Genetics of Age-related Macular Degeneration and Stargardt disease in South African populations

by

Johann Baard

BRDJOH015

MBChB, DA(SA)

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In fulfilment of the requirements for the degree

MSc(Med) in Human Genetics

University of Cape Town

October 2015

Supervisors: Miss Lisa Roberts, Professor Rajkumar Ramesar

Division of Human Genetics, University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Plagiarism declaration

I acknowledge that plagiarism is the act of using another researcher or person's work and claiming it as one's own. I thereby state that the work on which this dissertation is based is my personal original research (except where acknowledgements and referencing indicate otherwise) and that neither the whole dissertation or any part of it has been, is being, or will be submitted for another degree in any other university locally or abroad.

I have and will not allow anyone to copy my research and pass it on as their own. I empower the university to reproduce for the purposes of research either the whole or any portion of the contents of this dissertation in any manner whatsoever.

Signature: Signed by candidate

Date: 17 September 2015

Acknowledgements

I would like to express my thanks and gratitude to the following people who have enabled and facilitated me in making this research possible:

Firstly and foremost I would like to thank my bench and co-supervisor, Miss Lisa Roberts, for her patience, guidance and knowledge that she has shared with me during the last three years.

My supervisor, Professor Raj Ramesar, for his opportunity to study within this division and for the constant support and motivation throughout my masters.

Dr. Eric Albrecht, for his support and assistance in my recruitment effort within the Ophthalmology department in Grootte Schuur hospital.

And finally, the whole Division of Human Genetics, thank you for creating such a pleasant and supportive environment.

Table of Contents

Content **Page**

List of Figures

List of Tables

Abbreviations

°C	degrees Celsius
∆G	Gibbs free energy
µm	micrometer
µl	microliter
µM	micromolar
A	Adenine
A2E	N-retinyl-N-retinylidene ethanolamine
ABC	ATP-binding cassette
ABCA4	ATP-binding cassette, sub-family A, member 4
ACE	angiotensin-converting enzyme
aCGH	array-comparative genomic hybridization
AF	fundus auto fluorescence
AGE	advanced glycation end-product
AMD	Age-related macular degeneration
AOO	age of onset
APC	antigen presenting cell
ApoE	apolipoprotein E
ARMS2	age-related maculopathy susceptibility 2
ATR	all-*trans*-retinal
BM	Bruch's Membrane
bp	base pair
BP	branching point
BRB	blood-retinal barrier
BMI	body mass index
C	Cytosine
CFH	Complement Factor H
CN	choroidal neovascularization
CNV	copy number variants
CRD	Cone-rod Dystrophy
CV	consensus value
ddNTP	dideoxynucleotide triphosphate
dH_2O	distilled H_2O
dNTP	deoxynucleotide triphosphates
ECM	extracellular matrix
EDTA	Ethylenediaminetetraacetic acid

ELOVL4	Elongation of Very Long Chain Fatty Acids Protein 4
ESE	Exonic Splicing Enhancers
EtBr	Ethidium Bromide
G	Guanine
GA	geographic atrophy
GWAS	genome-wide association studies
HGMD	Human Gene Mutation Database
HGVS	Human Genome Variation Society
HMCN1	Hemicentin-1
HN-PCC	Hereditary nonpolyposis colon cancer
HSF	Human Splice Finder
HTRA1	high-temperature requirement factor A1
IDT	Integrated DNA Technologies
IL	interleukin
IPE	iris pigment epithelium
IS	inner segment
LD	linkage disequilibrium
mA	milliampere
MAF	minor allele frequency
MESA	Multi-Ethnic Study of Atherosclerosis
Mg^{2+}	Magnesium
$MgCl_2$	Magnesium Chloride
min	minutes
MLPA	multiplex ligation-dependent probe amplification
mM	millimolar
MMP	matrix metalloproteinases
MP	macular pigment
mRNA	micro-ribonucleic acid
NCBI	National Center of Biotechnology Information
ng	nanogram
nm	nanometers
NR-PE	N-retinylidene-PE
NTC	no template control
OCT	optical coherence tomography
OD	optical density
OS	outer segments
PCR	polymerase chain reaction

PE	phosphatidylethanolamine
PhD-SNP	Predictor of human Deleterious Single Nucleotide Polymorphisms
pmol	picomole
PON-P	Pathogenic-or-Not-Pipeline
POS	photoreceptor outer segments
PROM1	Prominin-1
qPCR	qualitative PCR
RDD	Retinal Degenerative Disorders
RDS	peripherin 2
RE	restriction endonucleases
RFLP	Restriction Fragment Length Polymorphism
ROS	reactive oxygen species
RP	retinitis pigmentosa
RPE	retinal pigment epithelium
rpm	revolutions per min
SA	South African
SAP	shrimp alkaline phosphatase
sec	seconds
SFD	Sorbys' fundus dystrophy
SIFT	Sorting Intolerant From Tolerant
SNAP	Screening For NonAcceptable Polymorphisms
SNP	single nucleotide polymorphisms
SR	Serine/Arginine-rich
STGD	Stargardt disease
T	Thymine
TGF-β	Transforming growth factor-beta
TIMP	tissue inhibitor of matrix metalloproteases
T_M	melting temperature
U	unit
UCT	University of Cape Town
USA	United States of America
UV	ultra-violet
V	volt
VEGF	vascular endothelial growth factor
VMD2	Vitelliform Macular Dystrophy 2
VOUS	Variant of unknown significance
W/V	Weight/volume

Abstract

Background: The Retinal Degenerative Diseases (RDD) Research Group in the Division of Human Genetics at UCT has for the past 25 years been intensively investigating a range of RDD phenotypes. Two points of particular note have emerged regarding Macular Degenerations (MD): (i) that more than 58% of juvenile MD, notably Stargardt Disease (STGD), in Caucasian populations may have the underlying causative genetic defect identified, while only 11% of the similar phenotype in indigenous African populations is resolved, and (ii) that the 'elderly' form of MD, i.e. age-related macular degeneration (AMD) has a remarkably lower incidence in the indigenous African population when compared to any other population group, and most notably the Caucasian (or European-derived) population/s. This study investigates the genetic factors underlying macular degeneration (MD) in our study cohort comprising various South African ethnolinguistic groups with particular focus on disease in juvenile and elderly indigenous Africans.

Materials and Methods: For the STGD part of the study, sequencing of the entire *ABCA4* coding and splice region (comprising 50 amplicons) was performed in three African STGD patients who were representative of three common haplotypes identified within the larger cohort of 36 patients. Pathogenicity predictive software, PON-P and Human Splice Finder (HSF), were used for *in silico* data analysis.
For the AMD subset: Available local indigenous southern African population-based genome-wide Single Nucleotide Polymorphism (SNP) chip (Affymetrix SNP6) data was used to identify SNPs within known AMD candidate genes in which allele frequencies were significantly different (i.e. 10 fold) between Caucasians and indigenous southern Africans. Nine SNPs occurring at higher frequencies within Africans compared to Caucasian controls were genotyped by SNaPshot PCR within a multi-ethnic AMD SA cohort. Minor allele frequencies (MAF) were compared using SHEsis.

Results: Sequencing of *ABCA4* in three African STGD patients produced 39 unique variants, out of which only one, (V643M), was deemed pathogenic. HSF predicted 22 of

these non-exonic variants to be 'possibly pathogenic', confounding analysis. No variants segregated with the common haplotypes.
Regarding the AMD cohort, eight SNPs in candidate AMD genes showed a decreased MAF in African AMD cases compared to controls, two of which (rs9621622 in *TIMP3* and rs17110714 in *ABCA4*), were statistically significant (p values of 9.95×10^{-4} and 1.04×10^{-2}, respectively).

Discussion and Conclusion: Although a number of variants were identified in the coding region of three haplotype-representative STGD subjects, only one variant proved pathogenic but did not co-segregate with the haplotype in the rest of the samples. It is possible that variants in regulatory regions not captured by the exonic screening might be involved, or that another gene may be implicated in the 'STGD-like' phenotype in the indigenous African subjects. In the second part of the study, the investigation of the African AMD cohort suggested that SNPs in *TIMP3* and *ABCA4* are associated with a decreased susceptibility, and may therefore plausibly be protective for AMD in indigenous Africans. Overall, however, this should be considered only a pilot study of macular degeneration in the indigenous African population, providing leads to larger scale studies of this group of disorders in this population group.

Chapter 1: Introduction

1.1 Anatomy of the Human Eye

The human eye, encased within the orbital cavity formed by the frontal, sphenoid, maxillary and zygomatic bones, comprises three distinct anatomical layers (Moore. 2013) (Figure 1.1). Structures within these three layers are discussed below.

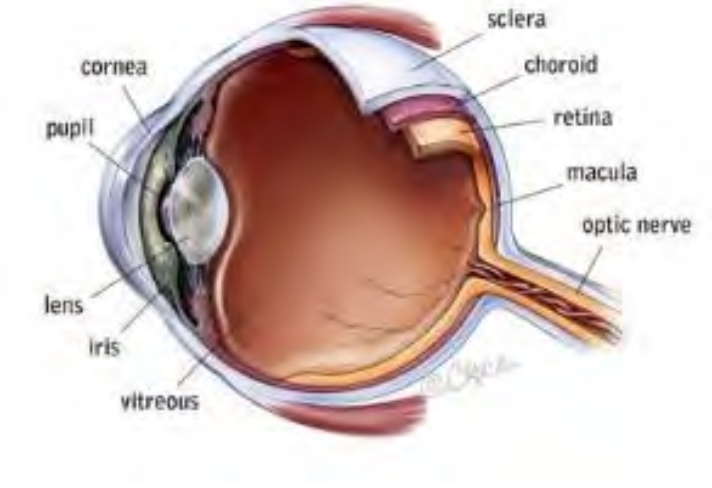


Figure 1.1: An illustration of a cross-sectional view of the human eye featuring important structures and the layers.
Reproduced from the website Bionicvision Australia:
http://bionicvision.org.au/eye/healthy_vision

1.1.1 The fibrous tunic

The outer layer, also termed the fibrous tunic, is formed posteriorly by the opaque sclera and anteriorly by the translucent cornea. The sclera serves as the collagenous, protective boundary providing insertion points for the extrinsic ocular muscles, enabling rotation of the eye. It adjoins anteriorly to the avascular dome-shaped cornea which acts as the eye's initial and main refractive structure, providing two-thirds of the eye's total

focusing power (http://webvision.med.utah.edu/book/part-i-foundations/gross-anatomy-of-the-ey/).

1.1.2 The uvea

The middle layer, also termed the vascular tunic or uvea, includes the choroid, ciliary body and iris.

1.1.2.1 Iris and ciliary body

Anteriorly, the iris is a diaphragm-like sphincter extending from the ciliary body that controls the aperture size of the pupil in order to regulate the amount of incoming light. The posterior stratum of the iris, the iris pigment epithelium (IPE), is coated in a high concentration of the pigment melanin and determines a person's eye colour. Darker coloured irises are due to additional melanin located in a fibro-vascular layer in the iris called the stroma.

Surrounding the iris is a muscular circular structure, the ciliary body that enables accommodation by shaping the lens anchored to it via suspensory ligaments. Epithelium lining the ciliary body produces aqueous humour which fills the anterior and posterior chambers, providing vital nutrients to the lens and cornea.

1.1.2.2 Choroid

The purpose of the extremely vascular, pigmented choroid is twofold (Mortazavi *et al*. 2014). Its melanin, produced by melanocytes, absorbs stray light that would otherwise reflect from the sclera, while its concentrated blood and lymphatic networks deliver nutrition to the retina. Histologically, the choroid is divided into four sections, namely the suprachoroid lamina, stroma, choriocapillaris and Bruch's Membrane (BM).The suprachoroid lamina is a fragile layer of pigmented connective tissue while the stroma mainly consists of medium- to large sized blood vessels. The choriocapillaris consists of

a solitary layer of fenestrated capillaries which constitute the majority of the choroid. The fenestrated endothelium of the choriocapillaris works in conjunction with BM and the retinal pigment epithelium (RPE) to form the outer part of the blood-retinal barrier (BRB) which functions to maintain a highly specialized milieu within the eye (Runkle *et al*. 2011).

1.1.2.2.1 Bruch's membrane

BM, the innermost layer of the choroid, is an acellular, pentalaminar structure primarily composed of extracellular matrix. As a transport medium, it functions to unite the RPE and choriocapillaris by exchanging biomolecules, nutrients and metabolic end-products between them. Structurally, it functions as a coil for the lens during accommodation. The BM suffers significant age-related alterations at a structural level causing reduced filtration-capability and loss of elasticity which contributes (pathologically) to macular and retinal disorders. These changes include collagen cross-linking, mineralization from calcium, iron and zinc deposits, advanced glycation end-products (AGEs) and lipid accumulation (http://www.mccormick.northwestern.edu/biomedical/markj/PDFofPapers/markj/Curcio2013.pdf).

1.1.3 The retinal tunic

The inner layer, or retinal tunic, consists of multi-layered neuronal tissue (Figure 1.2) containing photoreceptor cells that transduce the incoming electromagnetic light energy into electric nervous impulses. These action potentials are processed and carried along the ganglion cells of the optic nerve to the posterior pole of the occipital cortex where a visual representation of the surrounding environment is created.

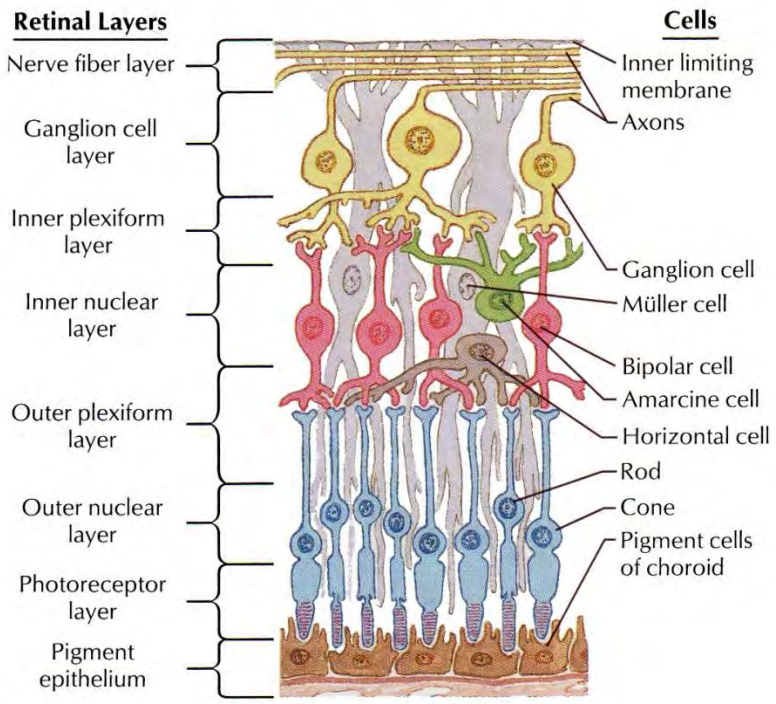


Figure 1.2: An illustration of the retina depicting the different layers and cell structures. Reproduced and adapted from the website MyHumanBody.ca: http://www.corpshumain.ca/en/Vue_en.php

The macula is an oval-shaped area, roughly 5.5 mm in diameter, and located 3.5 mm lateral to the optic disc, proximal to the midpoint of the retina. The darker discolouration of the macula on fundoscopy is due to its macular pigment (MP) concentration, particularly zeaxanthin. Centrally located within the macula is the fovea centralis, a small pit containing only closely packed cone cells specific for maximal visual acuity. In contrast to the rest of the retina, the macula's underlying choroid capillary bed and pigment epithelial layer is much thicker. The thicker pigment epithelial layer leads to a higher melanin content which is of clinical significance for macular disorders.

Morphologically, the retina sub-divides into ten distinct strata consisting of the RPE externally, and two synaptic or plexiform layers alternating between the photoreceptors, i.e. the inner nuclear layer and ganglion cells internally (http://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina/). The two plexiform layers, namely the inner- and outer-plexiform layers, serve to create synapses for communication both laterally and vertically between adjacent layers. They

also contain non-fenestrated capillary endothelial cells that comprise part of the inner BRB. The photoreceptor layer contains the two specialized neuronal cells termed rod and cone cells. Their cell bodies are situated in the outer nuclear layer just internal to the photoreceptor layer. The rest of the cells divide into the inner and outer segments, comprising the photoreceptor layer. The inner segment (IS) is packed with mitochondria and various organelles essential for cell metabolism. The outer segments (OS) contain stacked double membrane discs embedded with visual pigments to initiate phototransduction. The OS of cone cells are cone-like in structure and contain iodopsin as the visual pigment. In contrast, the OS of rod cells are elongated and tubular in structure with rhodopsin as the visual pigment. Rod cells vastly outnumber cone cells in the retina at a ratio of about 20:1 and are extremely sensitive to light, enabling scotopic or night vision. Cone cells are exclusively responsible for colour perception and are densely arranged within the fovea centralis, the central pit inside the macula, to permit central vision, at high resolution (Kawamura *et al*. 2008).

The inner nuclear layer contains amacrine, bipolar and horizontal cells which operate, both as a signal conduit between the photoreceptor and ganglion cells, and to improve the contrast sensitivity of the eye. The inner BRB is made up of the non-fenestrated retinal capillary endothelial cells within the plexiform layers of the retina.

Ganglion cells, collectively, transmit processed visual data as action potentials along long axons through the optic nerve to the brain. Moreover, the photosensitive capability of ganglion cells contribute to the pupillary light reflex and help regulate sleep-wake circadian rhythms.

1.1.3.1 The retinal pigment epithelium (RPE)

The basal layer of the retina, the RPE, is positioned between the photoreceptor retinal layer and the uvea. Its densely pigmented, monolayer of tightly packed hexagonal cells serves several functions, namely - light absorption, retinal storage and metabolism,

metabolic homeostasis, preservation of photoreceptor outer segment length and regulation of the immune privilege of the eye (Strauss *et al*. 2005).

1.1.3.1.1 Melanin

Light entering the eye is predominantly absorbed by pigments, namely, melanin and MP. Ocular melanin enhances visual acuity by: (i) absorbing scattered light reflecting from the sclera, (which would otherwise create false signals), and (ii) diminishing photo-oxidative stress. Melanin occurs in two different pigment cells; choroidal melanocytes and pigment epithelial cells, each with separate embryonic origins (Park *et al*. 2009). Pigment epithelial cells, in the RPE, contain mainly eumelanin which has a black-brown hue and has a comparable high density in the irises of all people. Conversely, choroidal melanin content differs significantly between different ethnic groups, with the highest concentrations occurring in Africans. Choroidal melanocytes, containing not only eumelanin but also pheomelanin which imparts a lighter reddish hue, maintain a constant melanin content throughout life. Unlike the choroid, melanin biosynthesis is absent in adult RPE cells, and the pigment decreases in this tissue with increasing age.

1.1.3.1.2 Pro-oxidative environment

Additionally, melanin functions as an anti-oxidant and free radical scavenger within the pro-oxidative milieu of the RPE. This stressful environment results firstly from the strong concentration of light energy focused through the lens onto the retina, and secondly, from the high venous oxygen saturation in the choroid capillaries. The high oxygen saturation results in both the choroid's high blood flow and low oxygen extraction which is a diminutive three percent when compared to the 50 % extraction of the retinal circulation (Linsenmeier *et al*. 2000). Beneficially, this high blood flow acts as a heat sink for all the heat generated from metabolic processes and light absorption, while simultaneously providing most of the retina's oxygen and glucose.

Another compounding element adding to the pro-oxidative environment is the unremitting load of free radicals entering the RPE from the phagocytosis of shed photoreceptor outer segments (POS). As mentioned previously, the POS are the external sections of the photoreceptor generated from stacks of double pleated membranes containing light sensitive visual pigments. These POS membranes are rich in polyunsaturated fatty acids making them prone to peroxidation. Old outer segments are pinched off and engulfed by the apical processes of the RPE on a diurnal basis. Approximately 15 % of the POS are phagocytized by the RPE, and replaced, on a daily basis (Hollyfield *et al*. 1977).

1.1.3.1.3 Macular pigments

The MP, lutein and zeaxanthin, are lipophilic carotenoids that function as anti-oxidants along with melanin. Besides quenching reactive oxygen species (ROS), they effectively protect the retina from damaging blue light due to their absorption spectrum, peaking at 460nm. MP are found mainly within the inner plexiform and outer ganglion layers within the retina, with the highest concentration occurring within the fovea inside the macula (Ahmed *et al*. 2005). It is this distribution of MP to which the macula owes its yellow colouration. Since humans are unable to synthesize carotenoids, the concentration of the macular pigments is entirely diet-dependent.

1.1.3.1.4 Visual cycle

Another vital recycling process within the RPE is the retinal visual cycle. The aim of this cycle is the circulation and metabolism of Vitamin A derivatives between the RPE and POS. 11-*cis*-retinal, a photoreactive chromophore, is a cofactor to the transmembrane protein, opsin, together forming the visual pigment rhodopsin, which is loosely packed within the central region of the POS (Buzhynskyy *et al*. 2011). Light induces isomerization of 11-*cis*-retinal to all-*trans*-retinal (ATR) which prompts a conformational transformation in opsin. This leads to G-protein coupled second messenger activation and subsequent neurotransmission (http://webvision.med.utah.edu/book/part-v-

phototransduction-in-rods-and-cones/phototransduction-in-rods-and-cones/). Via circular enzymatic pathways, ATR is re-isomerized to 11-*cis*-retinal within the RPE and returned to the POS via Retinal-binding protein 3, to replace the consumed chromophore.

1.1.3.1.5 *ABCA4* gene and protein

Integral to the retinal visual cycle is an outer segment rim protein, termed ATP-binding cassette, sub-family A, member 4 (ABCA4), which facilitates removal of ATR from POS following photobleaching. ABCA4, encoded by the *ABCA4* gene on chromosome 1p22, is a member in a superfamily of ATP-binding cassette (ABC) transporters localized within disk margins of both rod and cone outer segments. As one of the biggest protein classes, ABC transporters hydrolyze ATP to transport various substrates across the cell membrane. Once activated, ABCA4 transfers its substrates, N-retinylidene-PE (NR-PE) and phosphatidylethanolamine (PE), from the POS lumen to its cytoplasmic side. PE is a class of phospholipids found in all human cells. NR-PE then dissociates into PE and all-*trans* retinaldehyde which is subsequently reduced to ATR by retinol dehydrogenase. ATR is then transferred to the RPE where it is converted to 11-*cis*-retinal by retinoid isomerase. Failure to remove NR-PE results in formation of a fluorophore salt, N-retinyl-N-retinylidene ethanolamine (A2E), a toxic derivative of the visual cycle and primary component of lipofuscin. As a fluorophore, A2E sensitizes the retina to more blue light damage and its ionic properties result in further destabilization of cellular membranes (Shaban *et al*. 2002).

1.1.3.1.6 Lipofuscin

Lipofuscin refers to the autofluorescent, intra-cellular material that accumulates within the RPE, and is composed of chemically altered deposits of partially degraded POS. RPE cells are post-mitotic and therefore unable to dilute the intracellular lipofuscin, which accumulates with increasing age. Accumulation is aggravated by the fact that lipofuscin aggregates have proven resistant to cellular enzymatic breakdown. The

highest concentrations of lipofuscin occur within the macular RPE, with the exception of the fovea. Surprisingly, an exponential rate of lipofuscin build-up occurs in the first two decades of life. Thereafter, the increase in RPE lipofuscin correlates linearly with increasing age. Lipofuscin build up leads to RPE cellular dysfunction through cellular engorgement and leaching of toxic oxidized molecules into the cellular cytoplasm (Kennedy *et al*. 1995).

1.1.3.1.7 Immune regulation

Finally, the RPE contributes to the immune privilege of the retina by secreting several critical factors and by its tight junctions between neighbouring cells, contributing to the outer component of the BRB. The RPE contributes to the innate immunity of the retina by expression of Toll-like receptors which are conserved immune response molecules, able to sense pathogenic microbial molecules (Kumar *et al*. 2004). As a first line of defense against microbial invasion, RPE cells rapidly secrete pro-inflammatory cytokines such as interleukin (IL)-6 and IL-8.

When required, the RPE can also down regulate the immune response by a variety of immune suppressive factors such as IL-11 and transforming growth factor-beta (TGF-β). TGF-β is a multifaceted cytokine involved in cell growth, immune response and extracellular matrix (ECM) synthesis. When activated, TGF-β has been shown to induce RPE cells to release factors like pigment epithelium-derived factor and vascular endothelial growth factor (VEGF). Pigment epithelium-derived factor acts as an antiangiogenic and neuroprotective factor, inhibiting pathological vascularization in apoptotic conditions. VEGF is active in angiogenesis and chemotactic signalling recruiting both granulocytes and macrophages. Part of the angiogenic function of VEGF is to increase matrix metalloprotease activity. Conversely, the RPE also secretes different types of tissue inhibitors of matrix metalloproteases (TIMP), which are thought to be critical for preservation of BM. This ability to secrete antagonistic factors, when needed, illustrates how essential the RPE is for a homeostatic environment in the retina.

Lastly, the RPE is also involved in the adaptive immune response by functioning as an antigen presenting cell (APC) (Holtkamp *et al*. 2001).

1.2 Stargardt Disease

Stargardt disease (STGD), of which there are three types, is one of the most frequent causes of juvenile macular dystrophy and was first described in 1909 by a German ophthalmologist, Karl Stargardt (Stargardt *et al*. 1909). STGD typically presents within the first two decades of life with an incidence of approximately 1:10 000, with a particularly high carrier frequency in the general population of about two percent (Blacharski *et al*. 1988). Patients initially present with a severe debilitating central vision loss while maintaining their peripheral vision. There is also an associated loss of colour vision and reduced dark adaptation.

STGD type 1, which is the predominant type in populations studied to date, is inherited in an autosomal recessive pattern and mostly associated with mutations in *ABCA4.* STGD type 3, conversely, is inherited in an autosomal dominant fashion and caused by mutations in *ELOVL4* which encodes the 'Elongation of Very Long Chain Fatty Acids Protein 4'. STGD type 4, which is inherited as an autosomal dominant condition, occurs as a result of mutations in the *Prominin-1* (*PROM1*) gene. STGD type 2 is no longer used within STGD classification since the original mapping of the STGD phenotype in a family, to chromosome 13q34, was discovered to be incorrect.

1.2.1 Pathogenesis

As described above, a defective ABCA4 protein leads to the accumulation of A2E and lipofuscin in RPE cells of the macula. This accumulation leads to lysosomal dysfunction and photo toxicity in the RPE. This results in RPE cell death which secondarily, triggers photoreceptor damage and ultimate vision loss (Cideciyan *et al*. 2004). Clinically, STGD is observed as yellow-orange flecks around the macula , with the characteristic "beaten-bronze" appearance upon fundoscopy. A definitive diagnosis is made with a fluorescein

angiogram which shows a dark choroid with hyper fluorescence from the retinal flecks. The dark choroid is caused by the blocking of the choroidal fluorescence from the excessive RPE lipofuscin buildup. The causal gene for type 1 STGD, *ABCA4,* was identified in 1997 (Allikmets *et al*. 1997a). Surprisingly this gene was found to be expressed mainly within rod photoreceptors.

More than 800 mutations have been found within *ABCA4* since its discovery (Allikmets *et al*. 2007), although only 10 % of STGD patients are found to have frequent disease-associated *ABCA4* variants, and more than half these mutations have been reported once only (Zernant *et al*. 2011). Although most are missense mutations, a wide spectrum of genetic variants have been reported, ranging from insertion/deletions to uniparental disomies (Fingert *et al*. 2006). Additionally, *ABCA4* is known to exhibit high allelic diversity of its coding region; this is illustrated through its nucleotide diversity being nine to 400 times higher than other macular disease genes which were similarly studied (Webster *et al*. 2001). Screening of *ABCA4* has historically been inefficient with only one mutation found in 15 – 20 % of patients and no mutation in 15 % of patients (Zernant *et al*. 2011). Similar proportions were found when *ABCA4* was screened using next generation sequencing and array-comparative genomic hybridization (aCGH) in an African American STGD cohort (Zernant *et al* 2014). Mutations in *ABCA4* have also been shown to cause various other visual disorders including, cone-rod dystrophy (Cremers *et al*. 1998), retinitis pigmentosa (RP) (Cremers *et al*. 1998) and age-related macular degeneration (AMD). The first part of this study will focus on STGD, the juvenile form of macular degeneration, which is associated with mutations in *ABCA4*.

ABCA4 is a large gene comprising of 50 exons, and mutation screening using conventional Sanger sequencing is arduous. A commercial microarray, the ABCR400 chip, has been designed by Asper Ophthalmics in Estonia (http://www.asperbio.com/asper-ophthalmics) to simultaneously screen for all previously reported *ABCA4* mutations/variations (n>600). The microarray chip, which is frequently updated, efficiently identifies 65 – 75 % of all disease causing *ABCA4* alleles within each cohort screened. To date, on average, 40 % of a STGD cohort will have both

mutations identified (i.e. bi-allelic), with another 40 % having only one mutation and the remaining 20%, no mutations identified, when using the ABCR400 chip (Ernest *et al*. 2009; Klevering *et al*. 2004; Allikmets 2007). This array has proven successful for screening Caucasian South African (SA) patients with STGD, and seven common *ABCA4* mutations were found to occur frequently amongst the South African Afrikaner community (Roberts *et al*. 2012). The Afrikaner community is a Caucasian population that originated from European immigrants more than 300 years ago (September *et al*. 2004). These endogamous immigrants stem mainly from Holland, Germany and France and have been associated with a number of heritable genetic disorders other than STGD (Gevers at al. 1987).

In contrast to what has been found in the Caucasian population, the microarray screening has proven less successful for indigenous African SA patients with STGD. This is likely due to them having novel mutations, which are not tested for by the array.

1.3 Age-related Macular Degeneration (AMD)

AMD is a multi-factorial, neuro-degenerative disease of the macula in the elderly, set to become a major economic health burden due to the increase in global population and life expectancy rates. Symptomatically, it presents with a gradual loss of central vision which is essential for everyday activities such as reading and driving.

1.3.1 Epidemiology

Globally, AMD is the fourth largest cause of blindness following cataracts, glaucoma and blindness of undetermined causes, and accounts for five percent of the global blind population (Jonas *et al*. 2014). Within the global elderly population (over 50 years of age) in the developed world, AMD is the leading cause of irreversible visual impairment. Epidemiological studies on AMD worldwide have mainly focused on Caucasian and Asian populations, with studies on African (-derived) populations occurring outside of the African continent.

Although, anecdotally, it is assumed to be a rarity among indigenous southern Africans, AMD has been reported to be a significant cause of visual impairment in studies performed in populations in Kenya, Tanzania, Cameroon (Oye *et al*. 2007), Guinea, Zanzibar and Rwanda (Mathenge *et al*. 2007). Accurate comparison between studies is limited due to the use of different grading systems and the lack of retinal digital photography for all individuals. A recent cross-sectional study conducted on a Kenyan population, older than 50 years of age, using a modified version of the International Classification and Grading System for Age-Related Maculopathy and retinal photographs for those diagnosed, found the incidence of early and late AMD to be 11.2 and 1.2 %, respectively (Mathenge *et al*. 2013). Multiple population-based studies have reported incidence differences in AMD between ethnic groups. Despite limitations and differing diagnostic criteria in these studies, they all confirm AMD to be less prevalent in Africans with incidence rates ranging between 0.7 and 2.7 % compared to Caucasian incidence rates of between 2.1 and 10.9 % (Klein *et al*. 2006; Bressler *et al*. 2008; Leske *et al*. 2006; Erke *et al*. 2012). The difference in prevalence is also observed when Africans and Caucasians are studied from within the same geographical region (Friedman *et al*. 1999). This partly removes the environment as a confounding factor, suggesting that genetic protective factors may play a role in Africans, or that Caucasians are more genetically susceptible.

Further studies also highlight the geographical effect on the frequency of early and late AMD in the same ethnic group (Klein *et al*. 2004). Early AMD is more prevalent in Caucasian populations from Norway and Iceland while late AMD is reported to occur at a higher incidence in populations in the USA, Australia and Holland (Klein *et al*. 2004) . Paradoxically, it has been reported that although early signs of AMD are similar among African and Caucasian populations, the incidence of early and late AMD is higher in Caucasians (Klein *et al*. 2004).

1.3.1.1 Ethnic anatomical differences

Anatomical differences within the macula between ethnic groups have been hypothesized to influence susceptibility to AMD. These differences include: Iris colour, melanin content, lipofuscin and macular-pigment concentration. Iris colour, owing to pigment density in its anterior layer, determines the initial amount of light absorbed and therefore protects the retina from excessive light exposure and oxidative damage. This is evident in individuals with blue- or hazel-coloured irises, more common in Caucasians, who display significantly higher susceptibility to AMD (Nicolas *et al*. 2003).

Melanin concentration in the RPE, which peaks within the macula, is similar among Caucasian and African populations and decreases with age. Choroidal melanin content is significantly greater in Africans compared to Caucasians (Hu *et al*. 2008).

RPE-lipofuscin aggregation increases with age in all ethnic groups but considerably more so in Caucasians. This aggregation leads to thickening of the already thicker nerve fiber layer within the macula in Caucasians.

The MP, lutein and zeaxanthin, play important roles in filtering blue light and quenching free radicals. Racial differences in MP distribution were investigated among healthy subjects and a definite increased density in central macular pigment was observed in Africans compared to other groups (Wolf-Schnurrbusch *et al*. 2007).

1.3.2 Risk factors

AMD develops due to a complex interaction between environmental, genetic and aging factors. Aging is the greatest risk factor for developing AMD. Various age-related changes contribute to AMD, including loss of rod photoreceptors and RPE atrophy due to lipofuscin build up (Brink *et al*. 2002). Among modifiable factors, smoking plays the most significant role in AMD development (Smith *et al*. 2001). Smoking contributes to oxidative damage at the level of the RPE by freeing large quantities of ROS and other

toxic substances. This leads to accelerated atherosclerotic plaque formation and endothelial dysregulation within the RPE. Other modifiable factors include dietary supplementation which has been shown to alter AMD progression, to a certain extent. These dietary supplements include vitamins B6, B12, C and E, zinc and the macular pigments, lutein and xeazanthin (Age-Related Eye Disease Study Research Group *et al*. 2001). Clinical factors that increase susceptibility include hypertension, elevated body mass index (BMI) and hypercholesterolemia (Hogg *et al*. 2008).

1.3.3 Pathogenesis

Drusen, the hallmark of AMD, are extracellular aggregates primarily composed of lipid peroxidation products, AGE, protein crosslinks and various complement factors formed by the constant oxidative stress in the macula (Crabb *et al* 2002). Drusen accumulates with age, between the basement membrane of the RPE and BM, and is ubiquitous among individuals over 40 years of age. Clinically, drusen may be sub classified into hard and soft drusen. Hard drusen can be seen as small, hard sub-RPE yellowish deposits under 63 micrometers (μm) in diameter with well-defined borders. Small pockets of hard drusen deposits in the periphery can be seen as part of the normal aging process and are not considered pathological. Soft drusen are wider than 63 μm with indistinct margins, tend to coalesce to become confluent and represent localized detachment of the RPE.

