mutation or mutation combination responsible for the progression of disease phenotype from STGD to CRD.

One individual in the patient cohort was shown to be homozygous for the V256V mutation (RPS 133.4CHA). September et al. (2004) noted the V256V mutation could have a severe effect on the ABCA4 protein, being associated with arRP and resulting in an earlier AOO. The AOO for this patient was eight years. Since the patient was shown to carry two “severe” V256V mutations, an early AOO was not surprising. Due to the severe nature of this mutation, the fact that it had been previously associated with arRP and the fact that the individual carried two copies of this mutation, it could have been expected that the individual would develop arRP, a more severe AAR. However, this individual was confirmed to have STGD and not arRP. Research has shown phenotype variability between individuals even from the same family. This could be a sign of possible environmental factors and the influence of other genes on the clinical outcome

of an *ABCA4* genotype (Ernest et al. 2009). Therefore, it is important to revisit all patients in order to follow up on the reclassification of patient phenotypes.

As mentioned earlier, previous research defining the relationship between the various *ABCA4* mutations detected, and the resulting retinal disease has been limited by a variety of factors. Attempts at defining the genotype-phenotype correlation have been restricted by: i) the limited number of families with the same mutations, ii) the limited number of families with the same combination of mutations for comparison, and iii) inadequate phenotypic or ophthalmologic data. The combined screening results obtained from this current study and from previous microarray screening revealed a total of 23 different mutation combinations that were identified amongst a total of 106 individuals (who had a specified AOO) in the patient cohort. This was seen as a large group on which to conduct statistical analysis, however, very few individuals were detected with the same mutation combination, resulting in some mutation combinations being excluded from the data analysis. This had an effect on the strength of the analysis. If more individuals had been identified with the same mutation combination and were also found to have a self-reported AOO, this could have aided in increasing the strength of the statistical analyses. It is also important to note that investigations into the effect of the mutations independently, as well as the effect of each mutation in combination with one another, could also prove to be essential in order to understand the biological effect they could have on an individual's phenotype.

4.4 Concluding remarks and future plans

The assays designed for this project utilise the existence of seven common *ABCA4* mutations in the South African population, specifically in those individuals of Afrikaner descent, and may be employed in the future to target or direct testing for both research and diagnostic purposes. These mutations had a pick-up rate of approximately 40% in the tested patient cohort. This assay will potentially reduce costs to both the research

laboratory and patient. It is important to note, however, that both the ABCR400 microarray and the present analysis were not screening for previously uncharacterised mutations (i.e. through re-sequencing). Although the seven mutations were found in a large percentage of South African patients, in the future novel variants might also be found to be diseasing-causing in a large proportion of individuals. Therefore, it is possible that these seven mutations may not be the only relatively common *ABCA4* mutations in the South African population.

Ideally it would have been more beneficial to design one assay to screen for all seven of the common *ABCA4* mutations. However, due to technical problems encountered in trying to create methods to detect some of the mutations and due to time constraints, this was not accomplished. Future work could therefore involve creating the single multiplex diagnostic assay, in order reduce i) the required labour time, and ii) testing costs for both the patient and the laboratory itself. The incorporation of the seven mutations into a single test could become appealing in a diagnostic setting, where it could be promoted by counsellors, medical geneticists and ophthalmologists to patients as a potential risk test (Sohrab et al. 2010).

Previous research has suggested the presence of founder mutations in the South African population, of mainly Caucasian Afrikaner descent. For this reason the Afrikaner-specific sub-cohort was of particular interest in the current study. A total of eight heterozygous mutations (4.73%) were detected in the individuals of Afrikaner descent, with a total of 12 heterozygous mutations (4.46%) being detected in the total control cohort. As much as an emphasis was made to recruit "Afrikaner" controls, the Caucasian control cohort was not selected on the basis of excluding any "Afrikaner" ancestry, and the findings may represent a level of "admixture". The results from the total control cohort indicated the carrier frequency in this particular population to be 4.46 per 100 individuals. This information can be used in order to assist counsellors in relating risk to patients and their families during a genetic counselling session.

Future work regarding *ABCA4* mutations could involve the investigation of the frequencies of these seven mutations in other population groups, such as Mixed Ancestry and indigenous Black African, for comparison with the frequencies observed in the Caucasian/Afrikaner population.

The presence of founder mutations in the South African population, of mainly Caucasian Afrikaner descent, needs to be validated for ABCA4. In order to determine whether the *ABCA4* mutations are true founder mutations, linkage analysis using microsatellite markers and single nucleotide polymorphisms (SNPs) spanning the region of each of the seven common *ABCA4* mutations may be used in groups of unrelated individuals with the same common mutation. If the genotyping results reveal that the mutant alleles of unrelated individuals occur on the same genomic haplotype, this would confirm the mutation's founder status. However, this may not have any practical relevance or clinical utility.

Although IVS38-10T>C was one of the more commonly occurring mutations detected in the patient cohort, its role as a mutation was brought into question when its predicted pathogencity was being investigated. Although it has been thought to play a role in splicing of the ABCA4 protein, the splice prediction programme and previous studies conducted by other researchers indicated that it had no effect on protein activity. It could perhaps be in linkage disequilibrium with another *ABCA4* mutation which has yet to be identified, or it could affect ABCA4 protein function in another way. Due to the relatively large size of the *ABCA4* gene (50 exons), functional studies could prove challenging. Future research on IVS38-10T>C could involve performing a similar procedure to that conducted by Burn et al. (1995) using an Exon Trapping System in retinal-derived cell lines, instead of in COS 7 cells. This could possibly be used to determine the true nature of this variant. However, as mentioned above, previous research has not shown any indication that this variant has an effect on protein the activity of ABCA4. Although it has previously been proposed that it affects splicing of exon 39, it is important to keep in mind that other exons could be involved. A decision as to which exon(s) to investigate using an Exon Trapping System would therefore need

to be considered. The study by Burn et al. (1995) utilised COS 7 cells; since these are cell lines derived from monkey kidney cells, future work using retinal-derived cell lines, could be more beneficial. Although these cells might not mimic the human eye, in comparison to COS 7 cells, they could provide more insight into the functional implication of the IVS38-10T>C variant on the ABCA4 protein.

