MOLECULAR GENETIC STUDIES OF RECESSIVELY INHERITED EYE DISEASES

by

SARAH JOYCE

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Clinical and Experimental Medicine
The Medical School
University of Birmingham
July 2011

Abstract

Cataract is the opacification of the crystalline lens of the eye. Both childhood and later-onset cataracts have been linked with complex genetic factors. Cataracts vary in phenotype and exhibit genetic heterogeneity. They can appear as isolated abnormalities, or as part of a syndrome.

During this project, analysis of syndromic and non-syndromic cataract families using genetic linkage studies was undertaken in order to identify the genes involved, using an autozygosity mapping and positional candidate approach.

Causative mutations were identified in families with syndromes involving cataracts. The finding of a mutation in *CYP27A1* in a family with Cerebrotendinous Xanthomatosis permitted clinical intervention as this is a treatable disorder. A mutation that segregated with disease status in a family with Marinesco Sjogren Syndrome was identified in *SIL1*. In a family with Knobloch Syndrome, a frameshift mutation in *COL18A1* was detected in all affected individuals.

Analysis of families with non-syndromic autosomal recessive cataracts was also performed, using genome-wide linkage scans to identify homozygous candidate regions and to then sequence candidate genes within these regions. The identification of a potential putative mutation in one family in CDC25A illustrated the challenges of distinguishing between rare benign variants and pathogenic mutations.

Identification of novel genes involved in cataractogenesis will increase understanding of the pathways involved in cataract formation, and benefit affected families through genetic counselling, and, potentially personalised management.

Acknowledgements

I would like to thank my supervisors Professor Eamonn Maher and Dr Paul Gissen for giving me the opportunity to carry out this research and for their support during the entire process. So many people helped me in the lab, but I would particularly like to thank Dr Esther Meyer for unwavering support, often beyond the call of duty. Thanks also to Uncaar, Manju, Louise, Ania, Neil, Derek, Jane, Holly, Chris, Blerida, Shanaz, Nick, Dean, Dewi and Mike for advice and for answering my questions. I would also like to thank my family, especially my mum and my sister Kathleen, and my friends for all helping me through the tougher moments. I'm sure my dad would be proud. I am grateful for the families who agreed to take part in this project, and for Fight For Sight for funding.this work.

Declaration

In this project, some work was carried out by other people. The Affymetrix Genechip Human Mapping 10K Array was carried out by MRC Geneservice. The Genechip Human Mapping 250K SNP array was carried out by Louise Tee or Fatimah Rahman. Dr Michael Simpson conducted the exome sequencing in this project. Dr Esther Meyer performed linkage analysis for cataract families 1, 5 and 6. DNA extraction was carried out by West Midlands Regional Genetics Molecular Genetics Laboratory.

TABLE OF CONTENTS

List of Illustrations

List of Tables

Abbreviations

AMD	age-related macular degeneration
APS	ammonium persulphate
ASMD	anterior segment mesenchymal dysgenesis
bp	base pair
BFSP1	beaded fiber specific protein 1
BFSP2	beaded fiber specific protein 2
BLAST	basic local alignment search tool
BSA	bovine serum albumin
bZIP	basic region leucine zipper
C17orf42	chromosome 17 open reading frame 42
CCFDN	Cataract, Congenital, with Facial Dysmorphism and Neuropathy
CDC25A	cell division cycle 25 homolog A (S. pombe)
cDNA	complementary DNA
CEPH	Centre d'Etude du Polymorphisme Human
CHX10	CEH10 homeodomain-containing homolog
cM	centiMorgan
CNS	central nervous system
CNV	copy number variant
COL18A1	collagen type XVIII alpha 1
CRYAA	crystallin, alpha-A
CRYAB	crystallin, alpha-B
CRYBA1	crystallin, beta-A1
CRYBA4	crystallin, beta-A4
CRYBB1	crystallin, beta-B1
CRYBB2	crystallin, beta-B2
CRYBB2P1	crystallin, beta B2 pseudogene 1
CRYBB3	crystallin, beta-B3
CRYGC	crystallin, gamma-C
CRYGD	crystallin, gamma-D
CRYGEP	crystallin, gamma E, pseudogene
CRYGFP	crystallin, gamma F pseudogene
CRYGS	crystallin, gamma-S
CRYZP1	crystallin, zeta (quinone reductase) pseudogene 1#
CTDP1	C-terminal domain of RNA polymerase II subunit A, a phosphatase of, subunit 1
CTX	Cerebrotendinous Xanthomatosis
CYP27A1	cytochrome P450, family 27, subfamily A, polypeptide 1
DMSO	dimethyl sulfoxide
dNTP	deoxyribonuclease triphosphates
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid

EMEM	Eagle's Minimum Essential Medium
ER	endoplasmic reticulum
EVI2A	ectropic viral integration site 2A
EVI2B	ectropic viral integration site 2B
F	inbreeding coefficient
F8	oagulation factor VIII, procoagulant component
FACS	fluorescence activated cell sorting
FLVCR2	feline leukemia virus subgroup C cellular receptor family, member 2
FOXE3	forkhead box E3
FSC	forward scatter
GALK1	galactokinase 1
GJA1	gap junction alpha-1, connexin43
GJA3	gap junction alpha-3, connexin46
GJA8	gap junction alpha-8, connexin50
GCNT2	glucosaminyl (N-acetyl) transferase 2
FGF1	fibroblast growth factor 1
FGF2	fibroblast growth factor 2
FGF3	fibroblast growth factor 3
Fq	homozygous by descent
FYCO1	FYVE and coiled-coil domain containing 1
H	heterozygosity
hg18	human genome 18/build 36
HPE	holoprosencephaly
HSF4	heat shock transcription factor 4
IMAGE	integrated molecular analysis of genomes and their expression
INHA	inhibin, alpha
kb	kilobases
kDa	kiloDalton
KNO1	Knobloch Syndrome Type 1
KNO2	Knobloch Syndrome Type 2
KNO3	Knobloch Syndrome Type 3
LB	Lysogeny broth
LIM2	lens intrinsic membrane protein 2
MAF	V-MAF avian musculoaponeurotic fibrosarcoma oncogene homolog gene
Mb	megabases
$MgCl_2$	magnesium chloride
mRNA	messenger RNA
MSS	Marinesco-Sjögren Syndrome
MYO1D	myosin ID
NCBI	National Centre for Biotechnology Information
NF1	neurofibromin 1
NR	non recombinant
OMG	oligodendrocyte myelin glycoprotein

OMIM	Online Mendelian Inheritance in Man
PAX6	paired box gene 6
PBS	phosphate buffered saline
PBST	phosphate-buffered saline Tween
PCR	polymerase chain reaction
PIC	polymorphism information content
PITX3	paired-like homeodomain transcription factor 3
PLA2G6	phospholipase A2, group VI
PVDF	polyvinylidene fluoride
q	disease allele frequency
R	coefficient of relationship
RAB11-FIP4	RAB11 family interacting protein 4 (class II)
RFLP	restriction fragment length polymorphism
RIPA	radioimmunoprecipitation assay
RNA	ribonucleic acid
RNF135	ring finger protein 135
rpm	round per minutes
RX	retina and anterior neural fold homeobox
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SHH	sonic hedgehog
SIL1	SIL1 homolog, endoplasmic reticulum chaperone (*S. cerevisiae*)
SIX3	sine oculis homeobox 3
SNP	single nucleotide polymorphisms
SOX2	SRY-box 2
SRY	sex determining region Y
SSC	side scatter
TBE	Tris-borate-EDTA
TEMED	N,N,N′,N′-tetramethylethylenediamine
TUBA4A	tubulin alpha-4A gene
TUBA4B	tubulin, alpha 4b (pseudogene)
UCSC	University of California Santa Cruz
UNC119	unc-119 homolog (C. elegans)
V	volt
VSX2	visual system homeobox 2
WHO	World Health Organisation
WNT10A	wingless-type MMTV (mouse mammary tumour virus) integration site family, member 10A

Chapter 1
Introduction

1. INTRODUCTION

1.1 Overview

In this project, work was carried out to find genes responsible for non-syndromic cataracts, and some rare autosomal recessive disorders of which cataracts can be a feature, and were present in the families studied. The conditions investigated in this project were Knobloch Syndrome, Marinesco Sjögren Syndrome, and Cerebrotendinous Xanthomatosis. A strategy employing an autozygosity mapping with positional candidate gene identification approach was used. For non-syndromic cataracts, a locus on chromosome three was investigated, and functional analysis was carried out on a putative cataract gene. By identifying genes responsible for syndromic and non-syndromic cataracts in families, a greater understanding of the processes involved in their formation will be gained. It is important to remember the more immediate human benefit of gene identification for the families involved. With this knowledge, an accurate diagnosis can be given, along with genetic counselling, prenatal diagnosis and treatment, and in the future, tailored therapy.

1.2 The Eye

In humans, the eye is a sensory organ which enables light detection. The eye has an approximate diameter of 24mm, being roughly spherical in shape. 85% of its posterior exterior is encased in the protective sclera, which is white in colour. The transparent cornea completes the anterior surface. It is lined by a 7-8μm tear film. The ciliary epithelium secretes aqueous humor into the posterior chamber, which bathes the posterior surface of the cornea. Light entering the eye travels through the cornea and aqueous humor, then reaching the lens (Figure 1.1).

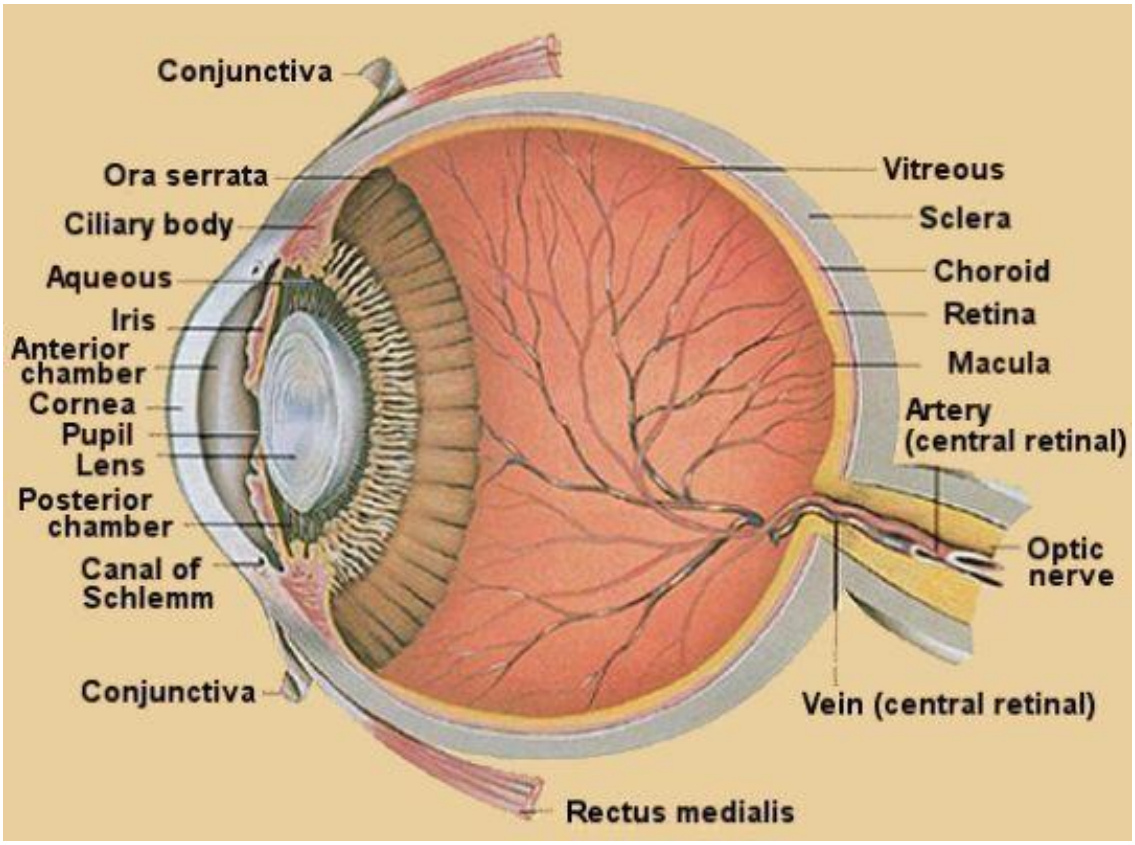


Figure 1.1: *Anatomy of the eye, from www.gbmc.org.*

The biconvex human lens is approximately 5mm thick and its diameter is approximately 9mm, although this changes due to accommodation (focusing on different objects at varying distances from the viewer) and maturation of the individual. It is positioned behind the iris, in the anterior segment of the eye. The lens is transparent, and focuses light onto the photoreceptor rich retina. The lens is capable of transmitting light of wavelength 390-1200nm, which far surpasses the range of visual perception, which only extends to 720nm (Hejtmancik, 2007). The mature lens is an avascular structure.

After the lens, light passes through the vitreous humor, a viscous substance which provides structural integrity to the eye, comprising around 90% of the eye's volume. The retina contains the single largest concentration of the body's sensory receptors,

where there are around 126 million photoreceptors (rods, cones and ganglion cells), accounting for approximately 70% of all receptors. Photoreceptors absorb light and convert it into an electrical signal via phototransduction, reaching the photoreceptor synapses as an impulse which is conveyed to the ganglion cells, and ultimately the brain (Berman, 1991). The retina is actually considered to be part of the brain because it develops from part of the forebrain in the embryo.

1.2.1 Eye Development

In humans, eye development commences during gastrulation (around day 16 of gestation). The anterior neural plate is the central area in the developing brain, and this bisects to form the optic vesicles. At this stage, the lens placode is a constituent of the surface ectoderm, and it must interact with each optic vesicle. After this, the optic cup is created by invagination of the neuroectoderm. The two layers of the optic cup become the neural retina and the retinal pigmented epithelium (Graw, 2003).

1.2.2 Lens Development in Humans

At approximately day 33 of gestation, in the 4mm embryo (Francis *et al*, 1999), the lens cup forms as the lens placode separates from the surface ectoderm. The lens cup develops into the lens vesicle, closing over, which is almost spherical in shape. The posterior segment and anterior pole of the lens cup are populated by epithelial cells, and at approximately day 44 of gestation, those in the posterior segment elongate to fill the lens vesicle's central cavity, forming primary lens fibres (Graw, 2003). At this stage, they lose their nuclei (Francis *et al*, 1999).

Secondary lens fibres are formed by the differentiation of mitotically active cells that migrate from the central to equatorial part of the lens epithelium. The secondary lens

cell fibres surround the primary lens fibres and an onion-like structure begins to form, as they lay down in concentric layers. After this differentiation, lens fibre cells no longer have mitochondria and nuclei (Graw, 2003).The movement of secondary lens fibres from the equatorial region to the outer lens cortex continues for the whole of the individual's life, but the rate decreases. These successive layers of lamellae cause the central fibres to become more compact (Francis *et al*, 1999).

The three main components of the adult lens are the lens capsule, lens fibres and the lens epithelium. The lens is enclosed by the transparent lens capsule, an elastic basement membrane largely made up of sulphated glycosaminoglycans and collagen (type IV). Zonular fibres link the capsule to the ciliary body, and when the zonular fibres are relaxed, the capsule is under less tension, and this causes the lens to become more globular in shape. The lens capsule is not of uniform width, being thinnest around the posterior pole, and widest around the equatorial region. The lens epithelium lies between the lens capsule and lens fibres in the anterior region of the lens. This simple cuboidal epithelium is where the exchange of metabolic products, water and ions occurs with the aqueous humor, a homeostatic mechanism which ensures that nutrients are available for active processes without altering internal lens volume. The lens epithelium differentiates into new transparent lens fibres, which comprise the most part of the lens. They are arranged vertically in layers, lamellae, stretching between the two poles, and are connected by gap junctions.

1.3 Eye Disease

The World Health Organisation (WHO) estimates that, globally, more than 161 million people are visually impaired, with 37 million of these being blind. With a visual

acuity of less than 3/60, an individual was considered to be blind. Low vision was included in the visual impairment category, and is defined as visual acuity (clarity of vision) of less than 6/18, but of equal to or greater than 3/60. (The numerator refers to the distance, in feet, of the subject from the Snellen chart. The denominator is the distance at which the average eye can read a certain line on the Snellen chart.) The WHO 2002 study into visual impairment (Resnikoff *et al*, 2004), used data from 55 countries, and found that 90% of visual impairment cases are in developing countries, with the largest age group (over 82%) being those over 50 years of age. Gender also appears to be a risk factor, with more females than males affected

Cataracts are the most common cause of visual impairment in most regions of the world, excluding developed countries (Figure 1.2). Other eye diseases causing visual impairment are glaucoma, age-related macular degeneration, diabetic retinopathy and trachoma (Resnikoff *et al*, 2004).

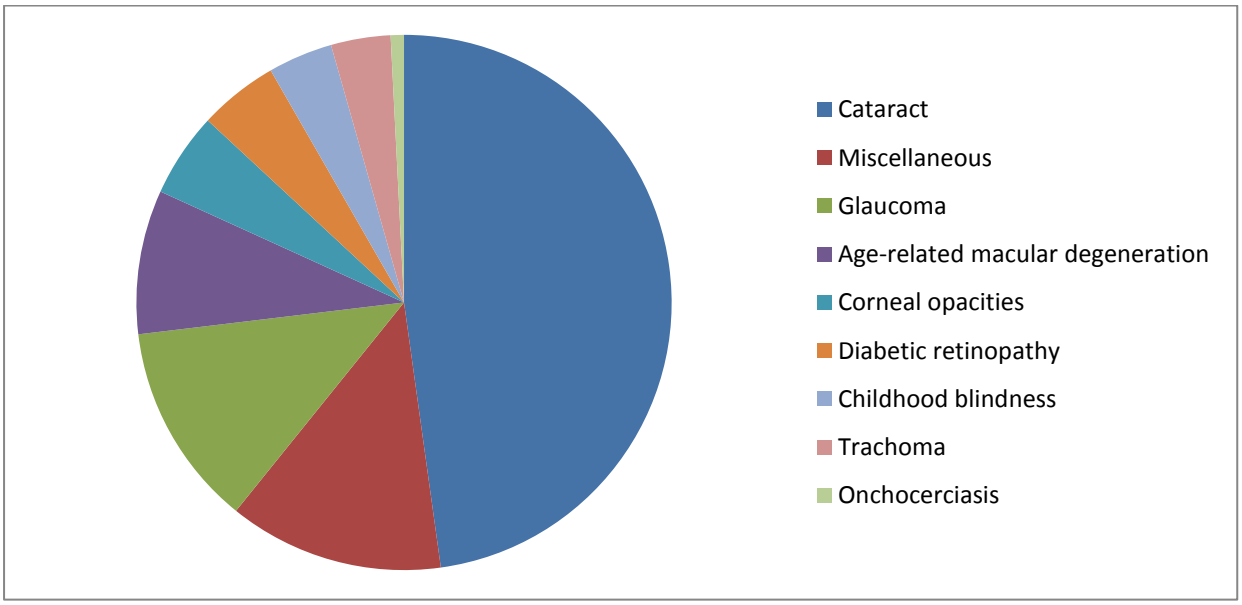


Figure 1.2: *Global causes of blindness as a percentage of total blindness in 2002. From Resnikoff et al (2004).*

Glaucoma is the second most prevalent cause of blindness, followed by age-related macular degeneration (AMD). In those developed countries where cataracts are not the main cause of blindness, it is AMD that is the most prevalent cause, and cases are increasing, in line with an increasingly weighted ageing population (Resnikoff *et al*, 2004). Trachoma refers to scarring of the cornea, resulting from bacterial infection by *Chlamydia trachomatis*. Onchocerciasis (river blindness) is caused by a host immune response to the death of the nematode worm *Onchocerca volvulus* in individuals who were infected through *Simulium* black fly bites.

Vitamin A deficiency has long been associated with visual impairment.

WHO data suggests that 1.4 million children under 15 are blind, and that half of these cases involve avoidable blindness. Treatment of cataract can restore vision, if corrective measures are initiated at a young age.

It is the aim of the WHO to eradicate avoidable blindness by 2020, and Vision 2020 (the Global Initiative for the Elimination of Avoidable Blindness) was set up in 1998 to achieve this.

1.3.1 Hereditary Eye Disease

1.3.1.1 The Genetics of Eye Development

A variety of genes play a role in eye development. Normal development relies on a complex interaction of signalling molecules and transcription factors which are encoded by these genes. Some of those which are the most vital for normal eye development are described below.

1.3.1.1.1 Paired Box Gene 6 (*PAX6*)

The human gene for paired box gene 6 (*PAX6*) is located on chromosome 11p13, and consists of 14 exons. It is a member of a family of genes encoding transcription factors which possess both a homeodomain and a paired domain (Graw, 2003).

When the mouse *PAX6* orthologue, *Pax6*, is expressed ectopically in *Drosophila*, ommatidial eyes are formed, and this was achieved in *Drosophila* legs, wings and antennae (Halder *et al*, 1995). In another study, *Xenopus Pax6* mRNA was injected into frogs, resulting in ectopic formation of normal eyes (Chow *et al*, 1999). Homozygous *PAX6* mutations result in eye absence, severe brain abnormalities, and are lethal (Hanson and Van Heyningen, 1995).

Most human *PAX6* mutations cause aniridia, but they have been associated with cataracts. Two *PAX6* mutations were identified within the same family (Glaser *et al*, 1994). The father had developed bilateral cataracts soon after birth. They progressively worsened, and were surgically removed at age 38 and 40. The mother's ocular anomalies were more severe, including aniridia, but she also presented with bilateral cataracts. A TCA (Ser 353) to TGA (stop) mutation in exon 12 of *PAX6* was found in the father, for which he was heterozygous. It is thought that this will cause PAX6 truncation in the middle of the PST domain (a region rich in proline, serine and threonine). In the mother, a mutation was identified in exon 6 of *PAX6*, for which she was heterozygous. It also involved a premature stop codon, with this being substituted for wild type Arg103. It is thought that this renders the polypeptide non-functional, since truncation would occur in the C terminal portion of the paired domain. Their daughter, the proband, inherited one of each of these

mutated alleles (a compound heterozygote) and died eight days after birth (Glaser *et al*, 1994).

Through experiments with the developing chicken lens, Duncan *et al* (1998) investigated the inverse relationship between Pax-6 and beta-crystallin genes, concluding that a likely effect of decreasing Pax-6 levels in lens fibre cells is to permit beta-B1 crystallin gene expression.

1.3.1.1.2 Sonic Hedgehog Gene (*SHH*)

The human sonic hedgehog gene, *SHH*, is a three exon gene which lies on chromosome 7q36. It has a role in *PAX6* regulation (van Heyningen and Williamson, 2002), and mutations in humans are most commonly known to manifest phenotypically as holoprosencephaly (HPE), with varying degrees of severity. HPE occurs when the eye field and forebrain do not separate into the right and left eye, and most affected individuals do not make it to term. SHH from the prechordal plate probably marks out the area for bisection (Roessler and Muenke, 2003).

SHH mutations causing HPE and cyclopia have also been observed in mice. Targeted *SHH* gene disruption has widespread effects, causing abnormal spinal column, limb and rib development (Chiang *et al*, 1996).

1.3.1.1.3 Sine Oculis Homeobox 3 Gene (*SIX3*)

The human gene sine oculis homeobox 3, *SIX3*, is located on chromosome 2p21. Its two exons span 4.4kb. In humans, *SIX3* mutations have been found to cause holoprosencephaly (Wallis *et al*, 1999).

When the mouse gene *Six3*, which codes for a homeodomain transcription factor, was expressed in the ear placode of the medaka fish, a lens developed (Oliver *et al*, 1996). This indicates the importance of *Six3* in normal eye development.

When *Six3* is overexpressed in chicken embryos, normal lens development is prevented. The lens placode does not undergo invagination, and α–crystallin expression is lost (Zhu *et al*, 2002).

1.3.1.1.4 SRY-Box 2 Gene (*SOX2*)

This single exon SRY (sex determining region Y)-box 2 gene, *SOX2*, on chromosome 3q26.3-q27, encodes a 317 amino acid transcription factor. Heterozygous *SOX2* mutations cause anophthalmia (eyes do not develop) in humans, with all of these apparently *de novo* mutations, inherited in an autosomal dominant manner (Fantes *et al*, 2003). This indicates that SOX2 plays a crucial role in eye development.

1.3.1.1.5 Retina and Anterior Neural Fold Homeobox Gene (*RX*)

The human retina and anterior neural fold homeobox gene, *RX*, is expressed early on in eye development, specifically during retinal development. This gene is found on chromosome 18q21.31 and encodes a homeodomain transcription factor.

The first human *RX* mutations were reported in 2004 (Voronina *et al*, 2004). DNA from 75 individuals with anophthalmia and microphthalmia was obtained, and one of these was found to be a compound heterozygote for *RX* mutations. No mutations were found in the remaining 74, or in 55 controls with normal eyes, although common polymorphisms were detected. The first mutation identified was a premature stop codon instead of a glutamine residue, Gln147X, in exon 2. The

second mutation involved a missense substitution of arginine for glutamine, Arg192Gln, in exon 3. Both mutations occurred in the protein's DNA-binding domain.

In *Rx*-deficient mice, anlagen do not form, *Pax6* is not expressed in the eye field, and eyes do not form (Zhang *et al*, 2000).

1.3.1.2 The Genetics of Lens Development

Lens development is initiated by regulating the formation of the molecular DNA binding complex which is composed of PAX6 and SOX2 on lens-specific enhancer elements (Kamachi *et al*, 2001). After this interaction has occurred, other genes and their products are necessary for normal lens development. Fibroblast growth factors FGF1, FGF2 and FGF3 are involved, as are the transcription factors PAX6, SIX3, and PITX3 (Francis *et al*, 1999). FGF transcripts are expressed in the optic cup and vesicle as they develop, and mouse models are suggestive of their essentiality in lens fibre differentiation (Robinson *et al*, 1998).

FOXE3, MAF, and *PITX3* are involved in lens development, and the *FOX*, *MAF* and *PITX* gene families contain some more pivotal genes involved in further eye development.

1.3.1.2.1 Forkhead Box E3 Gene (*FOXE3*)

FOXE3, forkhead box E3 gene, is found on chromosome 1p21, and comprises one exon. It codes for a transcription factor. The FOX genes are a family which share a 110 amino acid region, which was first noted in *Drosophila* as a DNA binding domain. *FOXE3* is expressed in human adult anterior lens epithelium, and mutations have been associated with anterior segment ocular dysgenesis and cataracts (Semina *et al*, 2001). A frameshift was identified in two affected individuals, causing

the predicted protein to gain 111 amino acids, and possess five abnormal terminal amino acids.

The murine ortholog *FoxE3* is located on chromosome 4, showing strong homology. In situ hybridisation has been used to demonstrate that *FoxE3* is expressed at day 9.5 in the embryonic development of mice, at lens placode induction. Throughout lens placode formation, *FoxE3* expression increased, and continued in the lens vesicle after detachment from the surface ectoderm. During lens fibre differentiation, expression was confined to undifferentiated cells on the anterior lens surface. Peak expression was from 9.5-10 days, ceasing after this time (Blixt *et al*, 2000).

1.3.1.2.2 V-MAF Avian Musculoaponeurotic Fibrosarcoma Oncogene Homolog Gene (*MAF*)

The MAF family of genes encode basic region leucine zipper (bZIP) transcription factors. The *MAF* gene, V-MAF avian musculoaponeurotic fibrosarcoma oncogene homolog, is located on chromosome 16q22-q23. It was identified via an avian retrovirus (AS42) transforming gene, the *v-maf* oncogene (Nishizawa *et al*, 2003). *MAF* gene products have a role in regulation of differentiation. Mutations in *MAF* have been implicated in autosomal dominant juvenile cataracts. An arginine to proline (Arg288Pro) substitution was found in affected individuals of a family, and it was not present in 217 individuals with other ocular anomalies or 496 unaffected control chromosomes (Jamieson *et al*, 2002). This mutation falls in the DNA binding domain of MAF, and is suspected to cause aberrant helical conformation.

1.3.1.2.3 Paired-like Homeodomain Transcription Factor 3 Gene (*PITX3*)

The PITX family of genes consists of three genes that encode transcription factors.

The paired-like homeodomain transcription factor 3 gene, *PITX3*, is a four exon gene located on chromosome 10q25. *PITX3* is expressed in the lens placode and developing lens pit. A 17bp insertion in exon 4 of *PITX3*, causing a frameshift, was identified in a patient with anterior segment mesenchymal dysgenesis (ASMD) and cataracts (Semina *et al*, 1998). The same group also identified a missense mutation (Ser13Asn) in exon 2 of *PITX3* in a family with bilateral congenital cataracts.

1.3.1.2.4. Crystallins

There are numerous crystallin genes. There are also numerous pseudogenes, for example, *CRYBB2P1, CRYGEP, CRYGFP* AND *CRYZP1*.

Crystallins are either taxon (or enzyme) specific, or ubiquitous. Vertebrate crystallins are referred to as ubiquitous, and they make up approximately 95% of human lens proteins. They are highly stable and water soluble. Lens transparency is largely due to their regular organisation in lens fiber cells.

There are three categories of mammalian crystallins: alpha, beta and gamma, although beta and gamma crystallins are also grouped together as a superfamily. This grouping is based on gel exclusion chromatography elution order (Graw, 2009).

The following crystallins are expressed in mammalian lens tissue: alpha-A, alpha-B, beta-B1, beta-B2, beta-B3, beta-A1/A3, beta-A2, beta-A4, gamma-A, gamma-B, gamma-C, gamma-D, gamma-E, gamma-F and gamma-S (Bloemendal *et al*, 2004). In human lenses though, the only crystallins detected with high levels of expression are gamma-C and gamma-D (Brakenhoff *et al*, 1990), but alpha-B crystallin has strong expression in lens epithelial cells (Bhat and Nagineni, 1989). Crystallin

expression is not only limited to the lens, but has been detected in retinal tissue, skeletal muscles, heart muscle, skin and other tissues (Bhat and Nagineni, 1989).

Alpha-crystallin is formed from acidic alpha-A and basic alpha-B crystallin polypeptides and plays a key role in preventing abnormal protein interactions, and acts as a molecular chaperone (Horwitz, 1992). There is evidence linking alpha-crystallin to inhibition of apoptosis. Alpha-A crystallin may play an anti-apoptotic role (Mehler *et al*, 1996). Alpha-B crystallin negatively regulates apoptosis (Kamradt *et al*, 2005). Overexpression of alpha-A and alpha-B crystallin in cell lines heightens tolerance of stress (Andley *et al*, 2000). They both play a role in cytoskeleton maintenance and repair (Liang and McRae, 1997; Vicart *et al*, 1998). They are related to the small heat-shock proteins (Caspers *et al*, 1995). There is a possibility that alpha-crystallins are involved in cell signalling pathways since they undergo *in vivo* phosphorylation and have autokinase activity (Kantorow and Piatigorsky, 1994). Little is understood about *in vivo* alpha-A and alpha-B crystallin substrates.

Beta-crystallins are composed of around five polypeptides.

Beta-gamma-crystallins are a protein family with the same general structure of two domains with a connecting peptide. The domains are made up of four Greek key motifs. These fold into a beta-sandwich structure, with the gamma-crystallins existing as monomers and the beta-crystallins forming complexes (Hejtmancik, 2007). Expression of beta and gamma crystallins has been detected additionally, in lens epithelial cells, and outside the lens (Wang *et al*, 2004). Beta-B2 crystallin has the most extensive expression pattern, and gamma-S crystallin is also expressed outside the lens (Andley *et al*, 2007). Much about the role of gamma and beta-crystallins *in vivo* remains to be elucidated.

Gamma-crystallin mutations tend to cause nuclear or zonular cataracts, presumably based upon when during development the insult exerted its effect. Nuclear cataracts suggest that this occurred early on, and zonular cataracts are a result of a later developmental insult (Hejtmancik, 2007). There has been a report of a gammaD-crystallin mutation that causes nuclear and coralliform cataracts (Gu *et al*, 2006) and of another mutation responsible for microcornea (Hansen *et al*, 2007).

1.3.1.2.5 Connexins

Connexins are a family of genes with over twenty members which share a similar protein structure consisting of four transmembrane domains, with three intracellular regions. These are the N terminus, cytoplasmic loop and C terminus. There are also two extracellular loops, E1 and E2 (Yeager and Nicholson, 2000). One connexon is composed of six connexins. To form a gap junction channel with a pore diameter of 10-15 angstroms, the extracellular loops from two contrary connexons must lodge in the plasma membrane, and hundreds of these pentalamellar structures localise to create gap junctions (Fleishman *et al*, 2004; Unger *et al*, 1999).

Three connexins are known to be structural components of intercellular gap junction channels in the lens, essential for transport in the lens. These are alpha-1 (connexin43 or *GJA1*), alpha-3 (connexin46 or *GJA3*) and alpha-8 (connexin50 or *GJA8*) (Gong *et al*, 2007). This group also demonstrated that alpha-3 connexin is required for lens transparency, and alpha-8 connexin is necessary for lens growth and transparency. Connexin alpha-1 expression is mostly confined to lens epithelial cells (Beyer *et al*, 1989). Connexin alpha-3 expression is mainly evident in lens fibre cells, and connexin alpha-8 expression is found in lens epithelial cells and lens fibre cells (Rong *et al*, 2002).

1.3.2 Cataract

A cataract is said to be present when the crystalline lens of the eye has undergone opacification. They show vast genetic heterogeneity and phenotypic variation.

1.3.2.1 Cataract Classification

Chronological Classification

Cataracts are considered to be congenital or infantile when present during the first year after birth. Those manifesting during life up to ten years of age are termed juvenile cataracts. Presenile cataracts are visible before 45 years of age, and senile (age-related) cataracts occur after this time. These are not absolute classification boundaries, so a degree of variation is encountered.

Congenital Cataracts

Congenital cataracts can occur as isolated anomalies, or they can manifest as a feature in a syndrome. A number of genes have been implicated in cataract formation, and mutations have been identified. The mode of inheritance can be autosomal recessive, autosomal dominant or X-linked.

Non-Genetic Causes of Cataracts

A significant proportion of congenital cataracts, estimated to be 8.3%-25% (Francois, 1982; Merin, 1991), are considered to be genetic in origin. Various environmental factors can give rise to the same phenotype.

1.3.2.2 Types of Cataracts

An attempt to classify the numerous cataract phenotypes seen in slit-lamp examinations of the lens has been made, although there is no standardised

nomenclature. Logical classification is hampered because, in the past, many cataract phenotypes have been named after the affected family or the ophthalmologist. Thefollowing descriptions of opacities are presented as compartmentalised by Reddy *et al* (2004).

Anterior polar cataract: Most anterior polar cataracts are not caused by genetic factors, although there are instances where this is the case. Anterior polar cataracts affect the anterior pole of the lens, mostly as symmetrical and discrete opacities (Figure 1.3A). A study of a four generation family was conducted, where 26 family members were examined, and 17 of these presented with cataract. Non-progressive cataracts were apparent a few months after birth, and in 16 cases were bilateral. Transmission of the defect showed autosomal dominant inheritance. The gene responsible is predicted to fall within a 13cM region between markers D17S489 and D17S796 on chromosome 17p (Berry *et al*, 1996).

Posterior polar cataract: This describes an opacification of the posterior pole of the lens (Figure 1.3B). Both progressive and static forms occur. Progressive posterior polar cataracts were seen in ten affected individuals of a Japanese family (Yamada *et al*, 2000). The mode of inheritance was autosomal dominant, although there was variation in expression. *BFSP1*, a known cataract gene, falls within the linked region (20p12-q12) described by Yamada *et al* (2000), but no mutations were found in the coding exons in the affected members.

Nuclear: Nuclear cataracts affect the foetal and/or embryonic lens nuclei (Figure 1.3C). These are phenotypically varied, with pulverulent nuclear cataracts, blue or punctuate dots or complete nuclear opacification having been documented. Nuclear

cataracts are often described as Coppock or Coppock-like (from the name of the family they were first noted in by Nettleship and Ogilvie in 1906), depending on when the defect manifests during development, although this practice is becoming less common. Coppock-like cataracts refer to those which exist in the nuclear region with a diameter of no more than 2.5mm. These are considered to be embryonal because, during the second month of gestation, this is the diameter of the nucleus, so the cause must have exerted its effect at this stage of development. At this point in lens development, the lens vesicle ceases to exist. Coppock cataracts refer to those with a nuclear diameter of up to 6mm. At birth, the lens has a diameter of 6mm, so cataracts of this type are indicative of a cause that was present during embryonal and foetal lens development. The foetal nucleus is said to have formed when secondary lens fibres are present.

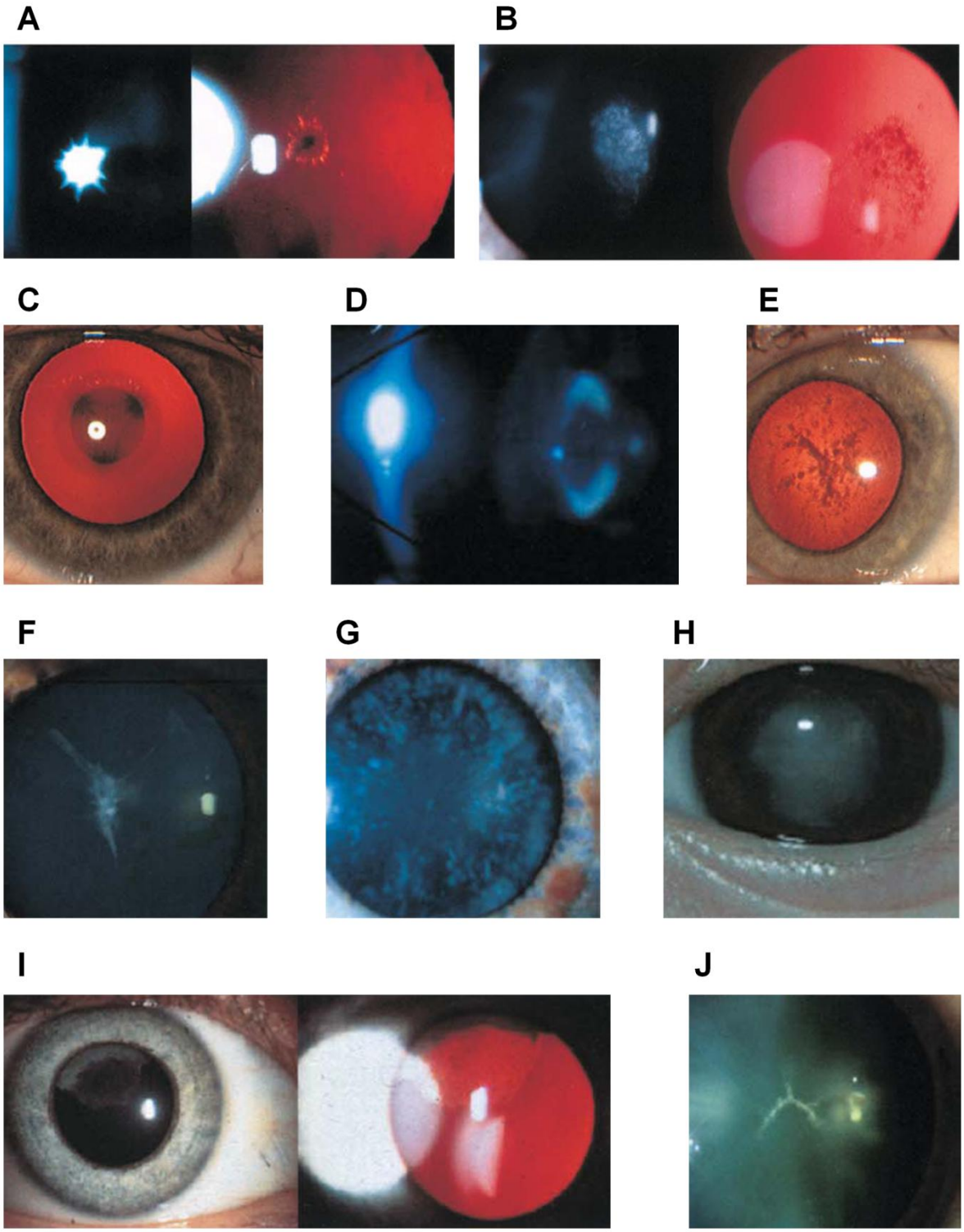


Figure 1.3: *Types of cataracts.* ***A:*** *Anterior polar cataract.* ***B:*** *Posterior polar cataract.* ***C:*** *Nuclear cataract.* ***D:*** *Lamellar cataract.* ***E:*** *Pulverulent cataract.* ***F:*** *Aceuliform-like cataract.* ***G:*** *Cerulean cataract.* ***H:*** *Total cataract.* ***I:*** *Cortical cataract.* ***J:*** *Sutural cataract. From Reddy et al (2004).*

Lamellar: Lamellar cataracts are those occurring in the lamellae (Figure 1.3D). Lamellae are formed from secondary lens fibres as they are laid down concentrically around the developing embryonal nucleus. The size of the opacity can give an indication as to when the cataract arose.

Pulverulent: These opacities speckle the lens in a scattered pattern, showing varying distribution (Figure 1.3E). They can be inherited in an autosomal dominant or autosomal recessive manner. Autosomal dominant inheritance of pulverulent cataract was demonstrated in a Brazilian family with an intronic donor splice junction mutation in betaA1-crystallin, present in all affected members (Bateman *et al*, 2000). Pras *et al* (2002) identified a mutation responsible for pulverulent cataracts in a consanguineous Iraqi Jewish family, which is discussed in more detail under 1.3.2.3.7.

Aceuliform: These rare cataracts consist of spiked processes which project from the nucleus, through the anterior and posterior cortex (Figure 1.3F).

Cerulean: Cerulean cataracts are progressive, being absent at birth, but becoming apparent during childhood. They consist of cerulean blue and white specks distributed across the lens in a scattered pattern, with a greater distribution density in the cortex (Figure 1.3G). Cuneiform opacities may be seen in the mid-periphery of the cortex as a result of this (Ionides *et al*, 1999).

Total: This type of cataract is most common in boys showing X-linked cataract inheritance (Reddy *et al*, 2004). Phenotypically, it is visible as an opacity of the entire foetal nucleus at birth and in the cortex in early infancy (Figure 1.3H).

Cortical: These are another rare inherited manifestation of cataract. The affected area of the lens is the outer part of the cortex, next to the lens capsule (Figure 1.3I). It is a less severe form, since it affects new secondary lens fibres in the equatorial region, causing minimal visual reduction.

Sutural: Sutural cataracts (those affecting the anterior and posterior sutures) (Figure 1.3J) usually present alongside other lenticular anomalies, although they are seen without other opacities in Nance-Horan syndrome, in female carriers of X-linked cataract (Bixler *et al*, 1984).

Polymorphic: This description broadly refers to asymmetric cataracts in the polar region of the lens which show great variability amongst affected members, even in the same family. This type of cataract was first classified in three unrelated families, showing autosomal dominant inheritance, having a phenotypic appearance as a bunch of grapes (Rogaev *et al*, 1996).

Another review suggests grouping cataracts according to phenotype as follows: total (mature, complete, Morganian or disk-like), polar (anterior polar, anterior pyramidal, anterior subcapsular, anterior lenticonus, posterior subcapsular, posterior lenticonus, or posterior cortical), zonular (lamellar, nuclear oil droplet, cortical, coronary, sutral, pulverulent, cerulean, or coralliform), or membranous (capsular) (Huang and He, 2010).

1.3.2.3 Genetics of Cataracts

Different genes are involved in cataract formation, but they fall into subgroups of genes, including those encoding crystallins (discussed in 1.3.2.3.2), beaded filament

structural proteins, lens intrinsic membrane proteins, heat shock factor proteins, glucosaminyl (N-acetyl) transferases and gap junction proteins.

1.3.2.3.1 Beaded Filament Structural Protein 1 Gene (*BFSP1*)

Beaded fiber specific proteins (BFSPs) are intermediate filament proteins, and are lens specific (Perng and Quinlan, 2005). They consist of a 12-15nm globular head and a 7-9nm filament, with the filament consisting of BFSP1 (115kDa protein, also known as filensin or CP115) alone, and the globular head being made up of both BFSP1 and BFSP2 (49kDa protein, also known as phakinin and CP49) (Goulielmos *et al*, 1996). It is likely that they play a major role in maintenance of lens opacity in humans, since *BFSP1* and *BFSP2* knock-out mice develop cataracts (Alizadeh *et al*, 2002 and 2003). *BFSP1* knock-out mice start to develop cataracts at around 2 months, progressively worsening over time (Alizadeh *et al*, 2003).

A study on a consanguineous Indian family by Ramachandran *et al* (2007) found a 3343bp deletion (c.736-1384_c.957-66del) in the beaded filament structural protein 1 gene *BFSP1* (located on chromosome 20p12.1) which included exon 6 (Ramachandran *et al*, 2007). This was predicted to cause a shift in the open reading frame. 19 family members were examined, with 11 of these being affected, and the mutation segregated in an autosomal recessive manner. Linkage to a 5.43Mb region on chromosome 20p between the markers D20S852 and D20S912 was shown, and *BFSP1* is located at 20p12.1. The deletion was not found in 50 controls from southern India. Most of the affected individuals in this family presented with cataracts at five years of age, although in one individual, cataracts were noted at two years.

1.3.2.3.2 Crystallins

Most crystallin genes, when, mutated, play a role in cataractogenesis (Table 1.1).

1.3.2.3.2.1 Crystallin, Alpha-A (*CRYAA*)

Both autosomal dominant and autosomal recessive cataracts can result from alphaA-crystallin mutations. A Trp9X substitution in *CRYAA* in a Persian Jewish family caused a chain termination, and cataracts were observed, inherited autosomal recessively (Pras *et al*, 2000). Since the affected individuals had the cataracts extracted before three months of age, no cataract phenotype information was available. Non–conservative missense mutations in alphaA-crystallin have been found in families exhibiting autosomal dominant cataract inheritance. It would appear that the surface charge is important for normal functioning, since mutations identified tend to involve the substitution of basic arginine to a neutral or hydrophobic amino acid (Hejtmancik, 2007). A missense mutation in *CRYAA* causing Arg116Cys was discovered in a family with autosomal dominantly inherited congenital zonular central nuclear cataracts (Litt *et al*, 1998).

1.3.2.3.2.2 Crystallin, Alpha-B (*CRYAB*)

Mutations in *CRYAB* cause autosomal dominant “discrete” cataracts and desmin-related myopathy (Vicart *et al*, 1998) because there is a sharp drop in chaperone activation, causing the protein to aggregate and precipitate (Bova *et al*, 1999). Vicart *et al* (1998) identified Arg120Gln mutations in a French family. In two patients with myofibrillar myopathy, mutations in *CRYAB* were detected: 464delCT and Gln151X (Selcen and Engel, 2003). No cataracts were noted with direct ophthalmoscopy. Slit-lamp investigation was not carried out.

1.3.2.3.2.3 Crystallin, Beta-A1 (*CRYBA1*)

A 3bp deletion causing delGly91 in *CRYBA1* segregated with disease status in a Chinese family, where affected members had autosomal dominant congenital nuclear cataract (Qi *et al*, 2004).

1.3.2.3.2.4 Crystallin, Beta-A4 (*CRYBA4*)

CRYBA4 was proven to have a role in cataractogenesis, when mutated, by Billingsley *et al* (2006).A heterozygous c.317T>C genotype was detected in members of an Indian family with autosomal dominant congenital lamellar cataract. This was predicted to cause Phe94Ser, replacing a hydrophobic amino acid with a hydrophilic one, causing protein instability. Leu69Pro was also noted in the index case, a de novo mutation, likely to be responsible for the other clinical manifestations described in this individual (microphthalmia and enophthalmia).

1.3.2.3.2.5 Crystallin, Beta-B1 (*CRYBB1*)

Both autosomal dominant and autosomal recessive transmission of disease status for cataracts linked to mutations in this gene have been reported.

Gly220X was predicted in a family with autosomal dominant pulverulent cataract, where a c.658G>T transversion segregated with disease status (Mackay *et al*, 2002). Pulverulent cataract phenotype has also been associated with mutations in *CRYGC* (Ren *et al*, 2000).

Cohen *et al* (2007) investigated the genetics of autosomal recessive congenital nuclear cataract in two apparently unrelated consanguineous Israseli Bedouin families. In affected individuals, they identified a single homozygous nucleotide

deletion, 168delG which caused a frameshift. This resulted in a missence sequence at amino acid 57, and termination of the polypeptide at codon 107.

1.3.2.3.2.6 Crystallin, Beta-B2 (*CRYBB2*)

Litt *et al* (1997) detected a single base substitution (C475T) in CRYBB2, causing Gln155X, and hence, premature termination of beta-b2 crystallin by 51 amino acid residues. This was detected in a family where affected individuals presented with congenital cerulean cataract. Four affected family members were heterozygous for the mutation, and an affected child of a first cousin union was homozygous for the change.

In a four generation Swiss family, Gill *et al* (2000) identified the same mutation as Litt *et al* (1997), showing autosomal dominant transmission, but in this family, affected individuals had Coppock-like cataracts. This finding highlights the variety of cataract phenotypes found from mutations within the same gene, and even from the same mutation.

1.3.2.3.2.7 Crystallin, Beta-B3 (*CRYBB3*)

A G493C nucleotide substitution was identified in *CRYBB3* in two Pakistani families, predicted to result in Gly165Arg. The affected family members displayed autosomal recessive congenital nuclear cataract (Riazuddin *et al*, 2005).

1.3.2.3.2.8 Crystallin, Gamma-C (*CRYGC*)

A C502T substitution in *CRYGC* caused Arg168Trp, resulting in congenital lamellar cataract in an affected mother and two affected children in an Indian family (Santhiya *et al*, 2002). Autosomal dominant inheritance was evident even though the family

was consanguineous, since affected individuals were found to be heterozygous for the mutation.

Autosomal dominant inheritance for a *CRYGC* mutation was described by Ren *et al* (2000) in a family where affected individuals had cataracts. This time, the cataract phenotype was noted as variable zonular pulverulent, with a range of phenotypes including subtle unilateral pulverulent cataract, and bilateral nuclear cataracts. The causative mutation is a 5bp deletion in exon 2.

1.3.2.3.2.9 Crystallin, Gamma-D (*CRYGD*)

Héon *et al* (1999) identified a G411A substitution, predicted to result in Arg58His. Individuals heterozygous for this change exhibited aculeiform cataracts in three families. One family was from Macedonia, and the other two were from Switzerland (Héon *et al*, 1998).

There has also been a report of a gammaD-crystallin mutation that causes nuclear and coralliform cataracts (Gu *et al*, 2006) and of another mutation responsible for microcornea (Hansen *et al*, 2007).

1.3.2.3.2.10 Crystallin, Gamma-S (*CRYGS*)

A Gly18Val substitution mutation was detected in exon 2 of *CRYGS* in a six generation Chinese family, with affected individuals displaying progressive polymorphic cortical cataracts (Sun *et al*, 2005). Known loci were excluded with microsatellite markers. This was followed by a genome wide scan which identified a 20.7cM region on chromosome 3 with a maximum lod score of 6.34 (θ=0.0) at D3S1602.

Gene	Location	Mode of Inheritance	Phenotype of Cataracts	Mutation	Reference
CRYAA	21q22.3	Autosomal recessive	Not described	Trp9X	Pras *et al*, 2000
		Autosomal dominant	Congenital zonular nuclear	Arg116Cys	Litt *et al*, 1998
		Autosomal dominant	Bilateral early onset (and iris coloboma)	Arg116Cys	Beby *et al*, 2007
CRYAB	11q22.3-q23.1	Autosomal dominant	Congenital posterior polar	Frameshift at codon 150 (450delA)	Berry *et al*, 2001
CRYBA1	17q11.2-q12	Autosomal dominant	Congenital nuclear	Gly91Del	Qi *et al*, 2004
CRYBA2	2q34-q36	No association with cataracts reported to date			
CRYBA4	22q11.2-q13.1	Autosomal dominant	Congenital lamellar	Phe94Ser	Billingsley *et al*, 2006
CRYBB1	22q11.2-q12.1	Autosomal recessive	Bilateral pulverulent	Gly220X	Cohen *et al*, 2007
		Autosomal dominant	Congenital nuclear	Frameshift at codon 57	Mackay *et al*, 2002
CRYBB2	22q11.2-q12.1	Autosomal dominant	Congenital cerulean	Gln155X	Litt *et al*, 1997
		Autosomal dominant	Coppock-like	Gln155X	Gill *et al*, 2000
CRYBB3	22q11.2-q12.2	Autosomal recessive	Congenital nuclear	Gly165Arg	Riazuddin *et al*, 2005
CRYGA	2q33-q35	No association with cataracts reported to date			
CRYGB	2q33-q35	No association with cataracts reported to date			
CRYGC	2q33-q35	Autosomal dominant	Congenital lamellar	Arg168Trp	Santhiya *et al*, 2002
		Autosomal dominant	Variable zonular pulverulent	5 bp insertion in exon 2	Ren *et al*, 2000
CRYGD	2q33-q35	Autosomal dominant	Aculeiform	Arg58His	Héon *et al*, 1999
CRYGS	3q27	Autosomal dominant	Progressive polymorphic cortical	Gly18Val	Sun *et al*, 2005

Table 1.1: *Selected examples of cataract phenotypes resulting from various crystallin gene mutations.*

1.3.2.3.3 Galactokinase 1 (*GALK1*)

Galactokinase catalyses the first stage of the galactose metabolic pathway, when galactose is phosphorylated, forming galactose-1-phosphate. Okano *et al* (2001) found an Ala198Val mutation in *GALK1* (referred to as the Osaka variant) in three non-consanguineous Japanese individuals referred to clinic for high galactose levels. The three patients were compound heterozygotes, all with one copy of Ala198Val, and then c.509-510delGT in patient one, Met1Ile in patient two, but the second mutation in the third patient was not described. Based on analysis of DNA from infants, they suggested that the Osaka variant had a 4.1% prevalence in the Japanese population (24 out of 582 alleles), and a 7.8% frequency ($p<0.023$) in Japanese individuals with bilateral cataract, and that this variant may play a role in late onset cataract. 2.8% of Korean individuals screened (8 out of 288 alleles) had the Osaka variant, but it was absent in all screened individuals from the USA (188 alleles in Caucasians, 10 alleles in Blacks).

Stambolian *et al* (1995) studied two families and found *GALK*1 mutations in both. In the first family, the female proband had cataracts as the result of homozygous Val32Met substitution (which lowers galactokinase activity). Cataracts were absent in the heterozygous parents. In the second family, the proband (offspring from a first cousin union) developed cataracts and galactosaemia by the age of one. A single base substitution was noted (T238G), which produced a nonsense codon (TAG) at amino acid 80 (Glu80X). The parents were heterozygous for this substitution and the child was homozygous.

When Ai *et al (*2000) cloned and disrupted *Galk1* in mice, galactose levels increased, but cataracts did not develop after six months on a high galactose diet. When these

mice were bred with mice expressing human *ALDR1*, and the resulting transgene expressed in *Galk1* deficient mice, cataracts were noted in day one postnatal mice.

1.3.2.3.4 Glucosaminyl (N-acetyl) Transferase 2 Gene (*GCNT2*)

This 3 exon glucosaminyl (N-acetyl) transferase 2 gene is located on chromosome 6p24.2 and there are three known isoforms (GCNT2A, -B and –C). GCNT2 (also known as IGNT) has a role in i/I blood group differentiation. The i/I blood group classification system refers to the linear i and branched I poly-N-acetyllactosaminoglycans. The i antigen is found in foetal erythrocytes, and has been converted to the I antigen in adult erythrocytes by the enzyme GCNT2 (Bierhuizen *et al*, 1993). Mutations in *GCNT2* give rise to structural abnormalities in the enzyme, which prevents formation of the branched I antigen. This results in the adult i blood group phenotype. It is not problematic unless blood transfusions are received, where it has been known to cause severe haemolytic anaemia, due to anti-I antibody (Chaplin *et al*, 1986).

A study looked at four distantly related Arab families from Israel, whose 13 analysed affected members presented with congenital cataract (Pras *et al*, in 2003). In the majority of cases, bilateral leukocoria had been noted before the child was one month old. Pras *et al* (2003) mapped a region of 13.0cM between D6S470 and D6S289. A homozygous substitution segregated with cataracts in the families, indicating autosomal recessive inheritance. There was a G to A substitution for the 58th base in exon 2 which resulted in premature stop codon formation: Trp328X for *GCNT2A*, Trp326X for *GCNT2B* and Trp328X for *GCNT2C*.

In humans, the three *GCNT2* isoforms have an identical exon 2 and exon 3, but exon 1 differs. The variants *GCNT2-1A* and *-1C* have a coding region of 925bp, and

919bp in *GCNT2-1B* which is the only expressed isoform in human lens-epithelium cells (Yu *et al*, 2003). Mutations in *GCNT2-1C* alone were not found to be sufficient to cause cataracts.

1.3.2.3.5 Gap Junction Alpha-8 Protein Gene (*GJA8*)

The gene for gap junction alpha-8 protein (GJA8) is located on chromosome 1q21.1. There are two known *GJA8* variants according to Ensemble Release 50 (http://www.ensembl.org). The first variant has two exons, with exon 1 being non-coding and the second exon consisting of 1,322 nucleotides, and the second variant has one exon with 1,302 coding nucleotides.

Research was conducted on a three-generation Russian family who had three members affected with zonular pulverulent cataract, where DNA was available from two affected members (Polyakov *et al*, 2001). The proband presented with cataracts aged three years. Markers D1S2696 and D1S252 were used to confirm linkage to a region containing *GJA8* which showed a T to G substitution for nucleotide 741. This causes isoleucine to be substituted for a methionine residue at position 247, and these amino acids are neutral and have non-polar side chains. This change was found to segregate in the family in an autosomal recessive manner, and was not found in 25 unrelated controls. Mutations in this gene had been previously reported to be involved in cataract formation, but the mode of inheritance was autosomal dominant (Shiels *et al*, 1998).

1.3.2.3.6 Heat Shock Transcription Factor 4 Gene (*HSF4*)

Heat shock response genes are activated by heat shock transcription factors in response to stresses such as heat and heavy metals. The 15 exon gene encoding

heat shock transcription factor 4, *HSF4*, is located on chromosome 16q21. In humans, there are two isoforms, HSF4a and HSF4b. HSF4b has an extra 30 amino acids (Smaoui *et al*, 2004).