As mentioned previously, the comparability of AMD studies is limited due to the different classification systems that have been used, ranging from the Rotterdam Eye Study Staging System (Klaver *et al*. 2001) to the International Classification and Grading System by the International Age-related Maculopathy Epidemiological Study Group (Bird *et al*. 1995). Although clinically categorized into four groups, the disease is generally divided into early or late AMD. Early, also termed dry, AMD is diagnosed with the presence of drusen and RPE abnormalities, including geographic atrophy (GA). Abnormalities of the RPE include hyper or hypopigmentation which signify mobilization of the pigment and loss of RPE cells. GA is defined as an oval area of

hypopigmentation exceeding 175 μm where choroidal vessels are more visible in surrounding areas. GA is characterized by an area of RPE cell death with overlying neural retina atrophy and leads to gradual visual loss.

Twenty percent of individuals with GA progress to late, also termed wet, AMD which is characterized by choroidal neovascularization (CN). CN, mediated by VEGF, is a process of vasculogenesis into the RPE and sub-retinal space through openings in BM. These abnormal vessels leak fluid and haemorrhage, leading to macular scar formation and sudden onset of vision loss (Bressler *et al*. 2009). Wet AMD is defined as the presence of any of the following: Serous (or haemorrhagic) sensory retinal detachment; neo-vascular membrane formation; pigment epithelial haemorrhages, or sub-retinal fibrin-like deposits

1.3.4 Genetics of AMD

AMD has been extensively studied for associated genetic and epigenetic factors. In contrast to other complex diseases, half of the genetic heritability of AMD can be accounted for by a few genetic loci (Swaroop *et al*. 2007). This is further illustrated through twin and familial studies, where risk of AMD in first-degree siblings of AMD individuals is double that for siblings of unaffected individuals (Swaroop *et al*. 2007). Similarly, concordance among mono-zygotic twins further supports an increased risk associated with a family history of AMD. Historically, identification of susceptibility loci relied on linkage studies in smaller families, and consistently showed positive signals on chromosomes 10q26 and 1q25-31, suggesting these as loci for susceptibility factors (Swaroop *et al*. 2007). Detection of AMD-associated variants near the *Complement Factor H* (*CFH*) locus on chromosome 1q31 led to the discovery of multiple associations with other complement genes such as *C2, C3, CFB* and *CFI* (Anderson *et al*. 2010).

The association on chromosome 10q26, has led to the discovery of three potential susceptibility genes, namely: *LOC387715/ARMS2* (age-related maculopathy susceptibility 2), *HTRA1/PRSS11* (high-temperature requirement factor A1) and

PLEKHA1. The *ARMS2* gene produces a protein of (as yet) unknown function, and which is present in the retina and placenta (Andreoli *et al*. 2009). Recently, the most significant discoveries have emerged from genome-wide association studies (GWAS). A meta-analysis of AMD-GWAS performed through collaboration of 18 centers revealed 19 different AMD susceptibility loci which fit into previously known pathways, and seven novel loci containing novel genes (Fritsche *et al*. 2013). Pathways associated with the identified susceptibility loci have been consistent with the proposed pathogenesis of AMD. These pathways consist of the complement system, oxidative stress, extra-cellular matrix, lipid regulation, and angiogenesis-signalling.

1.3.4.1 Protective alleles

After years of genetic research since the first gene, *CFH*, was associated with AMD, the full genetic landscape of AMD still remains unresolved. The susceptibility to AMD consists of an interaction between both risk and protective alleles, along with non-genetic factors such as the environment and increasing age. Some candidate genes, like *CFH* and *apolipoprotein E* (*ApoE)*, have even been shown to harbor both risk and protective alleles (Baird *et al*. 2004). Largely, studies have focused on identifying protective alleles in a specific ethnic group. Within Caucasian populations, a decreased frequency of the *ApoE* ε4 allele within the AMD group compared with a control cohort provides support for the hypothesis of *ApoE* ε4 as a protective gene/allele for AMD. Another investigation within Caucasian AMD family-based and control data sets found SNPs in *CFB* and *CC2* conferring a decreased susceptibility to AMD (Spencer *et al*. 2007). Occasionally, a protective factor is shown to act across multiple ethnic groups like the protective *CFH* haplotype that is associated with a *delCFHR1/CFHR3* mutation (Hageman *et al*. 2006).

Identifying a protective allele provides insight into possible gene and pathway associations that ultimately lead to a better understanding of the protective role of the allele in AMD within the ethnic group concerned. This is illustrated with the first demonstration of an Alu insertion element in the angiotensin-converting enzyme (*ACE*)

gene, providing protective effects for dry AMD (Hamdi *et al*. 2006). Although the correlation between blood levels of ACE and the pathophysiology of AMD is unknown, these findings help stimulate future research concerning AMD pathways (Hamdi *et al*. 2006). Similarly, the Q192R polymorphism in *PON1* showed decreased susceptibility for AMD, although the mechanism is still unknown (Pauer *et al*. 2010). In 2010, a GWAS found two intronic SNPs, rs429608 in *SKIV2L* and rs2679798 in *MYRIP,* as protective variants for AMD. The function of the protein that *SKIV2L* encodes for is yet to be confirmed and it shares its cytogenetic location on chromosome six with other complement genes. Furthermore, MYRIP is involved in RPE melanosome trafficking and maps to chromosome 3p21 (Kopplin *et al*. 2010).

Considering the large genetic heritability of AMD and the significant difference in prevalence of AMD between Africans and Caucasians, it is reasonable to hypothesize that genetic protective factors exist within the African population. As discussed earlier, the anatomic and environmental differences between these ethnic groups have been investigated to explain the discrepancy in prevalence, but none have provided sufficient evidence. Other clinical factors such as diabetes, BMI and inflammatory markers were analyzed among AMD individuals from different ethnic groups in the Multi-Ethnic Study of Atherosclerosis (MESA) cohort and also failed to explain these differences in prevalence (Klein *et al*. 2008). These results substantiate the need to explain these differences genetically. In a multi-ethnic study, Klein *et al*. (2008) tried unsuccessfully to use the *CFH* Y402H CC genotype to explain the differences in AMD prevalence between Caucasians and Africans. This adds to the solitary study done in Africa seeking an association between the Y402H variant and dry AMD, which was based on a limited cohort and failed to show a significant correlation (Ziskind *et al*. 2008).

No previous studies have been performed on the various ethnic African groups within SA for possible genetic protective factors against AMD. Firstly, African control genomic data from a local database were investigated to find SNP loci within AMD candidate genes with significantly higher frequencies when compared to a Caucasian control group. It is hypothesised that of these SNPs, any showing a decreased allele frequency

within an African AMD cohort when compared to the African control group are likely to be associated with a protective factor, either in the gene it resides in, or a genomic locus with a significant linkage disequilibrium (LD).

1.4 Aims and objectives

This pilot project aims to investigate the two macular degeneration disorders described, namely STGD and AMD, in the previously understudied indigenous SA population. The two disorders will be interrogated in two separate parts of the project.

The first part of this study aims to screen an African STGD cohort for novel mutations in the STGD-causal gene, *ABCA4*. The objective is to sequence the entire coding region of the *ABCA4* gene, comprising 50 exons, in three STGD individuals representing the three common haplotypes, in which mutations have not hitherto been identified. This will be done in order to identify pathogenic mutations which may account for a significant proportion of STGD in this population group.

The second part of this study aims to search for protective genetic factors that may explain the reduced susceptibility to AMD observed within the indigenous African populations in SA. The objective is to genotype variants within known AMD candidate genes that are more frequent in African populations compared to Caucasian populations. Any of these variants then found to exhibit a reduced frequency within an African AMD cohort, could represent candidate protection factors. Identification of genetic protection factors, and the functional roles they play, could ultimately have implications in the treatment of AMD.

Chapter 2: Materials and Methods

2.1 Patient recruitment

2.1.1 UCT Retinal Registry Database investigation

The Retinal Degenerative Disorders (RDD) database currently comprises phenotypic and demographic details of subjects with RDDs, as well as details of biological material available, and molecular genetic investigations carried out to date. A specific emphasis was on extracting information on individuals and families admitted to the study with either STGD or AMD phenotypes. Furthermore, an attempt was made to obtain data on the methodologies used in investigations to date, and to determine whether there was any bias in the ethnicity of individuals in whom disease-causing mutations had been detected. Following this, specific samples/subjects in which no disease-causing mutations had been detected to date were selected for further molecular investigations.

These subjects had already been recruited under the broader RDD project with informed consent obtained according to the tenets of the Declaration of Helsinki (2008). Ethics approval for this study was obtained from the University of Cape Town (UCT) Research Ethics Committee (UCT HREC REF: 432/2013) linked under the umbrella RDD project (UCT HREC REF 226/2010). All individual samples chosen were already stored within the Retinal Registry in the Division of Human Genetics at UCT, along with clinical information including ethnicity, date of birth, age of onset (AOO) and ophthalmological diagnosis. The three individual STGD patients were deduced as representative of the common haplotype groups will be referred to as the test samples hereafter. The remaining African STGD samples known to carry one or more of the common haplotypes will be referred to as the haplotype samples.

2.1.2 PHASE Haplotype analysis for STGD sample selection

Interrogation of the RDD database showed the ABCR400 microarray chip to be relatively unsuccessful in identifying mutations within the African STGD cohort (i.e. only 11% of those screened had both disease-causing alleles identified – compared to 58% of Caucasian STGD patients.). It was therefore decided to investigate the part of the indigenous southern African cohort without full molecular diagnosis for possible common *ABCA4* haplotypes. The decision to use haplotype analysis and not full cohort exome sequencing was in part due to resource constraints within this research setting. A haplotype refers to a combination of genotypes/alleles that associate statistically and are inherited *en bloc* or together. Novel *ABCA4* mutations, if present, could be expected to track with common haplotypes within the African STGD cohort. Therefore, screening of the individuals containing the common haplotypes presents a more cost-effective manner in which to elucidate these possible novel mutations.

Haplotype analysis was performed on the 32 unrelated patients within the African STGD cohort who had only one or no *ABCA4* mutations identified. Firstly, all genotype data of the 32 patients obtained from the ABCR400 microarray chip screening done previously (unpublished data) by Asper Ophthalmics (http://asperbio.com/asper-ophthalmics) was collected. Thereafter, all variants for each of the 32 patients occurring within *ABCA4* were extracted. These variants were further reduced to only those variants occurring in all of the 32 patients as different versions of the ABCR400 microarray had been used for the cohort in past years. This equated to a total of 17 SNPs shared among all 32 patients within *ABCA4* (Appendix 7).

Information containing number of individuals, number of SNPs, chromosomal location of each of the 17 variants in ascending order and genotypes were organised onto an Excel Spreadsheet (Microsoft Office Home and Student 2010) (Redmond, WA, USA) according to the basic format within the PHASE (Stephens *et al*. 2001) guide. Once formatted, the sheet was changed to an input file (.inp) before it was run through the PHASE (Stephens *et al*. 2001) program using the command prompt, after which the

output file was produced within the same directory. This work, although presented here for a fuller understanding of what was done, was carried out by another member of the RDD research team, Ms Lisa Roberts, prior to the current work being carried out.

2.1.3 AMD sample selection

In order to increase the AMD cohort size within the Retinal Registry, additional patients were recruited in collaboration with ophthalmologists in the Department of Ophthalmology at Groote Schuur Hospital in Cape Town. Ethics approval was granted by the UCT Research Ethics Committee (UCT HREC REF: 432/2013). A Confirmation of Diagnosis form (Appendix 6) was designed according to the International classification and grading system of AMD along with inclusion and exclusion criteria. Each consenting patient meeting the inclusion criteria underwent a 30° fundus photograph and a basic ocular examination, including a careful 78D fundal assessment by a registered ophthalmologist. After completion of the RDD Molecular Request form (Appendix 6) containing demographic, family and diagnostic information, 8ml (2 x 4ml) venous blood was collected in Ethylenediaminetetraacetic acid (EDTA) tubes. These samples were subsequently de-identified and added to the Retinal Registry within the Division of Human Genetics Database.

2.1.4 Selection of SNPs in candidate AMD genes

The aim of the second part of this study was to identify the possible protective genes associated with a decreased susceptibility for AMD in the African population in SA. In order to do this, data on allele frequencies was accessed from a study in which high density genome-wide SNPs had been assayed on five indigenous Southern African populations, including Zulu, Xhosa, Sotho/Tswana, Herero and San (Chimusa *et al*. 2015). Similar data in the public domain, from Caucasian populations (of which South African Caucasians were presumed to be a subset) were then compared for SNPs that displayed significantly high allele frequency differences in genes previously associated with AMD. An African AMD cohort was then genotyped for these selected SNPs and

those showing significantly decreased allele frequencies compared with the African control population were hypothesised to associate with a protective genetic factor. If multiple SNPs associated with the same gene it could be theorized that this gene plays an important protective role in the indigenous African population in SA.

Firstly, an up–to-date list of 24 genes associated with AMD (Appendix 4) was extracted from 'Genetics Home Reference' (http://ghr.nlm.nih.gov/condition/age-related-macular-degeneration/show/Related+Gene(s)) (Updated: June 2011). The variants in these genes were extracted from data previously generated by our laboratory (Chimusa *et al*. 2015). This data comprises indigenous SA African control groups genotyped with the Affymetrix (Santa Clara, CA, USA, 1992) Genome-Wide Human SNP Array 6.0. Effectively the database contains allelic frequencies for various ethnic groups namely: Zulu (n=17), Xhosa (n=28), Sotho (n=25), Herero (n=25) and Khoisan (n=24).

This local SNP 6 database contains more than 906 600 SNPs across the genome obtained from the Genome wide Human SNP array 6.0 by Affymetrix (Santa Clara, CA, USA, 1992). Data from the database was mined by using R (R core team, Vienna, Austria, 2013) (http://www.R-project.org/). The Caucasian control group was obtained from 1000 Genomes Browser (McVean *et al* 2012) (http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/) (Build 14, Oct 2013) using the BioMart data mining tool (Kinsella *et al* 2011) in Ensembl (Flicek *et al* 2014) (http://www.ensembl.org/biomart/martview/0619df929ccd736db0908ca9d26d23e9). The group used to represent the Caucasian control group included Utah residents with Northern and Western European ancestry (n=99). SNP data used for this specific study included all SNPs located in the selected AMD genes that were also present in both the 1000 Genomes (McVean *et al* 2012) Caucasian population and the local SNP 6 database. From these approximately 1600 variants, all those SNPs with a ten-fold difference in the African control group compared to the Caucasian control group were extracted.

This totalled 69 different variants located in 12 of the 24 selected AMD genes. The 69 variants were then further reduced by selecting only those SNPs that occurred in all the different ethnic groups within the SA African control group. From these, 10 SNPs were chosen from three different genes involved in proposed pathways that would support a plausible protective genetic role in the indigenous southern African population.

2.2 DNA extraction and determination of integrity

All samples recruited for the AMD cohort had DNA isolated by assigned researchers within the Division of Human Genetics at UCT using a modified salting out technique (Miller *et al*. 1988). Over time DNA samples fragment into smaller pieces due to freeze-thaw cycles and nuclease contamination, therefore it was necessary to determine the quality of all samples before further genetic analysis.

2.2.1 Spectrophotometry

Spectrophotometry measures DNA concentration and purity from contaminants such as protein and inorganic solvents. DNA concentration is calculated by optical density (OD) based on the absorbance of ultra-violet (UV) light by the sample at 260 nanometers (nm). The OD of the sample is converted to ng/µl using the Beer Lambert Law which states that absorbance is proportional to concentration at a specific wavelength. As proteins absorb UV light at a wavelength of 280 nm, the ratio of absorbance at 260 nm and 280 nm was used to exclude protein contamination. Values between 1.7 and 2.0 indicate minimal protein contamination. A second measure of purity is the ratio of absorbance at 230 nm and 260 nm. Values under 1.8 indicate the presence of contaminants such as EDTA and phenol.

The NanoDrop ND-1000 Spectrophotometer (Nanodrop Technologies Inc., Wilmington, United States of America [USA]) was used to perform spectrophotometry on all samples. Each sample measurement was repeated to obtain an average concentration,

used subsequently for dilutions to 100 nanogram (ng)/microliter (µl)) using SABAX distilled H_2O (dH_2O) (Adcock Ingram, Johannesburg, SA).

2.2.2 Agarose gel electrophoresis

All samples were subjected to agarose gel electrophoresis to determine DNA integrity. Gel electrophoresis is a separation method used to determine DNA molecular size. When voltage is applied across a gel, DNA migrates to the anode due to its negatively charged phosphate backbone. The rate of migration is directly proportional to size; hence smaller DNA fragments migrate faster. Smeared patterns indicate fragmentation while a distinct, high-molecular weight band points to an intact specimen. For accurate resolution, a two percent weight/volume (w/v) agarose (SeaKem® LE, Lonza, Switzerland) (Appendix 1) gel was prepared and treated with Ethidium Bromide (EtBr)(Sigma, ST Louis, MO, USA)(1 ng/µl). Wells were loaded with 200ng of the DNA sample and three µl loading dye (Appendix 1) to add density and facilitate observation. Adjacently, GeneRuler™ 100 base pair (bp) Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA) (Appendix 2) was loaded to provide a set of size standards for comparison.

Once loaded, the gel was electrophoresed within TBE (891.8 millimolar (mM) Tris, MW 121.34; 889.4mM Boric Acid; 19.9mM EDTA, pH8) (Appendix 1) at 120 volt (V) and 90 milliampere (mA) for 25 minutes (min). DNA visualization and image capturing was performed using the Uvipro Gold transilluminator (UVITec, Cambridge, UK) and UVIpro software v12.3 for Windows(Copyright 1999-2005). All subsequent agarose gel electrophoresis were carried out as per these described conditions, unless otherwise stated, and are referred to as the Standard Electrophoresis Conditions.

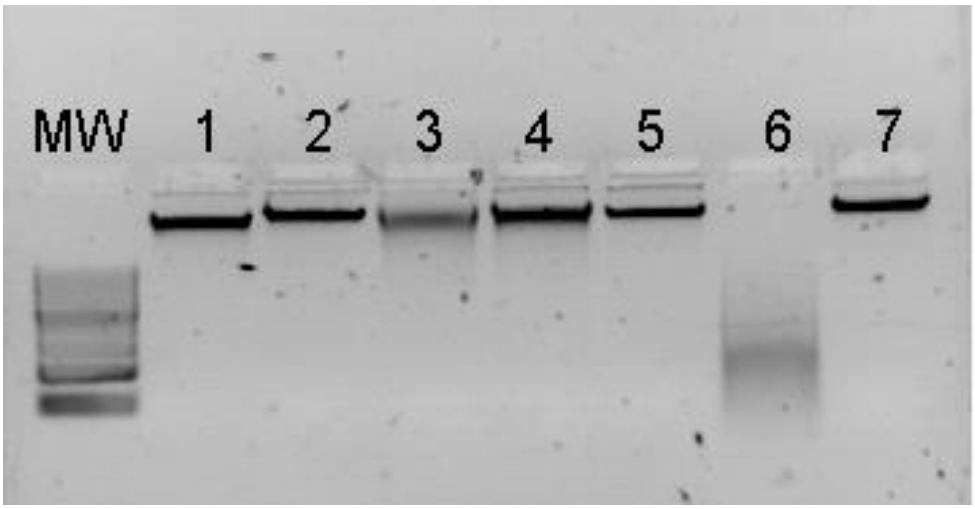


Figure 2.1: An example of an electrophoresis gel photograph depicting the DNA integrity of seven of the 14 haplotype samples. Lane MW: Molecular weight marker (GeneRuler™ 100bp Plus DNA Ladder; Thermo Scientific, Waltham, MA, USA). All samples showed a distinct, high-molecular weight band except Lane 6 which showed a smeared pattern indicating DNA degradation.

2.3 Primer Design

2.3.1 Gene Annotation

All gene sequences required for genetic analysis were annotated with the Annot v9 programme designed by Dr. G. Rebello (April 2006). These include *ABCA4*, *TIMP3* and *HMCN1*. In the case of *TIMP3*, whenever a SNP fell outside of the gene boundaries the sequence range upstream or downstream of the gene was increased to accommodate the SNP location. Two websites were used to obtain the full genomic sequence, namely: National Center of Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) and Ensembl (Flicek *et al* 2014) (http://www.ensembl.org/). Annotation accurately displays intron-exon boundaries, the most recent updated version of SNPs and genomic numbering of the gene sequence. The annotated sequence was used for sequence alignment, variation screening, primer alignment and calculation of the expected amplicon size for each exon of *ABCA4*.

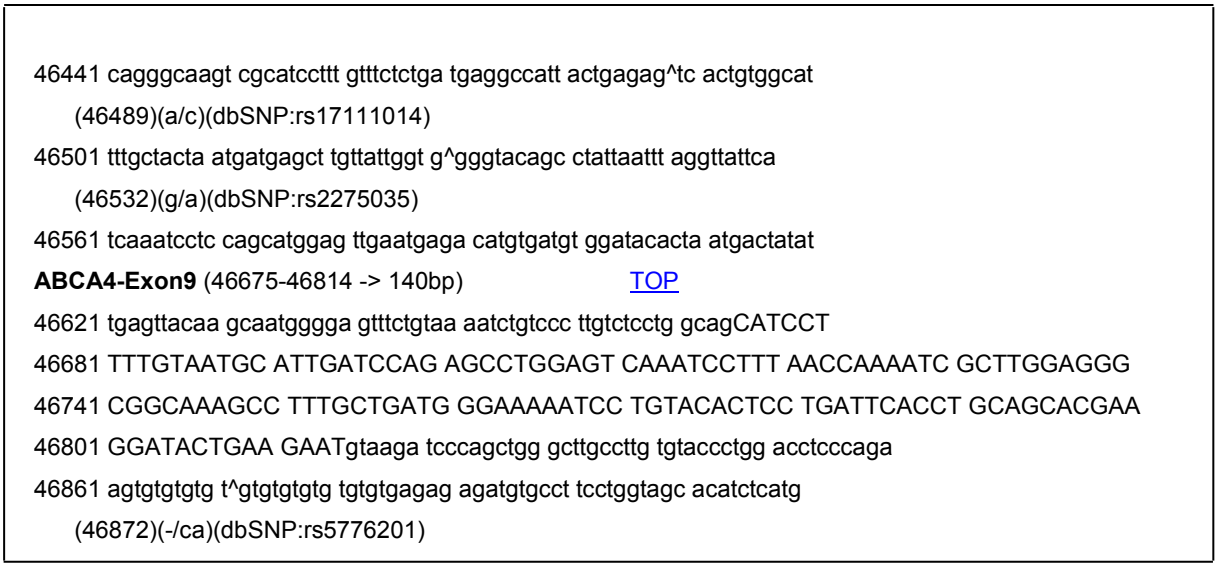


Figure 2.2: An example of the annotated *ABCA4* showing a region of exon nine. Exonic regions are indicated by uppercase letters and intronic regions by lowercase letters. SNP positions are indicated by a circumflex accent to the left of the nucleotide with information pertaining to nucleotide number, genotype and reference number within dbSNP shown below the line.

2.3.2 External Primer design

Primers are single-stranded synthetic oligonucleotides that bind externally to a specific region of choice in order to amplify that region by polymerase chain reaction (PCR). Primers were designed to flank at least 10-15 bp either side of each exon to include splicing regions. External primers were designed using NCBI/Primer-BLAST (Ye *et al* 2012) (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) which is a combination of Primer 3 (Untergasser *et al* 2012) (Whitehead Institute for Biomedical Research, 2004)(http://www.frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and NCBI BLAST (Camacho *et al* 2008) (Bethesda, USA) (http://www.ncbi.nlm.nih.gov/). Input data consisted of the genomic sequence of interest in FASTA format, retrieved from the gene annotation. The following parameters were considered when using the NCBI/Primer-BLAST (Ye *et al* 2012) program to maximize specific DNA amplification:

1. Primer length: Optimal primer length is 18-25 bp for adequate specificity
2. GC content: Preferably between 40 and 60 %.

3. Primer melting temperature (T_M): Reverse and forward primers with similar T_M within the range of 52 degrees Celsius (°C) -62°C were chosen.

Suitable primers within these parameters, and showing limited mismatches to unintended targets within the human genome were then analyzed with IDT (Integrated DNA Technologies) SciTools OligoAnalyzer 3.0 (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx) to screen for secondary structures. These structures were assessed on their stability, represented by the Gibbs free energy (ΔG), and include hairpins, homodimers and heterodimers. Primer pairs were excluded if dimer formation resulted in a ΔG of less than minus five and especially if it contained three prime end dimers and/or hairpins.

For each of the 50 exons of the *ABCA4*, external primers had been designed previously by researchers in the Division of Human Genetics at the University of Cape Town (UCT) (Appendix 5). From these external primers, working stocks of 20μM were diluted with SABAX dH_2O (Adcock Ingram, Johannesburg, SA) and stored at -4°C. Before commencing with optimization, each of the 50 primer pairs was evaluated for suitability according to the above criteria.

New external primers were designed according to the above criteria for exon nine of *ABCA4* due to an inability to optimize the PCR with the previously designed primers (Table 2.1). For each of the ten candidate SNPs in previously identified AMD-associated genes, showing high allelic frequency differences between Black and Caucasian populations, a new external primer pair was designed (Table 2.2). Gene regions 400 bp downstream and upstream from each SNP of interest were interrogated for the optimal primer pair. Whenever a SNP had more than one suitable primer pair, the amplicon size of other SNPs were considered. This assured that SNPs grouped within the same multiplex reaction had amplicon sizes that could easily be differentiated when performing agarose gel electrophoresis. All external primers were purchased from IDT via Whitehead Scientific (Pty) Ltd (Cape Town, SA).

Table 2.1: Information pertaining to the external primer pair for *ABCA4* exon nine

Exon	Primer	Length	Primer sequence	Size
9	Forward	20	5´-ttacaagcaatggggagttt-3´	467
	Reverse	20	5´-agagaattgtgaacccttcc-3´	

Table 2.2: Information relating to the external primer pairs for 10 SNPs in candidate AMD genes

SNP/Gene	Primer	Length	Primer sequence	Size
rs.17110714 ABCA4	Forward	20	5´-agacccgacaggactttgta-3´	581
	Reverse	20	5´-gtttacgacgaagcccatct-3´	
rs.17110878 ABCA4	Forward	20	5´-agactcaccccattcctgag-3´	203
	Reverse	20	5´-gagcaatcctgaaagccaga-3´	
rs.6703052 ABCA4	Forward	20	5´-acatcaggagacccagtttt-3´	195
	Reverse	20	5´-gaagcccagagcacatttat-3´	
rs.487906 ABCA4	Forward	20	5´-attcagcatgtcagagtcag-3´	511
	Reverse	20	5´-aggatcatacactgtggctt-3´	
rs.9621578 TIMP3	Forward	20	5´-cttggggtgaagtaaatggt-3´	299
	Reverse	20	5´-tcttgtgtcatgtgtaggct-3´	
rs.9621622 TIMP3	Forward	20	5´-caacctggcagaatccttag-3´	354
	Reverse	20	5´-attcccaaaagaaagcgtgt-3´	
rs.5998713 TIMP3	Forward	20	5´-cctgccaacctcacttagta-3´	169
	Reverse	20	5´-ttcttacccagcatcttcca-3´	
rs.1555494 HMCN1	Forward	20	5´-ctcatcccagaagtgacaca-3´	272
	Reverse	20	5´-aggaaaagtcaggtcagtgg-3´	
rs.1407428 HMCN1	Forward	21	5´-agtcttccatgttctctctgc-3´	571
	Reverse	20	5´-cgtagcattggggtgacaca-3´	
rs.6425006 HMCN1	Forward	20	5´-aagttggcaggagcacaaga-3´	327
	Reverse	20	5´-tgtttatggcactggggcaa-3´	

2.3.3 Internal Primer design

One internal primer was designed for each of the ten SNPs in candidate AMD genes (Table 2.3), along with the above mentioned external primers, for SNaPshot genotyping. The three prime tail of an internal primer binds a single base upstream or downstream from the SNP of interest, after which primer extension occurs with a single fluorescently labelled dideoxynucleotide triphosphate (ddNTP). This limits the choice of primer design to either the sense or antisense sequence immediately preceding the SNP. The presence of another SNP within either flanking sequence precluded it from selection. Final selection was based on the following parameters:

1. Primer length: A range of 13bp to 42bp with an ideal length of 25bp.
2. GC content: Between 20 and 80 % is acceptable. Ideally between 40 and 60 %.
3. Primer T_M: Primers multiplexed together should have similar melting temperatures.
4. Length modification: Primers multiplexed together need to differ by four to six nucleotides in order to distinguish between final SNaPshot products.
5. Secondary structures: Homodimers with less than minus nine ΔG and hairpins at high melting temperatures with involvement of the three prime end were deemed unacceptable.

The web based tool, IDT SciTools OligoAnalyzer 3.0 was used to select primers according to the above mentioned parameters. The Local BLAST application in BioEdit (Hall 1999) Sequence Alignment Editor v.7.0.5.2 (Ibis Biosciences, California, USA) was used to create a database containing the sequences of all SNP amplicons to be used for SNaPshot analysis. To assess for any non-specific binding to other amplicons within each multiplex reaction, each internal primer was blasted against the created database. Finally, all internal primers were aligned against each other using the web-based version of AutoDimer 1.0 (Vallone *et al*. 2004) (http://yellow.nist.gov:8444/dnaAnalysis/primerToolsPage.do) to ensure minimal dimerization within multiplex reactions. A text file with all the internal primer sequences was uploaded as input data. All default values were used except the total score required

which was set at three. This setting limits the output file to primer interactions that score three or more. Dimer formation is expected when internal primers bind at more than four consecutive nucleotide positions and when they specifically bind at the three prime end. Analysis in AutoDimer was repeated for each chosen internal primer once a random sequence five prime tail was added. Random tail sequences were generated using Sequence Manipulation Suite Version 2 (Stothard, 2000) (http://www.bioinformatics.org/sms2/).

Table 2.3: Internal primers for ten SNPs and the associated information.

SNP	Primer	Length	Primer sequence *
rs6703052	Forward	41	5´-agccagatga**ttttttagtatttttttcttcttcacatggaa** -3´
rs17110714	Forward	28	5´-**atggacatctagttcacataaatgctca-**3´
rs17110878	Forward	23	5´-**catgtatgtgctttcctgggcag**-3´
rs487906	Reverse	35	5´-gaagccacaccc**gataatttacatgaaaataccag** -3´
rs9621578	Reverse	25	5´-**agctagactggtgaattgggggaaa**-3´
rs5998713	Reverse	39	5´-gtaacggtaattgat**ctcagtaactaagagggcataaga** -3´
rs9621622	Reverse	25	5´-**tcccttattacccagtaggcaagctgttatc**-3´
rs1407428	Forward	33	5´-acagtgatat**aaagggaggctaagc tttggatt** -3´
rs1555494	Forward	31	5´-**cttgagcttgcacctgaccttcagt**-3´
rs6425006	Forward	37	5´-gtaacaaatcgagac**aagacaggcaa aggaaatcaag** -3´

*The sequence in bold represents the sequence designed to anneal to the complementary genomic DNA sequence. The rest of the sequence is the random selected nucleotide tail added to the primer.

2.4 Polymerase Chain Reaction

PCR involves the exponential amplification of a specific region of interest on any DNA sequence. The process firstly requires a pair of oligonucleotide primers complementary to a short sequence at both the five prime and three prime end of the target sequence. DNA polymerase, a heat resistant enzyme, then synthesizes a new strand of DNA complementary to the delineated target sequence using deoxynucleotide triphosphates (dNTPs). The process is repeated for 30 cycles through three different incubation temperatures, namely denaturation, annealing and extension, on a thermal cycler. The annealing temperature within the Standard PCR conditions changes depending on the length and composition of the specific primer pair for that exon. Optimization experiments for each exon were performed using Standard PCR cycling conditions and a Standard PCR reaction mixture as shown in Table 2.4 and 2.5 respectively, unless otherwise stated. Amplification was confirmed by agarose gel electrophoresis as described in Section 2.2.2. These reactions were performed on the GeneAmp® PCR System 9700 (Applied Biosystems, California, USA) unless otherwise indicated.

Table 2.4: Standard PCR Cycling conditions

Number of Cycles	Process	Temperature (°C)	Duration
1	Initial Denaturation	95	5 min
30	Denaturation	94	30 sec
	Annealing	exon/SNP specific	30 sec
	Elongation	72	40 sec
1	Final Elongation	72	7 min

Table 2.5: Standard PCR reaction mixtures

Reagents	Final Concentration
Forward Primer	10pmol
Reverse Primer	10pmol
dNTP's (Bioline, Michigan, USA)	200µM
Colourless GoTaq™ Reaction Buffer (Promega, Madison, USA)(1.5mM, pH8.5)	1 x
GoTaq™ DNA polymerase (Promega, Madison, USA)	0.5 units

2.4.1 Multiplex PCR

Multiplex PCR is an alternative PCR method whereby two or more loci are concurrently amplified in the same reaction. Advantages compared to singleplex PCR include both reduction in laboratory costs and time. All ten SNPs in candidate AMD genes were divided into three multiplex groups based upon amplicon size and primer compatibility (Table 2.6). All amplicons within each multiplex were selected to ensure a difference of at least 80 bp in size between them in order to improve visualization on agarose gel electrophoresis. Prior to multiplex grouping all internal primers and amplicons were checked for compatibility with AutoDimer 1.0 (Vallone and Butler 2004) to reduce the formation of unwanted heterodimers.

Multiplex PCR was conducted according to standard PCR reaction mixtures and cycling conditions as described in Table 2.4 and 2.5. Singleplex PCR, whereby a single loci is amplified in one reaction, was used for SNPs in candidate AMD genes within a multiplex group that eluded amplification and samples to be subsequently sequenced for genotype confirmation.

Table 2.6: Information relating to multiplex grouping of 10 SNPs in candidate AMD genes

Group	SNP	Amplicon size
1	rs.17110714	581
	rs.9621622	354
	rs.1555494	272
	rs.6703052	195
2	rs.1407428	571
	rs.6425006	327
	rs.17110878	203
3	rs.487906	511
	rs.9621578	299
	rs.5998713	169

2.4.2 PCR Optimization

PCR optimization was performed not only to maximize product yield but also to reduce amplification of non-specific products for accurate downstream DNA analysis. Various parameters, namely annealing temperature, Magnesium Chloride ($MgCl_2$) concentration and primer concentrations of each of the 50 *ABCA4* exons and for every multiplex group of SNPs in candidate AMD genes were altered during optimization. For each exon an optimized singleplex PCR was indicated by a single, bright band at the expected bp size while an optimized multiplex PCR had three or four equally bright bands at the expected bp size without any non-specific amplification. Further adjustments were introduced for samples showing distinct but weak bands. In these cases, DNA concentration was increased to 200ng/µl and the number of cycles in the Standard PCR cycling conditions was increased to 35.

2.4.1.1 Temperature-gradient PCR

Temperature gradient PCR identifies the annealing temperature at which a specific primer pair amplifies genomic DNA optimally. The temperature gradient was chosen to cover ±5°C from the calculated primer melting temperature for each exon and SNP primer pair. Five identical PCR reactions were set up spanning the temperature range of 10°C with a no template control (NTC) at the lowest annealing temperature. The NTC is added to monitor for contamination in the reaction and was run at the lowest temperature where non-specific amplification is generally more pronounced. PCR was conducted using 1µl of 100ng/µl DNA in a 25µl reaction volume with the reaction mixture (Table 2.5) and the remaining reaction volume was made up using SABAX dH_2O. All reactions were performed on the MultiGene cycler (Labnet International, Inc. Edison, NJ, USA) with standard cycling conditions (Table 2.4) except that the annealing temperature was set to span the 10°C interval range specific to that exon or SNP. The highest optimal annealing temperature was determined by agarose gel electrophoresis (Section 2.2.2). Any exons requiring further optimization were subject to $MgCl_2$ titrations or primer titrations.