For a test of genotype versus phenotype, the AOO of the disease was tested against the different paired combinations of *ABCA4* disease-causing mutations. However, despite AOO being the most accessible clinical measure, eight out of 114 individuals were excluded from the analysis as they did not have a specified or recorded AOO. Insufficient clinical data was found to impede the investigation of the genotype-phenotype correlation. If the AOO of disease for these eight individuals had been available, it may possibly have aided in increasing the strength of the statistical analyses. It should be noted, however, that the cohort of 106 patients included in the analyses consisted of some related individuals; it is therefore possible that other genetic and environmental factors may also influence the AOO of disease. In future, it will be useful to obtain consistent clinical information for each affected individual to strengthen the power of the statistical analyses, towards a better understanding of the genotype-phenotype correlation. Screening more affected individuals, particularly unrelated patients in order to minimise the effect of any familial factors, and identifying more individuals with the same paired combination of disease-causing variations in *ABCA4* may also aid in increasing the understanding of the genotype-phenotype model.

The exact function of each domain of the ABCA4 protein is still being elucidated. Recently, research has shown specific interaction of the ECD2 domain with all-*trans*-retinal (Biswas-Fiss et al. 2010). Furthermore it was shown that three *ABCA4* mutations (R1443H, W1408L, and C1488R) significantly alter the structure of the above-mentioned domain, as well as interfering with its normal functioning with all-*trans*-retinal. It is therefore important in the first instance to determine the function of each ABCA4 domain, followed by the effect of each mutation on domain functioning. In addition, retinal scanning could be employed to visualise the effect of different types of

ABCA4 mutations. For example, ophthalmologists could examine patients who have mutations known to be involved with protein misfolding and compare their scanning results with patients who have mutations known to result in protein truncation. Alternatively, ophthalmologists could also examine profile differences between homozygote and heterozygote individuals. In addition, choroidal imaging could be utilised to visualise the structural changes of the choriocapillaris related to RPE cell death (Yeoh et al. 2010). Although these techniques are costly, if they prove to be successful in developed countries, in the future they may be applied in emerging nations.

In the current study the presence of a significant cohort of South African individuals diagnosed as having STGD, with the same common mutations in various combinations, was essential in order to i) improve the understanding of the relationship between *ABCA4* mutation combination and patient phenotype and ii) develop a relevant model algorithm. In future, this may enhance the knowledge regarding the impact of each mutation or the specific pathogenic effect of each allele on the ABCA4 protein (Schindler et al. 2010). These results may well provide the basis of prognosis information to patients, thereby enhancing the clinical utility of identifying *ABCA4* mutations.

Until treatment options for AARs become available, recommendations that may limit the progression of these diseases include advising affected individuals to: i) limit their exposure to direct sunlight, and ii) avoid the intake of Vitamin A-containing products, as it is Vitamin A's derivatives that accumulate when the *ABCA4* gene is defective (Sohrab et al. 2010).

Finally, the recent RPE65 clinical trials for Leber's Congenital Amaurosis have been well publicised and could provide the link to possible *ABCA4* clinical studies (Bainbridge et al. 2008; Smith et al. 2009). However, various factors need to be considered for clinical studies, such as the criteria for patient selection including i) a patient's age and disease stage, when investigating an early-onset disease, and ii) the

level of genetic heterogeneity of the individuals included, i.e. the particular the genotype inclusion or exclusion criteria (Fishman 2005; Cideciyan et al. 2009). In addition, understanding the progression of vision loss in patients with *ABCA4* mutations by assessing the loss of sensitivity in both rod and cone cells could also play a vital role in potential therapeutic approaches (Cideciyan et al. 2009). The RPE65 clinical trial requirements and findings could aid in the selection of patients for future *ABCA4*-related studies. A clinically and molecularly defined cohort of patients will therefore be useful for *ABCA4*-gene replacement therapies. A gene-based therapy for STGD, known as StarGen, involves the use of a LentiVirus vector system, and is currently being investigated by the National Neurovision Research Institute in collaboration with Oxford Biomedica (http://www.oxfordbiomedica.co.uk). The initiation of the patient clinical trials is due to start at the end of 2010.

Chapter 5: References

ABI PRISM® SNaPshot™ Multiplex Kit Protocol (2000) Applied Biosystems

Aguirre-Lamban J, Riveiro-Alvarez R, Maia-Lopes S, Cantalapiedra D, Vallespin E, Avila-Fernandez A, Villaverde-Montero C, Trujillo-Tiebas M, Ramos C, Ayuso C (2008) Molecular analysis of the *ABCA4* gene for reliable detection of allelic variations in Spanish patients: identification of 21 novel variants. Br J Ophthalmol 93:614-621

Akiyama M (2006) Harlequin ichthyosis and other autosomal recessive congenital ichthyoses: The underlying genetic defects and panthomechanisms. J Dermatol Sci 42(2):83-89.

Allikmets R, Singh N, Sun H, Shroyer NF, Hutchinson A, Chidambaram A, Gerrad B, Baird L, Stauffer D, Peiffer A, Rattner A, Smallwood P, Li Y, Anderson KL, Lewis RA, Nathans J, Leppert M, Dean M, Lupski JR (1997a) A photoreceptor cell-specific ATP-binding transporter gene (*ABCR*) is mutated in recessive Stargardt macular dystrophy. Nat Genet 15: 236-246

Allikmets R, Shroyer NF, Singh N, Seddon JM, Lewis RA, Bernstein PS, Peiffer A, Zabriskie A, Li Y, Hutchinson A, Dean M, Lupski JR, Leppert M (1997b) Mutation of the Stargardt disease gene (*ABCR*) in age-related macular degeneration. Sci 277:1805-1807

Allikmets R, Wasserman WW, Hutchinson A, Smallwood P, Nathans J, Rogan PK, Schneider TD, Dean (1998) Organization of the *ABCR* gene: analysis of promoter and splice junction sequences. Gene 215:111-122

Allikmets R (2007) Stargardt disease: from gene discovery to therapy. In: Tombran-Tink J, Barnstable CJ (ed) Opthalmology research: Retinal degenerations: biology, diagnostics, and therapeutics. Humana Press Inc., U.S, 105-118

