Investigations of Cataract Genetics

by

Kathryn Penelope Burdon, BSc(Hons)

Submitted in fulfilment of the requirements for the Degree of

Doctor of Philosophy

University of Tasmania

(December 2003)

Statement

This thesis contains no material accepted for a degree or diploma by the university or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief, contains no material previously published or written by another person, except where due acknowledgement is made in the text of the thesis.

KPBurdon 15/4/04

Kathryn Penelope Burdon

Authority of Access Statement

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act 1968.*

KPBurdon 15/4/04

Kathryn Penelope Burdon

Abstract

Familial cataract is a heterogeneous disorder, characterised by opacity of the lens from an early age, often leading to blindness or severe visual disability. Approximately 20% of congenital cataracts are inherited and can occur in isolation or as part of over 200 genetic disorders. Many inherited cataracts also have an age of onset during childhood and as such are not truly congenital, but have similar aetiologies. There have been fourteen genes identified for isolated familial cataract, but many of the molecular causes remain unknown. Forty-five pedigrees with various forms of familial cataract have been collected from south-eastern Australia. This collection includes 40 with autosomal inheritance, four with X-linked inheritance and one with Nance-Horan Syndrome (NHS), an X-linked syndrome involving congenital cataract, dental abnormalities and mental retardation.

Linkage studies of the large pedigree with Nance-Horan Syndrome lead to the refinement of the critical region for this disorder to 1.3 Mb through informative recombination events. Coding regions of all six characterised genes in the region were screened for mutations and shown not to be involved. A novel candidate gene was identified in the draft human genome sequence using gene prediction programs, EST information and comparison with other genomes available in public databases, and the genomic structure of the gene was determined experimentally. The gene was subsequently found to be expressed in all human tissues examined, including tissues affected by NHS such as eye and brain. Additional NHS pedigrees were ascertained and mutations of this novel gene were identified by direct sequencing in five of the six NHS pedigrees. All mutations caused severe truncation of the predicted protein

and were not identified in controls. Hence, this novel gene, although its function remains unknown, is believed to be causative for Nance-Horan Syndrome.

Analysis of known cataract genes and loci by both linkage analysis and mutation detection techniques in pedigrees with autosomal dominant congenital cataract revealed three mutations in previously known cataract genes. These mutations segregate with the cataract phenotype and are thought to be responsible for the cataract observed in these pedigrees. In addition, linkage analysis in one large pedigree has excluded linkage to most known loci, providing evidence for the existence of an additional locus for isolated congenital cataract.

In summary, the molecular defect was determined in three large pedigrees with isolated autosomal dominant congenital cataract, and a novel gene was identified and shown to cause Nance-Horan Syndrome.

Acknowledgements

I would like to thank numerous people who have assisted me throughout my PhD, both professionally and socially. Firstly, my supervisor, Dr Michèle Sale who is an inspiration to me and encourages the best in me. A/Prof David Mackey and Dr Jamie Craig for their invaluable clinical knowledge and patient ascertainment. I would particularly like to thank Jamie for his input into the Nance-Horan Syndrome work for being the driving force behind this project and coordinating several laboratories to get the results. Also Dr M. Gabriela Wirth who ascertained and examined many of the pedigrees and Dr James Elder and Dr C. Gregory Keith who created the databases from which the index cases were identified. Also involved in the Nance-Horan Syndrome work were Dr James McKay and Dr Jozef Gecz. I have learned so much from working with these gentlemen and thank them for their expertise and knowledge of the sequence databases and resources. In the laboratory I would like once again to thank James and also Liesel FitzGerald for their help with the sequencing of huge candidate genes and literally hundreds of controls. I would never have got there without you. Shelly Brown was also invaluable, extracting DNA. A big thank you to Jim Stankovich for statistical advice and also to Tim Albion for writing his genetic data management system, fondly known as GRS. Thanks also to Dr Sallyanne Fossey for helpful last minute comments on the text and to Dr Adrian West who has been (and will remain) my friend and mentor. A special thanks must go to my office and bench partner and best friend Jac Charlesworth. She has kept me laughing for the last three years and made my experience as a PhD candidate a whole lot more fun. I would also like to thank my husband Andrew who has stood by me through the stress and anxiety of the last three years and is always there when I get home.

Particular note needs to be made of the assistance of the following individuals involved in the Nance-Horan Project:

Dr Jozef Gecz and Dr James McKay for computational analysis and sequence data mining.

Shiwani Sharma and Marie Shaw for expression analysis of the *NHS* gene

Liesel FitzGerald and Dr James McKay for for assistance with sequencing candidate genes and control individuals.

Dr Isabelle Russell-Eggit for ascertainment of additional NHS pedigrees.

Table of Contents

Figures and Tables

Chapter 1 Introduction

Chapter 3 Refinement of the Critical Region for Nance-Horan Syndrome

Chapter 4 – Identification of the Nance-Horan Syndrome Gene

Chapter 5 - The Investigation of Crystallin Genes in Autosomal Dominant Paediatric Cataract Pedigrees

Chapter 6 - Investigation of Known Cataract Loci in Four Large Pedigrees

Chapter 1

Introduction

1.1 Cataract

A cataract is defined as any opacity of the lens. The lens is a transparent structure within the eye, just behind the iris, which focuses light onto the retina (Figure 1.1). The retina then sends the image to the visual cortex for processing within the brain. If the lens becomes opaque, vision can become blurred as light is scattered inappropriately by the lens. Total opacity of the lens results in blindness as light cannot reach the retina. Cataract is the most common cause of blindness worldwide (Thylefors *et al.* 1995). There are both paediatric and adult-onset forms with adult-onset being the most common. This form is caused by environmental factors contributing to oxidative stress (Ottonello *et al.* 2000) but is also believed to have a strong genetic component (Hammond *et al.* 2000; Hammond *et al.* 2001; Heiba *et al.* 1995). Congenital cataract is a heterogeneous group of disorders involving lens opacities, affecting visual function from an early age. The terms congenital and infantile are used interchangeably in the clinical setting as the clinical management of the disorders is the same (Rahi and Dezateaux 2001), and many cases of familial paediatric cataract have similar aetiologies. Throughout this thesis, the term familial will be used as a later age of onset is observed in some pedigrees. Reports of the incidence of all forms of congenital cataract vary from 1 to 6 per 10,000 live births (Francis *et al.* 2000c; Watts *et al.* 2000) with a recent study indicating a cumulative incidence of around 3.5 per 10,000 live births by age 15 years (Rahi and Dezateaux 2001).

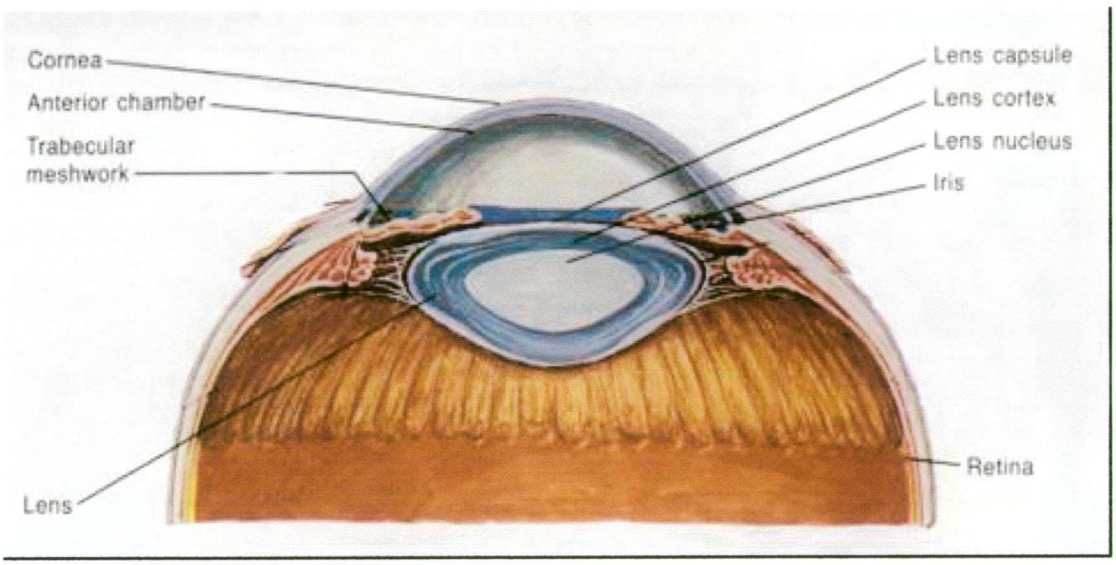


Figure 1.1: The anatomy of the anterior segment of the human eye (adapted from the website of the Ophthalmology Department, Hebrew University and Hadassah Medical Center; http://www.md.huji.ac.il/depts/ophtha/)

Familial cataract contributes greatly to blindness in children, costing the community both in expensive surgical treatments and long-term support for those remaining visually impaired.