For each multiplex PCR of SNPs in candidate AMD genes, a single annealing temperature was chosen within the overlapping range of potential annealing temperatures for each SNP within the group. Each multiplex reaction was then further optimized at the chosen annealing temperature with $MgCl_2$ titration experiments. An example of a temperature-gradient PCR is shown in Figure 2.3.

2.4.1.2 Magnesium and pH gradient

Magnesium (Mg^{2+}) serves as an essential co-factor for thermostable DNA polymerase. Optimal Mg^{2+} and pH concentration is essential for Taq DNA Polymerase activity and reduced non-specific amplification.

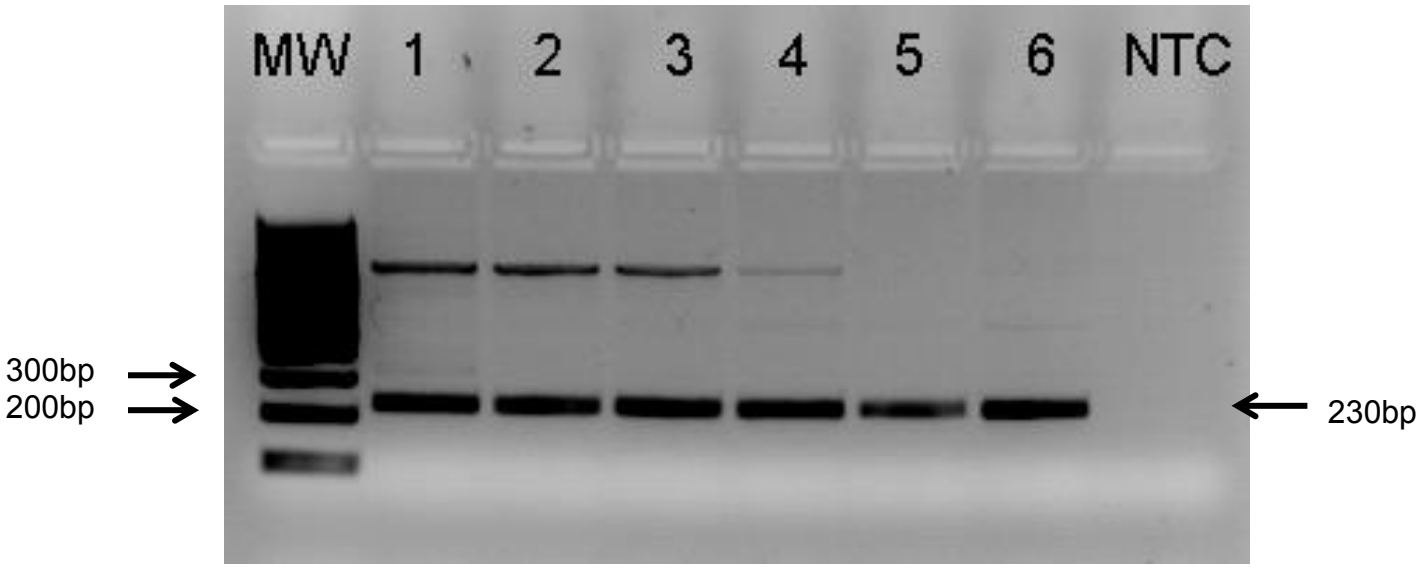


Figure 2.3: Electrophoresis gel photograph showing Temperature-gradient PCR for exon 5 of *ABCA4*. MW: GeneRuler™ 100bp Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA), Lane 1 – 6 represents identical PCR reactions differing only with respect to the annealing temperature. Lane 1: 55°C, Lane 2: 57.4°C, Lane 3: 58.9°C, Lane 4: 61°C, Lane 5: 63.4°C, Lane 6: 65°C, NTC: No-template control. Lane 5 indicates the optimal annealing temperature, as non-specific products are not amplified.

To determine the optimal Mg^{2+} and pH combination, an in-house PCR Optimization Kit created according to the Roche Optimization Kit manufacturer's instructions (Roche Diagnostics, Basel, Switzerland) was used. This kit contains 16 buffer solutions used to generate a $MgCl_2$ and pH gradient (Table 2.7). Using control DNA, 16 identical reactions were prepared with the standard PCR reaction mixture (Table 2.5) differing only by buffer solution and run at Standard PCR cycling conditions (Table 2.4). The PCR product of each reaction was then subjected to agarose gel electrophoresis (Section 2.2.2). The buffer generating a single, distinct band of DNA (or distinct bands in the case of a multiplex) at the expected molecular weight was chosen for further amplification and sequencing. Each multiplex group of SNPs in candidate AMD genes along with *ABCA4* exon 31 and 49 were subjected to $MgCl_2$ titration for complete optimization (Figure 2.4).

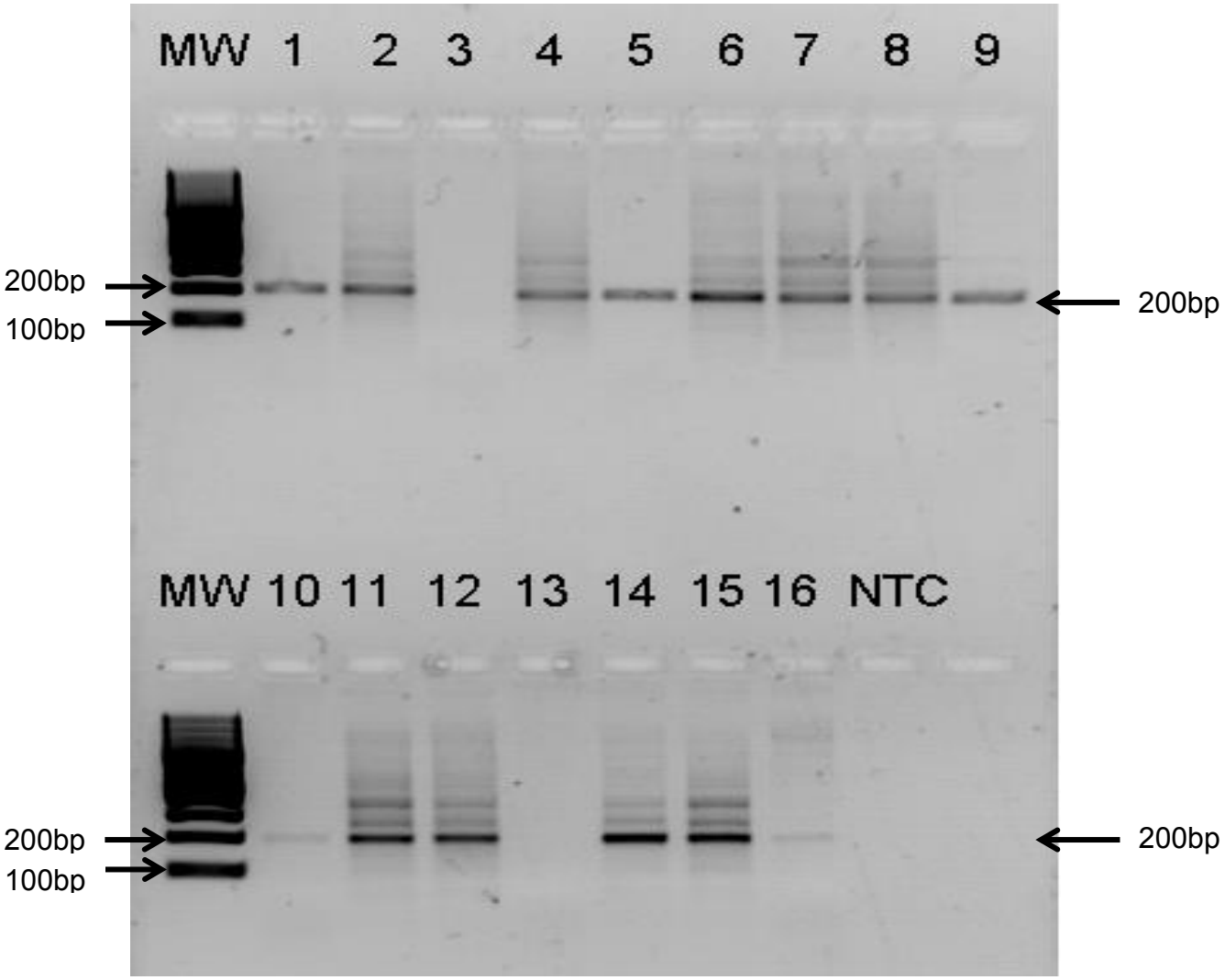


Figure 2.4: Electrophoresis gel photograph depicting a Magnesium and pH titration with 16 buffers from Roche PCR Optimisation kit for exon 49. MW: GeneRuler™ 100bp Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA). Lane numbers correspond to buffer numbers (Table 2.7), NTC: Non-template control. Buffer 1 was chosen for further amplification of exon 49.

Table 2.7: Roche Optimization kit buffers 1-16. Each buffer is numbered to show the corresponding $MgCl_2$ and pH concentration.

Buffer	pH	$MgCl_2$(mM)
1	8.3	1
2		1.5
3		2
4		2.5
5	8.6	1
6		1.5
7		2
8		2.5
9	8.9	1
10		1.5
11		2
12		2.5
13	9.2	1
14		1.5
15		2
16		2.5

2.4.1.3 Primer titration

High primer concentrations inhibit PCR reactions by primer-dimer formation and non-specific amplification. Primer titration involved four PCR reactions with Standard PCR reaction mixture (Table 2.5) differing only by primer concentration, which ranged from 10pmol to 4pmol in 2pmol intervals, run using Standard PCR cycling conditions (Table 2.4). The optimal primer concentration was selected as the concentration with the brightest, distinct band without any non-specific amplification. Only exon 30 of *ABCA4* required further optimization with primer titration.

Primer concentrations were adjusted to ensure equivalent band intensity of all SNPs in candidate AMD genes within a multiplex group. This adjustment was required for SNP rs17110878 within multiplex group two where the forward and reverse primer concentration was decreased from the standard 10pmol to 3pmol.

2.4.2 Cohort PCR amplification

Following optimization, each of the 50 *ABCA4* exons was amplified for all three test samples by PCR. The same optimized conditions for each exon was used for all three test samples and confirmed by agarose gel electrophoresis (Section 2.2.2). Similarly, all 154 AMD samples were amplified by PCR for each optimized multiplex SNP group of SNPs in candidate AMD genes. An example of an optimized PCR reaction is shown in Figure 2.5 and 2.6.

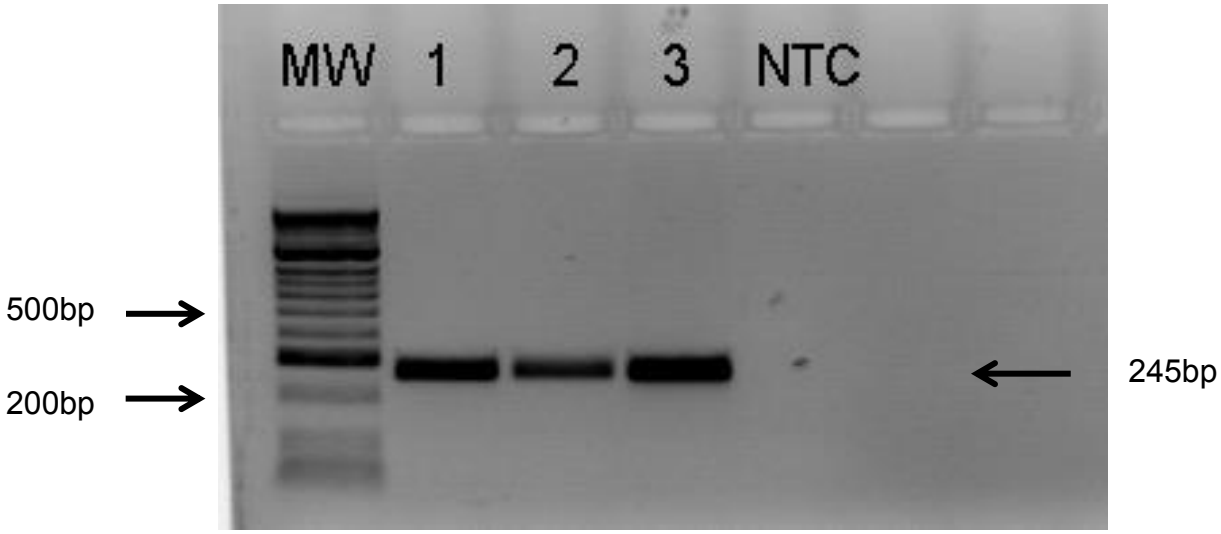


Figure 2.5: Electrophoresis gel photograph showing optimized PCR amplification for 3 test samples of exon 4 of *ABCA4*. MW: GeneRuler™ 100bp Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA). Lane 1 – 3 contain different samples and NTC the no-template control.

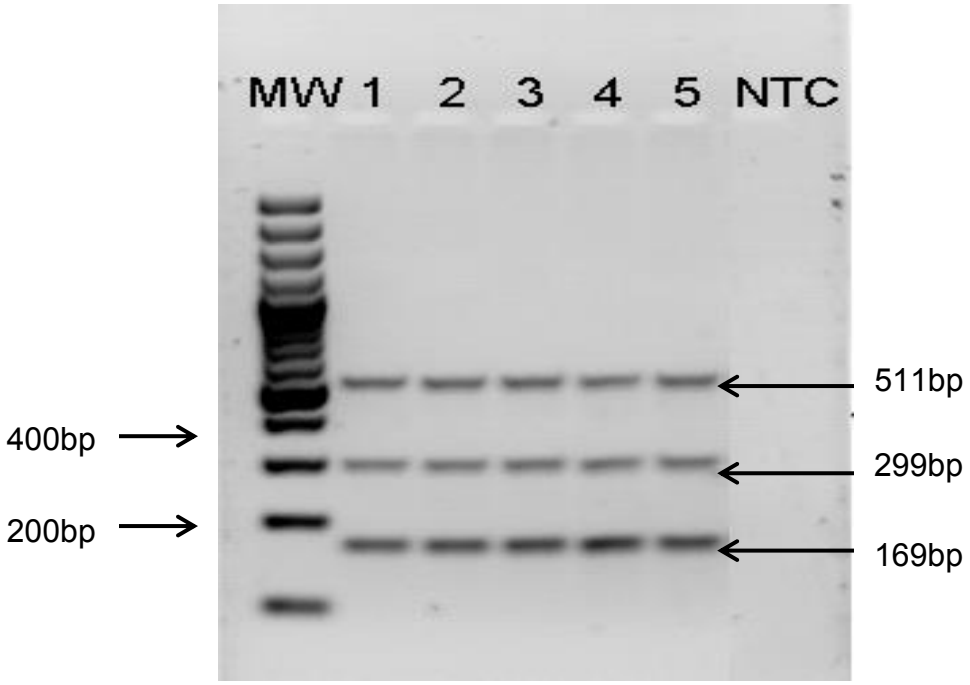


Figure 2.6: Electrophoresis gel photograph showing optimized PCR amplification for multiplex group 3 for SNPs in candidate AMD genes. MW: GeneRuler™ 100bp Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA). Lane 1 – 5 contain different samples and NTC the no-template control.

2.5 Direct cycle sequencing

Following PCR, all three test samples were subjected to direct cycle sequencing of 50 exons of ABCA4 in both the forward and reverse direction to determine the presence of any sequence variants. All exons were successfully sequenced in both directions producing satisfactory electropherogram data for variant calling except exon seven which only produced interpretable sequencing data from the forward primer. As the sequencing data from the forward primer completely covered exon seven and its splicing regions, troubleshooting efforts were curtailed for sequencing in the reverse direction. Possible factors leading to only one primer producing interpretable sequencing data includes poor primer design, a SNP within the reverse primer binding site and primer contamination. Both pre- and post-purification was performed prior to resolution of the sequence variants on the ABI Prism 3130xl Genetic Analyzer (Applied Biosystems, California, USA) automated sequencer.

2.5.1 PCR purification reaction

Enzymatic purification was used to purify PCR products before cycle sequencing. In order to maintain the ratio of dNTP's to ddNTP's for efficient cycle sequencing reactions shrimp alkaline phosphatase (SAP) was used to dephosphorylate unincorporated dNTP's. Exonuclease I degrades any unincorporated primers in a three prime to five prime direction that could possibly interfere with normal sequencing methods. Purification was performed in a 10µl reaction volume with 8.9µl PCR product, 1 unit (U) FastAP™ (ThermoFisher Scientific, Waltham, Massachusetts, USA), and 2 U Exonuclease I (ThermoFisher Scientific, Waltham, Massachusetts, USA). Purification was conducted on the GeneAmp® PCR System 9700 (Applied Biosystems, California, USA) and thermal cycling conditions consisted of an initial incubation step at 37°C for one hour, followed by an enzymatic inactivation step at 75°C for 15 minutes and a final inactivation step at 95°C for five minutes.

2.5.2 Cycle sequencing reaction

A cycle sequencing reaction is an adapted PCR technique that uses a single primer instead of two for linear amplification. Another difference is that it randomly incorporates fluorescently labelled ddNTP's into the three prime end of the amplified template fragments which halts the extension process. Sequencing reactions were prepared in a 20µl mixture of 4µl PCR product, 1x BigDye Termination Sequencing Buffer (Applied Biosystems, California, USA), 0.25x BigDye Termination mix (Applied Biosystems, California, USA) and either 20 or 40pmol primer (sense or antisense) depending on the length of the PCR product. For PCR products longer than 350 bp, 40pmol was used whereas 20pmol was sufficient for any products less than 350 bp. The sequencing cycling conditions comprised an initial denaturation step at 95°C for five min, followed by 30 cycles of 95°C for 30 seconds (sec), 50°C for 15 sec, and 60°C for four min. Cycle sequencing reactions were completed on the GeneAmp® PCR System 9700 (Applied Biosystems, California, USA).

2.5.3 Sequencing purification reaction

Ethanol precipitation was used to purify and concentrate the cycle sequencing products. The cycle sequencing products were added to 50µl absolute ethanol (Merck Chemicals, Gauteng, SA) and 2µl sodium acetate (3mM, pH5.5). Each sample was mixed and incubated overnight at -20°C. All samples were centrifuged at 10 000 revolutions per min (rpm) for 10min using the Microcentrifuge 5415D (Eppendorf, New York, USA). The supernatant was decanted before adding 50µl of 70% ethanol and a final centrifugation at 10 000rpm for 10 min. Again the supernatant was removed and the samples air dried for 60min in order to allow any excess ethanol to evaporate, as it would interfere with the sequencing.

2.5.4 Capillary electrophoresis

Each purified sequencing pellet was re-suspended in 10µl SABAX distilled water (Adcock Ingram, Johannesburg, SA). After careful pipette mixing, 5µl of the resuspended solution was mixed with 8µl of deionized injection solvent, Hi-Di™ Formamide (Applied Biosystems, Warrington, UK) in a MicroAmp® Optical 96-well Reaction Plate (Applied Biosystems). The reaction plate was loaded into the autosampler of the ABI Prism 3130xl Genetic Analyzer (Applied Biosystems) for capillary electrophoresis. The following settings were chosen for data capture: the analysis module used was 3130BigDyeTerminatorKBv5.2, the dye set was Z-BigDyeV3, and the run module was RapidSeq36_POP7_1. Electropherogram data was analysed by Sequencing Analysis v5.4 Software for Windows (Applied Biosystems, Copyright 2009) for viewing and editing.

2.6 Genotyping by SNaPshot

SNaPshot is a modified PCR genotyping method which can cost-effectively investigate numerous SNP in various locations in the human genome in a single multiplex reaction. All PCR amplicons, each containing a SNP of interest, are pooled within a multiplex

reaction with respective internal primers. DNA polymerase facilitates single base extension at the three prime end of the internal primer with available fluorescent labelled ddNTPs as described in section 2.3.3. Detection of colour peaks by capillary electrophoresis corresponding to specific incorporated ddNTPs determines the genotype for that sample (Table 2.8).

Table 2.8: Information pertaining to the fluorescently labelled dyes for each ddNTP used during SNaPshot genotyping.

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

2.6.1 Pre-SNaPshot purification

Prior to the SNaPshot reaction, all amplified PCR products were purified with shrimp alkaline phosphatase FastAP™ (ThermoFisher Scientific, Waltham, Massachusetts, USA) and Exonuclease I (E. Coli) (New England Biolabs Inc., USA) in order to remove unincorporated dNTP's and primers, respectively.
Each group multiplex purification reaction for SNPs in candidate AMD genes was conducted in a 20µl volume reaction consisting of 5µl PCR product, 1.5 U of FastAP™ (ThermoFisher Scientific, Waltham, Massachusetts, USA), 1 U of Exonuclease I (New England Biolabs Inc., USA), and SABAX distilled water (Adcock Ingram, Johannesburg, SA).

Singleplex purification reactions were conducted with equivalent reaction volumes to multiplex purification reactions except that 1U of FastAP™ (ThermoFisher Scientific, Waltham, Massachusetts, USA) was used instead of 1.5U. Purification cycling

conditions included an incubation step at 37°C for 60 min, followed by enzyme inactivation at 75°C for 15 min conducted on the GeneAmp® PCR System 9700 (Applied Biosystems, California, USA).

2.6.2 SNaPshot reaction

Following the initial purification step, separate multiplex SNaPshot reactions were performed for the three groups of SNPs in candidate AMD genes. For each group of SNPs in candidate AMD genes, internal SNaPshot primers were pooled in a ratio of 1:1:1:1 for group 1 and 1:1:1 for groups two and three, respectively. For every batch of multiplex reactions, fresh dilutions of pooled internal primers were used.

Multiplex SNaPshot reactions were conducted in a 10µl reaction volume, consisting of 5µl purified PCR product, 1µl pooled internal primers and 1µl SNaPshot Multiplex Ready Reaction Mix (Applied Biosystems, Warrington, UK). The remaining volume was made up using SABAX distilled water (Adcock Ingram, Johannesburg, SA).

Singleplex SNaPshot reactions were conducted with equivalent reaction volumes to multiplex SNaPshot reactions except that 1µl purified PCR product was used instead of 5µl. In order to recognize possible contamination, a negative control was included with every multiplex reaction with 1µl distilled water replacing the purified PCR product. Cycling conditions for both the multiplex and singleplex SNaPshot reactions were as follows: 25 cycles consisting of 10 sec at 96°C, 5 sec at 50°C, and 30 sec at 60°C. All experiments were performed on the GeneAmp® PCR System 9700 (Applied Biosystems, California, USA). All SNaPshot PCR products were stored at 4°C in dark conditions to reduce the loss of fluorescence post PCR.

2.6.3 Post-SNaPshot purification

Upon conclusion of the SNaPshot reaction, 1 U FastAP™ (ThermoFisher Scientific, Waltham, Massachusetts, USA) was added to each reaction to remove unincorporated

ddNTP's. This was followed by an incubation step at 37°C for 60 min and enzyme inactivation at 75°C for 15 min conducted on the GeneAmp® PCR System 9700 (Applied Biosystems, California, USA).

2.6.4 Capillary electrophoresis and genotype analysis

Subsequent to post-SNaPshot purification, the products were electrophoresed on the ABI Prism 3130xl Genetic Analyzer (Applied Biosystems). Firstly, a volume of 9µl of Hi-Di™ Formamide (Applied Biosystems, Warrington, UK) with 0.2µl GeneScan™ - 120LIZ™ size standard (Applied Biosystems, Warrington, UK) (Appendix 3) and 1µl of purified SNaPshot product was loaded on a MicroAmp® Optical 96-well Reaction Plate (Applied Biosystems).

Thereafter, the mixture was denatured for five min at 96°C on the Hybaid Touchdown TD-7500 Thermal Cycler (The Scientific Group, Milnerton, SA) and cooled on ice for two min. Hi-Di™ Formamide functions as a solvent that re-suspends DNA samples before electrokinetic injection while GeneScan™ - 120LIZ™ provides nine single-stranded LIZ® dye labeled fragments ranging from 15 to 120 nucleotides which creates a sizing curve ideal for SNaPshot genotype analysis. After loading the MicroAmp® Optical 96-well Reaction Plate within the autosampler, the following settings were chosen for data capture: the dye set was E5, and the run module was SNP_36_3V. Output data were examined using GeneMapper Software v4.1 (2009) (Applied Biosystems). Separate sets of bins and panels were setup for each group within GeneMapper Software v4.1 to assist in programmed allele detection.

Table 2.9: Expected and observed genotyping peak data from the multiplex grouping of 10 SNPs in candidate AMD genes. WT – wild type, Alt – Alternative allele.

SNP	Nucleotide Change	Expected Fragment Length(bp)	Observed Fragment Length(bp)	Allele peak (WT/Alt)
		Group 1		
rs.17110714	g.134151T>C	28	36.5/35	T/C
rs.9621622	c.-163+6142G>A	25	33/33.5	C/T
rs.1555494	c.900+5239A>G	31	38.5/38	A/G
rs.6703052	c.3523-508C>A	41	47.5/48	C/A
		Group 2		
rs.1407428	c.268+51430C>A	33	41/41	C/A
rs.6425006	c.269-18380T>C	37	43/42	T/C
rs.17110878	c.4667+544A>G	23	35/33	A/G
		Group 3		
rs.487906	c.5460+1537C>G	35	39/40.5	G/C
rs.9621578	c.711+2887A>G	25	35.5/37	T/C
rs.5998713	c.-1975G>T	39	44/45	C/A

2.6.5 Genotyping confirmation by cycle sequencing

A selection of samples for each SNP in the candidate AMD genes (representative of each individual genotype) were sequenced by direct cycle sequencing to confirm the accuracy of the SNaPshot genotyping results.

Sequencing was performed on the singleplex PCR product for the particular SNP and sample of interest in both the forward and the reverse direction as described in section 2.5.

2.7 Bioinformatic Analysis

2.7.1 *ABCA4* Sequence alignment and variant identification

All sequencing electropherogram data was uploaded onto BioEdit (Hall 1999) Sequence Alignment Editor v7.0.5.3 (Ibis Biosciences, California, USA) for further analysis and interpretation. The reference genomic sequence was uploaded from the annotated file (Section 2.3.1) in FASTA format, together with both the forward and reverse sequences generated from each respective amplicon. For multiple sequence alignment the ClustalW (Larkin *et al* 2007) (Kyoto University, Bioinformatics Center, Japan) application in BioEdit (Hall 1999) was used. Only good quality sequence data with tall, well-spaced peaks and minimal background interference were evaluated for sequence variations. Any sequence variations differing from the reference sequence were then further investigated with other bioinformatics tools.

The Human Genome Variation Society (HGVS) (http://www.hgvs.org/mutnomen/) contains nomenclature for the description of sequence variants. The most recent guidelines and recommendations provided were used to describe sequence variants, novel or previously identified, identified among the three test samples.

The Ensembl Project (Flicek *et al* 2014) (http://www.ensembl.org/) served as the genome browser of choice for further variation comparison and annotation. Ensembl Release 75 – February 2014 © version provided the comparative sequence in order to assess whether a variant was novel or previously identified. Variants identified as novel were further compared with the Exome Aggregation Consortium (ExAC) (Lek *et al* 2015) (http://exac.broadinstitute.org/) browser for verification. If previously identified, the SNP was further investigated for any gene or regulatory consequences using the SNP database, dbSNP (Sherry *et al* 2001) (http://www.ncbi.nlm.nih.gov/SNP/). Population allelic frequencies, especially among African populations, were extracted for every sequence variant from the 1000 Genomes Browser

(http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/) if available (McVean *et al* 2012).

2.7.2 Determination of pathogenicity

2.7.2.1 Missense variants

Missense mutations are non-synonymous variations where a base change results in a different codon that alters the amino acid sequence. To predict whether these mutations may be disease causing, the online predictor tool Pathogenic-or-Not-Pipeline (PON-P) (Olatubosun *et al* 2015) (http://bioinf.uta.fi/PON-P/index.shtml) was used. PON-P (Olatubosun *et al* 2015) incorporates five separate predictor tools to calculate the probability that missense mutations alter protein function and lead to disease. The five predictor tools are separated into either tolerance or stability predictors. Tolerance predictors include Predictor of human Deleterious Single Nucleotide Poly- morphisms (PhD-SNP)(version 2.0.6)(http://snps.biofold.org/phd-snp/phd-snp.html), Sorting Intolerant From Tolerant (SIFT)(version 4.0.3)(http://sift.jcvi.org/), PolyPhen-2 (version 2.0.22)(http://genetics.bwh.harvard.edu/pph2/) and Screening For NonAcceptable Polymorphisms (SNAP)(version1.0.8)(http://www.ngrl.org.uk/Manchester/ page/snap-screening-nonacceptable-polymorphisms). The only stability predictor used was I-Mutant (version 3.0.6) (http://folding.biofold.org/i-mutant/). Input included both the protein FASTA formatted sequence for each exon and the protein variations. PON-P (Olatubosun et al 2015) gives an output value between zero and one, with zero considered benign, and one considered pathogenic. A score of 0.5 is seen as uncertain. From this value the mutation was classified as either Neutral, Pathogenic or a 'Variant of unknown Significance' (VOUS). Bootstrapping is used to approximate the standard error of the prediction from which the prediction reliability is derived and included in the output data.

During our study, the designers of the PON-P (Olatubosun et al 2015) software released a faster and updated version namely, PON-P2 (Niroula *et al* 2015)

(http://structure.bmc.lu.se/PON-P2/) (Lund University, 2013). The identifier submission was chosen as input format to input the same four *ABCA4* missense variants using the Ensembl (Flicek *et al* 2014) identifier as reference. The output data, in terms of value and classification of these missense variations, changed between the two respective versions.

2.7.2.2 Synonymous and intronic variants

Aberrant splicing caused by exonic and intronic point mutations is an important mechanism leading to dysfunctional protein formation and genetic disease. In order to identify whether any variants lead to splicing defects, Human Splice Finder (HSF) Version 2.4.1 (13 October, 2010) (Desmet *et al* 2009) (http://www.umd.be/HSF/#) was used. Input data included both the wild-type and the mutant sequence for each *ABCA4* exon. Before analysis the first and last nucleotide of each exon was manually identified to define exon/intron boundaries. The following factors were considered when evaluating whether a variant may be pathogenic:

1. Donor/acceptor splice sites – Two separate matrices were used for prediction of potential donor/acceptor sites. Matrices derived from Shapiro and Senapathy (Shapiro *et al*. 1987) were used to derive a consensus value (CV) for each sequence and only sequences above a certain threshold value would qualify as potential sites. Output for these matrices was given as donor splice sites or acceptor splice sites. Any mutation that resulted in a reduction of 10 % or more in CV for any position, or a reduction of seven percent in position +4 for a specific sequence, was deemed to significantly affect splicing and hence should be further investigated as it could prove to be pathogenic (Desmet *et al*. 2009). A prediction algorithm modified from the MaxEnt script (Yeo *et al*. 2004) was also used to derive sequence predictions for splice site motifs. Output of these matrices was given as either 3´ or 5´ motifs.

2. Exonic Splicing Enhancers (ESE) - The spliceosome is guided by various intronic and exonic *cis*-elements to help define exon boundaries. Serine/Arginine-rich (SR) proteins

bind to these ESEs to promote exon definition. The binding of SR proteins (SRp40, SC35, SF2/ASF, SF2/ASF IgM/BRCA1 and SRp55) to their respective enhancer motifs were investigated from previously published algorithms on ESEfinder release 3.0 (2001-2006, Cold Spring Harbor Laboratory). From the output data only new or disrupted sites caused by the mutation were considered significant for further consideration. No definite guidelines were provided to confidently assign pathogenicity to any particular disrupted or newly created SR site but studies have shown that disruption in even one motif can result in disease, as shown by Cartegni *et al*. (2002) in spinal muscular atrophy. Therefore, in this study we ascribe significant weight to any variation that disrupts an ESE as possibly mutational.

3. Branching points - The branching point (BP) motif acts as an obligatory signal for pre-mRNA splicing. Only BP's within 100 nucleotides from the exon were considered for further processing. The threshold value of 67 was calculated using the consensus sequence YNYCRAY together with certain exclusion algorithms and a position weight matrix. Noteworthy variations reduced the threshold value below 67 and therefore resulted in a disrupted BP (Desmet *et al*. 2009).

2.8 Screening for possible pathogenic *ABCA4* variants in the STGD cohort

All variants tracking with the three common haplotypes indicated by PHASE (Stephens *et al*. 2001) analysis were pooled together to facilitate the determination of a possible novel mutation (Table 2.10). Haplotype four contained four synonymous changes, of which three were determined to be pathogenic by HSF along with two intronic variants namely c.5835+122T>G and c.5836 -11 G>A. Haplotype 12 contained three pathogenic variants as predicted by HSF - the exception being the c.2589 -39 T>C variant. Haplotype 29 had the only missense variants observed among all the haplotypes, although both were deemed benign after re-analysis by PON-P2 (Niroula et al 2015). The variant, c.3329 -34 A>G was deemed pathogenic by HSF, meaning all haplotypes contained at least one predicted pathogenic variation. Since no obvious mutational

changes were sequenced i.e. nonsense variants or intronic insertions/deletions within each haplotype, it was decided to rather screen the haplotype cohort for the three most probable pathogenic variants among all the variants found. The rationale for this decision is further discussed within the Discussion section.

Table 2.10: List of all variants occurring within each of the 3 common haplotypes.

Haplotype 4	Haplotype 12	Haplotype 29
c.3329 -58 C>T	c.2589 -38 G>C	c.3329 -34 A>G
c.5461-44 dupG	c.2589 -39 T>C	c.3602 T>G p.L1201R
c.5682 G>C p.L1894L	c.6282 +7 G>A	c.6764 G>T p.S2255I
c.5814 A>G p.L1938L	c.6249 C>T p.I2083I	c.5715-25A>C
c.5835+122T>G		
c.5844 A>G p.P1948P		
c.6069 T>C p.I2023I		
c.5836 -11 G>A		

Various factors were taken into account when choosing the most likely pathogenic variants for further genotyping. These factors include:

1. Clinical significance: any known association with STGD or previous determined pathogenicity by segregation analysis or functional studies.
2. Population frequencies: any mutant type alleles present in more than 10% of the normal African population as extracted from the Ensembl Genome Browser (Flicek *et al* 2014) were regarded as improbable to be pathogenic.
3. Splice site location: Variants situated within 10-15 bases of the intron/exon boundary especially on the five prime end were prioritized above variants situated far from the splice region for each exon.
4. PON-P results: The predicted class as indicated by the re-analysis by PON-P2 carried much significance in regard to pathogenicity determination.
5. HSF findings: Only variants with pathogenic changes as indicated within HSF parameters were considered specifically if changes involved branching points and splice sites.

The two missense variants which were found to be pathogenic, L1202R and V643M, according to PON-P (Olatubosun et al 2015) (the initial version used) were obvious choices. The third variant chosen was the synonymous variant, c.3831 G>A p.T1277T, in exon 26. HSF predicted it to alter exon structure and expression by forming a new branching and acceptor splice site, breaking existing five prime and three prime motifs and breaking a SF2/ASF enhancer site while forming two new SRp40 sites. Its position within the first 15 bases of the exon is critical to splicing of various transcripts and this supports its mutational effects as indicated by HSF. Reinforcing its pathogenic possibility is its relatively low minor allele frequency of three percent in the general population.