A large consanguineous Tunisian family were studied, with the researchers having obtained DNA from 63 family members, with 22 of these being affected with congenital total white cataract (Smaoui *et al*, 2004). Linkage was confirmed to a 1.8cM (4.8Mb) region flanked by the markers D16S3031 and D16S3095, which contains *HSF4*. A homozygous mutation in the 5' splice site of intron 12 (c.1327+4A>G) was found, which segregated in the family, and causes exon 12 skipping.

This study by Smaoui *et al* (2004) was the first to report an association between *HSF4* and autosomal recessive cataracts. *HSF4* mutations had previously been associated with autosomal dominant cataracts, such as Marner cataracts (Bu *et al*, 2002). An Arg120Cys substitution, caused by c.36C>T in exon 3 of *HSF4*, was shown to be responsible for autosomal dominant segregation of zonular stellate opacity with anterior polar opacity in a Danish family, first reported in 1949 (Marner, 1949).

1.3.2.3.7 Lens Intrinsic Membrane Protein 2 Gene (*LIM2*)

LIM2 is found on chromosome 19q13.4 and consists of five exons. It codes for a protein (lens intrinsic membrane protein 2, LIM2) of 173 amino acids which has an approximate molecular mass of 20kDa, containing four intramembrane domains (Arneson and Louis, 1998). It is generally considered to be the second most abundant lens fibre cell protein.

In an inbred Iraqi Jewish family, a homozygous mutation was found to segregate with pre-senile cataracts (Pras *et al*, 2002). A Phe105Val substitution in *LIM2* was found, caused by a T to G change. The parents in the family were first cousins, and their three affected children presented with cataracts between the ages of 20 and 51 years.

A Val15Gly mutation in the mouse *LIM2* ortholog, *Lim2*, gives rise to cataracts. This was demonstrated in To-3 mice (Steele *et al*, 1997). The inheritance is semidominant, as mice heterozygous and homozygous for this mutation develop cataracts.

1.3.2.3.8 Visual System Homeobox 2 Gene (*VSX2*)

The 5 exon visual system homeobox 2 gene, *VSX2* is located on chromosome 14q24.3. It is also known as *CHX10*, for CEH10 homeodomain-containing homolog, with expression in the retina and retinal neuroblasts. It codes for a 361 amino acid polypeptide of approximately 39 kDa (Percin *et al*, 2000). Percin *et al* (2000) looked at a Turkish consanguineous kindred with non-syndromic microphthalmia (OMIM 251600), cataracts and severe iris abnormalities. Two affected family members were studied. Markers were used to show homozygosity in affected members in the 6.3cM region flanking *VSX2,* between D14S77 and D14S61. A G>A substitution caused the arginine residue at 200 to be substituted for glutamine (Arg200Gln) in the two affected members. This change was not found in 365 mixed ethnicity controls with eye defects, or in 110 Turkish control chromosomes. This residue lies within the DNA recognition helix of the homeodomain. This study also employed gel mobility shift assays to show that Arg200Gln in *VSX2* reduces binding to the consensus binding site.

1.4 Disease Gene Identification

Identification of disease genes is essential for understanding disease pathogenesis and the pathways involved, and hence, potential therapeutic targets, resulting in improved patient care.

Although there are a number of methods that can be employed to identify disease genes, they essentially fall into two categories: those techniques identifying a disease locus and selecting candidate genes within this region (positional approach), and those that proceed in a positionally-independent manner (functional candidate gene approach). In practice, most research groups will amalgamate the two and use a positional-candidate gene identification approach.

1.4.1 Functional Candidate Gene Approach

It is possible to identify disease genes if the pathology of the disease is understood. Before genetic mapping information was readily available, this was the most frequently used approach to gene identification. Haemophilia A is an X-linked recessive disorder caused by mutations in *F8* (Xq28), and sufferers bruise easily and bleed excessively. *F8* was identified as the causative gene via a functional candidate approach (Gitschier *et al*, 1984). The biochemical basis for haemophilia A was understood inasmuch as patient serum samples had indicated factor VIII deficiency, so gene-specific oligonucleoties were designed to amplify and sequence cleaved factor VIII peptides.

Animal models are a useful tool for the functional approach as they are likely to react as a human would, but they are more appropriate for experimentation. Comparisonn of the genomes of other species to the human genome can be a useful approach in

candidate gene identification. The mouse genome exhibits 99% homology to the human genome (Waterson *et al*, 2002). Even less genetically similar species still serve as useful models for researchers. This means that a lot of genetic research is carried out in model systems such as *Danio rerio* (zebrafish), *Drosophila melanogaster* (fruitfly) and *Mus musculus* (mouse). The *Drosophila melanogaster* genome has been sequenced (Adams *et al*, 2000). Along with this readily available information, this organism is particularly suitable for research because around 75% of human disease genes have a fly ortholog (Reiter *et al*, 2001), it requires little space and care in the laboratory, gestation and maturation is relatively rapid and development from egg to adult fly can occur in one week, and many distinct morphological changes can be seen without the aid of a microscope. *Drosophila* has been used in studies of eye disease, including cataract (Azuma *et al*, 2000).

1.4.2 Positional Approach

With this method, a chromosomal region which contains the disease gene is defined. Candidate genes are selected from the region for further analysis, so one of the initial aims must be to ensure that the region under scrutiny is as small as possible, containing fewer genes for consideration. The Human Genome Project has generated large amounts of data with genetic and physical maps indicating the position of genes and markers, therefore facilitating this approach.

The Human Genome Project started in 1990. It aimed to provide the sequence of the nucleotide bases in the human genome, and to locate and identify all the genes. The human genome was the first vertebrate genome to be sequenced. In parallel to this, the mouse genome was also sequenced, with a draft genome released in 2002 (Waterston *et al* 2002). A draft human genome was released in 2001 by the

International Human Genome Sequencing Consortium (IHGSC) of 20 groups around the world, estimating that the human genome contained 30,000-40,000 protein coding genes, but had ~15,000 gaps, including ~10% of the euchromatin in the genome (Lander *et al* 2001). In 2004, the IHGSC released Build 35 of the human genome, with ~99% of the euchromatic genome annotated, consisting of 2,851,330,913 nucleotides, with approximately 1 error per 100,000 bases, and 341 gaps, which was a ~475-fold improvement on the draft human genome sequence (International Human Genome Sequencing Consortium, 2004).

There are still heterochromatic areas of the human genome which have not been sequenced, and these were not investigated by the human genome project.

Online databases (Ensembl, NCBI and UCSC Genome Browser, for example) serve as repositories for all this information. Linkage analysis is most often used to map Mendelian disorders to a specific locus. Once this region is defined, candidate genes can be chosen from within the chromosomal region, usually selected on the basis of expression data, phenotypes in animal models resulting from disruption of orthologs, gene function and homology.

1.4.2.1 Genetic Mapping

Genetic and physical maps have been created, and are invaluable to scientists engaged in mapping disease loci and identifying genes involved. There are many examples of these, and over time, more accurate maps have been constructed. The Généthon Map was constructed on the basis of 814 polymporphic (CA)n repeats in three generations of eight families, covering about 90% of the estimated human genome (Weissenbach *et al*, 1992). The Marshfield Map was created using information on almost one million genotypes in eight Centre d'Etude du

Polymorphisme Human (CEPH) families, using over 8,000 short tandem-repeat polymorphisms (Broman *et al*, 1998). The Icelandic deCODE Map was constructed with information from 5,136 microsatellite markers in 146 families, totaling 1, 257 meiotic events, making it immensely more accurate than previous maps (Kong *et al*, 2002). Linkage is the most common tool used for mapping, but some groups have used cytogenetics, such as in identifying rare translocations in Duchene Muscular Dystrophy (Worton and Thompson, 1988). Others have used association studies, for example, for haemochromatosis (Feder *et al*, 1996).

By mapping, researchers, aim to identify a small region of a chromosome that contains the gene of interest. Genome browsers such as Ensembl, NCBI and UCSC show the location of markers and genes along the chromosome. Genes within the region of interest can be selected as candidate genes for further investigation, considering known functions, paralogs, orthologs, and expression data.

In human genetic mapping, researchers discover how frequently loci are separated by recombination during meiosis, and the location of a locus in relation to another. The proportion of offspring who are recombinant (R), as opposed to non recombinant (NR) is represented by θ, the recombination fraction, calculated as follows:

$$\theta = \frac{R}{(R + NR)}$$

An allele is one form of a DNA segment or gene, at a defined chromosomal locus. Say that an individual is heterozygous at both loci of interest, with allele combinations G-1 G-2, H-1 H-2. If the individual's offspring inherit either combination of G-1 H-1 or G-2 H-2, then that child is non-recombinant, but those inheriting G-1 H-2 or G-2 H-1 are recombinant. The recombination fraction is a measure of the

proportion of recombinant offspring between the two loci, in this case, G and H, and never exceeds 0.5. If θ is 0.5, the loci are not linked. The closer two loci are, the less likely it is that there will be recombination to separate them, so θ decreases. When loci coincide, θ=0.

Loci on different chromosomes exhibit independent segregation. Loci on the same chromosome are said to be syntenic, and ignoring crossovers during prophase of meiosis I, would segregate together. In this family scenario, if sperm has allele G-1, there is 50% chance it will also have allele H-1 and 50% chance of also having allele H-2 instead, meaning that the average expected outcome would be for 50% of offspring to be recombinant and 50% of offspring to be non-recombinant.

Meiotic recombination happens less often when two loci are located in close proximity on a chromosome. More recombination is seen when loci are on separate chromosomes or far apart on the same chromosome (Morgan, 1911). Alleles in close proximity which tend to be inherited together are termed haplotypes. Two loci are said to be linked if recombination occurs in fewer than 50% of events. The distance between two loci can be measured by their genetic distance or their physical distance. If recombination between two loci occurs in 1% of meiosis events, they are said to be 1 centimorgan (cM) apart.

Single recombination events create two recombinant and two non-recombinant chromatids. Even though double recombination events can occur between two, three of four chromatids, the net result will still never exceed a recombination fraction of 0.5. It is not possible to arrive at the recombination fraction across a section of a genetic map, over multiple loci, by simple addition. This is achieved by employing

mapping functions. Haldene's function is used when crossovers happen randomly without exerting an influence on each other.

Haldene's function: $w = -\frac{1}{2}\ln(1 - 2\theta)$ or $\theta = \frac{1}{2}[1 - \exp(-2w)]$

map distance = w θ = recombination fraction

ln = logarithm to the base e exp = e to the power of

Crossovers are never entirely random events due to interference, as when chiasmata are formed between chromosomes the formation of another chiasma close by is prevented. Kosambi's function is one mapping function which allows for interference.

Kosambi's function: $w = \frac{1}{4}\ln[(1 + 2\theta) / (1 - 2\theta)]$ or $\theta = \frac{1}{2}[\exp(4w) - 1] / [\exp(4w) + 1]$

Genetic distances and physical distances do not correspond to one another since the physical distance is a measure of the number of bases, often measured in megabases (Mb). Recombination occurs more frequently in females, and more often near the telomere. Other areas in the genome, such as repetitive sequences, are more prone to recombination too, and hence are known as recombination hot-spots. Centromeric regions tend to see fewer recombination events.

1.4.2.2 Linkage Analysis

Linkage refers to the tendency of particular loci to be inherited together. The closer that two loci lie on a chromosome, the greater the likelihood that they will be inherited together during meiosis and escape recombination events which would cause them to be located in different daughter cells. With increasing distance between two loci on a chromosome, recombination events become more likely, and

hence, it is more likely that these loci will not be inherited together. Linkage occurs when recombination happens in less than 50% of events. This principle is used to conduct linkage analysis studies. Quantitatively, linkage analysis is a measure of whether or not θ (the recombination fraction) is significant; that is to say, if it differs significantly from 0.5. If θ=0.5,the two loci are not linked.

Linkage analysis only works well for Mendelian disorders with a single causative gene at one locus. In complex disorders with multiple risk alleles, it loses its power. There must be information on the marker status for at least one of the parents (who will be heterozygous) in order for the marker to be informative in the offspring. The mean heterozygosity (*H*) is usually sufficient to indicate marker informativeness, calculated by subtracting the sum of the square of each allele frequency from 1.

$$H = 1 - \sum_{i} p_i^2$$

It is important to note that members of consanguineous families are more likely to carry homozygous alleles, so heterozygosity (*H*) is decreased. The inbreeding coefficient (F) is introduced, taking into account the proportion of genes which are identical by descent, giving (1-F)*H*.

The polymorphism information content (PIC) also gives information about marker informity, and incorporates the significant proportion of children (0.5) who will be heterozygous offspring of parents who are heterozygous for the same allele combination, p_i being the frequency of the *i*th allele (Strachan and Read, 2001).

$$\mathrm{PIC} = 1 - \sum_{i=1}^{n} 2p_i^2 - \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2\, p_i^2\, p_j^2$$

1.4.2.3 LOD Scores and Critical Values

LOD scores, Z(θ) are a measure of the probability of linkage, and are required to test if θ is significant (Morton, 1955). It is the logarithm of the ratio between two assumptions: the loci are linked, H1 (recombination factor = θ), and the loci are not linked, H0 (θ = 0.5). In some cases, θ can be calculated by counting the recombinations, but only if the parental phase is known.

$Z(\theta) = \log_{10} [L(H_1)/L(H_0)$

$Z(\theta) = \log_{10} [L(\theta)/L(\theta=0.5)]$

θ is calculated at a range of values for each pedigree, and the most likely value for θ is the highest LOD score. LOD scores from multiple families can be combined by addition.

The likelihood ratio can be calculated when information on the phase of the parents is available, where a=those with known meiosis phase, b=non-recombinants and c=recombinants.

$(1 - \theta)^b \times \theta^c / (1/2)^a$

For autosomal loci, linkage is accepted when the LOD score is $Z_{max} \geq 3$, and linkage is rejected when $Z_{max} < -2$ (Morton, 1955). If Z falls between 3 and -2, it is inconclusive. For linkage to be accepted with X linked loci, Z_{max} must be ≥2.The corresponding significance values must not exceed 0.001 ($Z_{max} \geq 3$) and 0.01 ($Z_{max} < -2$). It has been estimated that only once in fifty random selections will two loci be linked purely by chance.

1.4.2.4 Two-Point and Multi-Point Linkage Analysis

LOD scores can be calculated by two-point and multi-point analysis, providing proof of linkage. Two-point linkage analysis is used to discern if one marker and one locus are linked, and is regarded as a way to gauge the location of the locus under investigation. Multi-point linkage analysis maps the disease locus in relation to a number of markers, and provides much more information than two-point analysis, gives a more accurate estimation of the disease locus, and yields higher LOD scores.

1.4.2.5 Markers

A haplotype is a set of linked alleles on a chromosome, and they can be used in mapping. A locus is the defined position of a particular DNA segment on a chromosome. The locus of a disease gene can be pinpointed using markers. Markers are tracked through pedigrees to elucidate inheritance of the markers within the family. For a marker to be informative, it must be polymorphic. The most suitable markers are those which are highly polymorphic in the population, so most people will be heterozygous. As part of the Human Genome Project, over 10,000 markers were overlaid on maps of the human genome (Broman *et al*, 1998). When looking at markers transmitted through the pedigree, they are informative if it is possible to discern whether or not recombination has occurred.

1.4.2.5.1 Restriction Fragment Length Polymorphisms (RFLPs)

RFLPs were the first markers to be developed and used in genetic studies (Grodzicker *et al*, 1974). These refer to different fragment sizes that result from the presence or absence of restriction sites, which cause differing polymorphic lengths of

homologous DNA. Temperature sensitive adenovirus mutations were correlated with differences in restriction fragment lengths and this information was used to place the mutations on a physical map. This technique was developed further to allow mapping of traits to loci without knowing the gene responsible. Restriction enzymes are used to digest DNA into fragments of known sizes which are then detected by hybridising them to radioactive probes using Southern blotting (Botstein *et al*, 1980). This approach is time consuming compared to other available methods, and often a particular locus is uninformative because the restriction enzymes are only able to recognise a specific variant.

1.4.2.5.2 Microsatellites

Microsatellites are used as genetic markers. They are short tandem repeat polymorphisms of a DNA sequence of 1-4 base pairs in length, with the longest extending up to approximately 0.1kb in total length. Tetranucleotide repeats are rarer, but preferable, as they tend to yield more definitive results, whereas dinucleotide repeats are more inclined to yield less robust results. They can be amplified by PCR with specific primers, using a fluorescently labelled forward primer, and the dye is detected during analysis. Many studies have been successful in identifying genes responsible for disorders using microsatellite markers to genotype DNA from affected individuals and family members. Genotyping with microsatellite markers (in combination with results from SNP arrays) was used to identify *FLVCR2* as the gene harbouring mutations responsible for Fowler Syndrome (Meyer *et al*, 2010).

1.4.2.5.3 Single Nucleotide Polymorphisms (SNPs)

SNPs are polymorphic variations at a specific nucleotide base. The aim of two projects, HapMap and 1000 genome project, has been to assemble a catalogue of variation within the human genome which will be made available to the global scientific community as it is generated.

The International HapMap Project (http://www.hapmap.org) began in 2002, aiming to find a minimum of one SNP per 5kb of the human genome, using 270 samples from four different populations: Yoruba of Nigeria, European ancestry in Utah, Han Chinese in Beijing, and Japanese in Tokyo (The International HapMap Consortium, 2007). The Human Genome Project had made sequence information available, and information about many SNPs had been deposited in online databases such as Ensembl and UCSC. By the end of the first phase of the project, ~1.3 million SNPs had been genotyped (The International HapMap Consortium, 2005).

The second phase of the project characterised 2.1 million more SNPs in the same samples, generating a HapMap with 1.14 SNP per kb average, meaning an average of 875bp between SNPs (The International HapMap Consortium, 2007).

A significant strength of the International HapMap project is that it provides high density information on SNPs in the individuals analysed (Collins and Tapper, 2011).

The 1000 genome project commenced in 2008. Itinvolves a consortium composed of the Wellcome Trust Sanger Instiute (Hinxton, UK), Beijing Genomics Institute (BGI) in Shenzen, China and National Human Genome Research Institute (NHGRI) in Maryland, USA, amongst others. It aims to describe >95% of the variation within the

genome with an allele frequency ≥1%, in the regions that can be analysed by existing sequencing technology (The 1000 Genomes Project Consortium, 2010).

The populations studied include Yoruba from Nigeria, Han Chinese in Beijing, European ancestry in Utah, Japanese in Tokyo, Chinese in Denver, British from England and Scotland, and Luhye in Kenya.

For the main project, approximately 2500 samples will be sequenced with 4X coverage. As more data from the 1000 Genomes Project becomes available, researchers will be able to make comparisons with data generated by their own genome-wide association studies to determine, computationally, the likely genotypes of their samples, without physically genotyping them. This will reduce the overall cost of genotyping. In the long run, this data will pave the way for clinical advances linked to individual susceptibility to disease and drug response (Via *et al* 2010).

Even though SNPs are biallelic, their density across the (human) genome makes them invaluable in genetic mapping studies, SNPs can be used for analysis by arrays, allowing large numbers to be included at the same time. SNPs have successfully been used in studies linking polymorphisms with disease risk. SNPs in *TNC* (coding for an extracellular matrix protein) have been associated with atherosclerosis and coronary artery disease (Minear *et al*, 2011).

Genome wide scans, to investigate Mendelian disorders, complex disorders and copy number variants (CNV), can be conducted using SNP arrays. The first of these to be developed was the Affymetrix Genechip® 10K Xba Array (Affymetrix Inc, Santa Clara, CA, USA), which contained 11,555 SNPs, and it is still used by many laboratories today to conduct mapping studies (Hattersley *et al*, 2010). Arrays with greater SNP density have now been developed. The Affymetrix Genome-Wide

Human SNP Array 6.0 contains over 906,600 SNPs and over 946,000 probes for CNV detection.

1.4.2.6 Autozygosity Mapping

Autozygosity mapping refers to genetic mapping of consanguineous families that have an autozygous region inherited from a common ancestor, hence homozygous by descent. The power of homozygosity was recognized by Smith (1953), since if consanguineous families are used, fewer families are usually required to provide proof of linkage. If nuclear families were to be used, large numbers would be required, with many affected siblings, but with rare disorders, it is often the case that a large enough study group cannot be found (Wong *et al*, 1986). This is an extremely useful approach for investigation of the genetic cause of autosomal recessive disorders nowadays, due to the increasing availability of information on the human genome and advances in molecular biology. Once genetic maps had been constructed, this methodology could be utilized, first of all with RFLPs (Lander and Botstein, 1987). Many genes involved in autosomal recessive disorders have now been identified using autozygosity mapping (for example, White *et al*, 2007).

1.4.2.6.1 Consanguinity

The coefficient of inbreeding (F) is the probability an individual will be homozygous by descent at any particular loucs (Wright, 1922). Consanguinity is generally regarded to be a union between second cousins, or more closely related individuals, giving a coefficient of inbreeding (F) of ≥0.156 in their offspring.

The coefficient of relationship (R) is the proportion of genes that a consanguineous couple expect to share by descent from a common ancestor.

Parental relationship	Coefficient of relationship (R)	Coefficient of inbreeding (F)
Siblings	1/2	1/4
Uncle-niece/ aunt-nephew/ double first cousin	1/4	1/8
First cousins	1/8	1/16
Double second cousins	1/16	1/32
Second cousins	1/32	1/64

Table 1.2: *Coefficients of relationship (R) and inbreeding (F).*

In some populations, there is a high incidence of consanguinity, which can usually be explained by perceived sociocultural benefits, including strengthening family ties. Economic considerations are sometimes a factor, as smaller dowries are usually required. It is usually the parents who make the choice of partner for their children. Religious tradition was not often given as a reason to enter into a consanguineous union. Women involved in consanguineous unions often feel that the marriages allow them to be manipulated, reinforcing the hierarchy within the family (Hussain, 1999).

There is a high incidence of autosomal recessive disorders amongst populations where consanguineous unions are prevalent, and this problem has been exacerbated in recent years since health care is improving, and more people survive to childbearing age, and pass on the allele associated with the autosomal recessive disorder.

A five year study was carried out in Birmingham to examine the effects of consanguinity (Bundey and Alam, 1993). 4934 children born in 1986 and 1987 were enrolled, with ~20% of these being of Pakistani origin. 7.9% of the genetic disorders encountered were those in the Pakistani populations, compared to 4.3% in the population of European descent. Only 0.4% of marriages between individuals of European descent were found to be consanguineous. 69% of British Pakistanis in the study were related, and 57% were first cousins.

1.4.2.6.2 Principles of Autozygosity Mapping

Figure 1.4 shows a pedigree of a consanguineous family, in which there are members affected with an autosomal recessive condition (IV:1 and IV:2). These are children from a union between first cousins, III:1 and III:2. They have both received two copies of the mutation from their common ancestor, I:2, which has been passed down through the generations, making them homozygous by descent for the disease gene. The region flanking the disease gene, indicated in black, is also homozygous in IV:1 and IV:2. In successive generations, the size of the autozygous region will decrease due to meiotic recombination. In consanguineous families, most autosomal recessive disease mutations are located within a homozygous region, and these can be identified using markers spaced along the genome (such as microsatellite markers and SNP arrays). When a homozygous segment in affected individuals is identified, candidate genes can be selected and sequenced.

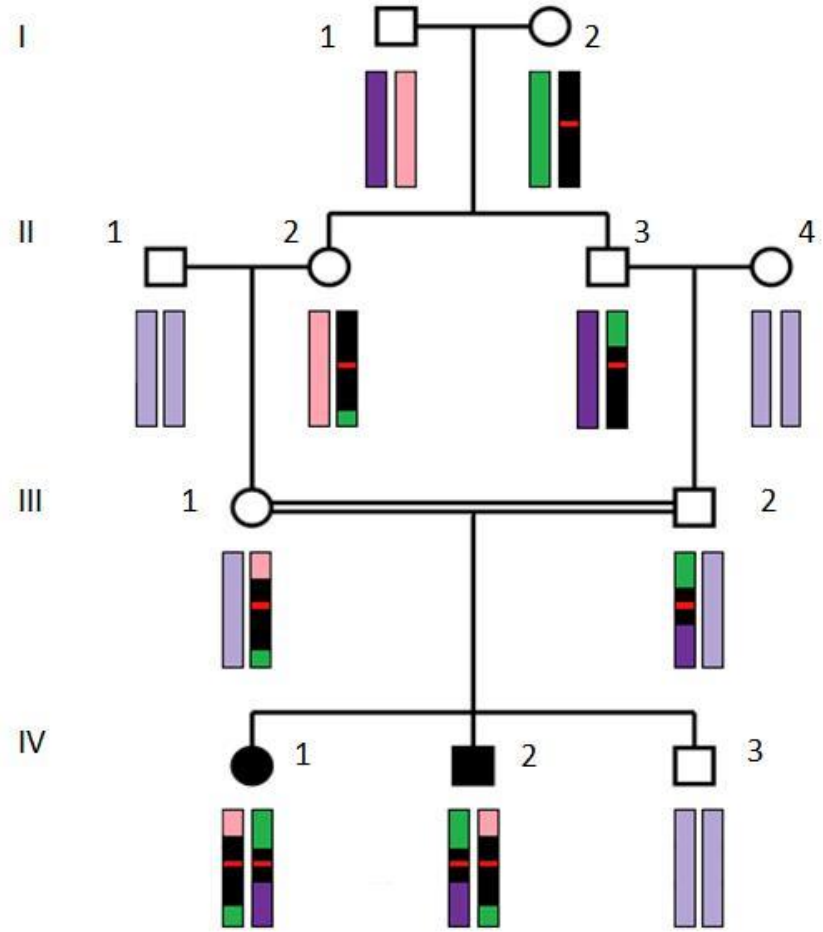


Figure 1.4: *The principle of autozygosity mapping. The red bar indicates a gene harbouring a mutation. Generation IV are products of a first cousin consanguineous union, and demonstrate how a mutation passed on from a common ancestor can cause disease affected status.*

1.4.2.6.3 Mathematical Principles of Autozygosity Mapping

For a child from a consanguineous union, the probability of being homozygous by descent (Fq) is equal to the disease allele frequency (q), multiplied by the coefficient of inbreeding (F), assuming the disease is in Hardy Weinberg equilibrium (Lander and Botstein, 1987). The probability of the child not being homozygous due to two randomly meeting disease alleles is:

$(1-F)q^2$

If the disease gene is in Hardy Weinberg equilibrium in the population, the proportion of affected individuals who are homozygous by descent (α) is:

$$\alpha = \frac{Fq}{Fq + (1 - F)q^2}$$

Assuming disease allele frequency (q) is small in comparison to the coefficient of inbreeding (F), α is more likely to be approximately equal to 1. This equation therefore shows the validity of assumption that for a rare recessive disorder, the disease gene will lie within a homozygous region.

This also explains why children with recessive disorders are likely to have parents who are closely related.The proportion of affected individuals with related parents will be:

$$\frac{Fq}{(1-F)q^2 + Fq}$$

1.4.2.6.4 The Power of Autozygosity Mapping

By using α and F, a LOD score can be generated. The probability of the region flanking a disease gene being homozygous by descent is α. The rest of the genome will be unlinked, and the probability of homozygosity by descent is represented by F. The ratio α:F is the odds in favour of linkage, if genotyping is fully informative. In an individual with an inbreeding coefficient (F) of 1/16 (offspring from a first cousin union or equivalent), with a region that is homozygous by descent, the odds ratio will be 1:16 in favour of linkage, corresponding to a LOD score of 1.204. A significant LOD score of 3.61 would be generated by detecting a shared homozygous region in just three separate first cousin unions, each having one affected child with a shared phenotype, assuming the marker is fully informative and α is approximately 1.

If more distantly related families are used in mapping (inbreeding coefficient decreases), the power to detect linkage increases, as does the maximum LOD score. A maximum LOD score of 2.41 can be obtained from an affected child from a second cousin union. In this case though, the region, homozygous by descent, will be smaller due to meiotic recombinations, so this necessitates more closely spaced markers (Lander and Botstein, 1987).

1.4.2.6.5 Pitfalls of Autozygosity Mapping

There are some considerations that researchers must bear in mind when using autozygosity mapping in disease gene identification.

Some mutations will only be present in one particular isolated community, so it may be impossible to find other families who have the same mutation, or even a mutation in the same gene. This is particularly problematic in the search for new cataract genes since one particular mutation may cause a variety of cataract phenotypes, and also, the same cataract phenotype may be the result of a mutation in one of a number of genes (locus heterogeneity). In these instances, it is more of a challenge for researchers to provide convincing proof of pathogenesis. If possible, a large consanguineous kindred with affected individuals should be obtained, investigated for regions of homozygosity, and then compared with regions of homozygosity in smaller families with the same phenotype. Functional studies are required to confirm this with putative mutations. Autosomal recessive diseases are rare, and it can prove difficult to find enough patients with a matching phenotype. Depending on the population group studied, it may be difficult to find families with more than one affected child due to small overall family sizes (Mueller and Bishop, 1983). In an attempt to surmount this obstacle, international collaborations are often set up.

Not all diseases in consanguineous kindreds are caused by homozygous mutations. Occasionally, disease may arise as a result of compound heterozygous mutations. A genome-wide scan which identifies homozygous regions in the genome will fail to find the causative mutations, as they will not lie in a homozygous region. This was the case in a consanguineous Jordanian Arab family with two boys affected with Karak Syndrome (pantothenate kinase associated neurodegeneration with early onset cerebellar ataxia), where inheritance appeared to be autosomal recessive or X linked, based on the pedigree, but no linkage to *PNAK2* was found (Mubaidin *et al*, 2003). Morgan *et al* (2006) found compound heterozygous mutations in *PLA2G6* in the two boys, selected for analysis because of phenotypic similarities to neuroaxonal dystrophies. The parents had different heterozygous alleles.

In a consanguineous kindred, it is still possible for allelic heterogeneity to occur, so that linkage will not be detected with markers when searching for homozygous regions. An example of this was highlighted in a study of enhanced S-cone syndrome (Miano *et al*, 2000).There were four affected individuals in the kindred, with homozygous mutations, but one of them was a compound heterozygote, so this was missed as the locus due to pooling. Also, autozygosity was detected which had no connection with the disease. The extent of inbreeding was also underestimated, resulting in a higher LOD score, making false-positive linkage more likely.

1.4.3 Exome Sequencing

Until exome sequencing was developed, there had been little progress in identifying Mendelian disease genes, in comparison to the progress made in the understanding of complex disorders, achieved via genome-wide association studies. Linkage studies are useful, but they are not always successful at identifying pathogenic

alleles. High-throughput sequence capture and next-generation sequencing advances have now made exome sequencing more accessible to researchers (Ku *et al*, 2010).

The causative gene for Kabuki syndrome, *MLL2*, was identified using exome sequencing (Ng *et al*, 2010). Massively parallel sequencing was undertaken for ten affected patients, each from a different family, and less stringent criteria were set, so that genetic heterogeneity could be identified. In *MLL2*, seven of the patients harboured nonsense of frameshift mutations, and two of the remaining patients were found to have mutations in *MLL2*.

Exome sequencing involves enrichment and sequencing of the exome (coding regions of the genome), Approximately 1% (30Mb) of the human genome consists of protein coding regions, dispersed in ~180,000 exons (Ng *et al*, 2009). Therefore, exome sequencing has some drawbacks. Non-coding sequences cannot be investigated, and also, not all exons are captured, Whole genome sequencing is currently too expensive for this approach to be used routinely when identifying disease genes.

1.4.4 Progress in Gene Identification

According to OMIM, in June 2011, the molecular basis for 2725 autosomal Mendelian phenotypes are known, along with 236 for X-linked disorders, 4 for Y-linked, and 28 for mitochondrial disorders.

Chapter 2
MATERIALS AND METHODS

2. MATERIALS AND METHODS

This chapter details the reagents and techniques used. Primer sequences are given in the relevant chapters, and further information can be found in the Appendix.

2.1 Materials

2.1.1 Chemicals

10X TBE (1X TBE dilution used)	Geneflow
Acetamide	Sigma
Acrylamide	Geneflow
Agarose (molecular grade)	Bioline
Albumin from bovine serum, minimum 98%	Sigma-Aldrich
Ammonium Persulphate (APS)	Sigma
Ampicillin	Sigma
β-mercaptoethanol	Sigma
BigDye Terminator Cycle Sequencing Kit version 3.1	Applied Biosystems
BioMix^{TM} Red	Bioline
Bio-x-act	Bioline
ColorPlus Prestained Protein Ladder P7711S	New England Biolabs
Competent cells, JM109	Promega
Complete Mini Protease Inhibitor Tablets	Roche

DC Protein Assay Reagents A, B and S	Bio-Rad
dNTPs (working dilution of 2mM)	Bioline
DMSO	Sigma
ECL Plus Western Blotting Detection System, Amersham ™	GE Healthcare
EDTA	Fisher Scientific
Eagle's Minimum Essential Medium (EMEM)	LGC Standards
EcoRI	New England Biolabs
Ethanol	Sigma
Ethidium Bromide	Sigma
ExoSAP-IT	Amersham Pharmacia
FastStart Taq DNA polymerase GCRich	Roche Diagnostics
Genescan-500 LIZ size standard	Applied Biosystems
Hi-Di Formamide	Applied Biosystems
Hi-Spec Additive	Bioline
Hyperladder I	Bioline
IMAGE Clone CDC25A	Source Bioscience
JM109 Competent Cells (high efficiency)	Promega
Kanamycin	Sigma
KpnI	New England Biolabs

LB Broth Medium Powder	Sigma
L-Glutamine	Invitrogen
Lipofectamine^{TM} 2000	Invitrogen
Magnesium Chloride ($MgCl_2$)	Roche
Methanol	Fisher
Micro CLEAN	Web Scientific
Milk (powdered)	Marvel
Mouse monoclonal anti-HA clone (HA-7)	Sigma-Aldritch
OptiBuffer	Bioline
PVDF Membrane	GE Healthcare
Paraformaldehyde (4%, in PBS)	Sigma
Phosphate buffered saline tablets (Dulbecco A) (PBS	OXOID
Polyclonal rabbit anti-mouse immunoglobulin HRP (P0161	DakoCytomation
Prestained Protein Broad Range Protein Ladder	New England Biolabs
Primers	Sigma-Genosys
ProLong® Gold Antifade Reagent	Invitrogen
Propidium iodide	Sigma
Protein Standard BSA (2mg/ml)	Sigma
QIA Prep Spin Mini Prep Kit	Qiagen

Quiagen-tip 500 columns	Qiagen
QIAQuick® Gel Extraction Kit	Qiagen
Quik Change® Lightning Site Directed Mutagenesis Kit	Agilent Technologies
ResGen v10 Mapping Set	ResGen
RNase A	Fermentas
Saran wrap	Dow
SDS (10%)	Fisher
SDS PAGE Tank Buffer (10X) Tris Glycine SDS (B9-0032)	Geneflow
Silver Effieciency Competent Cells	Bioline
SOC Medium	Invitrogen
T4-DNA Ligase	Promega
TEMED	Fisher
TOPRO®-3 iodide	Invitrogen
Tris Glycine Electroblotting Buffer (B9-0056)	Geneflow
TRIS powder (1.5M and 0.5M used)	Fisher
Triton X100 (0.1% in PBS)	Sigma
Tween-20	Sigma
Water, distilled (dH_2O) DNAse, RNAse free	Gibco Invitrogen

2.2 Methods

2.2.1 DNA

Genomic DNA used in this project was already extracted from blood or buccal swabs when received. Most extractions were carried out by West Midlands Regional Genetics Molecular Genetics Laboratory, using Gentra System's Puregene DNA Purification System, which involves salting-out precipitation.Cells are lysed, then protein is precipitated, followed by precipitation of DNA. DNA is washed with ethanol and then hydrated. In this state, DNA can be put into long term storage at -80°C.

2.2.2 Polymerase Chain Reaction (PCR)

A number of PCRs were carried out during the course of this study, with varying conditions. Specific primer details are contained within the appropriate chapters.

PCR allows specific DNA sequences to be amplified, employing a thermostable DNA polymerase, Taq polymerase. Taq polymerase was originally isolated from the organism *Thermus aquaticus*, a bacterium which is found in hot springs (Chien *et al*, 1976), although it was not until the 1980s that Kary Mullis developed PCR as a viable research technique for use in molecular biology.

For PCR, the following reagents are required: sample DNA (to act as a template), buffer, deoxyribonuclease triphosphates (dNTPs), a DNA polymerase (usually *Taq* polymerase), $MgCl_2$ and primers. The concentrations of these can be adjusted to achieve optimum conditions. Similarly, the cycling conditions may be altered to obtain required amplification.

Specific oligonucleotide primers are designed on opposite DNA strands, flanking the region of interest. Primer design is only possible with data on the nucleotide

sequence of the target DNA segment. It is assumed that there is a remote chance that this sequence occurs elsewhere in the genome, even for only one primer, and that there is even less possibility that the other primer's sequence will, by chance, occur nearby. This assumption is always tested using BLAST.

There are three main stages to PCR: denaturation, annealing, and elongation (or extension). Heat is used to denature the template duplexes, breaking the hydrogen bonds that hold the two strands of the double helix together, and giving rise to single strands. A temperature of around 95°C is employed. The temperature is then decreased to allow annealing of complementary primers to the single strands of DNA. The optimum temperature for this annealing step usually falls within the range of 50-65°C. Next, primer extension occurs along the single strands of the double helix, with dNTPs added in a 5' to 3' direction to the nascent strand. This is usually carried out at 72°C, close to the optimum temperature of the Taq polymerase. The steps are repeated numerous times, and the newly formed segments are used as templates in subsequent stages of the PCR. This is how the exponential increase in copies of sequence of interest is achieved. A final extension step is usually carried out at 72°C to allow complete extension of any remaining single strands.

PCR is an incredibly sensitive technique. To maintain accuracy, it is therefore necessary to ensure that there is no contamination of samples and reagents with foreign DNA, so a negative control is always included, which experiences the same conditions as the experimental PCR, and contains all the same reaction components apart from experimental DNA.

2.2.2.1 PCR Conditions

Although it was sometimes necessary to alter the standard procedure, the following general conditions were generally used to achieve satisfactory amplification.

The reactions used BioMix Red, which enables faster set-up since it already includes the polymerase enzyme BIOTAQ DNA polymerase, dNTPs, buffer and 1.5mM $MgCl_2$. A final reaction mixture of 25µl was made including:

2X BioMix Red	12.5µl
forward primer (5.0pmol)	0.5µl
reverse primer (5.0pmol)	0.5µl
dH_20	6.5µl
DNA (20ng/µl)	5.0µl

PCR was carried out on a Tetrad Thermal Cycler using the following conditions:

1. 95°C for 5 minutes
2. 95°C for 30 seconds
3. 57°C for 1 minute
4. 72°C for 1 minute
5. Stages 2-4 carried out 35 times
6. 72°C for 5 minutes

A heated lid condition was programmed to avoid evaporation of reagents.

If these conditions did not yield required product, the annealing temperature and the concentration of $MgCl_2$ could be altered. If there was non-specific priming, 5% [w/v] acetamide or GC rich solution was added to reactions to optimize results. Reactions could also be supplemented with DMSO to inhibit non-specific priming.

2.2.2.2 Primer Design

Primers for this project were on the whole designed using the program Primer3 version 0.4.4 at http://frodo.wi.mit.edu/. The sequence of the exon (or partial exon) of interest was copied from Ensembl (http://www.ensembl.org/index.html) and pasted into the Primer3 programme. Occasionally, it was necessary to design some primers by eye where the regions flanking the target sequence were inconducive to automatic primer design, being highly GC rich. Primers were designed to 18-26bp in length. An annealing temperature between 55-65°C was required, and could be predicted with the following formula:

Annealing temperature = 59.9 + 0.41 x (%GC) – 600/length

Primers were designed to amplify coding exons of genes of interest, including intron-exon boundaries, and were positioned so that the segments amplified were not much over 600bp in length, but were ideally 500bp. Any larger exons were covered with more than one pair of oligonucleotide primers, ensuring that overlap occurred.

2.2.3 Agarose Gel Electrophoresis

After PCR, DNA was separated based on product size by agarose gel electrophoresis. Phosphate groups in DNA are negatively charged, so DNA migrates through the gel medium towards the positive anode. This movement occurs at a slower rate with larger molecules, but is not always entirely dependent on this. Secondary structure can also affect the rate of migration.

1% (w/v) Agarose gels were made with 1.5g agarose powder and 150ml 1X TBE. These were melted in a 600W microwave, then left to cool to approximately 40°C.

Ethidium bromide was then added (3µl). It is an intercalating agent, and interacts with DNA, fluorescing with UV light. In this way, DNA can be visualized.

The gel was cast in a gel casting tray and wells were ensured by the positioning of well combs as the gel solidified, taking approximately 30-60 minutes. Gels were then placed in an electrophoresis tank so that samples could be loaded. No additional loading buffer was required due to the presence of BioMix Red. 5µl of product was loaded in the wells, and 3µl of Hyperladder I was added to the first well to allow size comparison. Hyperladder I features 14 fragments varying in size from 200 to 10,000bp.

The gels were run at 180V for 10-15 minutes. This depended on the size of the gel and the size of the fragment of interest. The gel was viewed on a UV transilluminator, using a wavelength of 260nm. It was photographed with a connected camera.

2.2.4 DNA Purification

PCR products of interest needed to be further purified in order for it to be suitable for sequencing reactions. This was achieved either by using ExoSAP or MicroCLEAN. Primers had to be removed so that only the forward or reverse primer was amplified in subsequent reactions.

2.2.4.1 ExoSAP-IT

ExoSAP-IT was used to clean up the PCR/DNA amplification product. For 5µl PCR product, 2µl ExoSAP-IT and 2µl dH_2O were added. This mixture was incubated at 30°C for 15 minutes, and then inactivated by heating at 80°C for 15 minutes. After this inactivation, sequencing was initiated within 30 minutes.

ExoSAP-IT is a combination of hydrolytic enzymes Exonuclease I and shrimp alkaline phosphatase. Exonuclease I degrades single stranded DNA (primers and superfluous PCR products), and shrimp alkaline phosphatase is responsible for hyrdolysing excess dNTPs.

2.2.4.2 MicroCLEAN

MicroCLEAN is another reagent designed for PCR/DNA clean up, and it was used independently of ExoSAP-IT clean up reactions. An equal volume of MicroCLEAN was added to PCR product (typically 2.5µl of each per forward or reverse primer). This mixture was then centrifuged for 40 minutes at 4000rpm, the supernatant removed, and the pellet re-suspended in dH_2O (5µl per primer reaction) to rehydrate the DNA.

2.2.5 Linkage Studies

During this project, two approaches were used to conduct linkage studies: microsatellite marker analysis, and SNP arrays.

2.2.5.1 Microsatellite Marker Analysis

Linkage analysis, using microsatellite markers was, carried out during the course of this study, with varying conditions, detailed within the appropriate chapters. Microsatellites are tandem repeats of 1-4bp of DNA, usually running for less than 0.1kb, and are markers with a known genomic location. Linkage analysis was performed using fluorescently labelled microsatellite markers and an ABI PRISM 3730 DNA sequencer. The microsatellite markers were part of ResGen v10 mapping set or ABI panels, labelled with dyes FAM (blue), HEX (yellow) and TET (green). The ResGen v10 mapping set consists of 405 microsatellite markers, with approximate

spacing of 9cM between markers. Novel microsatellite markers were also designed using microsatellite data on UCSC (GRCh37/hg19) (RepeatMasker; Variations and Repeats; Microsatellites), and Primer3.The forward primer was labelled with FAM. This was necessary when no known or in-laboratory markers for the region of interest were available. Stock primers for novel markers were made up with dH_2O according to the manufacturer's recommendation, and working dilutions were made up with 20µl forward primer 20µl reverse primer and 160 µl dH_20. A reaction mix of 10µl was prepared including:

2X BioMix Red	5.0µl
Forward primers (2.0pmol)	0.2µl
Reverse primers (2.0pmol)	0.2µl
dH_20	3.6µl
DNA (20ng/µl)	1.0µl

The PCR conditions were an initial denaturing step for 5 minutes at 95°C, followed by 30 cycles (95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds), then 72°C for 5 minutes. PCR products were diluted 1:5 to 1:15 with dH_20 and 10µl of a HiDi and LIZ mixture (1000 µl HiDi with 4 µl LIZ) was added to 1µl of the diluted PCR product. After centrifugation, the plate was run through a denaturing cyle (95°C for 5 minutes), and afterwards, was immediately placed on ice. Once cool, the plate was taken to be put on the ABI PRISM 3730 Sequencer. A negative control, using dH_20 instead of DNA, was included for each pair of primers to detect contamination. The ABI PRISM 3730 uses LIZ as a size standard.

Gene Mapper Software was used to analyse the data gathered from the computer attached to the sequencer. Genescan (Applied Biosystems) was used to identify the size of the products.

2.2.5.2 SNP Chip

SNP chip genome-wide linkage scans were performed by Louise Tee or Fatimah Rahman, using the GeneChip® Human Mapping 250K SNP Array. These microarrays permit analysis of 238,304 SNPs of an individual's DNA. Genotyping with these SNP chips was carried out as per the manufacturer's instructions. 250ng DNA was digested with StyI restriction enzymes (New England Biolabs), then ligated to StyI adaptors with T4 DNA Ligase (New England Biolabs). For each DNA sample, three PCR reactions were carried out, with primer 002 (Affymetrix) and Titanium DNA Amplification Kit (Clontech). A 1.5% agarose gel was used to run PCR products (200-1,100bp). A DNA Amplification clean-up kit (Clontech) was then used. DNA was then fragmented to yield products no larger than 200bp, labelled, and hybridised to the 250K SNP chip. A fluidics station was used to wash and stain the chips. An Affymetrix GeneChip Scanner 3000 with GCOS 1.3 software was used to scan the chips, and GTYPE 4.1 software was consequently employed to identify genotypes at the SNPs and chromosomal location.

Microsoft Office Excel 2007 was used to manipulated the generated data, and to identify regions of homozygosity ≥2Mb. Each chromosome's data was given a separate Excel sheet. The cells with heterozygous calls were filled with pale shading, and the homozygous calls were given two different darker shades. 'No calls' could not be discounted as representative of continuing homozygosity, so they were also shaded, lightly. By eye, the data was scanned (with magnification reduced to 15%), and regions of homozygosity ≥2Mb were noted and used to commence linkage analysis studies.

2.2.6 Sequencing

If a sequence of DNA is known, it can be detected using Direct Sequencing. One type of Direct Sequencing is Sanger Dideoxy sequencing, which uses dideoxynucleotides. Dideoxynucleotides have hydrogen at the 3' prime carbon where the hydroxyl group is in deoxynucleotides. This prevents phosphodisester bond formation between the dideoxynucleotides and nucleotide, terminating amplification. Each of the four dideoxynucleotides can be labelled with a different fluorophore, and therefore can be identified. Sequencing was carried out on an ABI 3730.

A 10µl reaction mix is made up with the following quantities:

Big Dye Reaction Mix	0.5µl
5X Sequencing buffer	2.0µl
Forward or reverse primer (4.0pmol)	2.0µl
dH_20	1.0µl
Purified DNA product	4.5µl

The cycling conditions were an initial denaturing step for 3 minutes at 96°C, followed by 30 cycles (96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes).

Precipitation steps are then carried out, with centrifugation after each addition. 1µl 250mM EDTA is added to sequencing reaction, spun briefly, and then 30µl 100% ethanol is added. This is centrifuged for 20 minutes at 2000rpm. The plate is then inverted and spun gently onto absorbent paper. 90µl of freshly made 70% ethanol is then added to sequencing reaction, and this is centrifuged for 10 minutes at 2000rpm. After this, the plate is again inverted and centrifuged at low speed onto absorbent paper. 10µl of HIDI is added to every well and it is denatured for five

minutes. The plate is then put onto ice immediately to stop the products reannealing. At this point the plate can be taken to the ABI 3730 automated sequencer.

Chromas Software is used to view the sequencing results after the plate has been run on the ABI 3730.

2.2.7 Next Generation Sequencing

Next generation sequencing is used to identify DNA sequences, and is often referred to as massively parallel sequencing. The genome is fragmented, and the fragments of DNA are ligated to adapters, in order to be read randomly during DNA synthesis (Zhang *et al*, 2011). Current next generation sequencing methods only produce short read lengths of 50-500bp, so adequate coverage is essential for accuracy. Coverage refers to the number of overlapping short reads for a specific section of the genome.

For next generation sequencing, samples were sent to Dr Michael Simpson at Guy's Hospital in London. DNA from individuals with cataracts was fragmented and Agilent SureSelect Whole Exome hybrid capture was used to enrich these fragments for exomic sequences. Biotinylated RNA probes of 120bp were used to cleave off DNA from the array, transcribe DNA and incorporate it. 3µg DNA yields a fragment of 150bp which is introduced into a Covaris sample processing tube, with ultra high frequency sound waves. The ends of the strands are repaired (adenine overhangs) with ligation and universal sequence adapters. This is then mixed with probes and left at 65°C overnight, then passed through magnetic beads which hold onto the biotin. DNA is bound to the probes, and eluted. This resulting library is sequenced with 76bp paired end reads across two lanes of the Illumina GaIIx flowcell (v2 chemistry) (Ostergaard *et al*, 2011). Reference genome hg18 was used for alignment. Optical and clonality duplicates were identified and not included in further

analysis. The SamTools software package was used to sort through SNPs and small insertion deletions, and sequence variations were compared with data from the 1000 Genomes Project and dbSNP131 to see if any variants detected by exome sequencing were novel, and hence, potentially pathogenic mutations (Ostergaard *et al*, 2011). At this point, the generated data was received, and novel variants in genes known to be expressed in the eye were selected for further analysis.

2.2.8 Cloning of PCR Products

Cloning was used in this project to obtain the sequence of a gene, *CDC25A*, in isolation, amplify it, and introduce it into cell lines for further investigation, and to allow manipulation of the gene from wild type to a specific mutant form.

2.2.8.1 IMAGE Clone

An IMAGE (integrated molecular analysis of genomes and their expression) clone of full length cDNA for *CDC25A* was obtained, in a pOTB7 vector, with resistance to chloramphenicol.

An incubator was set to 37°C, and an agar plate infused with chloramphenicol was inverted, and placed inside. A sterile inoculation loop was swiped across the top of the agar in the cryovial containing the IMAGE clone. This loop was then traced lightly across the surface of the agar in a zigzag pattern. The inverted plate was incubated overnight. During the morning of the following day, the agar plates were examined for growth of colonies.

20ml falcons were set up with 5ml LB medium and 6.25µl 20mg/ml chloramphenicol. A single, individual colony was removed from the agar plate with a pipette tip and transferred to one of the falcons. The falcons were placed in the incubator and the

rotor was switched on to160rpm. This was left to incubate overnight. A mini-prep (see 2.2.8.4.1) was carried out, followed by gel extraction (see 2.2.8.5).

The required section of DNA was obtained by PCR. A 1X reaction mix with plasmid preps of the IMAGE clone was set up:

Optibuffer	20.0µl
$MgCl_2$ (50mM)	10.0µl
dNTPs	4.0µl
Bio-x-act	8.0µl
Hi-spec additive	40.0µl
Forward primer (2.0pmol)	4.0µl
Reverse primer (2.0pmol)	4.0µl
dH_2O	94.0µl
Plasmid prep of IMAGE clone	2.0µl

The products were loaded onto 1% agarose gel and run for two hours at 80V. The required band was then excised from the gel in a dark room, and gel extraction (see 2.2.8.5) was carried out.

2.2.8.2 Vectors

Vectors are cloning vehicles consisting of DNA, which are used to contain a DNA sequence of interest (the target sequence). DNA sequences for entire genes, or DNA segments, can be inserted via recombinant DNA methods. Plasmid vectors contain double-stranded, circular DNA. This can be cleaved with specific endonucleases. Vectors typically contain many unique restriction sites. Directional cloning is made possible due to the presence of multiple cloning sites. The same two

restriction endonucleases are used to cleave both the target sequence and the vector, and these particular enzymes are chosen based on the sticky ends that will result, ensuring that target DNA can be inserted in the vector in the correct orientation. In order to insert the DNA sequence of interest, the ends are joined with DNA ligase to form a circular DNA loop again. The vectors consisting of recombinant DNA are introduced into cells, to transform them. Vectors have a replicon with an origin of replication, so they can replicate independently of the host cell's chromosomes when the host cell replicates. Progeny cells will contain a copy of the vector. One or more genes in the vector give them a detectable phenotype (usually resistance to a specific antibiotic). This allows individual colonies to be selected when transformed cells are plated out on a solid medium, such as agar, infused with a particular antibiotic. The amount of cells can then be scaled up in liquid culture, and the recombinant DNA harvested.

The *CDC25A* IMAGE clone was supplied with *CDC25A* located within a pOTB7 vector, with resistance to chloramphenicol. During this project, *CDC25A* was ligated into pCMV-HA and/or pEGFPC2 vectors.

2.2.8.3 Antibiotic Plates and LB Medium

16g powdered agar was added to 500ml dH_2O in glass conical flasks. Lids were loosely fastened with autoclave tape, and the bottles and contents were autoclaved. This was allowed to cool but not set. The bottle was infused with 500µl of 100mg/ml ampicillin or kanamycin, and this was aliquoted in 25ml volumes into plates. These were left to set at room temperature with lids ajar, and stored at 4°C long term.

LB medium (luria broth) was prepared in autoclaved flasks. 500ml distilled water and 10g LB broth medium powder were used per flask.

2.2.8.4 Plasmid DNA Purification

2.2.8.4.1 Miniprep of Plasmids

The LB medium was supplied with the CDC25A IMAGE clone, and incubated at 37°C. It was then checked to see if it had become cloudy overnight, indicating growth. 1.6ml medium was transferred (2x 800µl) into eppendorfs and centrifuged at 6.800g for three minutes, lysing the *E. coli* bacterial cultures containing the sequence of interest. To purify plasmid DNA, the Qiagen QIA Prep Spin Miniprep Kit was used, following the manufacturer's instructions. Rounds of centrifugation follow. In the spin column, the silica gel membrane permits adsorption of DNA. Impurities are removed as flow-through. Finally, a sample of purified DNA in water is eluted.

2.2.8.4.2 Maxiprep of Plasmids

In order to obtain sufficient amounts of plasmid DNA for transfection, 1ml bacterial culture was introduced to 500ml LB medium. A Qiagen Maxiprep was carried out as per the manufacturer's instructions, with Qiagen-tip 500 columns, capable of yielding up to 500µg DNA.

2.2.8.5 Gel Extraction

Gel extraction is necessary to obtain clean DNA, so that it can be used for further investigation. Gel extraction was carried out with a Qiagen QIAquick® Gel Extraction Kit. Bands of DNA are excised from an agarose gel under UV light, dissolved in high-salt buffer, adsorbed to the spin column membrane and subjected to rounds of centrifugation in spin columns, using washing to remove impurities, binding again, and elution of DNA in a low-salt buffer. This step is carried out after restriction

enzyme digests (see 2.2.8.6) are used to cut the plasmid, and the products are separated on an agarose gel.

2.2.8.6 Restriction Enzyme Digests

Restriction enzyme digests rely on the ability of restriction endonucleases to cut DNA when they recognise specific sites. This principle works well when attempting to distinguish between wild type and mutant sequences. It is also useful for cloning as it makes ligation possible and ensures that the gene of interest is cloned in the plasmid in the right direction. Both the gene fragment and plasmid must be cut. Restriction endonucleases recognise small sequences of 4-8bp, usually not falling on the axis of symmetry, resulting in DNA with sticky ends. When the recognition site falls on the axis of symmetry, blunt ends are produced. In many cases, two different restriction endonucleases will be used to cleave the vector so a segment of DNA between two sites is removed, and the vector cannot rejoin as the ends are not complementary.

A 1X master mix (30µl) was set up for the first restriction endonuclease, EcoRI, consisting of:

Buffer H	3.0µl
BSA	0.3µl
EcoRI	1.0µl
dH_2O	0.7µl
Gel extraction of gene or plasmid	25.0µl

These were left on a heating block set to 37°C for 90 minutes.

A KpnI master mix (20µl) was set up, 1X consisting of:

Buffer J	5.0µl

BSA	0.5µl
KpnI	1.0µl
dH_2O	13.5µl

and this was added to the 30µl EcoRI digest. These were left on the heating blocks, still at 37°C, for 90 minutes.

Total volumes were loaded on 1% agarose gels, loading 3µl hyperladder I in one lane, and run at 120V for two hours. In a dark room, the bands were excised from the gel and gel extraction (see 2.2.8.5) was carried out.

2.2.8.7 Ligation

DNA ligase is required to form a circular DNA loop again. Vectors and inserts were ligated in the following way. 10µl reactions were set up:

Rapid ligation buffer	5.0µl
T4-DNA ligase	1.0µl
Insert	3.0µl
Vector	1.0µl

Ligation was allowed to occur at room temperature for 10 minutes. In the meantime, polypropylene tubes were chilled on ice, and either JM109 Competent Cells or Silver Efficiency Competent Cells were thawed on ice. The cells were mixed by flicking them, and 100µl transferred to each chilled tube. The ligation reaction mixtures were then added to the appropriate labelled, chilled tube, and mixed by flicking. These were left on ice for 10 minutes. The chilled tubes were taken to a water bath which had been pre-heated to 42°C, and heat-shocked by immersion for 45 seconds taking extreme care not to shake the tubes. Immediately after this, the tubes were placed

on ice for 2 minutes. 900µl chilled SOC medium was added to each tube, and these were then incubated at 37°C in a shaker for one hour. 300µl of incubation was added to an appropriate antibiotic agar plate. When pEGFPC2 vector is used, kanamycin plates are required, and ampicillin plates for pCMV-HA vector. Sterile spreaders are used to plate out the mixture. The plates were returned to 37°C, first with the lids ajar for 20 minutes so that the plates would dry. The lids were then replaced and the plates were inverted and incubated at 37°C overnight. Single, individual colonies were selected from the agar's surface, and grown in LB medium. After this, minipreps (see 2.2.8.4.1) and maxipreps (see 2.2.8.4.2) were carried out.