1.2 Lens morphogenesis

The lens begins to form *in utero* at around 4 weeks. It is derived from the surface ectoderm which thickens to form the lens placode which then invaginates to form the lens pit. The lens pit closes over to form the lens vesicle (Figure 1.2). Primary fibre cells differentiate from the single layer of epithelial cells on the posterior side of the lens vesicle, filling the cavity as they elongate, forming the embryonic nucleus (Figure 1.2). Secondary fibres are formed by the division and differentiation of epithelial cells on the anterior side, and continue to be laid down throughout development and indeed throughout life, forming layers of fibre cells (Larsen 2001). The total lens present at birth is known as the nucleus and consists of the embryonic and foetal layers. The cells laid down after birth form the cortex (Larsen 2001). The points at which secondary fibres meet form Y shaped lines of optical discontinuity and are called the Y-sutures. They are present at the anterior and posterior poles of the lens (Francis *et al.* 1999). As fibre cells differentiate all organelles are broken down to ensure lens transparency, rendering a cell incapable of protein turnover (Francis *et al.* 2000c; Kannabiran and Balasubramanian 2000). The proteins present in primary fibre cells and most secondary fibre cells must remain stable for the life span of the person in order to maintain lens transparency.

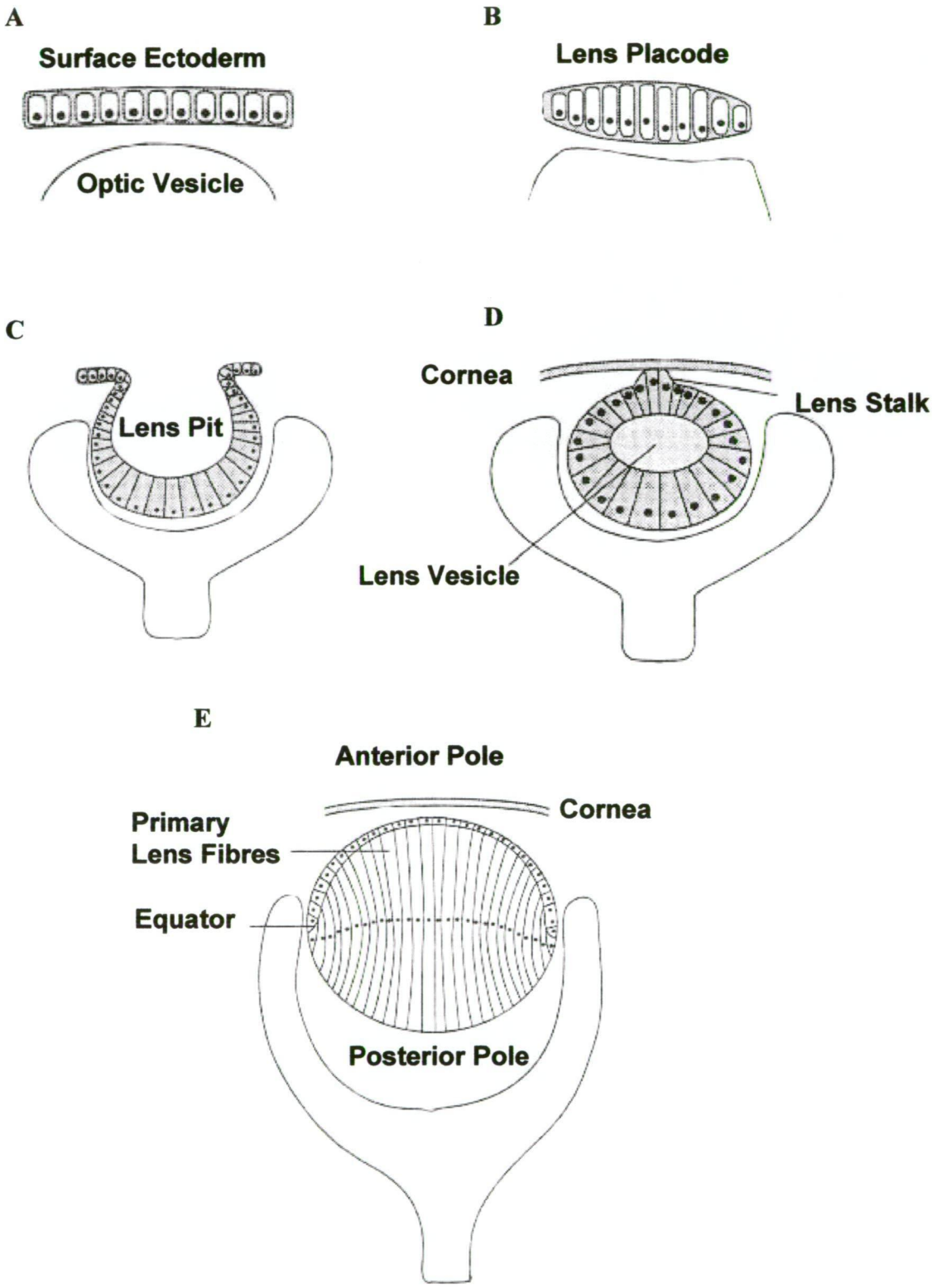


Figure 1.2. Lens Morphogenesis **A.** and **B.** Surface ectoderm thickens to form the lens placode. **C.** Placode invaginates to form the lens pit. **D.** Lens pit closes over to form the lens vesicle. **E.** Primary fibre cells differentiate from the epithelial cells, filling the cavity as they elongate. (From Francis PJ, Berry V, Moore AT, Bhattacharya S. (1999) Lens biology: development and human cataractogenesis *Trends in Genetics* 15(5):191-196.)

1.3 Molecular mechanisms of the lens

The majority of soluble lens protein is made up from the crystallin family. These proteins are resistant to degradation under physiological conditions. The β and γ-crystallins are primarily responsible for the high refractive index of the lens, required to focus light onto the retina. Their tertiary structure enable the molecules to associate into highly compact oligomers. This reduces the light scattering by internalizing hydrophobic residues (Francis *et al.* 1999). However, over time, they and other trace proteins in the lens do become unstable and unfold inappropriately. The α-crystallins help overcome this problem by acting as molecular chaperones, binding any misfolded protein. They belong to the family of small heat shock proteins which perform similar functions in other cells types under stressful conditions (Bova *et al.* 2000). α-crystallins are expressed early in fibre cell formation and are thought to be involved in the differentiation of the epithelial cell to the fibre cell (Boyle and Takemoto, 2000).

In order to help maintain protein stability an intricate network of cell to cell communication is present in the lens. Connexin proteins form hexamers in the cell membranes. These heteromeric oligomers are known as connexons. A connexon in one cell membrane interacts with that in another to form gap junctions which allow the transport of small metabolites directly between cells (Francis *et al.* 1999). Two connexin proteins, CX46 and CX50, are expressed in the lens. Major Intrinsic Protein of the lens (MIP) forms similar connections between cells to allow the movement of water across the membranes (Berry *et al.* 2000).