Anderson DH, Fisher SK, Steinberg RH (1978) Mammalian cones: Disc shedding, phagocytosis, and renewal. Assoc for Res in Vis and Ophthal 17(2):117-133

Bainbridge JWB, Smith AJ, Barker SS, Robbie S, Henderson R, Balaggan K, Viswanathan A, Holder GE, Stockman A, Tyler N, Petersen-Jones S, Bhattacharya SS, Thasher AJ, Fitzke FW, Carter BJ, Rubin GS, Moore AT, Ali RR (2008) Effect

of gene therapy on visual function in Leber's Congenital Amaurosis. N Engl J Med 358(21):2231-2239

Biswas-Fiss EE, Kurpad DS, Joshi K, Biswas SB (2010) Interaction of extracellular domain 2 of the human retina-specific ATP-binding cassette transporter (ABCA4) with all-*trans*-retinal. J Biol Chem 285(25):19372-19383

Boon CJF, Klevering BJ, Keunen JEE, Hoyng CB, Theelen T (2008) Fundus autofluorescence imaging of retinal dystrophies. Vis Res 48:2569-2577

Botha MC, Beighton P (1983) Inherited disorders in the Afrikaner population of southern Africa. Part I. Historical and demographic background, cardiovascular, neurological, metabolic and intestinal conditions. SA Med J 64:609-612

Briggs CE, Rucinski D, Rosenfeld PJ, Hirose T, Berson EL, Dryja TP (2001) Mutations in *ABCR* (*ABCA4*) in patients with Stargardt macular degeneration or cone-rod degeneration. Invest Ophthalmol Vis Sci 42(10):2229-2236

Brunham LR, Singaraja RR, Pape TD, Kejariwal A, Thomas PD, Hayden MR (2005) Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the *ABCA1* gene. PLOS Genetics 1(6):739-747

Bungert S, Molday LL, Molday RS (2001) Membrane topology of the ATP binding cassette transporter ABCR and its relationship to ABC1 and related ABCA tranporters. J Biol Chem 276(26):23539-23546

Burn TC, Connors TD, Klinger KW, Landes GM (1995) Increased exon-trapping through modifications to the pSPL3 splicing vector. Gene 161:183-187

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR (2003) ESEfinder: a web-based resource to identify exonic splicing enhancers. Nucl Acids Res 31(13):3568-3571

Cheong N, Madesh M, Gonzales LW, Zhao M, Yu K, Ballard PL, Shumann H (2006) Functional and trafficking defects in ATP binding cassette A3 mutants associated with respiratory distress syndrome. J Biol Chem 281(14)9791-9800

Cideciyan AV, Aleman TS, Swider M, Schwartz SB, Steinberg JD, Brucker AJ, Maguire AM, Bennett J, Stone EM, Jacobson SG (2004) Mutations in *ABCA4* result in accumulation of lipofuscin before slowing of the retinoid cycle: a reappraisal of the human disease sequence. Hum Mol Genet 13(5):525–534

Cideciyan AV, Swider M, Aleman TS, Tsybovsky Y, Schwartz SB, Windsor EAM, Roman, AJ, Sumaroka A, Steinberg JD, Jacobson SG, Stone EM, Palczewski K

(2009) *ABCA4* disease progression and a prposed strategy for gene therapy. Hum Mol Genet 18(5):931-941

Coles WH (1989) Ophthalmology: A diagnostic text. Chapter 4 Retina. Williams & Wilkins Baltimore, Maryland. Editor: Carol-Lynne Brown, printed in the USA ISBN 0-683-02056-0 60-137

Cremers FPM, van der Pol DJR, van Driel N, den Hollander AI, van Haren FJJ, Knoers NVAM, Tijmes N, Bergen AAB, Rohrschneider K, Blankenagel A, Pinckers AJLG, Deutman AF, Hoyng CB (1998) Autosomal recessive retinitis pigmentosa and con-rod dystrophy caused by splice site mutations in the Stargardt's disease gene *ABCR*. Hum Mol Genet 7(3): 355-362

Dean M, Allikmets R (1995) Evolution of ATP-binding cassette transporter genes. Curr Opin Genet & Dev 5:779-785

Dean M, Hamon Y, Chimini G (2001) The human ATP-binding cassette (ABC) transporter superfamily. J Lipid Res 42:1007-1017

Dean M, Allikmets R (2001) Complete characterisation of the human ABC gene family. J of Bioenerg & Biomembr 33(6):475-479

Ernest PJG, Boon CJF, Klevering BJ, Hoefsloot LH, Hoyng CB (2009) Outcome of *ABCA4* microarray screening in routine clinical practice. Mol Vis 15:2841-2847

Essner E, Roszka JR, Schreiber JH (1978) Phagocytosis and surface morphology in cultured retinal pigment epithelial cells. Assoc for Res in Vis and Ophthalmol 17(11):1040-1048

Fein A, Szuts EZ (1982) Photoreceptors, their role in vision. Cambridge University Press, United States of America

Ferrer-Costa C, Gelpi JL, Zamakola L, Parraga I, de la Cruz X, Orozco M (2005) PMUT: a web-based tool for the annotation of pathological mutations on proteins. Bioinformatics 21(14):3176-3178

Fishman GA (2005) Challenges associated with clinical trials for inherited and orphan clinical trials. Retina 25(8 Suppl):S10-12

Franze K, Grosche J, Skatchkov SN, Schinkinger S, Foja C, Schild D, Uckermann O, Travis K, Reichenbach A, Guck J (2007) Müller cells are living optical fibers in the vertebrate retina. Proc Natl Acad Sci USA 104(20):8287-8292

Fumagalli A, Ferrari M, Soriani N, Gessi A, Foglieni B, Martina E, Manitto MP, Brancato R, Dean M, Allikmets R, Cremonesi L (2001) Mutational scanning of the ABCR gene with double-gradient denaturing-gradient gel electrophoresis (DG-DGGE) in Italian Stargardt disease patients. Hum Genet 109:326-338

Gerber S, Rozet JM, van der Pol TJR, Hoyng CB, Munnich A, Blankenagel A, Kaplan J, Cremers FPM (1998) Complete exon-intron structure of the retina-specific ATP binding transporter gene (*ABCR*) allows for the identification of novel mutations underlying Stargardt disease. Genomics 48:139-142