2.2.8.8 Site-Directed Mutagenesis

The cloned wild type sequence of *CDC25A* must be expressed in a cell line. In order to compare phenotypes of the *CDC25A* wild type and mutant sequences, the mutant must be created as a construct within a plasmid, and expressed in a cell line. This is achieved via site-directed mutagenesis. This can be achieved using Agilent's QuikChange Lightning Site-Directed Mutagenesis Kit. Site-directed mutagenesis involves synthesis of the mutant strand via PCR. The DNA template is denatured, and the mutagenic primers are annealed (possessing the desired mutant sequence), extended, and incorporated with *Pfu*-based DNA polymerase. *Dpn* I is used to digest the template DNA, and then the resulting mutant construct is transformed into competent cells.

2.2.8.8.1 PCR of Site-Directed Mutagenesis

The volume of dH_2O to be added to stock primers (from Thermo Scientific) was calculated, and these stocks were run through a Nanodrop ND-100UV to confirm concentrations and to enable working dilutions of 25ng/µl to be made. A

QuikChange® Lightning Site-Directed Mutagenesis Kit (stored at -20°C), was used, and the manufacturer's instructions were followed. After addition of 1µl QuikChange Lightning Enzyme, PCR conditions were:

1. 2 minutes at 95°C
2. 20 seconds at 95°C
3. 10 seconds at 60°C
4. 3 minutes at 68°C
5. Stages 2-4 cycled 17 more times
6. 5 minutes at 68°C

2.2.8.8.2 Digest and Transformation

2µl DpnI enzyme was added to the amplification reaction, mixed by pipetting, and incubated at 37°C for 5 minutes. XL-10 Gold Ultracompetent Cells were taken from -80°C and thawed on ice. Polypropylene tubes were chilled on ice. 45µl thawed cells were transferred to each chilled polypropylene tube. 2µl β-mercaptoethanol was added to each, incubated on ice for 2 minutes, then 2µl DpnI-treated DNA was added. These were incubated on ice for 30 minutes. SOC medium was warmed by holding the bottle in a water bath heated to 42°C for 90 seconds. The tubes containing cells were pulse-heated in the water bath for 30 seconds. The tubes were then incubated on ice for 2 minutes. 500ml pre-warmed SOC medium was added, and the tubes incubated on a shaker set to 225-250rpm for one hour at 37°C. 250µl was transferred to each appropriate antibiotic-infused agar plate and spread with a sterile spreader. The plates were returned to the 37°C incubator with the lids ajar for 30 minutes, and the lids were then replaced, the plates inverted, incubated overnight. The following day, a single, individual colony was selected, and grown in

LB medium (see 2.2.8.3). Then a miniprep (see 2.2.8.4.1) and maxiprep (see 2.2.8.4.2) were carried out, followed by transfection (see 2.2.9.1.2).

2.2.9 Functional Work

2.2.9.1 Cell Culture

Cell culture experiments were carried out in a Class 2 biological safety cabinet, using plasticware from Starstedt.

Cells were cultured in EMEM, supplemented with10% fetal bovine serum (50ml aliquots stored at -20°C until required), 2% penicillin/streptomycin, and 2% L-glutamine, in an incubator maintained at 37°C with 5% CO_2.

CRL-11421 cell line was selected for functional work because it was derived from human lens cells. The cells came from a lens that was surgically removed from child under 1 year of age, who was undergoing surgery for retinopathy of prematurity. The cells were then transformed with an adenovirus 12 – SV40 virus hybrid, and then this cell line was made commercially available. CRL-11421 is an adherent cell line which grows in a monolayer.

2.2.9.1.1 Cell Storage and Thawing

To store lower passage numbers of cell lines, a T75 flask of cells was trypsinised, and then transferred to a falcon using aseptic technique. The falcons were centrifuged at 750g for 5 minutes. The cells were resuspended in 3.6ml EMEM and 0.4ml DMSO. 1ml volumes were transferred to cryovials, and the cells were frozen slowly in a vaporization chamber. They were then kept at -80°C overnight and then transferred to liquid nitrogen for long term storage.

In order to bring up a viable cell line from storage, cells were thawed inside their cryovials, suspending these in a 37°C waterbath. Cells were then centrifuged at 750g for 5 minutes. The supernatant was removed, EMEM added, and the cells were again centrifuged at 750g for 5 minutes. This supernatant was removed. New EMEM was added, the cells resuspended, and transferred to a T25 flask containing 5ml EMEM, and incubated at 37°C.

2.2.9.1.2 Transfection

Transfection is a nonviral mediated method used for gene transfer. It can be used to introduce vectors into cells, and there are a variety of ways that this can be achieved. Electroporation, calcium phosphate and artificial lipid vesicles are commonly used. In this project, lipofectamine^{TM} 2000 was used to transfect plasmid DNA into cell lines. Cells were grown to approximately 90% confluency in six well plates in EMEM, and then provided with glass cover slips, the medium changed to 2ml opti-MEM, and incubated at 37°C over night. A ratio of 1:3 for DNA to lipofectamine concentration was calculated for wild type DNA sequence, mutant DNA sequence, and empty vector, so that 12µl lipofectamine would be used per well with 4µg DNA. Lipofectamine was incubated at room temperature with opti-MEM (1.5ml opti-MEM and 72µl lipofectamine for a six well plate) for 5 minutes in eppendorfs prior to addition of DNA. DNA (4µg in 250µl opti-MEM for each well) and lipofectamine/opti-MEM (1:3, made up to 250µl for each well) were combined, mixed by inverting several times, and incubated for 20 minutes. 500µl of the resulting complexes were added to each appropriate well, and incubated at 37°C for 18 hours.

2.2.9.1.3 Cell Lysis

Cell lysis is a method which destroys the cell membrane in order to obtain the cell's contents. In this way, it is possible to acquire the protein of interest for analysis.

Six-well plates were used to grow cells to approximately 90% confluency. The media was removed and cells were washed briefly with PBS, using 0.5ml per well, tilting the plate to allow addition via the wells' walls. After removal of the PBS, 50µl RIPA buffer (5ml 1M Tris pH8.0; 3ml 3M sodium chloride; 1ml SDS; 200µl 0.5M EDTA; 5ml 10% deoxycolate; 10ml 10% igepal; all made up to 100ml with H_20; then 10ml transferred to a universal adding one tablet Complete Mini Protease Inhibitor) was added. Cells were removed from the bottom of the wells using a plastic cell scraper and the resulting suspension was transferred to relevant, labelled, 1.5ml eppendorfs for untransfected, wild-type transfected, mutant transfected, and empty vector transfected. The eppendorfs were closed and left on ice for 15 minutes to allow the buffer to work. These were centrifuged at 14,000rpm at 4°C for 20 minutes. The supernatant was transferred to new, labelled eppendorfs. This was stored at -20°C, or kept on ice if protein determination (see 2.2.9.2) was to be carried out immediately afterwards.

2.2.9.2 Protein Determination

This was carried out to quantify the protein content of the lysed cell samples. Serial dilutions of BSA protein standard in lysis buffer were set up along with lysed cell samples and RIPA buffer, using 5µl total volume per well. The standards were diluted as follows:

Concentration (mg/ml)	BSA, 2mg/ml (µl)	Lysis Buffer (µl)
0.4	4.0	16.0
0.8	8.0	12.0
1.2	12.0	8.0
1.6	16.0	4.0
2.0	20.0	0.0

Table 2.1: *Protein determination results.*

These dilutions were set up in duplicate, and lysed cells samples were analysed in triplicate. 1ml Bio-Rad DC Protein Assay Reagent A was combined with 20µl Bio-Rad DC Protein Assay Reagent S, and 25µl was added to each well in use, followed by 200µl Bio-Rad DC Protein Assay Reagent B, and this was left for 15 minutes before reading absorbances with Perkin Elmer Workout 2.5 Software on Perkin Elmer 2030 Multi Label Reader Victor x3 spectrophotometer at A690. This generated a standard curve of standards' absorbances against BSA absorbance, and used to determine protein concentration in lysed cell samples.

2.2.9.3 Western Blotting

Western blotting (or immunoblotting) is used to separate proteins in a polyacrylamide gel, based on size. Detergent is used to solubilise cells and the resulting lysate is separated with SDS-PAGE gel electrophoresis, keeping the proteins denatured. When current is applied, smaller proteins migrate more quickly. After transfer by electroblotting to a nitrocellulose or PVDF (polyvinylidene difluoride) membrane, antibody probes are used to detect the protein of interest. These primary antibodies must be able to react with SDS-solubilised proteins. Non-specific binding to the

membrane is prevented by blocking with proteins (such as non-fat milk in PBST). Labelled anti-immunoglobulin antibodies (labelled with an enzyme, usually horseradish peroxidise, or radioisotope), known as the secondary antibody, are then bound to these antibodies. This cleaves a chemiluminescent molecule, therefore causing luminescence, which is used to detect the protein of interest, and the signal strength is an indicator of protein concentration.

Due to protein concentrations obtained during the course of this project, gels in Western blotting were made up to accommodate 1.5mm 10 well combs. The proteins in the samples, obtained from cell lysis (see 2.2.9.1.3) were separated using gel electrophoresis and 10% denaturing SDS polyacrylamide gel in BioRad Mini-Protean II Electrophoresis Units.

The resolving gel was made up first in a 50ml falcon:

water	4.13ml
30% acrylamide	3.16ml
TRIS (1.5M, pH8.8)	2.50ml
10%SDS	100µl
TEMED	6µl
10% APS	100µl

APS and TEMED cause the gel to set between two glass plates.

The stacking gel was made up in a 50ml falcon:

water	2.94ml
30% acrylamide	660µl
TRIS (0.5M, pH6.8)	1.26ml

X% SDS	60µl
TEMED	10µl
10% APS	80µl

After the resolving gel had set, the stacking gel was poured on top, placing the comb carefully in the unset stacking gel between the glass plates.

Calculations were carried out with results from protein determination to decide the volumes of protein sample and dye to use per well, loading 5µg protein per well. 5X dye was used consisting of bromophenol blue (5mg/ml) and β-mercaptoethanol in a ratio of 1:2. Samples combined with dye were denatured at 100°C for five minutes on a heating block. Alongside the samples, one well was also loaded with 3µl Prestained Protein Broad Range Protein Ladder. The tank was filled with 1X running buffer, and gels were run for one hour at 150V.

The gel was prised from the glass plate and placed onto a piece of PDVF membrane (previously soaked in methanol, briefly, just before use). A sandwich for the transfer was set up in a tray of 1% transfer buffer (transfer buffer consisting of 150ml ethanol, 100ml 10X transfer buffer, 750ml water) as follows: sponge, two pieces of 3mm filter paper, PVDF membrane (cut diagonally at one corner to facilitate recognition of membrane direction), gel, two pieces of filter paper, sponge. Any air bubbles were squeezed out. This sandwich was lifted from the tray, kept in that orientation, and placed onto the clear side of the cassette, and closed. The black side of the cassette was slotted next to the black side of the running tank. An ice pack was also placed in the running tank, and the tank was filled with 1% transfer buffer. The transfer was carried out at 100V for 75 minutes. When the transfer had taken place the membrane was removed from the sandwich and blocked in 5% milk (0.75g milk

powder and 15 ml PBST mixed thoroughly first by vortexing) for at least one hour at room temperature, although this could be done at 4°C overnight.

The PVDF membrane was probed with primary antibodies specific to the protein of interest, diluting the antibody in 5% milk according to the required concentration. The membrane was incubated in primary antibody in a falcon on a tube roller for one hour. The membrane was washed three times, each for 5 minutes in PBST. The secondary antibody was made up to the required concentration (again diluted in 5% milk), and the membrane was incubated in this on a tube roller for one hour. The membrane was washed three times 5 minutes in PBST, and placed on saran wrap with the protein side facing up. ECL Plus Western Blotting Detection System Solution A (1ml) and Solution B (25µl) were mixed and used immediately, dripping the resulting solution onto the membrane. After 1 minute, excess ECL solution was dripped off, and the membrane was trapped in a new sheet of folded-over saran wrap. This was taped down in a hypercassette and taken to a dark room, where exposures of varying time intervals were taken using a SX-101A Developer, exploiting the chemiluminescence of the horseradish peroxidase (conjugated to the secondary antibody), activated by the ECL.

If the membrane was to be reprobed, it was first stripped by boiling it in a beaker of water for 10 minutes, and then stored in a falcon of PBST at 4°C until required, for reprobing with primary antibody.

2.2.9.4 Protein Densitometry

Protein densitometry allows measurement of the intensity of protein staining, and can give a measure of how stable proteins are.

The PVDF membrane was stripped and reprobed using mouse monoclonal anti-HA clone (HA-7) 1:1000 as the primary antibody, and polyclonal rabbit anti mouse immunoglobulin HRP 1:2000 as the secondary antibody.

Pictures of the membranes probed for HA were taken in the dark room, and then again for the same membrane when probed for actin. The images were scanned and the bands of staining were selected on Gene Tools software (Syngene) for spot blot analysis. Each band was compared to the background to discern the intensity of staining. The staining was normalised by dividing the value of the HA staining for the band by the actin staining for the band and multiplying this value by 100, and a graph was generated.

2.2.9.5 Immunofluorescent Staining

Immunofluorescent staining is the prelude to immulofluorescence microscopy. It is process where a protein of interest is stained with an appropriate fluorescent dye so that its subcellular location can be detected with a microscope. During this project, immunofluorescent staining was carried out after transfection of CRL-11421 cells, but cells were not lysed as they were for Western blotting.

The slide mountant, ProLong Gold Antifade Reagent was removed from -20°C at the start of the procedure. Cells on glass coverslips in six-well plates were washed briefly three times in PBS, taking care to ensure that cells did not dry out, and gently so that the cells were not removed from the coverslips. Cells were fixed in 4% paraformaldehyde in PBS for 20 minutes. Cells were washed three times 5 minutes in PBS on a shaker. Permeabilisation of cells was achieved by immersing them in 0.5ml 0.1% Triton X100 in PBS for 3 minutes. After this, cells were again washed three times 5 minutes in PBS. 1% bovine serum albumin in PBS (PBS-BSA) was

used to block the cells, by submersing them in this solution for 30 minutes. The glass cover slips were removed from the plates with forceps and transferred onto a sheet of hydrophobic wax and incubated with 100µl primary antibody for one hour. The required dilution of primary antibody was achieved by diluting stock solutions with PBS-BSA, as per the manufacturers' recommendations. The incubation was covered over to minimise light exposure. Coverslips were transferred to new wells and washed three times 5 minutes with PBS-BSA. During the washes, the required dilution of secondary antibody was prepared according to the manufacturers' instructions. PBS-BSA was used to block the cells, by submersing them in this solution for 30 minutes. The glass coverslips were removed from the plate wells with forceps and transferred onto a sheet of hydrophobic wax and incubated with 90µl secondary antibody for 45 minutes, covering them for the duration of the incubation. Coverslips were washed three times 5 minutes with PBS in new wells, covered with tin foil to exclude light. TOPRO®-3 iodide was diluted 1:1000 in PBS, and the coverslips were incubated with this for 2 minutes, and were then immediately washed three times 5 minutes with PBS, always excluding light whenever possible. 2.5µl ProLong Gold Antifade Reagent (molecular probe) was applied to glass microscope slides (pipette tips were slant-cut due to the reagent's viscosity), towards the centre. The coverslips were inverted, using forceps, and mounted, avoiding bubbles and movement once placed in the mountant. Slides could be left overnight, at room temperature, in the dark, until viewed with a confocal microscope.

2.2.9.5.1 Confocal Microscopy

A Leica Confocal Scanning Light Microscope SP2 AOBS (Leica, Germany), with lens 63 N/A 1.3, optical zoom 4, was used to visualise cells on the slides, and various

pictures of untransfected, wild-type transfected, mutant transfected, and empty vector transfected were taken.

2.2.9.6 Fluorescence Activated Cell Sorting (FACS)

Flow cytometry is used to analyse certain characteristics of microscopic particles, in particular, cells and chromosomes. FACS is a particular flow cytometry method used to separate cells labelled with fluorophores, based on their ability to scatter light and fluoresce. Initially, the cells pass through a fluorescence detector where the intensity is measured. Cells are introduced in a stream of droplets, with vibration adjusted to avoid more than one cell per droplet. An electrical charge can be assigned to each of these at this point (via a charging collar). As the droplets pass through the electrical deflection plates, they are sorted into separate containers, depending on whether they are charged or uncharged.

Cells in six well plates, transfected at least 18 hours before (see 2.2.9.1.2) were taken from 37°C incubation. The trypsin-EMEM media was removed, and the cells were washed with PBS. 200µl trypsin was added and the plate was returned to 37°C for 2 minutes. After this, 2ml media was added to inactivate the trypsin. The transfected cells were removed from the wells with plastic scrapers and transferred to appropriately labelled falcons. Centrifugation was carried out at 1250rpm for 5 minutes at 4°C. The supernatant was removed and discarded. 1ml PBS was added to each tube. The cells were resuspended by pipetting, and centrifuged again. The supernatant was removed again, taking care not to disturb the pellet. 1ml cold 70% ethanol was added drop by drop, whilst vortexing, to each falcon. The tubes were left on ice for 30 minutes. These were then centrifuged at 6000rpm for 5 minutes. The

supernatant was removed, 1ml PBS added, the cells resuspended, and centrifuged again.

30µl PI and 30µl RNase A were added to 3ml PBS. 500µl of this was added to half of the tubes. To the other half, 500µl of 30µl RNase A in 3ml PBS was added. The tubes were left at room temperature in the dark for 30 minutes. They were then processed on the Flow Cytometer, and results were analysed with FlowJo software to produce histograms. Gates were put on FSC (forward scatter) and SSC (side scatter) to separate debris from the selection. FSC corresponds to cell volume, and SSC corresponds to aspects of cell structure. Gates were put on FL3-H vs FL2-A to remove doublets.

Chapter 3
SYNDROMIC CATARACT FAMILIES

3. SYNDROMIC CATARACT FAMILIES

3.1 CEREBROTENDINOUS XANTHOMATOSIS (CTX)

3.1.1 Introduction

Cerebrotendinous Xanthomatosis (CTX; MIM 213700) is a disorder resulting from defective lipid-storage metabolism. It is rare, and inherited in an autosomal recessive manner. The clinical features include juvenile cataracts, tendon xanthomatosis and premature atherosclerosis. Abnormally large cholesterol and cholestanol deposits are found in most tissues during autopsy. Cerebellar ataxia usually manifests as the first stage of the progressive neurological degeneration that is a characteristic of CTX. There is a pseudobulbar stage, foreshadowing imminent death. CTX is clinically heterogeneous. It is often the case that characteristic features such as cataracts may not be present before neurological degeneration becomes apparent.

The first report of CTX was made in 1937, when affected cousins in their 30s were found to have cataracts, tendon xanthoma, xanthelasmata, and mental degeneration (Van Bogaert *et al*, 1937). Symptoms became apparent at ages 12-13. When autopsies were performed, the cousins were found to have deposits in the cerebellum's white matter and the cerebral peduncles.

Mutations in the gene cytochrome P450, family 27, subfamily A, polypeptide 1, *CYP27A1*, on chromosome 2q35, are responsible for CTX (Cali *et al*, 1991b). It is expressed in the brain, lung, liver, skin, small intestine, and numerous other tissues. *CYP27A1* is a nine exon gene, with parts of exon 1 and exon 9 being non-coding, giving a translation length of 531 amino acid residues. Cali's group identified a cysteine residue instead of the wild type arginine at position 446, and at position 362,

in two unrelated CTX patients. *CYP27A1* encodes sterol 27-hydroxylase (CYP27), a mitochondrial cytochrome P-450 enzyme. It is an integral component of the bile acid synthesis pathway. CYP27 converts cholesterol to cholic acid and chenodeoxycholic acid, which is the initial stage in side chain oxidation of sterol intermediates (Cali *et al*, 1991a). Cholic acid and chenodeoxycholic acid form part of a feedback loop, inhibiting cholesterol production, so in individuals with homozygous *CYP27A1* mutations, this leads to increased serum cholestanol levels. This is one characteristic tested for, diagnostically, along with a urine test. Individuals with CTX have higher levels of bile acids in their urine. A diagnosis can be made by showing evidence of elevated serum cholestanol and tendon xanthoma.

Early identification of CTX is vital, since disease progression can be slowed or halted by treatment with bile acids (Salen *et al*, 1975).

Aims: To identify the gene and pathogenic mutation responsible for cataracts in this family.

3.1.2 Patient DNA

A consanguineous family from Bangladesh presented with cataracts and learning difficulties (Figure 3.1). Since learning difficulties are often seen within the general population, the cataracts were initially classed as a non-syndromic anomaly. During this project, a mutation which segregated with disease phenotype was found in *CYP27A1*. This led to further clinical investigation, so that a diagnosis of CTX could be made for affected individuals.

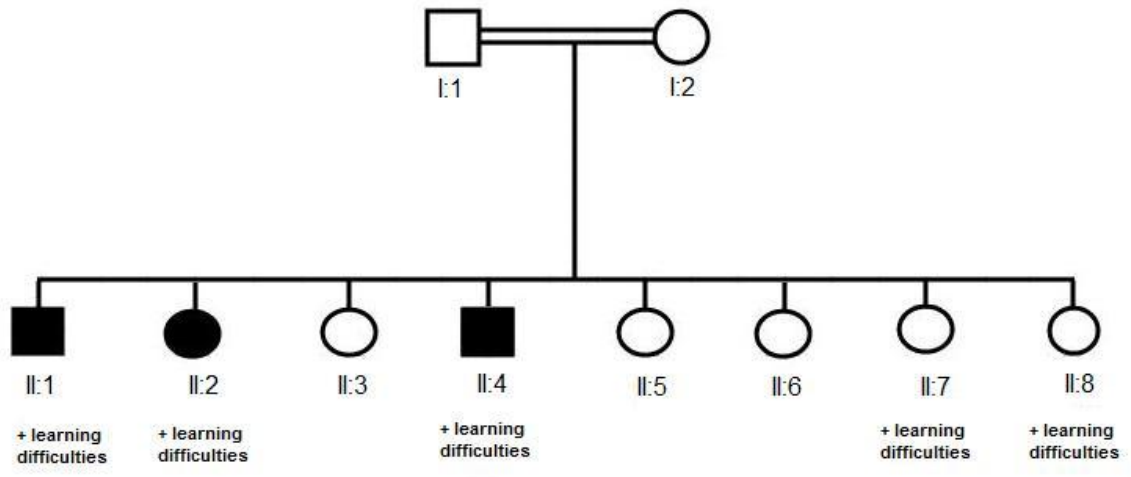


Figure 3.1: *Pedigree of the Bangladeshi family that presented with cataracts and learning difficulties.*

An incomplete medical history was made available. There was no information on the father (I:1). The mother (I:2) had questionable developmental delay, due to not having learnt Bengali or English until she was 3, but this could be due to changing country of residence. She had no cataracts. Affected child II:1 had developmental delay and bilateral pulverulent cataracts. He received lens implants at age 15 (right eye) and age 16 (left eye). His sister II:2 had developmental delay and learning difficulties. She presented with bilateral pulverulent cataracts, and received lens implants at age 14 (right eye) and age 15 (left eye). Affected child II:4 had bilateral pulverulent cataracts and learning difficulties. Visual acuity of 6/18 was noted in the right eye, and 6/9 in the left eye. An orthoptic examination of unaffected sibling II:7 revealed a visual acuity of 6/6 and was reported to have learning difficulties. The youngest sibling, II:8, had global developmental delay, microcephaly, hypotonia, hyperextensibility of the joints, and possibly had learning difficulties. There was no noted evidence of cataracts. The eldest child (not shown on the pedigree) died from gastroenteritis in Bangladesh, aged 1 year. There were three siblings (II:3, II:5 and II:6) without pathological findings. At this stage, it was uncertain if the cataracts and

developmental delay were a single condition with variable expression or if they were the result of different conditions.

3.1.3 Previous Work

DNA was available for all members of the family who were shown in the pedigree. Genome wide linkage analysis was carried out using 250K Affymetrix SNP Microarray, by Fatimah Rahman. Three continuous homozygous regions were found, as listed in the Table 3.1.

Chromosome	Location	Size(Mb)
2	218,994,785-222,589,683	3.6
6	107,406,110-127,463,213	20.0
16	31,511,273-47,017,059 (including centromere)	15.5

Table 3.1: *Homozygous regions detected in the Bangladeshi family.*

Linkage analysis was performed by Dr Esther Meyer using microsatellite markers located within regions of apparent homozygosity.

The 3.6Mb region on chromosome 2 from 218,994,785 to 222,589,683 contained the crystallin gene, crystallin beta A2, *CRYBA2* (219,556,371-219,563,156). Since many other crystallin genes are known to be involved in cataractogeneis when mutated, *CRYBA2* was prioritized as a candidate gene. It was amplified using the primers listed in Table 3.2, and then sequenced. No mutations were found.

3.1.4 Molecular Genetic Methods

Genes in the region on chromosome 2 were prioritised for sequencing if they were known to be expressed in the eye and/or lens, and if they had previously been reported to have a role in cataract formation. After primer optimisation with control DNA, DNA from the unaffected father (I:1) and affected child (II:1) was sequenced in order to identify any pathogenic changes.

3.1.4.1 Prioritised Candidate Genes

3.1.4.1.1 Indian Hedgehog (*IHH*)

IHH (Indian hedgehog homolog) is a three exon gene, the third of which is partially non-coding, giving a translation length of 411 residues. It is found on chromosome 2q33-35, and is expressed in the eye and lens, as well as the colon. It is also expressed in the kidney and liver (Marigo *et al*, 1995). There are no reports of *IHH* mutations causing cataracts, but there have been reports of involvement in brachydactyly (Gao *et al*, 2001) and acrocapitofemoral dysplasia (Hellemans *et al*, 2003). IHH has a role in cartilage differentiation, acts in the hedgehog signalling pathway, cholesterol binding, cell-cell signalling and cellular maturation.

Primers were designed and optimised for all three exons (Table 6.2), but were not sequenced due to results obtained when sequencing *CYP27A1* (discussed below).

3.1.4.1.2 Inhibin Alpha (*INHA*)

INHA (inhibin, alpha) is a two exon gene on chromosome 2q33-q36, with non-coding sections of both, having a translation length of 366 amino acids. It is expressed in the eye, brain and other tissues. There is no known association between cataracts and *INHA*. This gene has been demonstrated to play a role in tumour suppression in

poorly differentiated prostate tumours. *INHA* promoter hypermethylation and loss of heterozygosity in 42% of prostrate carcinomas implicates this gene in tumours (Schmitt *et al*, 2002).

Primer (Tabe 6.3) optimisation was carried out with control DNA, but family DNA was not sequenced due to *CYP27A1* results.

3.1.4.1.3 Tubulin Alpha 4a (*TUBA4A*)

The tubulin alpha-4A gene, *TUBA4A*, is located on chromosome 2q36.1, spanning 2853kb of DNA. There are three transcripts, all sharing a 223bp second exon and 149bp third exon, with a non-coding first exon. Two transcripts have 4 exons, and the other has five, containing the same sequence, but split over two exons, with a 4 bp intron. It is expressed in the eye, as well as the brain, testis and tonsils. There is a pseudogene, *TUBA4B*, on 2q36.

The tubulins are a family of genes, mostly comprised of pseudogenes. They make up microtubules, which are then involved in movement and mitosis (Lewis and Cowan, 1990).

All exons were amplified with control DNA, using the primers and conditions listed in Table 6.4.

3.1.4.1.4 Wingless-type MMTV Integration Site Family Member 10A (*WNT10A*)

The wingless-type MMTV (mouse mammary tumour virus) integration site family, member 10A gene, *WNT10A*, is located on chromosome 2q35, consisting of four exons. Expression has been noted in the lens, eye, prostate and placenta.

WNT gene family products are signalling proteins, many of which are said to be oncogenes. WNT10A overexpression has been noted in human carcinoma lines, for example in HL-60, derived from promyelocytic leukaemia (Kirikoshi *et al*, 2001).

A *WNT10A* mutation has also been associated with odono-onycho-dermal dysplasia. A premature stop codon (Glu233X) was detected in affected members from three families, causing the protein to consist of 232 amino acids instead of 417 (Adaimy *et al*, 2007).

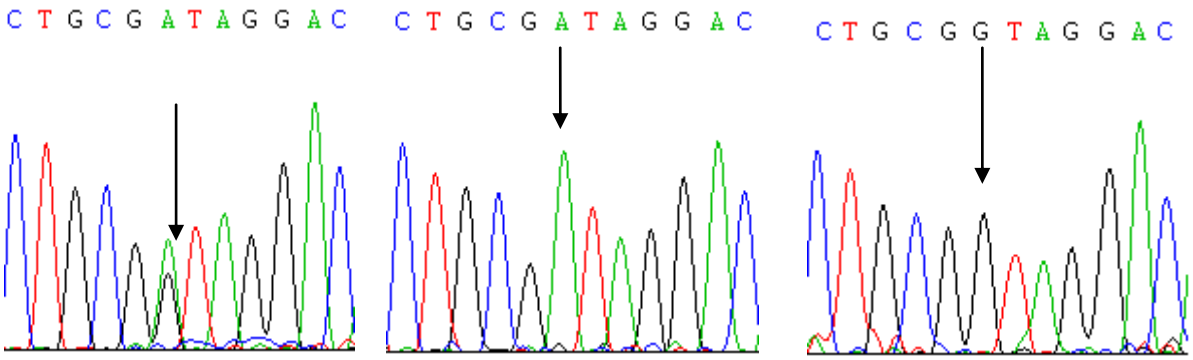
The first exon could not be amplified using the primer pair WNT10A-1F and 1R. All other exons were amplified with control DNA (Table 6.5).

3.1.4.1.5 Cytochrome P450 Family 27 Subfamily A Polypeptide 1 (*CYP27A1*)

This gene lies within the region identified through linkage analysis on chromosome 2. It was selected as a suitable candidate gene due to its known involvement in CTX, and hence cataractogenesis, when mutations are present. This gene was amplified by the primers listed in Table 3.7.

3.1.5 Results

CYP27A1 was considered to be the best candidate gene, due to its known involvement in cataractogenesis in CTX. Since a mutation, which segregated with cataract status, was identified, sequencing of the remaining candidate genes was not necessary. Direct sequencing of *CYP27A1* revealed a substitution of a guanine to an adenine at the first base in intron 6 (c.1184+1G>A, NM_000784.3; Figure 3.2). The mutation was predicted to affect a splice site and therefore the precise consequence of this change on protein level is difficult to establish.



A. Father (I:1) *B. Affected child (II:1)* *C. Control*

Figure 3.2: *Sequencing results illustrating the splice site mutation in intron 6 of CYP27A1.A. The father (I:1) is heterozygous A/G. B.The affected child (II:1) is homozygous G/G for the mutation. C. The sequence for the control is G/G.*

The other family members were then sequenced and the results are shown in Table 3.2.

Family Member	Cataracts	Genotype
I:1 (father)	no	A/G
I:2 (mother)	no	A/G
II:1	yes	A/A
II:2	yes	A/A
II:3	no	G/G
II:4	yes	A/A
II:5	no	A/G
II:6	no	G/G
II:7	no	G/G
II:8	no	A/G

Table 3.2: *Results of sequencing of CYP27A1 gene in Bangladeshi family.*

The mutation was not detected in 454 ethnically matched control chromosomes, which all showed wild type G at the intronic splice site.

3.1.6 Discussion

A family was referred for analysis, described as a non-syndromic cataract family. Cholestanol tests had not been done on the family by clinicians at this time. Molecular genetic analysis results indicated that a diagnosis of CTX may be

accurate. Cholestanol testing was carried out for some of the family members, with the results summarised in Table 3.3, confirming a diagnosis of CTX for affected members.

Family Member	Cataracts	Genotype	Cholestanol Level
II:1	yes	A/A	129
II:2	yes	A/A	62
II:4	yes	A/A	79
II:5	no	A/G	6
II:6	no	G/G	6

Table 3.3: *Cholestanol test results.*

The results for affected children II:1, II:2, and II:4 are classed as high, whereas the results for unaffected children II:5 and II:6 are normal cholestanol levels.

A possible effect of the mutation c.1184+1G>A is the insertion of intron 6 or the splicing out of exon 6, therefore leading to a change in protein sequence.

Due to the absence of alleles with genotype A in the first base in intron 6 of *CYP27A1* in all 454 ethnically matched control chromosomes, it is even more likely that the change is pathogenic, and causes CTX within this family.

A number of other mutations in *CYP27A1*, associated with CTX, have been described, including substitution, deletion, and splice site mutations. Shiga *et al* (1999) identified a splice site mutation in a Japanese family with CTX. The 39 year old male proband with first cousin parents presented with tendon xanthomas, ataxia, and mental retardation. His serum cholestanol level was 37.5µg/ml. A sequence change from G to A at the first nucleotide in intron 7 was detected, causing skipping of exon 7. This was observed in patient cDNA, where exon 8 directly followed exon 6. The proband was homozygous for this mutation, whereas all other family members were heterozygous. Transcripts from blood leukocytes were analysed,

revealing truncated transcripts alone in the patient, while the unaffected silblings had normal and truncated transcripts.

It has been noted that the first two nucleotides at splice donor sites (GT) are highly conserved amongst many eukaryotes (Padgett *et al*, 1986), and mutations at this location can disrupt splicing, since the 5' GT sequence facilitates cleavage and 3' splice site exon joining (Aebi *et al*, 1986).

The Bangladeshi family is an important clinical case because it was referred as a non-syndromic cataract family, but has been shown to be CTX, of which, cataracts are a feature. CTX manifests with variable phenotype, so clinicians should test for mutations in this gene to confirm or exclude CTX diagnosis in individuals that present with congenital and juvenile cataracts, especially those who also have learning difficulties. It highlights the importance of evaluating if CTX is an appropriate diagnosis since the condition can be treated if identified before irreversible brain damage. This finding shows the validity of autozygosity mapping and positional candidate gene identification to find causative mutations, even when the aetiology of the condition and diagnosis are uncertain.

3.2 MARINESCO-SJÖGREN SYNDROME (MSS)

3.2.1 Introduction

3.2.1.1 Marinesco-Sjögren Syndrome (MSS)

Marinesco-Sjögren Syndrome (MSS; OMIM #248800) is a rare autosomal recessive condition. It is of particular relevance to this study, as sufferers present with congenital bilateral cataracts. A diagnosis is made based on clinical symptoms and

brain MRI. Its mandatory features include cerebellar ataxia as a result of cerebellar atrophy, demonstrating decreased numbers of Purkinje fibres and granular cells. Muscle is gradually replaced by fat and connective tissue. Fibre size varies. There are atrophic and necrotic myofibres, rimmed vacuoles, and autophagic vacuoles with membranous whirls (Anttonen *et al*, 2005). Most MSS sufferers have dysarthia, short stature, developmental delay and mental retardation. Other symptoms include skeletal abnormalities, such as pigeon chest, and hypergonadotropic hypogonadism. Serum creatine kinase levels are often elevated.

Under the electron microscope, lysosomes can be seen to have whorled lamellar or amorphous inclusion bodies (Walker *et al*, 1985).This was observed in four affected individuals, from two nonconsanguineous families, one of Sicilian and Irish descent, and a family consisting of a “nonconsanguineous black couple” and their offspring. When conjunctival biopsies were conducted, elevated numbers of lysosomes in fibroblasts were found in six patients displaying typical MSS symptoms of congenital cataracts, mental retardation, delayed motor development and cerebellar ataxia (Zimmer *et al*, 1992). Vacuolar degeneration and abnormal mitochondria were also noted.

MSS has been linked to mutations in the gene SIL1 homolog endoplasmic reticulum chaperone, *SIL1*, on chromosome 5q31. Using DNA from a Finnish family with individuals with MSS, a 3.52Mb region on chromosome 5 was identified, which was then narrowed down to a 1.98Mb region between D5S500 and D5S2116 when haplotype analysis was conducted with two Swedish affected individuals (Anttonen *et al* 2005). This group identified a homozygous four-nucleotide duplication (c.506_509dupAAGA) in exon 6 of *SIL1* in the Finnish individuals, and the Swedish individuals were found to be compound heterozygotes, having the duplication found

in the Finnish individuals, and an intron 6 donor splice site mutation (c.645+2T>C). RT-PCR showed the duplication transcript to be of the expected 461 residues, but there were two shorter transcripts as a result of the splice site mutation. The shortest demonstrated higher expression, and a 64 amino acid in-frame deletion. The duplication was again detected as a homozygous mutation in a Norwegian family by this group, segregating with disease phenotype. In 96 Finnish controls, one carrier of the duplication mutation was detected. The group noted that exon 6 of *SIL1* is serine-rich, and that this could be an indication of encoded phosphorylation sites (Anttonen *et al*, 2005).

3.2.1.2 *SIL1*

The protein coded for by *SIL1* is a 54kD ER glycoprotein nucleotide exchange factor for HSP70 chaperone HSPA5. *SIL1* (SIL1 homolog, endoplasmic reticulum chaperone (*S. cerevisiae*) gene), located on 5q31, spans 346.83kb. It is ubiquitously expressed (Senderek *et al*, 2005). SIL1 was cloned via yeast two-hybrid screening of a liver cDNA library and database screening, and found to contain an N-terminal ER targeting sequence, two putative N-glycosylation sites, and a C-terminal ER retention signal (Chung *et al*, 2002).

Not all individuals diagnosed with MSS are found to have a *SIL1* mutation, perhaps demonstrating its genetic heterogeneity, although non-coding mutations could be causative.

3.2.1.3 Cataract, Congenital, with Facial Dysmorphism and Neuropathy (CCFDN)

CCFDN is considered to be distinct from MSS, even though there are many common features. Merlini's group looked at two members of a consanguineous Italian Gypsy kindred with MSS (Merlini *et al*, 2002). They had congenital cataracts, delayed motor development, ataxia. One patient also had acute recurrent myoglobinuria. This report argued that MSS with myoglobinuria, peripheral neuropathy, and CCFDN result from a common founder mutation. Using haplotype analysis with the Italian Gypsy kindred and a German family with MSS described elsewhere (Muller-Felber *et al*, 1998), Merlini's group showed segregation of MSS with demyelating neuropathy and myoglobinuria with the CCFDN region.

CCFDN causative mutations have been found in the *CTDP1* gene (C-terminal domain of RNA polymerase II subunit A, a phosphatase of, subunit 1) on chromosome 18q23. It codes for FCP1, a protein phosphatase, which is a vital part of eukaryotic transcription machinery. A C>T substitution in intron 6 of CTDP1 caused aberrant splicing so that an Alu insertion was present in processed mRNA, and the T allele segregated with disease phenotype in 52 families with 85 affected members (Varon *et al*, 2003).

3.2.2 Patient DNA

DNA was available from two families with MSS.

Family 1 is a consanguineous Pakistani family (Figure 3.3). The two identical twin daughters, II:1 and II:2, have cerebellar hypoplasia, and developed cataracts in the

fourth year of life. Their sister II:3 is unaffected, and the youngest brother II:4 presented with cerebellar hypoplasia, but is yet to develop cataracts.

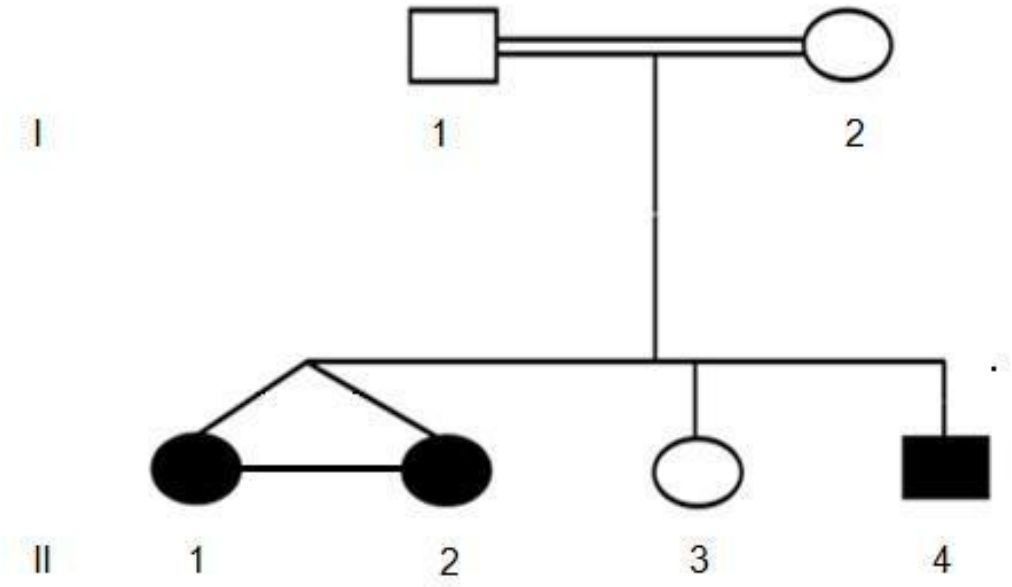


Figure 3.3: *Pedigree of Pakistani family 1 with MSS.*

Family 2 is of Pakistani origin (Figure 3.4). The proband, the eldest daughter II:1 presented at Birmingham Women's Hospital at eight months of age, and was found to have a structural brain anomaly, Dandy Walker variant, after MRI. She had bilateral cataracts, hypoplasia of the cerebellum, hypoplastic cerebellar vermis, global developmental delay, learning difficulties, convergent squint and nystagmus. The cataracts were described as "rapidly progressing intumescent cataracts". The right lens was removed and replaced with an intraocular implant, and a dense white cataract was observed in the left eye. Ultrasound was used to determine that there was no retinal detachment.

The brother, II:3, presented with hypoplasia of the cerebellum and hypoplastic cerebellar vermis. His lenses were clear but a sub-clinical lens change was noted.

He had hypermetrophic astigmatism, gross motor delay, nystagmus and a flattened occiput. In the brain, there was evidence of atrophy of the posterior fossa contents.

Due to these clinical observations, a “likely” diagnosis of MSS was given.

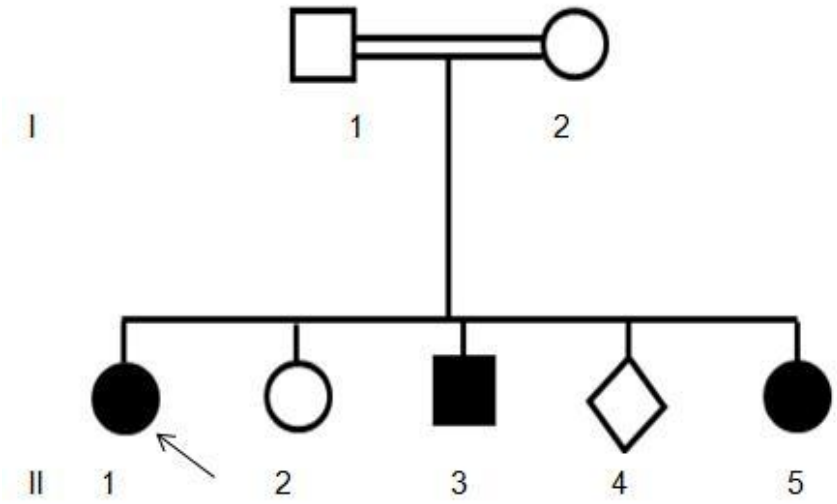


Figure 3.4: *Pedigree of Pakistani family 2 with MSS.*

3.2.4 Molecular Genetics Methods

Since *SIL1* mutations are associated with MSS, this gene was selected for initial molecular genetic analysis of DNA from all members of family 1 and family 2. Table 3.10 lists primers which were designed, and the PCR conditions used to amplify the coding exons of *SIL1* in Families 1 and 2. Sequencing was then carried out, as detailed in Chapter 2.

3.2.5 Results

Previous work carried out in the Birmingham Laboratory did not demonstrate linkage to the region on 5q31 in family 1, using microsatellite markers. In this project, the coding exons were amplified, but no mutations were found in the coding regions of

SIL1 in Family 1, highlighting the genetic heterogeneity of MSS. SIL1 interacts with HSPA5 (Anttonen *et al*, 2005), so markers flanking *HSPA5* on chromosome 9 were chosen, from 126,928,403 to 136,324,922. No common homozygous regions for affected individuals were found (Figure 3.5). It is unlikely that mutations in *HSPA5* are responsible for the apparent MSS symptoms in this family.

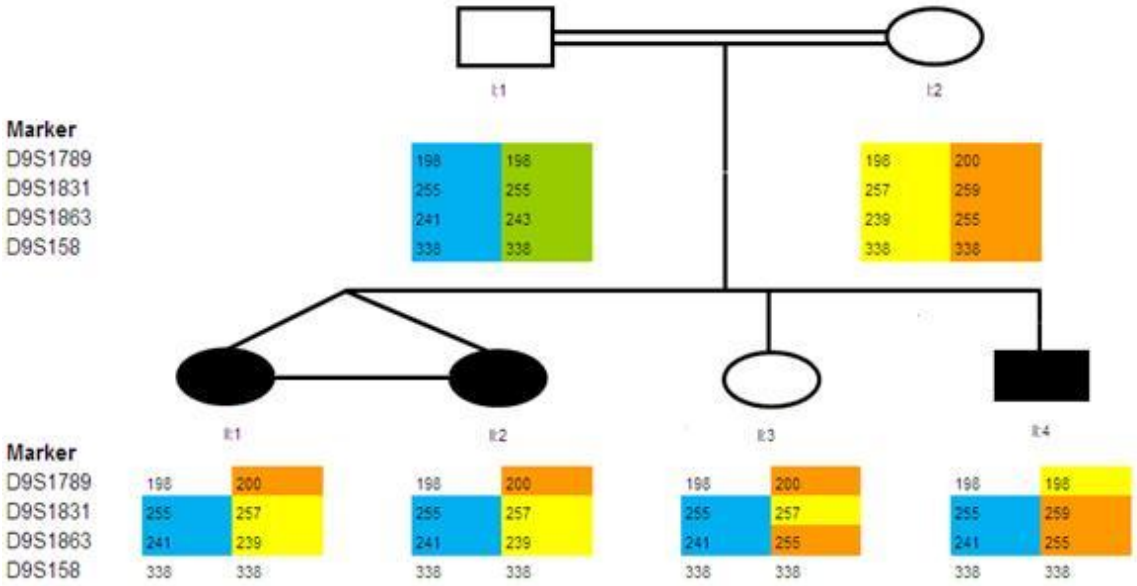


Figure 3.5: *Linkage analysis for microsatellite markers flanking HSPA5 on chromosome 9*

Since mutations in *CTDP1* (18:77,439,801-77,514,510 on GRCh37) have been implicated in CCFDN, a syndrome where symptoms are very similar to those of MSS, in-lab microsatellite markers near *CTDP1* were analysed, from 75,615,199 to 75,615,267. D18S1095 (at 75,319,246) and RH55591 (at 75,538,273) could not be optimised, and other markers selected for analysis were non-informative (Figure 3.6).

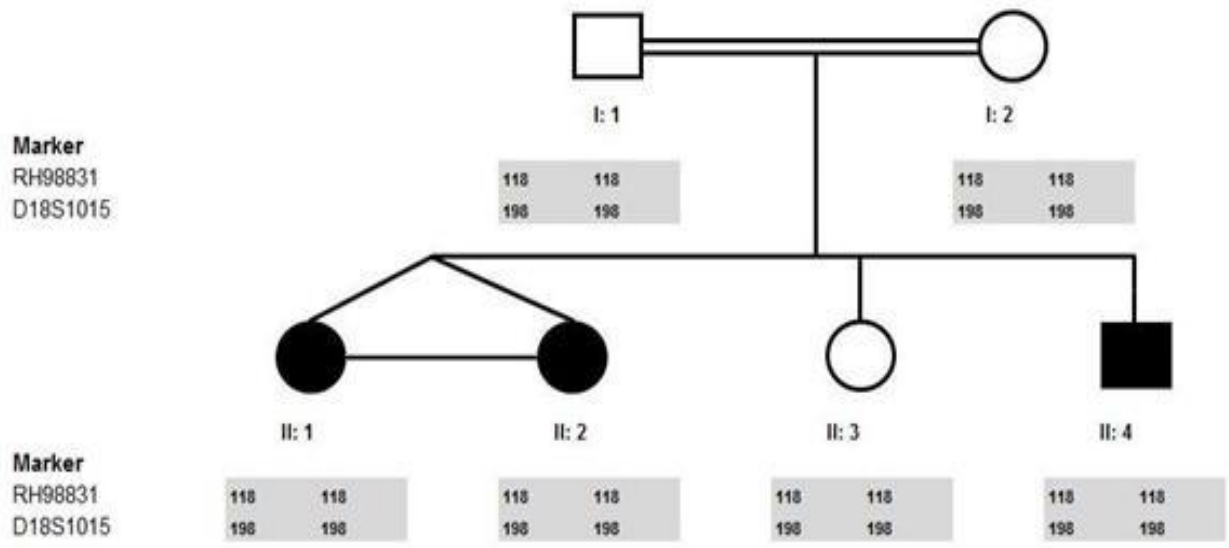


Figure 3.6: *Linkage analysis for microsatellite markers in the vicinity of CTDP1 on chromosome 18.*

In Family 2, a mutation was found when sequencing exon 11 of *SIL1*. There was a C>T substitution (c.1126C>T, NM_001037633.1; Figure 3.7), causing glutamine to be substituted for a stop codon (p.Gln376X), predicted to result in the loss of 86 amino acids.

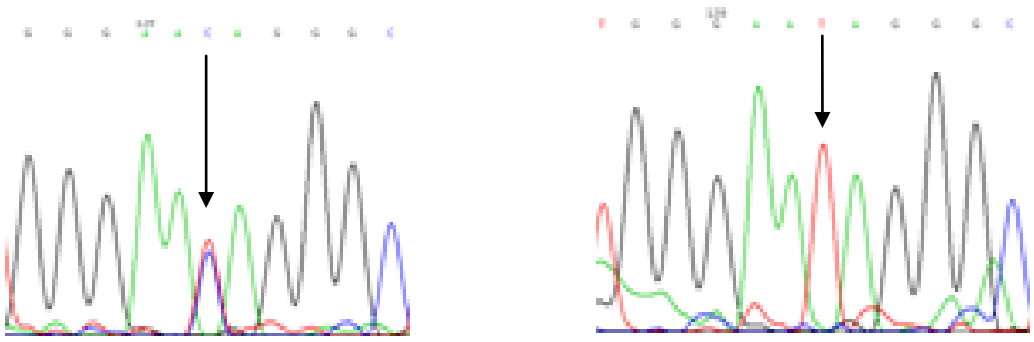


Figure 3.7: c.1126C>T in exon 11 of *SIL1*. *On the left, is the sequencing result for the mother of the family, showing that she is heterozygous for both the T and C allele. On the right is the sequencing result for the eldest affected daughter, showing that she is homozygous for the mutant T allele.*

Controls were not sequenced for this change in this family because mutations in *SIL1* are known to be responsible for MSS, and this is a nonsense mutation.

3.2.6 Discussion

A mutation in *SIL1* was identified in Family 2. Since the stop codon will cause the loss of 86 amino acids, it is likely that this is the mutation responsible for the MSS in affected individuals. Other nonsense mutations in SIL1 have been described. Arg111Ter has been noted in an Iranian family and a Turkish family (Senderek *et al,* 2005) . This mutation was also found in another Turkish family with MSS, where affected individuals had cataracts from 4 years of age (Anttonen *et al*, 2005).

In a Japanese family with three members affected with MSS, Takahata *et al* (2010) identified two novel mutations. One mutation was a 5bp homozygous deletion, del598-602(GAAGA), in exon 6 of all affected individuals. The father demonstrated heterozygosity for this mutation, but sequencing of the mother's DNA reveal this mutation. A ~58kb deletion in exon 6 was detected in the mother and the three affected individuals by array comparative genomic hybridisation and quantitative PCR. The mother was hemizygous for this deletion. All three affected individuals had both deletion mutations. With this in mind, some cases of MSS with unclear genetic basis after sequencing coding regions of SIL1 might not be explained by locus heterogeneity. Instead, deletions may be responsible.

Intronic substitutions have been associated with MSS, but so far, only alterations within the splice site have been found. The substitution c.1029+1G>A causes exon 9 skipping in affected members of a Bosnian family, and c.645+1G>A causes exon 6 skipping in a Turkish family (Senderek *et al*, 2005).

It is possible that a dominant intronic mutation is responsible for MSS in Family 1, even though to date there are no reports of dominant causative mutations, and extensive analysis with PCR and sequencing would be required to confirm or exclude this possibility.

3.3 KNOBLOCH SYNDROME

3.3.1 Introduction

This syndrome was first described in 1971 (Knobloch and Layer, 1971). It is a rare autosomal recessive disorder showing clinical heterogeneity, characterized by high myopia, vitreoretinal degeneration with retinal detachment, and occipital encephalocele. Other features accompanying the syndrome can include cataracts, subluxation of the lens, atypical palmar creasing, scalp defect, lung hypoplasia, flattened nasal bridge, joint hyperextensibility, unilateral duplicated renal collecting system and cardiac dextroversion (Cook *et al*, 1982, Czeizel *et al*, 1992, Passos-Bueno *et al*, 1994, Suzuki *et al*, 2002). Affected individuals experience progressive and irreversible loss of vision. Intelligence is normal.

Knobloch Syndrome was first noted in a family with 5 affected siblings (Knobloch and Layer, 1971). Although encephalocele is considered a characteristic feature of the syndrome, in some cases, congenital occipital scalp defects can be seen (Seaver *et al*, 1993).

Homozygosity mapping in a Brazilian family allowed a locus for Knobloch Syndrome to be mapped to a 4.3cM region on chromosome 21q22.3 (Sertie *et al*, 1993). A

splice site mutation was identified in the gene collagen type XVIII alpha 1, *COL18A1*, in this Brazilian family and in families from Hungary (Menzel *et al*, 2004). In the Hungarian families, one mutation was a 1bp insertion leading to a frameshift and premature stop codon, and the second was an amino acid substitution (Asp104Asn) of a conserved aspartic acid residue. Mutations in the *COL18A1* gene giving rise to Knobloch Syndrome are said to cause Knobloch Syndrome Type 1 (KNO1). At present, 17 different *COL18A1* mutations have been described in 19 families.

A second locus for Knobloch Syndrome was reported for a family from New Zealand who presented with features of the syndrome, but did not show linkage to *COL18A1* (Menzel *et al*, 2004). This unidentified locus has been designated KNO2.

Aims: To prioritise genes in the region between D17S1532 and D17S798 for sequencing, in order to identify any pathogenic mutations.

3.3.2 Patient DNA

A consanguineous Pakistani family shown in the pedigree presented with features associated with Knobloch Syndrome, at the Al-Shifa Trust Eye Hospital, Rawalpindi, Pakistan (Figure 3.9). Peripheral blood samples were obtained, after informed consent, from all family members. The standard phenol chloroform extraction procedure was used to obtain genomic DNA. A childhood history was obtained from parents of affected children. Affected individuals displayed high myopia with severe nystagmus, vitreoretinal degeneration and occipital scalp defect (Khaliq *et al,* 2007). At birth, the scalp defect was described as a swollen purple-red hairless area, and over time, this decreased in size and became firmer. Nightblindness became apparent between ages 2 to 4 years, degenerating to total blindness between 17 to 18 years. Three patients (V:1, V:7 and V:8) did not show occipital encephalocele on

brain imaging scans. There was mild attenuation of the retinal vasculature. Choroidal sclerosis with chorioretinal atrophy was observed. Cataract, lens subluxation, glaucoma and phthisis bulbi were also noted. Unlike classic Knobloch Syndrome, no retinal detachment was noted in this family, and the scalp defect was the only extraocular anomaly observed.

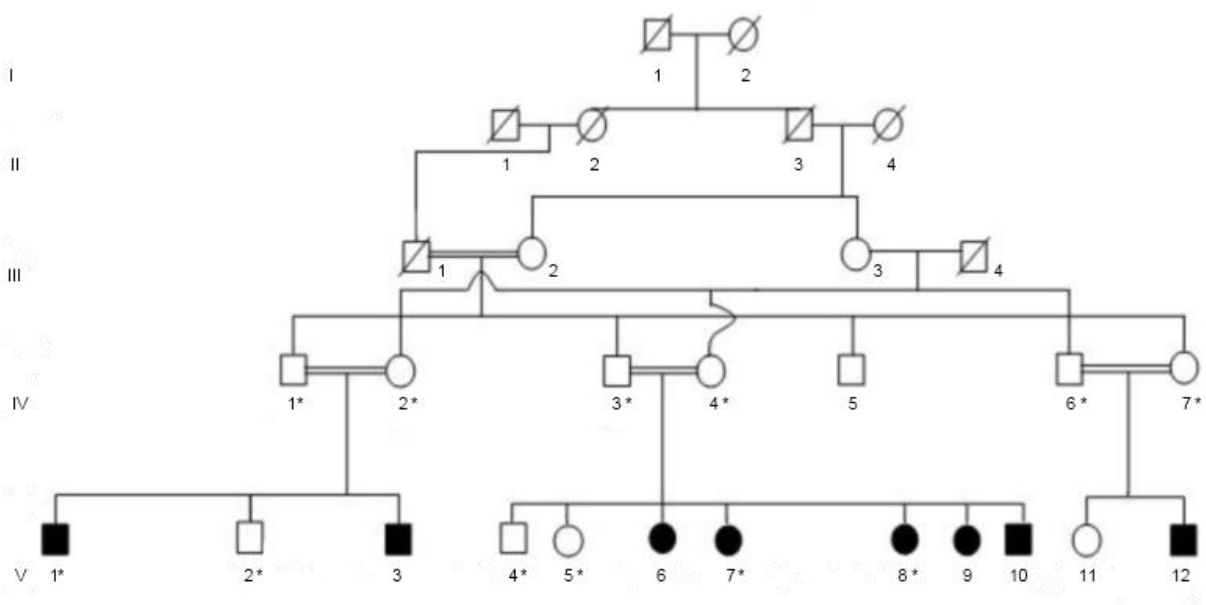


Figure 3.8: *Pedigree of family, with * indicating family members for whom DNA was available.*

D08.23732 is an additional “singleton”, from a Caucasian family, who presented with Knobloch Syndrome. A DNA sample was obtained, but no other clinical information was available for this individual.