2.8.1 Screening for L1201R by Restriction Fragment Length Polymorphism analysis

Restriction Fragment Length Polymorphism (RFLP) analysis is a genotyping technique based on the ability of restriction endonucleases (RE) to cleave DNA at specific recognition sites, usually four to six base pairs long. SNP's alter enzyme recognition sites and result in different DNA fragment lengths when incubated with RE's. To determine the suitability of RFLP as a genotyping technique, all three potentially pathogenic *ABCA4* variants were analyzed with Webcutter version 2.0 (Maarek *et al* 1997) (http://rna.lundberg.gu.se/cutter2). The wild type DNA sequence was entered as input data and analyzed with a linear sequence method. Output display included a map of all restriction sites caused by all RE in the database. This process was repeated for the mutant DNA sequence containing the variant change and then compared to the wild type sequence. Disrupted enzyme recognition sites were further investigated to determine the number of cleavage sites within the amplified sequence for the specific RE and the resulting fragment lengths.

After analysis of all three haplotype-associated variants using Webcutter version 2.0 (Maarek *et al* 1997), only the L1202R variant was found to be detectable by RFLP. *Msp*

I was shown to cleave once, at a site 201bp into the 287 bp amplicon on the sequence containing the L1201R variant, but not on the wild type sequence. The single cut resulted in sufficient fragment size difference to be viewed on a two percent w/v agarose gel (Figure 2.7) (Table 2.11). The other possible pathogenic mutations, V643M and T1277T, were run on Webcutter version 2.0 (Maarek *et al* 1997) to assess whether they could be detected by RFLP. The V643M mutation caused an alteration to two RE recognition sites in the amplicon sequence in comparison to the wild type sequence. Both these REs cleave the sequence at three other sites within the amplicon, thereby creating numerous fragments indistinguishable by agarose gel electrophoresis. The T1277T mutation created no changes in RE recognition sites.

Table 2.11: Information regarding the restriction endonuclease *Msp I* used to detect the presence of the L1201R variant.

PCR product size	Restriction Endonuclease	Recognition site
287 bp	*Msp I*	5´-agtcc*gggat-3´ 3´-tcagg*cccta-5´

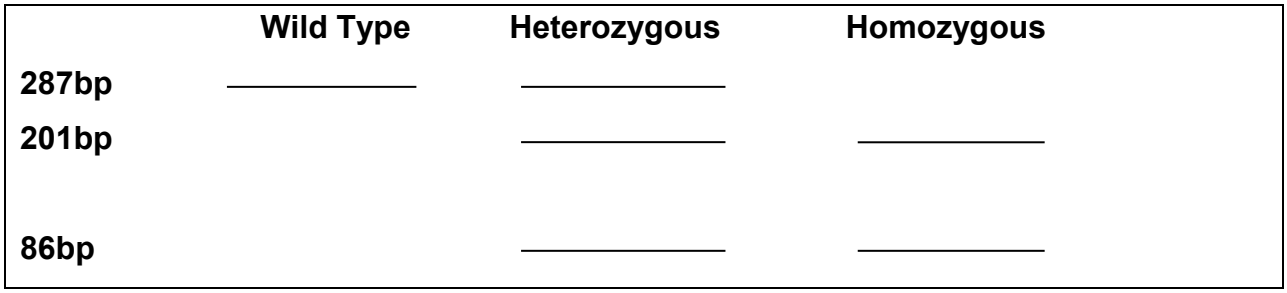


Figure 2.7: Illustration of expected DNA fragments for the restriction endonuclease digest screening of the L1201R mutation using *Msp I*. Wild type sequence without the L1201R mutation and both Heterozygous and Homozygous mutation sequences are shown.

All 14 of the haplotype samples were screened for the L1201R mutation by RFLP. *Msp I* (New England Biolabs, Ipswich, MA, USA) restriction endonuclease was obtained from the library of restriction enzymes stored within UCT's Division of Human Genetics at -

20°C. Firstly, a reaction mixture was prepared (Table 2.5). Aliquots of 8µl from the vortexed mixture were added into separate tubes containing 12µl PCR products from to equal a final reaction volume of 20µl. Individual tubes were incubated overnight at 37°C on a Bio TDB-100 (BioSan, Latvia) dry block thermostat before an additional three units *Msp I* enzyme and dH_2O was included for another 90 minutes.

Table 2.12: Reaction mixture used for RFLP analysis

Reagents	**Final Concentration**
Msp I	5 units (U)
NEB Buffer (New England Biolabs, Ipswich, MA, USA)	1 x
SABAX distilled H_2O	5.75µl

All haplotype samples were then subjected to agarose gel electrophoresis as per Standard Electrophoresis conditions except that the full 20µl post-digest reaction volume was used. L1201R variant positive and negative controls were added with the haplotype samples for comparison and to confirm success of the restriction digest.

2.8.2 Detection of T1277T and V643M mutation by direct cycle sequencing

Cycle sequencing was performed to screen the haplotype samples for the remaining mutations as they did not result in an altered RE recognition site. The mutation T1277T is located 18 bp into exon 26 and V643M 167 bp into exon 13. In order to screen for the presence of the mutations in all haplotype samples both exon 13 and 26 were sequenced by direct cycle sequencing. Firstly, both exons were amplified by PCR as per previously determined optimized conditions and confirmed by agarose gel electrophoresis. Amplified PCR products of all haplotype samples for both exons were then enzymatically purified and sequenced (Section 2.5) on the ABI Prism 3130xl Genetic Analyzer (Applied Biosystems). All electropherogram data were analyzed and screened for both mutations using Bioedit (Hall 1999) (Section 2.7.1).

Chapter 3: Results

3.1 Patient recruitment

3.1.1 Interrogation of the UCT Retinal Registry Database for STGD samples

The UCT Retinal Registry contains 234 STGD patients (212 families) divided ethnically into 178 Caucasian, 36 African, seven Mixed Ancestry (MA) and six Indian families, respectively. Of the 178 Caucasian probands, 104 (58.4 %) had their disease-causing mutations identified, compared to two (28.6 %) MA and two (33.3 %) Indian probands (Unpublished data). Only four of the 36 (11 %) African STGD probands in the UCT Retinal Registry that had been tested with the ABCR400 microarray had both mutated alleles of *ABCA4* identified, 11 (31%) had one mutated allele identified and 21 (58%) had no mutations identified (Unpublished data).

3.1.2 PHASE Haplotype analysis for STGD sample selection

Haplotype analysis of all 32 unrelated African STGD patients in which only one or no mutations were identified, showed a total of 35 different haplotypes. Haplotypes are assigned random numbers and three of these haplotypes, numbered 29, four and 12 respectively, occurred most frequently within the cohort.

Of these 32 African STGD patients, 17 carry at least one of the three common haplotypes determined by PHASE analysis (Stephens *et al*. 2001). Of these 17, five carry a combination of two of the most frequently observed haplotypes. Out of these five, three test samples were chosen to represent each unique combination of the common haplotypes (Table 3.1).

Table 3.1: Test samples representative of each unique combination of the three common haplotypes chosen for DNA sequencing to find novel *ABCA4* mutations. Haplotype numbers are those allocated by PHASE.

Test Sample	Haplotypes
RPM 1004.1	(12,29)
RPM 1075.1	(4,12)
RPS 1230.1	(4,29)

3.2 Sequence Analysis

All 50 *ABCA4* exons were successfully sequenced in both the forward and reverse direction except for exon seven which was only sequenced in the forward direction, in each of the three test samples. The electropherogram data produced was of sufficient quality for alignment and variant interpretation for all exons and splicing regions in the three test samples. Variations were accepted as genuine if present on both the forward and reverse strand for the specific exon. Heterozygous SNPs were called if a single peak position contained two overlapping peaks of fluorescence (Figure 3.1).

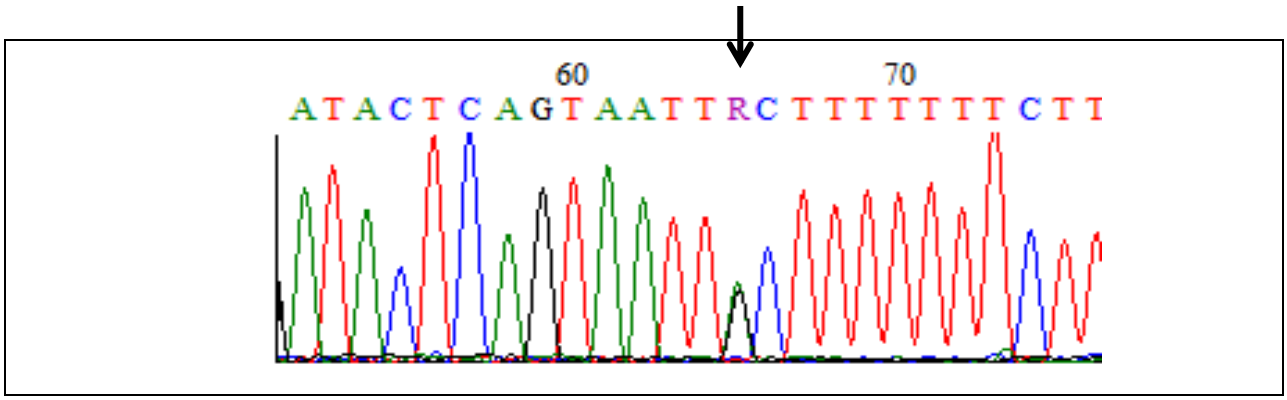


Figure 3.1: An electropherogram representing a segment of sequencing from the forward strand of exon 44 for test sample RPS 1230.1. The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The R peak at position 65 (indicated with an arrow) indicates the presence of the heterozygous SNP c.6006 -16G>A. An R peak is called when an A and a G nucleotide occur at the same position with green and black overlapping peaks, respectively.

Homozygous SNP's were called using the ClustalW (Larkin *et al* 2007) alignment sequence in Bioedit (Hall 1999) by comparing the annotated and the sample sequence (Figure 3.2).

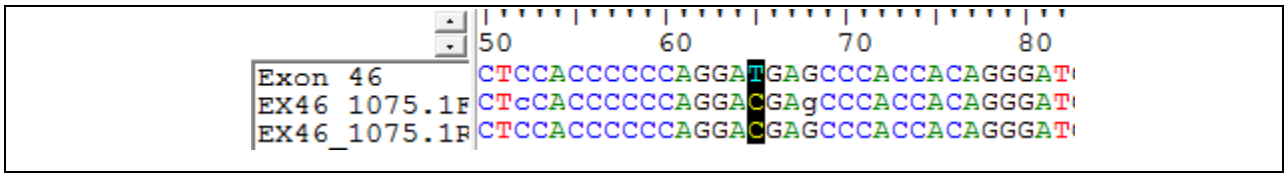


Figure 3.2: Multiple sequence alignment showing a homozygous change, c.6285 T>C, p.D2095D, in exon 46 for RPM 1075.1. The top sequence is uploaded from the annotated file and serves as the reference sequence for comparison to determine whether a variant is present. The second and third row of sequence corresponds to the forward and reverse sequence from the sample. The highlighted base at position 65 indicates where a C base is called instead of a T base indicating a homozygous change for that specific locus.

Heterozygous deletions and insertions cause multiple heterozygous peaks, as they result in a shifted sequence of one allele (Figure 3.3).

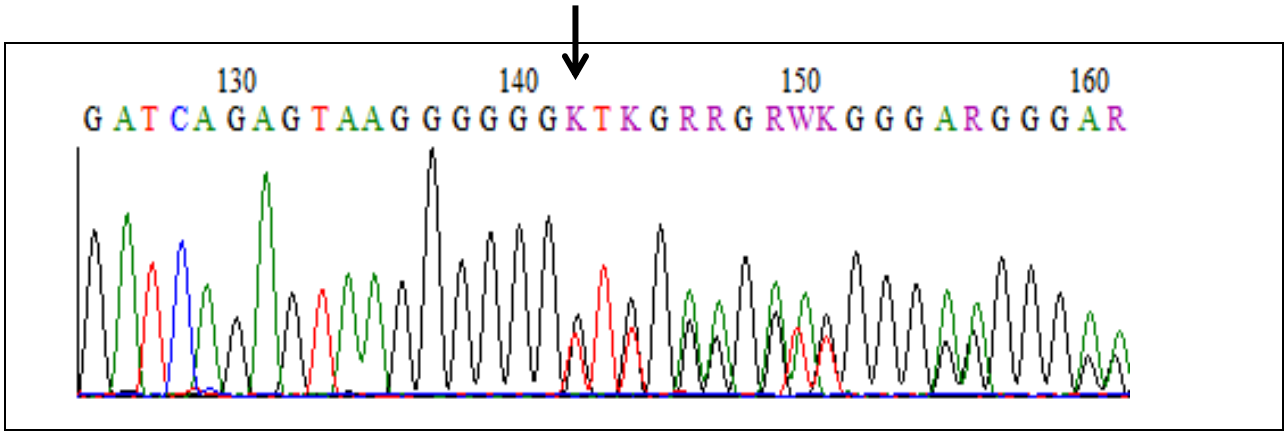


Figure 3.3: An electropherogram representing a segment of sequencing from the forward strand of exon 10 for test sample RPS 1230.1 showing a heterozygous insertion c.1356+5_1356+6 insG. The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The insertion is identified by normal sequence followed by multiple heterozygous peaks as the one allele is shifted by the insertion of a G nucleotide at position 142 (indicated with an arrow).

A total of 39 unique sequence variants were identified among the three test samples (Table 3.2). Of these 39, only three were considered novel as they have not been previously documented (according to our knowledge). Novelty was verified through comparison with the ExAc Browser (Lek *et al* 2015). Two of these novel variants, c.6148 -9 C>T and c.6148 -11 C>T, occur in all three test samples within the splicing region of exon 45 (Figure 3.4). The third novel variant, c.6480 +36 G>A, was found in the intronic region of exon 48 and was present only in RPS 1230.1 (Figure 3.5).

Appropriate concerns were raised whether these variants were to be considered real due to the noisy background on the electropherogram (Figure 3.4) apparent as multiple small and indeterminate peaks. Although plagued by similar background noise, the forward strand did show the same base call for both variants. The full list of sequence variants identified is shown in Table 3.3.

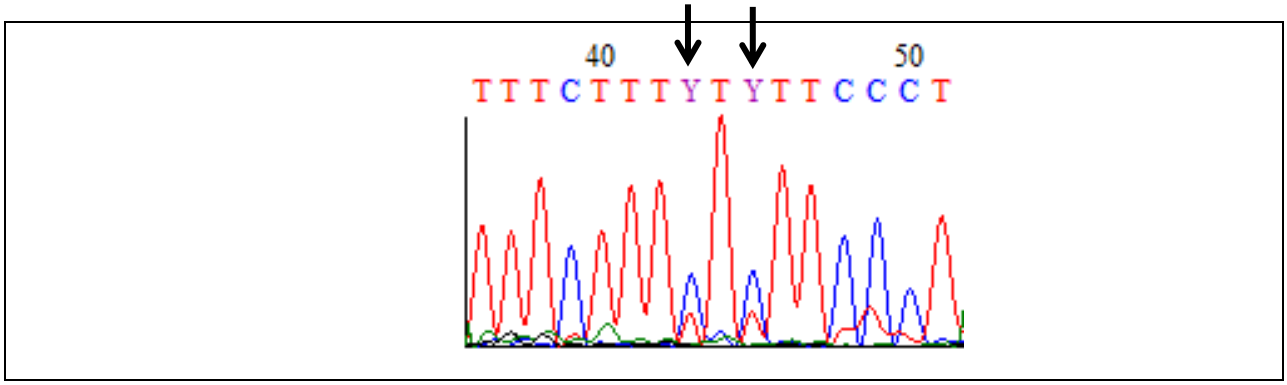


Figure 3.4: An electropherogram representing a segment of sequencing from the reverse strand of exon 45 for RPM 1004.1 showing two novel variants c.6148 -9 C>T and c.6148 -11 C>T that were identified in all three test samples. The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The base call Y at position 43 and 45 indicates a Thymine base simultaneously occurring with a Cytosine base.

Table 3.2: Total number of variations found for each of the three test samples after sequencing of all 50 exons of *ABCA4.*

	RPM 1004.1	RPM 1075.1	RPS 1230.1	All 3 TS*
Intronic	13	14	14	23
Missense	2	1	3	4
Insertion	1	2	2	2
Deletion	1	1	1	1
Synonymous	3	7	8	9
Novel	2	2	3	3
Total	20	25	28	39

***** The heading 'All 3 TS' indicates the total number of unique variants found between the three test samples for that specific type of variant.

Table 3.3: List of all unique variants found among the three test samples after sequencing all 50 exons of *ABCA4.* Test Sample indicates in which sample(s) the specific variant occurs. Test Sample 1 – RPM 1004.1, Test Sample 2 – RPM 1075.1, Test Sample 3 – RPS 1230.1. A total number of 39 unique variants were found.

Variant Type	Exon	Variant	Test Sample
Synonymous	2	c.141 A>G p.P47P	1,2,3
Intronic	3	c.302 + 26 A>G	1,2,3
Missense	6	c.677 G>T p.R226L	3
Intronic	7	c.858 + 14 T>C	1
Intronic	9	c.1239+66 G>A	1
Insertion	10	c.1356+5_1356+6 insG	1,2,3
Intronic	10	c.1240 -14 C>T	1,2,3
Intronic	10	c.1356+6 G>C	2
Missense	13	c.1927 G<A p.V643M	2
Intronic	18	c.2589 -38 G<C	1,2
Intronic	18	c.2589 -39 G<C	1,2
Intronic	21	c.2919 -14 T<A	2
Deletion	23	c.3329 -112delT	1,2,3
Missense	24	c.3602 T<G p.L1201R	1,3
Intronic	23	c.3329 -34 A>G	1,3
Intronic	24	c.3523-85C>T	2,3

Synonymous	26	c.3831 G>A p.T1277T	3
Intronic	33	c.4773 + 48 C>T	1,2,3
Intronic	34	c.4774 +42 A>G	3
Insertion	39	c.5461-51_5461-50insG	2,3
Synonymous	40	c.5682 G>C p.L1894L	2,3
Synonymous	41	c.5814 A>G p.L1938L	2,3
Intronic	41	c.5715-25A>C	2,3
Intronic	41	c.5835+122T>G	2,3
Intronic	42	c.5836 -11 G>A	2,3
Synonymous	42	c.5844 A>G p.P1948P	2,3
Synonymous	44	c.6069 T>C p.I2023I	2,3
Intronic	44	c.6006 -16G>A	3
Novel Intronic	45	c.6148 -9 C>T	1,2,3
Novel Intronic	45	c.6148 -11 C>T	1,2,3
Synonymous	45	c.6249 C>T p.I2083I	1,2
Intronic	45	c.6282 +7 G>A	1,2
Synonymous	46	c.6285 T>C p.D2095D	1,2,3
Novel Intronic	48	c.6480 +36 G>A	3
Missense	49	c.6764 G>T p.S2255I	1,3
Synonymous	49	c.6732 G>A p.V2244V	3
Intronic	49	c.6730 -3T>C	1
Intronic	49	c.6816 +28 G>C	1
Intronic	50	c.6817-85 C>T	3

3.3 Determination of pathogenicity of *ABCA4* variants

3.3.1 Missense variants

PON-P (Olatubosun et al 2015) was used to evaluate the pathogenicity of the four missense variants found in the three test samples (Table 3.4). As described previously (Section 2.7.2.1) a ‘probability of pathogenicity score’ of ‘0’ (zero) categorizes the variant as benign and ‘1’ (one) as pathogenic. Two of these variants, V643M and L1201R, were classified as pathogenic. The mutation, S2255I, was deemed benign and R226L considered a VOUS.

As the more reliable, accurate and updated version of PON-P (Olatubosun et al 2015) became available during the course of this study all four missense variants were re-analyzed with PON-P2 (Niroula et al 2015) (Table 3.5).

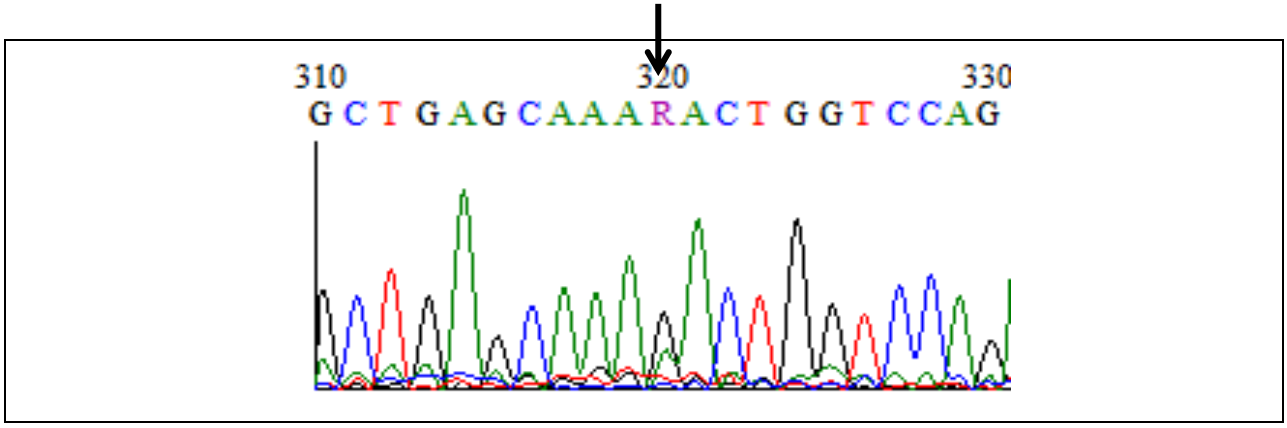


Figure 3.5: An electropherogram representing a segment of sequencing from the forward strand of exon 48 for test sample RPS 1230.1 showing the novel variant c.6480 +36 G>A. The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The R base called at position 320 occurs when both an Adenine and a Guanine base have equal fluorescence for that locus.

Table 3.4: List of all four missense variants identified in *ABCA4* of the three test samples with the respective PON-P results.

Exon	Mutation	Predicted class	Probability of pathogenicity	Test sample
6	c.677 G>T, p.R226L	Unknown	0.17	3
13	c.1928 G>A, p.V643M	Pathogenic	0.88	2
24	c.3602 T>G, p.L1201R	Pathogenic	0.88	1, 3
49	c.6764 G>T, p.S2255I	Neutral	0.11	1, 3

Table 3.5: List of all four missense variants identified in *ABCA4* of the three test samples with the respective PON-P2 results.

Exon	Mutation	Predicted class	Probability of pathogenicity	Test sample
6	c.677 G>T, p.R226L	Unknown	0.47	3
13	c.1928 G>A, p.V643M	Pathogenic	0.95	2
24	c.3602 T>G, p.L1201R	Unknown	0.61	1, 3
49	c.6764 G>T, p.S2255I	Neutral	0.06	1, 3

The probability of pathogenicity values changed for each variant but only mutation L1201R dropped significantly enough to alter the predicted class from pathogenic to unknown. The allelic frequencies of each missense variant were extracted from the 1000 Genomes (McVean *et al* 2012) African population data comprising of the 'Yoruba in Nigeria', 'Luhya in Kenya' and 'Americans of African descent'; which totalled 245 samples. The V643M mutation showed a C:T allelic ratio of 99:1. The L1201R and S2255I mutations showed a A:C allelic ratio of 91:9 and 52:48, respectively. The R226L mutation had no 1000 Genomes (McVean *et al* 2012) data available for these African populations. After considering both software pathogenicity predictions and population frequencies for each of the four missense mutations, only the V643M mutation was regarded pathogenic.

3.3.2 Synonymous and intronic variants

Pathogenicity prediction was done *in silico* for each synonymous and intronic variant through the Human Splice Finder Version 2.4.1 (Desmet *et al* 2009) (13 October, 2010) (http://www.umd.be/HSF/#). Results were restricted to variants with significant changes occurring in donor/acceptor splice sites, or affecting exonic splicing enhancers and branching points as these are the only parameters that provided pathogenicity thresholds (Table 3.6) (Desmet *et al*. 2009).These thresholds include a reduction of 10% in CV value in any position or seven percent in position four for donor/acceptor splice sites, any disruption of ESE sites and broken branching points caused by a threshold value under 67 (Section 2.7.2.2).

Table 3.6: List of variants with pathogenic changes in donor/acceptor splice sites, exonic splicing enhancers or branching points as predicted by Human Splice Finder. D/A – Donor/Acceptor, ESE – Exonic Splicing Enhancers.

Variant	D/A splice site	ESE	Branching points
c.141 A>G p.P47P		SRp40	AACCCAC
c.302+26 A>G	Acceptor 3' motif		

	5' motif		
c.1239+66 G>A	2 Donor	2 SRp55	
c.1356+5_1356+6insG	5´ motif		
c.1240-14 C>T	3´ motif		
c.1356+6 G>C	Acceptor 5´ motif		
c.2589 -38 G>C		SRp40, SC35	
c.2919 -14 T>A	3´ motif		
c.3329 -34 A>G	Acceptor Donor		
c.3523-85C>T		SF2/ASF (IgM-BRCA1) SF2/ASF	
c.3831 G>A p.T1277T	3´ motif 5´ motif	SF2/ASF	
c.4773 + 48 C>T		SC35	
c.5682 G>C p.L1894L	3´ motif	SC35	
c.5814 A>G p.L1938L	3´ motif		ATCTTAA
c.5835+122T>G	3´ motif Donor		
c.5836 -11 G>A		SC35, SRp55	
c.5844 A>G p.P1948P	Acceptor 5´ motif		
c.6249 C>T p.I2083I		SF2/ASF (IgM-BRCA1)	ATCGCAC
c.6282 +7 G>A		SC35, SRp40	
c.6480 +36 G>A	Acceptor		
c.6816 +28 G>C	Acceptor	SC35	
c.6817 -85 C>T		SC35,SRp40	

HSF predicted that 22 different variations resulted in possible pathogenic changes due to interference with splicing sites. This equates to a remarkable 59 % of all sequence variants found being deemed pathogenic, according to HSF. These results did not

include newly formed branching or SR protein binding sites as these were not deemed pathogenic by HSF. Due to the complexity of *in silico* mutational analysis especially in splicing regions, it was difficult to choose a clear variant for further mutational screening within the haplotype groups. Factors such as distance from the intron/exon boundary and allelic frequencies within African populations were considered when deciding upon candidates for mutational screening. Test sample one had nine predicted possible pathogenic changes according to HSF while test samples two and three had 16 and 12 changes, respectively. This overestimation of pathogenicity and inability to rank the severity between different pathogenic variants caused the screening for these variants in the extended cohort of samples with haplotypes to be abandoned.

3.4 Screening for possible pathogenic *ABCA4* variants in an extended cohort

3.4.1 Detection of L1201R mutation by RE *Msp I*

All 14 haplotype samples were subjected to gel electrophoresis after overnight incubation with RE *Msp I* in order to genotype for the L1201R mutation. Two haplotype samples namely RPM 398.1 and RPM 1168.1, showed banding patterns indicative of a heterozygous L1201R mutation (Figure 3.6). All other samples showed a wild type banding pattern and RE function was confirmed on each gel by a positive and negative control.

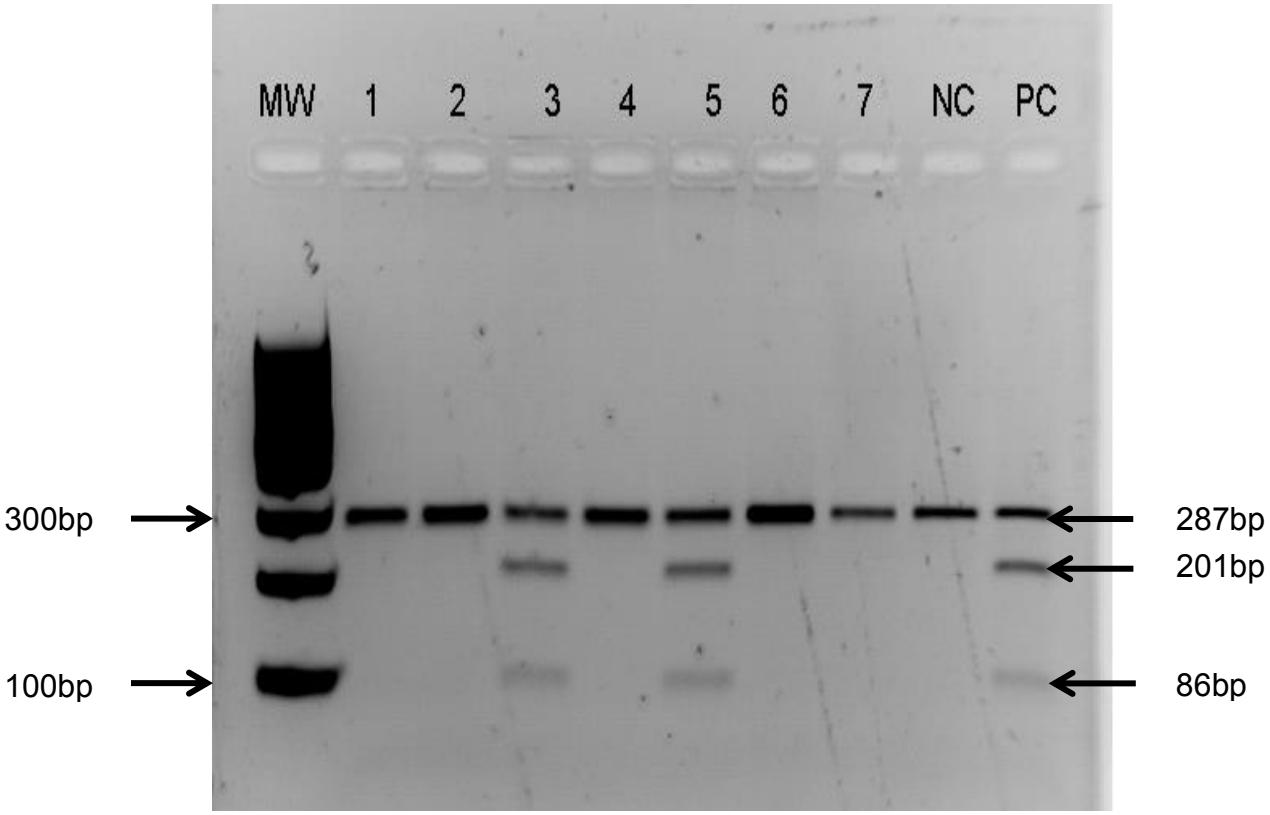


Figure 3.6: Electrophoresis gel image showing a restriction enzyme digest of seven haplotype samples with *Msp I* for mutation L1201R. MW: GeneRuler™ 100bp Plus DNA Ladder (Thermo Scientific, Waltham, MA, USA). Lane 1 – 7: Haplotype samples. NC: negative control. PC: positive control. Lane 3 (RPM 398.1) and 5 (RPM 1168.1) showed banding patterns consistent with a heterozygous L1201R mutation.

3.4.2 Detection of the V643M and T1277T mutation by direct cycle sequencing

All 14 haplotype samples were sequenced for the V643M and T1277T mutations. Both exons 13 and 26 showed sequence data of satisfactory quality for mutation screening. None of the 14 samples were found to be positive for the V643M mutation, while only one sample (RPM 960.2JIM), was found to be heterozygous for the T1277T mutation.

In summary, of the 36 African STGD cohort, only four patients (i.e. 11%) had a complete molecular diagnosis, i.e. two mutations, and 11 patients had only one mutation after microarray screening. After exon sequencing of *ABCA4* of three samples representative of the most common haplotypes within the STGD cohort, no new

definitive novel pathogenic mutations were found. All samples within the indigenous African STGD cohort containing any of the three common haplotypes were screened for the three most probable pathogenic variants found after sequencing and *in silico* pathogenicity prediction. One pathogenic mutation, L1201R, was re-classified as an 'Variant of Unknown Significance' according to the updated version of software prediction program, PON-P2 (Niroula et al 2015). In light of the above screening results and new pathogenic status of the L1201R mutation, only two patients (i.e. six percent) of the 36 African STGD cohort remain with two identified heterozygous mutations, ten patients have one heterozygous mutation and 24 patients still have no mutations identified.

3.5 Genotyping of SNPs in candidate AMD genes by SNaPshot

As explained in Section 2.1.3, the next part of the investigation involved identifying protective or susceptibility loci for AMD based on significant allelic frequency differences in genes previously associated with AMD.

Screening for the 10 SNPs was conducted on a cohort of 154 individuals diagnosed with AMD. Of these individuals, 127 were fully genotyped for all ten SNPs. Failure to genotype the full complement of SNPs for the remaining samples was most likely due to DNA degradation, and impurities. The 127 AMD individuals were further divided into Caucasians, African and Mixed Ancestry for further data analysis. Nine of the ten SNPs were incorporated into three different multiplex groups consisting of four (Figure 3.7), three (Figure 3.9) and two (Figure 3.11) SNPs respectively (Table 2.6), while genotyping for rs487906 SNP was performed using a 'singleplex' SNaPshot reaction (Figure 3.13).

Table 3.7 Allele frequencies within AMD cohort for each SNP in candidate AMD genes. Allele frequencies are shown according to ration of major allele/minor allele.

	Ethnic groups					
	Caucasian		Mixed Ancestry		Africans	
SNPs	Cases	Controls	Cases	Controls	Cases	Controls
rs9621622	.995/.005	.999/.001	1.00/.000	.880/.120	1.00/.000	.607/.393
rs9621578	.980/.020	.995/.005	.973/.027	.780/.220	.833/.167	.714/.286
rs5998713	.930/.070	.950/.050	.917/.083	.570/.430	.611/.389	.486/.514
rs1555494	.970/.030	.999/.001	.780/.220	.740/.260	.444/.556	.243/.757
rs1407428	1.00/.000	.995/.005	.917/.083	.820/.180	.667/.333	.614/.386
rs6425006	1.00/.000	.995/.005	.917/.083	.730/.270	.778/.222	.679/.321
rs17110714	.985/.015	.965/.035	.917/.083	.630/.370	.833/.167	.514/.485
rs6703052	.920/.080	.955/.005	.840/.160	.650/.350	.556/.444	.471/.529
rs17110878	1.00/.000	1.00/.000	.950/.050	.670/.330	.722/.278	.807/.193

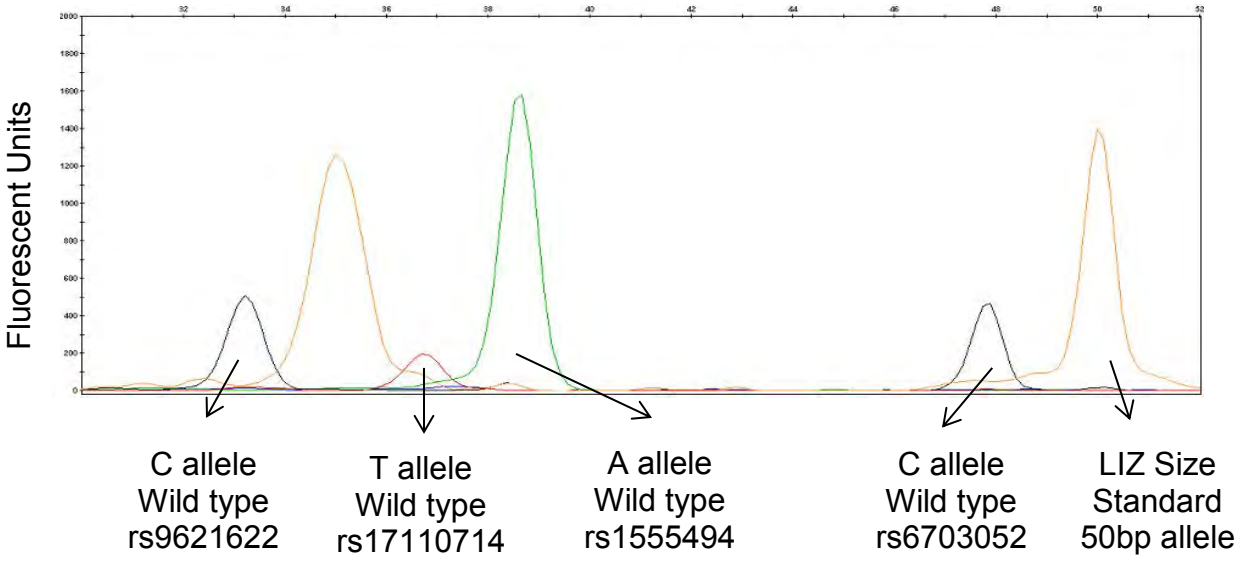


Figure 3.7: A SNaPshot reaction electropherogram for a sample that shows the wild type allele for all four SNPs within Multiplex group one namely, rs1555494, rs6703052, rs17110714 and rs9621622. GeneScan™ 120LIZ™ Size standard is shown at 35bp and 50bp. Black peak: cytosine dideoxynucleotide. Red peak: Thymine dideoxynucleotide. Blue peak: Guanine dideoxynucleotide. Green peak: Adenine dideoxynucleotide.