A lens specific cytoskeletal protein, Beaded Filament Structural Protein-2, forms part of the intermediate filament in the lens fibre cell cytoskeleton (Jakobs *et al.* 2000). On

electron microscopy, the cytoskeleton takes on a "beaded" appearance. These beads are thought to be α-crystallins (Weinreb *et al.* 2000). While the exact function of the beaded filament found only in lens is unknown, the interaction of α-crystallin with the filaments may be another mechnism by which α-crystallins protect the cell against stresses (Head *et al.* 2000).

1.4 Cataract phenotypes

Cataracts are generally classified according to their appearance and location within the lens.

1.4.1 Nuclear cataract

Nuclear cataracts are common (Figure 1.3A) and likely to be caused by a gene expressed early in development (Francis *et al.* 2000c). They affect both the embryonic and foetal nuclei. The phenotype can vary with some families presenting with a dense central opacity surrounded by fine dots, while others have dots distributed throughout the nucleus (Ionides *et al.* 1999).

1.4.2 Pulverulent cataract

A variation of the nuclear opacity is the pulverulent cataract (Figure 1.3B). This appears as a dust-like opacity with a 'pulverised' appearance. This phenotype varies more between family members than a nuclear opacity, and can even vary between eyes of the same patient. It is not restricted to the nucleus, although some families do present with a clear cortex (Ionides *et al.* 1999). The Coppock family was originally

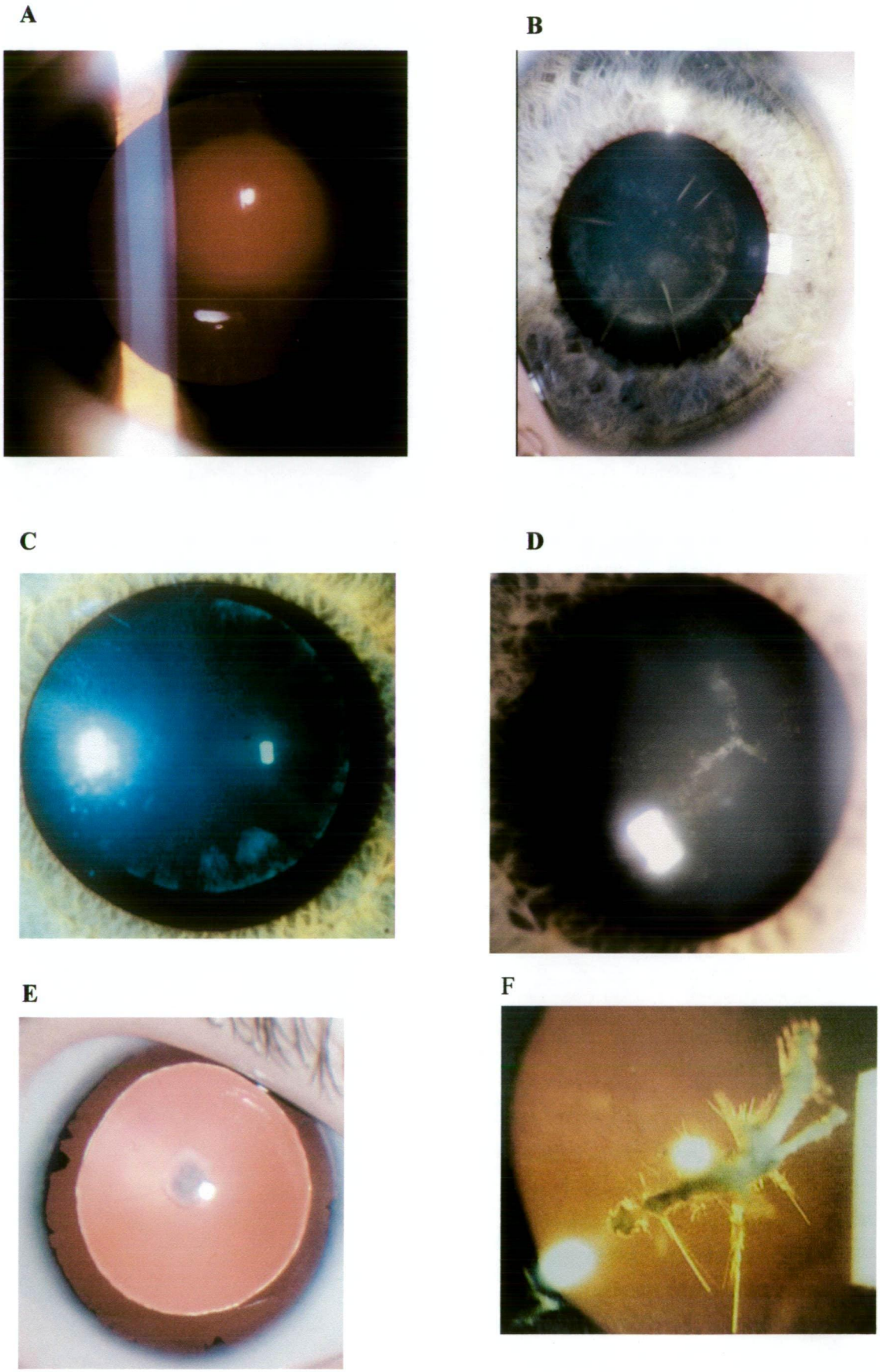


Figure 1.3. Examples of some cataract phenotypes: **A.** nuclear cataract; **B.** pulverulent cataract; **C.** lamellar cataract; **D.** sutural cataract; **E.** anterior polar cataract; **F.** coralliform cataract.

described by Nettleship and Ogilvie (1906) as possessing a central pulverulent opacity, confined to the embryonic nucleus. This phenotype is now referred to in the literature as a Coppock cataract. The Coppock-like cataract also involves the embryonic nucleus. The density of this opacity increases towards the periphery (Ionides *et al.* 1999; Francis *et al.* 2000c).

1.4.3 Lamellar cataract

Another common phenotype is the lamellar cataract (Ionides *et al.* 1999) (Figure 1.3C). It is confined to one of the concentric layers of fibre cells and represents a disruption for a short period of time during development (Francis *et al.* 2000c). These cataracts are sometimes referred to as zonular. A variation of the lamellar cataract is the sutural cataract which occurs around the posterior or anterior Y-sutures where the fibre cells join (Figure 1.3D) (Francis *et al.* 2000c).

1.4.4 Cortical cataract

Cortical cataract is rare as an autosomal dominant trait (Ionides *et al.* 1999). It is confined to the lens cortex and does not involve the nucleus. In contrast to lamellar cataract, it can affect a sector of the lens, rather than a layer (Ionides *et al.* 1999). An abnormality during the later stages of development is indicated (Francis *et al.* 2000c).

1.4.5 Polar cataract

Cataracts have also been described at both the anterior and posterior poles. Anterior polar cataracts have a minimal effect on vision (Figure 1.3E) (Ionides *et al.* 1999). Posterior polar cataracts are relatively common. They can be either stationary at the pole, or progress to have cortical extensions or lead to total cataract. This is the end

point of progression of many forms of cataract, but can also occur as a distinct phenotype (Ionides *et al.* 1999; Francis *et al.* 2000c).

1.4.6 Cerulean cataract

The Cerulean or "blue dot" cataract develops during early childhood and is not truly congenital. It consists of discrete opacities throughout the lens with a characteristic blue-white colour. Opacities are more numerous in the cortex, where they may form wedge-like shapes (Francis *et al.* 2000c).

1.4.7 Aculeiform cataract

A rare form of congenital cataract is the coralliform or aculeiform phenotype which consists of projections from the nucleus resembling coral (Figure 1.3F). This type originates from the foetal nucleus as well as post-natal regions, indicating a congenital opacity with progression (Héon *et al.* 1998; Francis *et al.* 2000c).