Gevers Wieland, Casciola LAF, Fourie AM, Sanan DA, Coetzee GA, van der Westhuyzen DR (1987) Defective LDL receptors that are common in a large population: familial hypercholesterolemia in South Africa. Biol Chem368:1233-1243

Green J, Britten N (1998) Qualitative research and evidence based medicine. Brit Med J 316:1230-1240

Grogan J, Suter R (1999) Sense organs in Biology 12: text book and creative work book. Parow East, Cape Town

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98

Hoti SL, Dhamodharan R, Subramaniyan K, Das PK (2009) An allele specific PCR assay for screening for drug resistance among *Wuchereria bancrofti* populations in India. Indian J Med Res 130: 193-199

Jaakson K, Zernant J, Kulm M, Hutchinson A, Tonisson N, Glavac D, Ravnik-Glavac M, Hawlina M, Meltzer MR, Caruso RC, Testa F, Maugeri A, Hoyng CB, Gouras P, Simonelli F, Lewis RA, Lupski JR, Cremers FPM, Allikmets R (2003) Genotyping microarray (gene chip) for the *ABCR* (*ABCA4*) gene. Hum Mut 22:395-403

Kawamura S, Tachibanaki S (2008) Rod and cone photoreceptors: Molecular basis of the difference in their physiology. Comp Biochem & Physiol (Part A) 150:369-377

Kaplan J, Gerber S, Larget-Piet D, Rozet JM, Dollfus H, Dufier JL, Odent S, Postel-Vinay A, Janin N, Briard ML, Frezel J, Munnich A (1993) A gene for Stargardt's disease (fundus flavimaculatus) maps to the short arm of chromosome 1. Nat Genet 5:308-311

Keller G, Warrack B (2000) Statistics for Management and Economics. Duxbury, USA

Klevering BJ, Yzer S, Rohrschnieder K, Zonneveld M, Allikmets R, van den Born LI, Maugeri A, Hoyng CB, Cremers FPM (2004) Microarray-based analysis of the *ABCA4* (*ABCR*) gene in autosomal recessive cone-rod dystrophy and retinitis pigmentosa. Eur J Hum Genet 12:1024-1032

Klevering BJ, Deutman AF, Maugeri A, Cremers FPM, Hoyng CB (2005) The spectrum of retinal phenotypes caused by mutations in the *ABCA4* gene. Graefe's Arch Clin Exp Ophthalmol 243:90-100

Kniazeva M, Chiang MF, Anduze AL, Zack DJ, Han M, Zhang K (1999) A new locus for autosomal dominant Stargardt-like disease maps to chromosome 4. Am J Hum Genet 64:1394-1399

Kwok S, Chang S, Sninsky JJ, Wang A (1994) A guide to the design and use of mismatched and degenerate primers. Genome Res 3:S39-S47

Latorra D, Arar K, Hurley M (2003a) Design considerations and effects of LNA in PCR primers. Molecular and Cellular Probes 17:253-259

Latorra D, Campbell K, Wolter A, Hurley JM (2003b) Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers. Hum Mut 22:79-85

Lewis RA, Shroyer NF, Singh N, Allikmets R, Hutchinson A, Li Y, Lupski JR, Leppert M, Dean M (1999) Genotype/phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, *ABCR*, in Stargardt disease. Am J Hum Genet 64:422-434

Loo TW, Clarke DM (2008) Mutational analysis of ABC proteins. Arch of Biochem & Biophys 476:51-64

Lorenz B, Preising MN (2005) Age matters- thoughts on a grading system for *ABCA4* mutations. Graefe's Arch Clin Exp Ophthalmol 243:97-89

Martinez-Mir A, Paloma E, Allikmets R, Ayuso C, del Rio T, Dean M, Vilageliu L, Gonzàlez-Duarte R, Balcells S (1998) Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene *ABCR*. Nat Genet 18:11-12

Maugeri A, van Driel MA, van de Pol DJR, Klevering BJ, van Haren FJJ, Tijmes N, Bergen AAB, Rohrschneider K, Blankenagel A, Pinckers AJLG, Dahl N, Brunner HG, Deutman AF, Hoyng CB, Cremers FPM (1999) The 2588G→C mutation in the *ABCR* gene is a mild frequent founder mutation in the Western European population

and allows the classification of *ABCR* mutations in patients with Stargardt disease. Am J Hum Genet 64:1024-1035

Maugeri A, Flothmann K, Hemmrich N, Ingvast S, Jorge P, Paloma E, Patel R, Roset JM, Tammur J, Testa F, Balcells S, Bird AC, Brunner HG, Hoyng CB, Metspalu A, Simonelli F, Allikmets R, Bhattacharya SS, D'Urso M, Gonzalez-Duarte R, Kaplan J, te Meerman GJ, Santos R, Schwartz M, Van Camp G, Wadelius C, Weber BHF, Cremers FPM (2002) The *ABCA4* 2588G→C Stargardt mutation: single origin and increasing frequency from Aouth-West to North-East Europe. Eur J Hum Genet 10:197-203

Michaelides M, Hunt DM, Moore AT (2003) The genetics of inherited macular dystrophies. J Med Genet 40:641-650

Miyazono S, Shimauchi-Matsukawa Y, Tachibanaki S, Kawamura S (2008) Highly efficient retinal metabolism in cones. Proc Natl Acad Sci USA 105(41):16051-16056

Molday RS (1998) Photoreceptor Membrane Proteins, Phototransduction, and Retinal Degenerative Diseases: The Friedenwald Lecture. Invest Ophthalmol Vis Sci 39(13):2493-2513

Molday LL, Rabin AR, Molday RS (2000) *ABCR* expression in foveal cone photoreceptors and its role in Stargardt macular dystrophy. Nat Genet 25: 257-258

Molday RS (2007) ATP-binding cassette transporter ABCA4: Molecular properties and role in vision and macular degeneration. J Bioenerg Biomembr 39:507–517

Molday RS, Zhong M, Quazi F (2009) The role of the photoreceptor ABC transporter ABCA4 in lipid transport and Stargardt macular degeneration. Biochim Biophys Acta 1791(7):573-583