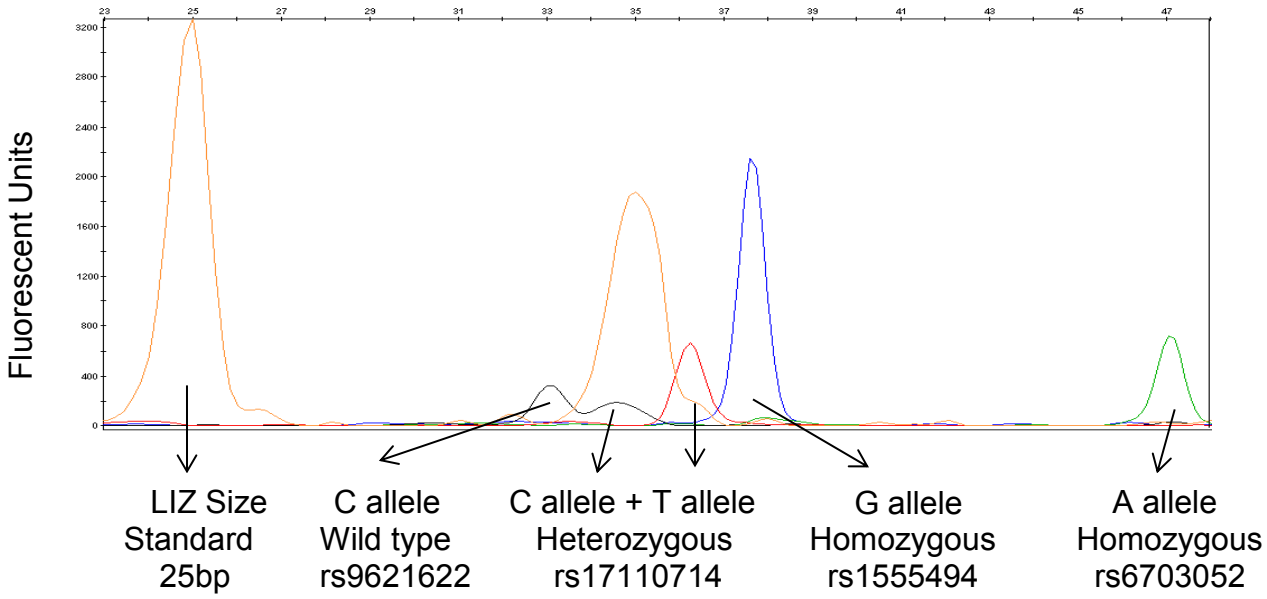


Figure 3.8: A SNaPshot reaction electropherogram of Multiplex group one for a sample that is homozygous for the alternative allele for both SNP rs1555494 and rs6703052, heterozygous for the rs17110714 SNP and wild type for the rs9621622 SNP. Size standard and colours of peaks as described in Figure 3.7

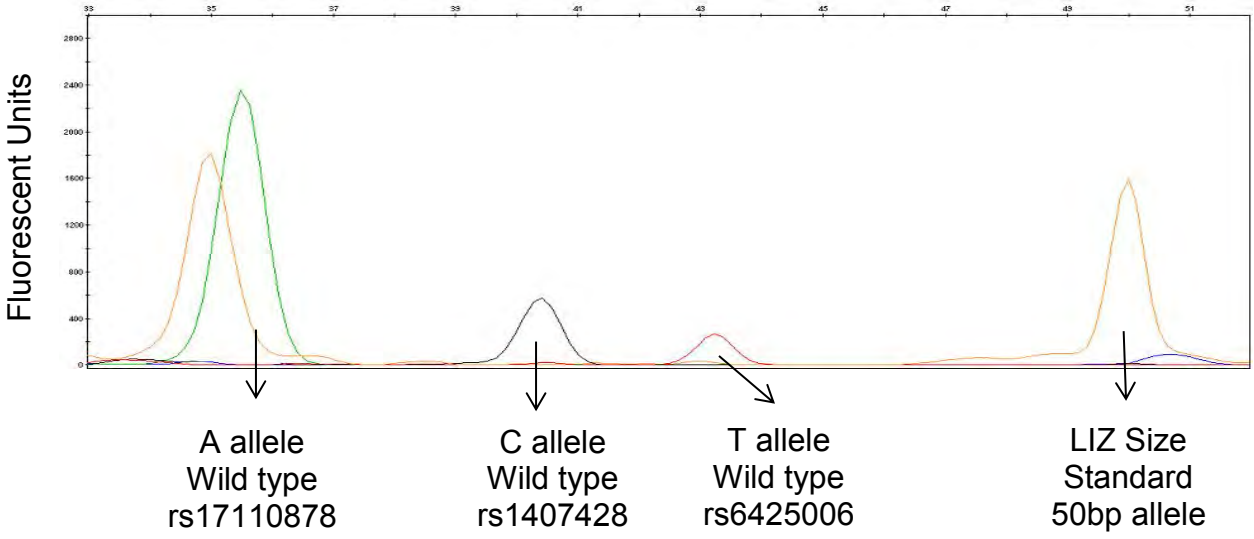


Figure 3.9: A SNaPshot reaction electropherogram for a sample that shows the wild type allele for all three SNPs within Multiplex group two namely, rs17110878, rs6425006 and rs1407428. Size standard and colours of peaks as described in Figure 3.7

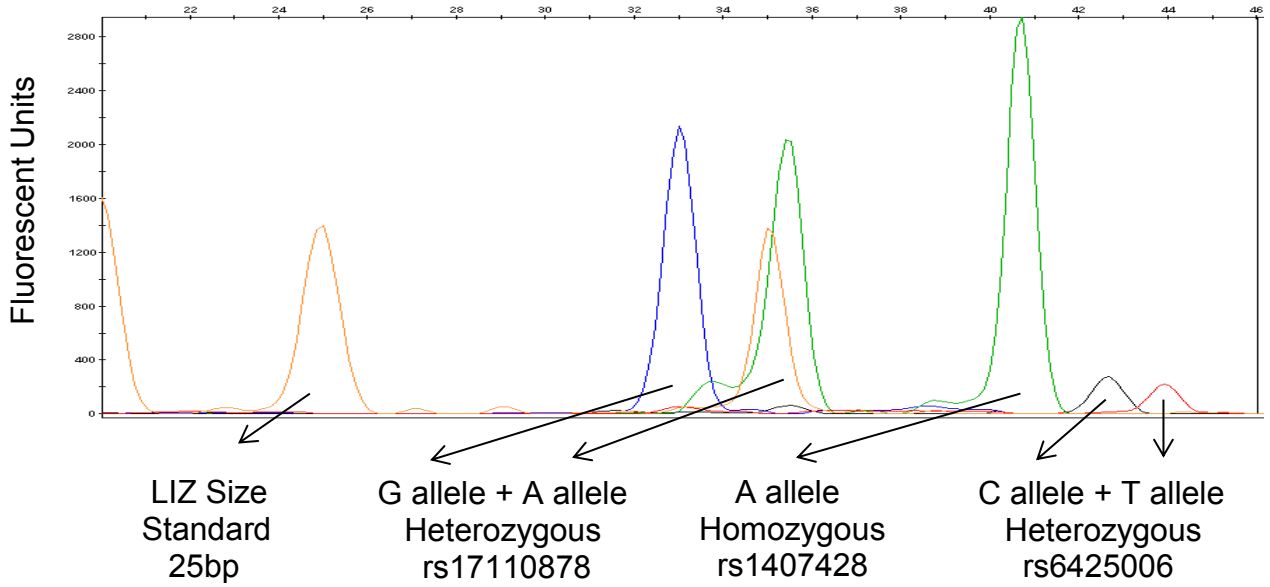


Figure 3.10: A SNaPshot reaction electropherogram of Multiplex group two for a sample that is heterozygous for the SNP rs17110878 and rs6425006 and homozygous for the alternative allele for SNP rs1407428. Size standard and colours of peaks as described in Figure 3.7

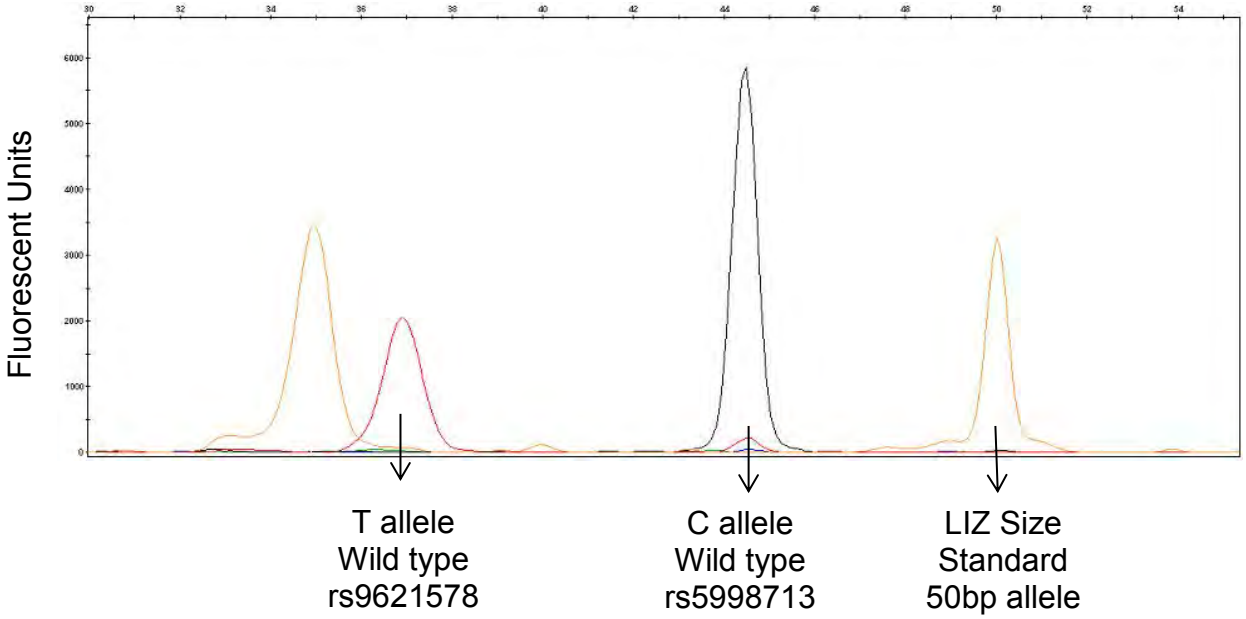


Figure 3.11: A SNaPshot reaction electropherogram for a sample that shows the wild type allele for two SNPs within Multiplex group three namely rs5998713 and rs9621578. Size standard and colours of peaks as described in Figure 3.7

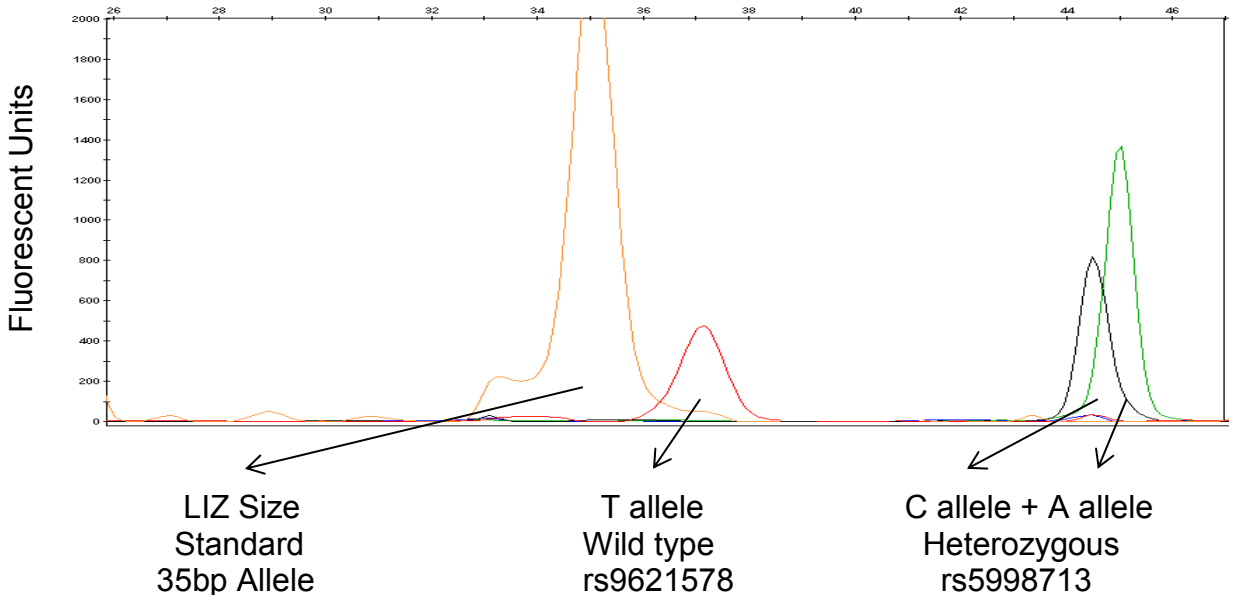


Figure 3.12: A SNaPshot reaction electropherogram of Multiplex group three for a sample that is heterozygous for the rs5998713 SNP and wild type for the rs9621578 SNP. Size standard and colours of peaks as described in Figure 3.7

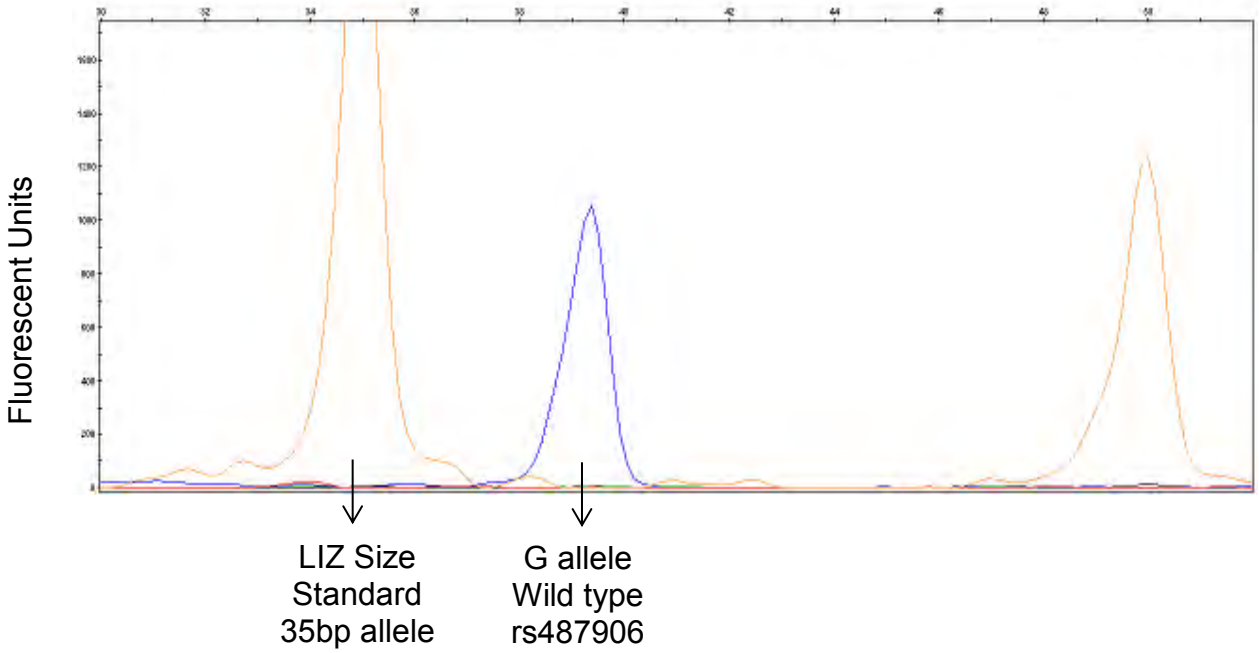


Figure 3.13: A SNaPshot reaction electropherogram for a sample that shows the wild type allele for the SNP rs487906. Size standard and colours of peaks as described in Figure 3.7

3.5.1 Identification of protective SNP alleles in candidate AMD genes

For this, the second part of the research, and to determine whether SNPs in the candidate AMD genes were protective against AMD in the African population, there had to be a significant difference (decrease) in the minor allele frequency (MAF) of the indigenous African AMD cohort versus the African controls. The online software analysis tool, SHEsis (Yong 2005) (http://analysis.bio-x.cn/myAnalysis.php), was used for analysis.

Single site analysis was used and input data for both cases and controls were entered in the required format. SNPs that showed a decreased MAF in the AMD cases compared to controls in Africans were considered as possibly protective SNPs. The Pearson's p value ($< 5 \times 10^{-2}$) within the output data was used to determine significance of these possibly protective SNPs, instead of the Fischer's p value, due to sample size.

3.5.2 Detection of rs9621622 in *TIMP3*

The heterozygous rs9621622 SNP in *TIMP3* was detected only once in 254 samples (Table 3.7) and verified by cycle sequencing (Figure 3.14). The difference in minor allele frequency between the Caucasian AMD and Caucasian Control cohort respectively (Table 3.7), is not significant with a p value of 5.56×10^{-1}. The difference in minor allele frequency between the African AMD and African Control cohort respectively (Table 3.7), is significant with a p value of 9.95×10^{-4}. The wild type allele i.e. the absence of the rs9621622 SNP, in a sample was represented on the electropherogram by a black peak, indicative of a C allele (Figure 3.7) (Figure 3.8). The presence of a heterozygous rs9621622 SNP in a sample was represented on the electropherogram by a black and red peak, indicative of a C and T allele, respectively. A homozygous rs9621622 SNP was represented by a single red peak, indicative of a T allele.

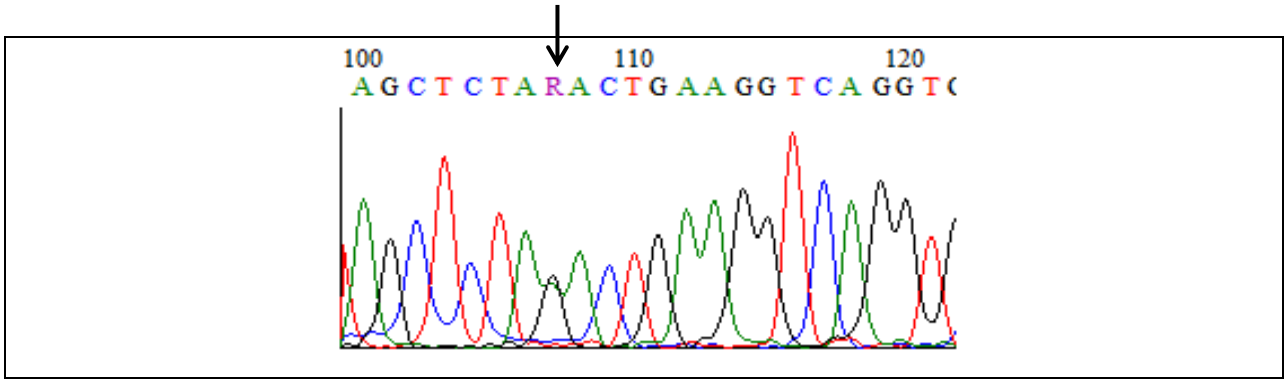


Figure 3.14: An electropherogram representing a segment of sequencing from the reverse strand showing the rs9621622 SNP within *TIMP3.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The R base called at position 107 (indicated with an arrow) represents the heterozygous rs9621622 SNP which occurs when both an A and a G base have equal fluorescence for that locus.

3.5.3 Detection of rs9621578 in *TIMP3*

The heterozygous rs9621578 SNP was detected in eight out of 254 samples (Table 3.7) and verified by cycle sequencing (Figure 3.15). The difference in minor allele frequency between the Caucasian AMD and Caucasian Control cohort respectively (Table 3.7), is not significant with a p value of 1.84×10^{-1}. The difference in minor allele frequency between the African AMD and African Control cohort respectively (Table 3.7), is not significant with a p value of 2.85×10^{-1}. The wild type allele i.e. the absence of the rs9621578 SNP, in a sample was represented on the electropherogram by a red peak, indicative of a T allele (Figure 3.11) (Figure 3.12). The presence of a heterozygous rs9621578 SNP in a sample was represented on the electropherogram by a red and black peak, indicative of a T and C allele, respectively. A homozygous rs9621578 SNP was represented by a single black peak, indicative of a C allele.

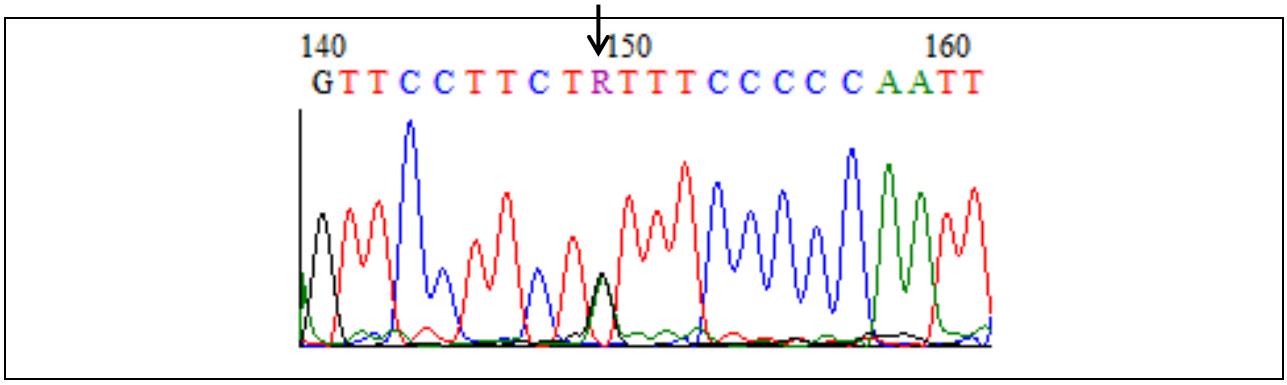


Figure 3.15 An electropherogram representing a segment of sequencing from the reverse strand showing the rs9621578 SNP within *TIMP3.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The R base called at position 149 (indicated with an arrow) represents the heterozygous rs9621578 SNP which occurs when both an A and a G base have equal fluorescence for that locus.

3.5.4 Detection of rs5998713 in *TIMP3*

The heterozygous rs5998713 SNP was detected in 22 out of 254 samples and in the homozygous state once (Table 3.7). Each separate genotype was confirmed by cycle sequencing (Figure 3.16) (Figure 3.17). The difference in minor allele frequency between the Caucasian AMD and Caucasian Control cohort respectively (Table 3.7), is not significant with a p value of 4.29×10^{-1}. The difference in minor allele frequency between the African AMD and African Control cohort respectively (Table 3.7), is not significant with a p value of 3.16×10^{-1}. The wild type allele i.e. the absence of the rs5998713 SNP, in a sample was represented on the electropherogram by a black peak, indicative of a C allele (Figure 3.11). The presence of a heterozygous rs5998713 SNP in a sample was represented on the electropherogram by a black and green peak, indicative of a C and an A allele, respectively (Figure 3.12). A homozygous rs5998713 SNP was represented by a single green peak, indicative of an A allele.

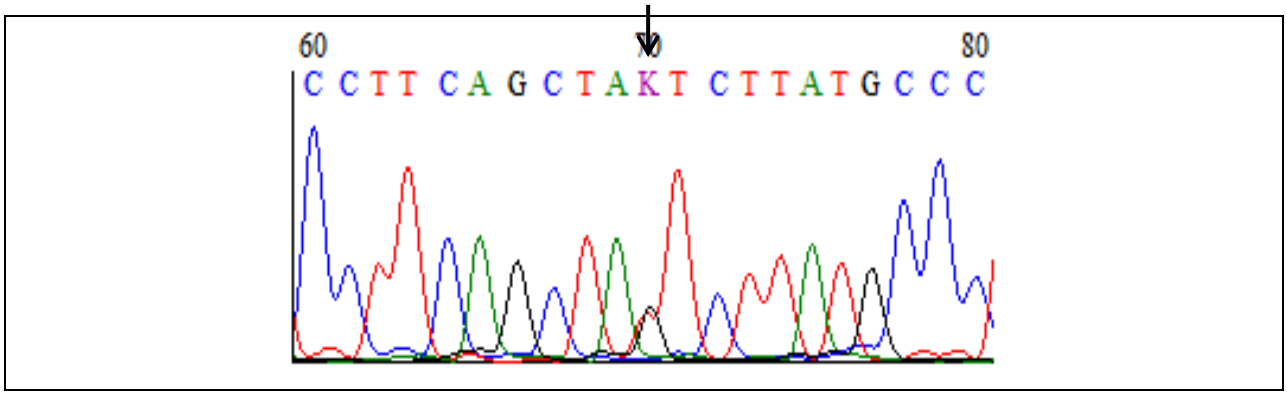


Figure 3.16 An electropherogram representing a segment of sequencing from the reverse strand showing the rs5998713 SNP within *TIMP3.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The K base called at position 70 (indicated with an arrow) represents the heterozygous rs5998713 SNP which occurs when both an A and a C base have equal fluorescence for that locus.

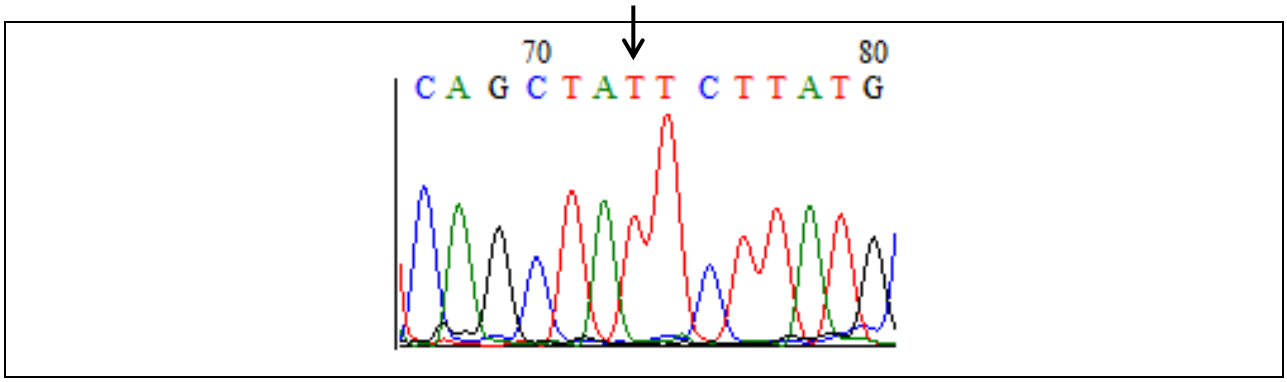


Figure 3.17 An electropherogram representing a segment of sequencing from the reverse strand showing the rs5998713 SNP within *TIMP3.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The T base called at position 73 (indicated with an arrow) represents the heterozygous rs5998713 SNP.

3.5.5 Detection of rs1555494 in *HMCN1*

The heterozygous rs1555494 SNP was detected in 16 out of 254 samples and in the homozygous state four times (Table 3.7). Each separate genotype was verified by cycle sequencing (Figure 3.18) (Figure 3.19). The difference in minor allele frequency

between the Caucasian AMD and Caucasian Control cohort respectively (Table 3.7), is not significant with a p value of 1.62 x 10^{-1}. The difference in minor allele frequency between the African AMD and African Control cohort respectively (Table 3.7), approaches significance with a p value of 6.84 x 10^{-2}. The wild type allele i.e. the absence of the rs1555494 SNP, in a sample was represented on the electropherogram by a green peak, indicative of an A allele (Figure 3.7). The presence of heterozygous rs1555494 SNP in a sample was represented on the electropherogram by a green and blue peak, indicative of an A and G allele, respectively. A homozygous rs1555494 SNP was represented by a single blue peak, indicative of a G allele (Figure 3.8).

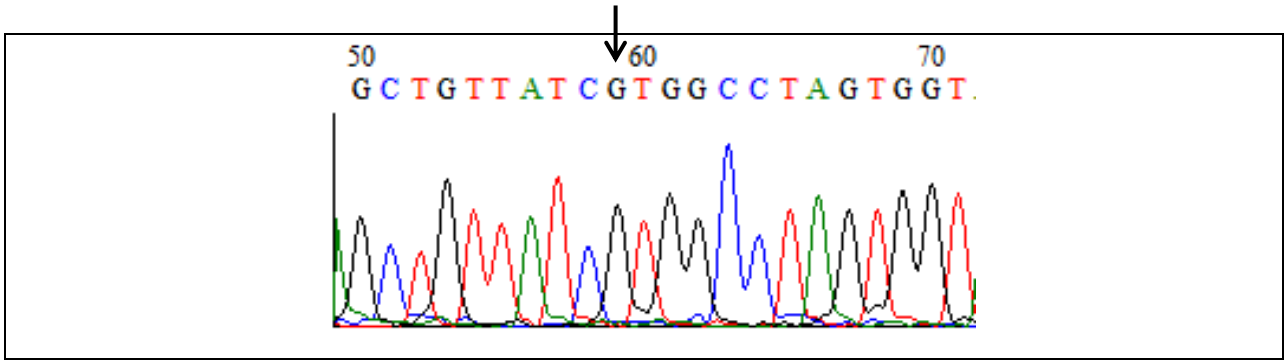


Figure 3.18: An electropherogram representing a segment of sequencing from the reverse strand showing the rs1555494 SNP within *HMCN1.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The G base called at position 59 (indicated with an arrow) represents the homozygous rs1555494 SNP.

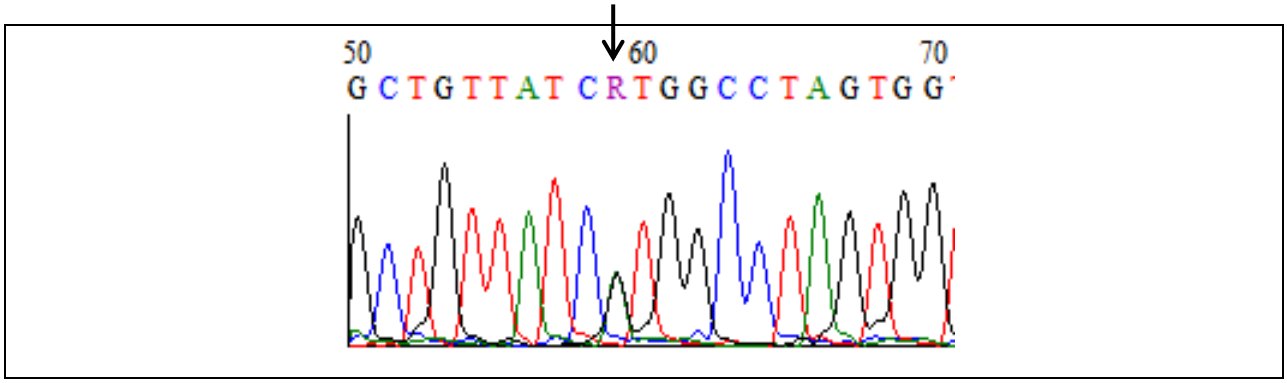


Figure 3.19: An electropherogram representing a segment of sequencing from the reverse strand showing the rs1555494 SNP within *HMCN1.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The R base called at position 59 (indicated with an arrow) represents the heterozygous rs1555494 SNP which occurs when both an A and a G base have equal fluorescence for that locus.

3.5.6 Detection of rs1407428 in *HMNC1*

The heterozygous rs1407428 SNP was detected in three out of 254 samples and in the homozygous state three times (Table 3.7). Each separate genotype was confirmed by cycle sequencing (Figure 3.20) (Figure 3.21).The difference in minor allele frequency between the Caucasian AMD and Caucasian Control cohort respectively (Table 3.7) is not significant with a p value of 3.12×10^{-1}. The difference in minor allele frequency between the African AMD and African Control cohort respectively (Table 3.7), is not significant with a p value of 6.66×10^{-1}. The wild type allele i.e. the absence of the rs1407428 SNP, in a sample was represented on the electropherogram by a black peak, indicative of a C allele (Figure 3.9). The presence of a heterozygous rs1407428 SNP in a sample was represented on the electropherogram by a black and green peak, indicative of a C and an A allele, respectively. A homozygous rs1407428 SNP was represented by a single green peak, indicative of an A allele (Figure 3.10).

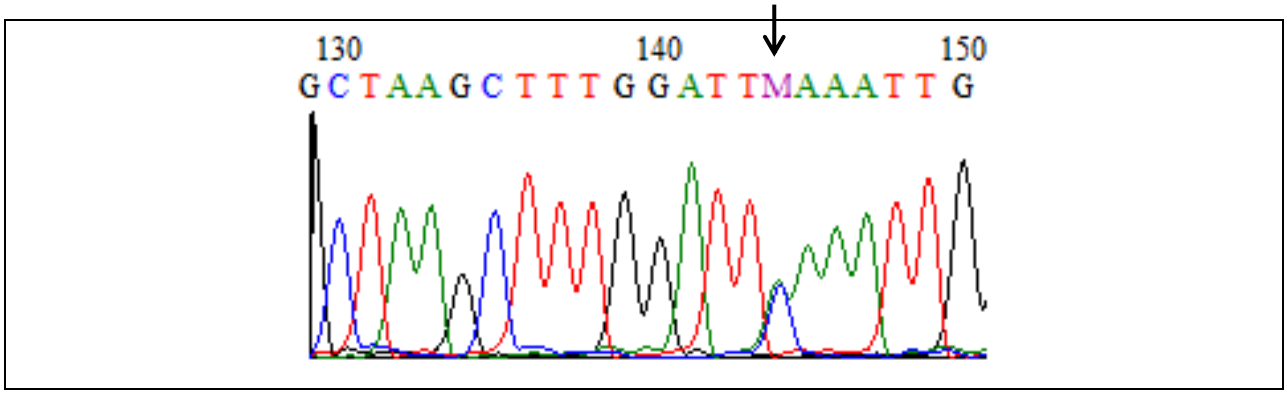


Figure 3.20 An electropherogram representing a segment of sequencing from the forward strand showing the rs1407428 SNP within *HMNC1.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The M base called at position 144 (indicated with an arrow) represents the heterozygous rs17110878 SNP which occurs when both an A and a C base have equal fluorescence for that locus.