1.5 Treatments

1.5.1 Surgery

The only treatment for cataract is surgery to remove the opacification, followed by visual correction with an implanted lens contact lenses and/or glasses. There are several types of surgery for removing an opacification. A lensectomy involves removing the lens, anterior capsule, cortex, nucleus and posterior capsule as well as the anterior vitreous (Figure 1.1). This technique ensures that the posterior capsule will not opacify following surgery (Eckstein *et al.* 1999). The alternative to this highly specialized surgery is a lens aspiration, which removes the anterior capsule, cortex and nucleus, leaving the posterior capsule in place. Lens aspiration often requires a secondary procedure to remove the posterior capsule, if it has opacified

following surgery. The capsule is sometimes punctured during the initial surgery to prevent this post-operative complication.

1.5.2 Complications of surgery

Other post-operative complications of surgery include retinal detachment and glaucoma. Retinal detachment is not a common complication. Eckstein *et al.* (1999), report only two detached retinas, both following lensectomy surgery in a study involving 65 children. However, the follow up period in this study was only 3 years. The estimated mean interval from surgery to retinal detachment is 22 years, (determined from a retrospective study of 341 congenital cataract patients who had undergone cataract surgery) making most studies of insufficient length to gain accurate figures (Francis *et al.* 2001). Eckstein *et al.* (1999) also reported that only two eyes developed secondary glaucoma (in different patients than the retinal detachments). Other reports indicate the incidence of secondary glaucoma following cataract surgery are much higher. Ariturk *et al.* (1998) report an incidence of 28.7% after a mean follow up of 4.5 years following surgery and 60% after 5-10 years. It appears the likelihood of observing secondary glaucoma is dependent on several factors including the surgical technique, age of the patient at the time of diagnosis and at surgery and length of the follow up period (Ariturk *et al.* 1998).

1.5.3 Outcomes

Visual outcomes for children following cataract surgery vary. Up to one third of congenital cataract patients will remain legally blind, with only 30%-50% of patients achieving a visual acuity of 20/40 or better (Francis *et al.* 2001). The main cause of poor visual acuity is deprivational amblyopia, particularly when the opacity is very

dense at birth. This condition occurs when one eye fails to be stimulated during visual development and becomes 'lazy'. It is thought that early intervention of these dense opacities will prevent amblyopia, while some milder cataracts may benefit from later intervention enabling primary intraocular lens placement (Francis *et al.* 2001). Insights into the development of the lens have been gained from studying the genetics of familial cataract. As the understanding of the molecular mechanisms increase, treatments can be better developed and tailored for individual needs.

1.6 Familial forms

For the 50% of cases of congenital cataract where a cause can be determined, the most common is a genetic abnormality (Francis *et al.* 2001). Non-genetic causes include intrauterine insults such as rubella and cytomegalovirus infection (Francis *et al.* 2001). There are approximately 200 genetic disorders that involve congenital cataract, including Down Syndrome (OMIM# 190685) and metabolic disorders such as galactosaemia (OMIM# 230400) and hyperferritinaemia cataract syndrome (OMIM# 600886), but the most common form is non-syndromic isolated cataract inherited in an autosomal dominant fashion (Francis *et al.* 1999). Autosomal recessive forms have been described, although mainly in consanguineous populations (Pras *et al.* 2000; Héon *et al.* 2001; Pras *et al.* 2001; Pras *et al.* 2002), and X-linked forms are also reported (Fraccaro *et al.* 1967; Francis *et al.* 2002).

To date, fourteen known genes have been described for autosomal dominant congenital (or familial) cataract (ADCC) (Glaser *et al.* 1994; Litt *et al.* 1997; Kannabiran *et al.* 1998; Litt *et al.* 1998; Semina *et al.* 1998; Shiels *et al.* 1998; Héon *et al.* 1999; Mackay *et al.* 1999; Stephan *et al.* 1999; Berry *et al.* 2000; Conley *et al.*

2000; Jakobs *et al.* 2000; Berry *et al.* 2001; Bu *et al.* 2002; Mackay *et al.* 2002). Each family studied so far has a mutation in only one of these genes. So although ADCC is genetically heterogeneous it is a monogenic Mendelian disorder. In addition to the known genes at least five other loci have been implicated (Eiberg *et al.* 1988; Armitage *et al.* 1995; Eiberg *et al.* 1995; Berry *et al.* 1996; Ionides *et al.* 1997; Ionides *et al.* 1998; Yamada *et al.* 2000; Vanita *et al.* 2001b). To date, two genes (Pras *et al.* 2000; Pras *et al.* 2002) and two loci (Héon *et al.* 2001; Pras *et al.* 2001) have been identified in autosomal recessive congenital cataract (ARCC). X-linked congenital cataract has been mapped to Xp22.3 in some families (Francis *et al.* 2002). The identified genes and loci detected in the absence of other syndromic features are listed in Table 1.1.

1.7 Cataract genes

1.7.1 Syndromic cataract genes, including Nance-Horan Syndrome

Congenital cataract is a feature of numerous developmental syndromes, including Lowe Oculocerebralrenal Syndrome (OMIM# 309000), Walker-Warburg Syndrome (OMIM# 236670), Stickler Syndrome (OMIM# 108300) and Cataract-Microcornea Syndrome (OMIM# 116150). One gene identified for a cataract syndrome is the transcription factor musculoaponuerotic fibrosarcoma oncogene (*MAF*) (Jamieson *et al.* 2002). The promoters of lens specific proteins carry binding sites for this factor which is involved in the regulation of their transcription. In one pedigree congenital cataract segregates at a translocation breakpoint within a genomic control region of this gene. Individuals with an unbalanced translocation also have Peter's anomaly (OMIM#603807) which includes corneal opacities and iris adhesions. In a second pedigree a point mutation (R288P) of the gene was found to segregate with congenital

cataract, microcornea and iris coloboma. This mutation is in the DNA binding domain of the protein (Jamieson *et al.* 2002).

A well documented phenotype is the Cataract-Dental Syndrome (OMIM# 302350) otherwise known as Nance-Horan Syndrome (NHS). This was first described independently in 1974 by Nance *et al.* (Nance *et al.* 1974) and Horan and Billson (Horan and Billson 1974) and consists of bilateral congenital cataract and characteristic dental anomalies in affected males (Walpole *et al.* 1990). The disorder is X-linked, with carrier females presenting with milder sutural cataracts with or without dental anomalies. The phenotype of the cataract in females is comparable to that seen in isolated X-linked congenital cataract (X-linked CCT) which has been reported as a separate disorder. Both NHS and X-linked CCT have been mapped to Xp22.3 (Francis *et al.* 2002; Toutain *et al.* 2002) and it has been suggested that they may in fact be allelic. Only three large families have been described for X-linked CCT. The diagnosis in smaller pedigrees is reliant on the phenotype and apparent transmission pattern, rather than confirmation by linkage studies. Therefore some small families identified with X-linked CCT may in fact have an autosomal dominant form. Investigation of NHS and identification of the causative gene are described in Chapters 3 and 4.