3.3.3 Previous Work

Initial mapping was carried out in Pakistan, in Professor Qasim Mehdi’s laboratory. Linkage analysis to 17q11.2 was confirmed in the Medical and Molecular Genetics laboratory in Birmingham by Dominic White, with a maximum lod score of 3.40 (θ=0.00) at markers D17S1307 and D17S1166. A minimum candidate region of

2.82cM (2.606Mb) was detected between markers D171532 and D17S798. It contained 43 genes, and this was thought to include the putative KNO3 locus. Dominic White undertook sequencing of candidate genes *MYO1D* and *RAB11-FIP4*, and Professor Mehdi's laboratory sequenced *UNC119* but no putative mutations were identified.

3.3.4 Molecular Genetic Methods

Genes were prioritised for sequencing depending on expression data available at the databases AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/) and Genatlas (http://www.dsi.univ-paris5.fr/genatlas/). Those with evidence of expression in the lens and/or eye were chosen. Genes were sequenced as described in the Materials and Methods chapter. Both forward and reverse primers were designed for coding exons of the genes *C17orf42, RNF135, OMG, EVI2A* and *EVI2B,* ensuring that they were within the introns, or outside the gene. DNA from family members IV:4 and V:8 was used initially for genotyping. If the mother, IV:4, was found to be heterozygous, and the affected child, V:8, homozygous for an atypical sequence, the other family members were sequenced to see if the homozygous mutant allele segregated with disease phenotype. PCR products were sequenced in both the forward and the reverse direction.

3.3.5 Results

The region on 17q11.2 contains 26 known genes, according to UCSC (http://genome.ucsc.edu/).

3.3.5.1 Chromosome 17 Open Reading Frame 42 (*C17orf42*)

This gene, chromosome 17 open reading frame 42 (*C17orf42*) is expressed in the eye, as well as in the small intestine, brain and lung. Its function is unknown. *C17orf42* consists of four exons, and 360 translated residues.

A known SNP (rs28539246) was identified in exon 3 of *C17orf42*. This fragment was sequenced in all family members for whom DNA was available (Table 3.4).

Family Member	Status	Genotype
IV:1	Unaffected	T/T
IV:2	Unaffected	T/A
V:1	Affected	T/A
V:2	Unaffected	T/T
IV:3	Unaffected	T/T
IV:4	Unaffected	T/A
V:5	Unaffected	T/A
V:7	Affected	T/T
V:8	Affected	T/T

Table 3:4: *rs28539246 segregation in Pakistani family with Knobloch syndrome.*

If this region segregated with disease status, it would be expected that affected individuals would have the same homozygous genotype.

3.3.5.2 Oligodendrocyte Myelin Glycoprotein (*OMG*)

OMG, the oligodendrocyte myelin glycoprotein gene, consists of two exons, the first of which is non-coding. The translation length is 44 residues. It lies within the first exon of *NF1*, but is transcribed from the opposite strand. It is expressed in the lens and eye, as well as in the brain and lung. The protein product plays a role in CNS myelination, potentiating neurite outgrowth inhibition (it is a NOGO- neurite outgrowth inhibitor- receptor ligand) and in axonal regeneration after brain injury.

A non-pathogenic change was found, rs11655238, where IV:4 was heterozygous T/C and the affected child V:8 was homozygous T/T. This is a SNP and does not disprove linkage.

3.3.5.3 Ecotropic Viral Integration Site 2A (*EVI2A*)

EVI2A, ectropic viral integration site 2A, is thought to be an oncogene in retrovirus-induced myeloid tumours. There are three different transcripts, only one of which contains an initial non-coding exon. It is located within intron 1 of *NF1*, but is transcribed on the opposite strand. There is evidence of expression in the eye, as well as in the brain, lung, spleen, kidney and uterus.

No sequence alterations were detected in I:4 and II:5.

3.3.5.4 Ecotropic Viral Integration Site 2B (*EVI2B*)

EVI2B, ectropic viral integration site 2B is a two exon gene, the first of which is non-coding. It is thought to have a role in melanocyte and keratinocyte differentiation. The protein product is thought to be proline rich and consist of 448 amino acids. Again, it lies within intron 1 of *NF1*, but is transcribed from the opposite strand. It is expressed in the eye, along with the brain, lung, spleen, kidney and uterus.

Two sequence variations were identified in exon 2 of *EVI2B*. The first was a coding SNP, rs12942186, where IV:4 was heterozygous G/A, and V:8 was homozygous G/G. The second was a 15bp deletion (del_TTTTAAAAAAAGTTA) in the non-coding region of the exon that the mother, I:4 was heterozygous for. The deletion was not evident in the sequencing results from II:5, the affected child.

3.3.5.5 Ring Finger Protein 135 (*RNF135*)

RNF135, the ring finger protein 135 gene, is expressed in the eye, as well as the lung, brain, prostate, kidney and placenta. It falls within the region of *NF1* commonly deleted in neurofibromatosis. The protein product has a role in ubiquitin ligase activity, and possibly in protein-protein interactions. There are two transcripts for this gene. The first consists of five exons, the first and fifth being partially non-coding. The transcript contains 432 amino acid residues. The second transcript consists of four exons, lacking the third 163bp coding exon found in the first transcript. The first and fourth exons are partially non-coding, and the transcript has a length of 210 amino acids. Mutations in this gene have been associated with Overgrowth Syndrome, where symptoms include macrophthalmia, myopia, learning disabilities, facial dysmorphism and hearing loss (Douglas *et al*, 2007).

A previously unreported change, c.299A>G (NG_011701.1), was found in this gene at amino acid position 100, causing a histidine residue to be substituted for an arginine residue. Initially, only DNA from IV:4 and V:8 was sequenced, and the change was found to be “possibly damaging” when using Polyphen (http://genetics.bwh.harvard.edu/pph/). The other family members were then sequenced, but the change did not segregate with disease phenotype (Table 3.5).

Family Member	Status	Genotype
IV:1	Unaffected	A/G
IV:2	Unaffected	A/G
V:1	Affected	A/G
V:2	Unaffected	A/A
IV:3	Unaffected	A/G
IV:4	Unaffected	A/G
V:5	Unaffected	A/A
V:7	Affected	G/G
V:8	Affected	G/G

Table 3.5: *Results of sequencing the novel change in RNF135.*

In V:8, heterozygous missense substitutions were also noted in *RNF135,* g.[39C>T;40C>A], c.[-13C>T;-12C>A] .

In light of the identification of SNPs rs28539246 at 26,250,127 in *C17orf42*, and of c.299A>G in *RNF135*, which did not segregate with disease phenotype, it seemed unlikely that the gene causing the Knobloch-like Syndrome will be found in this segment of the region. This narrowed down the candidate region identified by previous efforts, so the remaining region spanned from 26,351,053 to 28,292,780 on chromosome 17, a region of around 1.9Mb (according to March 2006 Build, NCBI36/hg18).

It was possible that a good candidate region was missed when the initial genome wide scan was undertaken, so affected individuals V:1, V:7 and V:8 were analysed by Louise Tee, using Affymetrix GeneChip Human Mapping 250K SNP Sty1 Array to see if a better candidate region can be identified. It was possible that a better candidate region was missed because the initial linkage study was conducted using a genome wide scan and microsatellite markers at 10-20cM intervals. The SNP Chip has been developed since this time, offering a more accurate picture due to its higher resolution. Homozygous regions in the three affected individuals were identified, including one that contained *COL18A1* (NM_130445.2), the original gene cited as being involved in Knobloch Syndrome. This region spanned 2.4Mb on chromosome 21 from 44,548,483 to 46,902,240 (March 2006 Build, NCBI36/hg18). Primers were designed to flank all coding exons in all versions of the gene. The Table 3.6 summarises the results of this analysis. Included in the Table is the result of sequence analysis of a Caucasian individual, D08.23732, who presented with Knobloch Syndrome, apparently a compound heterozygote.

In the Knobloch family from Pakistan, a 2bp deletion (CT) was detected in exon 40 (Ensembl transcript ID ENST00000400347), predicted to cause a frameshift mutation, NM_130445.2:c.3514_3515del, NP_569712.2:p.Leu1172Valfs*72 (Figure 3.9). In family member IV:4, this was a heterozygous change. The deletion was homozygous in V:8. The mutation segregated with disease status when available DNA from the other family members (IV:1, IV:2, IV:3, IV:4, IV:6, IV:7, V:2, V:4) was analysed. The three affected individuals tested were all homozygous for the mutation. Eight of nine unaffected individuals were heterozygous for the mutation, and one was homozygous wild type, and none were homozygous for the mutation. Using Superlink Online (http://bioinfo.cs.technion.ac.il/superlink-online/), the LOD score at recombination θ=0.00 was calculated as 2.24. 213 Asian controls (426 chromosomes) were sequenced using PCR product generated using primers flanking exon 40, and none demonstrate this change.

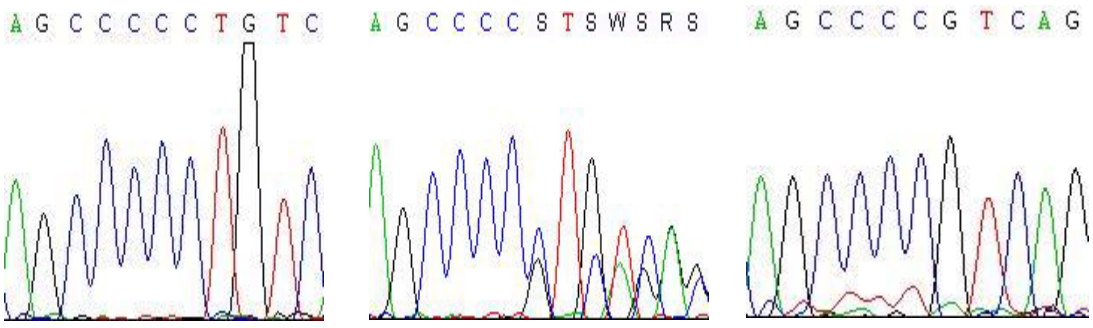


Figure 3.9: *Electropherograms showing c.3617_3618delCT status.*

In exon 38, IV:4 was found to be heterozygous G/A for SNPrs12483761, and V:8 was found to be homozygous A.

The affected Caucasian individual may be a compound heterozygote. In exon 1 (Ensembl transcript ID ENST00000359759), a heterozygous C/T SNP (rs8133886) was detected, along with a heterozygous previously unreported amino acid change,

p.Leu392Pro, which was found in 5 of 103 Caucasian control samples (all heterozygous for this change) so is likely to be an unreported SNP. In exon 8 (Ensembl transcript ID ENST00000342220), a previously unreported heterozygous cytosine insertion was found, but was detected in 5 out of 98 Caucasian controls as a heterozygous insertion. In exon 6 (Ensembl transcript ID ENST00000400347) heterozygous SNPs rs2230687 and rs2230688 were detected. Other heterozygous SNPs were found in ENST00000400347, in intron 10 (rs9979845) in exon 18 (rs11702425), and in intron 24 (rs35423701). A previously unreported sequence variation was found in exon 38, where cytosine was substituted for adenine, c.3239C>A, causing p.Pro1080Gln according to Ensembl, and P1077 according to NCBI and UCSC. This change was not found in 33 Caucasian controls. There is a previously unreported heterozygous cysteine to tyrosine substitution in exon 42 (ENST00000400347) which Polyphen predicts to be “probably damaging”, p.Cys1288Tyr, caused by c.3912G>A. This sequence variation was not found in 398 control Caucasian chromosomes.

3.3.6 Discussion

During the course of investigation, five candidate genes (*C17orf42, RNF135, OMG, EVI2A* and *EVI2B*) were excluded as Knobloch Sydrome candidates. KNO3, a novel Knobloch Syndrome locus was mapped to 17q11.2 (Khaliq *et al*, 2007) but the identification of a germline mutation in *COL18A1* (21q22.3) in the same family now permits rejection of the *KNO3* locus for the Knobloch Syndrome described in the Asian family (Joyce *et al*, 2010). The frameshift mutation identified (c.3514_3515del), predicted to cause p.Leu1175delinsValfsX71, is the likely culprit. It is perhaps of note that initially the family was described in clinic as having Knobloch-Like Syndrome, since affected members did not exhibit occipital encephalocele or cardiopulmonary

defects that are usually seen with Knobloch syndrome, even though most *COL18A1* mutations in Knobloch Syndrome are found in and after exon 30. It is possible that additional genetic factors, or environmental effects, are responsible for the atypical phenotype (Suzuki e*t al*, 2009).

This finding demonstrates that high-density SNP arrays are more reliable aids in candidate gene mapping than the microsatellite markers and autozygosity mapping approach. Initially, mapping did not highlight the 2.4Mb autozygous region, which contains *COL18A1*, because it lies inside a 6.7Mb interval between flanking markers D21S2055 and D21S1446. In consanguineous kindreds, the gene harbouring the autosomal recessive mutation usually resides in a large homozygous region. Woods *et al* (2006) found the mean homozygous segment length to be 26 cM (range 5-70 cM. Hence, small intervals may be found in some families.

Primer	*Sequence*	*Product size (bp)*	*Annealing Temp. (°C)*	*Additional*	*D08.23732*	*IV:4*	*V:8*
COL18A1-203pr1A-F	CCACCTCCAGGCACAGAG	559	64	acetamide	Ensembl rs914230. T allele. (ancestral=C)	Ensembl rs914230. T allele. (ancestral=C)	Ensembl rs914230. T allele (ancestral=C)
COL18A1-203pr1A-R	AGAGTGGTCCCATTCTCCTG						
COL18A1-203pr1B-F	AGGAGAACATTGCCGGTGT	593	64	acetamide	wt	wt	wt
COL18A1-203pr1B-R	GCTGTCCCTGGTTAAAGTGG						
COL18A1-203pr1C-F	ACTTCTGCACCCCCTGGT	478	64	acetamide	1) Htz CT SNP rs8133886. 2) Leu392Pro "possibly damaging" in 5/103 Caucasian Cs	wt	wt
COL18A1-203pr1C-R	GCTCCAACACGCATCCTG						
COL18A1-203pr1D-F	CCTTCCTCGCCTGGTTCT	291	57		wt	wt	wt
COL18A1-203pr1D-R	ACCTGTACCAGGCTCACGTC						
COL18A1-205pr8-F	CCAGCCTGGGACTCTGGA	407	60.4		Htz C ins. Caucasian controls: 93/98=wt 5/98=htz C ins	wt	wt
COL18A1-205pr8-R	AGCTCAGCTCACTCCCATGT						
COL18A1-205pr14-F	CCAAGCTTCCACGTTGGTTA	174	60		wt	wt	wt
COL18A1-205pr14-R	CTCAGTCCAAAGCAGTGCAG						
COL18A1-205pr15-F	TTTGCAGGGTTTAGAAAGCA	300	57		wt	wt	wt
COL18A1-205pr15-R	GTAGTAGGGCTGGGGCAAC						
COL18A1-205pr17-F	GATGGGCCACAGGTAGTGTT	203	57		wt	wt	wt
COL18A1-205pr17-R	GTTCAATGAAGGGGCATCTC						
COL18A1-204pr1+2-F	ACCGCGGCGGAGGAGGCAGCAT	varies	64	DMSO	wt	Insertion in intron 1: CTGCGGGG. Heterozygous	Insertion in intron 1: CTGCGGGG. Homozygous
COL18A1-204pr1+2-R	TGGGCAGGGCCGGGGCTGAA						

Primer	*Sequence*	*Product size (bp)*	*Annealing Temp. (°C)*	*Additional*	*D08.23732*	*IV:4*	*V:8*
COL18A1-204pr3A-F	GCCCTCCCAGACTCAGTTTC	499	57		wt	wt	wt
COL18A1-204pr3A-R	GTCCACTGGCCGACGAAG						
COL18A1-204pr3B-F	CAGGCCATGGTCTTGCTG	394	64	acetamide	wt	wt	wt
COL18A1-204pr3B-R	CACAGCTCCGTCCCTGTC						
COL18A1-204pr4-F	ATCTGGAGCTCAAGCAGCAC	241	57		wt	wt	wt
COL18A1-204pr4-R	GTCATAGACCTTGGCCCTGA						
COL18A1-204pr5-F	AGAGCAGCGTCCTTTGCTT	226	60.4		wt	wt	wt
COL18A1-204pr5-R	ACATGTGTCACCCCACTGC						
COL18A1-204pr6-F	TCACAAACCAAGCAAGTCTCC	278	60		Heterozygous rs2230687 (C=WT, G=var), rs2230688 (G=WT, C=var)	wt	wt
COL18A1-204pr6-R	CACCATCACACAAACACACG						
COL18A1-204-7new-F	CAGGACTGAAAGCGTTTGG	277	64		wt	wt	wt
COL18A1-204-7new-R	ACAGAGGGGCTTCATCAGG						
COL18A1-204pr9-F	TGGGGTGCATTTCCATGTAG	207	57		wt	wt	wt
COL18A1-204pr9-R	CTGGGTCGCGTTTAAAAAGA						
COL18A1-204pr10+11-F	GGGGAGGAGCACTGAGAGT	393	57		Intron 10 a/g rs9979845	wt	wt
COL18A1-204pr10+11-R	CAGTGCCTGCTTTTGAGGAG						
COL18A1-204pr12-F	CGTGGTTTCTCAGCTCCTTG	250	57		wt	wt	wt
COL18A1-204pr12-R	CTGAATCTTGGGGGTCCAT						

Primer	*Sequence*	*Product size (bp)*	*Annealing Temp. (°C)*	*Additional*	*D08.23732*	*IV:4*	*V:8*
COL18A1-204pr13-F	GTGGTGGTCATCCCTGGT	284	57		wt	wt	wt
COL18A1-204pr13-R	ggtgttcacccctcactcc						
COL18A1-204pr14-F	aaatcgttttagtggccgaat	228	57		wt	wt	wt
COL18A1-204pr14-R	gggctattcctggcatttc						
COL18A1-204pr15-F	ttcctctgtccactgtgctg	233	57		wt	wt	wt
COL18A1-204pr15-R	gatggacagatggcaggagt						
COL18A1-204pr17-F	gaggggtccttccctaagaa	202	57		wt	wt	wt
COL18A1-204pr17-R	tggctctcagggacactctc						
COL18A1-204pr18-F	tactagcgggcttttcttgc	247	57		rs11702425 (T/C)	wt	wt
COL18A1-204pr18-R	gcaaacttcccaggatgaat						
COL18A1-204pr19-F	aactcacccttcccttcacc	244	64		wt	wt	wt
COL18A1-204pr19-R	aaacaaagcagtggcaggaa						
COL18A1-204pr20-F	agagaagtccaggccatcg	241	57		wt	wt	wt
COL18A1-204pr20-R	atagtcccttcagggctcca						
COL18A1-204pr21-F	cggtcgggaaataaagaacc	206	57		wt	wt	wt
COL18A1-204pr21-R	gccaaaccttctgggatgt						
COL18A1-204pr22-F	gtggactctggggtcctg	262	60		wt	Apparent unseen heterozygous frameshift in intron 21	wt
COL18A1-204pr22-R	gtcctgacctgggggatgt						

Primer	*Sequence*	*Product size (bp)*	*Annealing Temp. (°C)*	*Additional*	*D08.23732*	*IV:4*	*V:8*
COL18A1-204pr23+24-F	aagtcgctcgagtccagttg	321	57		rs35423701 in intron 24	wt	wt
COL18A1-204pr23+24-R	ctgctctgtgctaccaggtg						
COL18A1-204pr25-F	ctgtcgggggagatggag	226	57		wt	wt	wt
COL18A1-204pr25-R	cctcccaacagtggtcatct						
COL18A1-204pr26-F	gtgaggctttgtggggaag	222	60		wt	wt	wt
COL18A1-204pr26-R	agccagaccctcaggaagac						
COL18A1-204pr27-F	ctcagagaggctgccaggt	183	60		wt	wt	wt
COL18A1-204pr27-R	cacagcacaagctcagcag						
COL18A1-204pr28-F	cccaactcgtggtcaagg	209	64		wt	wt	wt
COL18A1-204pr28-R	actcccctctgggcatgt						
COL18A1-204pr29-F	cgggaaggtgtctggtaatc	239	57		wt	wt	wt
COL18A1-204pr29-R	ataagtggcctctgggagga						
COL18A1-204pr30-F	tgcctggtcaggtttctgat	232	56	GC rich	wt	wt	wt
COL18A1-204pr30-R	ggctctgtgtgacagccaag						
COL18A1-204pr31-F	atacaggggctggtgcttct	244	57		wt	wt	wt
COL18A1-204pr31-R	ccactgggggagtctcact						
COL18A1-204pr32-F	ccaggcactagggcatttc	249	57		wt	wt	wt
COL18A1-204pr32-R	ccaggacctgcttgtggt						

Primer	*Sequence*	*Product size (bp)*	*Annealing Temp. (°C)*	*Additional*	*D08.23732*	*IV:4*	*V:8*
COL18A1-204pr33-F	ctaccgcgaaatggctagaa	216	57		wt	wt	wt
COL18A1-204pr33-R	ccagcgggatgtacaggac						
COL18A1-204pr34-F	gctccggaagcttctgact	393	57		wt	wt	wt
COL18A1-204pr34-R	acacggggacagggacat						
COL18A1-204pr35+36-F	gtgggggtttctcaggctat	547	57		wt	wt	wt
COL18A1-204pr35+36-R	ggtccgtgggagagtgtct						
COL18A1-204pr37-F	ctgaaacgggcattccttc	290	57		wt	wt	wt
COL18A1-204pr37-R	gagcctctcgccttcctta						
COL18A1-204pr38-F	gagggagaggtgggtgct	450	56	acetamide	Hetz unknown SNP 23rd base, C to A. P1080Q Ensembl. P1077 NCBI/UCSC. Not in 33 Caucasian controls	wt (heterozygous GA for rs12483761)	wt (homozygous A rs12483761)
COL18A1-204pr38-R	aggaccacgtgttgttcctc						
COL18A1-204pr39-F	ctgcttgccagttcagagc	430	57		wt	wt	wt
COL18A1-204pr39-R	ctgggccctcttctgctg						
COL18A1-204pr40-F	cccctcagtgtgtcacttgc	410	64		wt	Htz NM_130445.2:c.3514_3515del, NP_569712.2:p.Leu1172Valfs*72. Not in 213 Asian controls	Hmz NM_130445.2:c.3514_3515del, NP_569712.2:p.Leu1172Valfs*72.. Not in 213 Asian controls
COL18A1-204pr40-R	ggagttcaccccagaggtc						
COL18A1-204pr41-F	cacatccacaccccacatc	296	57		wt	wt	wt
COL18A1-204pr41-R	ggggaaactgcagataggag						

Primer	***Sequence***	***Product size (bp)***	***Annealing Temp. (°C)***	***Additional***	***D08.23732***	***IV:4***	***V:8***
COL18A1-204pr42-F	agcggcctctgccctaag	386	66	acetamide	Hetz Cys1291Tyr (Ensembl), Cys1288Tyr (NCBI/UCSC), Polyphen: probably damaging. Not in 199 Caucasian controls	wt	wt
COL18A1-204pr42-R	GAAAGTATGGCAGCCAGGTC						

Table 3.7: *Details of primer and conditions used in molecular genetic analysis of COL18A1 in the kindred, and the singleton case D08.23732, and results of this analysis. Key: wt=wild type, htz=heterozygous, hmz=homozygous.*

Chapter 4
Non-Syndromic Cataract Families

4. NON-SYNDROMIC CATARACT FAMILIES

4.1 General Introduction

Isolated cataracts can occur as an inherited condition, without any other detectable characteristics to suggest an underlying syndrome diagnosis. Non-syndromic cataracts may manifest as a result of autosomal dominant, autosomal recessive or X-linked recessive inheritance.

Careful consideration was given to the grouping of the families with non-syndromic cataracts studied during this project. The families could not be grouped by cataract phenotype, because in many cases, this was not available in the clinical information supplied. Some affected individuals had cataract surgery abroad, before their genetic predisposition was noted in clinics in the UK. In some cases, cataract phenotype was simply not noted or provided. Additionally, as discussed in 1.3.2.3.2.6**,** mutations in the same gene can sometimes cause a variety of cataract phenotypes. Grouping by ethnic origin was also considered as in some cases a founder mutation can be detected in apparently unrelated families with a common ethnic origin. However, equally, there are examples of families with shared ethnicity and location who prove to have different mutations (or genes) despite a shared phenotype. Hence it was decided not to make unsupported assumptions about underlying locus heterogeneity but to re-evaluate various hypotheses as results were accrued.

All positions referred to in this chapter are based on the February 2009 GRCh37/hg19 build of the human genome, unless otherwise stated.

4.2 Cataract Family 1 and *CDC25A*

Aim: To map the locus for a gene causing cataracts in a Turkish family and then to investigate the involvement of *CDC25A* as a candidate gene.

4.2.1 Patient DNA

DNA was extracted from blood samples. DNA was available from all family members, consisting of unaffected parents I:1 and I:2, and the three affected children II:1, II:2 and II:3 (Figure 4.1).

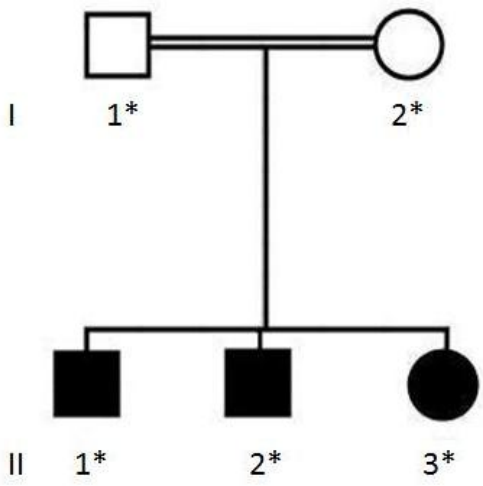


Figure 4.1: *Pedigree of Cataract Family 1 (of Turkish origin).*

4.2.2 Molecular Genetic Methods

Experimental procedures were carried out as detailed in Chapter 2.

4.2.2.1 Genome-Wide Scan

To identify shared regions of homozygosity between the affected individuals, a genome-wide scan was carried out with the Affymetrix 10K SNP microarray.

4.2.2.2 Microsatellite Markers

Regions of homozygosity >2Mb were probed with microsatellite markers. Family DNA was amplified by PCR, and then genotyped with an ABI 3730 Genetic Analyser. Data obtained here was analysed with the GeneMapper programme.

4.2.2.3 Primer Design and Sequencing

Candidate genes were selected for analysis. Dr Esther Meyer had previously investigated *GPX1* as a candidate. Mutation analysis was carried out for *RHOA*, *LAMB2*, *TREX1*, *CSPG5*, *CCDC12*, *TMIE* and *CDC25A*.

4.2.2.4 Functional Work

To investigate the effect of a putative mutation in *CDC25A*, wild type and mutant constructs were cloned using specific plasmid primers (Table 4.8) and introduced into the CRL-11421 cell line. Changes at the protein level were analysed by Western blotting, FACS and immunofluorescent staining.

4.2.3 Results

4.2.3.1 Genome-Wide Scan

The genome-wide scan was performed for individuals II:1, II:2, and II:3. One potential region of extended homozygosity was detected on chromosome 3 between rs2056321 (45,083,133) and rs1444185 (58,686,977). These SNPs define the extent of the homozygous region, because at least one of the affected individuals is heterozygous at this position.

4.2.3.2 Microsatellite Markers

This region of interest was investigated further by Dr Esther Meyer to confirm presence or absence of homozygosity at various polymorphic microsatellite markers. Linkage to the region on chromosome 3 was confirmed.

4.2.3.3 Candidate Genes

The linked region on chromosome 3 from rs2056321 (45,083,133) to rs1444185 (58,686,977) overlapped with a previously reported region for three Arab families with autosomal recessive cataracts from D3S1768 (34,624,330) to D3S2409 (49,418,084) (Pras *et al*, 2001). The search for the causative gene was therefore focussed on the overlap region from 49,418,084 to 58,686,977, which contains 169 genes.

Candidate genes were selected based on evidence of expression in the eye and/or lens. Previously, *GPX1* was investigated, and no pathogenic exonic mutations were identified. *RHOA*, *LAMB2*, *TREX1*, *CSPG5*, *CCDC12*, *TMIE* and *CDC25A* were chosen for mutation analysis. Analysis of coding exons of *RHOA*, *LAMB2*, *TREX1*, *CSPG5*, *CCDC12*, and *TMIE* revealed no potential mutations but a previously unreported variant was detected in *CDC25A*.

4.2.3.3.1 *CDC25A*

A putative mutation, c.137_139del (referring to NM_001789.2; Figure 4.2), was detected in exon 1 of *CDC25A*, this was predicted to cause the deletion of two amino acids and the insertion of a new amino acid (p.Thr46_Val47delinsIle) in the gene

product. This variant segregated with disease status. Affected individuals II:1, II:2 and II:3 were all homozygous for the 3bp deletion, whereas the unaffected parents I:1 and I:2 were heterozygous. Analysis of DNA from ethnically matched control chromosomes (316 Asian, 26 Caucasian, and 450 Turkish) did not reveal this deletion.

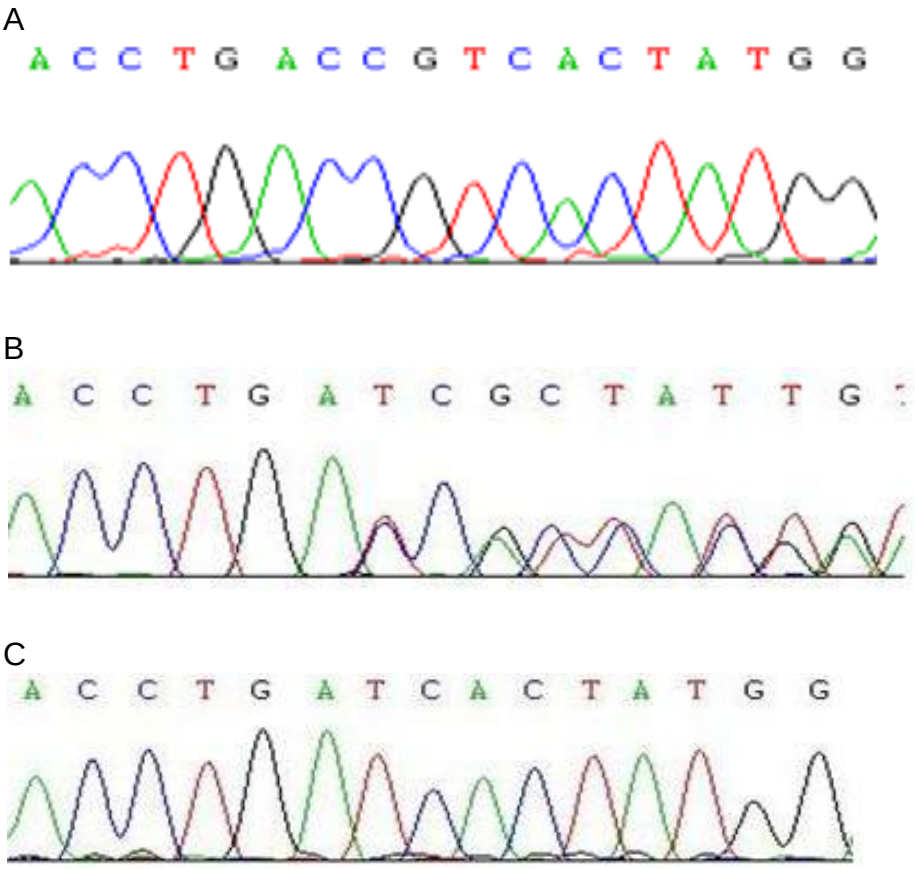


Figure 4.2: *Electropherograms showing homozygosity and heterozygosity for p.Thr46_Val47delinsIle, and wild type sequences. A=control sample showing wild type sequence. B=parent DNA showing heterozygosity for p.Thr46_Val47delinsIle. C= affected child, homozygous for p.Thr46_Val47delinsIle.*

A sequence alignment was carried out, which showed these bases to be highly conserved in mammals, but not in chickens (Figure 4.3).

```
CDC25A, H.sapiens        1 MELGPEPPHRRRLLFACSPPPASQPVVKALFGASAAGGLSPVTNLTVTMD 50
CDC25A, P.troglodytes    1 MELGPEPPHRRRLLFACSPPPASQPVVKALFGTSAAGGLSPVTNLTVTMD 50
CDC25A, C.lupus          1 MELGPEPPHRRRLLFACSPPPAPQPVVKALFGAPASGGLSPVTSLTVTMD 50
CDC25A, B.taurus         1 MELGPEPPHRRRLLFACSPPPAPQPVVKALFGTPAAGGLSPVTNLTVTMD 50
Cdc25a, M.musculus       1 MELGPEPPHRRRLFFACSPTPAPQPTGKMLFGASAAGGLSPVTNLTVTMD 50
Cdc25a, R.norvegicus     1 MELGPEPPHRRRLLFTCSPTPAPQPTGKVQFGASRAGGLSPVTNLTVTMD 50
CDC25A, G.gallus         1 MDPAPAASYRRRLLHLS--PASPAAVVKSLFPAE----LSPVSDLRLTME 44
```

Figure 4.3:*Sequence conservation of the T and V nucleotide (highlighted) involved in p.Thr46_Val47delinsIle across species.*

DNA from an affected individual from the pedigrees published by Pras *et al* (2001) was obtained, but mutation analysis revealed no putative mutations in the coding regions of *CDC25A*.

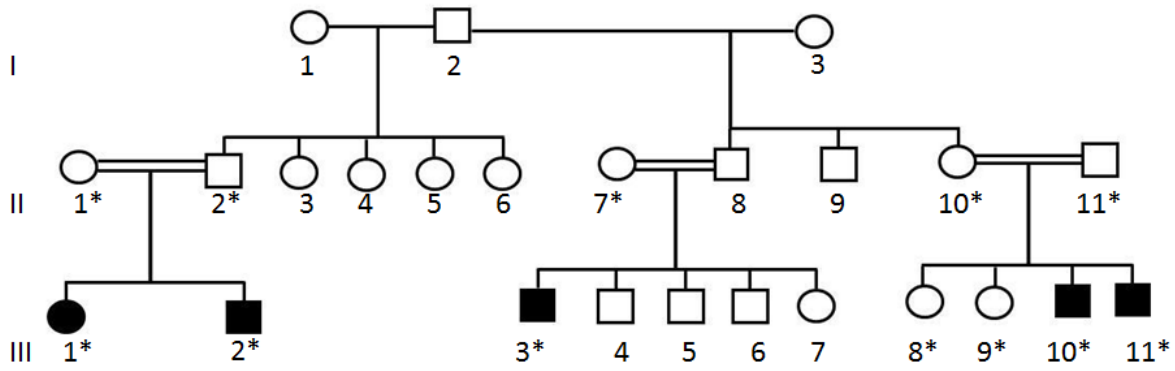


Figure 4.4: *Pedigree for cataract family 2. Affected members presented with non-syndromic cataracts.*

Another cataract family (Cataract family 2) which was of Turkish ethnicity was ascertained (Figure 4.4). PCR and sequencing of exon 1 of *CDC25A* was performed in family 2 but did not reveal the c.137_139del variant. Mutation analysis was then undertaken for the remaining exons of the gene, and in intron 5, III:11 was found to be heterozygous for rs3731507, thus excluding this gene in this family. Furthermore, coding exons of *CRYBB3*, *CRYBB1*, *CRYBB2* and *CRYBA4* were sequenced for II:10 and III:10 (see Tables 6.21-6.24). No sequence changes were detected for *CRYBB3*.

In exon 6 of *CRYBB1* (NM_001887.3), II:10 was heterozygous for c.755A>G, and III:10 was homozygous for c.755A>G (see Table 4.1). This exon was analysed for the remaining family members for whom DNA was available and all affected family members were homozygous G/G for c.755A>G. An unexpected genotype of A/A was detected for II:7, since the affected child III:3 was G/G and possible explanations for the apparent non-maternity would include mistyping or sample mixup or rare occurrences such as uniparental disomy. Though the c.755A>G variant segregated with disease status it was unclear if the variant was pathogenic as the predicted amino acid change p.252Lys>Arg was considered to be likely to be benign change by the Polyphen prediction program. This is a conservative change since both lysine and arginine are basic, hydrophilic amino acids with positively charged side chains. This sequence change has not been previously reported as a SNP. This residue is conserved in chimpanzee, dog, cow and mouse, but not in chicken or zebrafish.

Family member	**Status**	**Genotype**
II:1	Unaffected	A/G
II:2	Unaffected	A/G
II:7	Unaffected	A/A
II:10	Unaffected	A/G
II:11	Unaffected	A/G
III:1	Affected	G/G
III:2	Affected	G/G
III:3	Affected	G/G
III:8	Unaffected	A/G
III:9	Unaffected	A/A
III:10	Affected	G/G
III:11	Affected	G/G

Table 4.1: *Genotypes for cataract family 2 and c.755A>G in CRYBB1.*

Gene	Individual	SNP	Genotype
CRYBB1	II:10	rs576564656	C/G
CRYBB1	III:10	rs576564656	G/G
CRYBB2	II:10	c.306+67T>A	A/T
CRYBB2	III:10	c.306+67T>A	T/T
CRYBB2	II:10	rs4049504	C/T
CRYBB2	III:10	rs4049504	C/T
CRYBA4	II:10	rs16982454	C/A
CRYBA4	III:10	rs16982454	A/A
CRYBA4	II:10	rs7609505	T/C
CRYBA4	III:10	rs7609505	T/T

Table 4.2: *Genotypes for SNPs detected in CRYBB1, CRYBB2 and CRYBA4*.

In summary, no definite mutations were detected in the candidate genes tested in cataract family 2.

4.2.3.4 Functional Work

To further evaluate the significance of the c.137_139del mutation in *CDC25A*, functional work was carried out with the cell line CRL-11421, which was first derived from human lens, grown in EMEM, supplemented with10% fetal bovine serum.

4.2.3.4.1 Western Blotting

CRL-11421 cells were grown on glass cover slips to ~80% confluency and transfected with plasmid DNA (pCMVHA+CDC25A and pCMHVA+CDC25A p.Thr46_Val47delinsIle and pCMVHA empty vector as a control) using Lipofectamine, in a ratio of 1µg DNA: 3µl Lipofectamine. After 16 hours, cells were

lysed with RIPA buffer. Protein determination was carried out with Bio-Rad DC Protein Assay, using Perkin Elmer Workout 2.5 Software on Perkin Elmer 2030 Multi Label Reader Victor X3, measuring absorbance at A690. Calculations were performed too so that 5µg could be loaded per well in the polyacrylamide-SDS gel (see 2.2.9.3). Sigma-Aldrich monoclonal anti-HA clone HA-7 (mouse) was diluted 1:1000 in PBST to detect HA antigen. Primary antibodies were detected with polyclonal rabbit anti-mouse immunoglobulin HRP, diluted 1:2000 in PBST. Enhanced chemiluminescence was used to detect the secondary antibody. Pictures of the chemiluminesence were taken. CDC25A is a 56kD protein. Then the membrane was stripped and reprobed for actin (42kD). The primary antibody, Sigma mαβ-actin was diluted 1:15,000 in PBST, and the secondary antibody, polyclonal rabbit anti-mouse from Dako Cytomation was diluted 1:20,000 in PBST. Enhanced chemiluminesence was used to detect the secondary antibody. All experiments were carried out in triplicate.

Protein Densitometry was carried out to show protein stability levels.

4.2.3.2.1.1 Western Blotting Results

Although by eye there appeared to be a reduction in intensity for the mutant vs the wild type construct, it was necessary to reprobe for actin to confirm this and to check for equal well loading, and this was checked with protein densitometry (see 4.2.3.4.1.2).

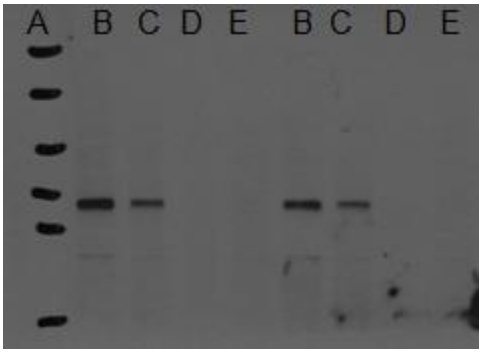


Figure 4.5: *A=protein ladder. B=pCMVHA-CDC25A. C=pCMVHA-CDC25A p.Thr46_Val47delinsIle mutant. D=pCMVHA (untransfected). E=pCMVHA (empty vector).*

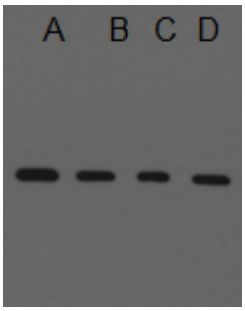


Figure 4.6: *Probing for actin, A=pCMVHA-CDC25A. B=pCMVHA-CDC25A p.Thr46_Val47delinsIle mutant. C=pCMVHA untransfected). D=pCMVHA (empty vector).*

4.2.3.2.1.2 Protein Densitometry Results

Normalised results still clearly show a reduction in protein stability from wild type to mutant. This was carried out as an average for three blots (Table 4.3).

	Intensity per volume: CDC25A	**CDC25A - blank**	**Intensity per volume: actin**	**Actin – blank**	**Normalised (CDC25A/ Actin)**	**Percentage of protein compared to wild type**
Wild Type	573393	78875	377872	44576	1.76945	55.81671 (normalised mutant/ normalised wild type x100
Mutant	516428	21910	355480	22184	0.987649	
Blank	494518	0	333296	0	0	

Table 4.3: *Protein densitometry result*

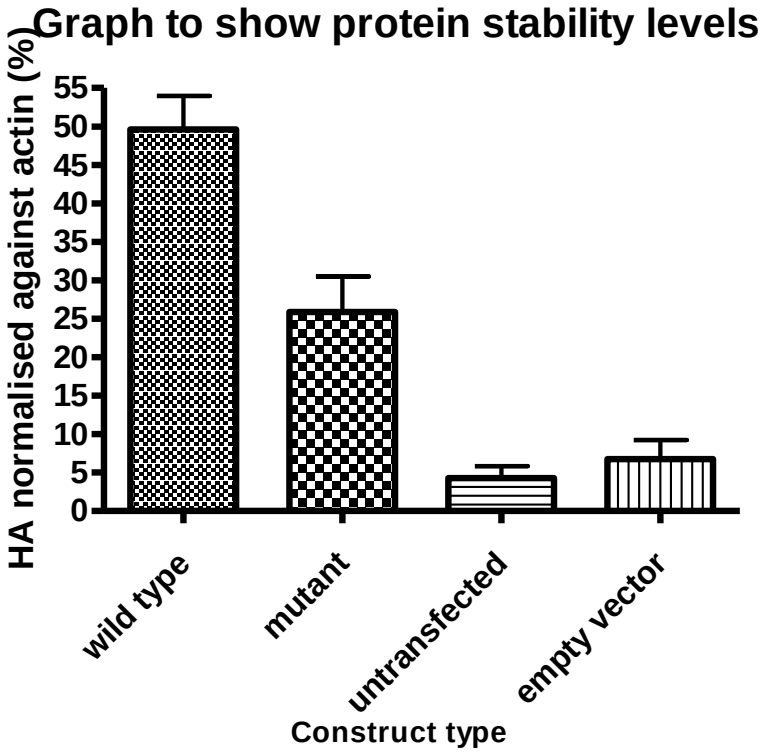


Figure 4.7: *Protein stability levels.*

Data displayed in graph form (Figure 4.7) clearly shows the reduction in stability between wild type and mutant CDC25A that was present in the various CRL-11421 populations.

4.2.3.4.2 Immunofluorescent Staining

As for Western Blotting (see 4.2.3.2.1), CRL-11421 cells were grown on glass cover slips to ~80% confluency and transfected with plasmid DNA (pCMVHA+CDC25A and pCMHVA+CDC25A p.Thr46_Val47delinsIle and pCMVHA empty vector as a control) using Lipofectamine, in a ratio of 1µg DNA: 3µl Lipofectamine. These were left for 24 hours. Cells were washed in PBS and then 4% paraformaldehyde in PBS for 20 minutes, and then permeabilised in 0.1% Triton X100.The primary antibody, anti-HA, was diluted 1:200 in PBS-BSA. After incubation, the primary antibody was detected with 1:200 goat anti-mouse 488 alexia in PBS-BSA. These were then

incubated for two minutes with TOPRO-3 Iodide, excluding light. A drop of Pro Long Gold Antifade Reagent was placed on glass slides and the cover slips were inverted, placed on top, and this was allowed to dry, excluding light. Cells were visualised with a Leica Confocal Scanning Light Microscope SP2 AOBS (Leica, Germany), using lens 63 N/A 1.3, optical zoom 4.

4.2.3.4.2.1 Immunofluorescent Staining Results

Figures 4.8, 4.9 and 4.10 show randomly selected samples of CRL-11421 cells which were observed with the confocal microscope after immunofluorescent staining (4.2.3.4.2). TOPRO-3 Iodide stains nucleic acids, and is a carbocyanine with far red fluorescence. The alexa fluor 488 dye fluoresces green, and shows CDC25A localisation, due to the cross-reaction between protein, primary antibody and secondary antibody.

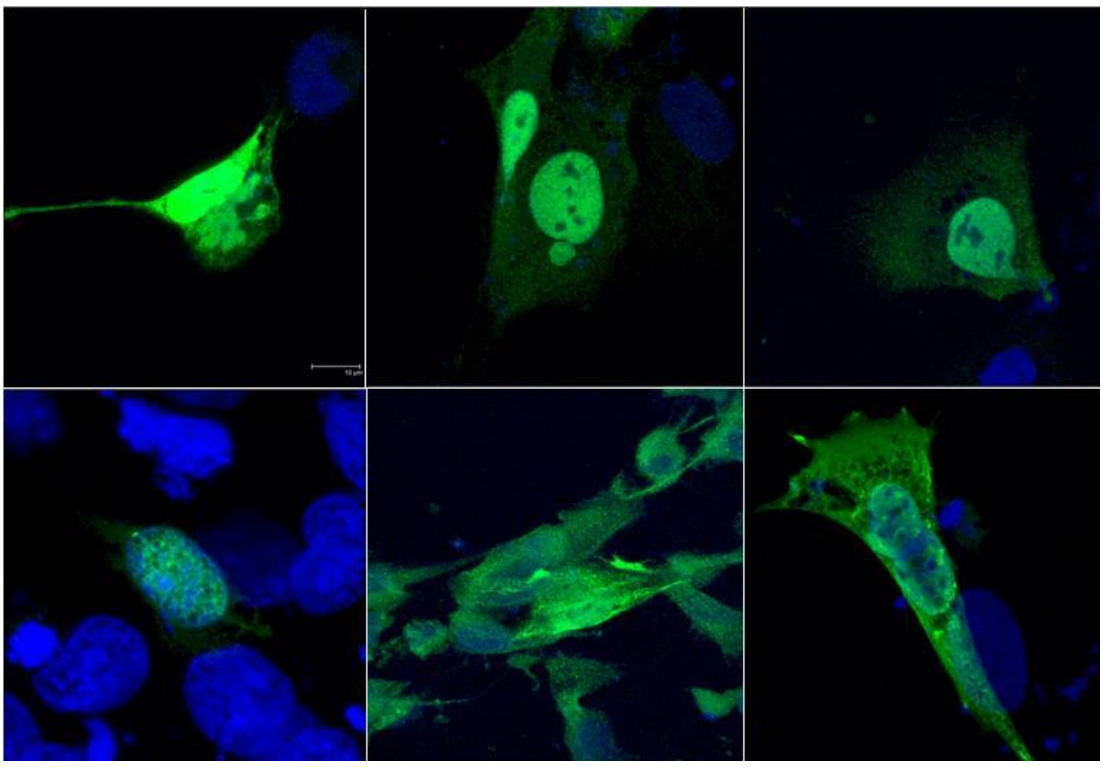


Figure 4.8: *Transfection of wild type CDC25A into CRL-11421 (size bar 10µm).*

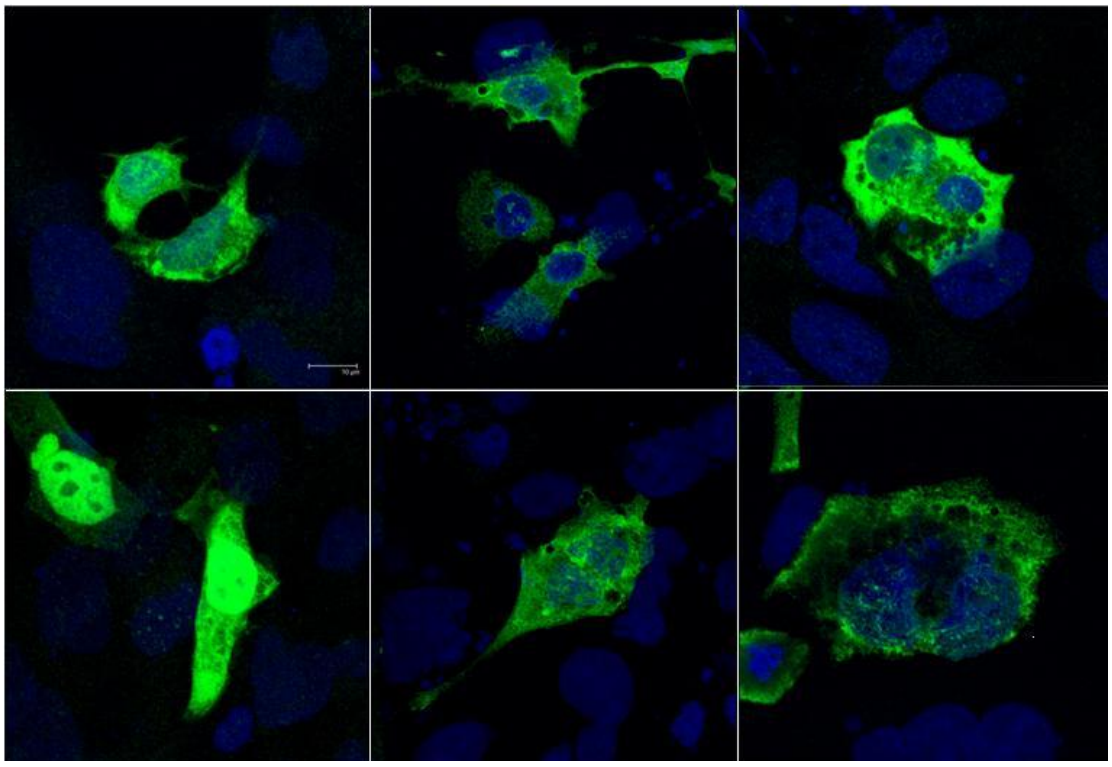
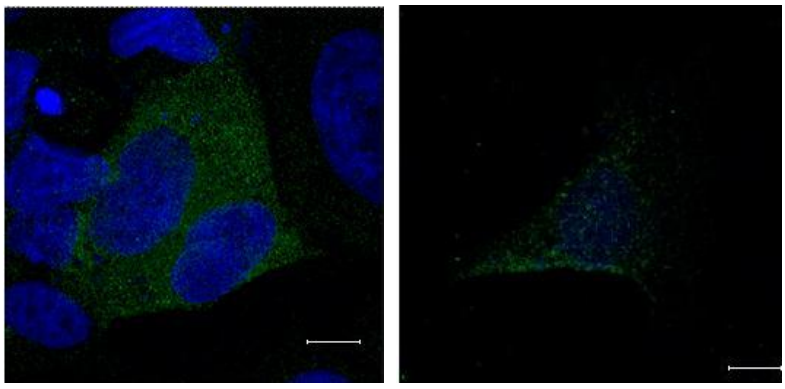


Figure 4.9: *Transfection of mutant p.Thr46_Val47delinsIle CDC25A into CRL-11421 (size bar 10μm).*



A B

Figure 4.10: *A=Transfection of empty pCMVHA vector into CRL-11421. B= Untransfected CRL-11421 (size bars 10μm).*

4.2.3.4.2.2 CDC25A Localisation

To elucidate the intracellular localisation of CDC25A, and compare differences between wild type and mutant p.Thr46_Val47delinsIle CDC25A, the images were compared (Figures 4.8 and 4.9). There is nuclear and cytoplasmic staining (with

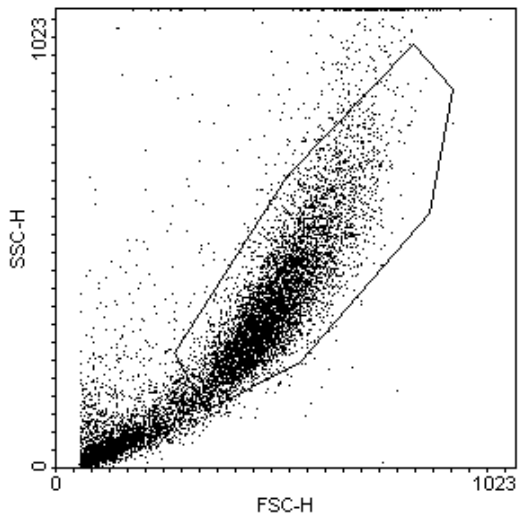
some areas unstained suggesting the presence of vacuoles) in both cases. Cell morphology in CRL-11421 was seen to vary even within the untransfected population (most probably due to cell cycle stage and cell division), so morphological variation was expected within wild type-transfected and mutant-transfected cells, although there did seem to be a tendency for mutant-transfected cells to have a more compact shape with a more serrated appearance to their cell membrane.

4.2.3.4.3 FACS

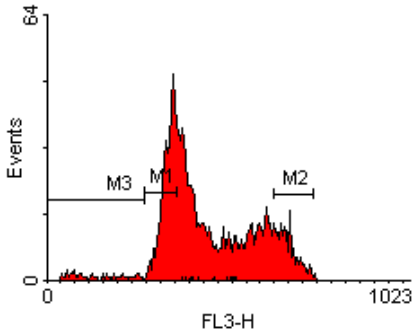
FACS results indicate that there is no particular difference between mutant and wild-type CDC25A transfectants with regard to the proportion of cells in different stages of the cell cycle in CRL-11421 cells (See Table 4.4 and Figures 4.11-4.14). In the histograms in Figures 4.12-4.15, M3 represents the proportion of dead cells, M1 the proportion of cells in G_0/G_1 phase, and M2 represents those in G_2 phase. The area between M1 and M2 is representative of cells in S phase. The dot-plot displays events, or cells, and those that are dead or clumped are gated out so that they can be excluded from further analysis. FSC refers to forward scatter, and SSC refers to side scatter. Since CDC25A stops damaged DNA from going from G_1 to S phase, and from G_2/M, the lack of difference could be due to a lack of damaged DNA.

Cell cycle stage	Wild type CDC25A	Mutant CDC25A	Empty vector	Untransfected
% Dead cells	2.09	1.86	2.33	1.98
% G1/G0	62.64	59.24	46.32	58.88
% G2/M	24.14	24.50	33.98	26.14
% S	11.13	14.40	17.37	13.00

Table 4.4: *FACS results indicating the percentage of CRL-11421 cells within cell cycle stages.*



A

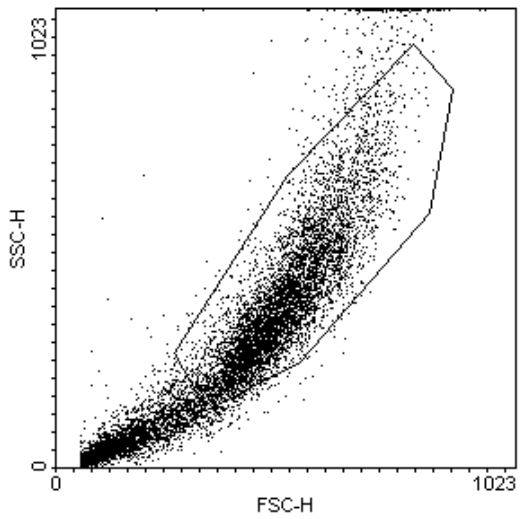


B

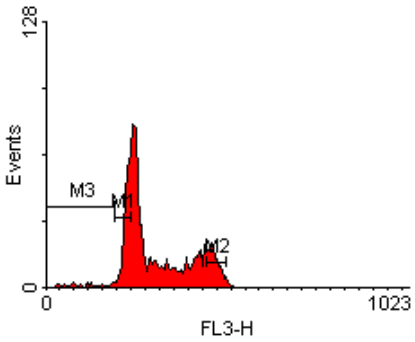
M	Low, High	Events	% Total	% Gated	G Mean	CV	Peak, Value
0	0,1023	5111	51.11	100.00	451.55	28.43	50,358
1	277,369	1596	15.96	31.23	343.65	5.31	50,358
2	638,749	617	6.17	12.07	674.81	3.87	17,686
3	0,277	107	1.07	2.09	142.52	50.92	3,78

C

Figure 4.11: *FACS results for CDC25A wild type transfected CRL-11421. A: dot-plot of cells; B: histogram of proportion of cells in stages of the cell cycle; C :Values generated as a quantitate representation of FACS results*



A

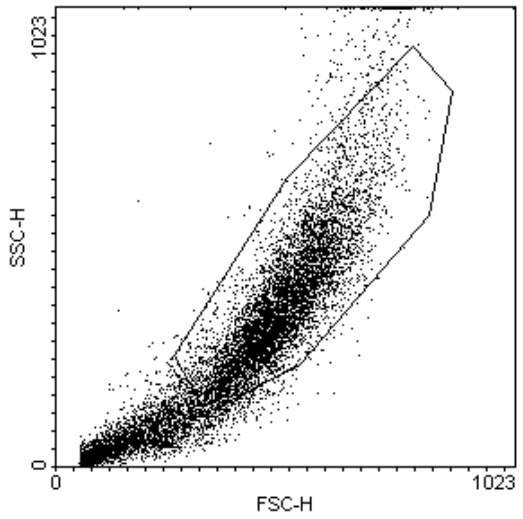


B

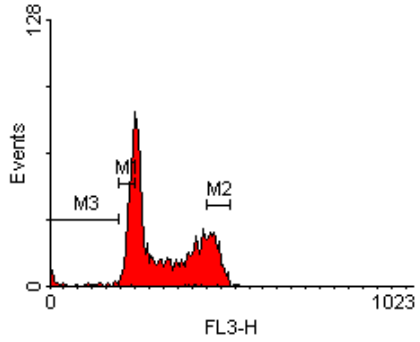
M	Low, High	Events	% Total	% Gated	G Mean	CV	Peak, Value
0	0,1023	4905	49.05	100.00	309.41	29.93	79,246
1	196,246	1453	14.53	29.62	232.90	4.25	79,246
2	450,507	601	6.01	12.25	471.34	3.05	22,466
3	0,196	91	0.91	1.86	123.23	43.31	3,196

C

Figure 4.12: *FACS results for CDC25A mutant transfected CRL-11421. A: dot-plot of cells; B: histogram of proportion of cells in stages of the cell cycle; C: Values generated as a quantitate representation of FACS results*



A

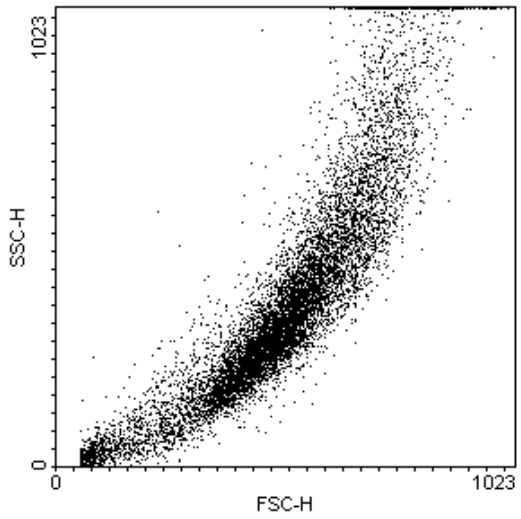


B

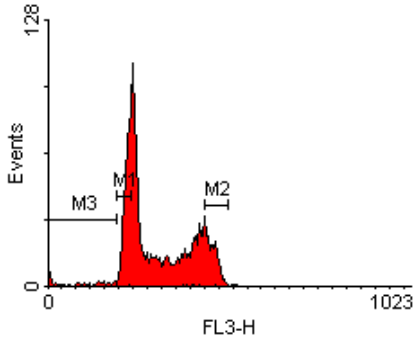
M	Low, High	Events	% Total	% Gated	G Mean	CV	Peak, Value
0	0,1023	5539	55.39	100.00	316.91	31.26	84,242
1	195,241	1283	12.83	23.16	228.60	4.12	72,240
2	440,507	941	9.41	16.99	463.85	3.34	26,462
3	0,195	129	1.29	2.33	65.43	106.73	8,8

C

Figure 4.13: *FACS results for CRL-11421 transfected with empty vector. . A: dot-plot of cells; B: histogram of proportion of cells in stages of the cell cycle; C: Values generated as a quantitate representation of FACS results*



A



B

M	Low, High	Events	% Total	% Gated	G Mean	CV	Peak, Value
0	0,1023	6067	60.67	100.00	307.33	30.57	108,240
1	195,239	1786	17.86	29.44	226.35	4.06	87,239
2	440, 507	793	7.93	13.07	461.60	3.22	34,440
3	0,195	120	1.20	1.98	107.14	65.13	7,6

C

Figure 4.14: *FACS results for untransfected CRL-11421. A: dot-plot of cells; B: histogram of proportion of cells in stages of the cell cycle; C: Values generated as a quantitate representation of FACS results*

4.2.4 Discussion

In a Turkish family, a putative deletion mutation, c.137_139del (p.Thr46_Val47delinsIle), was found in *CDC25A* which is located within a linked region on chromosome 3. This variant segregated with disease status.