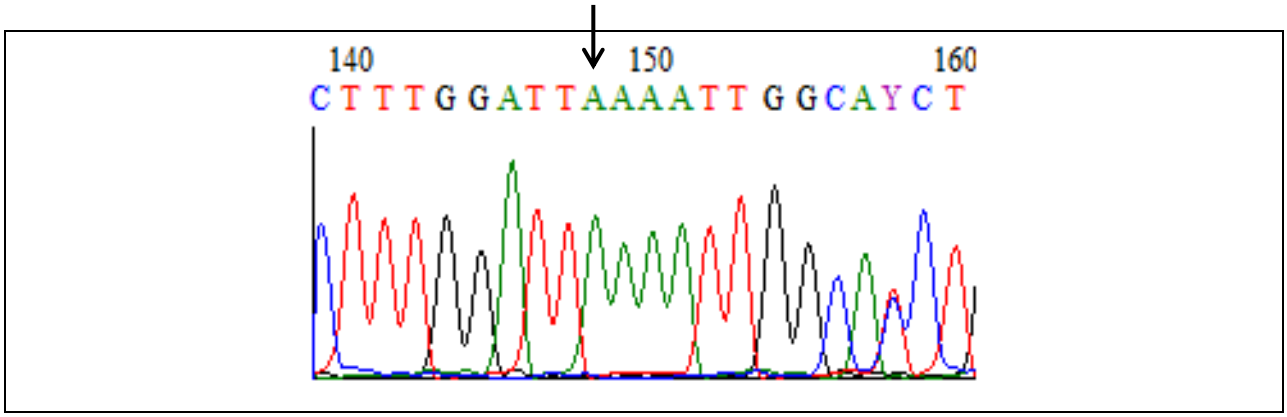


Figure 3.21 An electropherogram representing a segment of sequencing from the forward strand showing the rs1407428 SNP within *HMNC1.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The A base called at position 148 (indicated with an arrow) represents the heterozygous rs17110878 SNP.

3.5.7 Detection of rs6425006 in *HMNC1*

The heterozygous rs6425006 SNP was detected in seven out of 254 samples (Table 3.7) and verified by cycle sequencing (Figure 3.22). The difference in minor allele frequency between the Caucasian AMD and Caucasian Control cohort respectively (Table 3.7), is not significant with a p value of 3.12×10^{-1}. The difference in minor allele

frequency between the African AMD and African Control cohort respectively (Table 3.7), is not significant with a p value of 3.91 x 10^{-1}. The wild type allele i.e. the absence of the rs6425006 SNP, in a sample was represented on the electropherogram by a red peak, indicative of a T allele (Figure 3.9). The presence of a heterozygous rs6425006 SNP in a sample was represented on the electropherogram by a red and black peak, indicative of a T and C allele, respectively (Figure 3.10). A homozygous rs6425006 SNP was represented by a single black peak, indicative of a C allele.

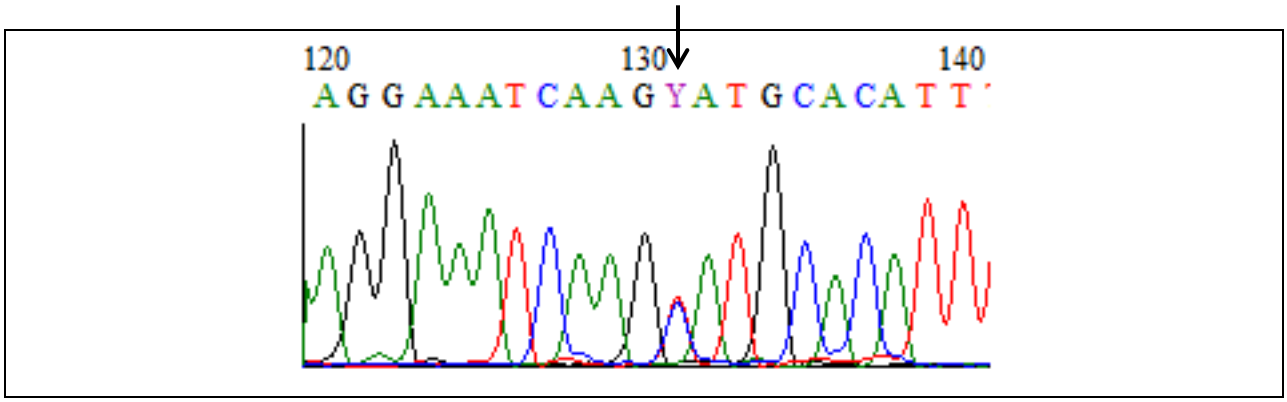


Figure 3.22 An electropherogram representing a segment of sequencing from the forward strand showing the rs6425006 SNP within *HMNC1.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The Y base called at position 131 (indicated with an arrow) represents the heterozygous rs6425006 SNP which occurs when both a T and a C base have equal fluorescence for that locus.

3.5.8 Detection of rs17110714 in *ABCA4*

The heterozygous rs17110714 SNP was detected in nine out of 254 samples (Table 3.7) and verified by cycle sequencing (Figure 3.23). The difference in minor allele frequency between the Caucasian AMD and Caucasian Control cohort respectively (Table 3.7), is not significant with a p value of 2.00 x 10^{-1}. The difference in minor allele frequency between the African AMD and African Control cohort respectively (Table 3.7), is significant with a p value of 1.04 x 10^{-2}. The wild type allele i.e. the absence of the rs17110714 SNP, in a sample was represented on the electropherogram by a red peak, indicative of a T allele (Figure 3.7). The presence of a heterozygous rs17110714 SNP in

a sample was represented on the electropherogram by a red and black peak, indicative of a T and C allele, respectively (Figure 3.8). A homozygous rs17110714 SNP was represented by a single black peak, indicative of a C allele.

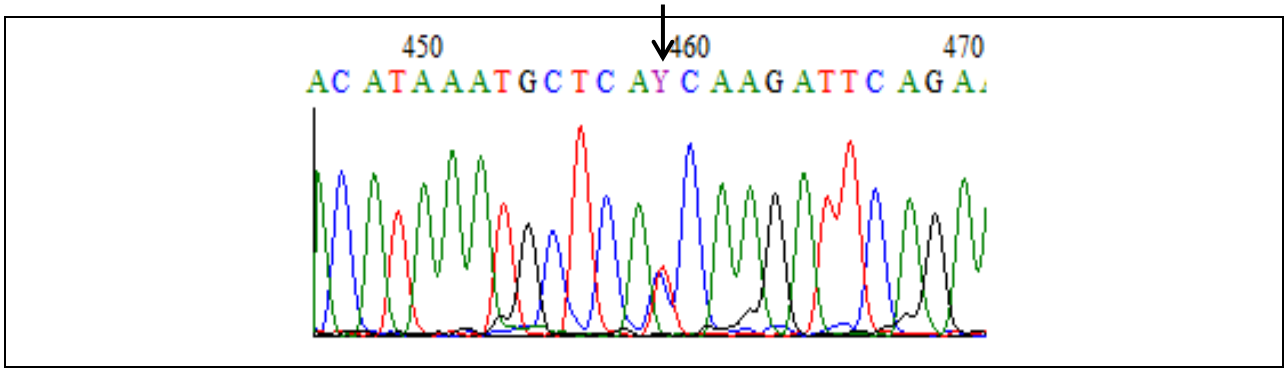


Figure 3.23: An electropherogram representing a segment of sequencing from the reverse strand showing the rs17110714 SNP within *ABCA4.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The Y base called at position 459 (indicated with an arrow) represents the heterozygous rs17110714 SNP which occurs when both a C and a T base have equal fluorescence for that locus.

3.5.9 Detection of rs6703052 SNP in *ABCA4*

The heterozygous rs6703052 SNP was detected in 26 out of 254 samples and in the homozygous state twice (Table 3.7). Each separate genotype was verified by cycle sequencing (Figure 3.24) (Figure 3.25). The difference in minor allele frequency between the Caucasian AMD and Caucasian Control cohort respectively (Table 3.7), is not significant with a p value of 1.63×10^{-1}. The difference in minor allele frequency between the African AMD and African Control cohort respectively (Table 3.7), is not significant with a p value of 5.01×10^{-1}. The wild type allele i.e. the absence of the rs6703052 SNP, in a sample was represented on the electropherogram by a black peak, indicative of a C allele (Figure 3.7). The presence of a heterozygous rs6703052 SNP in a sample was represented on the electropherogram by a black and green peak, indicative of a C and an A allele, respectively. A homozygous rs6703052 SNP was represented by a single green peak, indicative of an A allele (Figure 3.8).

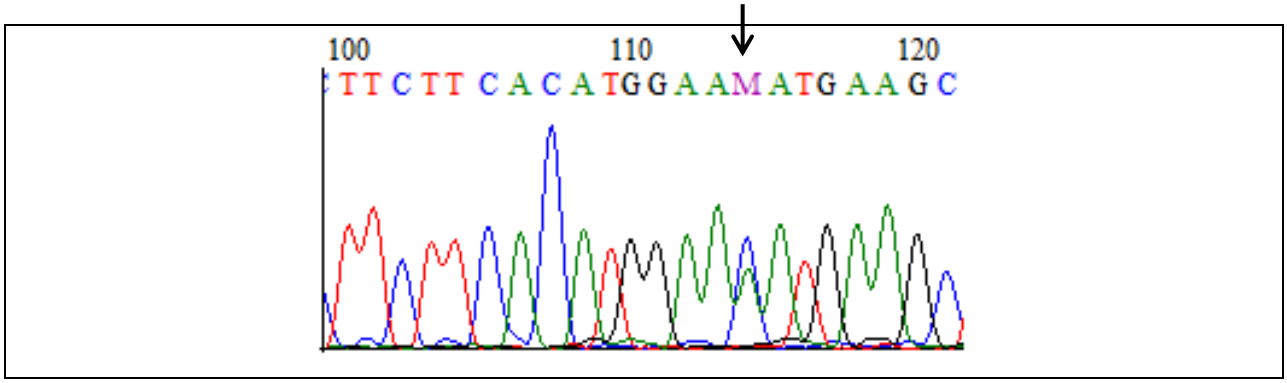


Figure 3.24: An electropherogram representing a segment of sequencing from the reverse strand showing the rs6703052 SNP within *ABCA4.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The M base called at position 114 (indicated with an arrow) represents the heterozygous rs6703052 SNP which occurs when both an A and a C base have equal fluorescence for that locus.

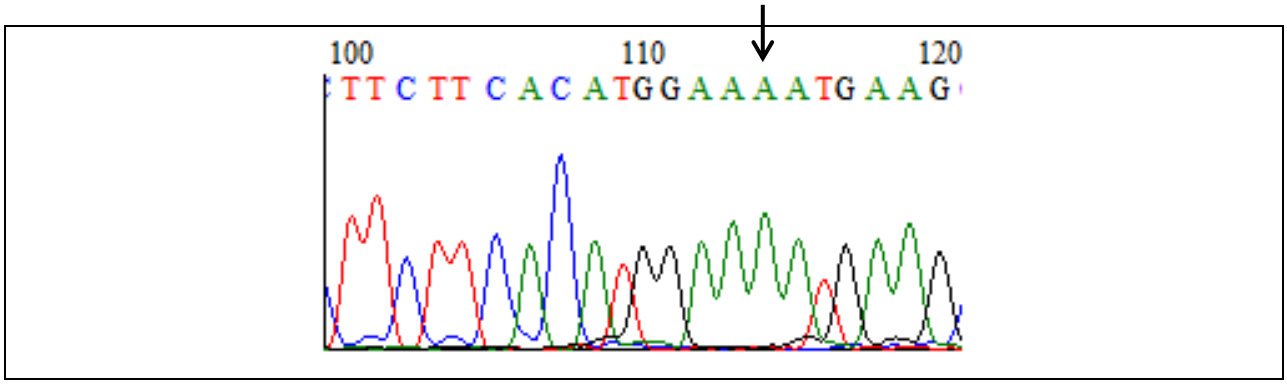


Figure 3.25: An electropherogram representing a segment of sequencing from the reverse strand showing the rs6703052 SNP within *ABCA4.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The A base called at position 114 (indicated with an arrow) represents the homozygous rs6703052 SNP.

3.5.10 Detection of rs17110878 in *ABCA4*

The heterozygous rs17110878 SNP was detected in five out of 254 samples, and once in the homozygous state (Table 3.7). Each separate genotype was confirmed by cycle sequencing (Figure 3.26) (Figure 3.27). The difference in minor allele frequency

between the Caucasian AMD and Caucasian Control cohort respectively (Table 3.7), is not applicable since no minor alleles were genotyped in either cohort. The difference in minor allele frequency between the African AMD and African Control cohort respectively (Table 3.7), is not significant with a p value of 3.98×10^{-1}. The wild type allele i.e. the absence of the rs17110878 SNP, in a sample was represented on the electropherogram by a green peak, indicative of an A allele (Figure 3.9). The presence of a heterozygous rs17110878 SNP in a sample was represented on the electropherogram by a green and blue peak, indicative of an A and a G allele, respectively (Figure 3.10). A homozygous rs17110878 SNP was represented by a single blue peak, indicative of a G allele.

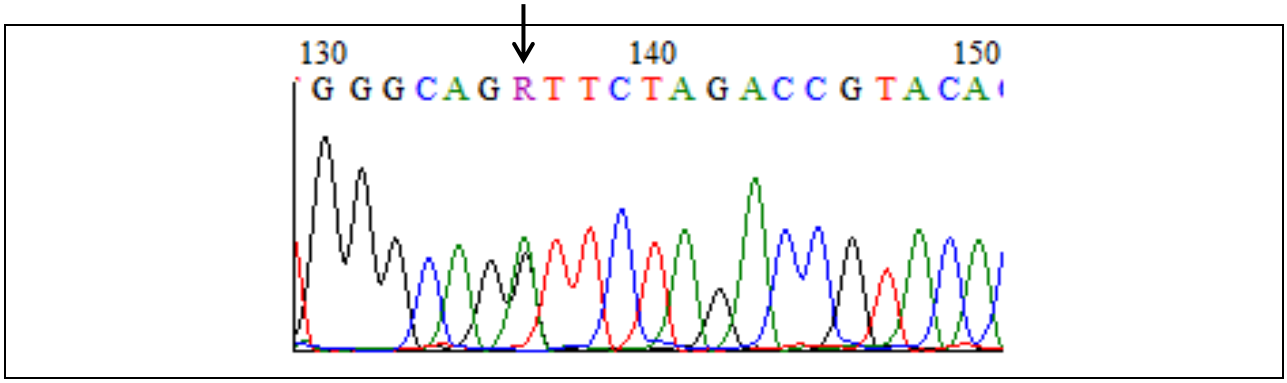


Figure 3.26: An electropherogram representing a segment of sequencing from the forward strand showing the rs17110878 SNP within *ABCA4.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The R base called at position 136 (indicated with an arrow) represents the heterozygous rs17110878 SNP which occurs when both an A and a G base have equal fluorescence for that locus.

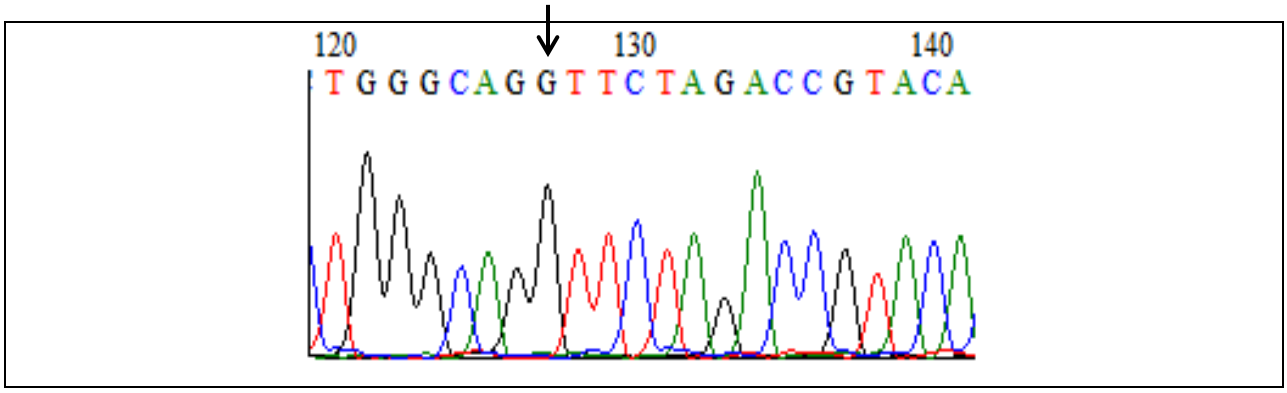


Figure 3.27: An electropherogram representing a segment of sequencing from the forward strand showing the rs17110878 SNP within *ABCA4.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The G base called at position 127 (indicated with an arrow) represents the homozygous rs17110878 SNP.

3.5.11 Detection of rs487906 in *ABCA4*

The heterozygous rs487906 SNP was detected in 92 out of 254 samples and in the homozygous state once.

The wild type allele i.e. the absence of the rs487906 SNP, in a sample was represented on the electropherogram by a blue peak, indicative of a G allele (Figure 3.13). The presence of a heterozygous rs487906 SNP in a sample was represented on the electropherogram by a blue and black peak, indicative of a G and C allele, respectively. A homozygous rs487906 SNP was represented by a single black peak, indicative of a C allele.

However, of the six samples, covering all three possible genotypes, were sequenced in order to verify the genotyping results. Of the six samples, only one showed corresponding genotype results for cycle sequencing and SNaPshot reactions. SNaPshot genotyping for the remaining conflicting samples showed a heterozygous (C/G) genotype while cycle sequencing showed a homozygous Guanine genotype on the reverse strand (Figure 3.28). The forward strand was not used for sequence

analysis due to the intronic variant, rs375387843, situated upstream from rs487906 SNP producing a deletion to frameshift the sequence and rendering it uninterpretable.

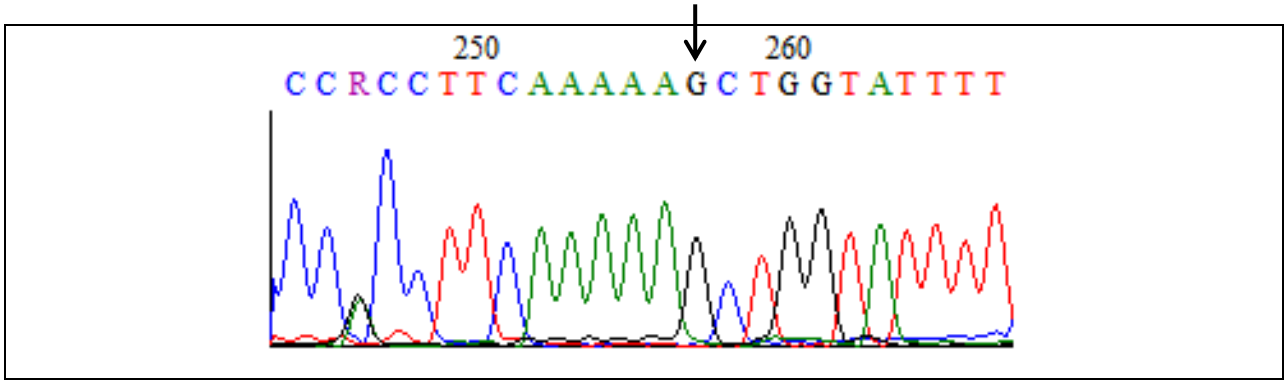


Figure 3.28 An electropherogram representing a segment of sequencing from the reverse strand showing the rs487906 SNP within *ABCA4.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The G base called at position 257 (indicated with an arrow) represents the homozygous rs487906 SNP.

Ideally, an internal primer from the forward strand could be designed to resolve the incongruous genotype results but a SNP situated 11 bp upstream from the rs487906 SNP would impede hybridization of the internal primer to the template. Possible reasons for the genotype discrepancy include internal primer design and poor DNA quality. Reasons for incorrect nucleotide extension with regard to internal primers include binding to other sequences in the amplified segment or hairpin and primer dimer formation enabling hybridization to an adjacent sequence. Degraded DNA could also explain the ambiguous results although it seems unlikely due to the same samples demonstrating accurate genotyping with the other SNPs in candidate AMD genes (Hansen *et al*. 2007). All SNaPshot genotype results were of satisfactory peak size with no or minimal background noise hence eliminating other causes of extraneous peaks such as incomplete removal of PCR primers and ddNTPs. Due to the genotype discrepancy, the rs487906 SNP genotype results were not incorporated into the final data analysis.

Further analysis of the sequencing results showed that four of the six samples contained two intronic variants 12 bp from the forward external primer. The first variant, rs375387843, caused a C/- allele feature truncation while the second variant, rs71588510, produced a CG/TA allele substitution (Figure 3.29).

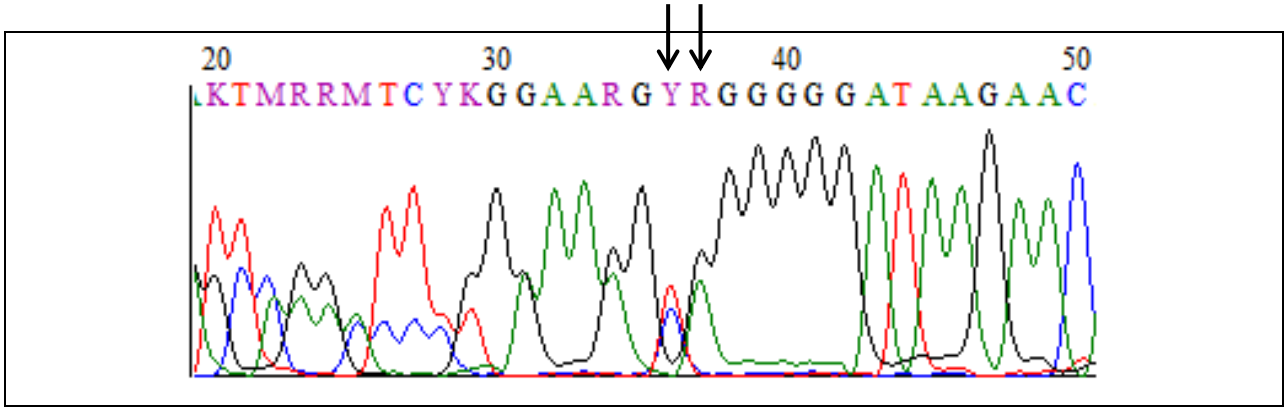


Figure 3.29 An electropherogram representing a segment of sequencing from the reverse strand showing the rs375387843 and rs71588510 SNPs within *ABCA4.* The actual sequence called is indicated above the peaks with each colour representing a different nucleotide. The Y and R base called at position 36 and 37 respectively (indicated with an arrow) represents the heterozygous CG/TA allele substitution. The heterozygous rs375387843 SNP at position 35 represents the heterozygous -/G allele substitution which is identified by multiple heterozygous peaks followed by normal sequence as one allele is shifted by the deletion of a G nucleotide.

Webcutter version 2.0 (Maarek *et al* 1997) (http://rna.lundberg.gu.se/cutter2) was used to determine the suitability of RFLP as a genotyping technique to potentially genotype the AMD cohort for the two intronic variants, rs375387843 and rs71588510. The combination of the two intronic variants altered two RE recognition sites but resulted in an inadequate fragment size difference to be accurately discerned by agarose gel electrophoresis.

Chapter 4: Discussion

In this study, macular degeneration, in its 'juvenile' and 'elderly' form, namely STGD and AMD, was investigated among the indigenous southern African population respectively. In the first part of the study, following a SNP-based phased haplotyping of a cohort of indigenous South African subjects, the entire coding region of *ABCA4* was screened in a small STGD cohort in pursuance of novel mutations to elucidate the missing heritability within this group. In the second part of this study, candidate AMD genes, in whom there were significant SNP allelic frequency differences between Caucasian and indigenous southern Africans were investigated as potential protective factors within an ethnically diverse AMD cohort to explain the considerably lower prevalence of AMD among indigenous southern Africans. Sections 4.1 to 4.1.5 pertain to the discussion concerning the first part of the study, while sections 4.2 to 4.2.4.1 relate to the second part of this study.

4.1 Novel *ABCA4* mutations in indigenous Africans with STGD

Vast numbers of mutations have been identified within *ABCA4* since its recognition as the causative gene for STGD (Allikmets *et al*. 1997a). Commercial microarrays, like the ABCR400, were developed to screen populations for numerous mutations simultaneously (Jaakson *et al*. 2003). Microarray chip screening in the past proved successful within the South African Caucasian population by detecting the causative mutations in 58.4 % of subjects. Seven common founder mutations emerged in the Afrikaner sub-cohort from this screening, resulting in a new diagnostic SNaPshot assay, the 'Quick 7' (Roberts *et al*. 2012), which is currently offered by the National Health Laboratory Services (NHLS) in South Africa. Subsequent screening of SA indigenous African STGD patients with the ABCR 400 microarray was disappointing, as only 11% had their causative mutations (homozygous or compound heterozygous) identified. To ascertain whether novel mutations were present within the black southern African STGD patient group, PHASE (Stephens *et al*. 2001) analysis of 17 SNPs across *ABCA4*

(which was available through these samples having been screened using the ABCR400 array) indicated three common haplotypes within the African STGD group. The aim of this study was to discover novel mutations within the African STGD cohort by exploratory sequencing of the entire coding region of *ABCA4* in three STGD patients representing each of the three common haplotypes.

Four previously-reported missense mutations were detected during sequencing of *ABCA4* among the three African STGD patients. *ABCA4* currently has 761 reported mutations of which 517 are missense/nonsense mutations as indicated in the Human Genetic Mutation Database (HGMD) (www.hgmd.org). Interestingly, the NCBI variation viewer (www.ncbi.nlm.nih.gov/variation/view/) asserts that only 18 of these 517 are definitely pathogenic. Determination of pathogenicity in the present study involved multiple processes ranging from database and literature reviews to analysis through prediction software, namely PON-P (Olatubosun et al 2015) (http://bioinf.uta.fi/PON-P/index.shtml), PON-P2 (http://structure.bmc.lu.se/PON-P2/) (Lund University, 2013) and Human Splice Finder (Desmet *et al* 2009) (13 October, 2010) (http://www.umd.be/HSF/#). Deciding upon pathogenicity is a complex process and involves consideration of multiple factors.

The R226L mutation in *ABCA4*, although not novel, was listed among the missense variants within the HGMD database. This mutation was originally reported in Russian and Moldovian Hereditary Nonpolyposis Colon Cancer (HNPCC) families (Maliaka *et al*. 1996). Both PON-P (Olatubosun et al 2015) and PON-P2 (Niroula et al 2015) classified its pathogenicity as ‘unknown’ and this mutation has not previously been reported in SA. No minor allele frequency data for the African populations was reported for this variant in the 1000 Genomes (McVean *et al* 2012) database. R226L did not track with any of the haplotypes and was only detected in one test sample, namely RPS1230.1. In view of the above factors R226L was considered unlikely to be pathogenic within this particular African STGD cohort.

The V643M mutation (Briggs *et al*. 2001) was the only mutation still classified as pathogenic after re-analysis with PON-P2 (Niroula et al 2015). The wild type allele frequency is 99 % in African populations, screened in the 1000 Genomes project (representing Yoruba, Luhya and Americans of African descent), and the mutant allele was identified only in patient RPM 1075.1 after screening the whole Haplotype cohort. The mutation is listed in HGMD and associated with the STGD phenotype. Although classified as pathogenic by PON-P2, V643M was deemed benign within an African American STGD cohort based on similar allele frequencies among cases and controls in a population of African American descent (Zernant *et al*. 2014). Of all four missense variants identified in this study, V643M is the most likely pathogenic mutation as supported by software prediction and it is the only mutation of the four predicted to be pathogenic by Stone, using a substitution matrix called Blosum62 (Stone 2003).

The L1201R mutation was first described when sequencing 150 recessive STGD families of northern and central European descent (Lewis *et al*. 1999). The L1201R mutation is listed similarly to V643M in HGMD, but differs in that its predictive pathogenicity was altered after re-analysis with PON-P2 (Niroula et al 2015). Ten percent of the African populations (1000 Genomes) screened, display the L1201R mutant allele which is remarkably similar to the allele frequency of 12.5 % within our Haplotype cohort. L1201R also displayed a high minor allele frequency (7.5 %) within an African American control population of 300 and identified as a non-pathogenic variant within 10 unrelated African American STGD patients within the same study (Utz *et al*. 2013a). Similarly, L1201R was deemed benign due to a high allele frequency in the general African American population when screened for in a cohort of 44 African American STGD patients (Zernant *et al*. 2014). Taking into account this high frequency within the African population, and considering that five to 12 percent of the general population screened positive for at least one *ABCA4* mutation (Jaakson *et al*. 2003; Rivera *et al*. 2000; Riveiro-Alvarez *et al*. 2009) makes it an unlikely pathogenic mutation within African populations.

This notion is supported by software prediction, but should be confirmed by performing family co-segregation analysis. Co-segregation analysis forms an essential part in pathogenicity confirmation, especially in cases where a frequent disease-associated allele also shows a higher than expected allele frequency within a control population. This was shown with the R2107H mutation in an African American STGD cohort (Zernant *et al*. 2014). Co-segregation analysis was not possible due to lack of DNA from additional family members of any of the three Bs at the time of this study. This should be regarded as a limitation to pathogenicity prediction in this study.

The final missense variant, S2255I, was previously reported among a European cohort of cone-rod dystrophy patients and is listed under that same phenotype in HGMD (Maugeri *et al*. 2000). A 'caution' to its classification as a pathogenic STGD mutation is the observation that the wild-type and mutant alleles occur with equal frequency within the African populations (1000Genomes). This observation is supported by the fact that two of the three test samples contain this variation, in a heterozygous state.

Therefore, after incorporating all of the above facts, viewed within the context of a limited STGD cohort, V643M is classified as the only pathogenic missense mutation detected among the three test samples within this study. Further conclusive deductions regarding the pathogenicity of the other three missense mutations could only be made if functional protein and segregation studies are performed. As missense variants constitute the largest proportion of all mutations in *ABCA4*, their pathogenic prediction is important, especially as we detected no obvious disease-causing mutations such as deletions, insertions or nonsense mutations within the exons.

4.1.1 Pathogenic predictor software accuracy

With the change in predictive class of the L1201R mutation, questions arose regarding the accuracy of pathogenic predictor software. Missense predictor software differs from other predictor tools in that they focus on predicting protein structural changes. These changes include the structure, stability and folding of the altered protein product. One of

the obstacles hampering predictor accuracy is that all missense predictor tools use different training datasets and machine learning methods, leading to bias. PON-P (Olatubosun et al 2015) overcomes this by incorporating many predictors whether they are structural or sequence-based. This observation is expected in view of evolutionary forces actively selecting against deleterious variants. Thusberg *et al*. (2011) compared all missense predictor tools against large mutation databases for outcomes such as accuracy, sensitivity and specificity with most predictor tools scoring in excess of 70 % in accuracy. Although missense predictor tools are becoming more reliable and accurate (Galehdari *et al*. 2013), they still lack the ability to account for the pathogenicity of mutations affecting cellular pathways such as transcriptional regulation, micro-ribonucleic acid (mRNA) splicing and exon silencing.

Determination of pathogenicity of the four missense variants identified in this study was mainly based on the outcome of PON-P2 prediction when all other factors were considered. In summary, only the V643M mutation was deemed to be disease-causing. More novel and known missense variants were expected especially considering the allelic heterogeneity of *ABCA4* and that this particular (indigenous African) ethnic group has not been investigated previously. The fact that only three samples were sequenced could have contributed to the low number of missense variants identified. Nonetheless, the fact that missense mutations were detected in each sample is encouraging.

4.1.2 Missing heritability in the STGD cohort

No mutational screening process is complete without Copy Number Variants (CNV) screening. CNVs are structural alterations in the human genome which include deletions, duplications, translocations and insertions of chromosomal regions ranging from one kilobase to several megabases in size. The PCR-based methods and sequencing used in this study are unable to detect CNV mutations, hence the need for further screening with techniques such as multiplex ligation-dependent probe amplification (MLPA). Bauwens *et al*. (2013) reported a single deletion among 48 patients with incomplete molecular diagnosis using a combination of qualitative PCR

(qPCR) and massive parallel sequencing, while Yatsenko *et al*. (2003) detected deletions in two out of 308 STGD subjects. In a study of 44 STGD patients using aCGH, no large CNVs were found in the *ABCA4* locus (Zernant *et al*. 2014). Although these studies indicate the rarity of CNVs as the causative mutation in most STGD cohorts, it must still be considered as a possible cause of STGD among our African cohort. Testing the whole Haplotype cohort for CNV mutations would be much more effective in comparison to screening just the three test samples, considering the rarity of CNVs in STGD in the literature to date.

An additional possibility for the lack of definitive mutations identified is that the African STGD cohort may contain causative mutations in a gene or gene(s) other than *ABCA4*. Autosomal dominant Stargardt-like macular dystrophy, known also as STGD type 3, has similar clinical features to the other forms of STGD but segregates in an autosomal dominant fashion (Stone *et al*. 1994). The causative mutation was mapped to *ELOVL4,* which is involved in the synthesis of very long chain saturated and polyunsaturated fatty acids, and located on chromosome 6q14. The role of *ELOVL4* was confirmed by subsequent studies in other families showing a dominant inheritance pattern for inherited macular degeneration (Griesinger *et al*. 2000; Zhang *et al*. 2001). STGD type 4 also shows a similar phenotype to STGD type 1. Yang *et al*. (2008) showed the causative mutation to be a missense mutation in *PROM1* after an autosomal dominant STGD phenotype was identified within a Caribbean family (Kniazeva *et al*. 1999).

Boon *et al*. (2007) reported a causative mutation within the peripherin 2 (*RDS)* gene for one-fifth of his patients with assumed STGD, after screening of *ABCA4. RDS* produces a protein that is essential for outer segment disk morphogenesis and leads to a slower macular degeneration than STGD. These studies illustrate the complexity in studying inherited macular degeneration disorders – not only are there major overlapping phenotypic features, but also genotypic heterogeneity among seemingly similar phenotypes. Extensive clinical diagnostic workup is therefore essential, especially when diagnosing any early onset macular degeneration.

Braun *et al*. (2013) showed that mutations can be detected within non-exomic and splicing regions by using RNA sequencing and functional analysis for individuals where only one mutant *ABCA4* allele had been identified. Although it is possible that the missing mutant alleles for our cohort are within the intronic region of *ABCA4*, the lack of additional familial samples with which to perform segregation analysis, and the large number of possible pathogenic intronic variants as indicated by HSF, complicates the choice of variant for subsequent functional analysis. In view of these challenges, it was prudent to only sequence exons and splice regions across the cohort compared to deep intronic sequencing. Since it has been shown that intronic variants up to 1Mb from the exon can be disease causing, it remains possible but improbable that the intronic regions of *ABCA4* harbour the missing mutant alleles for this cohort.

All intronic and synonymous variants were analysed with HSF to assess pathogenicity, yielding a total of 22 out of a possible 35 variants that were predicted to have a significant effect. Only three of the 22 variants were situated within the acceptor splicing region, of 10 to 15 bp, 5´ of the exon. Surprisingly, intronic variants in this cohort showed significant alterations as deep as 122 bp from the exon. If taken at face value, these results would mean that test sample three has a total of 16 pathogenic intronic and synonymous changes. For ease of interpretation, any significant change within this study will be assumed to be possibly pathogenic and therefore deserving further consideration. This creates a challenging situation as how to differentiate between these 16 possible pathogenic changes as it would be improbable for any allele to contain so many mutated variants. Logically, one would focus on variants occurring within the presumed splicing region, but again no family data is available to test whether these variants co-segregate. Another difficulty arises when attempting to rank pathogenic predicted variants according to the particular element disturbed within the spliceosome, whether it be branching points or ESEs. These factors are not adequately addressed by the designers of the HSF suite of programmes in order for one to confidently consider a variant as definitely pathogenic. This is understandable as ESE analyses are incorporated within HSF output data and these predictions are based upon different matrices and algorithms.