1.7.2 Autosomal dominant and recessive congenital cataract genes

Any protein that is expressed in the lens is a candidate gene for congenital cataract. Many of these have already been shown to be involved in the disorder. The most highly expressed genes in the lens are the crystallins. These proteins are present in the lens in high concentrations, creating a suitable refractive index to allow the focussing

Table 1.1: Genes, mutations and mapped loci associated with congenital cataract phenotypes

Locus	Gene	Mutation	Phenotype	No. of families	Reference
1pter-p36.1			Progressive, central and zonular	1	(Eiberg *et al.* 1995)
			Posterior polar	1	(Ionides *et al.* 1997)
1q21.1	CX50	E48K	Zonular pulverulent	1	(Berry *et al.* 1999)
		P88S	Zonular pulverulent	1	(Shiels *et al.* 1998)
		I247M	Zonular pulverulent	1	(Polyakov *et al.* 2001)
2q33.3	CRYGC	T5P	Coppock-like	1	(Héon *et al.* 1999)
		235insGCGGC	Zonular pulverulent	1	(Ren *et al.* 2000)
		R168W	Lamellar	1	(Santhiya *et al.* 2002)
	CRYGD	R14C	Punctate progressive	1	(Stephan *et al.* 1999)
		P24T	Lamellar	1	(Santhiya *et al.* 2002)
		R37S	Crystalline	1	(Kmoch *et al.* 2000)
		R58H	Aculeiform	1	(Héon *et al.* 1999)
		R156X	Central nuclear	1	(Santhiya *et al.* 2002)
3p22.2-p21.3			Recessive	3	(Pras *et al.* 2001)
3q22.1	BFSP2	ΔE233	Cortical	1	(Jakobs *et al.* 2000)
		R287W	Progressive	1	(Conley *et al.* 2000)
9q13-q22			Recessive, progressive pulverulent	1	(Héon *et al.* 2001)
10q24.32	PITX3	S13N	Congenital cataract	1	(Semina *et al.* 1998)
		550delG	Congenital cataract	1	(Yang *et al.* 2002)
		655ins17pb	ASMD and cataract	1	(Semina *et al.* 1998)
		655ins17pb	Posterior Polar	1	(Finzi *et al.* 2002)
11p13	PAX6	G18W	Congenital cataract	1	(Wolf *et al.* 1998)
11q23.1	CRYAB	R120G	Desmin related myopathy and cataract	1	(Vicart *et al.* 1998)
		450delA	Posterior polar	1	(Berry *et al.* 2001)
12q13.3	MIP	E134G	Lamellar, foetal nucleus	1	(Berry *et al.* 2000)
		T138R	Progressive polymorphic	1	(Berry *et al.* 2000)
13q12.11	CX46	N63S	Zonular pulverulent	1	(Mackay *et al.* 1999)
		1137insC	Pulverulent	1	(Mackay *et al.* 1999)
		P187L	Pulverulent	1	(Rees *et al.* 2000)
15q21-q22			Central pouchlike, sutural	1	(Vanita *et al.* 2001b)
16q22.1	HSF4	A20D	Lamellar	1	(Bu *et al.* 2002)
		I87V	Lamellar	1	(Bu *et al.* 2002)
		L115P	Lamellar	1	(Bu *et al.* 2002)
		R120C	Marner's cataract	1	(Bu *et al.* 2002)
17p13			Anterior polar	1	(Berry *et al.* 1996)
17q11.2	CRYBA1	Splice site	Zonular sutural	2	(Kannabiran *et al.* 1998; Bateman *et al.* 2000)
17q24			Cerulean	1	(Armitage *et al.* 1995)
19q13.33	LIM2	F105V	Recessive, pulverulent cortical	1	(Pras *et al.* 2002)
20p12-q12			Posterior polar	1	(Yamada *et al.* 2000)
21q22.3	CRYAA	W9Stop	Recessive	1	(Pras *et al.* 2000)
		R116C	Zonular, central nuclear	1	(Litt *et al.* 1998)
22q11.23	CRYBB2	Q155X	Cerulean	1	(Litt *et al.* 1997)
		Q155X	Coppock-like	1	(Gill *et al.* 2000)
		Q155X	Sutural, punctate	1	(Vanita *et al.* 2001a)
22q12.1	CRYBB1	G220X	Pulverulent	1	(Mackay *et al.* 2002)
Xp22.3			Total in males, Sutural in females	2	(Fraccaro *et al.* 1967; Francis *et al.* 2002)

of light onto the retina (Kannabiran and Balasubramanian 2000). This family of proteins has three subtypes, known as α, β and γ. They are expressed in a sequential pattern during development, with the α-crystallins expressed early, followed by the β- and γ-crystallins as the fibre cell matures (Boyle and Takemoto 2000). Mutations of the α-crystallins have been identified in a nuclear cataract (Litt *et al.* 1998), a posterior polar phenotype (Berry *et al.* 2001), and also a recessive form (Pras *et al.* 2000). Mutations of both β- and γ-crystallins have been shown to cause a range of congenital cataract phenotypes (Litt *et al.* 1997; Kannabiran *et al.* 1998; Héon *et al.* 1999; Stephan *et al.* 1999; Gill *et al.* 2000; Ren *et al.* 2000; Mackay *et al.* 2002). Investigations of crystallin genes are described in Chapter 5.

Other proteins are also integral in maintaining lens transparency. Connexins are involved in cell homeostasis by the transport of small molecules between cells (Francis *et al.* 1999; Kannabiran and Balasubramanian 2000). *Connexin 46* (*CX46*) and *Connexin 50* (*CX50*) are expressed in human lens fibre cells and mutations in both have been linked to zonular pulverulent congenital cataract (Shiels *et al.* 1998; Berry *et al.* 1999; Mackay *et al.* 1999; Rees *et al.* 2000). Connexin gene investigations are presented in Chapter 6.

Major Intrinsic Protein of the lens (*MIP*) a member of the aquaporin family (AQ0) of transmembrane water channels, is responsible for the transport of water between fibre cells (Berry *et al.* 2000). Two mutations have been identified in this gene, one in a pedigree with polymorphic progressive cataract and the other with stable lamellar cataract (Berry *et al.* 2000).

The Beaded Filament Structural Protein-2 (*BFSP2*) is a cytoskeletal protein (Conley *et al.* 2000; Jakobs *et al.* 2000). Mutations of this gene have been associated with nuclear and sutural cataracts. BFSP2 plays an important structural role during early development and later growth of the lens. This is indicated by the differing phenotypes associated with two mutations of this gene.

The Marner cataract consists of zonular stellate and anterior polar opacities (Marner 1949). This phenotype was mapped to 16q22.1 (Eiberg *et al.* 1988; Marner *et al.* 1989). Linkage to this region was also demonstrated in a Chinese pedigree with lamellar cataract (Bu *et al.* 2002). The candidate gene Heat Shock Factor 4 (*HSF4*) was investigated in the Chinese pedigree and a causative mutation identified. Missense mutations were also detected in two other Chinese pedigrees with lamellar cataract (Bu *et al.* 2002). Following this discovery, the *HSF4* gene was investigated in a Danish kindred with Marner cataract and the causative mutation was identified (Bu *et al.* 2002).

PAX6 and *PITX3* and are also transcription factors and are involved in eye development. Mutations in both have been associated with syndromic forms of cataract such a aniridia (Wolf *et al.* 1998) and anterior segment dysgenesis (Semina *et al.* 1998), although congenital cataract has been found to occur in isolation as well.

1.8 Summary

Congenital and familial cataract are a very heterogeneous group of disorders. A total of 14 genes have been identified and five loci described for isolated autosomal dominant congenital/familial cataract. Autosomal recessive cataract has been linked

to four separate loci, for which two of these genes have been described, and X-linked cataract has been mapped to Xp22.3. As the sequencing of the human genome is completed, the gene identification process is becoming more rapid. The identification of genes involved in disease has contributed to our understanding of lens biology and development through the detailed functional investigations of cataract-causing mutations such as those in α-crystallins that affect the chaperone function (Kumar *et al.* 1999; Shroff *et al.* 2000) and in *MIP* that affect protein trafficking (Francis *et al.* 2000b). This knowledge will continue to increase as novel mutations and genes are identified. The collection of large multigenerational pedigrees is crucial to the process of gene identification and therefore biological understanding of lens development and disease processes.

1.9 Aims

While many genes for congenital cataract are well described, there remain many loci for which the gene is unknown and also families in which none of these genes or loci explain the molecular causes of their disease. The heterogeneity of the disorder has slowed progress in the understanding of congenital cataract as studies have been limited to linkage-based analyses of single large pedigrees. It is still unknown how many loci are involved in inherited paediatric cataract, and to what extent the loci contribute to later-onset cataract. Our clinical collaborators have identified and collected samples from 45 congenital or familial cataract pedigrees from south-eastern Australia. The primary objective of this study was identification of causative mutations for the cataract phenotype in these families. The family collection included the pedigree which provided the original Nance-Horan Syndrome description and several other X-linked pedigrees which have facilitated the identification and

characterisation of the gene responsible for NHS. Using linkage analyses and mutation screening approaches we investigated the known genes and loci in the pedigrees with autosomal or unknown inheritance.