CDC25A, the cell division cycle 25 homolog A (*S. Pombe*) gene*,* is part of the CDC25 family of phosphatases and is involved in the cell cycle, being required for G_1/S and G_2/M transition during the cell cycle (Ray and Kiyokawa, 2007). In humans, there are two other CDC25 genes: *CDC25B* and *CDC25C*. There are four main stages in the cell cycle: G_1 phase, S phase, G_2 phase and M phase. G_0 is the gap or resting phase where cells are in quiescence. M phase includes mitosis and cytokinesis.

CDC25A is expressed in the eye, liver and spleen, brain, uterus, kidney, skin, lung and heart tissues, amongst others (Aceview at NCBI).

When DNA is damaged, CDC25A is degraded, and so cells with chromosomal abnormalities cannot proceed through the cell cycle. During the normal course of events, CDC25A activates CDC2, a cyclin-dependent kinase, and this is achieved by removing two phosphate groups. It has been shown that ultra-violet light and ionizing radiation exposure in human cells causes CDC25A degradation, which is ubiquitin and proteasome dependent (Mailand *et al*, 2000). The cell cycle is arrested by 2 hours after exposure to ultraviolet radiation. (There is another DNA damage checkpoint, mediated by p53/p21, which comes into play several hours later.) Progression to S phase was arrested due to continued inhibition of CDK2 by tyrosine phosphorylation and activated CHK1 protein kinase. Ultraviolet light irradiation has long been associated with cataractogenesis. It has been proposed that the effect is

exerted by the creation of highly reactive oxygen species, which cause oxidative stress (Varma *et al*, 2011). It is possible that components of an identical or a similar pathway are affected if this putative mutation is responsible for cataractogenesis in this family. When CDC25A was overexpressed, more DNA damage was evident, and fewer cells survived, indicating that CDC25A-mediated cell cycle arrest had been compromised.

It could be expected that a gene with such a crucial role in the cell cycle would, if mutated, cause more devastating and widespread effects than the phenotype noted in this family: non-syndromic cataracts alone. This could, however, be explained by alternative splicing. There are eight alternative splicing patterns for CDC25A (according to ASTD), and exon 1, where p.Thr46_Val47delinsIle occurs, is only found in splice patterns 1, 4 and 5 (which corresponds to exon 2b in the Genecards representation in Figure 4.15), perhaps decreasing the potential impact had a similar mutation occurred in a more widely included section of the gene. Interestingly, Karagoz *et al* (2010) reported that an exon 3 CDC25A missense substitution (p.Ser88Phe) was associated with breast cancer susceptibility in humans but the same variant in mice was associated with embryonic lethality (in the homozygous state) (Bahassi *et al*, 2011).

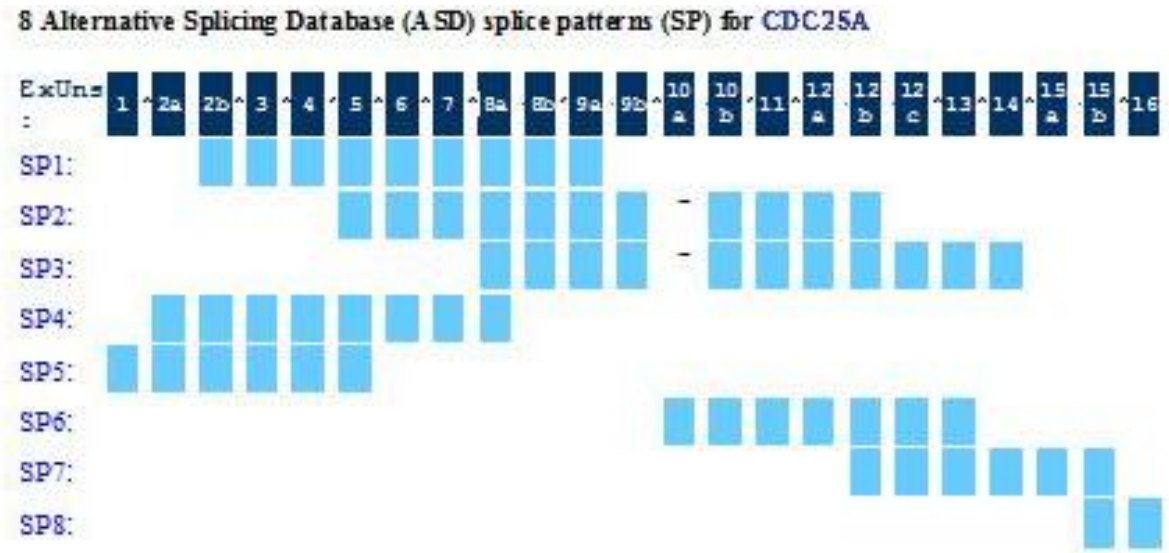


Figure 4.15: *Splicing patterns for CDC25A from Genecards.*

Chen *et al* (2011) investigated the cause of autosomal recessive congenital cataracts in twelve consanguineous Pakistani families, confirming linkage to 3p21-p22 with a summed LOD score of 33.42. Mutations in *FYCO1*, FYVE and coiled-coil domain containing 1 gene (45,959,395-46,037,307), segregated with disease status, and nine different mutations were identified. A mutation that segregated with affected status was also identified in the Israeli family that was also investigated during this project (Pras *et al*, 2001). FYCO1 is a member of the PI(3)P-binding protein family, mediating plus-end directed vesicle transport in microtubules, being associated with colocalisation to the autophagosome exterior. Mutations in *FYCO1* appear to be a significant cause of cataracts within the Pakistani population. It is interesting to note that *FYCO1* maps within the candidate autozygous region identified in Cataract Family 1. However, as *FYCO1* findings were published after laboratory work for this thesis finished, it is not known whether Cataract Family 1 harbours a *FYCO1* mutation or not. Chen's group did not investigate *CDC25A* as a candidate gene.

4.2.5 Conclusion

Any future work involving elucidation of the cause of cataractogenesis in cataract family 1 would first need to test for a mutation in *FYCO1* (Chen *et al*, 2011). The preliminary investigations undertaken to try and evaluate the significance of the CDC25A p.Thr46_Val47delinsIle variant illustrate the difficulty of establishing whether rare genetic variants are pathogenic or coincidental. Thus the variant was not detected in normal controls and segregated with disease status but the genetic information was insufficient to prove pathogenicity. Functional investigations were generally uninformative though there was some suggestion that the variant might possibly cause protein instability. If *CDC25A* was to be further investigated as a candidate gene for cataracts, it would be helpful to analyse a large number of recessively inherited cataract families to try to identify additional mutations. (The number of available families and the presence of locus heterogeneity limited the study.) In addition, further functional studies might be undertaken to compare the effects of wild-type and p.Thr46_Val47delinsIle variants on CDC25A-related functions. Thus the CellTiter 96® AQueous Non-Radioactive Cell Proliferation Assay (Promega) can be used to determine the number of viable, proliferating cells. It uses a tetrazolium compound (MTS), and an electron coupling reagent (phenazine methosulfate), and the formation of a formazan product can be measured colorimetrically at 490nm. There is a direct proportional relationship between the number of living cells and the amount of formazan formed. This could be used to compare differences between wild type-transfected and mutant-transfected CRL-11421 *in vitro*.

BrdU staining is another method that could be used to discover differences between the cell populations. 5-bromodeoxyuridine is used as a mutagen, and also as a

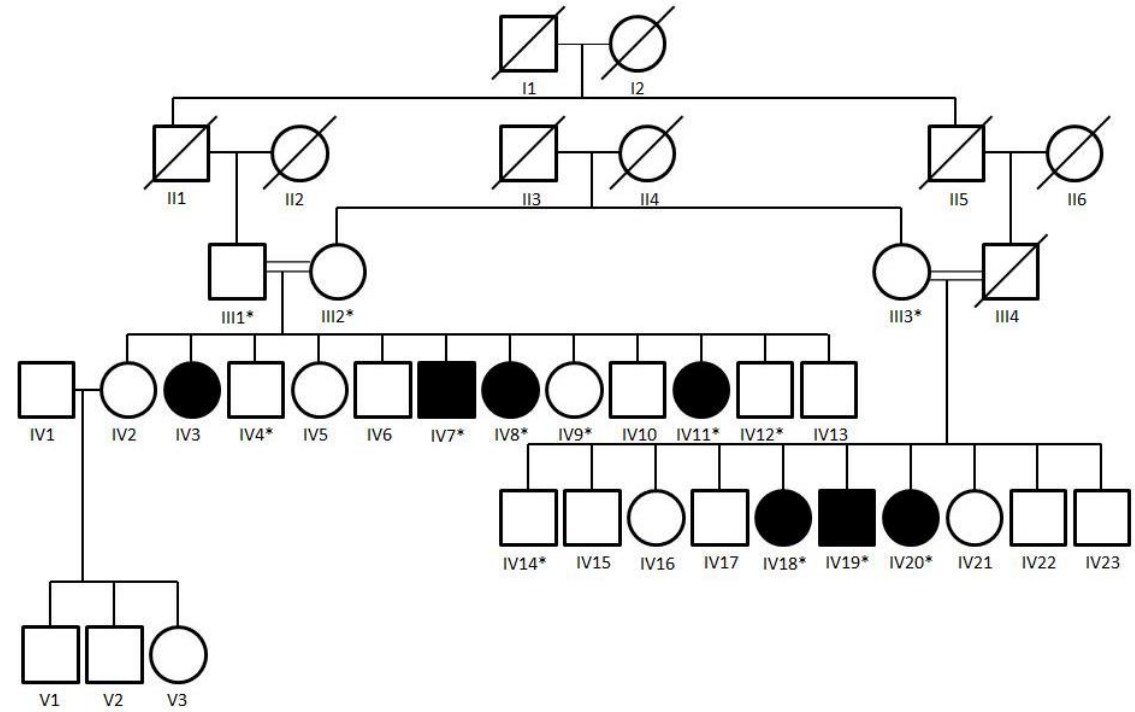
marker for DNA synthesis, due to the differing staining patterns that can be observed. It is a thymidine analogue, capable of inducing point mutations because of its tendency to tautomerization. It will pair with G instead of A when in the enol form. The protocol involves fixing cells in ethanol, resuspending in HCl with pepsin, centrifugation steps, incubation with primary and secondary antibodies, and treatment with propidium iodide.

4.3 Investigation of Cataract Family 3 and Family 4 (Omani Kindreds)

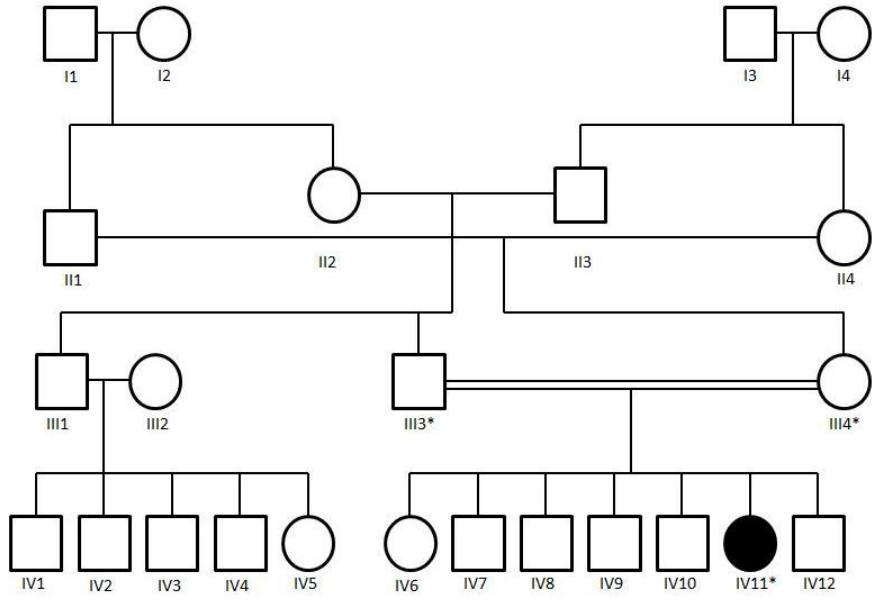
Aims: To locate shared regions of homozygosity, and to prioritise candidate genes for sequencing

4.3.1 Patient DNA

DNA was available from families in extended kindreds from Oman (Figure 4.16). Affected individuals presented with cataracts and exhibited no other clinical symptoms. The second kindred were said to be related to the first kindred as ‘distant cousins’. Peripheral blood samples and buccal swabs were obtained, after informed consent, from family members, in villages in Oman by Dr Derek Lim. Blood samples were collected from Kindred 1 family members III:1, III:2, IV:4, IV:7, IV:8, IV:9, IV:11, IV:12, III:3, IV:14, IV:18, IV:19, and IV:20, and Kindred 2 family members III:3, III:4 and IV:11. The standard phenol chloroform extraction procedure was used to obtain genomic DNA in the Birmingham Women’s Hospital Diagnostic Laboratories. Buccal swabs were taken from Kindred 1 IV:10 and IV:13, and Kindred 2 IV:7 and IV:8.



A



B

Figure 4.16: *A. Cataract Family 3 - Omani Kindred 1; B. Cataract Family 4 - Omani Kindred 2.*

4.3.2 Molecular Genetic Methods

4.3.2.1 SNP Chip

Genome-wide linkage scans were performed by Louise Tee, using the GeneChip® Human Mapping 250K SNP Array on DNA from Kindred 1 (IV:7, IV:8, IV:11, IV:12, IV:14 and IV:19). IV:19 was repeated since the call rate fell below the 93% call rate on the first attempt, having a call rate of 92%.

4.3.2.2 Microsatellite Markers

Microsatellite markers from ResGen or ABI panels or novel microsatellite markers were used. Details of these are provided in the Appendix (Table 6.36).

4.3.2.3 PCR

Forward and reverse primers were designed to flank all coding exons (and some non-coding exons) of the known cataract genes *CRYBB2* (Table 6.21), *CRYBA4* (Table 6.22), *CRYBB3* (Table 6.23), *BFSP1* (Table 6.25), *LIM2* (Table 6.26), *GCNT2* (Table 6.27), *GJA8* (Table 6.28), *CRYBB1* (Table 6.24), *HSF4* (Table 6.29), *CRYAA* (Table 6.30) and *GALK1* (Table 6.31). PCR was carried out with conditions described in the Materials and Methods chapter, unless noted otherwise in the tables in this chapter.

Primers were also designed to amplify coding regions of the following candidate genes: *LENEP* (Table 6.32), *CRYBG3* (Table 6.33), and *HSPB6* (Table 6.34).

CRYBG3 on chromosome 3 was chosen as a candidate gene because it is located from 97,595,819 to 97,663,810 and so lies in the region (74,919,074-115,297,798) identified by the genome wide scan. It was an ideal candidate gene, coding for a

crystallin, with evidence of expression in the eye (Aceview). The *in vivo* function is unknown.

LENEP (chromosome 1,154,966,062 to 154,966,791) was selected as a candidate gene because it is in the chromosome 1 region identified by genome wide scan, 150, 657,906-165,802,112. There is also evidence of lens expression (Aceview).

HSPB6 was chosen as a candidate gene. According to build NCBI36/hg18 (release 54, May 2009) this gene was located at 40,937,336-40,939,799 According to GRCh37/hg19, it is more accurately located on chromosome 19 from 36,245,470 to 36,247,930, placing it outside the chromosome 19 region detected by the genome wide scan from 40,459,706 to 58,857,408. There is evidence of expression in the eye, and three of four isoforms contain Hsp20/alpha crystallin family domains (Aceview). HSPB6 is involved in smooth muscle relaxation modulation (Tessier *et al* 2003).

4.3.2.4 Sequencing

Sequencing was carried out as described in the Materials and Methods chapter (see 2.2.6).

4.3.3 Results

DNA concentrations from the buccal swabs taken from Kindred 1 IV:10 and IV:13, and Kindred 2 IV:7 and IV:8, were too low, and as such, DNA could not be amplified via PCR for any of the exons attempted.

4.3.3.1 Genome-Wide Scan

After SNP chip analysis (based on build GRCh36/hg18), fifteen regions of homozygosity in Kindred 1 members IV:8 and IV:11 greater than 2Mb were considered possible candidate regions (Table 4.5). Two of these regions were not analysed with microsatellite markers since the third affected member, IV:7, was extremely heterozygous, and in this consanguineous family, the transmission of affected status was initially thought to be most likely to be autosomal recessive with all affected members being homozygous for the same mutation.

Chromosome	Start	End	Notes
1	105,464,488	107,660,193	Affected IV:7 is quite heterozygous
1	150, 657,906	165,802,112	Unaffected IV:12 and IV:14 mostly heterozygous. Affected IV:7 heterozygous in parts
2	131,671,717	144,316,267	Affected IV:7 is quite heterozygous
3	74,919,074	115,297,798	Affected IV:7 is quite heterozygous
6	88,878,455	91,450,041	Affected IV:7 is mostly heterozygous
6	95,956,306	100,444,677	Affected IV:7 is mostly heterozygous
7	116,265,885	119,025,178	Affected IV:7 is mostly heterozygous
11	31,389,403	34,497,225	Affected IV:7 is mostly heterozygous
11	38,567,581	40,614,578	Affected IV:7 is mostly heterozygous
14	87,629,745	94,999,444	Affected IV:7 is mostly heterozygous
16	27,176,280	47,670,664	Unaffected have similar genotype
19	17,270,289	37,787,896	Affected IV:7 is heterozygous
19	40,459,706	58,857,408	Affected IV:7 is heterozygous, but from 44.6 has same genotype as IV:8 and IV:11
20	39,967,656	41,904,898	Affected IV:7 is mostly heterozygous
22	35,860,982	41,897,898	Affected IV:7 is quite heterozygous

Table 4.5: *Regions analysed with microsatellite markers in the Omani Kindreds. Grey shading indicates that the regions were not selected for analysis with microsatellite markers.*

4.3.3.2 Microsatellite Markers

4.3.3.2.1 Chromosome 19

Fine mapping analysis with microsatellite markers for the region on chromosome 19 from 40,459, 706 to 58, 857, 408 was completed first (Figure 4.17) since the known cataract gene *LIM2* is located in this region from 56,574,976 to 56,583,009 (NCBI36/hg18). In the NCBI37/hg19 database release, *LIM2* is located 51,883,164-51,891,197. Twenty-nine microsatellite markers were used to investigate this region, from Chr19r2-16xAT (novel marker) (41,093,388) to D19S572 (58,797,162). D19S220 (43,123,394) showed linkage for all affected family members except kindred 2 family member IV:20. In kindred 1 family 1, a common homozygous haplotype was observed for all affected members (IV:7, IV:8 and IV;11) between Chr19r2-16xAT (novel marker) (41,093,388) and Chr19r2-15xGT (novel marker) (55,468,671), and Chr19-15xAG (novel marker) (55,215,067) and D19S571 (57,968,784). In kindred 1 family 2, this was observed from D19S220 (43,123,394) to D19S421 (43,562,945), and from D19S47 (45,006,014) to Chr19r2-15xTTA (novel marker) (51,999,196), but only in affected individuals IV:18 and IV:19, not in affected individual IV:20. Homozygous regions were noted in kindred 2 IV:11 from Chr19r2-15xAC41 (novel marker) (41,435,658) to D19S220 (43,123, 394), and from Chr19r2-15xTAT (44,221,282) to Chr19r2-22xGT44 (novel marker) (44,669,155).

The region was not found to be linked, so analysis of the other 12 regions, using fine mapping with microsatellite markers, was carried out.

4.3.3.2.2 Chromosome 1

In chromosome 1 region 1 (105,464,488 to 107,660,193), linkage was excluded due to microsatellite marker segregation and haplotypes (Figure 4.18). Seven microsatellite markers were analysed across this region from D1S485 (105,069,360) to D1S2726 (110,896,304).

In chromosome 1 region 2 (150, 657,906 to 165,802,112), sixteen microsatellite markers across the region were analysed. Two common homozygous regions were found in kindred 1 family 2, and in kindred 2. The regions extend between D1S498 (148,114,568) and Chr1-26xAC (novel marker) (153,111,478), and D1S1653 (156,199,398) and D1S1677 (161,826,325), although haplotypes were not identical, and the calls for Homozygosity extended as far as D1S1589 (172, 527,725) in kindred 2 family member IV:11. In kindred 1 family 1, unaffected individuals IV:4 and IV:9 shared a common homozygous region with identical haplotypes as those of affected individuals IV:7, IV:8 and IV:11 from Chr1-22xGT (novel marker) (152,009,159) to D1S2673 (163,312,611). Assuming that the same mutation is responsible for cataractogenesis in all affected individuals, this region was not linked (Figure 4.19).

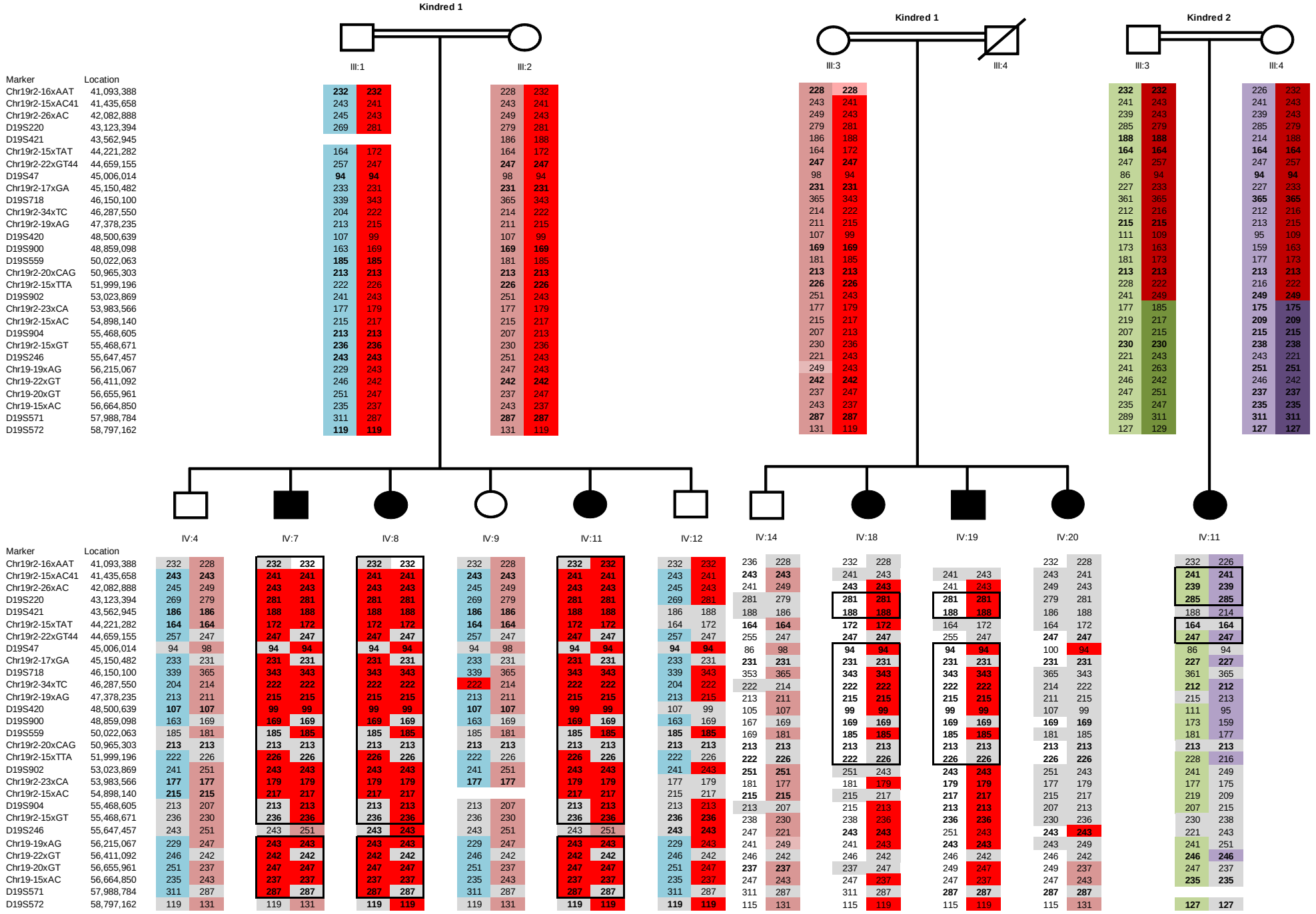


Figure 4.17: *Chromosome 19 region 2 (40,459,706-58,857,408). Affected individuals are shaded black. Markers displayed in order of physical location. Boxed areas contain homozygous haplotypes.*

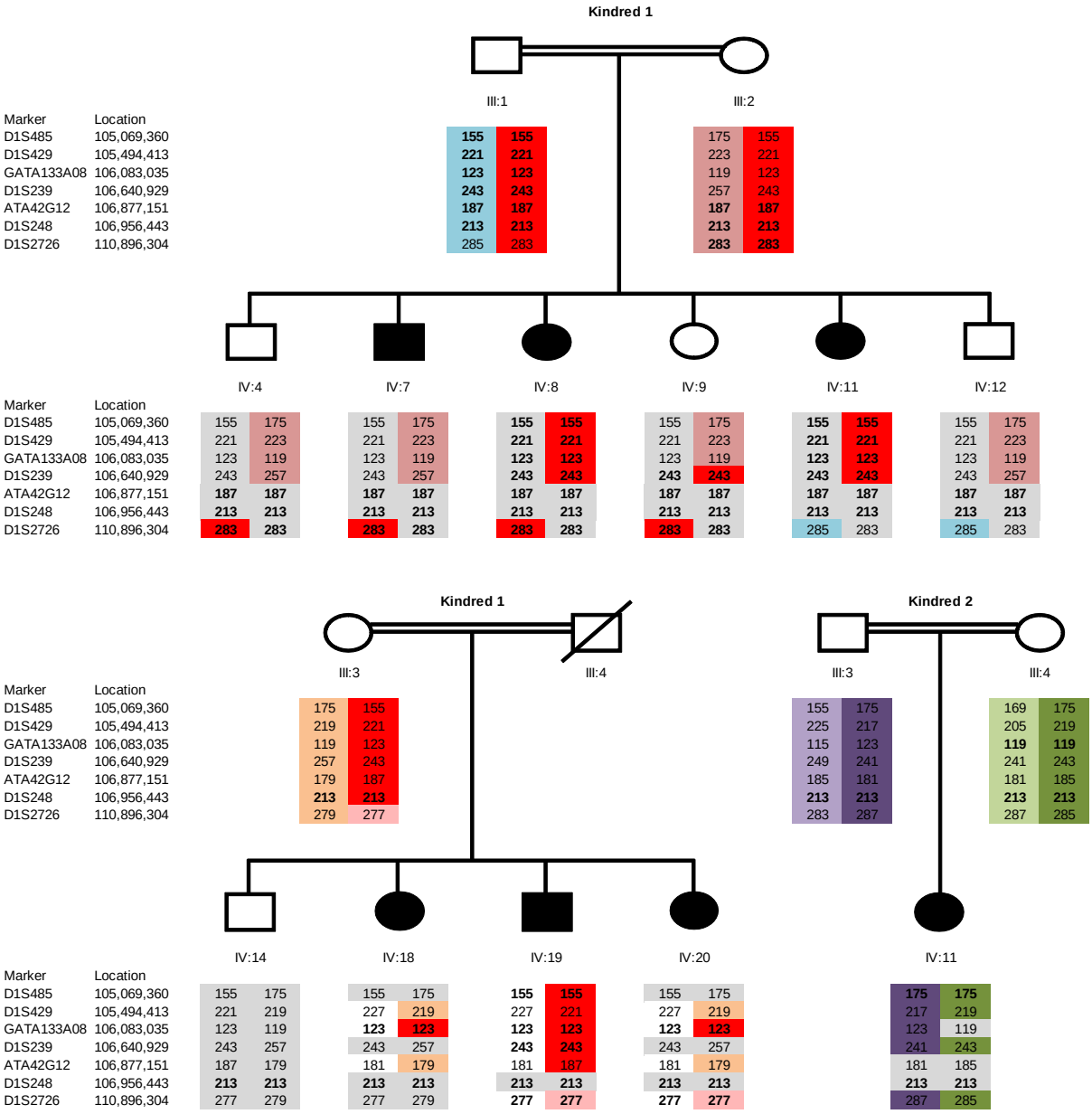


Figure 4.18: *Chromosome 1 region 1 (105,464,488 to 107,660,193). Affected individuals are shaded in black. Markers are displayed in order of physical location.*

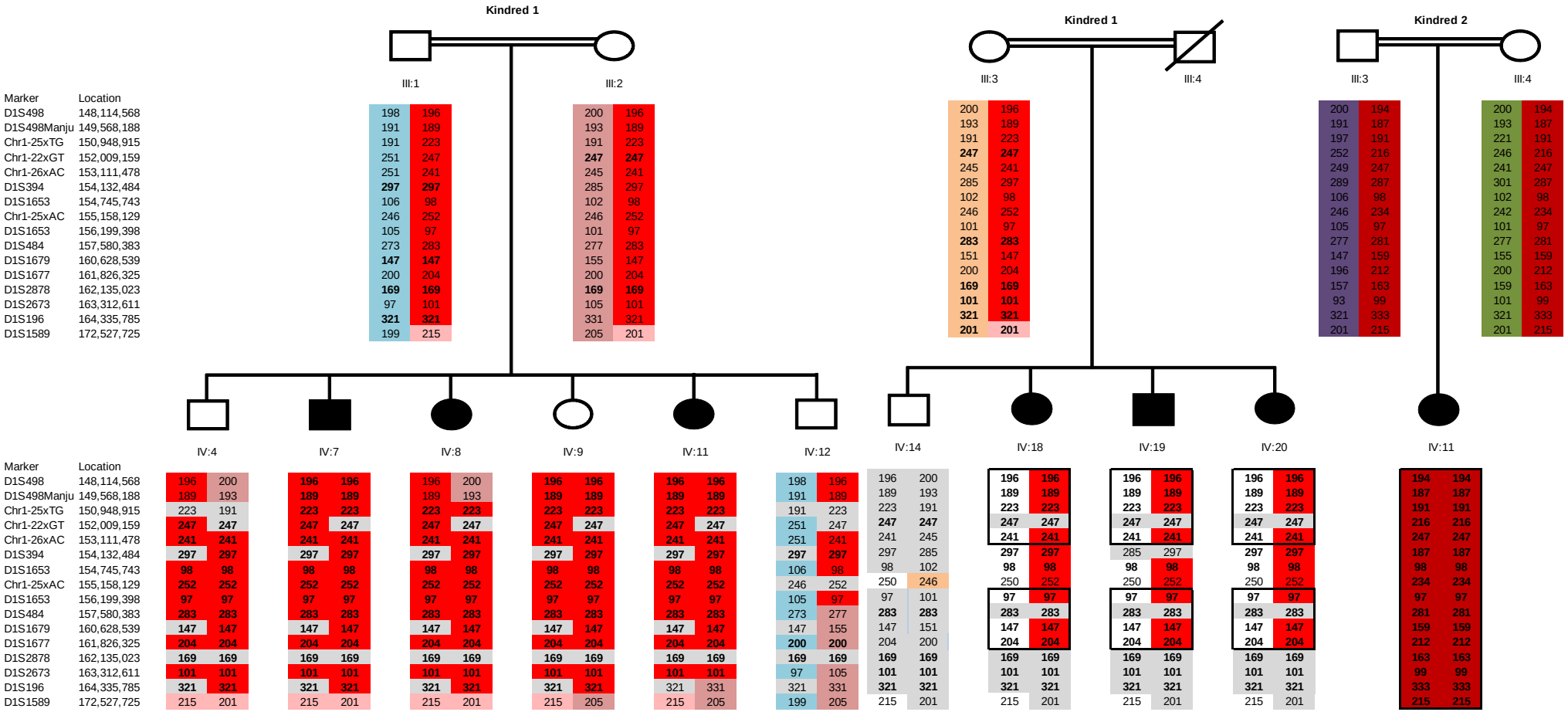


Figure 4.19: *Chromosome 1 region 2 (150, 657,906 to 165,802,112). Affected individuals are shaded in black. Markers are displayed in order of physical location. Black boxes around haplotypes indicate a common homozygous region, with identical haplotypes within the same family*.

4.3.3.2.3 Chromosome 2

For the region on chromosome 2 (131,671,717 to 144,316,267), eleven microsatellite markers were analyzed. This region was not linked in kindred 1, but in kindred 2, IV:11, homozygosity was identified from D2S112 (132,925,167) to D2S132 (144,683,977) (Figure 4.20).

4.3.3.2.4 Chromosome 3

For the region on chromosome 3 (74,919,074 to 115,297,798), 26 microsatellite markers were analysed (Figure 4.21). In kindred 1 family 1, a common homozygous haplotype was found in all three affected family members, IV:7, IV:8 and IV:11. This was identified from Chr3-16xTTA (novel marker) (75,301,460) to Chr3-16xTC (novel marker) (114,880,547). In kindred 2, IV:11 had a homozygous region from D3S3045 (108,472,546) to Chr3-16xTC (novel marker) (114,880,547).

4.3.3.2.5 Chromosome 6

Chromosome 6 region 1 extends from 88,878,455 to 91,450,041 (Figure 4.22). Four microsatellite markers were analysed from Chr6r1-15xATA (novel marker) (88.893,581) to Chr6-26xCA (novel marker) (91,342,860). No linkage was found. Affected individuals IV:8 and IV:11 in kindred 1 family 1 share a common homozygous haplotype across this region, but the other affected individual IV:7 is heterozygous, having inherited two different alleles. This region was excluded.

Chromosome 6 region 2 is between 95,956,306 and 100,444,677 (Figure 4.23). Five microsatellite markers were analysed, from Chr6-30xTA (novel marker) (95,957,151) to Chr6-22xTG (novel marker) (99,826, 744), but no linkage was detected. Affected individuals IV:8 and IV:11 in kindred 1 family 1 again share a common homozygous

haplotype across this region, but again, the other affected individual IV:7 is heterozygous, having inherited two different alleles. This region was excluded.

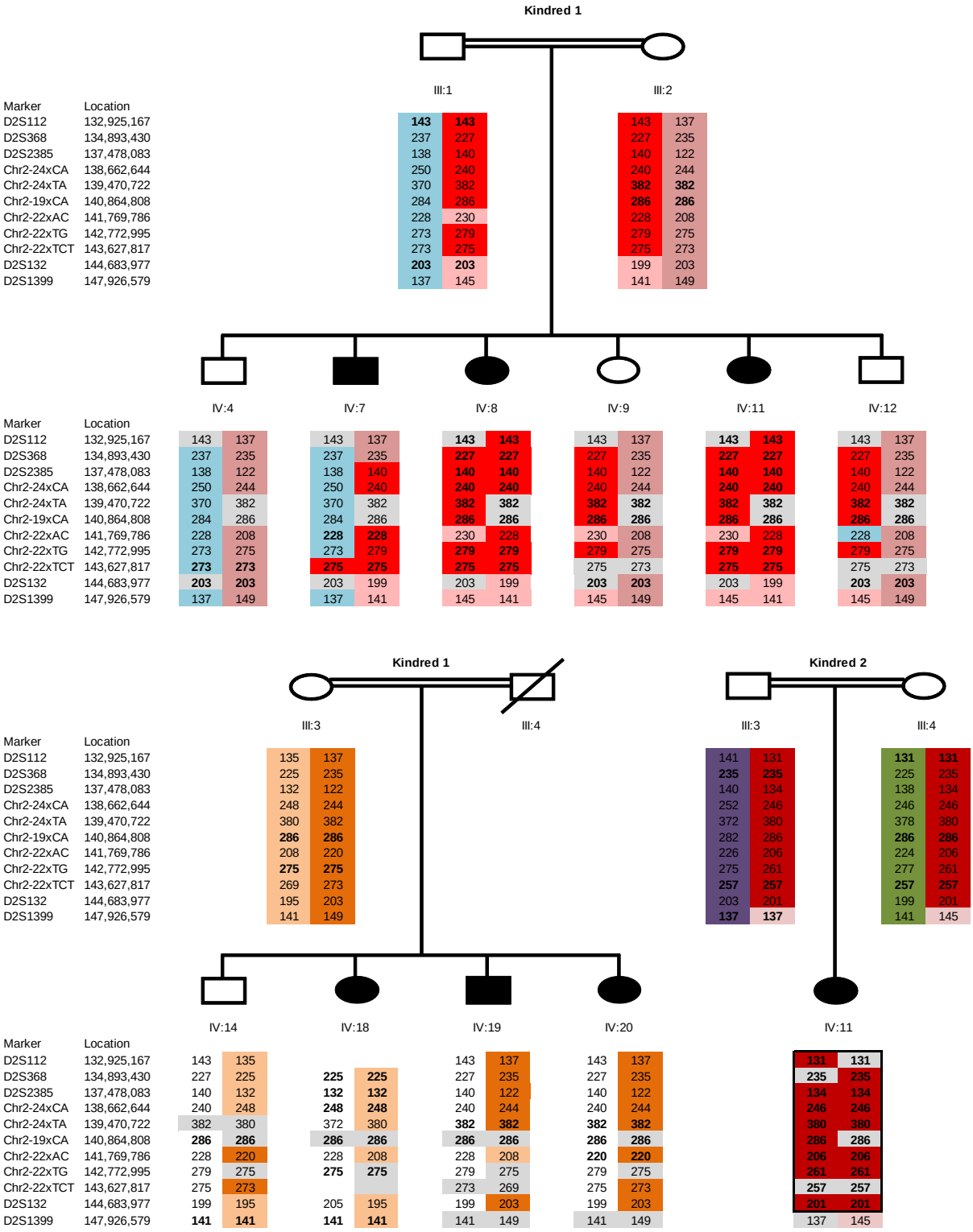
4.3.3.2.6 Chromosome 7

The region on chromosome 7 extended from 116,265,885 to 119,025,178 (Figure 4.24). Three microsatellite markers were analysed across this region from 17xATT (novel marker) (117,254,117) to 29xTA (novel marker) (118,671,349) but no common homozygous haplotype was found. A region of homozygosity for kindred 2 family member IV:11 was detected from 17xATT to 23xAT (novel marker) (117,998,458).

4.3.3.2.7 Chromosome 11

Chromosome 11 region 1 (31,389,403 to 34,497,225) (Figure 4.25) was investigated with four microsatellite markers from Chr11r1-20xAC (novel marker) (31,577,606) to Chr11r1-18xAC (novel marker) (34,192,263). No common homozygous haplotypes were detected across all three families, but Chr11r1-20xAC was linked for kindred 1 family 1.

Chromosome 11 region 2 (38,567,581 to 40,614,578) (Figure 4.26) was investigated with three microsatellite markers from Chr11r2-29xGT (novel marker) (38,838, 292) to Chr11r2-18xTG (novel marker) but no common homozygous haplotypes were found.



Marker	Location	III:1		III:2		IV:4		IV:7		IV:8		IV:9		IV:11		IV:12	
D2S112	132,925,167	**143**	**143**	143	137	143	137	143	137	**143**	**143**	143	137	**143**	**143**	143	137
D2S368	134,893,430	237	227	227	235	237	235	237	235	**227**	**227**	227	235	**227**	**227**	227	235
D2S2385	137,478,083	138	140	140	122	138	122	138	140	**140**	**140**	140	122	**140**	**140**	140	122
Chr2-24xCA	138,662,644	250	240	240	244	250	244	250	240	**240**	**240**	240	244	**240**	**240**	240	244
Chr2-24xTA	139,470,722	370	382	**382**	**382**	370	382	370	382	**382**	**382**	**382**	**382**	**382**	**382**	**382**	**382**
Chr2-19xCA	140,864,808	284	286	**286**	**286**	284	286	284	286	**286**	**286**	**286**	**286**	**286**	**286**	**286**	**286**
Chr2-22xAC	141,769,786	228	230	228	208	228	208	**228**	**228**	230	228	230	208	230	228	228	208
Chr2-22xTG	142,772,995	273	279	279	275	273	275	273	279	**279**	**279**	279	275	**279**	**279**	279	275
Chr2-22xTCT	143,627,817	273	275	275	273	**273**	**273**	**275**	**275**	**275**	**275**	275	273	**275**	**275**	275	273
D2S132	144,683,977	**203**	**203**	199	203	**203**	**203**	203	199	203	199	**203**	**203**	203	199	**203**	**203**
D2S1399	147,926,579	137	145	141	149	137	149	137	141	145	141	145	149	145	141	145	149

Marker	Location	III:3		III:4		III:3		III:4		IV:14		IV:18		IV:19		IV:20		IV:11	
D2S112	132,925,167	135	137			141	131	**131**	**131**	143	135			143	137	143	137	**131**	**131**
D2S368	134,893,430	225	235			**235**	**235**	225	235	227	225	**225**	**225**	227	235	227	235	**235**	**235**
D2S2385	137,478,083	132	122			140	134	138	134	140	132	**132**	**132**	140	122	140	122	**134**	**134**
Chr2-24xCA	138,662,644	248	244			252	246	246	246	240	248	**248**	**248**	240	244	240	244	**246**	**246**
Chr2-24xTA	139,470,722	380	382			372	380	378	380	382	380	372	380	**382**	**382**	**382**	**382**	**380**	**380**
Chr2-19xCA	140,864,808	**286**	**286**			282	286	**286**	**286**	**286**	**286**	**286**	**286**	**286**	**286**	**286**	**286**	**286**	**286**
Chr2-22xAC	141,769,786	208	220			226	206	224	206	228	220	228	208	228	208	**220**	**220**	**206**	**206**
Chr2-22xTG	142,772,995	**275**	**275**			275	261	277	261	279	275	**275**	**275**	279	275	279	275	**261**	**261**
Chr2-22xTCT	143,627,817	269	273			**257**	**257**	**257**	**257**	275	273			273	269	275	273	**257**	**257**
D2S132	144,683,977	195	203			203	201	199	201	199	195	205	195	199	203	199	203	**201**	**201**
D2S1399	147,926,579	141	149			**137**	**137**	141	145	**141**	**141**	**141**	**141**	141	149	141	149	137	145

Figure 4.20: *Chromosome 2 (131,671,717 to 144,316,267) Affected individuals are shaded in black. Markers are displayed in order of physical location. The black box around haplotypes in IV:11 indicates a region of homozygosity*.

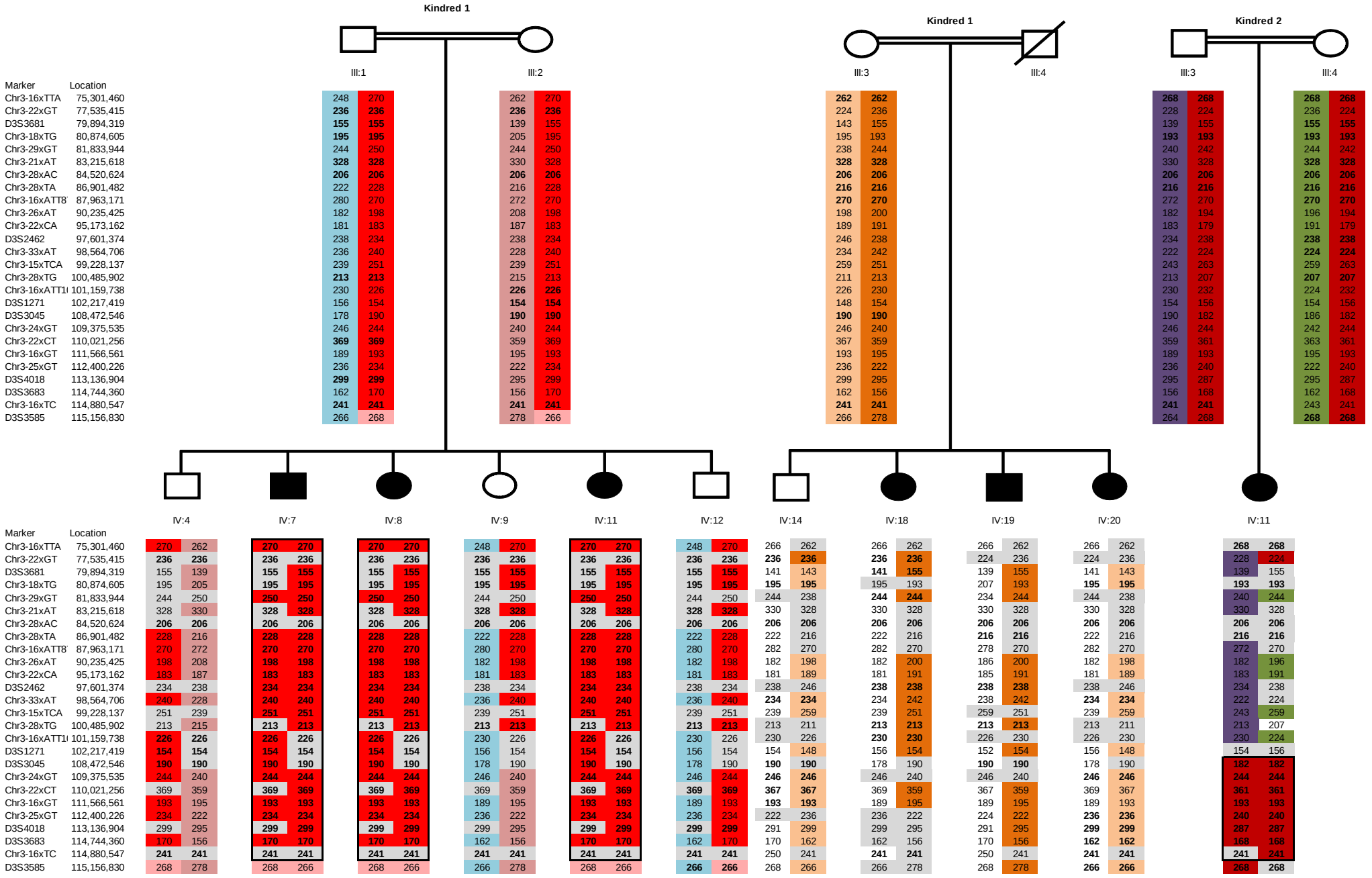


Figure 4.21: *Chromosome 3 (74,919,074 to 115,297,798). Affected individuals are shaded in black. Markers are displayed in order of physical location. Black boxes around haplotypes indicate homozygous linked regions.*

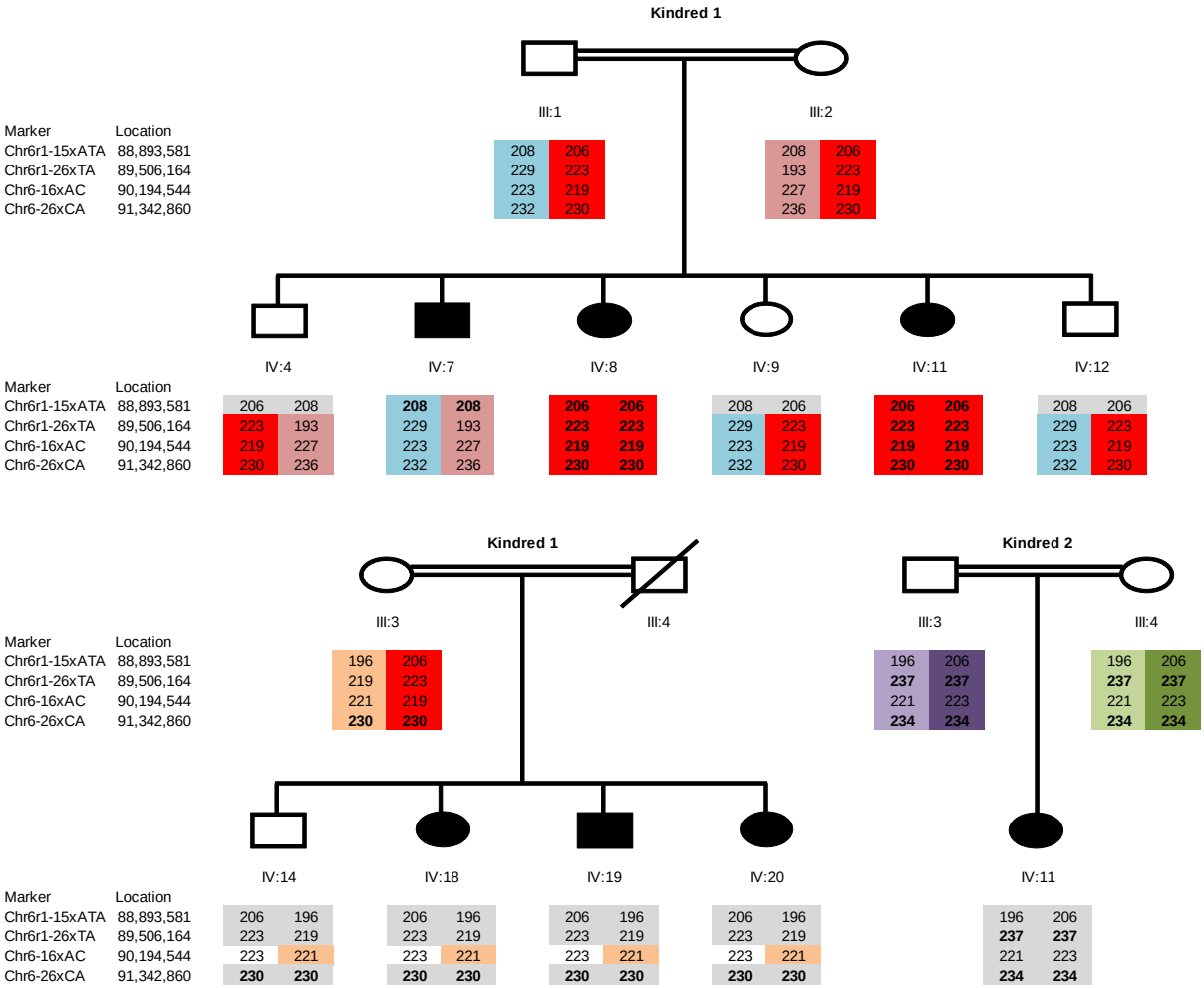


Figure 4.22: *Chromosome 6 region 1 (88,878,455 to 91,450,041). Affected individuals are shaded in black. Markers are displayed in order of physical location.*

4.3.3.2.8 Chromosome 14

The region on chromosome 14 (87,629,745 to 94,999,444) was investigated with seven microsatellite markers from D14S68 (87,697,387) to D14S1434 (94,377,875) (Figure 4.27). No common homozygous haplotype was found, although kindred 2 family member IV:11 was homozygous from D14S68 to D14S280 (91,252,619).

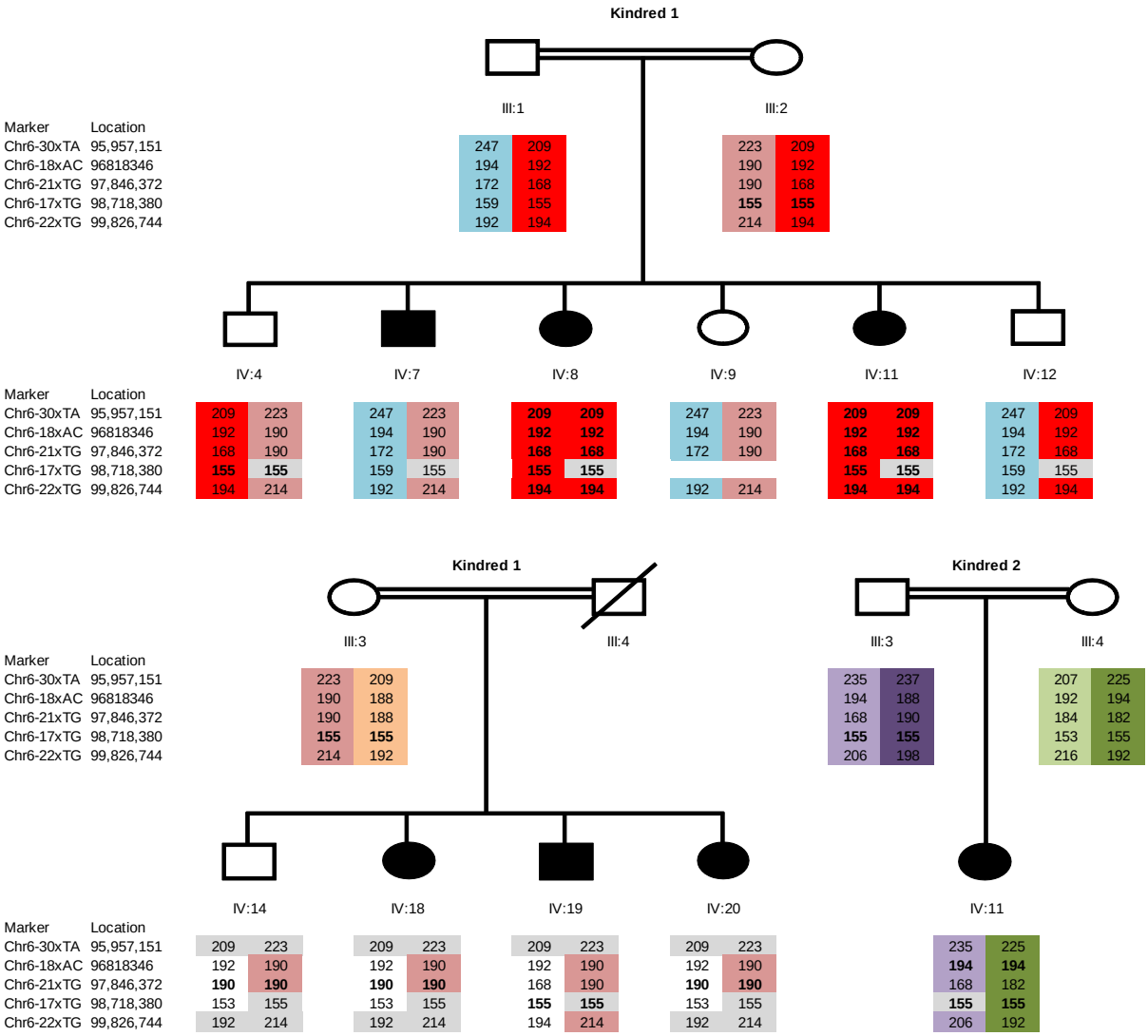


Figure 4.23: *Chromosome 6 region 2 (95,956,306 and 100,444,677). Affected individuals are shaded in black. Markers are displayed in order of physical location.*

4.3.3.2.9 Chromosome 16

The region on chromosome 16 (27,176,280 to 47,670,664) was investigated with six microsatellite markers, from D16S685 (30,575,220) to Chr16-23xAC (novel marker) (46,392,429). No common homozygous haplotype was detected amongst affected individuals (Figure 4.28).