Mouritzen P, Nielson AT, Pfundheller HM, Choleva Y, Kongsbak L, Møller S (2003) Single nucleotide polymorphism genotyping using locked nucleic acid (LNA^{TM}). Expert Rev Mol Diagn 3(1):27-38

Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. Nucl Acids Res 31(13)3812-3814

Oswald AH, Goldblatt J, Sampson G, Clokie R, Beighton P (1985) Retinitis Pigmentosa in South Africa. SA Med J 68:863-866

Passerini I, Sodi A, Giambene B, Mariottini A, Menchini U, Torricelli F (2009) Novel mutations of the ABCR gene in patients with Stargardt disease. Eye 1-7

Patel N, Adewoyin T, Chong NV (2008) Age-related macular degeneration: a perspective on genetic studies. Eye 22:768-776

Pennisi E (1997) Gene found for the fading eyesight of old age. Sci 277:1765-1766

Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucl Acids Res 30(17)**:** 3894-900

Ramesar R, September A, Rebello G, Greenberg J, Goliath R (2001) Migratory history of populations and its use in determining research directions for retinal degenerative disorders. In: Anderson RE, LaVail MM, Hollyfield JG (ed) New insights into retinal degenerative diseases, Kluwer Academic/Plenum Publishers, pp335-338

Rivera A, White K, Stöhr H, Steiner K, Hemmich N, Grimm T, Jurklies B, Lorenz B, Scholl HPN, Apfelstedt-Sylla E, Weber BHF (2000) A comprehensive survey of sequence variation in the *ABCA4* (*ABCR*) gene in Stargardt disease and age-related macular degeneration. Am J Hum Genet 67:800-813

Roberts LJ, Ramesar RS, Greenberg J (2009) Clinical utility of the ABCR400 Microarray: Basing a genetic service on a commercial gene chip. Arch Ophthalmol 127(4):549-554

Rosenberg T, Klie F, Garred P, Schwartz M (2007) N965S is a common *ABCA4* variant in Stargardt-related retinopathies in the Danish population. Mol Vis 13:192-1969

Rozet JM, Gerber S, Ghazi I, Perrault I, Ducroq D, Souied E, Cabot A, Dufier JL, Munnich A, Kaplan J (1999) Mutations of the retinal specific ATP binding transporter gene (*ABCR*) in a single family segregating both autosomal recessive retinitis pigmentosa RP19 and Stargardt disease: evidence of clinical heterogeneity at this locus. J Med Genet 36:447-451

Rust S, Rosier M, Funke H, Real J, Amoura Z, Piette JC, Deleuze JF, Brewer HB, Duverger N, Denélfe P, Assmann G (1999) Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. Nat Genet 22:352-355

Samuels ML, Witmer JA (1999) Statistics for the Life Sciences. Prentice-Hall, Inc., USA

Saurin W, Hofnung M, Dassa E (1999) Getting in or out: Early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. J Mol Evol 48:22-41

Schindler EI, Nylen EL, Ko AC, Affatigato LM, Heggen AC, Wang K, Sheffield VC, Stone EM (2010) Deducing the pathogenic contribution of recessive *ABCA4* alleles in an outbred population. Hum Mol Genet doi: 10.1093/hmg/ddq284

Schütt F, Davies S, Kopitz J, Holz FG, Boulton ME (2000) Photodamage to human RPE cells by A2-E, a retinoid component of lipofuscin. Invest Ophthalmol Vis Sci 41(8):2303-2308

September, AV (2003) The molecular investigation of Stargardt disease in South Africa. PhD Dissertation, University of Cape Town

September AV, Vorster AA, Ramesar RS, Greenberg LJ (2004) Mutation spectrum and founder chromosomes for the *ABCA4* gene in South African patients with Stargardt disease. Invest Ophthalmol Vis Sci 45(6):1705-1711

Shroyer HF, Lewis RA, Allikmets R, Singh N, Dean M, Leppert M, Lupski JR (1999) The rod photoreceptor ATP-binding cassette transporter gene, *ABCR*, and retinal disease: from monogenic to multifactorial. Vis Res 39:2537-2544

Shroyer NF, Lewis RA, Yatsenko AN, Lupski JR (2001) Null missense *ABCR* (*ABCA4*) mutations in a family with Stargardt disease and Retinitis Pigmentosa. Invest Ophthalmol Vis Sci 42(12):2757-2761

Smit AL, van Dijk DE (1970) Receptor organs and sensations. Introduction to modern biology, 1st edn. Maskew Miller LTD., Cape Town, pp506-514.

Smith AJ, Bainbridge JW, Ali RR (2009) Prospects for retinal gene replacement therapy. Trends in Genet 25(4):156-165

Sohrab MA, Allikmets R, Guarnaccia MM, Smith T (2010) Preimplantation genetic diagnosis for Stargardt disease. Am J Ophthalmol 149(4):651-655

Stone EM, Nicholas BE, Kimura AE, Weingeist TA, Drack A, Sheffield VC (1994) Clinical features of Stargardt-like dominant progressive macular dystrophy with genetic linkage to chromosome 6q. Arch Ophthalmol 112:763-772

Stone EM, Webster AR, Vandenburgh K, Streb LM, Hockey RR, Lotery AJ, Sheffield VC (1998) Allelic variation in *ABCR* associated with Stargardt disease but not with age-related macular degeneration. Nat Genet 20:328-329

Sullivan JM (2009) Focus on molecules: *ABCA4* (*ABCR*) – An import-directed photoreceptor retinoid flipase. Exp Eye Res 89:602-603

Sun H, Nathans J (1997) Stargardt's ABCR is localized to the disc membrane of retinal rod outer segments. Nat Genet 17:15-16

Sun H, Smallwood PM, Nathans J (2000) Biochemical defects in *ABCR* protein variants associated with human retinopathies. Nat Genet 26:242-246

Thomas PD, Campbell MJ, Kerjariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechnia A (2003) PANTHER: A library of protein families and subfamilies indexed by function. Genome Res 13:2129-2141

Thomas PD, Kejariwal A (2004) Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: Evolutionary evidence for differences in molecular effects. Proc Nat Acad Sci USA 101(43):15398-15403