Chapter 2

Materials and Methods

2.1 Patient ascertainment

2.1.1 Ethics approval

Ethics approval for this study was obtained from the Human Research Ethics Committees of the Royal Children's Hospital, Melbourne, the Royal Victorian Eye and Ear Hospital, Melbourne, and the University of Tasmania, Hobart, and adhered to the tenets of the Declaration of Helsinki.

2.1.2 Familial cases

The Royal Children's Hospital and the Royal Victorian Eye and Ear Hospital, Melbourne, Australia, have undertaken to ascertain and recruit as many pedigrees as possible with congenital or familial cataract from south-eastern Australia. These facilities have a referral base for almost all paediatric cataracts in the Australian states of Victoria and Tasmania (Wirth *et al.* 2002). The records of all paediatric cataract patients over the 25 years prior to 2000 were reviewed. The incidence of paediatric cataract was determined to be 2.2 per 10,000 live births in this population. The study involved the review of 421 cases belonging to 391 families. 342 children had no known family history of congenital cataract while 79 children from 54 families had a family history. These 54 families were contacted to take part in a study into the genetic causes of congenital cataract. Genealogical research revealed that ten of these families were related and formed four larger pedigrees. Three others either could not be contacted or declined to participate. Buccal swabs were therefore collected from a total of 45 families for DNA extraction. The families are caucasian from European

(primarily British) descent. The breakdown of inheritance patterns included 34 families with autosomal dominant inheritance, two with autosomal recessive, five were considered to be X-linked (including one with NHS) and four were of uncertain inheritance. These 45 pedigrees (shown in full in Appendix 1) form the basis of this study which aimed to identify the genes responsible for the cataract phenotype in these families.

2.1.3 Control individuals

Unaffected controls for allele frequency estimation and mutation screening were ascertained from nursing homes in Launceston, Tasmania. All individuals were ophthalmically examined and found to be free of ophthalmic disorders, including any form of cataract. Although the controls are residents of south-eastern Australia and are caucasian from European descent, they were recruited from a more restricted sampling and are elderly and disease-free, so they are not representative of the population at large. However, they are ideal for investigating whether the mutations identified are also found in cataract-free individuals, and therefore likely to be common polymorphisms without causative effects.

Allele frequencies were also obtained from the database maintained by the Centre d'Etude Polymorphisme Humain (CEPH) (http://www.cephb.fr/). These frequencies were obtained from a collection of Caucasian pedigrees from Utah, U.S.A. and France.

2.2 Genetic material

Buccal mucosa swabs were the most common sample type collected. DNA was extracted from the swabs using the PureGene DNA isolation kit (Gentra Systems).

All reagents were supplied in the kit unless otherwise indicated. The buccal mucosa of participants was scraped with a cytology brush which was then immersed in 300 µl of cell lysis solution in a screw-cap tube. Brushes were transported in this state, then stored at 4°C until extraction. Samples were incubated overnight at 55°C with 1.5 µl of 20 mg/ml Proteinase K (Sigma-Aldrich). After cooling to room temperature, 1.5 µl of 5 mg/ml RNaseA was added, the tube inverted several times to mix, and the sample incubated at 37°C for 15 minutes. The sample was again allowed to cool to room temperature before the addition of 100 µl of Protein Precipitation Solution, mixing by inversion and incubation on ice for 5 minutes. The sample was centrifuged at 13,000 rpm for 3 minutes and the supernatant added to a clean tube containing 200µl of isopropanol (Sigma-Aldrich) and 0.5 µl of 20 mg/ml glycogen (Sigma-Aldrich). The tube was again mixed by inversion and incubated at room temperature for 5 minutes before centrifuging at 13,000 rpm for 5 minutes. The supernatant layer was then discarded and 300 µl of 70% molecular grade ethanol (Sigma-Aldrich) added to the pellet. The sample was again centrifuged at 13,000 rpm for 1 minute, the supernatant discarded and the pellet allowed to air dry. Once completely dry, the pellets were resuspended in 20 µl of DNA Hydration Solution with heating to 65°C for 1 hour. All DNA samples were stored at 4°C.

Whole blood was collected in 10 mL EDTA tubes from consenting individuals in pedigrees ctas16 (III:5, III:8, III:9, IV:3, IV:5, IV:6, IV:7, V:4, V:5, V:7 and VI:1), ctas17 (IV:7, V:3, V:8, VI:1), ctas35 (I:1), ceeh42 (III:2, IV:1) and crch2551 (IV:2, V:3, V:4, V:5, V:7). DNA was extracted by technician Michele Brown using the Nucleon BACC3 kit (Amersham Pharmacia Biotech). Four times the volume of the sample of Reagent A was added and mixed for 4 minutes at room temperature. The

samples were centrifuged at 1300g for 5 minutes and the supernatant discarded. The pellet was resuspended in 2 mL of Reagent B and transferred to a clean tube before adding 15 µl of 50 µg/ml RNase and incubating at 37°C for 30 minutes. Protein was precipitated by the addition of 500 µl of sodium perchlorate solution and the samples mixed by inversion. Two ml of chloroform was added and mixed by inversion before the addition of 300 µl of Nucleon resin without remixing the phases. The samples were centrifuged at 1300g for 3 minutes. The upper phase was transferred to a clean tube and 2 volumes of cold absolute ethanol added. The sample was mixed by inversion until the precipitate appeared, followed by centrifuging at 4000g for 5 minutes. The supernatant was discarded. The pellet was washed with cold 70% ethanol and centrifuged again at 4000g for 5 minutes. The pellets were air dried and resuspended in 100 µl of TE buffer.

2.3 General PCR conditions

All PCR reactions were conducted using Corbett Research 960C cooled thermal cyclers, unless otherwise noted. 10x PCR buffer was prepared, containing 100 mM Tris HCl, 500 mM KCl, and variable concentrations of $MgCl_2$, ranging from 10-35 mM. PCR reactions were carried out in a total volume of either 10 or 30 µl and contained between 1.0 and 3.5 mM final Mg^{2+} concentration, 0.2 mM dNTP (Promega), 0.8 µM of each primer and 0.005 U/µl *Taq* DNA Polymerase (Qiagen). Reactions underwent 30 cycles of 94°C for 30 seconds, primer pair specific annealing temperature for 30 seconds and 72°C for 30 seconds followed by 72°C for 10 minutes. Annealing temperatures and optimal Mg^{2+} concentrations, listed in Appendix 3, were determined empirically for each primer pair by cycling of a range of temperatures and Mg^{2+} concentrations in a PC-690G gradient thermal cycler

(Corbett Research) and assessment of the optimal conditions by visualisation on agarose gel containing 0.3 ng/ml ethidium bromide.

2.4 Primer Extension Preamplification (PEP)

This PCR-based technique was used as a method of whole genome amplification to provide sufficient DNA for multiple PCR-based analyses from buccal mucosa swabs. Each reaction contained 50 ng genomic DNA, 2.0 mM Mg^{2+}, 0.2 mM dNTP (Promega), 20 pmol/µl polyN 15mer primer (Operon technologies) and 5U *Taq* DNA polymerase (Qiagen). Reactions underwent 50 cycles of 94°C for 1 minute, 37°C for 2 minutes and 55°C for 4 minutes followed by a final extension of 72°C for 10 minutes. Products were electrophoresed on 1% agarose gel containing 0.3 ng/ml ethidium bromide to confirm the presence of high molecular weight product. Samples were diluted 1:8 with distilled water for use in PCR.