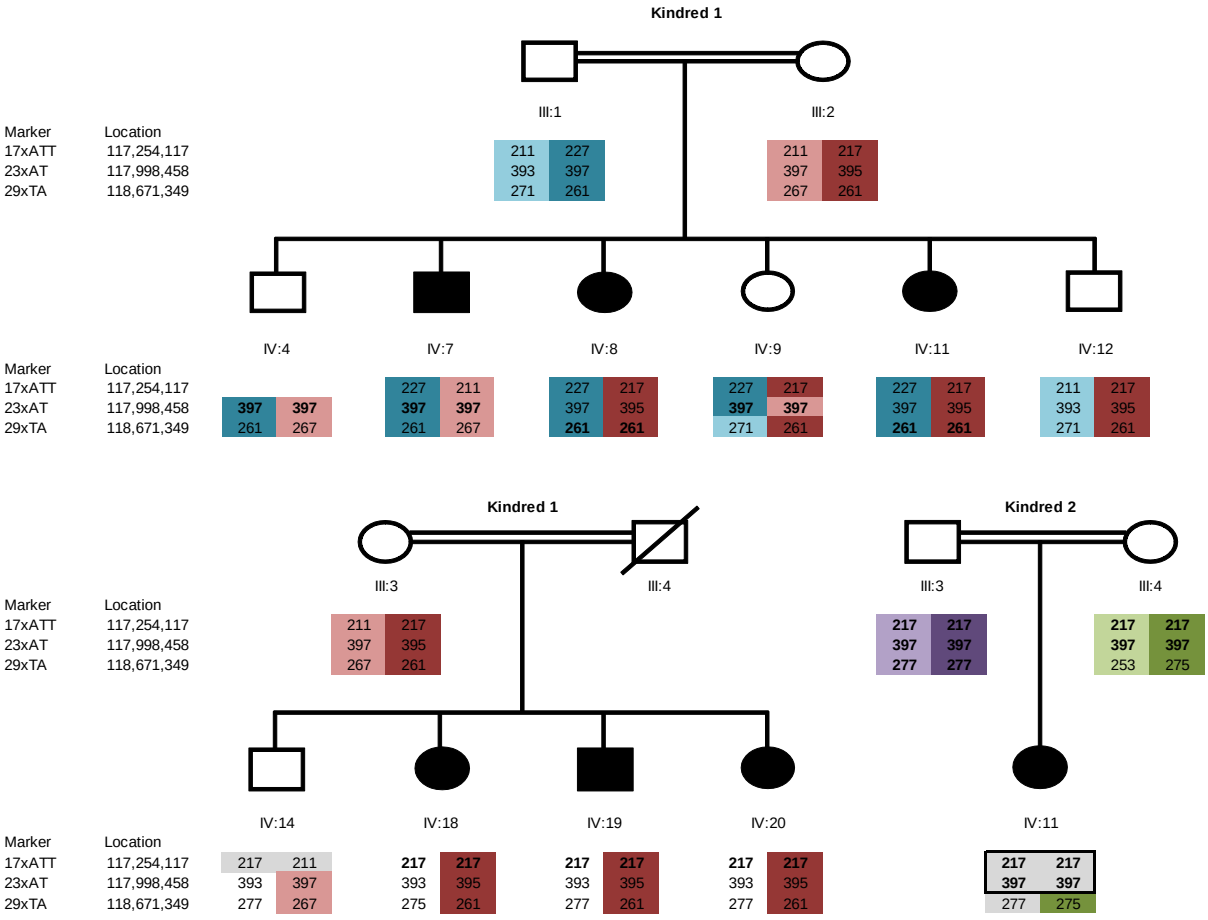


Figure 4.24: *Chromosome 7 (116,265,885 to 119,025,178). Affected individuals are shaded in black. Markers are displayed in order of physical location*.

4.3.3.2.10 Chromosome 22

The region on chromosome 22 from 35,860,982 to 41,897,898 was investigated with seven microsatellite markers (Figure 4.29) from D22S445 (35,895,844) to D22S1151 (41,865,323), but no common haplotype was noted in affected individuals.

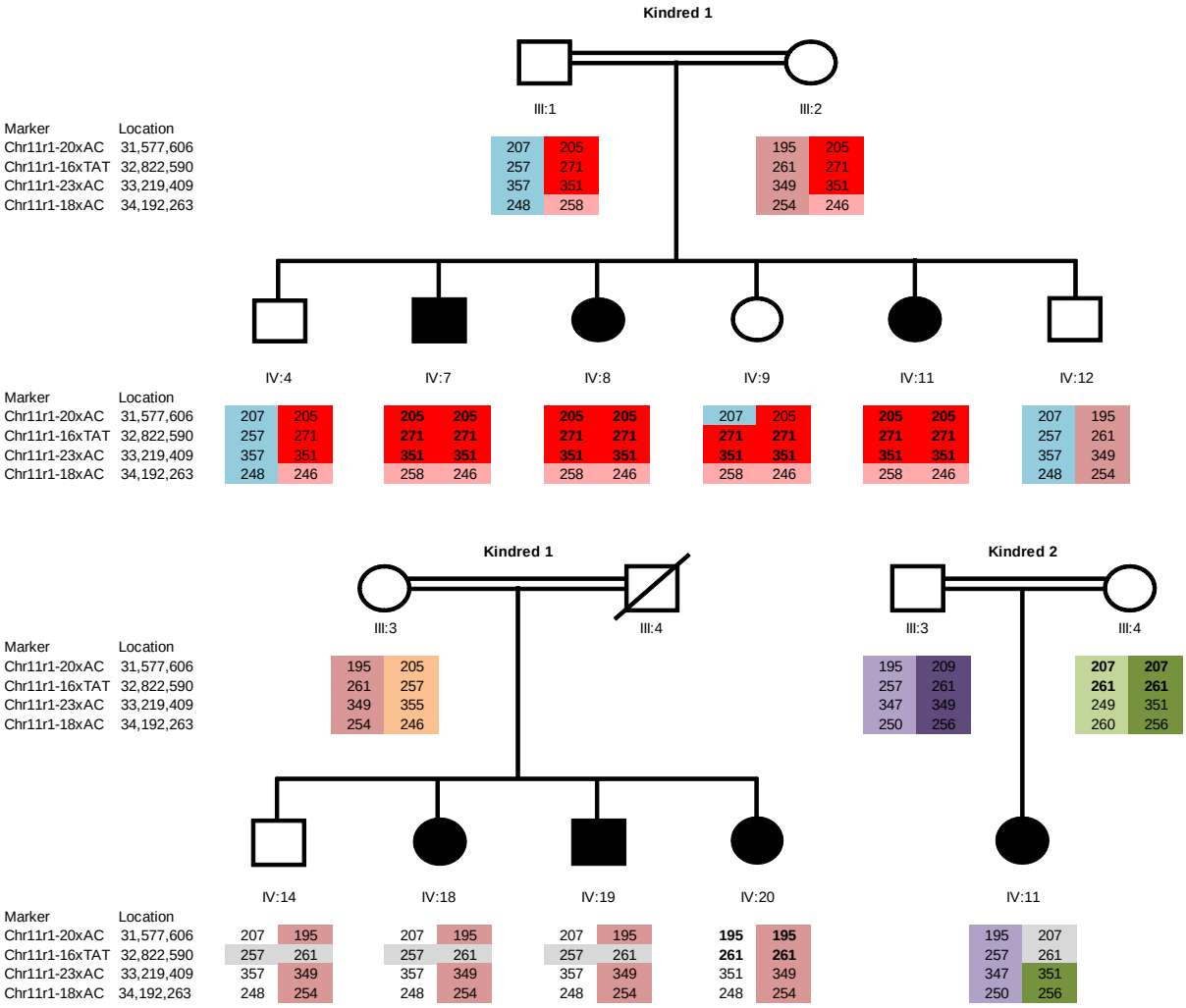


Figure 4.25: *Chromosome 11 region 1 (31,389,403 to 34,497,225). Affected individuals are shaded in black. Markers are displayed in order of physical location.*

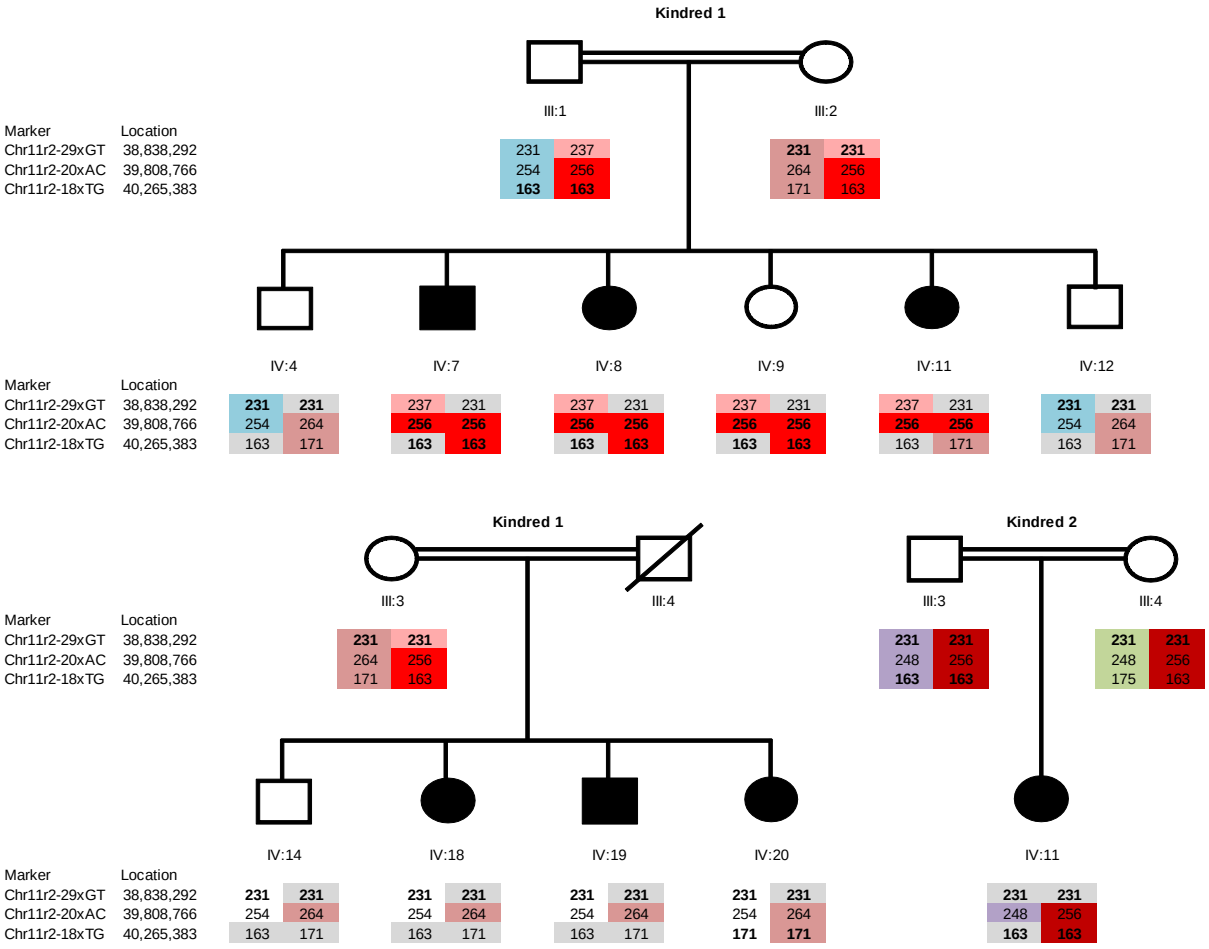


Figure 4.26: *Chromosome 11 region 2 (38,567,581 to 40,614,578). Affected individuals are shaded in black. Markers are displayed in order of physical location.*

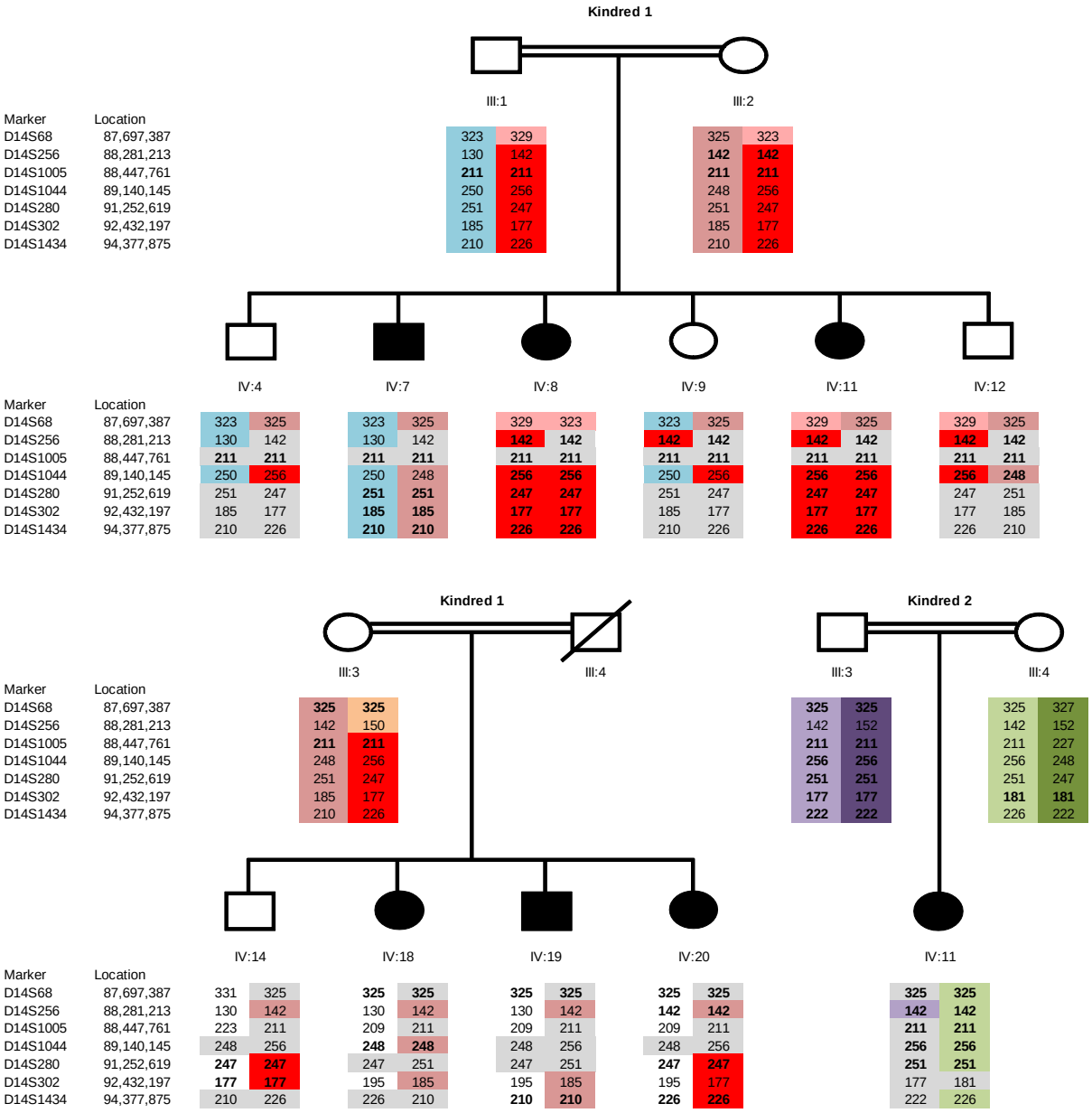


Figure 4.27: *Chromosome 14 (87,629,745 to 94,999,444). Affected individuals are shaded in black. Markers are displayed in order of physical location.*

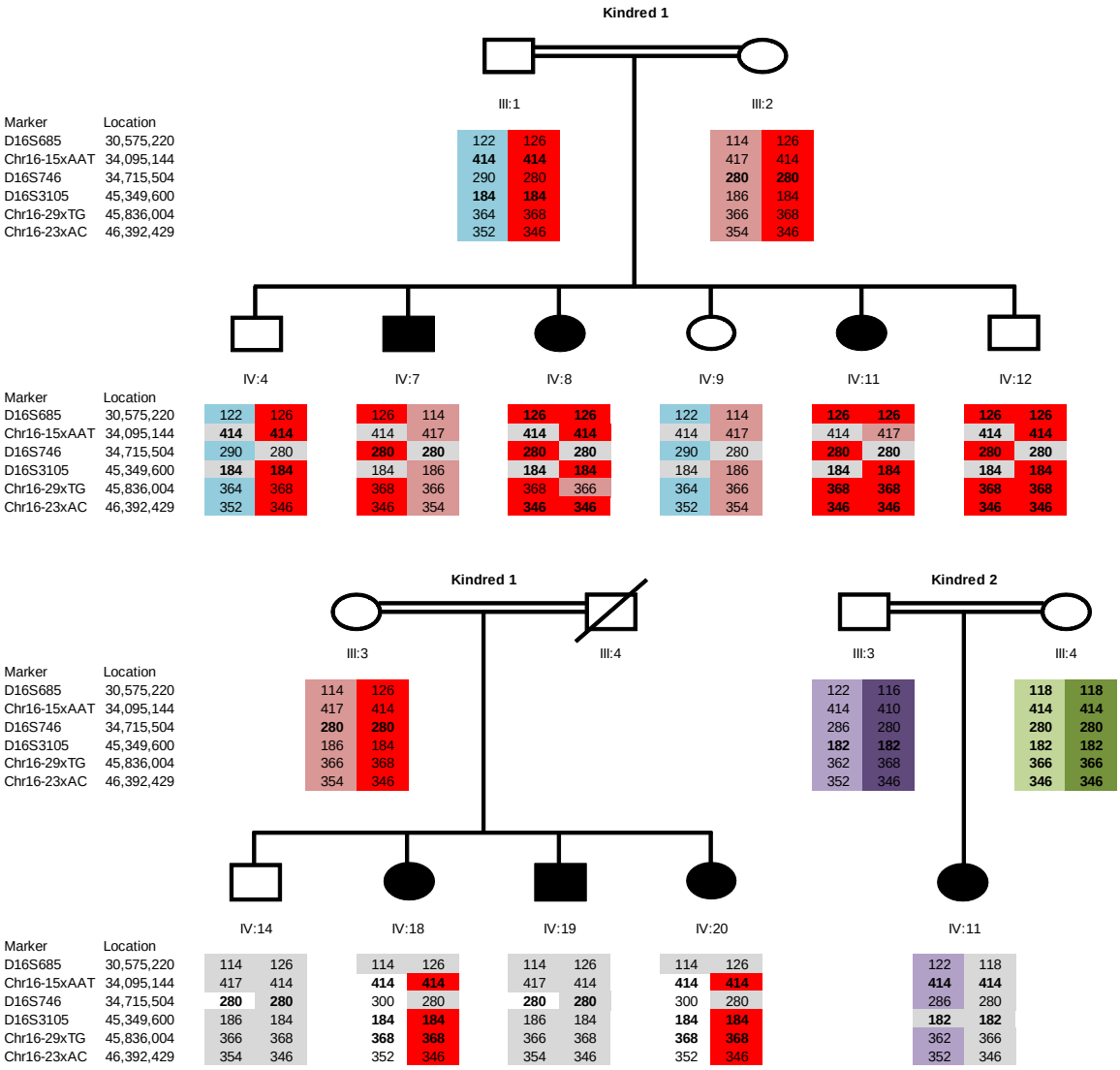


Figure 4.28: *Chromosome 16 (27,176,280 to 47,670,664). Affected individuals are shaded in black. Markers are displayed in order of physical location.*

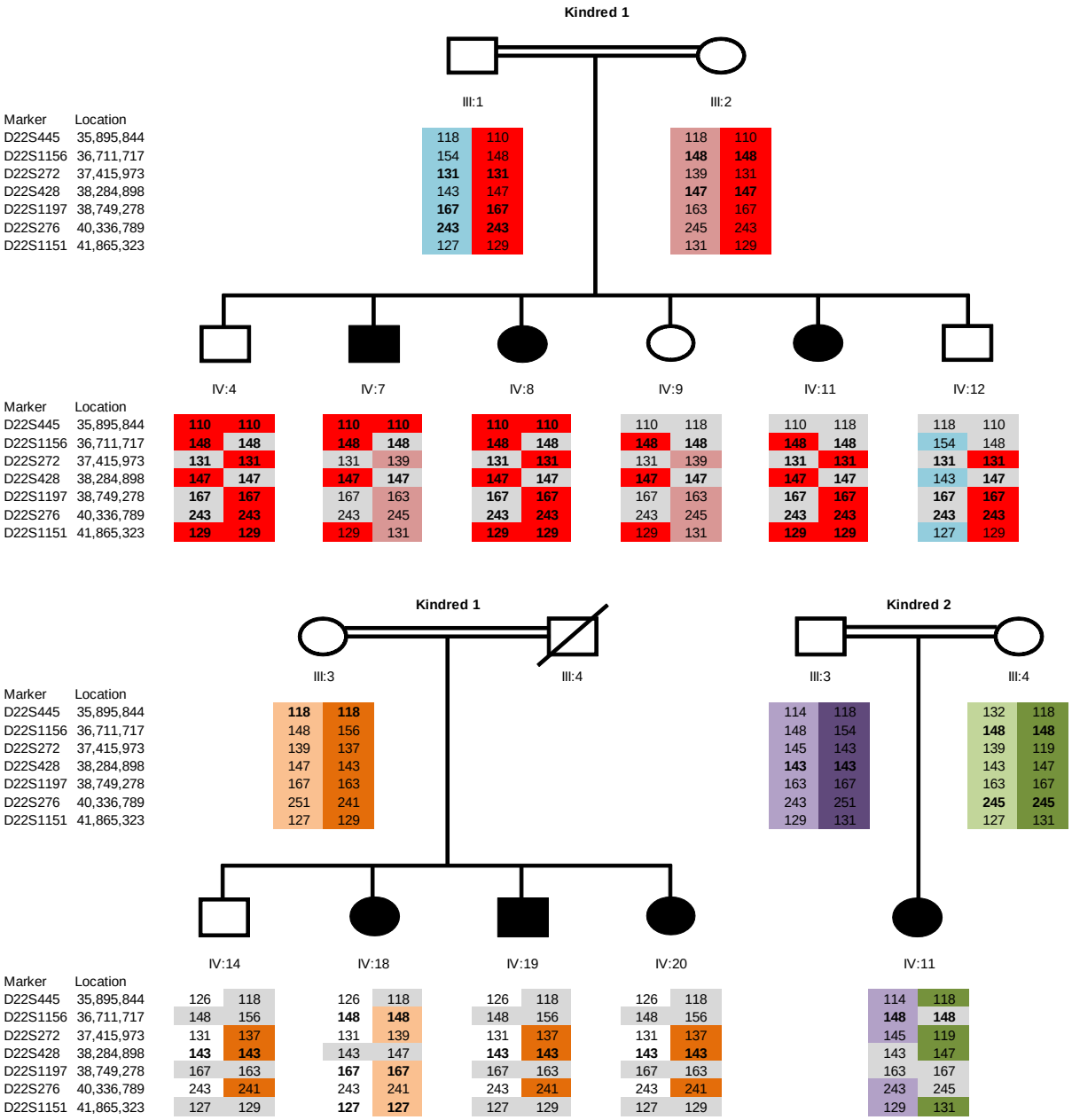


Figure 4.29: *Chromosome 22 (35,860,982 to 41,897,898). Affected individuals are shaded in black. Markers are displayed in order of physical location.*

4.3.3.4 Sequencing of Known Cataract Genes

The following known cataract genes were sequenced for kindred 1 family members III:1, IV:7, II:3 and IV:18, and kindred 2 family members (unless otherwise noted in Table 4.5), using primers to amplify exons and intron/exon boundaries as detailed previously: *CRYBB3* (Table 6.23), *BFSP1* (Table 6.25), *LIM2* (Table 6.26), *GCNT2*

(Table 6.27), *GALK1* (Table 6.31), *GJA8* (Table 6.28), *CRYBB1* (Table 6.24), *HSF4* (Table 6.29), *CRYAA* (Table 6.30), *CRYBB2* (Table 6.21), *CRYBA4* (Table 6.22) and *BFSP2* (Table 635).

Although many known SNPs were identified, no putative pathogenic mutations were detected (Table 4.6).

	Kindred 1 Family 1		**Kindred 1 Family 2**		**Kindred 2 Family 1**	
	III:2 (P)	**IV:8 (A)**	**III:3 (P)**	**IV:19 (A)**	**III:3 (P)**	**IV:11 (A)**
CRYBB3-1	WT	WT	WT	WT	WT	WT
CRYBB3-2	WT (rs 2269672 TT)	WT (rs=TT)	WT (rs=CT)	WT (rs=CT)	WT (rs=TT)	WT (rs=CT)
CRYBB3-3	WT	WT	WT	WT	WT	WT
CRYBB3-4	WT	WT	WT	WT	WT	WT
CRYBB3-5	WT (hmz for rs960378 c.337C>G, p.His113 Asp	WT (hmz for rs960378 c.337C>G, p.His113 Asp	WT (htz for rs960378 c.337C>G, p.His113 Asp	WT (hmz G)	WT (htz for rs960378 c.337C>G, p.His113Asp	WT (htz for rs960378 c.337C>G, p.His113Asp
CRYBB3-6	WT	WT	WT	WT	WT	WT
BFSP1-1	WT	WT	WT	WT	*WT	WT
BFSP1-2	WT	WT	WT	WT	*WT	WT
BFSP1-3	WT	WT	WT	WT	*WT	WT
BFSP1-4	WT	WT	WT	htz c.576-42G>A	*htz c.576-42G>A	WT
BFSP1-5	WT	WT	WT	WT	*WT	WT
BFSP1-6	WT	WT	WT	htz c.776-45A>G	*htz c.776-45A>G	hmz c.776-45A>G
BFSP1-7	WT	WT	WT	WT	*htz AG rs6080719	hmz Grs6080719
BFSP1-8A	WT	WT	WT	WT	*WT	WT
BFSP1-8B	WT	WT	WT	WT	*htz AG rs6136118	hmz G rs6136118
BFSP1-8C	hmz G rs6080718	htz AG rs6080718	hmz G rs6080718	hmz G rs6080718	*hmz G rs6080718, htz CT rs6080717	hmz G rs6080718, hmz T rs6080717
BFSP1-T2-1F	WT	WT	WT	WT	*WT	WT
LIM2-2	WT	***WT	WT	WT	*WT	WT
LIM2-3	WT	***WT	WT	WT	*WT	WT
LIM2-4	WT (htz CT int4 rs2547318)	***WT (hmz int4 T rs2547318)	WT (htz CT int4 rs2547318)	WT (hmz int4 T rs2547318)	*WT (hmz int4 T rs2547318)	WT (hmz int4 T rs2547318)
LIM2-5	x	***WT	x	WT	*WT	WT

	Kindred 1 Family 1		Kindred 1 Family 2		Kindred 2 Family 1	
	III:2 (P)	**IV:8 (A)**	**III:3 (P)**	**IV:19 (A)**	**III:3 (P)**	**IV:11 (A)**
GCNT2-001-1A	**hmz C	***htz TC rs2230906	htz TC rs2230906	htz TC rs2230906	hmz C	hmz C
GCNT2-001-1B	**WT	***WT	WT	WT	WT	WT
GCNT2-001-1C	**WT	***WT	WT	WT	WT	WT
GCNT2-001-2	**WT	***WT	WT	WT	WT	WT
GCNT2-001-3	**WT	***WT	WT	WT	WT	WT
GCNT2-004-1A	**WT	***WT	WT	WT	WT	WT
GCNT2-004-1BF	**WT	WT	WT	WT	WT	WT
GCNT2-004-1C	**hmz G rs539351	***hmz G rs539351	hmz G rs539351	hmz G rs539351	hmz G rs539351	hmz G rs539351
GCNT2-006-3A	**htz CT rs557441	***htz CT rs557441, htz AC -26 from start of ex3 coding	hmz T rs557441, htz AC -26 from start of ex3 coding	hmz T rs557441, htz AC -26 from start of ex3 coding	hmz T rs557441	hmz T rs557441
GCNT2-006-3B	**WT	***WT	WT	WT	WT	WT
GCNT2-006-3C	**WT	***htz E>K298, G>A892, unknown SNP	htz E>K298, G>A892, unknown SNP	htz E>K298, G>A892, unknown SNP	WT-wt	WT-wt
GCNT2-010-3	**WT	***WT	htz AT rs66794874	htz AT rs66794874	WT	WT
GCNT2-201-1	**wt: E41, G121	***htz unkn SNP E/K41, A/G121	htz unkn SNP E/K41, A/G121	htz unkn SNP E/K41, A/G121	wt: E41, G121	wt: E41, G121
GJA8-1	WT	WT	WT	WT	WT	WT
GJA8-MID-1	WT	WT	WT	WT	WT	WT
GJA8-2	WT	WT	WT	WT	WT	WT
GJA8-3F	WT	WT	WT	WT	WT	WT
CRYBB1-1	WT	WT	WT	WT	WT	WT
CRYBB1-2	WT	WT	WT	WT	WT	WT
CRYBB1-3	WT	WT	WT	WT	WT	WT
CRYBB1-4	WT	WT	WT	WT	WT	WT
CRYBB1-5	WT	***WT	WT	WT	WT	WT
CRYBB1-6	WT	WT	WT	WT	WT	WT
CRYBB1-201-1	WT	WT	WT	WT	WT	WT
HSF4-3	WT	WT	WT	WT	WT	WT
HSF4-4+5	WT	WT	WT	WT	WT	WT
HSF4-6	WT	WT	WT	WT	WT	WT
HSF4-7+8	WT	WT	WT	WT	WT	WT

	Kindred 1 Family 1		Kindred 1 Family 2		Kindred 2 Family 1	
	III:2 (P)	**IV:8 (A)**	**III:3 (P)**	**IV:19 (A)**	**III:3 (P)**	**IV:11 (A)**
HSF4-9	WT	WT	WT	WT	WT	WT
HSF4-10+11	WT	WT	WT	WT	WT	WT
HSF4-12+13	WT	WT	WT	WT	WT	WT
HSF4-14	WT	WT	WT	WT	WT	WT
HSF4-15	WT	WT	WT	WT	WT	WT
HSF4-T2-10	WT	WT	WT	WT	WT	WT
HSF4-T2-11	WT	WT	WT	WT	WT	WT
CRYAA-1	hmz T rs872331	***htz TC rs872331	hmz T rs872331	hmz T rs872331	*htz TC rs872331	hmz C rs872331
CRYAA-2	WT	***WT	WT	WT	*WT	WT
CRYAA-3	WT	***WT	WT	WT	*WT	WT
CRYAA-003-1	WT	***WT	WT	WT	*WT	WT
CRYAA-004-1	WT	***WT	WT	WT	*WT	WT
CRYBB2-2	WT	WT	WT	WT	WT	WT
CRYBB2-3	WT	WT	WT	WT	WT	WT
CRYBB2-4	WT (hmz T int4+67)	WT (htz AT)	WT (htz AT)	WT (htz AT)	WT (htz AT)	WT (htz AT)
CRYBB2-5	WT	hmz C rs2330991; hmz A rs233092; hmz C rs4049504	hmz C rs2330991; hmz A rs233092; hmz C rs4049504; htz AG rs4049505 int5	htz AG re4049505; intron 5 +34 C>T not on Ens/NCBI	WT	hmz C rs2330991; hmz A rs233092; hmz C rs4049504
CRYBB2-6	WT	WT	WT	WT	WT	wt-hetz GtoA ENSSNP11913674/ rs8140949
CRYBA4-1	WT	WT	WT	WT	WT	WT
CRYBA4-2	WT	WT	WT	WT	WT	WT
CRYBA4-3	WT	WT	WT	WT	WT	WT
CRYBA4-4	WT	WT	WT	htz TC intron 4	WT	WT
CRYBA4-5	WT	WT	WT	WT	WT	WT
CRYBA4-6	WT	WT	WT	WT	WT	WT
LENEP-1 F	WT	WT	WT	WT	WT	WT
CRYBG3-1A	WT	WT	WT	WT	WT	WT
CRYBG3-1B	WT	WT	WT	WT	WT	WT
CRYBG3-1C	WT	WT	WT	WT	WT	WT
CRYBG3-2+3	WT	WT	WT	WT	WT	WT
CRYBG3-4F	WT	WT	WT	WT	WT	WT

	Kindred 1 Family 1		Kindred 1 Family 2		Kindred 2 Family 1	
	III:2 (P)	**IV:8 (A)**	**III:3 (P)**	**IV:19 (A)**	**III:3 (P)**	**IV:11 (A)**
CRYBG3-5	WT	WT	WT	WT	WT	WT
CRYBG3-6	WT	WT	WT	WT	WT	WT
CRYBG3-7	WT	WT	WT	WT	WT	WT
CRYBG3-8	WT	WT	WT	WT	WT	WT
CRYBG3-9	WT	WT	WT	WT	WT	WT
CRYBG3-10+11	WT	WT	WT	WT	WT	WT
CRYBG3-12	WT	WT	WT	WT	WT	WT
CRYBG3-13	WT	WT	WT	WT	WT	WT
CRYBG3-14	WT	WT	WT	WT	WT	WT
CRYBG3-15	WT	WT	WT	WT	WT	WT
CRYBG3-16	WT	WT	WT	WT	WT	WT
CRYBG3-17	hmz H926 rs4857302 (C,C), hmz G rs832089 int17	hmz H926 rs4857302 (C,C), hmz G rs832089 int17	htz H926, N926 (C,A) rs4857302, htz GA rs832089 int17	hmz N926 rs4857302 (A,A), hmz G rs832089 int17	hmz N926 rs4857302 (A,A), hmz A rs832089 int17	htz H926, N926 (C,A), htz rs832089 int17
CRYBG3-18	WT	WT	WT	WT	WT	WT
CRYBG3-19	WT	WT	WT	WT	WT	WT
CRYBG3-201-1	WT	WT	WT	WT	WT	WT
CRYBG3-201-2	WT	WT	WT	WT	WT	WT
CRYBG3-004-1	WT	WT	WT	WT	WT	WT
HSPB6-1F	WT	***WT	WT	WT	WT	WT
HSPB6-001-2+3	WT	***WT	WT	WT	WT	WT
HSPB6-201-4	WT	***WT	WT	WT	WT	WT

Table 4.6: *Sequencing results for the Omani Kindreds. (Htz=heterozygous; Hmz=homozygous; WT=wild type; X= no result; *=DNA from kindred 2 family 1, III:4; **=DNA from kindred 1 family 1, III:1; ***= DNA from kindred 1 family 1,IV:7).*

4.3.3.5 Whole Exome Sequencing

Since mapping was not able to identify a shared region of homozygosity in the Omani kindreds it was considered that the relevant cataract gene mutation might be in a small region of homozygosity (or, despite the consanguinity, one or more of the affected cases could be a compound heterogote). To test these hypotheses (a) sequencing was undertaken of known cataract genes and (b) DNA from kindred 1

family 1 affected individual IV:11 was sent to Dr Michael Simpson at Guy's Hospital for whole exome sequencing.

43 novel homozygous changes were identified (Table 4.8). These consisted of 34 missense changes, 6 synonymous changes, 1 nonsense change, 1 insertion, and 1 UTR change.

Investigative focus was directed on genes found in regions from SNP chip analysis. Primers were designed for the changes identified in *CATSPERB*, *ARL13B*, *NUP210L, ARMC5*, *CRYGA*, *SPAG17* and *ITFG*, to examine segregation within the kindreds. Segregation of the novel *ARL13B* change (P866R), was detected, but only within kindred 1 family 1 (Table 4.7).

Gene	Chromosome, Location, Exon	Nucleotide	Protein	K1F1 III:1 (Un)	K1F1 III:2 (Un)	K1F1 IV:4 (Un)	K1F1 IV:7 (Aff)	K1F1 IV:12 (Un)	K1F1 IV:8 (Aff)	K1F1 IV:11 (Aff)	K1F1 IV:9 (Un)	K1F2 III:3 (Un)	K1F2 IV:14 (Un)	K1F2 IV:19 (Aff)	K1F2 IV:18 (Aff)	K1F2 IV:20 (Aff)	K2 III:3 (Un)	K2 III:4 (Un)	K2 IV:11 (Aff)
CATSPERB	14,91146577-91146578 (ex21)	C2597G	P866R	CG	CG	CG	CC (wt)	CG	GG	GG	CG	CG	CG	CC (wt)	CC (wt)	CG	CC (wt)	CC (wt)	CC (wt)
ARL13B (005)	3,95252408-95252409 (ex9)	G1193T	G398V	GT	GT	GT	TT	GT	TT	TT	GT	GG (wt)	GG (wt)	GG (wt)	x	GG (wt)	x	GG (wt)	GG (wt)
NUP210L (001)	1,152251399-152251400 (ex34)	A4274G	N1575S	x	x	AG	GG	AG	AA (wt)	GG	GG	AG	AG	x	AA (wt)	GG	AA (wt)	AA (wt)	AA (wt)
ARMC5 (201)	16,31383364-31383365 (ex4)	C1520T	P507L	CT	TT	x	x	x	TT	TT	CC	CT	CT	CT	x	x	CC	CC	x
CRYGA	2,208736185-208736186 (ex2)	G239A	R80H	GG (wt)	AG	AG	AG	AG	GG (wt)	AG	AG	GG (wt)	GG (wt)	GG (wt)	GG (wt)	GG (wt)	GG (wt)	GG (wt)	GG (wt)
SPAG17	1,118386088-118386089 (ex21)	A2914G	K972E	AG	AG	AG	GG	AG	AG	GG	GG	x	AG	GG	AG	GG	AA (wt)	AA (wt)	AA (wt)
ITFG	16,46020265-46020266 (ex6)	G605A	R202Q	AG	AG	AG	AG	x	AA	AA	GG	AG	AG	x	x	AG	GG	GG	GG

Table 4.6: *Segregation of novel changes in the Omani kindreds. Un=unaffected. Aff=affected. X=no result. wt=wild type. K=kindred. F=family. Highlighted cells indicate segregation of the novel ARL13B change, but only within kindred 1 famiily 1.*

Ch.r	Start	Stop	Sequence	Gene	Accession no.	DNA/protein position	Effect
1	47676596	47676597	GCAGACTTAG C CGAGGACGAG	FOXD2	NM_004474	[c.203, p.68]	missense [GCC:A GtC:V c]
1	53514966	53514967	TCGGCCTCGT C CGATTTGTCT	LRP8	NM_001018054	[c.868, p.290]	missense [GAC:D tAC:Y U]
1	118386088	118386089	GCGTTATCCT T GCCTTTCTTT	SPAG17	NM_206996	[c.2914, p.972]	missense [AAG:K gAG:E c]
1	152251399	152251400	TACAGTTGAA T TGAGGGTATT	NUP210L	NM_001159484	[c.4724, p.1575]	missense [AAT:N AgT:S c]
1	153441519	153441520	TCTCCAGCCC A TCGACCTCCG	THBS3	NM_007112	[c.498, p.166]	synonymous[GAT:D GAc:D c]
1	154832851	154832852	CAGCTCTCCA A GTCTTTGTTT	GPATCH4	NM_015590	[c.391, p.131]	synonymous[TTG:L cTG:L c]
1	156716356	156716357	TGCAGATCTT G GCAGAAAACC	OR10R2	NM_001004472	[c.66, p.22]	synonymous[TTG:L TTa:L c]
1	158123040	158123041	GTTCTTTTTG G ATCTGCTGCT	CCDC19	NM_012337	[c.651, p.217]	synonymous[ATC:I ATt:I c]
1	158917643	158917644	TGAGGCAGAC C GTGCCATTCT	CD48	NM_001778	[623-624 207-208]	synonymous[ACG:T ACc:T c]
1	159430062	159430063	CCTGAGCCAG T TGGGCAGTCC	ADAMTS4	NM_005099	[1725-1726 575-576]	missense [ACT:T gCT:A c]
1	163905171	163905172	CAGGTGTGGT C GGTTGATGAG	ALDH9A1	NM_000696	[1069-1070 356-357]	missense [CGA:R CaA:Q c]
3	47433894	47433895	GGCGAGGGAA G GGGAGCCTTT	SCAP	NM_012235	[2872-2873 957-958]	missense [CCT:P CaT:H c]
3	95252408	95252409	GAGCCTCTTG G TGAAACACAT	ARL13B	NM_001174150	[1192-1193 397-398]	missense [GGT:G GtT:V U]
3	95252408	95252409	GAGCCTCTTG G TGAAACACAT	ARL13B	NR_033427		UTR
3	112780633	112780634	TTACTTAAAG A TAGAGTCAAG	CD96	NM_005816	[613-614 204-205]	missense [GAT:D GgT:G U]
4	8671993	8671994	ATCCCCGCCC G GATGAAGAGG	CPZ	NM_001014447	[1708-1709 569-570]	missense [CGG:R CaG:Q c]
4	20228275	20228276	CCATGCCAGG C GATCAAGTGC	SLIT2	NM_004787	[4252-4253 1417-1418]	missense [GCG:A GtG:V c]
5	139999322	139999323	AGGGCCGCCT C AGGCCCACGG	TMCO6	NM_018502	[26-27 8-9]	synonymous[CTC:L CTg:L c]
7	102503061	102503062	CGGGGCGACC G CGAGCTCGGG	ARMC10	NM_001161009	[106-107 35-36]	missense [CGC:R CaC:H U]
8	139675587	139675588	CTTCATGTAC G CCGGGGGCAT	COL22A1	NM_152888	[4468-4469 1489-1490]	missense [GCG:A GtG:V c]
9	129627905	129627906	CCACTCGAGC G TGCGGCCCAT	ENG	NM_000118	[577-578 192-193]	missense [ACG:T AtG:M c]
11	60366278	60366279	TGTGGAGTTC G AGTCGAACCC	CCDC86	NM_024098	[105-106 35-36]	nonsense [GAG:E tAG:* U]
11	65070852	65070853	ACCGCGACAG G CCCCGCCCCC	LTBP3	NM_001130144	[2221-2222 740-741]	missense [GCC:A GgC:G c]
12	88268920	88268921	GGCTTGGAAC T TACTGAAGCC	DUSP6	NM_001946	[412-413 137-138]	missense [AAG:K AgG:R c]
14	52689249	52689250	GCCGCTCTCA C CCTCGCTGTA	DDHD1	NM_001160147	[316-317 105-106]	missense [GGT:G GcT:A c]
14	91146577	91146578	TTTTTTTACC G GCAAAGTTTT	CATSPERB	NM_024764	[2596-2597 865-866]	missense [CCG:P CgG:R U]
16	2957438	2957439	CCGCCGGCCG C GATTGGGGGT	KREMEN2	NM_024507	[1090-1091 363-364]	missense [GCG:A GtG:V c]
16	29725562	29725563	GGCCCCCCCC T TCCCCGTGCT	MAZ	NM_001042539	[39-40 13-14]	missense [TTC:F gTC:V c]
16	31383364	31383365	CAACGCACTC C GGGCCGCAGC	ARMC5	NM_001105247	[1519-1520 506-507]	missense [CCG:P CtG:L U]
16	46020265	46020266	ATGTGGAATT C GCATTTTACT	ITFG1	NM_030790	[604-605 201-202]	missense [CGA:R CaA:Q c]
17	6315207	6315208	CACTCACTGC G GGAGGCACCC	PITPNM3	NM_001165966	[1512-1513 504-505]	missense [CGC:R tGC:C U]
17	6328323	6328324	CAAGGAAACC C GGTACAGTTC	PITPNM3	NM_001165966	[178-179 59-60]	missense [CGG:R CaG:Q c]
17	7180781	7180782	GCTGAGATGA C GGTCAAGCTG	ACAP1	NM_014716	[4-5 1-2]	missense [ACG:T AtG:M c]
17	7693514	7693515	TTCTGCCCAG C CCACACCCCC	KDM6B	NM_001080424	[3183-3184 1061-1062]	missense [CCC:P tCC:S U]
17	31209608	31209609	GGGGCTGAGT G CACGTGCATC	C17orf66	NM_152781	[972-973 324-325]	missense [CAC:H tAC:Y c]
17	54188482	54188482	CCGAACCCGA - ACCCGAGTCC	PPM1E	NM_014906	[125-125 41-42]	aa insertion [755 757]
18	54737255	54737256	TGACACCAGC C TCCCCAGCGT	ZNF532	NM_018181	[756-757 252-253]	missense [CTC:L gTC:V c]
18	75760789	75760790	GCTGTGGGTG C GGGCAGGGCG	KCNG2	NM_012283	[1386-1387 462-463]	missense [CGG:R tGG:W c]
19	1828329	1828330	GGGCAGCGCT C GTAGAGCGCC	FAM108A1	NM_001130111	[801-802 267-268]	missense [GAG:E cAG:Q c]
19	13780760	13780761	GGCGGCTACT A CGGGGCCAGC	ZSWIM4	NM_023072	[823-824 274-275]	missense [TAC:Y TgC:C U]
19	16297693	16297694	CCGCCTGCGT C CCCGCTGGAG	KLF2	NM_016270	[742-743 247-248]	missense [TCC:S TaC:Y U]
19	50779726	50779727	TCACGTTGAG C CGGCGGGAGG	OPA3	NM_001017989	[135-136 45-46]	missense [GCT:A tCT:S c]
19	55863318	55863319	GGCCGCCCCC A CCAGGGCGGC	SHANK1	NM_016148	[3709-3710 1236-1237]	missense [GTG:V GgG:G U]

Table 4.8: *Novel changes identified in whole exome sequencing of Omani Kindred 1 Family 1, IV:11.*

4.3.4 Discussion

Despite Kindred 1 being a very large consanguineous kindred with multiple affected individuals, a putative mutation in a candidate cataract gene could not be identified. Although autozygosity mapping is a powerful approach to gene identification, it is not completely sensitive and under certain circumstances the responsible gene may not be identified. Although we suspected that the Omani kindreds might share a common founder mutation (as this has proven to be the case in other recessively inherited diseases in this population) we looked at each family individually so that we were not led astray by interfamilial locus heterogeneity. Usually all affected members of a consanguineous kindred with an autosomal recessively inherited disorder will harbour the same homozygous mutation. However occasionally, two mutations may be segregating in the same family. Thus it was possible that in these kindreds studied, the cataract phenotype may result from compound heterozygous mutations.

Previously in the Maher laboratory, in a consanguineous Jordanian Arab family where two boys were affected with Karak Syndrome (an autosomal recessive disorder) (Mubaidin *et al*, 2003), compound heterozygous mutations in *PLA2G6* were found in the two boys (Morgan *et al*, 2006). Similarly, Forshew *et al* (2005) found two TYR mutations in a large consanguineous family of Pakistani origin with oculocutaneous albinism.

4.4 Cataract Family 5

Aims: To identify a pathogenic mutation that causes cataracts in a Pakistani family

4.4.1 Patient DNA

Cataract Family 5 are of Pakistani origin (Figure 4.30), displaying autosomal recessive congenital cataracts in two of the three children (II:1 and II:2). They underwent surgery to remove these. Clinical examination did not reveal other eye abnormalities or systemic problems. No cataracts were detected in I:1, I:2 or II:3.

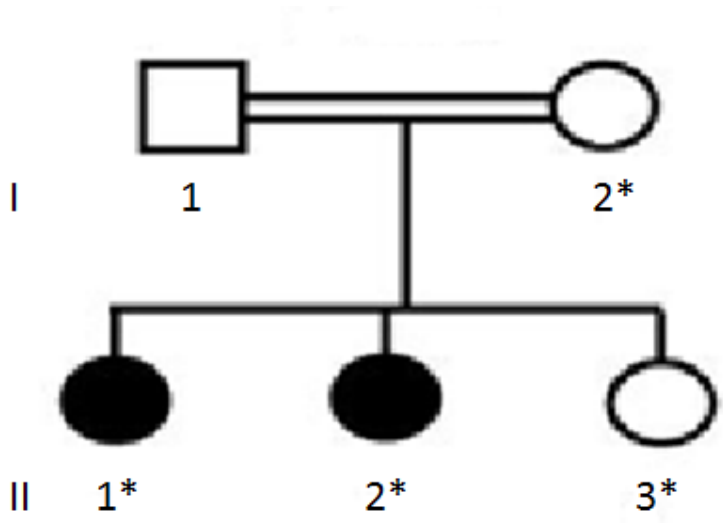


Figure 4.30*: Pedigree of the Pakistani cataract family 5.*

4.4.2 Molecular Genetic Methods

Experimental procedures were carried out as detailed in Chapter 2.

Primers were designed to amplify the coding regions of the known cataract gene *BFSP2,* located on chromosome 3q32.1 (Table 6.35) and sequencing was carried out.

4.4.3 Results

4.4.3.1 Genome-Wide Scan

Previous work in the laboratory with a genome-wide linkage scan (Affymetrix Genechip Human Mapping 10K Array Version 1) on II:1, and on affected individuals from another family of Pakistani origin identified a 38cM homozygous region on chromosome 9 between D9S301 (73,802,954) and D9S910 (101,623,741) (9q13-q22).

Further investigation in the other Pakistani family could not be carried out due to a lack of available DNA. The region overlaps with the 14cM CAAR locus, located on 9q13-q22 between markers D9S1123 (80,425,352) and D9S257 (90,290,735), which was identified in a family from an isolated region of Switzerland with autosomal recessive early-onset progressive pulverulent cataracts (ARPCs) (Héon *et al*, 2001), and previous work identified no putative mutations in the sequencing of exons and intron/exon boundaries of the following genes in the region: *ALDH1A1, RASEF, GCNT1* and *UBQLN1*.

Since these two Pakistani families had an overlapping region on chromosome 9, DNA from affected member II:1 from cataract family 5, and from one affected individual in the other Pakistani family was analysed on a 250k SNP array to look for an identical homozygous region on this chromosome. No identical homozygous region was identified. In Cataract Family 5, in addition to the chromosome 9 region, ten regions of extended homozygosity, >2Mb, were identified.

Chromosome	Start	End	Size (Mb)
1	rs1924569 (58,801,506)	rs1414065 (71,073,756)	12.3
2	rs11127467 (12,994)	rs1710503 (2,520,736)	2.5
2	rs7563021 (6,448,402)	rs3747516 (18,113,508)	11.6
3	rs1969262 (122,863,207)	rs1116439 (147,599,246)	24.7
5	rs6862891 (31,748,611)	rs17552447 (73,001,379)	41.3
8	rs7814384 (116,262,669)	rs199208 (140,712,922)	24.4
9	rs10781059 (71,219,807)	rs10817865 (118,978,682)	47.8
10	rs10047370 (2,628,101)	rs12248552 (8,916,831)	6.3
14	rs2067644 (24,142,675)	rs10144249 (30,923,371)	6.8
14	rs1286293 (91,310,904)	rs3825564 (102,921,447)	11.6

Table 4.9 :*Homozygous regions detected in the genome wide scan on II:1 from cataract family 5.*

4.4.3.2 Candidate Genes

The region on chromosome 3 contains the known autosomal dominant cataract gene *BFSP2* (Conley *et al*, 2000), so the coding exons and intron/exon boundaries were sequenced in parent I:2 and patient II:2. Sequencing of *BFSP2* in I:2 and II:2 revealed homozygosity at rs2276737 for patient II:2, and heterozygosity for parent I:2. At rs2737717 and rs10563564, both I:2 and II:2 were found to be heterozygous, but no pathogenic changes were identified.

4.4.4 Discussion

No putative mutation was found. It is likely that the causative gene lies in the region on chromosome 9, and the same gene or mutation may be responsible for the cataract phenotype in the Swiss CAAR family. The region on chromosome 3 is

unlikely to be the correct region, in light of the finding that I:2 and II:2 were heterozygous for rs rs2737717 and rs10563564. The SNP chip analysis was only carried out on II:1, and sequencing of *BFSP2* was completed for II:2, so this individual is likely to be heterozygous in this region, which would have been identified if SNP chip analysis had been carried out on II:2 DNA..

4.5 Cataract Family 6

Aim: To identify a pathogenic mutation that causes cataracts in a Pakistani family

4.5.1 Patient DNA

Cataract Family 6 are of Pakistani origin (Figure 4.31), displaying isolated autosomal recessive congenital cataracts in two of the five children (II:3 and II:4). No further clinical information was available.

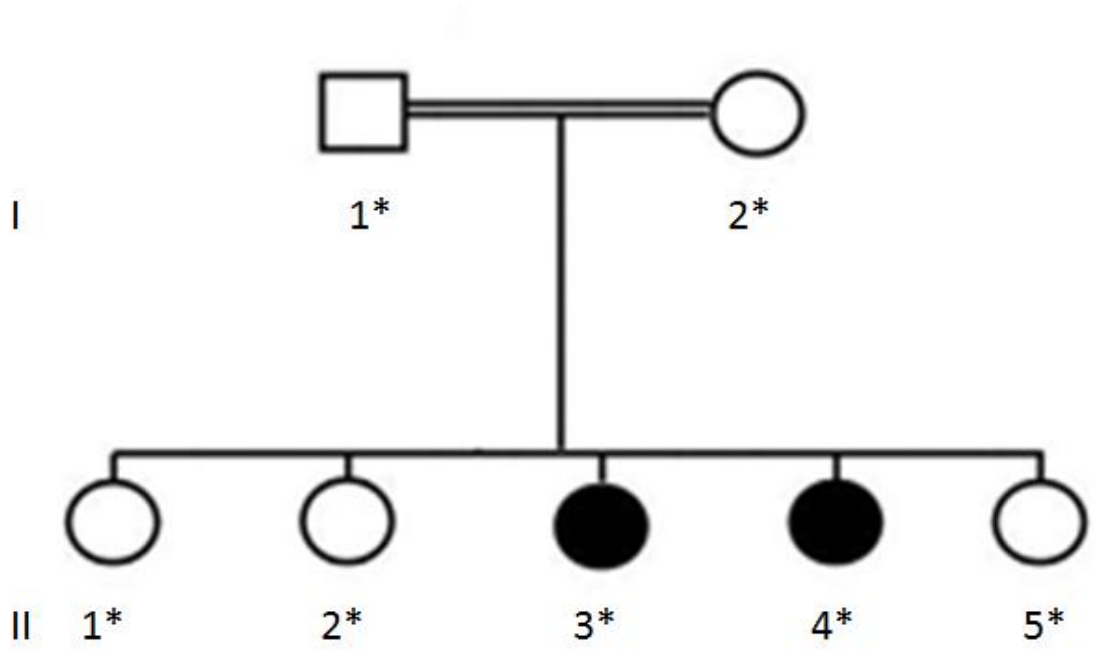


Figure 4.31: *Cataract family 6. Individuals with cataracts are shaded in black. * indicates family members from whom DNA was available.*

4.5.2 Molecular Genetic Methods

An Affymetrix GeneChip Human Mapping 10K Array (version 1) was performed by MRC Geneservice.

4.5.3 Results

4.5.3.1 Genome-Wide Scan

The genome-wide scan was performed for individuals II:3 and II:4. Three potential region of extended homozygosity was detected on chromosome 2, 8 and 15. The region on chromosome 2 from rs1079417 to rs1344063 was uninformative with microsatellite markers. The region on chromosome 8 from rs902619 to rs344278 was uninformative with microsatellite markers, as was the region on chromosome 8 from rs1172 to rs911. A region on chromosome 15 was detected between rs2042613 and rs1875084.

4.5.3.2 Microsatellite Markers

Regions of interest were investigated further by Dr Esther Meyer to confirm presence or absence of homozygosity at various polymorphic microsatellite markers. Linkage to the region on chromosome 15 was confirmed and all markers for the region on chromosome 2 and 8 were uninformative. Microsatellite markers more closely defined the region on chromosome 15, indicating that it lies between D15S988 and D15S984, extending ~10Mb.

4.5.3.3 Whole Exome Sequencing

DNA from II:3 was set to Dr Michael Simpson at Guy's Hospital for whole exome sequencing. 51 novel homozygous changes were identified, comprising of 49 missense changes, 1 nonsense change and 1 insertion/deletion (Table 4.11).

Due to prior SNP chip results, changes in the regions identified on chromosome 15, 2, and 8 were selected for further analysis. Changes in *C15orf39* (Arg1029Lys) and *SNUPN* (Gly137Asp) segregated with cataract phenotype (Table 4.10). Polyphen predicted the Arg1029Lys change in C15orf39 to be tolerated, but predicted the Gly137Asp change in *SNUPN* to be damaging. Gly137 in *SNUPN* is highly conserved across species (Figure 4.32), even in plants such as rice (*O. sativa*) and thale cress (*A. thaliana*) (data not shown).

Gene	Chromosome, Location	Nucleotide	Protein	I:1 (Un)	I:2 (Un)	II:1 (Un)	II:2 (Un)	II:3 (Aff)	II:4 (Aff)	II:5 (Un)
C15orf53	15,36776152-36776153,	CGG-CaG	R18Q	GA	GA	GA	AA	AA	AA	AA
LCMT2	15,41409732-41409733	GCG-tCG	A83S	GT	GT	GT	TT	TT	TT	TT
MAP1A	15,41604191-41604192	GTC-aTC	V1077I	GA (+ GA rs1060939)	GA (+ GG rs1060939)	GA (+ GG rs1060939)	AA (+ GG rs1060939)	AA (+ GG rs1060939)	AA (+ GG rs1060939)	AA (+ GG rs1060939)
DUOX1	15,43231942-43231943	GTG-aTG	V1121M	htz c.G3361A (V1121M); htz intron 25 NM_017434.3:c.3194-34G>A; htz GA rs1706804	htz c.G3361A (V1121M); hmz intron 25 wtG; htz GA rs1706804	htz c.G3361A (V1121M); hmz intron 25 wtG; htz GA rs1706804	hmz c.A3361 (V1121M); hmz intron 25 wtG; hmz A rs1706804	hmz c.A3361 (V1121M); hmz intron 25 wtG; hmz A rs1706804	hmz c.A3361 (V1121M); hmz intron 25 wtG; hmz A rs1706804	hmz c.A3361 (V1121M); hmz intron 25 wtG; hmz A rs1706804
SLC30A4	15,43564760-43564761	AAC-AAa	N394K	CA	CA	CA	AA	AA	AA	AA
SECISBP2L	15,47117119-47117120	ATT-gTT	I55V	AG	AG	AG	GG	GG	GG	GG
CA12	15,61418154-61418155	GAC-aAC	D264N	GA	GA	GA	AA	AA	AA	AA
C15orf39	15,73290451-73290452	AGA-AaA	R1029K	GA	GA	GA	GA	AA	AA	GA
SNUPN	15,73689045-73689046	GGT-GaT	G137D	GA	GA	GA	GA	AA	AA	GA
KLHL25	15,84113042-84113043	GCC-aCC	A335T	GA	GA	AA	GA	AA	AA	GA
RSAD2	2,6935604-6935605	GTC-aTC	V75I	GA	AA	AA	GA	GG	GG	GA

Table 4.10 : *Segregation of novel changes in Cataract Family 6. Un=unaffected. Aff=affected. X=no result. wt=wild type. K=kindred. F=family. Htz=heterozygous. Hmz=homozygous*

Highlighted cells indicate segregation of genotype with cataract phenotype. Lighter text indicates anomalous result.