Thompson DA, Gal A (2003) Vitamin A metabolism in the retinal pigment epithelium: genes, mutations, and diseases. Progr Retinal & Eye Res 22:683-703

Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. BioTechniques 37:226-231

Valverde D, Riveiro-Alvarez R, Bernal S, Jaakson K, Baiget M, Navarro R, Ayuso C. (2006) Microarray-based mutation analysis of the *ABCA4* gene in Spanish patients with Stargardt disease: Evidence of a prevalent mutated allele. Mol Vis 12:902-908

Van Driel MA, Maugeri A, Klevering BJ, Hoyng CB, Cremers FPM (1998) *ABCR* unites what opthalmologists divide(s). Ophthalmic Genet 3:117-122

WAVE System Operator's Guide

Weber BHF (1998) Recent advances in the molecular genetics of hereditary retinal dystrophies with primary involvement of the macula. Acta Anat 162:65-74

Webster AR, Héon E, Lotery AJ, Vandenburgh K, Casavant TL, Oh KT, Beck G, Fishman GA, Lam BL, Levin A, Heckenlively JR, Jacobson SG, Weleber RG, Sheffield VC, Stone ED (2001) An analysis of allelic variation in the *ACBA4* gene. Invest Ophthalmol Vis Sci 42(6):1179-1189

Wisznieski W, Zaremba CM, Yatsenko AN, Jamrich M, Wensel TG, Lewis RA, Lupski JR (2005) *ABCA4* mutations causing mislocalisation are found frequently in patients with severe retinal dystrophies. Hum Mol Genet 14(19):2769-2778

Xiao W, Oefner PJ (2001) Denaturing High-Performance Liquid Chromatography: A Review. Hum Mut 17:439-474

Yeoh J, Rahman W, Chen F, Hooper C, Patel P, Tufail A, Webster AR, Moore AT, DaCruz L (2010) Choroidal imaging in inherited retinal disease using the technique of enhanced depth imaging optical coherence tomography. Graefes Arch Clin Exp Ophthalmol doi: 10.1007/s00417-010-1437-3

Zhang K, Bither PP, Park R, Donoso LA, Seidman JG, Seidman CE (1994) A dominant Stargardt's macular dystrophy locus maps to chromosome 13q34. Arch Ophthalmol 112:759-764

Zhang K, Kniazeva M, Hutchinson A, Han M, Allikmets R (1999) The ABCR gene in recessive and dominant Startgardt diseases: A genetics pathway in macular degeneration. Genomics 60:234-237

Zhong M, Molday LL, Molday RS (2008) The role of the c-terminus of the photoreceptor ABCA4 transporter in protein folding, function and retinal degenerative diseases. J Biol Chem 284(6):3640-3649

Websites

http://www.biochem.ubc.ca/fac_research/faculty/molday.html (accessed on 3 March 2009)

http://illustration.stevetatler.co.uk/illustration6.htm (accessed on 5 August 2009)

http://medgadget.com/archives/2006/04/nonlightsensiti_1.html (accessed on 5 August 2009)

http://www.ganfyd.org/index.php?title=Image:SchematicPhotoreceptorsRetina.png (accessed on 3 March 2009)

http://www.ahaf.org/macular/about/understanding/normal-macula.html (accessed on 18 October 2010)

http://www.healthyfellow.com/150/vitamins-for-macular-degeneration/ (accessed on 18 October 2010)

http://www.medterms.com/script/main/art.asp?articlekey=10018 (accessed on 7 October 2009)

http://www.stlukeseye.com/Conditions/Stargardts.asp (accessed on 12 September 2009)

http://www.asperbio.com (accessed on 11 August 2010)

http://www.oxfordbiomedica.co.uk (accessed on 11 August 2010)

The *Human Genetics* Journal convention of referencing was used for this thesis.

Appendix 1: DNA consent form

REQUEST FOR MOLECULAR STUDIES (DNA) BLOOD FORM

DIVISION OF HUMAN GENETICS, WERNHER & BEIT NORTH

FACULTY OF HEALTH SCIENCES, UNIVERSITY OF CAPE TOWN, OBSERVATORY, 7925

TEL: 021 406-6299 FAX: 021 406-6826 EMAIL: jacquie.greenberg@uct.ac.za

SURNAME: _____ *NAME:* _____

Mother:

DATE OF BIRTH: ____/ ____/_____ DETAILED ETHNIC ORIGIN: **Father:**_____

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

SEX: FEMALE - ☐ MALE - ☐ Number of Children ___ Number of affected family members: ___

CONTACT ADDRESS: _____

_____**CODE** _____

TEL: …….. FAX: …… E-mail ……..

NAME OF REFERRING DOCTOR: _____ HOSPITAL: _____

FAX: _____ TEL_____E-mail: _____

ADDRESS: _____

REASON FOR REFERRAL (CLINICAL DIAGNOSIS):

AFFECTED ☐ AT RISK ☐ CARRIER ☐ SPOUSE ☐ QUERY ☐ UNAFFECTED ☐

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

ADDITIONAL FAMILY HISTORY _____ **Pedigree on separate page please**

ADDITIONAL DISORDERS (APPARENT OR PREVIOUSLY TREATED): _____

RELEVANT CLINICAL DETAILS: _____

PHYSICAL DISABILITY ☐ MENTAL RETARDATION ☐ DEAFNESS ☐ IMPAIRED VISION ☐
NIGHT BLINDNESS ☐ AGE OF ONSET: _____ OTHER: _____

Have samples from this patient been sent to a DNA lab before? Yes / No / Don't Know. If "Yes", Where: _____

Medical Aid _____ Medical Aid No._____

For Laboratory use only: DNA number: _____ *Vol.Blood:* _____ *(ml)*

Other: _____ *Date Received: Year:* _____ *Month:*____ *Day:*____ *Computer Index No:* _____

SPECIMEN TUBES REQUIRED: 2 X 4 ml 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.