2.5 Fluorescent genotyping

The forward PCR primer for each marker was labelled at the 5' end with either 6-FAM, TET, HEX or NED (Sigma-Aldrich). PCR reactions were carried out in 10 µl reaction volumes. Fragments labelled with 6-FAM or TET underwent 25 amplification cycles, while those with HEX or NED underwent 35 cycles. One µl of each PCR reaction was pooled into an 8 µl final volume, with between 3 and 7 markers per pool, such that markers that had a similar expected size in the same pool were labelled with different colours. Pools were electrophoresed using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems) with POP4 polymer (Applied Biosystems). Data were collected automatically and analysed with the Genescan® and Genotyper® software (Applied Biosystems).

2.6 Single-Stranded Conformational Polymorphism Analysis (SSCP)

PCR used for SSCP analysis of the crystallin genes in Chapter 5 primers are listed in Appendix 2. Both forward and reverse primers were end-labeled with γ^{32}P-ATP in reactions containing 0.2 U/μl T4 polynucleotide kinase (New England Biolabs), 3 μM of primer, 1X PNK Buffer (New England Biolabs) and 1.6 μCi/μl of γ^{32}P-ATP (Geneworks). Reactions were incubated at 37°C for 30 minutes and terminated at 70°C for 10 minutes. Each exon was amplified by PCR in 10 μl reaction volumes as described above, except that each reaction contained 0.7 μM of each unlabeled primer and 0.11 μM of each labeled primer.

Five μl of the PCR products were added to 35 μl of SSCP Stop Solution (95% deionised formamide, 10 mM NaOH, 0.25% Bromophenol blue, 0.25% Xylene Cyanol, all reagents from Sigma-Aldrich), denatured at 95°C for 2 minutes and snap cooled on ice. 4 μl was loaded onto a 0.4mm, 0.5X Mutation Detection Enhancement (MDE) Gel (Edwards Instrument Co.) and electrophoresed in 0.6X TBE buffer at 8W for 16 hours at room temperature. 1X TBE buffer contained 0.1 M Trizma base, 0.1 M Boric acid and 2 mM EDTA (all from Sigma-Aldrich). The gel was transferred to 3MM paper (Whatman) and exposed to Kodak Biomax MR film (Amersham Pharmacia Biotech) in the presence of Cronex Lightning Plus intensifying screens (Dupont) at -80°C for 24 hours, before being developed by hand with D-19 developer (Kodak) and fixed with Hypam Rapid Paper and Film Fixer (Ilford). Films were visually examined for mobility shifts.

2.7 DNA Sequencing

2.7.1 Protocol for Beckman CEQ2000

The sequencing of candidate genes for NHS described in Chapter 3 was carried out using a CEQ2000 automated DNA sequencer (Beckman Coulter). PCR products were sequenced with forward and reverse primers used to generate the PCR product (Appendix 2). Fragments were PCR amplified in 30 µl volumes, purified with Ultra Clean PCR clean up kit spin columns (MoBio) and cycle sequenced with QuickStart Ready Reaction Mix (Beckman Coulter) in 10 µl reactions consisting of 4 µl Ready Reaction Mix, 3.2 pmol of primer and 30-90 ng of PCR product. Samples underwent 30 cycles of 95°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. Precipitation was carried out with 1 µl of 20 mg/ml glycogen, 2 µl of 100 mM EDTA, 2 µl of 3M sodium acetate (pH 5.2) and 50 µl of 100% molecular grade ethanol (all from Sigma-Aldrich). Samples were immediately centrifuged at 13,000 rpm for 20 minutes and the supernatant discarded. The pellet was washed with 200 µl of 70% ethanol and then centrifuged at 13,000 rpm for 5 minutes. The supernatant was again discarded and the wash step repeated. The pellets were vacuum dried and resuspended in Sample Loading Solution (Beckman Coulter) before electrophoresis on the CEQ2000.

2.7.2 Protocol for ABI PRISM 310 Genetic Analyzer

The sequencing of the *NHS* gene and crystallin and connexin genes in Chapters 4, 5 and 6 was carried out on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Fragments were prepared as above but cycled with Big Dye Terminator Ready Reaction Mix (Applied Biosystems) over 25 cycles. Products were precipitated with 2 µl of 3M sodium acetate (pH 4.2) and 50 µl of 100% molecular grade ethanol (Sigma-Aldrich). Following a 15 minute incubation at room temperature, samples

were centrifuged at 13,000 rpm for 20 minutes. The supernatant was discarded and the pellet washed with 200 µl of 70% molecular grade ethanol. Samples were then centrifuged at 13,000 rpm for 5 minutes and the supernatant discarded. The pellet was dried by a 1 minute incubation at 95°C with the lid open and was resuspended in 15 µl of Template Suppression Reagent (Applied Biosystems). Samples were electrophoresed on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems) with POP6 polymer (Applied Biosystems). Data were collected automatically and analysed using Sequence Analysis® (Applied Biosystems) and Sequencher® (GeneCodes) software.

2.8 Denaturing High Performance Liquid Chromatography (dHPLC)

Each fragment to be analysed by dHPLC was amplified in a 30 µl PCR reaction with the primers listed in Appendix 2. When searching for recessive mutations, known wild type DNA must be added to the sample to allow the formation of heteroduplexes between mutant and wild type strands. As the mutation causing this dominant disease should be heterozygous in affected individuals, no additional wild type DNA was added to reactions as heteroduplexes will form naturally. The melt temperature of the fragment was calculated at the Stanford dHPLC Melt Program Website (http://insertion.stanford.edu/melt.html). The best temperature for visualisation of the mutation was found experimentally using positive controls identified by sequencing, starting with the calculated temperature and moving up and down in 1°C increments until the clearest difference between wild type and mutant samples was seen. Samples were injected onto a Helix dHPLC column packed with 1000 Å alkylated silica (Varian). The column had an internal diameter of 3.0 mm and was 75 mm long. The acetonitrile concentration was increased from 45% to 68% by increasing the flow

rate of Buffer B (Varian) while decreasing the flow rate of Buffer A (Varian). Data were collected automatically and analysed using the Star Reviewer software (Varian).

2.9 Reverse Transcription Polymerase Chain Reaction (RT-PCR)

RNA from the cortex and cerebellum of an adult human male obtained post-mortem was reverse transcribed in 10 µl reactions containing 1X first strand buffer (Invitrogen), 10 mM dithiothreitol (Invitrogen), 0.5 mM dNTP (Promega), 1 µM gene specific reverse primer and SuperscriptIII RNase H^- Reverse Transcriptase (Invitrogen). Reactions were incubated at 37°C for 1 hour. The resulting cDNA was then diluted by ½ for use in 30 µl PCR reactions as described in section 2.4 except 35 cycles with an annealing temperature of 58°C were used. Where non-specific products or smears were obtained, the reactions were repeated with higher annealing temperatures. Products were purified with Ultra Clean PCR clean up kit spin columns (MoBio) or excised from the gel and extracted with a Qiaquick Gel Extraction Kit (Qiagen) before sequencing as described in Chapter 2.7.2.

2.10 5' Random Amplification of cDNA Ends (5'RACE)

This technique was carried out using the GeneRacer Ready cDNA kit (Invitrogen). The RNA used for the RT-PCR experiments was known to be degraded and was thought unlikely to contain full-length transcripts which are necessary for 5'RACE. Pre-prepared cDNA from adult human brain was supplied. Prior to purchase the RNA had been treated to ligate an oligo of known sequence (5'-CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA-3') to the 5' end of all full length mRNA, then reverse transcribed to produce cDNA. This cDNA was used as the template in PCR reactions using the supplied primer (5'-

GCACGAGGACACUGACAUGGACUGA-3') targeted to the ligated oligo and gene-specific primers targeted to predicted exons 3, 4, 5, and 6 of the novel gene (Appendix 2). Nested reactions were also carried out using the nested primer supplied (5'-GGACACTGACATGGACTGAAGGAGTA-3'). The supplied protocol was followed for PCR. Platinum HiFi *Taq* Polymerase (Invitrogen) was used, as advised. The product obtained was cleaned with Ultra Clean PCR clean up kit spin columns (MoBio) and sequenced directly as described above for other PCR products.