SNUPN,*H.sapiens*	126	VGKRALIVASR**G**STSAYTKSGYCVNRFSSLLPGGNRRNS-TA--KDYTIL	172
SNUPN,*C.lupus*	127	VGKRALIVASR**G**STSAYTKSGYCVNRFSSLLPGGNKRNSTTA--KDYTIL	174
SNUPN,*B.taurus*	127	VGKRSLIVASQ**G**LTSAYTKSGYCVNTFPSLLPGGNRRNSTTE--KDYTIL	174
Snupn,*M.musculus*	127	VGKRALIVASR**G**STSAYTKSGYCVNRFSSLLPGGNRRNSTTA--KDYTIL	174
Snupn,*R.norvegicus*	127	VGKRALIVASR**G**STSAYTKSGYCVNRFSSLLPGGNRRNSTTA--KDYTIL	174
SNUPN,*G.gallus*	128	VGKRALVVASR**G**STAAYTKSGFCVNRFPSLLPGGNRHNTMNE--KVICLA	175
Snupn,*D.rerio*	137	VGKRSLVVASK**G**STTSYTKSGYCVNRFPSLIPGGNRHNSALG--KDYTIL	184
F23F1.5,*C.elegans**	125	VGKRTLVVASR**G**FTVAYNKGGREVSRFQSRLPGGNTR--AKN--QAWTIL	170

Figure 4.32: *Sequence alignment of Gly137Asp in SNUPN across species highlighted in red. (For C. elegans, F23F1.5 is a hypothetical protein).*

4.5.4 Discussion

A potential causative mutation was detected in SNUPN in this Pakistani family, which may be responsible for the cataract phenotype. Polyphen predicts this G137D change to be damaging. Further work to investigate this putative mutation would include sequencing ethnically matched controls, and controls from other populations to discern whether or not this is actually just a rare SNP that happens to segregate within the family. Functional work could be carried out, expressing both the mutant and wild type sequences in the lens derived cell line CRL-11421.

Not much is known about the implications of the specific role of *SNUPN*. There is evidence of expression in the eye, as well as in the brain, liver, uterus, kidney, colon and many more tissues (Aceview). There are two known transcripts. ENST00000308588 is a nine exon gene of 1513bp and 360 amino acids, and ENST00000371091 is a ten exon gene of 1913bp and 402 amino acids (Ensembl). The protein coded for by this gene, SPN1 is a nuclear import receptor for cytoplasmically assembled, m(3)G-capped spliceosomal U small nuclear riboproteins (snRNPs) (Monecke *et al*, 2009).

Chromosome	Start	Stop	Codon	Gene	Transcript ID	DNA/protein	Effect
1	170895266	170895265	ATG-AcG	FASLG	ENST00000340030	M101T	missense
1	184274619	184274620	GGT-GtT	HMCN1	ENST00000367492	G1963V	missense
2	6935604	6935605	GTC-aTC	RSAD2	ENST00000382040	V75I	missense
2	21085625	21085626	GGC-GtC	APOB	ENST00000233242	G2540V	missense
2	71681465	71681466	ATC-cTC	DYSF	ENST00000258104	I1277L	missense
2	74630168	74630169	GGG-GaG	LOXL3	ENST00000264094	G176E	missense
2	95305518	95305519	GGG-aGG	PROM2	ENST00000317620	G137R	missense
2	198354793	198354794	TCC-TaC	BOLL	ENST00000321801	S21Y	missense
2	201980409	201980410	CTC-gTC	TRAK2	ENST00000332624	L83V	missense
3	180443718	3180443719	AAC-AAg	KCNMB3	ENST00000314235	N169K	missense
3	180534503	180534504	AAT-AgT	ZNF639	ENST00000326361	N353S	missense
3	182170731	182170732	AGT-gGT	FXR1	ENST00000305586	S414G	missense
4	6354985	6354986	TCA-cCA	WFS1	ENST00000226760	S855P	missense
4	75123159	75123160	AGC-AGa	CXCL3	ENST00000296026	S11R	missense
4	111763062	111763063	ATG-ATa	PITX2	ENST00000306732	M1I	missense
5	154190576	154190577	ATT-gTT	C5orf4	ENST00000326080	I115V	missense
7	6787494	6787495	CGT-tGT	RSPH10B2	ENST00000297186	R541C	missense
7	72649567	72649568	CCC-tCC	MLXIPL	ENST00000313375	P495S	missense
8	42350955	42350956	ATT-AcT	DKK4	ENST00000220812	I165T	missense
9	15584042	15584043	GCG-GtG	C9orf93	ENST00000297641	A183V	missense
9	15713729	15713730	ACT-gCT	C9orf93	ENST00000297641	T493A	missense
9	18940906	18940907	AAG-tAG	FAM154A	ENST00000380534	K23*	nonsense
9	97678153	97678154	GCG-cCG	C9orf102	ENST00000288985	A16P	missense
9	101032519	101032520	CGT-CaT	SEC61B	ENST00000223641	R95H	missense
9	115162686	115162687	CGG-CaG	BSPRY	ENST00000238359	R127Q	missense
9	129551319	129551320	TGC-TaC	SH2D3C	ENST00000314830	C377Y	missense
9	134770896	134770897	AAA-AcA	TSC1	ENST00000298552	K630T	missense
9	108731578	108731581	indel	ZNF462	ENSG00000148143	AA_DELETION	indel
10	3162055	3162056	GTG-aTG	PFKP	ENST00000381075	V569M	missense
12	12831388	12831389	GCC-cCC	APOLD1	ENST00000326765	A126P	missense
12	46479003	46479004	GCA-GtA	HDAC7	ENST00000310824	A30V	missense
12	122863982	122863983	CGC-CaC	DNAH10	ENST00000339192	R855H	missense
14	64280076	64280077	CCG-CtG	PLEKHG3	ENST00000247226	P1132L	missense
15	36776152	36776153	CGG-CaG	C15orf53	ENST00000318792	R18Q	missense
15	41409732	41409733	GCG-tCG	LCMT2	ENST00000305641	A83S	missense
15	41604191	41604192	GTC-aTC	MAP1A	ENST00000300231	V1077I	missense
15	43231942	43231943	GTG-aTG	DUOX1	ENST00000321429	V1121M	missense
15	43564760	43564761	AAC-AAa	SLC30A4	ENST00000261867	N394K	missense
15	47117119	47117120	ATT-gTT	SECISBP2L	ENST00000261847	I55V	missense
15	61418154	61418155	GAC-aAC	CA12	ENST00000178638	D264N	missense
15	73290451	73290452	AGA-AaA	C15orf39	ENST00000360639	R1029K	missense
15	73689045	73689046	GGT-GaT	SNUPN	ENST00000308588	G137D	missense
15	84113042	84113043	GCC-aCC	KLHL25	ENST00000337975	A335T	missense
16	2061858	2061859	CCT-tCT	TSC2	ENST00000219476	P674S	missense
16	20459495	20459496	CCA-CtA	ACSM2	ENST00000329697	P537L	missense
16	23613897	23613898	GAG-aAG	ERN2	ENST00000256797	E663K	missense
16	29949266	29949267	CAG-CcG	FAM57B	ENST00000380495	Q28P	missense
18	19743219	19743220	CGA-CaA	LAMA3	ENST00000269217	R765Q	missense
19	7586122	7586123	CGG-gGG	KIAA1543	ENST00000160298	R909G	missense
20	60327005	60327006	GGC-aGC	LAMA5	ENST00000370691	G2380S	missense
22	22892769	22892770	AAA-AgA	CABIN1	ENST00000263119	K1724R	missense

Table 4.11: *Novel changes identified in whole exome sequencing of Pakistani Cataract Family 6, II:3.*

4.6 Cataract Family 7

Aim: To identify a pathogenic mutation that causes cataracts in a Pakistani family

4.6.1 Patient DNA

Cataract family 7 are of Pakistani origin. I:1 and I:2 are first cousins. II:2 developed sudden onset cataracts at three years of age (Figure 4.33).

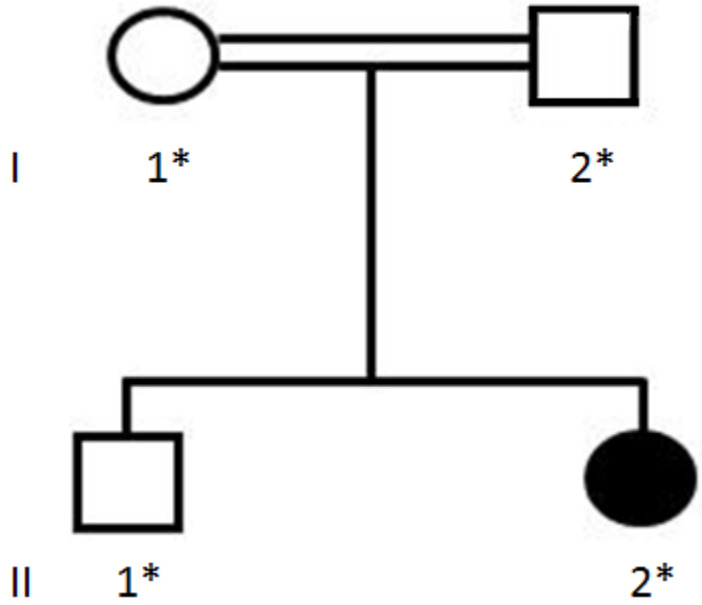


Figure 4.33: *Cataract family 7. Individuals with cataracts are shaded in black. * indicates family members from whom DNA was available.*

4.6.2 Molecular Genetic Methods

II:2 was run on 250K SNP chip, and regions of homozygosity, however small, around known cataract genes were identified.Primers for many known cataract genes were already present in the laboratory, and time available for investigation into the cause of cataractogenesis in this family was limited.

4.6.3 Results

4.6.3.1 Genome-Wide Scan

Small homozygous regions around known genes were found and noted. A homozygous call was made for 145,846,565, and *GJA8* is found across this position from 145,841,560-145,848.017 on chromosome 1. *BFSP2* (134,601,480-134,676,746 on chromosome 3) is contained within a homozygous region with homozygous calls at least from 134,598,678-134,677,146. *CRYGS* (187,738,926-187,744,861 on chromosome 3) is also located in a homozygous region extending at least from 187,713,436-187,776,514.

4.6.4 Discussion

This family's DNA was received towards the end of the project, leaving little time for analysis. Since homozygosity around *GJA8*, *BFSP2*, and *CRYGS* was detected by genome wide scan, the next step would be to sequence the coding regions of these genes to look for pathogenic mutations.

4.7 Cataract Family 8

Aim: To identify a pathogenic mutation that causes cataracts in an Omani family

4.7.1 Patient DNA

Cataract family 8 are of Omani origin. Two siblings (II:2 and II:3) presented with cataracts. Their sister(II:1) and parents (I:1 and I:2) were unaffected (Figure 4:35).

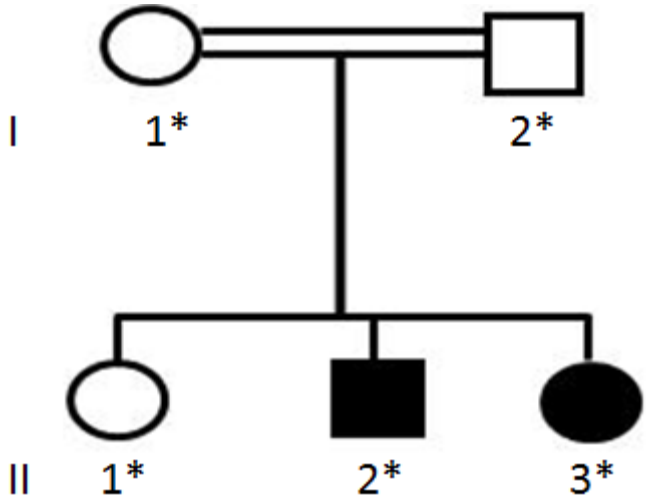


Figure 4.34: *Cataract family 8. Individuals with cataracts are shaded in black. * indicates family members from whom DNA was available.*

4.7.2 Molecular Genetic Methods

A 250K genome wide scan was performed on DNA from II:2 and II:3, and common homozygous regions were noted. Known cataract genes were sequenced to identify any mutations.

4.7.3 Results

4.7.3.1 Genome-Wide Scan

Seven regions of shared homozygosity were identified in II:2 and II:3 (Table 4:12).

Chromosome	Start	Stop
1	116,258,116	116,966,263
4	116,770,717	119,842,915
6	56,713,291	63,944,147
7	118,424,891	120,877,689
7	130,881,572	134,219,717
16	31,488,031	46,754,604
17	71,886,214	73,788,078

Table 4:12: *Regions of homozygosity shared by affected members of Cataract Family 8.*

4.7.3.2 Sequencing of Known Cataract Genes

Sequencing of known cataract genes *CRYBB3*, *BFSP1*, *LIM2*, *GCNT2*, *CRYBB2*, *CRYBA4*, *GJA8*, *CRYBB1*, *HSF4* and *CRYAA* was undertaken. Although no putative pathogenic mutations were detected, numerous SNPs were detected. Results are are summarised in Table 4.13.

	Cataract Family 8		**Cataract Family 9**	
	I:1 (P)	**II:3 (A)**	**III:1 (P)**	**IV:7 (A)**
CRYBB3-1	WT	WT	WT	WT
CRYBB3-2	WT	WT	WT	WT
CRYBB3-3	WT	WT	WT	WT
CRYBB3-4	WT	WT	WT	WT
CRYBB3-5	WT (htz for rs960378 c.337C>G, p.His113Asp	WT (htz for rs960378 c.337C>G, p.His113Asp	WT (hmz G)	WT (hmz G)
CRYBB3-6R	WT	WT	WT	WT
BFSP1-1	WT	WT	WT	WT
BFSP1-2	WT	WT	WT	WT
BFSP1-3	WT	WT	WT	WT
BFSP1-4	htz c.576-42G>A	hmz c.576-42G>A	htz c.576-42G>A	htz c.576-42G>A
BFSP1-5	WT	WT	WT	WT
BFSP1-6	htz c.776-45A>G	hmz c.776-45A>G	htz c.776-45A>G	htz c.776-45A>G
BFSP1-7R	htz AG rs6080719	hmz G rs6080719	htz AG rs6080719	htz AG rs6080719
BFSP1-8A	WT	WT	WT	WT
BFSP1-8BR	htz AG rs6136118	hmz G rs6136118	htz AG rs6136118	hmz A rs6136118

	Cataract Family 8		Cataract Family 9	
	I:1 (P)	**II:3 (A)**	**III:1 (P)**	**IV:7 (A)**
BFSP1-8C	hmz G rs6080718, htz CT rs6080717	hmz G rs6080718, hmz Trs6080717	1) htz AG rs6080718, P583, NG_012423.1:g.6963 8A>G 2) ATT>ATA on F and R. Still I636, not on Ens/NCBI 3)htz CA rs6105762 intronic	hmz G rs6080718
BFSP1-T2-1	WT	WT	WT	WT
LIM2-2	WT	WT	WT	WT
LIM2-3	WT	WT	WT	WT
LIM2-4 R	WT (homz int4 T rs2547318)	messy (homz int4 C rs2547318)	WT (hetz CT int4 rs2547318)	WT (hetz CT int4 rs2547318)
LIM2-5	WT	WT	WT	WT
GCNT2-001-1A	htz TC rs2230906	htz TC rs2230906	hmz C	htz CT rs2230906
GCNT2-001-1B	WT	WT	WT	WT
GCNT2-001-1C	WT	WT	WT	WT
GCNT2-001-2	WT	WT	WT	WT
GCNT2-001-3	WT	WT	WT	WT
GCNT2-004-1A	WT	WT	WT	WT
GCNT2-004-1B	WT	WT	WT	WT
GCNT2-004-1C	hmz G rs539351	hmz G rs539351	hmz G rs539351	hmz G rs539351
GCNT2-006-3A	WT	WT	WT	WT
GCNT2-006-3B	WT	WT	WT	WT
GCNT2-006-3C	WT	WT	WT	WT
GCNT2-010-3	WT	WT	WT	WT
GCNT2-201-1	wt: E41, G121	wt: E41, G121	wt: E41, G121	wt: E41, G121
GALK1-201-1	x	x	x	x
GALK1-201-2	x	x	WT	WT
GALK1-201-3	x	x	WT	x

	Cataract Family 8		Cataract Family 9	
	I:1 (P)	**II:3 (A)**	**III:1 (P)**	**IV:7 (A)**
GALK1-201-4	x	x	WT	WT
GALK1-201-5	x	x	WT	WT
GALK1-201-6+7+8	x	x	WT	WT
GALK1-202-2	x	x	WT	x
GALK1-202-3+4	x	x	WT	WT
GJA8-1	WT	WT	WT	WT
GJA8-MID-1	WT	WT	htz rs3766503: NM_005267.3:C804C>T, NM_034400.4:g.921503C>T	hmz for rs3766503
GJA8-2	WT	WT	WT	WT
GJA8-3	WT	WT	WT	WT
CRYBB1-1	WT	WT	WT	WT
CRYBB1-2	WT	WT	WT	WT
CRYBB1-3	WT	WT	WT	WT
CRYBB1-4	WT	WT	WT	WT
CRYBB1-5	WT	WT	WT	WT
CRYBB1-6	WT	WT	WT	WT
CRYBB1-201-1	WT	WT	WT	WT
HSF4-3	WT	WT	WT	WT
HSF4-4+5	WT	WT	WT	WT
HSF4-6	WT	WT	WT	WT
HSF4-7+8	WT	WT	WT	WT
HSF4-9	WT	WT	WT	WT
HSF4-10+11	WT	WT	WT	WT
HSF4-12+13	WT	WT	WT	WT
HSF4-14	WT	WT	WT	WT
HSF4-15	WT	WT	WT	WT

	Cataract Family 8		Cataract Family 9	
	I:1 (P)	**II:3 (A)**	**III:1 (P)**	**IV:7 (A)**
HSF4-T2-10	WT	WT	WT	WT
HSF4-T2-11	WT	WT	WT	WT
CRYAA-1	hmz T rs872331	hmz T rs872331	htz TC rs872331	X
CRYAA-2	WT	WT	WT	WT
CRYAA-3F	WT	WT	WT	WT
CRYAA-003-1	WT	WT	WT	WT
CRYAA-004-1	WT	WT	WT	WT
CRYBB2-2	WT	WT	WT	WT
CRYBB2-3	WT	WT	WT	WT
CRYBB2-4	WT (htz AT)	WT (htz AT)	WT (htz AT)	WT (htz AT)
CRYBB2-5	wt htz AG rs4049505	wt htz AG rs4049505,htz AG int5 rs55700037	WT	WT
CRYBB2-6	WT	WT	wt-hetz GtoA ENSSNP11913674/ rs8140949	wt-hetz GtoA ENSSNP11913674 / rs8140949
CRYBA4-1	WT	WT	hmz A int1 rs16982454	htz CA int1 rs16982454
CRYBA4-2	WT	WT	WT	WT
CRYBA4-3	WT	htz c.108G>A, still V	WT	WT
CRYBA4-4	WT	WT	WT	WT
CRYBA4-5	WT	WT	WT	WT
CRYBA4-6	WT	WT	WT	WT

Table 4.13: *Sequencing results for the Omani families. (Htz=heterozygous; Hmz=homozygous; WT=wild type; X= no result).*

4.7.4 Discussion

No putative mutations were identified in Cataract Family 8. Seven homozygous regions were identified, in total containing hundreds of genes. It is impractical to investigate all of these, especially in light of no obvious candidate genes in these regions. It would be useful to obtain more DNA from Omani families with cataracts, with the aim to compare regions of homozygosity in affected individuals, identify a suitable candidate region and identify a small number of candidate genes for sequencing, based on evidence of expression in the eye, or links to known cataract genes.

4.8 Cataract Family 9

Aim: To identify a pathogenic mutation that causes cataracts in an Omani family

4.8.1 Patient DNA

Cataract family 9 are of Omani origin (Figure 4:36). III1 has retinal disease and cataract, which were noted at age 40 (although from the pedigree it is not clear how long she had been afflicted with these conditions). The affected children presented with cataract at age 9 (IV:5), age 7 (IV:7), age 8 (IV:11). III:2 has myopia, but no cataracts (no DNA available). IV:3 and IV:9 (identified at age 7 for IV:9) have myopia, but no cataracts were observed.

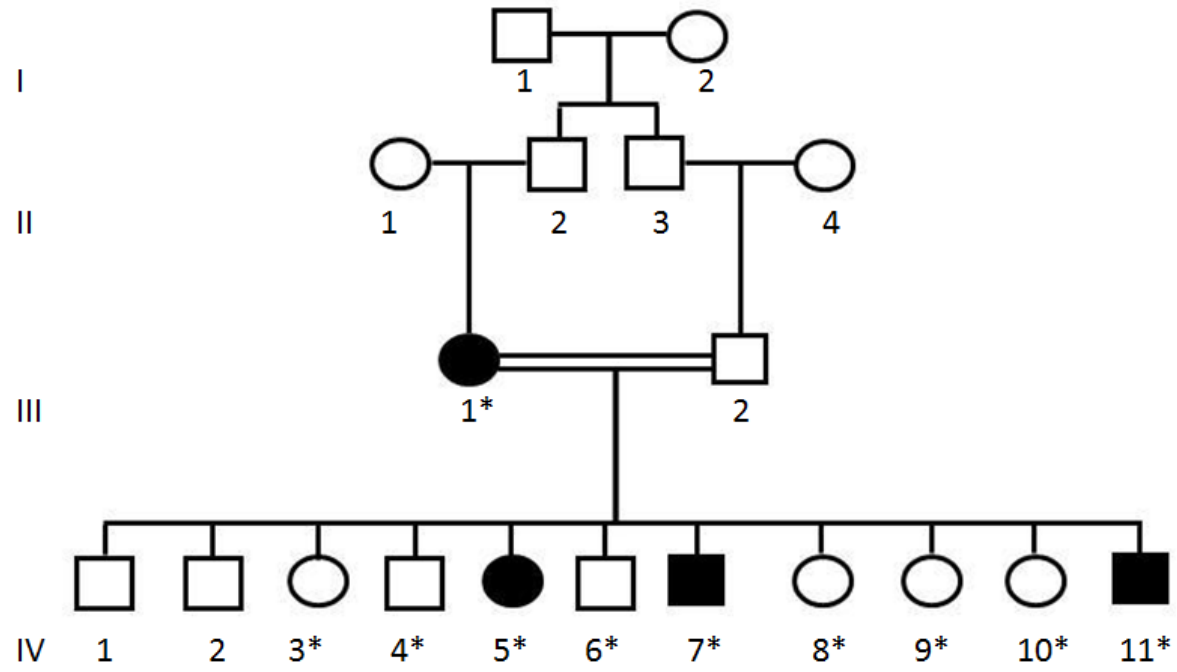


Figure 4.35: *Cataract family 9. Individuals with cataracts are shaded in black. * indicates family members from whom DNA was available.*

4.8.2 Molecular Genetic Methods

A 250K genome wide scan was carried out on DNA from affected individuals IV:7 and IV:11. Sequencing of known cataract genes *CRYBB3*, *BFSP1*, *LIM2*, *GCNT2*, *CRYBB2*, *CRYBA4*, *GJA8*, *CRYBB1*, *HSF4, GALK1* and *CRYAA* was carried out.

4.8.3 Results

4.8.3.1 Genome-Wide Scan

The 250K genome wide scan identified 14 common homozygous regions in IV:7 and IV:11 (Table 4.14).

Chromosome	Start	Stop
1	48,839,231	52,951,964
1	112,091,506	155,434,695
1	230,893,966	238,771,340
6	57,021,031	63,599,624
7	86,208,080	88,697,030
7	119,918,519	120,410,220
8	32,585,951	35,010,158
8	89,055,865	105,755,886
10	22,187,429	24,640,014
10	78,467,676	80,612,626
15	23,825,206	66,294,326
16	64,964,335	66,860,911
22	Small regions ~1-1.5Mb towards the end	

Table 4.14: *Regions of homozygosity shared by affected members IV:7 and IV:11 of Cataract Family 9.*

4.8.3.2 Sequencing of Known Cataract Genes

No putative pathogenic mutations were detected, numerous SNPs were detected. Results are summarised in Table 4.11.

4.8.4 Discussion

To narrow down the homozygous regions in this family, it would be useful to conduct a 250K genome wide scan on DNA from affected individual IV:5. Comparison of regions found in other Omani families would also aid identification of the most suitable candidate regions if homozygosity.

4.9 Conclusion

Gene identification studies with SNP arrays and exome sequencing can be successful. There are many published instances of genes that have been identified

using this approach. For example, *TGM6*, when mutated, was identified as a novel causative gene of spinocerebellar ataxia (Wang *et al*, 2010).

Currently, exome sequencing does not cover all exons in equal depths, and it is not possible to discount the possibility of a mutation within a non-coding region. When genome sequencing becomes available, many more questions about the genotype behind certain phenotypes are sure to be answered.

The 1000 Genome Project is a step in the right direction in elucidating sequence changes that can more accurately be described as mutations, Not all ethnic groups are included in this venture (currently, there are 28 populations being sequenced in phase 2 of the project), so when sequence changes are identified, their possible deleterious or benign nature cannot be discerned simply based on 1000 Genome data.

Functional studies can be useful in confirming that sequence changes are, in fact, damaging mutations, but this is more useful after a single (or few) suspects are identified, because functional studies are arduous, time-consuming and expensive. In a situation similar to that of the Omani Kindreds, so many sequence changes can be identified. Once the dogma that there must be a single causative autosomal recessively segregating mutation is abandoned, and possibility that there may be compound heterozygotes or different genes playing a role in the same family, a very good candidate must be identified before embarking on functional studies.

Chapter 5
References

5. REFERENCES

5.1 Web References

AceView http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/

ASTD http://www.ebi.ac.uk/asd/c

ECgene http://genome.ewha.ac.kr/ECgene/

Ensembl http://www.ensembl.org/index.html

Genatlas http://www.dsi.univ-paris5.fr/genatlas/

Genecards http://www.genecards.org

Mutalyzer http://www.mutalyzer.nl/2.0/check

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

Polyphen http://genetics.bwh.harvard.edu/pph/

Primer3 http://frodo.wi.mit.edu/

UCSC http://genome.ucsc.edu/

5.2 References

Adaimy L, Chouery E, Megarbane H, Mroueh S, Delague V, Nicolas E, Belguith H, de Mazancourt P, Megarbane A. Mutation in WNT10A is associated with an autosomal recessive ectodermal dysplasia: the odonto-onycho-dermal dysplasia. Am. J. Hum. Genet. 2007;81: 821-828.

Adams, M.D., Celniker, S.E., Holt, R.A. *et al.* (2000) The genome sequence of Drosophila melanogaster. **Science.** 287 (5461): 2185-2195

Aebi, M., Horning, H., Padgett, R.A., Reiser, J. and Weissmann, C. (1986) Sequence requirements for splicing of higher eukaryotic nuclear pre mRNA. **Cell.** 47: 411-415

Ai, Y., Zheng, Z., O'Brien-Jenkins, A. *et al.* (2000) A mouse model of galactose-induced cataracts. **Hum Mol Genet.** 9: 1821-1827

Alizadeh, A., Clark, J.I., Seeberger, T. *et al.* (2002) Targeted genomic deletion of the lens-specific intermediate filament protein CP49. **Invest Ophthalmol Vis Sci.** 43 (12): 3722-3727

Alizadeh, A., Clark, J., Seeberger, T. *et al.* (2003) Targeted deletion of the lens fiber cell-specific intermediate filament protein filensin. **Invest Ophthalmol Vis Sci.** 44 (12): 5252-5258

Andley, U.P., Song, Z., Wawrousek, E.F., Fleming, T.P. and Bassnett, S. (2000) Differential protective activity of alpha A- and alphaB-crystallin in lens epithelial cells. **J Biol Chem.** 275 (47): 36823-36831

Andley, U.P. (2007) Crystallinsin the eye: Function and pathology. **Progress in Retinal and Eye Research.** 26: 78-98

Anttonen, A.K., Mahjneh, I., Hämäläinen, R.H. *et al.* (2005) The gene disrupted in Marinsco-Sjögren syndrome encodes SIL1, an HSPA5 cochaperone. **Nat Genet.** 37: 1309-1311

Arneson, M.L. and Louis, C.F. (1998) Structural arrangement of lens fiber cell plasma membrane protein MP20. **Exp Eye Res.** 66 (4): 495-509

Azuma, N., Hirakiyama, A., Inoue, T., Asaka, A. and Yamada, M. (2000) Mutations of a human homologue of the Drosophila eyes absent gene (EYA1) detected in patients with congenital cataracts and ocular anterior segment anomalies. **Hum Mol Genet.** 9 (3): 363-366

Bateman, J.B., Geyer, D.D., Flodman, P. *et al.* (2000) A new betaA1-crystallin splice junction mutation in autosomal dominant cataract. **Invest Ophthalmol Vis Sci.** 41 (11): 3278-3285

Berman, E.R. (1991) Biochemistry Of The Eye. **Plenum Press**

Beby, F., Commeaux, C., Bozon, M. *et al.* (2007) New phenotype associated with an Arg116Cys mutation in the CRYAA gene: nuclear cataract, iris coloboma, and microphthalmia. **Arch Ophthalmol.** 125 (2): 213-216

Berry, V., Ionides, A.C., Moore, A.T. *et al.* (1996) A locus for autosomal dominant anterior polar cataract on chromosome 17p. **Hum Mol Genet.** 5 (3): 415-419

Berry, V., Francis, P., Reddy, M.A. *et al.* (2001) Alpha-B crystallin gene (CRYAB) mutation causes dominant congenital posterior polar cataract in humans. **Am J Hum Genet.** 69: 1141-1145

Beyer, E.C., Kistler, J., Paul, D.L. and Goodenough, D.A. (1989) Antisera directed against connexin43 peptides react with a 43-kD protein localized to gap junctions in myocardium and other tissues. **J Cell Biol.** 108 (2): 595-605

Bhat, S.P. and Nagineni, C.N. (1989) Alpha B subunit of lens-specific protein alpha-crystallin is present in other ocular and non-ocular tissues. **Biochem Biophys Res Commun.** 158: 319-325

Bierhuizen, M.F.A., Mattei, M-G. and Fukuda, M. (1993) Expression of the developmental I antigen by a cloned human cDNA encoding member of a beta-1,6-N-acetylglucosaminyltransferase gene family. **Genes Dev.** 7: 468-478

Billingsley, G., Santhiya, S.T., Paterson, A.D. *et al.* (2006) CRYBA4, a novel human cataract gene, is also involved in microphthalmia. **Am J Hum Genet.** 79: 702-709

Bixler, D., Higgins, M. and Hartsfield, J. Jr (1984) The Nance-Horan syndrome: a rare X-linked ocular-dental trait with expression in heterozygous females. **Clin Genet.** 26 (1) :30-35

Blixt, A., Mahlapuu, M., Aitola, M. *et al.* (2000) A forkhead gene, FoxE3, is essential for lens epithelial proliferation and closure of the lens vesicle. **Genes Dev.** 14: 245-254

Bloemendel, H., de Jong, W., Jaenicke, R. *et al.* (2004) Ageing and vision: structure, stability and function of lens crystallins. **Prog Biophys Mol Biol.** 86: 407-485

Bova, M.P., Yaron, O., Huang, Q. *et al.* (1999) Mutation R120G in alphaB-crystallin, which is linked to a desmin-related myopathy, results in an irregular structure and defective chaperone-like function. **Proc Natl Acad Sci USA.** 96 (11): 6137-6142

Brakenhoff, R.H., Aarts, H.J., Reek, F.H., Lubsen, N.H. and Schoenmakers, J.G. (1990) Human gamma-crystallin genes. A gene family on its way to extinction. **J Mol Biol.** 216: 519-532

Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L. and Weber, J.L. (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. **Am J Hum Genet.** 63 (3): 861-869

Bu, L., Jin, Y., Shi, Y. *et al.* (2002) Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. **Nat Genet.** 31 (3): 276-278

Bundey, S. and Alam, H. (1993) A five-year prospective study of the health of children in different ethnic groups, with particular reference to the effect of inbreeding. **Eur J Hum Genet.** 1 (3): 206-219

Cali, J.J. and Russell, D.W. (1991a) Characterization of human sterol 27-hydroxylase. A mitochondrial cytochrome P-450 that catalyzes multiple oxidation reaction in bile acid biosynthesis. **J Biol Chem.** 266 (12): 7774-7778

Cali, J.J., Hsieh, C-L., Francke, U. and Russell, D.W. (1991b) Mutations in the bile acid biosynthetic enzyme sterol 27-hydroxylase underlie cerebrotendinous xanthomatosis. **J Biol Chem.** 266: 7779-7783

Caspers, G.J., Leunissen, J.A. and de Jong, W.W. (1995) The expanding small heat-shock protein family, and structure predictions of the conserved "alpha-crystallin domain". **J Mol Evol.** 40 (3): 238-248

Chaplin, H., Hunter, V.L., Malecek, A.C., Kilzer, P. and Rosche, M.E. (1986) Clinically significant allo-anti-I in an I-negative patient with massive hemorrhage. **Transfusion.** 26: 57-61

Chen, J., Ma, Z., Jiao, X. *et al.* (2011) Mutations in FYCO1 Cause Autosomal-Recessive Congenital Cataracts. **Am J Hum Genet.** 88 (6): 827-838

Chiang, C., Litingtung, Y., Lee, E. *et al.* (1996) Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. **Nature.** 383 (6599): 407-413

Chien, A., Edgar, D.B. and Trela, J.M. (1976) Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. **J Bact.** 174: 1550-1557

Chow, R.L., Altmann, C.R., Lang, R.A. and Hemmati-Brivanlou, A. (1999) Pax6 induces ectopic eyes in a vertebrate. **Development.** 126: 4213-4222

Chung, K.T., Shen, Y. and Hendershot, L.M. (2002) BAP, a mammalian BiP-associated protein, is a nucleotide exchange factor that regulates the ATPase activity of BiP. **J Biol Chem.** 277: 47557-47563

Cohen, D., Bar-Yosef, U., Levy, J. *et al.* (2007) Homozygous CRYBB1 deletion mutation underlies autosomal recessive congenital cataract. **Invest Ophthal Vis Sci.** 48: 2208-2213

Cook, G.R. and Knobloch, W.H. (1982) Autosomal recessive vitreoretinopathy and encephaloceles. **Am J Ophthalmol.** 94 (1): 18-25

Conley, Y.P., Erturk, D., Keverline, A. *et al.* (2000) A juvenile-onset, progressive cataract locus on chromosome 3q21-q22 is associated with a missense mutation in the beaded filament structural protein-2. **Am J Hum Genet.** 66 (4): 1426-1431.

Crow, Y.J., Hayward, B.E., Parmar, R. *et al.* (2006) Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutières syndrome at the AGS1 locus. **Nat Genet.** 38 (8): 917-920

Czeizel, A.E., Goblyos, P., Kustos, G., Mester, E. and Paraicz, E. (1992) The second report of Knobloch syndrome. **Am J Med Genet.** 42: 777-779

Demetrick, D.J. and Beach, D.H. (1993) Chromosome mapping of human CDC25A and CDC25B phosphatases. **Genomics.** 18: 144-147

Douglas, J., Cilliers, D., Coleman, K. *et al.* (2007) Mutations in RNF135, a gene within the NF1 microdeletion region, causes phenotypic abnormalities including overgrowth. **Nature Genet.** 39: 963-965

Duncan, M.K., Haynes, J.I. 2nd, Cvekl, A. and Piatigorsky, J. (1998) Dual roles for Pax-6: a transcriptional repressor of lens fiber cell-specific beta-crystallin genes. **Mol Cell Biol.** 18 (9): 5579-5586

Fantes, J., Ragge, N.K., Lynch, S.A. *et al.* (2003) Mutations in SOX2 cause anophthalmia. **Nat Genet.** 33 (4): 461-463

Feder, J.N., Gnirke, A., Thomas, W. *et al*. (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. **Nat Genet.** 13 (4): 399-408.

Fleishman, S.J., Unger, V.M., Yeager, M. and Ben-Tal, N. (2004) A Calpha model for the transmembrane alpha helices of gap junction intercellular channels. **Mol Cell.** 15 (6): 879-888

Forshew, T., Khaliq, S., Tee, L. *et al.* (2005) Identification of novel TYR and TYRP1 mutations in oculocutaneous albinism. **Clin Genet.** 68 (2): 182-184

Francis, P.J., Berry, V., Moore, A.T. and Bhattacharya, S. (1999) Lens biology: development and human cataractogenesis. **Trends Genet.** 15 (5): 191-196

Francois, J. (1982) Genetics of cataract. **Ophthalmologica.** 184: 61-71

Gao, B., Guo, J., She, C. *et al.* (2001) Mutations in IHH, encoding Indian hedgehog, cause brachydactyly type A-1. **Nature Genet.** 28: 386-388

Gill, D., Klose, R., Munier, F.L. *et al.* (2000) Genetic heterogeneity of the Coppock-like cataract: a mutation in CRYBB2 on chromosome 22q11.2. **Invest Ophthalmol Vis Sci.** 41 (1): 159-165

Gitschier, J., Wood, W.I., Goralka, T.M. *et al.* (1984) Characterization of the human factor VIII gene. **Nature.** 312: 326-330

Glaser, T., Jepeal, L., Edwards, J.G. *et al.* (1994) PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. **Nature Genet.** 7: 463-471

Gong, X., Cheng, C. and Xia, C.H. (2007) Connexins in lens development and cataractogenesis. **J Membr Biol.** 218 (1-3): 9-12

Goulielmos, G., Gounari, F., Remington, S. *et al.* (1996) Filensin and phakinin form a novel type of beaded intermediate filaments and coassemble de novo in cultured cells. **J Cell Biol.** 132 (4): 643-655

Graw, J. (2003) The genetic and molecular basis of congenital eye defects. **Nat Rev Genet.** 4 (11): 876-888

Graw, J. (2009) Genetics of crystallins: cataract and beyond. **Exp Eye Res.** 88: 173-189

Gu, F., Li, R., Ma, X.X. *et al.* (2006) A missense mutation in the gammaD-crystallin gene CRYGD associated with autosomal dominant congenital cataract in a Chinese family. **Mol Vis.** 12: 26-31

Halder, G., Caliaerts, P. and Gehring, W.J. (1995) Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. **Science.** 267: 1788-1762

Hanson, I. and Van Heyningen, V. (1995) Pax6: more than meets the eye. **Trends Genet.** 11 (7): 268-272

Hansen, L., Yao, W., Eiberg, H. *et al.* (2007) Genetic heterogeneity in microcornea-cataract: five novel mutations in CRYAA, CRYGD, and GJA8. **Invest Ophthalmol Vis Sci.** 48 (9): 3937-3944

Hattersley, K., Laurie, K.J., Liebelt, J.E. *et al.* (2010) A novel syndrome of paediatric cataract, dysmorphism, ectodermal features, and developmental delay in Australian Aboriginal family maps to 1p35.3-p36.32. **BMC Med Genet.** 11: 165

Hejtmancik, J.F. (2008) Congenital cataracts and their molecular genetics. **Semin Cell Dev Biol.** 19 (2): 134-149

Hellemans, J., Coucke, P.J., Giedion, A. *et al.* (2003) Homozygous mutations in IHH cause acrocapitofemoral dysplasia, an autosomal recessive disorder with cone-shaped epiphyses in hands and hips. **Am J Hum Genet.** 72: 1040-1046

Héon, E., Liu, S., Billingsley, G. *et al.* (1998) Gene localization for aculeiform cataract, on chromosome 2q33-35. **Am J Hum Genet.** 63 (3): 921-926

Héon, E., Priston, M., Schorderet, D.F. *et al.* (1999) The gamma-crystallins and human cataracts: a puzzle made clearer. **Am J Hum Genet.** 65 (5): 1261-1267

Héon, E., Paterson, A.D., Fraser, M. (2001) A progressive autosomal recessive cataract locus maps to chromosome 9q13-q22. **Am J Hum Genet.** 68 (3): 772-777

Horwitz, J. (1992) Alpha-crystallin can function as a molecular chaperone. **Proc Natl Acad Sci USA.** 89: 10449-10453

Huang, B. and He, W. (2010) Molecular characteristics of inherited congenital cataracts. **Eur J Hum Genet.** 53: 347-357

Ionides, A., Francis, P., Berry, V. *et al.* (1999) Clinical and genetic heterogeneity in autosomal dominant cataract. **Br J Ophthalmol.** 83 (7): 802-808

Jamieson, R.V., Perveen, R., Kerr, B. *et al.* (2002) Domain disruption and mutation of the bZIP transcription factor, MAF, associated with cataract, ocular anterior segment dysgenesis and coloboma. **Hum Mol Genet.** 11: 33-42

Joyce, S., Tee, L., Abid, A. *et al.* (2010) Locus heterogeneity and Knobloch syndrome. **Am J Med Genet. A.** 152A (11): 2880-2881

Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. and Kondoh. H. (2001) Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. **Genes Dev.** 15 (10): 1272-1286

Kamradt, M.C., Lu, M., Werner, M.E. *et al.* (2005) The small heat shock protein alpha B-crystallin is a novel inhibitor of TRAIL-induced apoptosis that suppresses the activation of caspase-3. **J Biol Chem.** 280 (12): 11059-11066

Kantorow, M. and Piatigorsky, J. (1994) Alpha-crystallin/small heat shock protein has autokinase activity. **Proc Natl Acad Sci USA.** 91: 3112-3116

Karagoz, I.D., Ozaslan, M., Cengiz, B. *et al.* (2010) CDC25A gene 263C/T, -350C/T, and -51C/G polymorphisms in breast carcinoma. **Tumour Biol.** 31: 597-604

Khaliq, S., Abid, A., White, D.R. *et al.* (2007) Mapping of a novel type III variant of Knobloch syndrome (KNO3) to chromosome 17q11.2. **Am J Med Genet A.** 143A (23): 2768-2774

Kirikoshi, H., Sekihara, H. and Katoh, M. (2001) WNT10A and WNT6, clustered in human chromosome 2q35 region with head-to-tail manner, are strongly coexpressed in SW480 cells. **Biochem Biophys Res Commun.** 283: 798-805

Knobloch, W.H. and Layer, J.M. (1971) Retinal detachment and encephalocele. **J Pediat Ophthal.** 8: 181-184

Kong, A., Gudbjartsson, D.F., Sainz, J. *et al.* (2002) A high-resolution recombination map of the human genome. **Nat Genet.** 31 (3): 241-247

Lander, E.S. and Botstein, D. (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. **Science.** 236 (4808):1567-1570

Lewis, S.A. and Cowan, N.J. (1990) Tubulin genes: structure, expression, and regulation. In: Avila, J. (ed.): **Microtubule proteins.** Boca Raton: CRC Press Inc. pp. 37-66

Liang, P. and MacRae, T.H. (1997) Molecular chaperones and the cytoskeleton. **J Cell Sci.** 110 (Pt 13): 1431-1440

Litt, M., Carrero-Valenzuela, R., LaMorticella, D.M. *et al.* (1997) Autosomal dominant cerulean cataract is associated with a chain termination mutation in the human beta-crystallin gene CRYBB2. **Hum Mol Genet.** 6: 665-668

Litt, M., Kramer, P., LaMorticella, D.M. *et al.* (1998) Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. **Hum Mol Genet.** 7 (3): 471-474

Mackay, D.S., Boskovska, O.B., Knopf, H.L.S., Lampi, K.J. and Shiels, A. (2002) A nonsense mutation in CRYBB1 associated with autosomal dominant cataract linked to human chromosome 22q. **Am J Hum Genet.** 71: 1216-1221

Mailand, N., Falck, J., Lukas, C. *et al.* (2000) Rapid destruction of human Cdc25A in response to DNA damage. **Science.** 288 (5470): 1425-1429

Marigo, V., Roberts, D.J., Lee, S.M. *et al.* (1995) Cloning, expression, and chromosomal location of SHH and IHH: two human homologues of the Drosophila segment polarity gene hedgehog. **Genomics.** 28 (1): 44-51

Marner, E. (1949) A family with eight generations of hereditary cataract. **Acta Ophthal.** 27: 537-551

Mehlen, P., Schulze-Osthoff, K. and Arrigo, A.P. (1996) Small stress proteins as novel regulators of apoptosis. Heat shock protein 27 blocks Fas/APO-1- and staurosporine-induced cell death. **J Biol Chem.** 271 (28): 16510-16514

Merin, S. (1991) Inherited Cataracts. Inherited Eye Diseases. New York, NY: Marcel Dekker Inc. pp:86-120

Merlini, L., Gooding, R., Lochmuller, H. *et al.* (2002) Genetic identity of Marinesco-Sjogren/myoglobinuria and CCFDN syndromes. **Neurology.** 58: 231-236

Menzel, O., Bekkeheien, R.C., Reymond, A. *et al.* (2004) Knobloch syndrome: novel mutations in COL18A1, evidence for genetic heterogeneity, and a functionally impaired polymorphism in endostatin. **Hum Mut.** 23 (1): 77-84

Miano, M.G., Jacobson, S.G., Carothers, A. *et al.* (2000) Pitfalls in homozygosity mapping. **Am J Hum Genet.** 67 (5): 1348-1351

Monecke, T., Güttler, T., Neumann, P. *et al.* (2009) Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. **Science.** 324 (5930): 1087-1091

Morgan, N.V., Westaway, S.K., Morton, J.E. *et al.* (2006) PLA2G6, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron. **Nat Genet.** 38 (7): 752-754

Morgan, T.H. (1911) Random segregation versus coupling in Mendelian inheritance. **Science.** 34: 384

Morton, N.E. (1955) Sequential tests for the detection of linkage. **Am J Hum Genet.** 7 (3): 277-318

Mubaidin, A., Roberts, E., Hampshire, D. *et al.* (2003) Karak syndrome: a novel degenerative disorder of the basal ganglia and cerebellum. **J Med Genet.** 40 (7): 543-546

Mueller, R.F. and Bishop, D.T. (1993) Autozygosity mapping, complex consanguinity, and autosomal recessive disorders. **J Med Genet.** 30 (9): 798-799

Muller-Felber, W., Zafiriou, D., Scheck, R. *et al.* (1998) Marinesco-Sjogren syndrome with rhabdomyolysis: a new subtype of the disease. **Neuropediatrics.** 29: 97-101

Naz, S., Giguere, C.M., Kohrman, D.C. *et al.* (2002) Mutations in a novel gene, TMIE, are associated with hearing loss linked to the DFNB6 locus. **Am J Hum Genet.** 71: 632-636

Nettleship and Ogilvie (1906) A peculiar form of hereditary congenital cataract. **Trans. Ophthal. Soc. U.K.** 191-207

Ng, S.B., Turner, E.H., Robertson, P.D. *et al.* (2009) Targeted capture and massively parallel sequencing of 12 human exomes. **Nature.** 461 (7261): 272-276

Ng, S.B., Bigham, A.W., Buckingham, K.J. *et al.* (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. **Nat Genet.** 42 (9): 790-793

Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K.T. and Kawai, S. (1989) v-maf, a viral oncogene that encodes a "leucine zipper" motif. **Proc Natl Acad Sci USA.** 86 (20): 7711-7715

Okano, Y., Asada, M., Fujimoto, A. *et al.* (2001) A genetic factor for age-related cataract: identification and characterization of a novel galactokinase variant, 'Osaka,' in Asians. **Am J Hum Genet.** 68: 1036-1042

Oliver, G., Loosli, F., Köster, R., Wittbrodt, J. and Gruss, P. (1996) Ectopic lens induction in fish in response to the murine homeobox gene Six3. **Mech Dev.** 60: 233-239

Ostergaard, P., Simpson, M.A., Brice, G. *et al.* (2011) Rapid identification of mutations in *GKC2* in primary lymphoedema using whole exome sequencing combined with linkage analysis with delineation of the phenotype. **J Med Genet.** 48 (4): 251-255

Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Splicing of messenger RNA precursors. **Annu Rev Biochem.** 55: 1119-1150

Percin, E.F., Ploder, L.A., Yu, J.J. (2000) Human microphthalmia associated with mutations in the retinal homeobox gene CHX10. **Nat Genet.** 25: 397-401

Perng, M.D. and Quinlan, R.A. (2005) Seeing is believing! The optical properties of the eye lens are dependent upon a functional intermediate filament cytoskeleton. **Exp Cell Res.** (15)305: 1-9

Polyakov, A.V., Shagina, I.A., Khlebnikova, O.V. and Evgrafov, O.V. (2001) Mutation in the connexin 50 gene (GJA8) in a Russian family with zonular pulverulent cataract. **Clin Genet.** 60 (6): 476-478

Pras, E., Frydman, M., Levy-Nissenbaum, E. *et al.* (2000) A nonsense mutation (W9X) in CRYAA causes autosomal recessive cataract in an inbred Jewish Persian family. **Invest Ophthalmol Vis Sci.** 41 (11): 3511-3515

Pras, E., Levy-Nissenbaum, E., Bakhan, T. et al. (2002) A missense mutation in the LIM2 gene is associated with autosomal recessive presenile cataract in an inbred Iraqi Jewish family. **Am J Hum Genet.** 70 (5): 1363-1367

Pras, E., Raz, J., Yahalom, V. *et al.* (2004) A nonsense mutation in the glucosaminyl (N-acetyl) transferase 2 gene (GCNT2): association with autosomal recessive congenital cataracts. **Invest Ophthalmol Vis Sci.** 45 (6): 1940-1945.