Although ‘ESEfinder’ estimates have been shown to successfully predict aberrant splicing, there are instances when known splicing effects fail to associate with a sequence change that causes a reduction of the prediction score (Anczukow *et al*. 2008). When interpreting results regarding splicing aberrations it is critical for supplementary expression studies, especially of a quantitative nature, as not all aberrant splice products are produced in amounts substantial enough to influence protein function. Whenever further functional analysis or segregation studies are planned pertaining to a certain variant, it is important to use multiple predictor tools when finalizing a decision (Spurdle *et al*. 2008).

Considering all the above mentioned factors it is judicious to use the HSF predictions as a guide towards choosing a variant for further screening within our cohort rather than as a pathogenic predictor tool with diagnostic implications. Deciding which variant to investigate further was based upon multiple factors including variant position, population frequency and previous associations with equal weight given to HSF predictions.

In a resource-finite setting as with the current study, accurate haplotype inference is vital in disease mapping (Rieder *et al*. 1999). Ideally, all 32 African STGD patients with no mutations identified would have been sequenced for novel mutations in *ABCA4*. The high cost and labour intensity of sequencing made computational haplotype reconstruction with PHASE (Stephens *et al*. 2001) an applicable alternative in this setting. PHASE (Stephens *et al*. 2001) uses a Bayesian inference method to assign haplotypes within a cohort consisting of genotyped data with no phase information and the accuracy of this method has been previously investigated by comparing it to observed values from 15 SNPs in *Growth Hormone 1* (*GH1*) promoter region in 154 recruits (Adkins *et al*. 2004) . Correct haplotypes were allocated to individuals nearly 90% of the time and 87% when unambiguous individuals were excluded. Accuracy is reduced with increasing amount of heterozygous SNPs especially more than five, according to the same study. Therefore, although possible, it is unlikely that PHASE (Stephens *et al*. 2001) incorrectly identified the common haplotypes in this African

STGD cohort of 32 patients. Possibly, the failure to discover novel mutations in *ABCA4* could also be attributed to disease-causing mutations tracking along other haplotypes within the cohort not assigned to any of the three sequenced individuals.

Additionally, the possibility of reduced penetrance for certain variants in the African population, i.e. the L1201R mutation could be considered as it occurs at an allele frequency of ten percent in African populations and 12.5 % within the Haplotype cohort in this study. Penetrance refers to the percentage of individuals with a certain genotype that displays the associated phenotype. Penetrance can range from complete to incomplete to even pseudo-incomplete if clinical signs have not yet appeared at the time of examination. Different mechanisms, as discussed below, influence the degree of penetrance within a certain ethnic or population groups (Cooper *et al*. 2013).

4.1.3 Screening of the Haplotype Cohort

Variants were divided among the three common haplotypes, and for Haplotype four and 12, more than half the variants were considered significant changes. Only three variants tracked with Haplotype 29. Selecting a single variant per haplotype to further screen among the cohort was not straightforward due to the same concerns that applied when analyzing the HSF predictions. Owing to the lack of obvious exonic mutations and over-signifying the effects of the splicing and silent variants, it was concluded that no obvious variants correlated with the haplotypes, and thus the haplotype grouping of variants was abandoned. The three most likely pathogenic variants were then selected in order to determine their presence within the Haplotype cohort. The variants selected were L1201R and V643M, and the synonymous variant T1277T. The missense variants were clear choices, for reasons explained above, while the T1277T variant displayed significant changes in HSF, exhibited a low minor allele population frequency and was situated close to the donor splicing site. Ascertaining the frequency of these variants in even such a small cohort still provides insights into the possible pathogenicity of these variants especially as this specific ethnic group has not previously been investigated.

The V643M mutation allele was not present within the Haplotype cohort which was expected considering its one percent population frequency within African control populations extracted from the 1000 Genomes (McVean *et al* 2012) resource. With such a small cohort, no meaningful deductions can be made concerning this mutation within the African STGD population in SA.

The L1201R variant allele frequency in African control and Haplotype cohort groups was 10 and 12.5 % respectively, as mentioned above. This confirms the conjecture that this variant seems to be a normal, benign polymorphism within this population. The same assumption made for the L1201R variant applies to the T1277T variant as the genotype frequencies for African and haplotype groups were three and six percent, respectively. These results consolidate our beliefs regarding each variant's non-pathogenic status.

Ultimately no sample in the haplotype cohort, or the test cohort, could be characterised with two bi-allelic pathogenic *ABCA4* mutations, hinting at other genes underlying the retinal dystrophies manifesting as STGD in this population group. Whole exome sequencing is recommended in order to elucidate the genetic basis of disease.

4.1.4 STGD study limitations

The three test samples were diagnosed using only fundal assessment and 30° fundus photograph, whereas had they received additional fundus autofluorescence (AF) imaging and optical coherence tomography (OCT) they might have been diagnosed differently with a disorder within the STGD differential such as Fundus Albipunctatus, Retinitis Punctate Albescens or Best Disease. This lack of extensive clinical data results in improper gene targeting, as for example in the case of Best disease, one would have to screen the Vitelliform Macular Dystrophy 2 gene for mutations.

The lack of familial data for all three test samples again prevents confirmation of the autosomal recessive pattern expected in these STGD patients, adding to the diagnostic

dilemma and supports the recommendation to screen other genes in order to complete the molecular diagnosis for this cohort.

A noted limitation with the predictor tool, ESEfinder, is that it recognizes disruption or creation of potential ESE-binding sites but gives no information regarding which sites are most likely to be used under normal physiological or pathophysiological conditions. These sites are also poorly defined and influenced by multiple factors such as strength and distance from the splice site (Fairbrother *et al*. 2004).

4.1.5 Future work

Notwithstanding the infrequency of positive CNV mutational screening in STGD cohorts, it is still an integral component of comprehensive screening within undiagnosed cohorts i.e. this current study. Testing the whole Haplotype cohort for CNV mutations would be much more effective in comparison to screening just the three test samples. CNV analysis forms part of the future plans for further elucidating the causal mutations within this cohort. In future it would be recommended to either do exome or full genome sequencing on all 14 haplotype samples. Whole exome sequencing would cover all known MD candidate genes, and could also allow the identification of novel candidate genes. However, a large number of variants is expected to occur in each sample, therefore DNA from multiple affected and unaffected family members is required in order to analyse segregation of variants with the disease. Importantly, however, emphasis also does need to be placed on the desirability of detailed ophthalmological phenotyping of subjects and their relatives who end up being part of the study.

4.2 Protective genetic factors in indigenous Africans with AMD

Past epidemiological studies have categorically shown elderly Caucasian populations to be more susceptible to AMD than their African counterparts (Klein *et al*. 2006; Bressler *et al*. 2008). This difference persists even when both groups are derived from the same geographical region, further suggesting that genetic factors are responsible (Friedman

et al. 1999). AMD is a complex disease resulting from interplay of numerous modifiable and non-modifiable risk factors. To date, none of these risk factors have been able to fully account for the discrepancy in prevalence of AMD between these two ethnic groups (Klein *et al*. 2008).

As a complex disease, AMD is unique in that approximately half of its genetic heritability is accounted for by only a few susceptibility loci (Swaroop *et al*. 2007). Some of these susceptibility loci fail to associate with AMD in replicate studies within other ethnic groups, suggesting that modifier or protective genetic factors play a role within these groups. Mostly, protective factors track exclusively within an ethnic group, like the *ApoE* ε4 allele and SNPs in the complement genes, *CFB* and *CC2*, within Caucasians (Baird *et al*. 2004; Spencer *et al*. 2007). Occasionally, an unlikely factor confers protection across multiple ethnic groups, such as the *CFH* haplotype, linking to a *delCFHR1/CFHR3* mutation (Hagemann *et al*. 2006). As the group exhibiting the lowest prevalence of AMD among all ethnic groups, Africans present the ideal cohort in which to identify candidate protective genetic factors. This study aimed to establish whether protective genetic factors exist within the African population in SA in selected AMD candidate genes.

The AMD cohort consisted of patients of Caucasian, Mixed Ancestry and African ethnicities. A large proportion of this cohort was obtained from the Retinal Registry in the Division of Human Genetics at UCT, previously recruited for the RDD project. The rest of the cohort was recruited in collaboration with Ophthalmologists in the Department of Ophthalmology at Groote Schuur Hospital in Cape Town. A single African AMD patient was recruited during the year and a half long recruitment drive. This serves to illustrate not only the anecdotal low prevalence of AMD in SA Africans, but also the difficulties with AMD research within this ethnic group in SA.

4.2.1 Troubleshooting

Only samples fully genotyped for all ten SNPs in candidate AMD genes were used for further analysis. Singleplex PCR was used to genotype SNPs that were not amplifying within their multiplex groups. Generally, the longer amplicons (>350 bp) within each multiplex group were resistant to amplification even after repeated attempts. This was most likely attributed to DNA degradation that inevitably occurred in samples stored for extended periods of time. Or it could be simply due to the physicochemical principles which favour the preferential and complete amplification of the shorter amplicons. Either way, because of the lack of amplification of the larger amplicons, 18 % of samples were not included in the final data analysis. The resolution of one SNP, rs487906, within multiplex group three proved problematic when attempting amplification, even after satisfactory optimisation. Subsequently, it was amplified using singleplex PCR leaving multiplex group three with only two SNPs.

After screening the entire AMD cohort, SNP genotypes were confirmed with direct cycle sequencing. Not less than five percent of samples for each SNP were cycle sequenced in order to accurately validate genotypes. All displayed genotypes for each SNP were included within the samples selected for sequencing. Of all the SNPs in candidate AMD genes screened, only one showed conflicting cycle sequencing and SNaPshot results, namely rs487906, the same SNP removed from multiplex group three. Owing to this discrepancy, SNP rs487906 results were not included in the final data analysis. Technical reasons for this include either incorrect primer design or DNA quality variation. For the remaining nine SNPs, data analysis was performed using the online software program SHEsis (Yong 2005) to determine whether there was a significant difference in the MAF of the AMD cohort for each SNP.

Another technical factor to consider is the possibility of incorrect SNP genotyping, although SNaPshot provided clear results. This conflict in genotype evaluation has previously been assessed with similar techniques and equipment as was used in this study (Hansen *et al*. 2007). Using large sample sizes they confirmed a genotype discrepancy of between six and 15 % after cycle sequencing using the ABI Prism 3130xl Genetic Analyzer (Applied Biosystems). Again due to financial constraints,

confirmation of SNaPshot genotypes involved sequencing of roughly five to ten percent of the cohort of each SNP assuring inclusion of each possible genotype. Due to the small sample size of the African cohort, incorrectly assessed genotypes could alter the p value sufficiently to alter significance. However, the concordance of the sequencing and genotyping results in this particular study show this was not likely to be the case, other than for SNP rs487906 which was discarded from further data analysis.

4.2.2 Proposed role of protective AMD variants

The first step in data analysis was to compare the MAF of the selected SNPs, between the Caucasian control and the AMD Caucasian cohort, respectively. By demonstrating that no significant differences exist between these MAFs, it consolidated the premise that these SNPs could confer protection within the African population only, by virtue of their overrepresentation in that particular group. The MAFs of all nine SNPs showed no significant difference between Caucasian control and AMD cohorts. Both the Caucasian control and AMD cohorts had similar sample numbers, 100 and 98, respectively. The second step was then to compare the MAF between the African control and the African AMD cohorts respectively. Contrary to the Caucasian cohorts, the African control cohort comprised 70 samples while the African AMD cohort contained only nine samples. Eight of the nine SNPs showed a decreased trend in the MAF in the AMD cohort, but only two of these, rs9621622 and rs17110714, showed a statistically significant decrease in cases, when using the threshold p value of 0.05. Another SNP, rs1555494, located in the intronic region in *HMCN1* also showed a decreased MAF in the AMD cohort with a p value of 0.068. SNP rs9621622 is situated within the intronic region of *TIMP3* while rs17110714 is located 838 bp 3′ downstream of *ABCA4*.

TIMP3 (together with a range of other TIMPs, and the Matrix Metalloproteinases or MMPs) is involved in regulation of the ECM and plays a pivotal role in drusen buildup, which in turn accelerates the onset of AMD. Earlier studies have uncovered susceptibility SNPs within (European) Caucasian and Indian populations (Ristau *et al*. 2014; Kaur *et al*. 2010), but neither risk nor protective alleles have been discovered

within African populations. *ABCA4*, the causative gene for STGD type 1, has previously been postulated to associate with AMD but contradictory results from studies within other population groups have caused uncertainty among the research community (Allikmets *et al*. 1997b; Rivera *et al*. 2000). This study provides the first suggestion of a protective SNP within *ABCA4* and *TIMP3*, especially within the SA indigenous African population.

4.2.2.1 Protective SNP rs9621622 and *TIMP3*

The ECM presents a system of endogenously produced macromolecules, consisting mainly of adhesive and structural fibrous proteins interlinked with polysaccharide glycosaminoglycans. Apart from functioning as the adhesive between cells, the ECM controls multiple cellular functions, including migration, adhesion and proliferation (Teti *et al*. 1992). Integral to the homeostasis of the ECM is the continual balance between synthesis and degradation process. This balance is maintained via the interplay between MMPs and TIMPs.

The zinc-dependent MMPs ability to catabolize structural proteins like elastin and collagen enables essential physiological functions such as tissue remodeling, repair and angiogenesis. To prevent uncontrolled degradation of the ECM, TIMPs regulate MMP activity when inflammatory and tissue remodeling processes are activated. In contrast to other TIMPs, TIMP3 binds to the ECM once released from the RPE and is present in BM under physiological conditions (Teti *et al*. 1992). Within BM, TIMP3 both regulates the rate of ECM turnover and restricts CN, once initiated.

One of the early factors in AMD pathogenesis is thickening of BM which occurs as part of normal aging. Overexpression of *TIMP3* resulting in excessive deposits within soft drusen contributes to the age-related BM thickening (Kamei *et al*. 1999). Excessive TIMP3 results in retardation of BM renewal due to its MMP inhibitive properties. The increase in BM thickening in turn leads to a decrease in permeability and ultimately in

RPE atrophy. Subsequent to RPE atrophy, *TIMP3* expression is vastly reduced resulting in unopposed neovascularisation from MMPs (Kamei *et al*. 1999).

Considering the role of TIMP3 within the pathogenesis of AMD it is possible that a protective SNP, such as rs9621622, could alter *TIMP3* expression within the SA African population. Decreased expression might lead to less accumulation within BM while maintaining sufficient levels to regulate MMP activity. Alternatively, the intronic SNP rs9621622 could be associated with an undiscovered protein-altering variation that results in reduced translation or alternate isoforms of the TIMP3 protein. Kobayashi *et al*. (2013) have previously illustrated this through a genome-wide association study in a Japanese population. This study showed that the C allele of an intronic SNP, rs130293, in *TIMP3* associated with resistance to high-altitude pulmonary edema and suggested a functionally reduced *TIMP3* as the responsible factor. Bearing in mind the complexity of ECM turnover within the pathogenesis of AMD, it is likely that SNP rs9621622 is merely one of many role players contributing to the protective effect detected in Africans, or it itself may be in linkage disequilibrium with a functionally relevant genomic variant.

4.2.2.2 Protective SNP rs17110714 and *ABCA4*

Photon capture within the retina is possible due to the continuous renewal and circulation of Vitamin A derivatives between the RPE and POS. ABCA4, an ABC importer, is located on the outer margins of photoreceptors and functions to flip NR-PE and PE from the POS lumen to the cytoplasmic side. This process prevents the toxic accumulation of the pyridinium compound, A2E, in the RPE as lipofuscin. The initial facet in this process is the ATP binding to the internal nucleotide binding domains which leads to modification of the transmembrane-domain ligand binding site enabling binding of NR-PE and ATR with high affinity. Hydrolysis of ATP results in a conformation change of the channel and subsequent movement of the substrates intracellularly (Shaban *et al*. 2002).

Retinal lipofuscin deposits are ubiquitous across all ethnicities and starts accumulating within the first two decades of life indicating that clearance mechanisms are sub-optimal. A2E forms part of lipofuscin deposits and therefore optimal ABCA4 function is crucial to limit accelerated lipofuscin deposition. Clinical symptoms only arise after a certain threshold of deposition is reached. It is therefore sensible to suggest that certain *ABCA4* genetic factors provide protection against AMD through minor enhancements in the process of substrate removal from POS. These could include: an increased rate of ATP-ase activity, increased binding affinity to substrates, increased transfer rate at lower substrate concentrations and reduction in phospholipid asymmetry within photoreceptor membranes (Quazi *et al*. 2012).

Therefore, these two SNPs are hypothesised to be candidate protective factors for AMD within the SA African population. The locations of these two SNPs serve to provide focus for future research to further assist in elucidating the full genetic contribution of AMD in the African population in SA, and possibly the rest of Africa.

4.2.3 AMD study limitations

The discovery of these two protective SNPs, albeit statistically significant, should be critically viewed when considering the limitations within this study. A cohort sample size of only nine AMD African patients is the most obvious factor impacting the conclusions made within this study. This limitation was expected, as the African AMD patients are infrequent within Ophthalmology clinics in SA. This is further hampered by the majority of African patients in SA being situated within rural areas where a lack of infrastructure prevents healthcare access. A further confounding factor is that the life expectancy in 2014 in SA was 61.2 years which is below the recruitment age for AMD patients (http://beta2.statssa.gov.za/publications/P0302/P03022014.pdf).

Another shortcoming owing to the sub-optimal recruitment, within selected cohorts is the lack of distinction between AMD subtypes, namely dry and wet AMD. Epidemiological studies have shown that the same ethnic group can display dissimilar prevalence rates

of dry AMD depending on the geographical area studied (Rosenberg *et al*. 1987a; Rosenberg *et al*. 1987b). Genome wide linkage studies that stratified families between AMD phenotypes (dry and wet AMD) have shown different genetic loci associated with each phenotype (Majewski *et al*. 2003; Abecasis *et al*. 2004). This current study is therefore unable to distinguish whether these two protective SNPs play a role within dry or wet AMD which is a significant limitation for future research based on this work, as treatment currently only exists for wet AMD.

Only one patient of the African AMD cohort was recruited specifically for this study in collaboration with Ophthalmologists at the Ophthalmology Department of Grootte Schuur Hospital in Cape Town. This patient's diagnosis was optimally confirmed and investigated as described in section 2.1.3, which is not the case with the other eight patients within the African AMD cohort. Interestingly, this patient also showed the least heterozygosity for the nine SNPs in candidate AMD genes among the nine African AMD patients with only two of these nine SNPs occurring in the heterozygous state. Some of these African AMD patients, accessed from within the UCT Retinal Registry, had sub-optimal diagnostic investigations, sometimes consisting only of fundoscopy by a trained Ophthalmologist, but without retinal photographs. This same limitation applies to the Caucasian and Mixed Ancestry AMD cohorts although a larger proportion of these patients were recruited optimally when compared to the African AMD cohort as described above.

Ideally, AMD case and control cohorts of each ethnic-group ought to be derived from the same geographic region to decrease the likelihood of genetic confounders. The African AMD and control cohort were both taken from equivalent SA African groups, the only difference being the proportions of each ethnolinguistic sub-group within each cohort. Conversely, data for the Caucasian control group was derived from the 1000 Genomes CEU population. This group is of Northern and Western European descent while the AMD Caucasian cohort was recruited mostly out of the Western Cape of SA. However, upon deeper examination of the allele frequencies between 'cases' and

'controls' as shown in Table 3.7, one can gauge that this choice of the 1000 Genome Caucasian data as controls was legitimate.

Another possible limitation is that SNP selection was restricted by those represented by the Human SNP 6.0 microarray from Affymetrix (Santa Clara, CA, USA, 1992), which was used to genotype the African control cohort. This chip genotyped a relatively large number of SNPs genome-wide (i.e. 906 600), the data available for the Caucasian controls was effectively that from whole genome sequencing. Therefore, the choice of SNPs displaying an increased frequency was limited to those available within the SNP 6.0 chip. In mitigation, however, almost always, even with the availability of whole genome sequencing data, exploratory studies such as the one presented here rely on subsets of data that are comparable and workable. The 906 000 SNPs representing indigenous southern African populations, may simply be seen as 'tagging reagents' being adequate to interrogate regions of the genome for allelic differences – and these regions are effectively the whole genome.

4.2.4 Future work

Funding and time constraints allowed only ten SNPs to be genotyped in this pilot AMD study, but ideally all variants (69) that displayed an increased frequency within African control group compared to Caucasian controls should be investigated. Future work would include replicating this study using larger, more clinically defined cohorts to improve statistical power and assess whether these findings replicate. It could also provide clarification on SNPs showing marginal significance such as rs1555494 within *HMCN1*. Candidate genes have historically been selected based on their association with increased susceptibility to a disease which is also the case with AMD. Candidate protective factors or modifiers could possibly exist outside of these candidate genes and to find these factors would require use of genome-wide studies and large case cohorts. Future work could include using genome sequencing data when performing these population-based comparative studies to find protective factors located within the genome outside of candidate genes. This would be incomplete without fully sequenced

Caucasian and African control populations from within the same geographical area as case cohorts to provide a more accurate comparison.

Considering the possible effect of SNP rs9621622 on *TIMP3* expression, it would be prudent to quantify and compare BM TIMP3 levels among Caucasian and African cohorts with immuno-histological staining techniques. Illustrating a significant lower TIMP3 concentration in BM of African controls compared to Caucasian controls and AMD cases would provide tentative support for the hypothesis that SNP rs9621622 associates with a protective marker within *TIMP3*.

4.2.4.1 Implications within Mixed Ancestry population

Within the Mixed Ancestry (MA) AMD cohort (18), a trend of increasing heterozygosity was noted when compared to the Caucasian AMD cohort (section 3.5). The MA population forms the major ethnic group (49.6 %) in the Western Cape of SA, from where the AMD cohort was recruited for this study (http://beta2.statssa.gov.za/publications/P0302/P03022014.pdf). The different (increased) MAF within the MA cohort is to be expected as this group represents admixture from four different populations namely: Xhosa, Europeans, South Asians and Indonesians (Patterson *et al*. 2010). In addition to elucidating the various ethnic contributions to the MA group in the Western Cape, Patterson *et al*. (2010) showed the viability of using the MA population for admixture gene mapping. Compared to association and linkage studies, admixture mapping has greater statistical power and requires fewer genetic markers to detect loci that contribute to ethnic discrepancy in complex disease susceptibility (Patterson *et al*. 2010). Due to the MA group's complexity, admixture mapping techniques could present technical difficulties but further investigation within this population is needed to increase the understanding of the genetic factors in AMD.

4.3 Conclusion

In conclusion, a number of variants were identified in screening of *ABCA4* in individuals exhibiting three unique haplotypes around this gene – however only one of the variants was validated as disease causing through various pathogenicity prediction tools. Apart from the possibility of phenotypic heterogeneity, it is possible that the disease causing mutation may lie in one of the other two known STGD-genes (*ELOVL4* or *PROM1*), or it may hint at the existence of a novel gene. Furthermore, this pilot study discovered two possibly protective SNPs for AMD within the SA indigenous African population. It is tempting to speculate that understanding the genetic factors that protect the SA African population from AMD is the initial step in the process of developing locally relevant therapies for individuals affected and predisposed to AMD, thereby enhancing the quality of life for patients and their families.

References

Abecasis GR, Yashar BM, Zhao Y, Ghiasvand NM, Zareparsi S, Branham KEH, Reddick AC, Trager EH, Yoshida S, Bahling J, Filippova E, Elner S, Johnson MW, Vine AK, Sieving PA, Jacobson SG, Richards JE, Swaroop A (2004) Age-related macular degeneration: a high-resolution genome scan for susceptibility loci in a population enriched for late-stage disease. Am J Hum Genet. 74: 482–494.

Adkins RM (2004) Comparison of the accuracy of methods of computational haplotype inference using a large empirical dataset. BMC Genetics 22: doi:10.1186/1471-2156-5-22.

Adkins RM, Krushkal J, Tylavsky FA (2011) Racial Differences in Gene-Specific DNA Methylation Levels are Present at Birth. Birth Defects Res A Clin Mol Teratol. 91: 728–736.

Age-Related Eye Disease Study Research Group (2001) A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8. Arch Ophthalmol. 119: 1417-1436.

Ahmed SS, Lott MN, Marcus DM (2005) The macular xanthophylls. Surv Ophthalmol 50: 183–193).

Allikmets R (2007) Stargardt disease: from gene discovery to therapy. Retinal Degenerations: Biology, Diagnostics and Therapeutics. Humana Press; 2007: 105–118.

Allikmets R, Shroyer NF, Singh N, Seddon JM, Lewis RA, Bernstein PS, Peiffer A, Zabriskie NA, Li Y, Hutchinson A, Dean M, Lupski JR, Leppert M (1997b) Mutation of

the Stargardt Disease Gene in Age-Related Macular Degeneration. Science 277: 1805-1807.

Allikmets R, Singh N, Sun H, Shroyer NF, Hutchinson A, Chidambaram A, Gerrard 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.

Anczuków O, Buisson M, Salles MJ, Triboulet S, Longy M, Lidereau R, Sinilnikova OM, Mazoyer S (2008) Unclassified variants identified in BRCA1 exon 11: Consequences on splicing. Genes Chromosomes Cancer 47: 418-426.

Anderson DH, Radeke MJ, Gallo NB, Chapin EA, Johnson PT, Curletti CR, Hancox LS, Hu J, Ebright JN, Malek G, Hauser MA, Rickman CB, Bok D, Hageman GS, Johnson LV (2010) The pivotal role of the complement system in aging and age-related macular degeneration: Hypothesis re-visited. Progress in Retinal and Eye Research 29: 95–112.

Andreoli MT, Morrison MA, Kim BJ, Chen L, Adams SM, Miller JW, DeAngelis MM, Kim IK (2009) Comprehensive Analysis of Complement Factor H and LOC387715/ARMS2/HTRA1 Variants With Respect to Phenotype in Advanced Age-Related Macular Degeneration. American Journal of Ophthalmology 148: 869–874

Apte SS, Mattei MG, Olsen BR (1994) Cloning of the cDNA of metalloproteinase-3 (TIMP-s) and mapping of the TIMP3 gene to chromosome 22. Genomics 19: 86-90.

Baird PN, Guida E, Chu DT, Vu HT, Guymer RH (2004) The epsilon2 and epsilon4 alleles of the apolipoprotein gene are associated with age-related macular degeneration. Invest Ophthalmol Vis Sci. 45: 1311-1315.

Bauwens M, Van Cauwenbergh C, De Jaegere S, Lefever S, D'haene B, Pattyn F, Leroy B, De Baere E, Coppieters F (2013) A dual approach for comprehensive genetic testing of ABCA4 in Stargardt disease. Invest Ophthalmol Vis Sci 54: 3372.

Bird AC, Bressler NM, Bressler SB, Chisholm IH, Coscas G, Davis MD, de Jong PT, Klaver CC, Klein BE, Klein R (1995) An international classification and grading system for age-related maculopathy and age-related macular degeneration. Surv Ophthalmol 39: 367–374.

Blacharski PA (1988) Retinal Dystrophies and Degenerations: Fundus flavimaculatus. 135-159.

Boon CJF, van Schooneveld MJ, den Hollander AI, van Lith-Verhoeven JJC, Zonneveld-Vrieling MN, Theelen T, Cremers FPM, Hoyng CB, Klevering BJ (2007) Mutations in the peripherin/RDS gene are an important cause of multifocal pattern dystrophy simulating STGD1/fundus flavimaculatus. Br J Ophthalmol 91: 1504–1511.

Braun TA, Mullins RF, Wagner AH, Andorf JL, Johnston RM, Bakall BB, Deluca AP, Fishman GA, Lam BL, Weleber RG, Cideciyan AV, Jacobson SG, Sheffield VC, Tucker BA, Stone EM (2013) Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. Hum Mol Genet. 22: 5136-5145.

Bressler SB (2009) Introduction: Understanding the role of angiogenesis and antiangiogenic agents in age-related macular degeneration. Ophthalmology 116: S1-7.

Bressler SB, Muñoz B, Solomon SD, West SK; Salisbury Eye Evaluation (SEE) Study Team (2008) Racial differences in the prevalence of age-related macular degeneration: the Salisbury Eye Evaluation (SEE) Project. Arch Ophthalmol 126: 241- 245.

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

Brink UT, Terman A (2002) Lipofuscin: mechanisms of age-related accumulation and influence on cell function. Free Radical Biology and Medicine 33: 611–619.

Buzhynskyy N, Salesse C, Scheuring S (2011) Rhodopsin is spatially heterogeneously distributed in rod outer segment disk membranes. J Mol Recognit. 24: 483-489.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., & Madden T.L. (2008) BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Cartegni L, Kraimer AR (2002) Disruption of an SF2/ASF dependent Exonic Splice Enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat Genet 30: 377-384.

Chimusa ER, Meintjies A, Tchanga M, Mulder N, Seioghe C, Soodyall H, Ramesar R (2015) A genomic portrait of haplotype diversity and signatures of selection in indigenous southern African populations. PLoS Genet. 11(3):e1005052.

Cideciyan AV, Aleman TS, Swider M, Schwartz SB, Steinberg JD, Brucker AJ, Maguire AM, Bennet 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: 525-34 .

Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H (2013) Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet 132:1077–1130.

Crabb JW, Miyagi M, Gu X, Shadrach K, West KA, Sakaguchi H, Kamei M, Hasan A, Yan L, Rayborn ME, Salomon RG, Hollyfield JG (2002) Drusen proteome analysis: an approach to the etiology of age-related macular degeneration. Proc Natl Acad Sci USA. 99: 14682-14687.

Cremers FP, van de Pol DJ, van Driel M, den Hollander AI, van Haren FJ, Knoers NV, Tijmes N, Bergen AA, Rohrschneider K, Blankenagel A, Pinckers AJ, Deutman AF, Hoyng CB (1998) Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. Hum Mol Genet. 7(3): 355-62

Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acid Research 37: e67.

Ebermann I, Phillips JB, Liebau MC, Koenekoop RK, Schermer B, Lopez I, Schafer E, Roux AF, Dafinger C, Bernd A, Zrenner E, Claustres M, Blanco B, Nurnberg G, Nurnberg P, Ruland R, Westerfield M, Benzing T, Bolz HJ (2010) PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. J Clin Invest 120, 1812–1823 (2010).

Erke MG, Bertelsen G, Peto T, Sjølie AK, Lindekleiv H, Njølstad I (2012) Prevalence of age-related macular degeneration in elderly Caucasians: the Tromsø Eye Study. Ophthalmology 119: 1737-1743.

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

Fairbrother WG, Yeo GW, Yeh R, Goldstein P, Mawson M, Sharp PA, Burge CB (2004) RESCUE-ESE identifies candidate exonic splicing enhancer in vertebrate exons. Nucleic Acids Res. 32: 187-190.

Fairley C, Zimran A, Phillips M, Cizmarik M, Yee J, Weinreb N, Packman S (2008) Phenotypic heterogeneity of N370S homozygotes with type I Gaucher disease: an analysis of 798 patients from the ICGG Gaucher Registry. J Inherit Metab Dis 31: 738–744.

Fingert JH, Eliason DA, Phillips NC, Lotery AJ, Sheffield VC, Stone EM (2006) Case of Stargardt disease caused by uniparental isodisomy. Arch Ophthalmol 124: 744–745.

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt S, Johnson N, Juettemann T, Kähäri AK, Keenan S, Kulesha E, Martin FJ, Maurel T, McLaren WM, Murphy DN, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, Riat HS, Ruffier M, Sheppard D, Taylor K, Thormann A, Trevanion SJ, Vullo A, Wilder SP, Wilson M, Zadissa A, Aken BL, Birney E, Cunningham F, Harrow J, Herrero J, Hubbard TJ, Kinsella R, Muffato M, Parker A, Spudich G, Yates A, Zerbino DR, Searle SM. Ensembl 2014. Nucleic Acids Res. 42: D749-55.

Friedman DS, Katz J, Bressler NM, Rahmani B, Tielsch JM (1999) Racial differences in the prevalence of age-related macular degeneration: the Baltimore Eye Survey. Ophthalmology 106: 1049- 1055.

Fritsche LG, Chen W, Schu M, Yaspan BL, Yu Y, Thorleifsson G, Zack DJ, Arakawa S, Cipriani V, Ripke S, Igo RP Jr, Buitendijk GH, Sim X, Weeks DE, Guymer RH, Merriam JE, Francis PJ, Hannum G, Agarwal A, Armbrecht AM, Audo I, Aung T, Barile GR, Benchaboune M, Bird AC, Bishop PN, Branham KE, Brooks M, Brucker AJ, Cade WH, Cain MS, Campochiaro PA, Chan CC, Cheng CY, Chew EY, Chin KA, Chowers I, Clayton DG, Cojocaru R, Conley YP, Cornes BK, Daly MJ, Dhillon B, Edwards AO, Evangelou E, Fagerness J, Ferreyra HA, Friedman JS, Geirsdottir A, George RJ, Gieger C, Gupta N, Hagstrom SA, Harding SP, Haritoglou C, Heckenlively JR, Holz FG, Hughes G, Ioannidis JP, Ishibashi T, Joseph P, Jun G, Kamatani Y, Katsanis N, N

Keilhauer C, Khan JC, Kim IK, Kiyohara Y, Klein BE, Klein R, Kovach JL, Kozak I, Lee CJ, Lee KE, Lichtner P, Lotery AJ, Meitinger T, Mitchell P, Mohand-Saïd S, Moore AT, Morgan DJ, Morrison MA, Myers CE, Naj AC, Nakamura Y, Okada Y, Orlin A, Ortube MC, Othman MI, Pappas C, Park KH, Pauer GJ, Peachey NS, Poch O, Priya RR, Reynolds R, Richardson AJ, Ripp R, Rudolph G, Ryu E, Sahel JA, Schaumberg DA, Scholl HP, Schwartz SG, Scott WK, Shahid H, Sigurdsson H, Silvestri G, Sivakumaran TA, Smith RT, Sobrin L, Souied EH, Stambolian DE, Stefansson H, Sturgill-Short GM, Takahashi A, Tosakulwong N, Truitt BJ, Tsironi EE, Uitterlinden AG, van Duijn CM, Vijaya L, Vingerling JR, Vithana EN, Webster AR, Wichmann HE, Winkler TW, Wong TY, Wright AF, Zelenika D, Zhang M, Zhao L, Zhang K, Klein ML, Hageman GS, Lathrop GM, Stefansson K, Allikmets R, Baird PN, Gorin MB, Wang JJ, Klaver CC, Seddon JM, Pericak-Vance MA, Iyengar SK, Yates JR, Swaroop A, Weber BH, Kubo M, Deangelis MM, Léveillard T, Thorsteinsdottir U, Haines JL, Farrer LA, Heid IM, Abecasis GR; AMD Gene Consortium. (2013) Seven new loci associated with age-related macular degeneration. Nature Genetics 45 433–439.

Galehdari H, Saki N, Mohammadi-Asl J, Rahim F (2013) Meta-analysis diagnostic accuracy of SNP-based pathogenicity detection tools: a case of UTG1A1 gene mutations Int J Mol Epidemiol Genet. 4: 77–85.

Gevers W, Casciola LAF, Fourie AM, Sanan DA, Coetzee GA, van der Westhuyzen DR (1987) Defective LDL receptors that are common in a large population: familial hypersholesterolemia in South Africa. Biol Chem 368: 1233-1243.