COURIER SERVICES CAN BE ARRANGED BY THE DIVISION OF HUMAN GENETICS UPON REQUEST

BLOODS ARE CODED ON ARRIVAL IN LABORATORY ACCORDING TO FAMILY NAME

RESULTS ARE GIVEN TO PARTICIPANTS ACCORDING TO ESTABLISHED PROTOCOL WHEN AVAILABLE

CONSENT FORM REQUIRED FOR DNA ANALYSIS AND STORAGE

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 Research Ethics Committee, provided that any information from such research will remain confidential

4. The results of the analysis carried out on this sample of stored biological material 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's name or other doctors involved in my care: _____

Names of family members: _____

Other persons: _____

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.
- 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 IN A LANGUAGE THAT I UNDERSTAND, AS WELL AS THE QUESTIONS THAT WERE ANSWERED BY: _____

DATE _____ /_____ /____ **PLACE:** _____

PATIENT'S SIGNATURE: _____ **WITNESS CONSENT:** _____

Appendix 2: Reagents, buffers and protocols

10x TBE buffer (stock):

216g (0.89M) Tris (B&M Scientific cc, Cape Town, SA)

110g (0.89M) Boric Acid (ICN Biomedicals Inc., USA)

14.8g (0.04M) EDTA (BDH Electron® Laboratory Supplies, UK)

Made up to 2L with dH_2O

1x TBE buffer (working stock):

A 10 X dilution of the TBE buffer (stock) was made up using dH_2O

1% agarose gel (100ml):

1g agarose (Lonza, Seakem® LE, Rockland, USA)

100ml 1 x TBE

6µl (10mg/ml) EtBr (Sigma-Aldrich, USA)

1.5% agarose gel (100ml):

1.5g agarose (Lonza, Seakem® LE, Rockland, USA)

100ml 1 x TBE

6µl (10mg/ml) EtBr (Sigma-Aldrich, USA)

2% agarose gel (100ml):

2g agarose (Lonza, Seakem® LE, Rockland, USA)

100ml 1 x TBE

6µl (10mg/ml) EtBr (Sigma-Aldrich, USA)

Loading Dye:

0.125g (0.25%) Bromophenol blue (Merck KgaA, Germany)

40g (40% w/v) sucrose (Merck Pty Ltd., SA)

pH 8.0

Made up to 50 ml with SABAX dH_2O (Adcock Ingram, Johannesburg, SA)

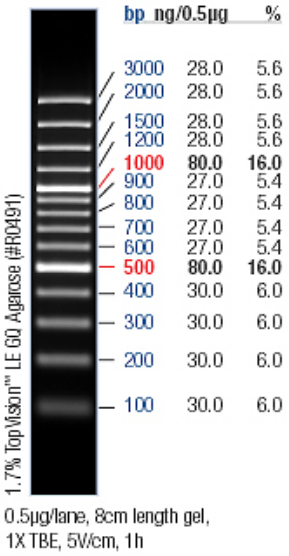
Molecular weight marker:

GeneRuler™ 100bp DNA Ladder Plus (Fermentas Life Sciences, Hanover, USA):

25µl molecular weight marker stock solution

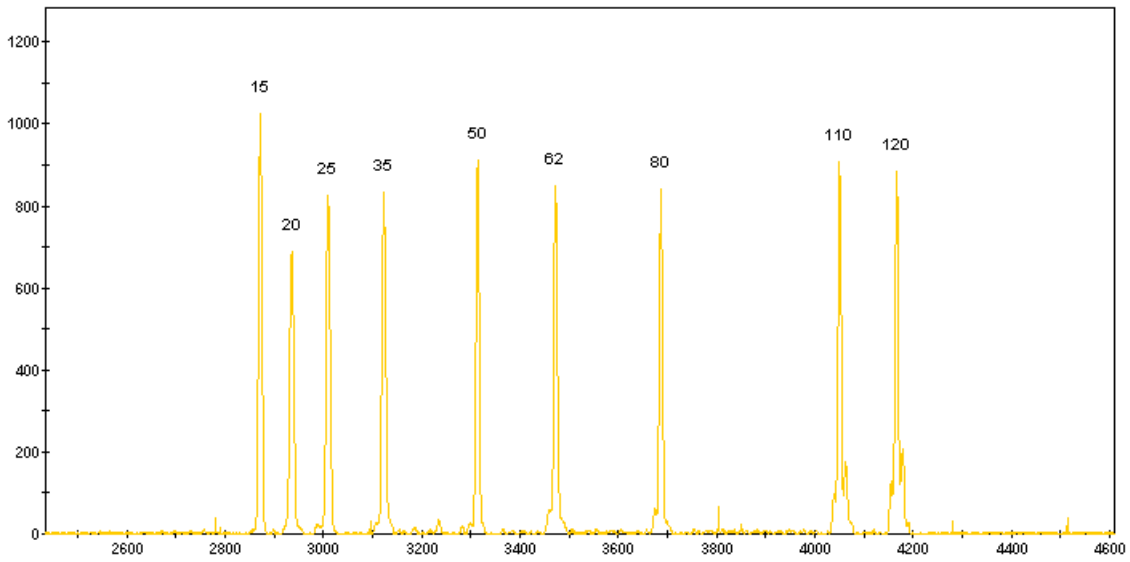
225µl SABAX dH_2O (Adcock Ingram, Johannesburg, SA)

250µl loading dye



GeneScan™ -120LIZ™ size standard

(Applied Biosystems, Warrington, UK)



Alternative PCR purification procedure used for exon 5 (V256V) PCR products

General procedure for NucleoFast 96 PCR (Machery-Nagel AgmH & Co. KG, Düren, Germany) - April 2008/Rev. 03

Standard protocol for the purification of PCR products under vacuum:

(Designed for PCR reaction volumes of between 20-100µl)

1. PCR products are transferred to the NucleoFast 96 PCR plate.
2. PCR contaminants are removed by ultrafiltration. This is performed by placing the NucleoFast 96 PCR plate on a suitable vacuum manifold where a vacuum is then applied. A typical vacuum has to be applied for 10-15 min for a PCR product volume of 50-100µl. Once the products have passed through the plate completely, the vacuum is applied for an additional 30-60 seconds.
3. Vacuum is released for 60-90 seconds, a volume of 100µl of RNase-free dH_2O is dispensed into each well on the plate and the vacuum is applied until the water has passed through the membrane. Thereafter the vacuum is applied for an additional 30-60 seconds. (This is an optional wash step).