Qi, Y., Jia, H., Huang, S. *et al.* (2004) A deletion mutation in the beta-A1/A3 crystallin gene (CRYBA1/A3) is associated with autosomal dominant congenital nuclear cataract in a Chinese family. **Hum Genet.** 114: 192-197

Ramachandran, R.D., Perumalsamy, V. and Hejtmancik, J.F. (2007) Autosomal recessive juvenile onset cataract associated with mutation in BFSP1. **Hum Genet.** 121: 475-482

Ray, D. and Kiyokawa, H. (2007) CDC25A levels determine the balance of proliferation and checkpoint response. **Cell Cycle.** 6 (24): 3039-3042

Reddy, M.A., Francis P.J., Berry, V,. Bhattacharya S.S., Moore, A.T. (2004) Molecular genetic basis of inherited cataract and associated phenotypes. **Sur Ophthalmol.** 49 (3): 300-315

Reiter, L.T., Potocki, L., Chien, S., Gribskov, M. and Bier, E. (2001) A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. **Genome Research.** 11 (6): 1114-1125

Ren, Z., Li, A., Shastry, B.S. *et al.* (2000) A 5-base insertion in the gamma-C-crystallin gene is associated with autosomal dominant variable zonular pulverulent cataract. **Hum Genet.** 106: 531-537

Resnikoff, S., Pascolini, D., Etya'ale, D. *et al.* (2004) Global data on visual impairment in the year 2002. **Bull World Health Organ.** 82 (11): 844-851

Riazuddin, S.A., Yasmeen, A., Yao, W. *et al.* (2005) Mutations in betaB3-crystallin associated with autosomal recessive cataract in two Pakistani families. **Invest Ophthalmol Vis Sci.** 46 (6): 2100-2106

Robinson, M.L., Ohtaka-Maruyama, C., Chan, C.C. (1998) Disregulation of ocular morphogenesis by lens-specific expression of FGF-3/int-2 in transgenic mice. **Dev Biol.** 198 (1): 13-31

Roessler, E. and Muenke, M. (2003) How a Hedgehog might see holoprosencephaly. **Hum Mol Genet.** 12 Spec No 1: R15-25

Rogaev, E.I., Rogaeva, E.A., Korovaitseva, G.I. *et al.* (1996) Linkage of polymorphic congenital cataract to the gamma-crystallin gene locus on human chromosome 2q33-35. **Hum Mol Genet.** 5 (5): 699-703

Rong, P., Wang, X., Niesman, I. *et al.* (2002) Disruption of Gja8 (alpha8 connexin) in mice leads to microphthalmia associated with retardation of lens growth and lens fiber maturation. **Development.** 129 (1): 167-174

Salen, G., Meriwether, T.W. and Nicolau, G. (1975) Chenodeoxycholic acid inhibits increased cholesterol and cholestanol synthesis in patients with cerebrotendinous xanthomatosis. **Biochem Med.** 14 (1): 57-74

Santhiya, S.T., Manohar, M.S., Rawlley, D. *et al.* (2002) Novel mutations in the gamma-crystallin genes cause autosomal dominant congenital cataracts. **J Med Genet.** 39: 352-358

Schmitt, J.F., Millar, D.S., Pedersen, J.S. *et al.* (2002) Hypermethylation of the inhibin alpha-subunit gene in prostate carcinoma. **Mol Endocr.** 16: 213-220

Seaver, L.H., Joffe, L., Spark, R.P., Smith, B.L. and Hoyme, H.E. (1993) Congenital scalp defects and vitreoretinal degeneration: redefining the Knobloch syndrome. **Am J Med Genet.** 46: 203-208

Selcen, D. and Engel, A.G. (2003) Myofibrillar myopathy caused by novel dominant negative alpha B-crystallin mutations. **Ann Neurol.** 54 (6): 804-810

Senderek, J., Krieger, M., Stendel, C. *et al.* (2005) Mutations in SIL1 cause Marinesco-Sjögren syndrome, a cerebellar ataxia with cataract and myopathy. **Nat Genet.** 37 (12):1312-1314

Semina, E.V., Ferrell, R.E., Mintz-Hittner, H.A. *et al.* (1998) A novel homeobox gene PITX3 is mutated in families with autosomal-dominant cataracts and ASMD. **Nature Genet.** 19: 167-170

Semina, E.V., Brownell, I., Mintz-Hittner, H.A., Murray, J.C. and Jamrich, M. (2001) Mutations in the human forkhead transcription factor FOXE3 associated with anterior segment ocular dysgenesis and cataracts. **Hum Mol Genet.** 10: 231-236

Sertié, A.L., Quimby, M., Moreira, E.S. *et al.* (1996) A gene which causes severe ocular alterations and occipital encephalocele (Knobloch syndrome) is mapped to 21q22.3. **Hum Mol Genet.** 5 (6): 843-847

Shiels, A., Mackay, D., Ionides, A. *et al.* (1998) A missense mutation in the human connexin50 gene (GJA8) underlies autosomal dominant "zonular pulverulent" cataract, on chromosome 1q. **Am J Hum Genet.** 62 (3): 526-532

Shiga, K., Fukuyama, R., Kimura, S., Nakajima, K. and Fushiki, S. (1999) Mutation of the sterol 27-hydroxylase gene (CYP27) results in truncation of mRNA expressed in leucocytes in a Japanese family with cerebrotendinous xanthomatosis. **J Neurol Neurosurg Psychiatry.** 67 (5): 675-677

Smaoui, N., Beltaief, O., BenHamed, S. *et al.* (2004) A homozygous splice mutation in the HSF4 gene is associated with an autosomal recessive congenital cataract. **Invest Ophthalmol Vis Sci.** 45 (8): 2716-2721

Smith, C.A.B. (1953) The detection of linkage in human genetics. **JR Stat Soc B.** 15: 153-184

Stambolian, D., Ai, Y., Sidjanin, D. *et al.* (1995) Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts. **Nat Genet.** 10: 307-312

Steele, E.C. Jr, Kerscher, S., Lyon, M.F. *et al.* (1997) Identification of a mutation in the MP19 gene, Lim2, in the cataractous mouse mutant To3. **Mol Vis.** 3: 5

Strachan, T. and Read, A.P. (1999) **Human Molecular Genetics 2**. BIOS Scientific Publishers Ltd, Second edition, reprinted 2001

Sun, H., Ma, Z., Li, Y. *et al.* (2005) Gamma-S crystallin gene (CRYGS) mutation causes dominant progressive cortical cataract in humans. **J Med Genet.** 42 (9): 706-710

Suzuki, O.T., Sertié, A.L., Der Kaloustian, V.M. *et al.* (2002) Molecular analysis of collagen XVIII reveals novel mutations, presence of a third isoform, and possible genetic heterogeneity in Knobloch syndrome. **Am J Hum Genet.** 71 (6): 1320-1329

Suzuki, O., Kague, E., Bagatini, K. *et al.* (2009) Novel pathogenic mutations and skin biopsy analysis in Knobloch syndrome. **Mol Vis.** 15: 801-809

Takahata, T., Yamada, K., Yamada, Y. *et al.* (2010) Novel mutations in the SIL1 gene in a Japanese pedigree with the Marinesco-Sjögren syndrome. **J Hum Genet.** 55 (3): 142-146

Talamas, E., Jackson, L., Koeberl, M. *et al.* (2006) Early transposable element insertion in intron 9 of the Hsf4 gene results in autosomal recessive cataracts in lop11 and ldis1 mice. **Genomics.** 88 (1): 44-51

Tessier, D.J., Komalavilas, P., Panitch, A., Joshi, L. and Brophy, C.M. (2003) The small heat shock protein (HSP) 20 is dynamically associated with the actin cross-linking protein actinin. **J Surg Res.** 111: 152-157

Unger, V.M., Kumar, N.M., Gilula, N.B. and Yeager, M. (1999) Three-dimensional structure of a recombinant gap junction membrane channel. **Science.** 283 (5405): 1176-1180

Van Bogaert, L., Scherer, H.J. and Epstein, E. (1937) **Une forme cerebrale de la cholesterinose generalisee.** Paris: Masson (pub.)

van Heyningen, V. and Williamson, K.A. (2002) PAX6 in sensory development. **Hum Mol Genet.** 11 (10):1161-1167

Varma, S.D., Kovtun, S. and Hegde, K.R. (2011) Role of Ultraviolet Irradiation and Oxidative Stress in Cataract Formation. Medical Prevention by Nutritional Antioxidants and Metabolic Agonists. .**Eye Contact Lens.** [Epub ahead of print]

Varon, R., Gooding, R., Steglich, C. *et al.* (2003) Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome. **Nat Genet.** 35: 185-189

Vicart, P., Caron, A., Guicheney, P. *et al.* (1998) A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy. **Nat Genet.** 20 (1): 92-95

Voronina, V.A., Kozhemyakina, E.A., O'Kernick, C.M. *et al.* (2004) Mutations in the human RAX homeobox gene in a patient with anophthalmia and sclerocornea. **Hum Mol Genet.** 13: 315-322

Walker, P.D., Blitzer, M.G. and Shapira, E. (1985) Marinesco-Sjögren syndrome: evidence for a lysosomal storage disorder. **Neurology.** 35 (3): 415-419

Wallis, D.E., Roessler, E., Hehr, U. *et al.* (1999) Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. **Nat Genet.** 22 (2): 196-198

Yamada, K., Tomita, H., Yoshiura, K. *et al.* (2000) An autosomal dominant posterior polar cataract locus maps to human chromosome 20p12-q12. **Eur J Hum Genet.** 8 (7): 535-539

Wang, J.L., Yang, X., Xia, K. *et al.* (2010) TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. **Brain.** 133 (Pt 12): 3510-3518

Wang, X., Garcia, C.M., Shui, Y.B. and Beebe, D.C. (2004) Expression and regulation of alpha-, beta-, and gamma-crystallins in mammalian lens epithelial cells. **Invest Ophthalmol Vis Sci.** 45 (10): 3608-3619

Weissenbach, J., Gyapay, G., Dib, C. *et al.* (1992) A second-generation linkage map of the human genome. **Nature.** 359 (6398): 794-801

White, D.R., Ganesh, A., Nishimura, D. *et al.* (2007) Autozygosity mapping of Bardet-Biedl syndrome to 12q21.2 and confirmation of FLJ23560 as BBS10. **Eur J Hum Genet.** 15 (2): 173-178

Wong, F.L., Cantor, R.M. and Rotter, J.I. (1986) Sample-size considerations and strategies for linkage analysis in autosomal recessive disorders. **Am J Hum Genet.** 39:25-37

Woods, C.G., Cox, J., Springell, K. *et al.* (2006) Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. **Am J Hum Genet.** 78: 889-896

Wright, S. (1922) Coefficients of inbreeding and relationship. **Am Nat.** 56: 330

Yeager, M. and Nicholson, B.J. (2000) Structure and biochemistry of gap junction. **Adv Mol Cell Biol.** 30: 31-98

Yu, L.C., Twu, Y.C., Chou, M.L. *et al.* (2003) The molecular genetics of the human I locus and molecular background explain the partial association of the adult i phenotype with congenital cataracts. **Blood.** 101 (6):2081-2088

Zenker, M., Tralau, T., Lennert, T. *et al.* (2004) Congenital nephrosis, mesangial sclerosis, and distinct eye abnormalities with microcoria: an autosomal recessive syndrome. **Am J Med Genet.** 130A: 138-145

Zhang, J., Chiodini, R., Badr, A. and Zhang, G. (2011) The impact of next-generation sequencing on genomics. **J Genet Genomics.** 38 (3): 95-109

Zhang, L., Mathers, P.H. and Jamrich, M. (2000) Function of Rx, but not Pax6, is essential for the formation of retinal progenitor cells in mice. **Genesis.** 28 (3-4): 135-142

Zhu, C.C., Dyer, M.A., Uchikawa, M. *et al.* (2002) Six3-mediated auto repression and eye development requires its interaction with members of the Groucho-related family of co-repressors. **Development.** 129 (12): 2835-2849

Zimmer, C., Gosztonyi, G., Cervos-Navarro, J., von Moers, A. and Schröder, J.M. (1992) Neuropathy with lysosomal changes in Marinesco-Sjögren syndrome: fine structural findings in skeletal muscle and conjunctiva. **Neuropediatrics.** 23 (6): 329-335

Chapter 6
Appendix

Primer	***Sequence***
CRYBA2-1F	CCTCAGCACTTTCTGTTGTC
CRYBA2-1R	TTTCCCATGTTGGCAGGTCT
CRYBA2-2F	AAGGGTTTGGGGTTGGAGAA
CRYBA2-2R	TCTTGGAGACACCCTTCTGA
CRYBA2-3F	GACAGCTTCCACAGCTTGTGA
CRYBA2-3R	CAGTCTGACTCCATCATCTC
CRYBA2-4F	GCTGCTGGTGGAGAACATAG
CRYBA2-4R	TTCGTAGCCAAGGATGCTGG

Table 6.1: *Primers used for sequencing coding regions of CRYBA2 gene.*

Primer	***Sequence***	***Primer length (bp)***	***Product size***	***Annealing temperature***
IHH-1F	CATCAGCCCACCAGGAGA	18	459	57°C +acetamide
IHH-1R	CCAGCCAGTCGAGAAAATGT	20		
IHH-2F	CGCCTACACCTGCACCTC	18	400	57°C
IHH-2R	CTACTCCTCCTGCCCATGC	19		
IHH-3A F	GTTGTGACCAGGAGGCTGAG	20	525	57°C
IHH-3A R	AGCAGGATGCCACCACAT	18		
IHH-3B F	GACACTGGTGGTGGAGGATG	20	469	57°C
IHH-3B R	GAGAGAGGGGTCAACAACCA	20		
IHH-3C F	AGTCATAGAGCTGCAAGCTGAG	22	459	57°C
IHH-3C R	CAGCAGAAGGGGGAGCTTA	19		
IHH-3D F	TGTGTTCTGGCCAATGTGAC	20	324	57°C
IHH-3D R	AAGGTGAGCCAGGCCTCT	18		

Table 6.2: *Primers for IHH gene.*

Primer	***Sequence***	***Primer length (bp)***	***Product size***	***Annealing temperature***
INHA-1F	GTGGGAGATAAGGCTCATGG	20	561	57°C
INHA-1R	CATGCTGTGCCTTGCTTTT	19		
INHA-2A F	AGAGGAGGGGTGCCAGGT	18	594	57°C
INHA-2A R	TGCTACTCTGTGGCAGTTGG	20		
INHA-2B F	CTCAACTCCCCTGATGTCCT	20	477	57°C
INHA-2B R	CTTTCCTCCCAGCTGATGAT	20		

Table 6.3: *Primers for INHA gene.*

Primer	*Sequence*	*Primer length (bp)*	*Product size*	*Annealing temperature*
TUBA4A-1F	GCAGCTAGCGCAGTTCTCA	20	150	57°C
TUBA4A-1R	AGCATGCCTGGAAGAAGT	20		
TUBA4A-2F	GAGACTCGCAAGGTGAGG	20	352	57°C
TUBA4A-2R	TCTCCAGGGAGAAGCTTTG	20		
TUBA4A-3F	TGAGTCCTTCCATGCATCT	20	293	57°C
TUBA4A-3R	TTCCTGACCATTAGCACAG	22		
TUBA4A-4A F	TACGCTTTATGGAGCCAGG	20	593	57°C
TUBA4A-4A R	AAAGCAGGCATTGGTGATC	20		
TUBA4A-4B F	CGCCTCATTAGCCAAATTG	20	599	57°C
TUBA4A-4B R	TCTTCTCCCTCATCCTCGT	20		
TUBA4A-4C F	TCAACTACCAGCCTCCCAC	20	395	57°C
TUBA4A-4C R	GAAGGCAGCAGAAGCTCA	20		

Table 6.4: *Primers for TUBA4A gene.*

Primer	*Sequence*	*Primer length (bp)*	*Product size*	*Annealing temperature*
WNT10A-1F	AGTCCCACTGGGCTGTGA	18	249	57°C
WNT10A-1R	TCTACCCCAGCAAGAGCATC	20		
WNT10A-2F	TGTTAGATGGGGCTCTCCTG	20	380	57°C
WNT10A-2R	CTGAGGTGGAGATGCTGGA	19		
WNT10A-3F	CTGGAGAATGGGGTGTCAAG	20	493	57°C
WNT10A-3R	GGGTAGGCCCTAGGCAAAG	19		
WNT10A-4A	GGAGTGGGTTTCAGAAGCAG	20	494	57°C
WNT10A-4A	GTCTGGCGCAGGATGTTGT	19		
WNT10A-4B	CGCCGACCTGGTCTACTTC	19	329	57°C
WNT10A-4B	CCAAGACCGTAAGCCTCAGA	20		

Table 6.5: *Primers for WNT10A gene.*

Primer	*Sequence*	*Primer length (bp)*	*Product size(bp)*	*Annealing temperature*
CYP27A1-1A F	CCCTCCAGGGATCAGATG	20	480	57°C
CYP27A1-1A R	CTTTGGGTCGAGTGCTGA	20		
CYP27A1-1B F	GTCCCTCTCCTGCAGCTC	18	500	57°C
CYP27A1-1BR	TCACCTCTAGCCTCTCTGT	22		
CYP27A1-2F	CCTCACCTCATCTCCTGTC	20	392	57°C
CYP27A1-2R	AGACCATCAGGCTCAGAG	20		
CYP27A1-3F	GGTGGACTTCAGGGTGAG	20	387	57°C
CYP27A1-3R	GGAGGCAGAACCAGGACA	20		
CYP27A1-4F	GTCAGGGAGAGGTTGTGC	20	388	57°C
CYP27A1-4R	TGGTAGCATGAGGGAAAA	20		
CYP27A1-5F	TCCTTGGAGATCATGACTT	22	290	57°C

CYP27A1-5R	AGGGAACTGGTTCAGGTT	20		
CYP27A1-6F	CCCATTACTGGCCTCTTTC	20	281	57°C
CYP27A1-6R	GATTCCCTCCCCACAAAGA	19		
CYP27A1-7F	ATGGGAGAGGTAGGGGAG	20	240	57°C
CYP27A1-7R	GAAGATTGGGCAGCATGA	20		
CYP27A1-8newF	ATTGCCCAGGAGTGCCCT	20	397	57°C
CYP27A1-8newR	TTGTGTGTTTGCCATCCAC	20		
CYP27A1-9F	CACAGGGTAGGAGTGTGC	20	499	57°C
CYP27A1-9R	TGACTCACTGGAGGAGAG	20		

Table 6.6: *Primers for CYP27A1 gene.*

Primer	***Sequence***	***Annealing Temperature (°C)***	***Annealing Time (seconds)***
SIL1-1F	GCAGGCGTGAAGCTATCTCT	64	60
SIL1-1R	AGGAGGAGGCCCGGAAG		
SIL1-2F	GTTAGTAGGGATGCGGCAGT	57	30
SIL1-2R	AGGAGAACATTCATTGGTGA		
SIL1-3F	GGGAACCAAGCATTAGCTTTT	59	60
SIL1-3R	ACAACAGCCCTGGTGACAG		
SIL1-4F	GAAAAGAACGTGCCACCCTA	57	30
SIL1-4R	GGGACAAAGGACATATGAGT		
SIL1-5F	CGACGACCTTCCAAAGACCT	57	30
SIL1-5R	CATGAATTTGGGTAACATTTTAAGT		
SIL1-6F	ACTGTGAGCCACCAAGCCTA	57	30
SIL1-6R	CCATTCTCGCAAAGACAACA		
SIL1-7F	TGTGAGTCTTGGAGCAGCAG	57	30
SIL1-7R	GCTTAGGGAAGGGGGATTTTA		
SIL1-8F	AACAGCTTTCTGTCCCTGA	57	30
SIL1-8R	CAAGTGTACCCTTCTTCCAAGC		
SIL1-9F	GGCTCAGTTCCCTGACCAT	58.4	30
SIL1-9R	AGGAACCCTTTCCTTTCAGG		
SIL1-10F	GGACTAATTTGGGGCACCTC	57	30
SIL1-10R	AAACAAGTGACGTCCGCAGT		
SIL1-11A F	GCTCCCTGTGTGTTCTCTCC	57	30
SIL1-11A F	AGAGGATGGCCTTCAGGTTT		
SIL1-11B F	AGCTTGCTGAAGGAGCTGAG	59	60
SIL1-11B R	CGATAGAGGGGCAGTGAAT		

Table 6.7: *Primers for SIL1 gene*

Primer	*Sequence*	*Primer length (bp)*	*Product size*	*Annealing temperature*
C17orf42- 1F	GCTTGGACAACAACCTTTCC	20	200	57°C
C17orf42- 1R	GCCTCTAGGGTCAGAGGCTAA	21		
C17orf42- 2F	GCTTGCAACTAACCCCAAAG	20	600	57°C
C17orf42- 2R	CAATGCCCAGAAGAGTGCTT	20		
C17orf42- 3F	CCCAGAAGGTGGAGCTTGTA	20	359	57°C
C17orf42- 3R	TCCTTTTGGGACCAGAAGATT	21		
C17orf42- 4A F	GACAGAGAGGGGACCAAACA	20	491	57°C
C17orf42- 4A R	CACTGCTAATTCATAGAAGGCAAT	24		
C17orf42- 4B F	GGGGAAGCATTTTGAACTGA	20	499	57°C
C17orf42- 4B R	GCTGGGATTATCGGCATAAG	20		

Table 6.8: *Primers for C17orf42 gene.*

Primer	*Sequence*	*Primer length (bp)*	*Product size*	*Annealing temperature*
OMG- 1F	TTGTCCATTTTGTTGCGTGT	20	298	57°C
OMG- 1R	AAAAGAAAAAGGAAAAAGATTCCA	24		
OMG- 2A F	ACCCATGCAGATGCCTAAAC	20	489	57°C
OMG- 2A R	GAGGACAACCTTTTCCAGCA	20		
OMG- 2B F	CCTCGGTCTCTGTGGAACAT	20	488	57°C
OMG- 2B R	GATGATATTTGGGTAGAACATGGA	24		
OMG- 2C F	TTTGACCAACTCTTTCAGTTGC	22	500	57°C
OMG- 2C R	GGTGTTGGGGATGATTTTGT	20		
OMG- 2D F	TCCACAGAGACTATCAATTCACAT	24	500	57°C
OMG- 2D R	AGCAAGTACCAAGACATTGTGC	22		
OMG- 2E F	GAAAGCACTCCTCCCTGATG	20	395	57°C
OMG- 2E R	GGTGAGGTCATTTTGATTCTGC	22		

Table 6.9: *Primers for OMG gene.*

Primer	*Sequence*	*Primer length(bp)*	*Product size*	*Annealing temperature*
EVI2A- 1F	ATGCCAGTGGCCAACTAGAG	20	328	62°C
EVI2A- 1R	TTTGATAGCATTTCCTCGAACA	22		
EVI2A- 2F	TTTTGTTTGCATAACTCAACGTA	23	298	56°C
EVI2A- 2R	TTGCCTGGTTAAGAACCCTA	20		
EVI2A- 3A F	AGGCATAAGCTTGTAGTTTACTTGT	25	600	57°C
EVI2A- 3A R	AACCACAGTTGATAGAAATAGAAAGG	26		
EVI2A- 3B F	CACAAGCAAAAGTCATGGTGA	21	492	57°C
EVI2A- 3B R	TGCAAGGTCTACCATGTCAAA	21		

Table 6.10: *Primers for EVI2A gene.*

Primer	*Sequence*	*Primer length(bp)*	*Product size*	*Annealing temperature*
EVI2B- 1F	CTTCCGGAACCAGCTATTGA	20	300	57°C
EVI2B- 1R	AACCACAAAAACCTGTTATAGACATC	26		
EVI2B- 2A F	TGCCAGGAATCATAATGAAAA	21	577	57°C
EVI2B- 2A R	AGGGATCTGGACAGATGATGA	21		
EVI2B- 2B F	CCAATAGCCAACACCTCCTC	20	584	57°C
EVI2B- 2B R	TTGGGGTTGTTGGAGTCTTC	20		
EVI2B- 2C F	CTGATGGAGAAACCCCTGAC	20	594	57°C
EVI2B- 2C R	GAGGTGGCAGGGATTCATTA	20		
EVI2B- 2D F	AATCATTTCCACCGCTTGAC	20	732	57°C
EVI2B- 2D R	CCCAACCCTTCAGTTGATGT	20		

Table 6.11: *Primers for EVI2B gene.*

Primer	*Sequence*	*Primer length (bp)*	*Product size*	*Annealing temperature*
RNF135- 1A F	GTCTTTCCAGCTGGGCACT	19	441	51°C +acetamide
RNF135- 1A R	AGGTCCTGCAGTAGCGTGTT	20		
RNF135- 1B F	GACCTCGGCTGCATCATC	18	391	56°C
RNF135- 1B R	CCACATGGGAAAGGGTCA	18		
RNF135- 2F	GTGGTTCCTGGGTCCAGTTT	20	254	57°C
RNF135- 2R	GCAATCCCTAGCCCTATTCC	20		
RNF135- 3F	TTAAAAACTTGCTGGGAATTAAA	23	367	56°C
RNF135- 3R	CCCCAGAGAACAAAGCCATA	20		
RNF135- 4F	TGCCTTGACCATGATGTTTG	20	233	51°C
RNF135- 4R	ATATTCTCGAAGCCCGCTCT	20		
RNF135- 5A F	TGGACCATCAAAAGATGACC	20	646	57°C
RNF135- 5A R	CCTGATTGATGACAGGGAGA	20		
RNF135- 5B F	CTTTGTACCCTGCCTTCTGG	20	500	58.4°C
RNF135- 5B R	CTTGCAAGGTCAAGCTGATACT	22		
RNF135- 5C F	TGAACCCTGGACATTCCAAA	20	491	57°C
RNF135- 5C R	CTGCCCTCAGGTTCAGTGAT	20		

Table 6.12: *Primers for RNF135 gene.*

Primer	**Sequence**	**Product Size bp**	**Annealing Temp. °C**
RHOA- 1F	CGGCTGCAATGATTGGTTA	485	57
RHOA -1R	GAGTCAGGTCAAGGGTCAGG		
RHOA- 2F	CCAAAGCATGTGTCATCCTG	295	57
RHOA- 2R	CATGCTCCATAAATATTCTAACATGG		
RHOA- 3F	TCAATGGTCTGCCATTTCAA	296	57
RHOA- 3R	ACTTGATGCCCAGAAACCAG		
RHOA- 4F	TGCAATTTCACTGAGGTTCTTG	300	57

RHOA- 4R	CCTGTGAAGGTTCCTGCTG		
RHOA- 5AF	ACCGACGAGCAAAACTGTCT	474	57
RHOA -5AR	ACGGGTTGGACATCGTTAAT		
RHOA- 5BF	TGCTACCAGTATTTAGAAGCCAAC	537	57
RHOA- 5BR	CCAATACACTTTCTTTGAGGATGA		
RHOA- 5CF	TGCCACTTAATGTATGTTACCAAAA	600	57
RHOA- 5CR	CTCTCCAGGGTGAGCTAGGC		

Table 6.13: *RHOA Primers.*

Primer	**Sequence**	**Product Size bp**	**Annealing Temp. °C**
LAMB2-1F	TTGGGGGCACAATAAAGATT	494	57
LAMB2-1R	AAGGGCAACTACCAGGACCTA		
LAMB2-2F	AGGTGGAGCCACTGGATAG	300	57
LAMB2-2R	GGGATTAGAATCAGTGCCTCAG		
LAMB2-3F	CTGGGGACAGGGTGTGAC	292	57
LAMB2-3R	GGACAGGAATCGGAAGTCAA		
LAMB2-4F	GTGGGGCAAAAGTGAGCAG	239	57
LAMB2-4R	ATGCTCAAGGAGGCTGTGTT		
LAMB2-5F	CAGTCACCCCCACTCTCTGT	400	57
LAMB2-5R	TGGAGAAGCAGGGAGTTCAT		
LAMB2-6F	TGGAAGCTGGGGATGATG	192	57
LAMB2-6R	GAGGCCCCAAATAGTCCCTA		
LAMB2-7F	CGGGTGGCAGTGTATAGGAG	394	57
LAMB2-7R	GGCAGGTAGGTTGTCAGCAC		
LAMB2-8F	GCTTCAGACGGTTGGGATAG	292	57
LAMB2-8R	TAGGGGCTTGACCAACTAGC		
LAMB2-9F	AGCCCCTACCCACAAATTCT	395	57
LAMB2-9R	CTAACCCCAATTTCCTGCAA		
LAMB2-10F	GGACTGGCAGTGAGCTAGTTG	298	57
LAMB2-10R	AGGAGTAAGAGGACAGGGGTTA		
LAMB2-11F	GAACATGAGCACTGCCACAG	227	57
LAMB2-11R	CAGCCCTGTGCTCTAAGGA		
LAMB2-12F	GTGGTTGGGATTTCCTGCT	248	57
LAMB2-12R	CTGGAGGGAACTCTGTGGTT		
LAMB2-13F	GGATGGGGAAAGATCCTGAG	300	57
LAMB2-13R	ATCACCCCTATCCCCTCAAC		
LAMB2-14F	CTGTGGCAGAAGAGGCAAAG	287	57
LAMB2-14R	AACGGGATGATGGAGGAGAT		

LAMB2-15F	ATCTGGGGCTGAACCTCTG	280	57
LAMB2-15R	GAATGAGCTGTGGGGTCTTC		
LAMB2-16F	GGAGGCCCAGATGATTTGTA	274	57
LAMB2-16R	ACCCCTGCTATCCCTCAAGT		
LAMB2-17F	AGTGCTCTGGCAATGCCTAC	323	57
LAMB2-17R	AGGGTTGGGCCAAGTCAG		
LAMB2-18F	TCTGTGATGTGGGTTGATGG	292	57
LAMB2-18R	AGCAGGTCCAGAAGGAGGAG		
LAMB2-19F	TCCTGACCTTCCCACCTAGA	427	57
LAMB2-19R	GCAGGGAGGAGGAAGAAGAA		
LAMB2-20F	ATGAGTGGGTTGGTTGGTG	300	57
LAMB2-20R	TTAAGAGGAAGCTGTGAGTGCT		
LAMB2-21F	CGTGGGATCAGGGTGAGTT	348	57
LAMB2-21R	GTCAAGAGGAGGCCAAGAAG		
LAMB2-22F	CTGAATTGGGGTTCTGCTTC	353	57
LAMB2-22R	AAGAGGCTAAGCCCTGGAGA		
LAMB2-23F	GCAGATCCCTGGTTAGATGG	249	57
LAMB2-23R	GTCACTGTGGGGCATTCTG		
LAMB2-24F	GCCAGGGTAGATGGAGGACT	500	57
LAMB2-24R	CTGGTCCACTGGGCCTCT		
LAMB2-25F	GCGGAAGGGCCTAAGAATAC	352	57
LAMB2-25R	CTAAATTGGGCAGAGGCAGA		
LAMB2-26F	TAGCCAACTAGGCAGGTGGA	381	57
LAMB2-26R	CAGGCACAGAACAAGTCAGG		
LAMB2-27F	GGCAGGTGGGTGTAGATGTT	469	57
LAMB2-27R	ATACAGACACCGGGAGTTGC		
LAMB2-28F	CCTGAGTCCATACCCCAAAA	362	57
LAMB2-28R	GGAGAAGGGGGAATCAGAAC		
LAMB2-29F	CCTATCCCTGACACCTGGTT	275	57
LAMB2-29R	GGGTGAGCCAAAGGTTACAC		
LAMB2-30F	AGGACATCAGTGGGACAAGG	293	57
LAMB2-30R	CCATCTTCCACCCCTACCT		
LAMB2-31F	GGAGAAGGTAGGGGTGGAAG	299	57
LAMB2-31R	CTAGGAAGGGCAGGTGTCAG		
LAMB2-32F	TAAGAAATGGGGCACCAGAG	374	57
LAMB2-32R	CGAGGTTCCCCTGGAGACT		

Table 6.14: *LAMB2 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
TREX1-2AF	ATGTGCTGGTCCCACTAAGG	485	57
TREX1-2AR	TGCTCAGACCTGTGATCTCG		
TREX1-2BF	ACCTCCCACAGTTCCTCCA	567	57
TREX1-2BR	TTGGTCCTAGCAGAGGCTGT		
TREX1-2CF	CAGCCTAGGCAGCATCTACA	484	57
TREX1-2CR	AGGAGCCAATTGACCACTCA		

Table 6.15: *TREX1 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
CSPG5-1F	AGCGAGGGGTCTGTCTGTT	497	57
CSPG5-1R	CTCACAGCTCCACCTGTCC		
CSPG5-2AF	CTGGGTTGACCAGGTCTAGG	588	57
CSPG5-2AR	TAAGTCAGCTCTGGCCCTTG		
CSPG5-2BF	CAGCTTCTGAACTCCCCAAG	577	57
CSPG5-2BR	TTTTCACTGGAGGCCAAGTC		
CSPG5-2CF	CACTGGGAAGCCTGGTCTG	345	57
CSPG5-2CR	TTGAGTCGGCCTCACATACA		
CSPG5-3F	CTAGATGGCCTTCGGAACAG	393	57
CSPG5-3R	AAGTCCCTGGATCAGTTGGA		
CSPG5-4F	ATTGTCATGTTTTGCCGTGA	295	57
CSPG5-4R	TTCTGAAATAGACATCAGACATGG		
CSPG5-5F	CGCTGATATTATTATTTGGTTGC	289	57
CSPG5-5R	GATAATGTTTCTTCCCCTACCC		

Table 6.16: *CSPG5 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
CCDC12-1F	CCGCGAGACATTTATCCACT	250	57
CCDC12-1R	CCCTTTGCTCTCCTCAAGC		
CCDC12-2F	CCACGTGGACTGGAAGTAGG	277	57
CCDC12-2R	CAGGGCAGACACTGAACAGA		
CCDC12-3F	AATTCACGGGCTATGCAAAT	246	57
CCDC12-3R	AGAGAGCAAAGCCTGCTGAG		
CCDC12-4AND5F	CAGTGTCCCTGCAGAGTGG	464	57
CCDC12-4AND5R	CATGCTGGCACCAAGAGTC		
CCDC12-6AND7F	ACCCAATGCCACATCAGG	371	57
CCDC12-6AND7F	AAGACCATCCTCTGCAGGAC		

Table 6.17: *CCDC12 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
TMIE-1F	ATGATGCTCGCTGACTACCC	291	57
TMIE-1R	AAAAAGGCGTCCCCTCTC		
TMIE-2F	GGCCCTTCTCAGCTGGTT	240	57
TMIE-2R	GGGTGCTGTGTTCCCACT		
TMIE-3F	CCATTCCTTGGGTCTCTGAA	289	57
TMIE-3R	GAGCAGAGGAACAGGGTGAC		
TMIE-4F	GTCAGACCCCAGGACCTTGT	249	57
TMIE-4R	GAGCTCAGAGTCCAACATGC		

Table 6.18: *TMIE primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
CDC25A-1A F	AAGAGGTGTAGGTCGGCTTG	400	57
CDC25A-1A R	AGCTCCCGCTCCCTCTTC		
CDC25A-1B F	GCGGAGGCAGAGGAAGAG	295	57
CDC25A-1B R	GTTTCCTGGCCTGATCTCTC		
CDC25A-2 F	GCCTGGCTGCTTCTTAATCT	286	57
CDC25A-2 R	TTTATCACCCAATGCCATGA		
CDC25A-3 F	AAAGACCCAGGTACTTTGTTTAGTG	245	57
CDC25A-3 R	GGTAAATCCAGGGCTCCAGT		
CDC25A-4 F	CTGTACCCCCAAGCTACAGG	211	57
CDC25A-4 R	TCCTATCAAGGTCATGCTGGT		
CDC25A-5 F	TTGAGAACTGAATTTTCCACTGA	229	57

CDC25A-5 R	ACTGACAGCCTGTGGGTTTC		
CDC25A-6 F	CTGAGGCCTGAGTTGGTTTT	249	57
CDC25A-6 R	CAATTCTGGATGAAAGAAGGTG		
CDC25A-7 F	GTGACCCATTCAAGGAAATG	270	57
CDC25A-7 R	CACATGCCTTCTCCAGAGGT		
CDC25A-8 F	TTGTTCAGGGAGAAACACACC	217	57
CDC25A-8 R	ACTCTGTCCAGTGCCCCAAT		
CDC25A-9 F	AATCGGCAGTGTCTGTTGG	399	57
CDC25A-9 R	AGCAAAGAAAATAAATGAACCACT		
CDC25A-10 F	CGAGACATGCTCCAAGTGAA	250	57
CDC25A-10 R	GACAGGCTGAGTGGGATTCT		
CDC25A-11and12 F	GCCAGACAAAACAGTATAAGTGTGA	400	57
CDC25A-11and12 R	TCCTCCCCACTTTAAAACCA		
CDC25A-13 F	GCTGAATCATTACTTCATGTGTGA	250	57
CDC25A-13 R	GTCCCACAAAAGCATCATGG		
Primer	**Sequence**	**Product Size bp**	**Annealing Temp. °C**
CDC25A-14 F	CTTGTTGCTGTGCTGGAAAA	243	57
CDC25A-14 R	ATGACCACCAGACAGATCCA		
CDC25A-15A F	GAGAGGAAACCAGGTTGTGG	593	57
CDC25A-15A R	CTTTGCTATTTGGCCAGCAG		
CDC25A-15B F	GAGAAGTTACACAGAAATGCTGCT	529	57
CDC25A-15B R	ATGCTGCCTCAAGTGCTTTC		
CDC25A-15C F	GTTCCATGACATTGGCTGGT	539	57
CDC25A-15C R	TCTCTCTAGAATCAATGAAGTCTGC		
CDC25A-15D F	TTCAAAAGAACACCTCATCCAA	560	57
CDC25A-15D R	TGGCCATCCCAAACACTAAT		

Table 6.19: *CDC25A primers.*

Primer	**Sequence**	**Length**	**Product Size bp**
CDC25A-P1F	ATGGAACTGGGCCCGGAGCCCCCGC	25	634
CDC25A-P1R	ACAAAGTGGCTGTCACAGGTGACTG	25	
CDC25A-P2F	GGACTCCAGGAGGGTAAAGATCTCT	25	633
CDC25A-P2R	GGCAAACTTGCCATTCAAAACAGATGCC	28	
CDC25A-P3F	CCATTGAGAACATTTTGGACAATGACCC	28	602
CDC25A-P3R	TCAGAGCTTCTTCAGACGACTGTAC	25	

Table 6.20: *Plasmid Primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
CRYBB2-2F	GTCTCAAGGCCCCACAGAG	214	64
CRYBB2-2R	AGAAGGGCAACTAAGTCTGGG		
CRYBB2-3F	CCACGTCTACCACCCAGTTC	276	64
CRYBB2-3R	CACAACTTGATCACCTCACCC		
CRYBB2-4F	GCTTTGGGCACAGCGATGTTCTG	750	64
CRYBB2-4R	GGCCCCTTCCTGGTCCCCA		
CRYBB2-5F	AGTGGTCATAGACACGTAGTGGGTGCAC	700	64
CRYBB2-5R	CTGTTCCCAAACTTAGGGACACACGC		
CRYBB2-6F	GGCTTCACCCTTCCTAGTGG	526	64
CRYBB2-6R	CTGTTGCCCTACTCCTAGCAC		

Table 6.21: *CRYBB2 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
CRYBA4-1F	CTTGCCATTGGCTCTTCAGC	253	57
CRYBA4-1R	CACAGTGACATCCCTAGGAC		
CRYBA4-2F	TGGATCCTGACCAGGTCCAA	206	64
CRYBA4-2R	GGATTCATGGGGACCTGAAC		
CRYBA4-3F	TGCGGATCTCCACCTTTTT	250	56
CRYBA4-3R	GGGAGAGGGGACCTAGAATATC		
CRYBA4-4F	CTGTGACCGTTCTAGACCCA	290	57
CRYBA4-4R	TTCCGAAGTGCCCACATGAG		
CRYBA4-5F	GGAAGGCTGATTTGGGAGAC	307	57
CRYBA4-5R	TACACCTACCCCTCCCAGTA		
CRYBA4-6F	TGTGTTGATCACCATGCTGG	570	64+ acetamide
CRYBA4-6R	AATGACCCAGAAGTCTGCCT		

Table 6.22: *CRYBA4 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
CRYBB3-1F	AGTGAAGTGGCCTGGATGGA	248	62.9
CRYBB3-1R	ATGCGCGTCCCACAGAAACT		
CRYBB3-2F	GCAAGGCTCTGTAGCTCATC	255	57+ DMSO
CRYBB3-2R	TTGGCTGGGAAGATGACCCT		
CRYBB3-3F	GATGCAGAAAGCAGAGGGTA	305	64
CRYBB3-3R	CAGATCTAGAGCTCAGACTG		
CRYBB3-4F	GGTTCAAGGTCAGCAGCTCT	305	62.9

CRYBB3-4R	AATGAGCTGCCTGACCAGCT		
CRYBB3-5F	ATCTGGAGCCTCCTTGACCT	275	64
CRYBB3-5R	CTGTTTGGAGCCAGAGGCAT		
CRYBB3-6F	CAGAGTGCATGGTCAGATGC	495	64
CRYBB3-6R	CTGGGAGTCTCAGAACACTC		

Table 6.23: *CRYBB3 primers.*

Primer	**Sequence**	**Product Size bp**	**Annealing Temp. °C**
CRYBB1-1F	TGCTGCTGTCTGCAGTGATC	303	57
CRYBB1-1R	CATGGCACTGCTATCTCTGC		
CRYBB1-2F	AAACGAGCTCCAAGGAGGCT	405	57
CRYBB1-2R	TGCGGAGGAGTAAGAGGTGA		
CRYBB1-3F	ATGTGACTCCTGCACTGCTG	286	62.9+ acetamide
CRYBB1-3R	GCAGCTACTGTTGTGTGGTC		
CRYBB1-4F	TTGCCAGGGGAGAGAGAAAG	288	62.9
CRYBB1-4R	CTCAGACTTACACCTGCCCT		
CRYBB1-5F	TCATCTCTCTCGCTCCACAC	300	62.9
CRYBB1-5R	AGCCTCTGATTCTGCCTGTG		
CRYBB1-6F	TCAATGAAGGACAGGCTGGT	438	62.9
CRYBB1-6R	GAGGAAGTCACATCCCAGTA		
CRYBB1-201-1F	GTTGTGGAGGGTTGGATTTG	194	57
CRYBB1-201-1R	TCACTGCTGTTTTGCTTTGG		

Table 6.24: *CRYBB1 primers.*

Primer	**Sequence**	**Product Size bp**	**Annealing Temp. °C**
BFSP1-1F	ACTGCTGATCGGCTCTGG	600	60.4+ GCrich
BFSP1-1R	ACCCCTGCATGGTAGCAG		
BFSP1-2F	AAAGGAGAGGGCATCGTACC	238	64
BFSP1-2R	AACCTGCACTTCCACCATTC		
BFSP1-3F	TATACCCGGCCAGCTTTACC	232	57
BFSP1-3R	CCTCGGCTTACCTGATCAAA		
BFSP1-4F	TGTCCATTCCTGTTCTCATCT	250	57
BFSP1-4R	GCCCTTCCCTGGGAGTCT		
BFSP1-5F	GCACCTTCTCTGCCCTTTTC	235	57
BFSP1-5R	GCGTAACACCTCCATGAAAC		
BFSP1-6F	CAGGTTCCTTTTCCTGGTGA	400	57

BFSP1-6R	CACATGCCCACTGCCTATAA		
BFSP1-7F	CATTTTTGCTCAACATTAACGA	230	57
BFSP1-7R	GAAGAGAGCCGCTTGGTTT		
BFSP1-8AF	ATAGGCGTTTGTTTCCGAGA	499	57
BFSP1-8AR	GTAATTGGCATCCCCTGTGA		
BFSP1-8BF	AGTCCCCTGACACAAGAAGG	459	57
BFSP1-8BR	GGGTCTCCTGGACTCTTCGT		

Table 6.25: *BFSP1 primers.*

Primer	**Sequence**	**Product Size bp**	**Annealing Temp. °C**
LIM2-2F	CCCAGTTCCTCCCTTCAAGT	480	57
LIM2-2R	GTCAGGAAATGCCACCTCTC		
LIM2-3F	GAAAGGGGCAAGGCATTT	296	57
LIM2-3R	TTGGGGTTTGAGATGAGAGC		
LIM2-4F	CCACCTCCAAAATCACACC	250	57
LIM2-4R	TGTGGGACACCCTGTCATC		
LIM2-5F	GGGATACCCAGGGAGAAAGA	221	64+DMSO
LIM2-5R	CTCCTCTTCAGTGGCCTCAC		

Table 6.26: *LIM2 primers.*

Primer	**Sequence**	**Product Size bp**	**Annealing Temp. °C**
GCNT2-001-1AF	GGGCAGGAGTGAGTGGAGTA	500	57
GCNT2-001-1AR	TTTTTCATCCACATGAACACAG		
GCNT2-001-1BF	GCAAGGAATACTTGACCCAGA	491	64
GCNT2-001-1BR	CAGGTGCTCTTGGTGGACAT		
GCNT2-001-1CF	TTTCTGCCTTCGAGGTCTCA	500	57
GCNT2-001-1CR	TTCCAGTTTTCATCTTCTACTTCC		
GCNT2-001-2F	CTGGAGGAAACAGAAGATGG	399	62.9
GCNT2-001-2R	AAGCCTCTGGCTTCTCCAAT		
GCNT2-001-3F	ACCATTGGCACAGTTGTAGTT	383	57
GCNT2-001-3R	CGAAGGCAAAAGTCTCTCACA		
GCNT2-004-1AF	CAAGACTGGCAAGAGAAGCA	486	51+DMSO
GCNT2-004-1AR	TTTCTCATCCACGTGAACACA		
GCNT2-004-1BF	ATCACCTTTGCGAAGTGTCC	487	57
GCNT2-004-1BR	TCGCTTAATTGCATGGTCAG		
GCNT2-004-1CF	GACTCCAGGCTGACCTGAAC	565	57
GCNT2-004-1CR	TTGCTACAAGCAACCAAACTAA		
GCNT2-006-3AF	CACGAGGATGATTTCGGAAC	499	57
GCNT2-006-3AR	GTCGCCTTCTGATCCAGGT		
GCNT2-006-3BF	CATGGTTCGAAGCCACTATG	486	57

GCNT2-006-3BR	TGGTTTAACAGTTCTTGGTGGA		
GCNT2-006-3CF	CTTGTGGCCTCTGAAGTTCC	481	57
GCNT2-006-3CR	TTTTCCACTCTTTTCAAAGCAA		
GCNT2-010-3F	TCATTCCCTATTTAATTTGTCCTTT	250	57
GCNT2-010-3R	GGTCATCTCATTCATCACACCA		
GCNT2-201-1F	CGCTGTTGGACGGACTAAAT	360	57
GCNT2-201-1R	CCGTAAAGAAGTTTGTCCCATT		

Table 6.27: *GCNT2 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
GJA8-1F	ATATCTGCTCAGTCGTGCAG	320	62.9
GJA8-1R	GAGGGTTTTCAGAGTTGGAG		
GJA8-MID-1F	GGCAGCAAAGGCACTAAGAA	574	57
GJA8-MID-1R	GAGCGTAGGAAGGCAGTGTC		
GJA8-2F	CGTTCTGGCAACTTGGAAAG	791	57
GJA8-2R	CAACTCCATCACGTTGAGGA		
GJA8-3F	CAAGAGACACTGCCTTCCTA	834	57
GJA8-3R	GAGGACAGGAGACAGAAGAT		

Table 6.28: *GJA8 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
HSF4-3F	GCAAACGCAGCACTTTCC	270	57
HSF4-3R	GGCATGGGTGTTCACTGAC		
HSF4-4+5F	CTGGACCCAAGAGTGAGCAT	494	64
HSF4-4+5R	AGTCCCAGCACCCCTCCT		
HSF4-6F	GGAATGAGCAAAGAGGAGGA	298	64
HSF4-6R	AAGGGGCTGCCCAGAACT		
HSF4-7+8F	CAGCCTCGCCATTCTGTG	474	64
HSF4-7+8R	CGGTGAAGGAGTTTCCAGAG		
HSF4-9F	GGAGGGAGGGAAGTGCAG	249	64
HSF4-9R	GGGAGAAGGAACAGGCTCTC		
HSF4-10+11F	TCCTCCCTCAAACCTTCTCC	647	57
HSF4-10+11R	AAGCTTTGTGGGCTGGTAAG		
HSF4-12+13F	TTGATGCATCTGGGTTCCTT	430	57
HSF4-12+13R	GGCTTTGGCTCACAGAGAAC		
HSF4-14F	CAGTGATTTCCCAGCTGTCC	244	57
HSF4-14R	AACTCAGGATTTCCCTGCAA		
HSF4-15F	GCTCTCCTTCCCTGAAGAAA	354	62.9
HSF4-15R	AGGGATAGTCGGGGTAGTGG		

HSF4-T2-10F	TCAAACCTTCTCCCTTAGACCT	300	57
HSF4-T2-10R	CTCTACCCCTACAGCCATCTG		
HSF4-T2-11F	CTGGCACTGCTCAAAGAAGA	285	60.4+DMSO
HSF4-T2-11R	GAAGCTTTGTGGGCTGGTAA		

Table 6.29: *HSF4 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
CRYAA-1F	CCTTAATGCCTCCATTCTGC	395	56
CRYAA-1R	CAGAGTCCATCGCTCTCCAC		
CRYAA-2F	AGGTGACCGAAGCATCTCTG	334	64
CRYAA-2R	AAGGCATGGTGCAGGTGT		
CRYAA-3F	GCAGCTTCTCTGGCATGG	376	64
CRYAA-3R	GGGAAGCAAAGGAAGACAGA		
CRYAA-003-1F	GGTTCCTCTGGTCTCCTGAGT	298	64
CRYAA-003-1R	CTGGGGAGGACCCTGTTTAT		
CRYAA-004-1F	TGACCATAGCCAAACAACCA	495	64
CRYAA-004-1R	AAGGCATGGTGCAGGTGT		

Table 6.30: *CRYAA primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
GALK1-201-1F	GAACCGGCTGAGGTCTGG	491	64 +GCrich
GALK1-201-1R	CTTCCAACGTGGGGAACAG		
GALK1-201-2F	ACTGTGGAGGCATCAGAACC	394	57
GALK1-201-2R	CTCTACAAGCCTTCCCCACA		
GALK1-201-3F	GGTTGGTGGCTTCTGACAAT	350	57
GALK1-201-3R	ACCCCCTCCATAAGGCATAG		
GALK1-201-4F	CAGGTGGTCCCAGCTTCTAC	385	57
GALK1-201-4R	CTGGGGTGGAGTTACAATGG		
GALK1-201-5F	CTCCTGGGTGGGAGTGTG	394	64 +acetamide
GALK1-201-5R	TTGAAGGGACTGGGGAGAG		
GALK1-201-6+7+8F	CCACCCTTCACCGTCCAG	685	64 +acetamide
GALK1-201-6+7+8R	ACCCTCACCGTGTGCTGT		
GALK1-202-2F	CTGGTGCTGCCTATGGTGA	566	57
GALK1-202-2R	TGTGCACTCCCCAAGCTG		
GALK1-202-3+4F	CCTAGAGAGGCCCAGTGAAC	378	57
GALK1-202-3+4R	GCCTCCACAGTCATCAGGTT		

Table 6.31: *GALK1 primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
LENEP-1 F	GGCCTAGAACCTTCCCTTTG	390	57
LENEP-1 R	TACTCTACCCTGCCCACCTG		

Table 6.32: *LENEP primers.*

Primer	Sequence	Product Size bp	Annealing Temp. °C
CRYBG3-1aF	TGCTGAAGGGAGTGTTGAAG	540	57
CRYBG3-1aR	TTGTACCTTTTGGCCTCCAC		
CRYBG3-1bF	TACGAAGAGCCCCTTCAAGA	572	57
CRYBG3-1bR	GAAATTCCAACGCTCTCCTG		
CRYBG3-1cF	GGAGGAGGAGGCAGCAGTAT	490	57
CRYBG3-1cR	GCTTGGCCAAAACATTGACT		
CRYBG3-2+3F	TGGATACAAGACTGCTAAGAAAATG	600	57
CRYBG3-2+3R	TCGAAAAGTAGAGGAAGGAGAA		
CRYBG3-4F	TGGAAGATATCCCAGCTTAAAAA	367	57
CRYBG3-4R	AACAGCACAAAAATTAGCAATGA		
CRYBG3-5F	TCCAGGACATTGAGAAGACTG	335	57
CRYBG3-5R	CTGCAGTCTTTCTTGGGTGA		
CRYBG3-6F	TGTGATGGAGGATTTTGTTTT	300	57
CRYBG3-6R	GCATGAGCAAACATCATGG		
CRYBG3-7F	TCTTGCCCCATCTACATTGA	471	57
CRYBG3-7R	CCTCCATCTTTGCAGACTGA		
CRYBG3-8F	TCCCTGTACATTTTGTGGTCTTT	294	57
CRYBG3-8R	TGCTTTGCTTAGCACATTTTTG		
CRYBG3-9F	TGGAGATGAACCACTATGAGCTT	300	57
CRYBG3-9R	CAGGAGTCATGTCCAATTTTTG		
CRYBG3-10+11F	CCCATCAGTATGTGATATTTTGC	570	57
CRYBG3-10+11R	TGACATCATTCAACAACTCCACT		
CRYBG3-12F	TACCCCAGCCTGGATAACAG	250	57
CRYBG3-12R	GAGAAAGATTTGAGAGAAAGGGTA		
CRYBG3-13F	AAAGGTTTTTATTTGCTCGTGA	299	57
CRYBG3-13R	TGCCTGTTGAAGATAAATTTTAAGG		
CRYBG3-14F	CCTACCCCAATTCCCACAG	294	57
CRYBG3-14R	CCAACTTCACATTATCATTGGTG		
CRYBG3-15F	CCTGTGTTCAGACTGGGATG	281	57
CRYBG3-15R	TGGTTTGCAACTCTGCAACT		
CRYBG3-16F	CCCTATAAGGCAAAGTGCAGA	284	57
CRYBG3-16R	TCAGATGCTGTGGTTTATGACA		
CRYBG3-17F	AGCTTTCCTCCTCTGGAAGG	316	57
CRYBG3-17R	TGCCAGTTGATTTGACAGGA		

CRYBG3-18F	CCTAAGATAAAAGGGACCTAATTCA	309	57
CRYBG3-18R	CCCAAAATGAGGAGATTTGG		
CRYBG3-19F	TGAACAGAAGCTTGTGAAAATTG	250	57
CRYBG3-19R	TTTCCACGTCCACAAGACAA		
CRYBG3-201-1F	TTCCCCAGACACAAGGATAGA	173	57
CRYBG3-201-1R	ATGCACAAGAATGGTGGTGA		
CRYBG3-201-2F	TGATTTGGAATCATGGGTCTC	228	57
CRYBG3-201-2R	TGACATCCCAGTCTGAACACA		
CRYBG3-004-1F	CAAGCTAAATGTGAAAAGCAAAC	226	57
CRYBG3-004-1R	TTCTGACTTGTGGAACTGCAA		

Table 6.33: *CRYBG3 primers*

.**Primer**	**Sequence**	**Product Size bp**	**Annealing Temp. °C**
HSPB6-1F	GCACAGGGCGCACTATAAAT	359	57
HSPB6-1R	CAAGGCCAACGAACATTCTC		
HSPB6-001-2+3F	CTCCGGAGGCCAGAGTGA	594	66 +acetamide
HSPB6-001-2+3R	GACAGTCCTTGGCGCACT		
HSPB6-201-4F	GAGGGCGTCCTGTCCATC	349	64
HSPB6-201-4R	GACTGTCGGCTGAGGGTTAG		

Table 6.34: *HSPB6 primers.*

Primer	**Sequence**	**Size**	**Temperature °C**
BFSP2-1A F	CACATGACCCTGCATCAAAG	575	57
BFSP2-1A R	CCCTAGGTCCTCAACAGCAC		
BFSP2-1B F	AGGACCAATGCCATGAGTG	450	64
BFSP2-1B R	GCCTCAGCCTACTCACAACC		
BFSP2-2 F	TGTGCCTAACACAAGTCATGG	247	57
BFSP2-2 R	CAGGCTCCAGAACAGCATTA		
BFSP2-3 F	TCACATAATTGGCCTGTTGG	358	57
BFSP2-3 R	CGCAATGTGGAGAGTCAGAA		
BFSP2-4 F	TCTGTGAAGCCTGTGTCTGG	383	57
BFSP2-4 R	TTTCTTCGTCTGCACATTGACT		
BFSP2-5 F	GAGGCAGTGGAGTGGTGATT	323	57
BFSP2-5 R	GGGAATCCCCTGGAAACTAA		
BFSP2-6 F	TGCCACTGCTCTGCACAC	389	57
BFSP2-6 R	TTAAATACTTCTACAGTGTTTGGAAAA		
BFSP2-7 F	TTGTTCCAAAGGCCAGATTC	452	57
BFSP2-7 R	TTCCAGAAGCCTTCAAGCAT		

Table 6.35: *BFSP2 primers.*

Marker	Chr.	Start	Stop	Size	Tag
D1S485	1	105,069,360	105,069,668	157-181	FAM
D1S429	1	105,494,413	105,494,742	213-225	TET
GATA133A08	1	106,083,035	-	114-134	HEX
D1S239	1	106,640,929	106,641,320	242-258	TET
ATA42G12	1	106,877,151	-	178-196	FAM
D1S248	1	106,956,443	106,956,768	191-211	TET
D1S2726	1	110,896,304	-	280-294	NED
D1S498	1	148,114,568	-	187-209	NED
D1S498Manju	1	149,568,188	149,568,378	183-205	FAM
*Chr1-25xTG	1	150,948,915	150,949,365	223	FAM
*Chr1-22xGT	1	152,009,159	152,009,203	246	FAM
*Chr1-26xAC	1	153,111,478	153,111,529	250	FAM
D1S394	1	154,132,484	154,132,911	289	FAM
D1S1653	1	154,745,743	154,746,066	104	FAM
*Chr1-25xAC	1	155,158,129	155,158,179	247	FAM
D1S1655	1	156,199,398	-	96-120	HEX
D1S484	1	157,580,383	-	272-286	VIC
D1S1679	1	160,628,539	-	144-176	FAM
D1S1677	1	161,826,325	-	184-212	FAM
D1S2878	1	162,135,023	-	148-176	NED
D1S2673	1	163,312,611	-	148-176	NED
D1S196	1	164,335,785	-	321-337	VIC
D1S1589	1	172,527,725	-	199-220	HEX
D2S112	2	132,925,167	132,925,417	136-150	TET
D2S368	2	134,893,430	134,893,756	218-244	TET
D2S2385	2	137,478,083	137,478,431	128-150	FAM
*Chr2-24xCA	2	138,662,644	138,662,692	249	FAM
*Chr2-24xTA	2	139,470,722	139,470,770	390	FAM
*Chr2-19xCA	2	140,864,808	140,864,845	289	FAM
*Chr2-22xAC	2	141,769,786	141,769,786	212	FAM
*Chr2-22xTG	2	142,772,995	142,773,039	242	FAM
*Chr2-22xTCT	2	143,627,817	143,627,884	273	FAM
D2S132	2	144,683,977	144,684,275	189-213	FAM
D2S1399	2	147,926,579	-	266-310	FAM
*Chr3-16xTTA	3	75,301,460	75,301, 509	270	FAM
*Chr3-22xGT	3	77,535,415	77,535,458	231	FAM
D3S3681	3	79,894,319	-	121-161	VIC
*Chr3-18xTG	3	80,874,605	80,874,641	199	FAM
*Chr3-29xGT	3	81,833,944	81,834,001	250	FAM
*Chr3-21xAT	3	83,215,618	83,215,659	345	FAM
*Chr3-28xAC	3	84,520,624	84,520,680	207	FAM
*Chr3-28xTA	3	86,901,482	86,901,538	225	FAM
*Chr3-16xATT87	3	87,963,171	87,963,218	281	FAM
*Chr3-26xAT	3	90,235,425	90,235,476	208	FAM
*Chr3-22xCA	3	95,173,162	95,173,205	186	FAM
D3S2462	3	97,601,374	97,601,614	232-254	FAM
*Chr3-33xAT	3	98,564,706	98,564,772	248	FAM
*Chr3-15xTCA	3	99,228,137	99,228,183	241	FAM
*Chr3-28xTG	3	100,485,902	100,485,958	228	FAM
*Chr3-16xATT101	3	101,159,738	101,159,785	228	FAM
D3S1271	3	102,217,419	-	84-104	VIC
D3S3045	3	108,472,546	-	172-208	FAM
*Chr3-24xGT	3	109,375,535	109,375,582	241	FAM
*Chr3-22xCT	3	110,021,256	110,021,299	375	FAM
*Chr3-16xGT	3	111,566,561	111,566,592	199	FAM
*Chr3-25xGT	3	112,400,226	112,400,275	233	FAM
D3S4018	3	113,136,904	113,137,228	282-302	FAM

Marker	Chr.	Start	Stop	Size	Tag
D3S3683	3	114,744,360	114,744,531	160-174	FAM
*Chr3-16xTC	3	114,880,547	114,880,578	239	FAM
D3S3585	3	115,156,830	115,157,098	227-271	FAM
*Chr6r1-15xATA	6	88,893,581	88,893,626	209	FAM
*Chr6r1-26xTA	6	89,506,164	89,506,216	218	FAM
*Chr6-16xAC	6	90,194,544	90,194,576	227	FAM
*Chr6-26xCA	6	91,342,860	91,342,911	238	FAM
*Chr6-30xTA	6	95,957,151	95,987,211	247	FAM
*Chr6-18xAC	6	96,818,346	96,818,781	193	FAM
*Chr6-21xTG	6	97,846,372	97,846,614	181	FAM
*Chr6-17xTG	6	98,718,380	98,718,413	159	FAM
*Chr6-22xTG	6	99,826,744	99,826,788	201	FAM
*Chr7-17xATT	7	117,254,117	-	223	FAM
*Chr7-23xAT	7	117,998,458	-	397	FAM
*Chr7-29xTA	7	118,671,349	-	270	FAM
*Chr11r1-20xAC	11	31,577,606	31,577,646	215	FAM
*Chr11r1-16xTAT	11	32,822,590	32,822,638	271	FAM
*Chr11r1-23xAC	11	33,219,409	33,219,454	356	FAM
*Chr11r1-18xAC	11	34,192,263	34,192,299	247	FAM
*Chr11r2-29xGT	11	38,838,292	38,838,349	239	FAM
*Chr11r2-20xAC	11	39,808,766	39,808,806	250	FAM
*Chr11r2-18xTG	11	40,265,383	40,265,419	164	FAM
D14S68	14	87,697,387	-	318-346	NED
D14S256	14	88,281,213	88,281,346	133-156	FAM
D14S1005	14	88,447,761	88,447,972	204-230	FAM
D14S1044	14	89,140,145	89,140,529	219-261	FAM
D14S280	14	91,252,619	-	238-258	FAM
D14S302	14	92,432,197	92,432,521	178-196	FAM
D14S1434	14	94,377,875	-	212-232	HEX
D16S685	16	30,575,220	30,575,577	123	FAM
*Chr16-15xAAT	16	34,095,144	34,095,188	419	FAM
D16S746	16	34,715,504	34,715,996	285	FAM
D16S3105	16	45,349,600	45,349,998	185-195	FAM
*Chr16-29xTG	16	45,836,004	45,836,062	373	FAM
*Chr16-23xAC	16	46,392,429	46,392,475	357	FAM
*Chr19r2-16xAAT	19	41,093,388	41,093,437	231	FAM
*Chr19r2-15xAC41	19	41,435,658	41,435,687	244	FAM
*Chr19r2-26Xac	19	42,082,888	42,082,940	249	FAM
D19S220	19	43,123,394	-	267-291	FAM
*Chr19r2-15xTAT	19	44,221,282	44,221,328	186	FAM
*Chr19r2-22xGT44	19	44,659,155	44,659,198	250	FAM
D19S47	19	45,006,014	45,006,185		FAM
*Chr19r2-17xGA	19	45,150,482	45,150,515	231	FAM
*Chr19r2-34xTC	19	46,287,550	46,287,617	228	FAM
*Chr19r2-19xAG	19	47,378,235	47,378,273	210	FAM
D19S420	19	48,500,639	-	95-117	NED
D19S900	19	48,859,098	28,859,425	157-177	FAM
D19S559	19	50,022,063	-	162-198	HEX
*Chr19r2-20xCAG	19	50,965,303	50,965,364	228	FAM
*Chr19r2-15xTTA	19	51,999,196	51,99,242	218	FAM
D19S902	19	53,023,869	-	237-273	FAM
*Chr19r2-23xCA	19	53,983,566	53,983,611	182	FAM
*Chr19r2-15xAC	19	54,898,140	54,898,169	233	FAM
D19S904	19	55,468,605	55,458,843	210-224	FAM
*Chr19r2-15xGT	19	55,468,671	55,468,701	233	FAM
D19S246	19	55,647,457	55,647,733	185-229	FAM
*Chr19-19xAG	19	56,215,067	47,378,273	210	FAM

Marker	Chr.	Start	Stop	Size	Tag
*Chr19-22xGT	19	56,411,092	56,411,535	239	FAM
*Chr19-20xGT	19	56,655,961	56,656,401	243	FAM
*Chr19-15xAC	19	56,664,850	56,665,279	234	FAM
D19S571	19	57,988,784	-	287-319	NED
D22S445	22	35,895,844	35,896,190	110-130	FAM
D22S1156	22	36,711,717	36,711,872	130-162	FAM
D22S272	22	37,415,973	37,415,973	132-150	FAM
D22S428	22	38,284,898	38,285,044	147-155	FAM
D22S1197	22	38,749,278	38,749,445	168	FAM
D22S276	22	40,336,789	40,337,164	241-263	FAM
D22S1151	22	41,865,323	41,865,641	129-137	FAM

***Table 6.36*:** Microsatellite markers.

Chapter 7
Publications

Bourkiza, R., Joyce, S., Patel, H. *et al*. (2009) Cerebrotendinous xanthomatosis (CTX): an association of pulverulent cataracts and pseudo-dominant developmental delay in a family with a splice site mutation in CYP27A1- a case report. **Ophthal Genet.** 31 (2): 73-76

Joyce, S., Tee, L., Abid, A. *et al*. (2010) Locus heterogeneity and Knobloch syndrome. **Am J Med Genet. A.** 152A (11): 2880-2881