Griesinger IB, Sieving PA, Ayyagari R (2000) Autosomal dominant macular atrophy at 6q14 excludes CORD7 and MCDR1/PBCRA loci. Invest. Ophthal. Vis. Sci. 41: 248-255.

Hageman GS, Hancox LS, Taiber AJ, Gehrs KM, Anderson DH, Johnson LV, Radeke MJ, Kavanagh D, Richards A, Atkinson J, Meri S, Bergeron J, Zernant J, Merriam J, Gold B, Allikmets R, Dean M; AMD Clinical Study Group (2006) Extended haplotypes in the complement factor H and CFH-related family of genes protect against age-related

macular degeneration: Characterization, ethnic distribution and evolutionary implications. Annals of Medicine 38: 592-604.

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

Hamdi HK, Reznik J, Castellon R, Atilano SR, Ong JM, Udar N, Tavis JH, Aoki AM, Nesburn AB, Boyer DS, Small KW, Brown DJ, Kenney MC (2006) Alu DNA polymorphism in ACE gene is protective for age-related macular degeneration. Biochemical and Biophysical Research Communications 295: 668–672.

Hansen LL, Madsen BE, Pedersen K, Wiuf C (2007) Conflicting results in genotype assessment. BioTechniques 43: 756-762.

Hogg RE, Woodside JV, Gilchrist SECM, Graydon R, Fletcher AE, Chan W, Knox A, Cartmill B, Chakravarthy U (2008) Cardiovascular Disease and Hypertension Are Strong Risk Factors for Choroidal Neovascularization. Ophthalmology 115: 1046–1052.

Hollyfield JG, Besharse JC, Rayborn ME (1977) Turnover of rod photoreceptor outer segments I. Membrane addition and loss in relationship to temperature. J Cell Biol. 75: 490-506.

Holtkamp GM, Kijlstra A, Peek R, de Vos AF (2001) Retinal Pigment Epithelium-immune System Interactions: Cytokine Production and Cytokine-induced Changes. Progress in Retinal and Eye Research 20: 29–48.

Hu DN, Simon JD, Sarna T (2008) Role of Ocular Melanin in Ophthalmic Physiology and Pathology. Photochem Photobiol. 84: 639-644.

Jaakson K, Zernant J, Külm 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 FP, Allikmets R (2003) Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. Hum Mutat. 22: 395–403.

Jonas JB, Bourne RR, White RA, Flaxman SR, Keeffe J, Leasher J, Naidoo K, Pesudovs K, Price H, Wong TY, Resnikoff S, Taylor HR; Vision Loss Expert Group of the Global Burden of Disease Study (2014) Visual Impairment and Blindness Due to Macular Diseases Globally: A Systematic Review and Meta-Analysis. American Journal of Ophthalmology 158: 808–815.

Kamei M, Hollyfield JG (1999) TIMP-3 in Bruch's Membrane: Changes during Aging and in Age-Related Macular Degeneration. Invest. Ophthalmol. Vis. Sci. 40: 2367-2375.

Kaur I, Rathi S, Chakrabarti S (2010) Variations in TIMP3 are associated with age-related macular degeneration. PNAS 107: 112-113.

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

Kennedy C, Rakosczy PE, Constable IJ (1995) Lipofuscin of the retinal pigment epithelium: A review. Eye 9: 763–771.

Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, Kersey P, Flicek P (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) 2011: bar030.

Klaver CC, Assink JJ, van Leeuwen R, Wolfs RC, Vingerling JR, Stijnen T, Hofman A, de Jong PT (2001) Incidence and progression rates of age-related maculopathy: the Rotterdam Study. Invest Ophthalmol Vis Sci 42: 2237–2241.

Klein R, Klein BE, Knudtson, MD, Wong TY, Cotch MF, Liu K, Burke G, Saad MF, Jacobs DR Jr. (2006) Prevalence of Age-Related Macular Degeneration in 4 Racial/Ethnic Groups in the Multi-ethnic Study of Atherosclerosis. Ophthalmology 113: 373–380.

Klein R, Knudtson MD, Klein BE, Wong TY, Cotch MF, Liu K, Cheng CY, Burke GL, Saad MF, Jacobs DR Jr, Sharrett AR (2008) Inflammation, Complement Factor H, and Age-Related Macular Degeneration: The Multi-Ethnic Study of Atherosclerosis. Ophthalmology 115: 1742–1749.

Klein R, Peto T, Bird A, Vannewkirk MR (2004) The epidemiology of age-related macular degeneration. American Journal of Ophthalmology 137: 486–495.

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

Kniazeva M, Chiang MF, Morgan B, 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.

Kobayashi N, Hanaoka M, Droma Y, Ito M, Katsuyama Y, Kubo K, Ota M (2013) Polymorphisms of the Tissue Inhibitor of Metalloproteinase 3 Gene Are Associated with Resistance to High-Altitude Pulmonary Edema in a Japanese Population: A Case Control Study Using Polymorphic Microsatellite Markers. PLoS One. 8: e71993.

Kopplin LJ, Igo RP Jr, Wang Y, Sivakumaran TA, Hagstrom SA, Peachey NS, Francis PJ, Klein ML, SanGiovanni JP, Chew EY, Pauer GJ, Sturgill GM, Joshi T, Tian L, Xi Q, Henning AK, Lee KE, Klein R, Klein BE, Iyengar SK (2010) Genome-wide association

identifies SKIV2L and MYRIP as protective factors for age-related macular degeneration. Genes and Immunity 11: 609–621.

Kumar MV, Nagineni CN, Chin MS, Hooks JJ (2004) Innate immunity in the retina: Toll-like receptor (TLR) signaling in human retinal pigment epithelial cells. Journal of NeuroImmunology 153: 7-15.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21): 2947-2948.

Leistritz DF, Pepin MG, Schwarze U, Byers PH (2011) COL3A1 haploinsufficiency results in a variety of Ehlers-Danlos syndrome type IV with delayed onset of complications and longer life expectancy. Genet Med 13: 717–722.

Lek M, Karczewski K, Minikel E, Samocha K, Banks E, Fennell T, O'Donnell-Luria A, Ware J, Hill A, Cummings B, Tukiainen T. Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv. 1: 030338.

Leske MC, Wu SY, Hennis A, Nemesure B, Yang L, Hyman L, Schachat AP; Barbados Eye Studies Group (2006) Nine-year incidence of age-related macular degeneration in the Barbados Eye Studies. Ophthalmology 113: 29- 35.

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.

Li Z, Clarke MP, Barker MD, McKie N (2005) TIMP3 mutation in Sorsby's fundus dystrophy: molecular insights Expert Reviews in Molecular Medicine 7: 1-15.

Lim KP, Yip SP, Cheung SC, Leung KW, Lam ST, To CH (2009) Novel PRPF31 and PRPH2 mutations and co-occurrence of PRPF31 and RHO mutations in Chinese patients with retinitis pigmentosa. Arch Ophthalmol 127: 784–790.

Linsenmeier RA, Padnick-Silver L (2000) Metabolic dependence of photoreceptors on the choroid in the normal and detached retina. Invest Ophthalmol Vis Sci. 41: 3117-3123.

Maarek Y, Shaul I (1997) WebCutter: A System for Dynamic and Tailorable Site Mapping. Computer Networks and ISDN Systems 29: 1269–1279.

Majewski J, Schultz DW, Weleber RG, Schain MB, Edwards AO, Matise TC, Acott TS, Ott J, Klein ML (2003) Age-related macular degeneration: a genome scan in extended families. Am J Hum Genet. 73: 540–550.

Maliaka YK, Chudina AP, Belev NF, Alday P, Bochkov NP, Buerstedde JM (1996) CpG dinucleotides in the hMSH2 and hMLH1 genes are hotspots for HNPCC mutations. Hum Genet. 97: 251-255.

Mathenge W, Bastawrous A, Peto T, Leung I, Foster A, Kuper H (2013) Prevalence of Age-Related Macular Degeneration in Nakuru, Kenya: A Cross-Sectional Population-Based Study. PLoS Med. 10: e1001393.

Mathenge W, Nkurikiye J, Limburg H, Kuper H (2007) Rapid assessment of avoidable blindness in Western Rwanda: blindness in a postconflict setting. PLoS Med 4: e217.

Maugeri A, Klevering BJ, Rohrschneider K, Blankenagel A, Brunner HG, Deutman AF, Hoyng CB, Cremers FPM (2000) Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. Am J Hum Genet. 67: 960-966.

McVean et al, The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56-65.

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleatic cells. Nucleic Acids Research 16: 55404.

Moore, K. L. Clinically Oriented Anatomy, p (2013).

Mortazavi MM, Griessenauer CJ, Adeeb N, Deep A, Bavarsad Shahripour R, Loukas M, Tubbs R, Tubbs RS (2014) The choroid plexus: a comprehensive review of its history, anatomy, function, histology, embryology, and surgical considerations. Child's Nervous System 30: 205-214.

Nicolas CM, Robman LD, Tikellis G, Dimitrov PN, Dowrick A, Guymer RH, McCarty CA (2003) Iris colour, ethnic origin and progression of age-related macular degeneration. Clinical & Experimental Ophthalmology 31: 465–469.

Niroula A, Urolagin S, Vihinen M (2015) PON-P2: prediction method for fast and reliable identification of harmful variants. PLoS One. 10(2): e0117380.

Olatubosun A, Väliaho J, Härkönen J, Thusberg J, Vihinen M (2012) PON-P: integrated predictor for pathogenicity of missense variants. Hum Mutat. 33: 1166-74.

Oye JE, Kuper H (2007) Prevalence and causes of blindness and visual impairment in Limbe urban area, South West Province, Cameroon. Br J Ophthalmol 91: 1435–1439.

Park HY, Kosmadaki M, Yaar M, Gilchrest BA (2009) Cellular mechanisms regulating human melanogenesis. Cell. Mol. Life Sci. 66: 1493–1506.

Patterson N, Petersen DC, van der Ross RE, Sudoyo H, Glashoff RH, Marzuki S, Reich D, Hayes VM (2010) Genetic structure of a unique admixed population: implications for medical research. Human Molecular Genetics 19: 411-419.

Pauer GJ, Sturgill GM, Peachey NS, Hagstrom SA; Clinical Genomic And Proteomic AMD Study Group (2010) Protective Effect of Paraoxonase 1 Gene Variant Gln192Arg in Age-Related Macular Degeneration. American Journal of Ophthalmology 149: 513–522.

Quazi F, Lenevich S, Molday RS (2012) ABCA4 is an N-retinylidene-phosphatidylethanolamine and phosphatidylethanolamine importer. Nature Communications 3: doi:10.1038/ncomms1927.

Rieder MJ, Taylor SL, Clark AG, Nickerson DA (1999) Sequence variation in the human angiotensin converting enzyme. Nat Genet. 22: 59-62.

Ristau T, Paun C, Ersoy L, Hahn M, Lechanteur Y, Hoyng C, de Jong EK, Daha MR, Kirchhof B, den Hollander AI, Fauser S (2014) Impact of the Common Genetic Associations of Age-Related Macular Degeneration upon Systemic Complement Component C3d Levels. PLoS One. 9: e93459.

Riveiro-Alvarez R, Aguirre-Lamban J, Lopez-Martinez MA, Trujillo-Tiebas MJ, Cantalapiedra D, Vallespin E, Avila-Fernandez A, Ramos C, Ayuso C. (2009) Frequency of ABCA4 mutations in 278 Spanish controls: an insight into the prevalence of autosomal recessive Stargardt disease. Br J Ophthalmol. 93(10):1359-64.

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

Roberts LJ, Nossek CA, Greenberg LJ, Ramesar RS (2012) Stargardt macular dystrophy: common ABCA4 mutations in South Africa--establishment of a rapid genetic test and relating risk to patients. Mol Vis. 18: 280-289.

Rochette J, Le Gac G, Lassoued K, Férec C, Robson KJ (2010) Factors influencing disease phenotype and penetrance in HFE haemochromatosis. Hum Genet 128: 233–248.

Rosenberg T (1987a) Prevalence and causes of blindness in Greenland. Arctic Med Res 46: 13–17.

Rosenberg T (1987b) Prevalence of blindness caused by senile macular degeneration in Greenland. Arctic Med Res 46: 64–70.

Runkle EA, Antonetti DA (2011) The Blood-Brain and Other Neural Barriers Methods in Molecular Biology: The Blood-Retinal Barrier: Structure and Functional Significance 686: 133-148.

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. Inverst Ophthalmol Vis Sci 45(6); 1705-1711

Shaban H, Richter C (2002) A2E and Blue Light in the Retina: The Paradigm of Age-Related Macular Degeneration. Biological Chemistry 383: 537–545.

Shapiro MB, Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15: 7155-7174.

Shen W, Gu Y, Zhu R, Zhang L, Zhang J, Ying C (2013) Copy number variations of the F8 gene are associated with venous thromboembolism. Blood Cells Mol Dis 50: 259–262.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29: 308-11.

Smith W, Assink J, Klein R, Mitchell P, Klaver CC, Klein BE, Hofman A, Jensen S, Wang JJ, de Jong PT (2001) Risk factors for age-related macular degeneration: Pooled findings from three continents. Ophthalmology 108: 697–704.

Spencer KL, Hauser MA, Olson LM, Schmidt S, Scott WK, Gallins P, Agarwal A, Postel EA, Pericak-Vance MA, Haines JL (2007) Protective effect of complement factor B and complement component 2 variants in age-related macular degeneration. Hum Mol. Genet. 16: 1986-1992.

Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM; IARC Unclassified Genetic Variants Working Group. (2008) Prediction and assessment of splicing alterations: Implications for clinical testing. Human Mutation 29: 1304-1313.

Stargardt K (1909) Über familiäre, progressive Degeneration in der Makulagegend des Auges. Graefe's Arch Clin Exp Ophthalmol 71: 534-550.

Stephens M, Donnelly P, Smith NJ (2001) A New Statistical Method for Haplotype Reconstruction from Population Data. Am. J. Hum. Genet. 68: 978–989.

Stone EM (2003) Finding and interpreting genetic variations that are important to ophthalmologists. Trans Am Ophthalmol Soc. 101: 437–484.

Stone EM, Nichols BE, Kimura AE, Weingeist TA, Drack A, Sheffield VC (1994) Clinical features of a Stargardt-like dominant progressive macular dystrophy with genetic linkage to chromosome 6q. Arch. Ophthal. 112: 765-772.

Stothard P (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28: 1102-1104.

Strauss O (2005) The retinal pigment epithelium in visual function. Physiological Reviews 85: 845–881.

Swaroop A, Branham KE, Chen W, Abecasis G (2007) Genetic susceptibility to age-related macular degeneration: a paradigm for dissecting complex disease traits Hum Mol Genet 16: 174–182.

Teti A (1992) Regulation of cellular functions by extracellular matrix. J Am Soc Nephrol. 2: 83-87.

Thiadens AA, Roosing S, Collin RW, van Moll-Ramirez N, van Lith-Verhoeven JJ, van Schooneveld MJ, den Hollander AI, van den Born LI, Hoyng CB, Cremers FP, Klaver CC (2010) Comprehensive analysis of the achromatopsia genes CNGA3 and CNGB3 in progressive cone dystrophy. Ophthalmology 117 825–830.

Thusberg J, Olatubosun A, Vihinen M (2011) Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat. 32: 358-368.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Research 40(15):e115

Utz VM, Beight CD, Marino MJ, Hagstrom SA, Traboulsi EI (2013b) Autosomal dominant retinitis pigmentosa secondary to pre-mRNA splicing-factor gene PRPF31

(RP11): review of disease mechanism and report of a family with a novel 3-base pair insertion. Ophthalmic Genet. 34: 183-188.

Utz VM, Chappelow AV, Marino MJ, Beight CD, Sturgill-Short GM, Pauer GJ, Crowe S, Hagstrom SA, Traboulsi EI (2013a) Identification of three ABCA4 sequence variations exclusive to African American patients in a cohort of patients with Stargardt disease. Am J Ophthalmol. 156(6): 1220-1227.

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

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 EM (2001) An analysis of allelic variation in the ABCA4 gene. Invest Ophthalmol Vis Sci. 42: 1179–89.

Weiss G (2010) Genetic mechanisms and modifying factors in hereditary hemochromatosis. Nat Rev Gastroenterol Hepatol 7: 50–58.

Wolf-Schnurrbusch UE, Röösli N, Weyermann E, Heldner MR, Höhne K, Wolf S (2007) Ethnic differences in macular pigment density and distribution. Invest. Ophthalmol. Vis. Sci. 48: 3783-3787.

Yang Z, Chen Y, Lillo C, Chien J, Yu Z, Michaelides M, Klein M, Howes KA, Li Y, Kaminoh Y, Chen H, Zhao C, Chen Y, Al-Sheikh YT, Karan G, Corbeil D, Escher P, Kamaya S, Li C, Johnson S, Frederick JM, Zhao Y, Wang C, Cameron DJ, Huttner WB, Schorderet DF, Munier FL, Moore AT, Birch DG, Baehr W, Hunt DM, Williams DS, Zhang K (2008) Mutant prominin 1 found in patients with macular degeneration disrupts photoreceptor disk morphogenesis in mice. J. Clin. Invest. 118: 2908-2916.

Yatsenko AN, Shroyer NF, Lewis RA, Lupski JR (2003) An ABCA4 genomic deletion in patients with Stargardt disease. Human Mutation 21: 636-644.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden T (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134.

Yeo G, Burge CB (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 11: 377-394.

Yong Y, Lin HE (2005) SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. Cell Research. 15: 97–98.

Zarbin MA (2004) Current Concepts in the Pathogenesis of Age-Related Macular Degeneration. Arch Ophthalmol. 122: 598-614.

Zhang K, Kniazeva M, Han M, Li W, Yu Z, Yang Z, Li Y, Metzker ML, Allikmets R, Zack DJ, Kakuk LE, Lagali PS, Wong PW, MacDonald IM, Sieving PA, Figueroa DJ, Austin CP, Gould RJ, Ayyagari R, Petrukhin K (2001) A 5-bp deletion in ELOVL4 is associated with two related forms of autosomal dominant macular dystrophy. Nature Genet. 27: 89-93.

Ziskind A, Bardien S, van der Merwe L, Webster AR (2008) The Frequency of the H402 Allele of CFH and Its Involvement with Age-Related Maculopathy in an Aged Black African Xhosa Population. Ophthalmic Genetics 29: 117-119.

Zernant J, Schubert C, Im KM, Burke T, Brown CM, Fishman GA, Tsang SH, Gouras P, Dean M, Allikmets R (2011) Analysis of the ABCA4 gene by next-generation sequencing Invest Ophthalmol Vis Sci. 2011 52(11): 8479-87

Zernant J, Collison FT, Lee W, Fishman GA, Noupuu K, Yuan B, Cai C, Lupski JR, Yannuzzi LA, Tsang SH, Allikmets R (2014) Genetic and clinical analysis of ABCA4-associated disease in African American patients. Hum Mutat. 35(10): 1187-1194.

References were according to the Journal of Human Genetics

Online References

http://analysis.bio-x.cn/myAnalysis.php

http://beta2.statssa.gov.za/publications/P0302/P03022014.pdf

http://bioinf.uta.fi/PON-P/index.shtml

http://bionicvision.org.au/eye/healthy_vision

http://folding.biofold.org/i-mutant/

http://genetics.bwh.harvard.edu/pph2/

http://ghr.nlm.nih.gov/condition/age-related-macular-degeneration/show/Related+Gene(s))

http://global.britannica.com/EBchecked/topic/506498/rod

http://lancastria.net/blog/2010/11/22/page/2

http://rna.lundberg.gu.se/cutter2

http://sift.jcvi.org/

http://snps.biofold.org/phd-snp/phd-snp.html

http://structure.bmc.lu.se/PON-P2/

http://webvision.med.utah.edu/book/part-i-foundations/gross-anatomy-of-the-ey/

http://www.asperbio.com/asper-ophthalmics

http://www.bioinformatics.org/sms2/

http://www.mccormick.northwestern.edu/biomedical/markj/PDFofPapers/markj/Curcio2013.pdf

http://www.corpshumain.ca/en/Vue_en.php

http://www.ensembl.org/

http://www.ensembl.org/biomart/martview/0619df929ccd736db0908ca9d26d23e9

http://www.frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi

http://www.hgvs.org/mutnomen/

http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx

http://www.lighthouse.org/about-low-vision-blindness/vision-disorders/age-related-macular-degeneration-amd/

http://www.ncbi.nlm.nih.gov/

http://www.ncbi.nlm.nih.gov/SNP/

http://www.ncbi.nlm.nih.gov/tools/primer-blast/

http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/

http://www.ngrl.org.uk/Manchester/ page/snap-screening-nonacceptable-polymorphisms

http://www.R-project.org/).

http://www.scienceofdme.org/learn/

http://www.umd.be/HSF/#

http://yellow.nist.gov:8444/dnaAnalysis/primerToolsPage.do

Appendices

Appendix 1:Reagents

10X TBE buffer (stock)

216g (0.89M) Tris (B&M Scientific cc, Cape Town, SA)
110g (0.89M) Boric Acid (ICN Biomedical Inc., USA)
14.8g(0.04M) EDTA (BDH Electron® Laboratory Supplies, UK)
Made up to 2L with H_2O

1X TBE buffer (working stock)

A 10X dilution of the TBE buffer (stock) was made up using H_2O

2% agarose gel (50ml)

1g agarose (Lonza,Seakem® LE, Rockland, USA)
50ml 1 X TBE buffer
5µl (10mg/ml) EtBr (Sigma-Aldrich, USA)

Loading Dye

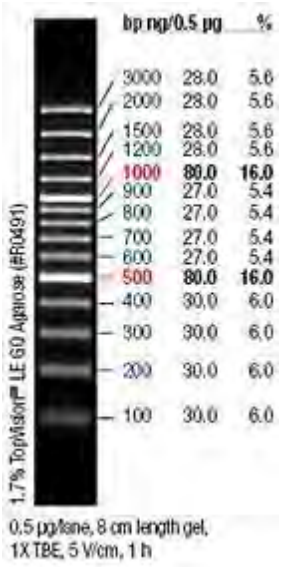
0.125g (0.25%) Bromophenol Blue (Merck KgaA, Germany)
40g (40% w/v) sucrose (Merck Pty Ltd., SA)

pH 8.0

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

Appendix 2: Molecular weight marker

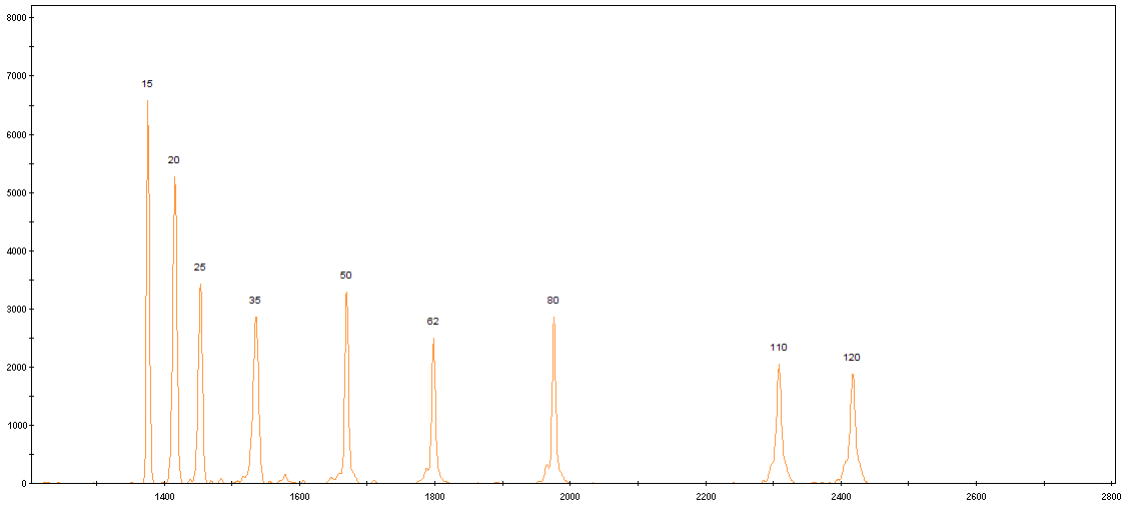
GeneRuler™ 100bp DNA Ladder Plus (Fermentas Life Sciences, Hanover, USA): 25µl molecular weight marker stock solution



Appendix 3: Genescan™ size standard

Genescan™ -120LIZ® (Applied Biosystems, Warrington, UK)

Fluorescent peaks are labelled with size (bp) corresponding to different size fragments.



Appendix 4: AMD candidate genes

The list of 24 genes associated with AMD that was extracted from Genetics Home Reference (http://ghr.nlm.nih.gov/condition/age-related-macular-degeneration/show/Related+Gene(s)) (Updated: June 2011).

- ABCA4
- APOE
- ARMS2
- ASPM
- BEST1
- C2
- C3
- CETP
- CFB
- CFH
- CFHR2
- CFHR4
- CFHR5
- CFI
- CX3CR1
- ELOVL4
- ERCC6
- F13B
- FBLN5
- HMCN1
- HTRA1
- LIPC
- MAP2

- TIMP3

Appendix 5: External primers for *ABCA4*

Exon	F-PrimerSequence	R-PrimerSequence
1	5' GACCAATCTGGTCTTCGTG 3'	5' GTTTATTTGCTCCACACCTC 3'
2	5' TAGCACCACTGAACTTTCTC 3'	5' AAGGCCCAGACCAAAGTCTC 3'
3	5' CCTGCTTGGTCTCCATGAC 3'	5' ACGTGAAGGGGTGTGCAAC 3'
4	5' AGCATGAGATATTATTACTG 3'	5' AGTCTCTCCATAGGTGAGG 3'
5	5' GACCCATTTCCCCTTCAAC 3'	5' AGGCTGGGTGCTTCCCTC 3'
6	5' GGTGTCTTTCCTACCACAG 3'	5' AGGAATCACCTTGCAATTGG 3'
7	5' GTGCCTATGTGTGTATATACC 3'	5' ATAAGTGGGGTAAATGGTGG 3'
8	5' CGCAGCTGGATTAAGGATTG 3'	5' CCCAGGTTTGGTTTCACCTAG 3'
9	5' AAGCAATGGGGAGTTTCTG 3'	5' GAGATGTGCTACCAGGAAG 3'
10	5' GATAGGGGCAGAAAAGACAC 3'	5' CAAGTGGAACTTTGTTGCCC 3'
11	5' CTAAGCAGAGCAGTGACTG 3'	5' ACTTGACTTGCTAAGGGAG 3'
12	5' AGCCCTTATCCTGTGACTTTC 3'	5' TCCAGTCTCAATCCCTTTCTC 3'
13	5' AGCTATCCAAGCCCGTTCC 3'	5' CCATTAGCGTGTCATGGAG 3'
14	5' CCTGTTTTCCTTTCCCTCCATC 3'	5' TCTTTGAGTGTCTCCCACGTTG 3'
15	5' AGGCTGGTGGGAGAGAGC 3'	5' AGTGGACCCCCTCAGAGG 3'
16	5' CTGTTGCATTGGATAAAAGGC 3'	5' GATGAATGGAGAGGGCTGG 3'

17	5' CTGCGGTAAGGTAGGATAGGG 3'	5' CACACCGTTTACATAGAGGGC 3'
18	5' CTCTCCCCTCCTTTCCTG 3'	5' GTCAGTTTCCGTAGGCTTC 3'
19	5' TGGGGCCATGTAATTAGGC 3'	5' TGGGAAAGAGTAGACAGCCG 3'
20	5' ACTGAACCTGGTGTGGGG 3'	5' TATCTCTGCCTGTGCCCAG 3'
21	5' GTAAGATCAGCTGCTGGAAG 3'	5' GAAGCTCTCCTGCTCCAAGC 3'
22	5' CTCTTCCTCACCCTCCACAGC 3'	5' GCTAGGGCTGCAGTGAGA 3'
23	5' AGACACTGTGTGTGGCAATG 3'	5' CCTGTGTGAGTAGCCATGTC 3'
24	5' ATGTGTTGACTACACTTGGCAG 3'	5' GCATCACAACAGGACACACC 3'
25	5' GGTAACCTCACAGTCTTCC 3'	5' GGGAACGATGGCTTTTTGC 3'
26	5' CTTGTGGTTACTGTGTGCTTACG 3'	5' TCGAACTCAGGTGGTCCATC 3'
27	5' GAGATCCAGACCTTATAGGC 3"	5' ACTGAGCTCAGCTAAACACCG 3'
28	5' CCACCAGGGGCTGATTAG 3'	5' CCCAAACCCACAGAGGAG 3'
29	5' GTTGCATGATGTTGGCACG 3'	5' CTTAGGACAGGGGCGCG 3'
30	5' GTCAGCAACTTTGAGGCT 3'	5' TCCCTCTGTGGCAGGCAG 3'
31	5' TTGAAGTGAGAACTAGGGAGATGC 3'	5' TGTTGAATGGGGCCCTCAAATCAG 3'
32	5' GAAAGTTAACGGCACTGC 3'	5' CATGGCTGTGAGGTGTGC 3'
33	5' GCACAGACCAGATGCAGAAG 3'	5' GAGGCCTCTCTAGTGATAGG 3'
34	5' GCTTAACTACCATGAATGAG 3'	5' ATTCCTTGCTAGATTTCAGC 3'
35	5' GCAGCGTCTCAGATGTCCTC 3'	5' AAGAGTGGAGAAGGTGACAAG 3'

36	5' GTATCTTCTCCTCCTTCTGC 3'	5' ACACACAAGCTCCACCTTG 3'
37	5' TTGCAGAGCTGGCAGCAG 3'	5' CCACCAGGCTTCTCTCTTCAG 3'
38	5' GGAATGGAATGTGGAACTCC 3'	5' ACACATACTCTACTATCCTAC 3'
39	5' GCCCCACCCGCTGAAGAG 3'	5' TCCCAGCTTTGGACCCAG 3'
40	5' AGCTGGGGCGGCTGAAG 3'	5' TGCCCTGAGCTGCCCAC 3'
41	5' GGACACTGTACAGCCAGC 3'	5' GACGAGTTATAACACAGGG 3'
42	5' GTCACAGTTCTCAGTCCGG 3'	5' GGAGGAGAGGCAGGCAC 3'
43	5' CTTACCCTGGGGCCTGAC 3'	5' CTCAGAGCCACCCTACTATAG 3'
44	5' GAAGCTTCTCCAGCCCTAGC 3'	5' TGCACTCTCATGAAACAGGC 3'
45	5' GTTTGGGGTGTTTGCTTGTC 3'	5' ACCTATTTCCCCAACCCAAGAG 3'
46	5' GAAGCAGTAATCAGAAGGGC 3'	5' GCCTCACATTCTTCCATGCTG 3'
47	5' CACATCCCACAGGCAAGAG 3'	5' AGGTGGATCCACAGAAGGC 3'
48	5' ATTACCTTAGGCCCAACCAC 3'	5' ACACTGGGTGTTCTGGACC 3'
49	5' GTGTAGGGCGCTGTTTTCCTG 3'	5' GCTCTGAGCCAAGGAACTG 3'
50	5' AGTGCCTCAGCTGAGTGC 3'	5' CCAGCTGCCCAGAGTTCC 3'

Appendix 6: Recruitment forms

Confirmation of diagnosis form

Tel: +27 72 879 0988
Fax: +27 21 650 2010
Email: brdjoh015@myuct.ac.za

Division of Human Genetics
Level 3, Wernher and Beit North
Institute of Infectious Disease and Molecular Medicine
Faculty of Health Sciences
University of Cape Town
Observatory, 7925
South Africa

AMD CONFIRMATION OF DIAGNOSIS

To be completed by Ophthalmologist – PLEASE PRINT CLEARLY

Name of Patient:

Date of Birth: YYYY-MM-DD Tel/Cell: Fax: ( )

Address:

Gender: M F Ethnic Group: Asian | Black | Coloured | Indian | White

Smoker: Y N Previous cataract surgery: Y N

Please complete the following classification:

Diagnosis	Stage		Description
Age-related Macular Degeneration		0	No AMD *or* Small, hard drusen (<63µm)
		1	Soft, distinct drusen (≥63µm) *or* Only pigmentary abnormalities
		2	Soft indistinct drusen (≥125 µm) *or* Reticular drusen only, *or* Soft distinct drusen(≥63µm) with pigmentary abnormalities
		3	Pigmentary abnormalities with either Soft indistinct drusen (≥125 µm) *or* Reticular drusen
		4a	Geographic atrophy (GA)
		4b	Evidence of neovascular "wet" AMD

Doctor's Name:

Contact number: Fax number:

Signature: Date: YYYY-MM-DD

PLEASE RETURN THE COMPLETED FORM TO:
Dr. Johann Baard (brdjoh015@myuct.ac.za / fax: 021 650 2010)

May 2019

Request for molecular studies form

REQUEST FOR MOLECULAR STUDIES FORM

DIVISION OF HUMAN GENETICS, WERNHER & BEIT NORTH
FACULTY OF HEALTH SCIENCES, UNIVERSITY OF CAPE TOWN, OBSERVATORY, 7925
TEL: 021 406-6299 FAX: 021 406-6926 EMAIL: jacquie.greenberg@uct.ac.za

PATIENT DETAILS

SURNAME: _____ NAME: _____

DATE OF BIRTH: ___/___/_____ SEX: FEMALE - ☐ MALE - ☐ ETHNICITY: _____

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

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

CONTACT ADDRESS: _____

_____ CODE _____

TEL: ..... FAX: ..... E-mail .....

REFERRAL SOURCE

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

NAME OF REFERRING DOCTOR: _____ REFERRING FACILITY: _____

FAX: _____ TEL: _____ E-mail: _____

ADDRESS: _____

REASON FOR REFERRAL (CLINICAL DIAGNOSIS)

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

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

FAMILY HISTORY INFORMATION

ADDITIONAL FAMILY HISTORY _____

ADDITIONAL DISORDERS (APPARENT OR PREVIOUSLY TREATED): _____

RELEVANT CLINICAL DETAILS: _____

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

Appendix 7: 17 *ABCA4* SNPs for Haplotype Analysis using PHASE

SNP	Ensembl Location	Distance (bp)	
Exon 6 - 635G>A, R212H rs6657239	94564483	-	
Exon 10 - 1268A>G, H423R rs3112831	94544234	20 249	
Exon 10 - 1269C>T, H423H rs4147831	94544233	1	
IVS10+5delG, rs4147887	94544135	98	
Exon 19 - 2828G>A, R943Q rs1801581	94512565	31 570	
Exon 28 - 4203C>A/T, P1401P rs1801666	94496602	15 963	
IVS33+48C>T, (c.4773+48 C>T) rs472908	94487354	9 248	
Exon 40 - 5603A>T, N1868I rs1801466	94476467	108 87	
Exon 40 - 5682G>C, L1894L rs1801574	94476388	79	
Exon 41 - 5814A>G, L1938L rs4147857	94474328	2 060	
Exon 42 - 5843C>T, P1948L rs56142141	94473846	482	
Exon 42 - 5844A>G, P1948P rs2275029	94473845	1	
Exon 44 - 6069C>T, I2023I rs1762114	94471075	2 770	
Exon 45 - 6249C>T, I2083I rs1801359	94467447	3 628	
Exon 46 - 6285T>C, D2095D rs1801555	94466659	788	
IVS48+21C>T, rs1800699	94463396	3 263	
Exon 49 - 6764G>T, S2255I rs6666652	94461717	1 679	