Ethanol precipitation protocol used for cycle sequencing purification:

1. A volume of 2µl sodium acetate (3mM, pH5.5) was added to the sequencing product.
2. A 2 X volume of absolute ethanol (Merck Chemicals (Pty) Ltd, Gauteng, SA) was also added.
3. This was followed by an incubation period at -20°C for more than 30 min.
4. The samples were then centrifuged at 10 000 revolutions per minute (rpm) for 10 min.
5. The supernatant was removed.
6. A 2 X volume of cold 70% ethanol was added to the tubes.
7. The samples were then centrifuged at 10 000rpm for 10 min.
8. The supernatant was removed.
9. The samples were then air-dried for approximately 60 min.
10. The samples were re-suspended in 10µl SABAX dH_2O (Adcock Ingram, Johannesburg, SA).

Appendix 3: Sequence annotations of the seven exons interrogated in the study

The exonic regions are written in capital letters, whereas the intronic regions are written in lower case.

The external primary PCR primers (sense and antisense) are highlighted in pink.

The AS-PCR primers (excluding random nucleotide tails) are indicated by the bold nucleotides.

The internal Snapshot primers (excluding random nucleotide tails) are underlined.

The mutations are indicated by the bold red font.

Exon 5

PCR amplicon size: 230 bp
The c.454C>T (R152X) mutation is indicated by the bold red font.

exon 5 sense primer

gacccatttccccttcaacaccctgttctttctcttattcatatgtagGAAGAGGAATA**C>T****GAATAAGGGA**

TATCTTGAAAGATGAAGAAACACTGACACTATTTCTCATTAAAAACATCGG

CCTGTCTGACTCAGTGGTCTACCTTCTGATCAACTCTCAAGTCCGTCCAGAGC

AGgtaggggatgtcactggccagtggtccctggagggagggaagcacccagcct

exon 5 antisense primer

Exon 6

PCR amplicon size: 294 bp
The c.768G>T (V256V) mutation is indicated by the bold red font.

exon 6 sense primer

ggtgtctttcctaccacaggcagtttctagtgttgcctcttctccctgcagTTCGCTCATGGAGTCCCGGAC

CTGGCGCTGAAGGACATCGCCTGCAGCGAGGCCCTCCTGGAGCGCTTCATCA

TCTTCAGCCAGAGACGCGGGGCAAAGACGGTGCGCTATGCCCTGTGCTCCCT

CTCCCAGGGCACCCTACAGTGGATAGAAGACACTCTGTATGCCAACGTGGA

CTTCTTCAAGCTCTTCCGTGT**G>T**gtaagggaggggtttggctgctcgccaattgcaaggtgattcct

exon 6 antisense primer

Exon 13

PCR amplicon size: 267 bp
The c.1885C>T (R602W) mutation is indicated by the bold red font.

exon 13 sense primer

agctatccaagcccgttccatcctttgtccctctgtgtcttctcagGTATTGGGATTCTGGTCCCAGAGCT

GATCCCGTGGAAGATTTC**C>T**GGTACATCTGGGGCGGGTTTGCCTATCTGCA

GGACATGGTTGAACAGGGGATCACAAGGAGCCAGGTGCAGGCGGAGGCTCC

AGTTGGAATCTACCTCCAGCAGATGCCCTACCCCTGCTTCGTGGACGATTCgt

gagtctgaagttcgcgatcctcctccatgacacgctaatgg

exon 13 antisense primer

Exon 17

PCR amplicon size: 232 bp
The c.2588G>C (G863A) mutation is indicated by the bold red font.

exon 17 sense primer

ctgcggtaaggtaggatagggagactccttcgactttctctgtttatttgtctctatttttag**G>C****AGACTATGGAAC**

CCCACTTCCTTGGTACTTTCTTCTACAAGAGTCGTATTGGCTTGGCGGTGAA

Ggtgagtcctttaaaacacaaatcttaatgtttgaaatcaactccttgggctctgtgcaagatgtatatggatcacagaggtggcc

ctctatgtaaacggtgtg

exon 17 antisense primer

Exon 30

PCR amplicon size: 286 bp
The c.4469G>A (C1490Y) mutation is indicated by the bold red font.

exon 30 sense primer

gtcagcaactttgaggctgattatggaatattttcttgtcttccatgagGGAGTACCCCTGTGGCAACTCAA

CACCCTGGAAGACTCCTTCTGTGTCCCCAAACATCACCCAGCTGTTCCAGAA

GCAGAAATGGACACAGGTCAACCCTTCACCATCCTGCAGGT**G>A**CAGCACC

AGGGAGAAGCTCACCATGCTGCCAGAGTGCCCCGAGGGTGCCGGGGGCCTC

CCGCCCCCCCAGgtacctgacctccaaacaacggggccccaggtctgcctgccacagaggga

exon 30 antisense primer

Exon 39

PCR amplicon size: 245 bp
The c.5461-10T>C (IVS38-10T>C) mutation is indicated by the bold red font.

exon 39 sense primer

gccccacctgctgaagagaggggggtggggtttgccccgtttccaacagtcctacttc**t>c**ctgtttcagACGCTGCT
CAGGTTCAACGCCGTGCTGAGGAAGCTGCTCATTGTCTTCCCCCACTTCTGC
CTGGGCCGGGGCCTCATTGACCTTGCACTGAGCCAGGCTGTGACAGATGTCT
ATGCCCGGTTTGgtgggtggtagccgaggcccatggagcatgggcctgggtccaaagctggga

exon 39 antisense primer

Exon 44

PCR amplicon size: 287 bp
The c.6079C>T (L2027F) mutation is indicated by the bold red font.

exon 44 sense primer

gaagcttctccagccctagctctatggtcatccctccactccttgaaggatactcagtaattgctttttttcttgcagTATTTTA
ACCAATATTTCTGAAGTCCATCAAAATATGGGCTACTGTCCTCAGTTTGATG
CAATTGATGAGCTG**C>T**TCACAGGACGAGAACATCTTTACCTTTATGCCCGG
CTTCGAGGTGTACCAGCAGAAGAAATCGAAAAGgtgaaaaatgttttgttgtggccacatagga
gtctggttaattacaagcctgtttcatgagagtgca

exon 44 antisense primer