2.11 Computer based methods

2.11.1 Linkage simiulations and analysis

Linkage simulations were carried out by SLINK (Ott 1989; Weeks *et al.* 1990) and are described in sections 3.3.2 and 5.2.2. Two-point linkage analysis was carried out with the FASTLINK package (Lathrop *et al.* 1984, with modification by Cottingham *et al.* 1993) and is described in detail in sections 3.3.3 and 5.2.4. Multipoint linkage analysis and haplotype estimation were carried out with X-linked Genehunter Plus (Kong and Cox 1997) and is described in section 3.3.3.

2.11.2 Database mining

Microsatellites for linkage analysis were obtained from the Genome Database (http://www.gdb.org). Map locations were determined from the location database (http://cedar.genetics.soton.ac.uk/public_html/ldb.html) and the physical locations from UCSC genome browser (http://genome.ucsc.edu).

Database tools used for the identification of the novel *NHS* gene are indicated in Chapter 4.

Chapter 3

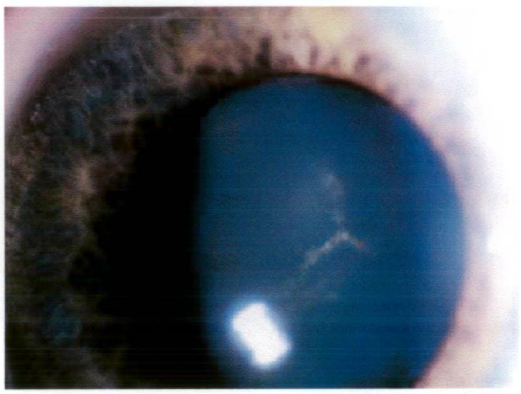
Refinement of the Critical Region for Nance-Horan Syndrome

3.1 Background

3.1.1 Nance-Horan Syndrome

Nance-Horan Syndrome (NHS [OMIM #302350]) was first described in 1974 as an X-linked syndrome involving congenital cataract and dental anomalies (Horan and Billson 1974; Nance *et al.* 1974). Ophthalmological findings in affected males include bilateral severe congenital cataract involving the foetal nucleus and posterior Y-suture with variable zonular extensions into the posterior cortex and requiring surgery. Microcornea, nystagmus and microphthalmia have also been reported in some pedigrees (Lewis *et al.* 1990; Stambolian *et al.* 1990; Walpole *et al.* 1990). Dental abnormalities include screwdriver blade-shaped incisors, supernumerary maxillary incisors (meisiodens) and diastema (Walpole *et al.* 1990). Several families were reported to have lateral brachymetacarpalia (Lewis *et al.* 1990) and some affected males display mental retardation (Toutain *et al.* 1997a). Characteristic dysmorphic facial features, which may be subtle, include large anteverted pinnae, long narrow face and prominent nose and nasal bridge (Lewis *et al.* 1990; Walpole *et al.* 1990). The disorder appears to be inherited in a codominant fashion, with heterozygous females often expressing some level of phenotype but rarely as severely as males. Features in females include posterior Y-sutural cataracts with little or no loss of vision, and characteristic dental abnormalities (Walpole *et al.* 1990; Zhu *et al.* 1990). Typical characteristics are depicted in Figure 3.1.

A



B

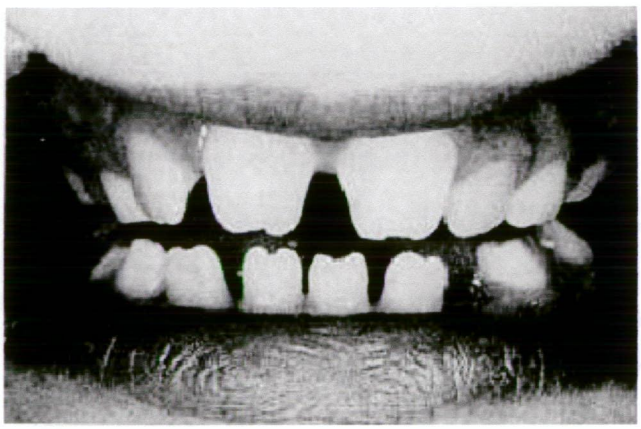


Figure 3.1. Clinical features of Nance-Horan Syndrome **A.** Examples of sutural cataract observed in females. **B.** Characteristic dental anomalies including screwdriver blade shaped notched incisors and diastema. (B. reproduced with permission of the publisher from Horan MB, Billson FA (1974) X-linked cataract and hutchinsonian teeth *Aust Paediat J* 10:98-102)

Initially, NHS was mapped to a region of approximately 25cM on Xp21.1-p22.3 (Lewis *et al.* 1990; Zhu *et al.* 1990) which was substantially refined to 16cM between DXS237 and DXS451 (Stambolian *et al.* 1990). Progress on the genetics of NHS then slowed with only three more studies published between 1990 and 2000 (Bergen *et al.* 1994; Toutain *et al.* 1997b; Zhu *et al.* 1998), when our study commenced. These studies ultimately reduced the critical region to 3.5cM of the genetic map, or 5.8Mb of the physical map, between STR markers DXS1053 and DXS443.

3.1.2 X-linked cataract

X-linked total congenital cataract (X-linked CCT [OMIM#302200]) without dental anomalies has also been reported, although its existence as an entity separate from NHS has been debated. Several families with varying phenotypes have been reported. Fraccaro *et al.* (1967) described a pedigree with cataract that appeared to be X-linked. Males presented with bilateral nuclear cataract while females had posterior sutural cataract with little effect on vision. This family demonstrated possible linkage to the Xg blood group on the X chromosome (Fraccaro *et al.* 1967).

In 1969 Krill *et al.* reported a family with an apparently X-linked progressive cataract. Males presented initially with sutural opacities and punctate opacities of the nucleus and cortex that progressed to almost total cataract in older men. The females had a non-progressive sutural cataract. Intrafamilial variation between females was described and postulated to be caused by random X inactivation. Only one affected male had offspring and his one son was said to be unaffected although he was not examined by the authors. Because of the rarity of affected males with offspring in the

described pedigrees, it is difficult to definitively classify these pedigrees as X-linked in the absence of linkage studies.

Recently, a pedigree was reported in which the males presented with severe cataract that required removal in the first few months of life. Some affected males also displayed cardiac abnormalities. Affected females had a milder central nuclear opacity not requiring treatment until the sixth decade (Francis *et al.* 2002). Again, none of the affected males had offspring, but linkage to the X chromosome was suspected based on the differences in phenotype between males and females, which are not reported in families with ADCC. Detailed linkage studies revealed significant linkage to the X chromosome, with the critical region being defined as less than 2.5cM between markers DXS9902 and DXS999 (Francis *et al.* 2002). This is within the 3.5cM NHS region between DXS1053 and DXS443 reported by Zhu *et al.* (Zhu *et al.* 1998). As these two disorders are now known to map to the same region, it is not certain whether isolated X-linked cataract is a separate disorder or allelic with NHS (Francis *et al.* 2002). Both disorders appear to be phenotypically variable between families and may in fact represent a spectrum with the same underlying molecular cause.

3.2 Aims

We aimed to identify the gene for NHS and assess its role in isolated X-linked cataract. We first used a linkage approach to confirm the linkage to Xp22.13 and investigated recombination events to refine the region. We then screened all characterised genes in the critical region for mutations in NHS patients.

3.3 Methods

3.3.1 Pedigrees

Cataract pedigree ascertainment and recruitment is described in Chapter 2.1.2.

3.3.2 Simulations

Simulations using the SLINK program (Ott 1989; Weeks *et al.* 1990) using a single marker with four equally frequent alleles over 200 replicates were carried out on all five pedigrees believed to be X-linked to determine the power of the study for linkage analysis.

3.3.3 Linkage analysis

Seven STR markers across 12cM of the X chromosome, encompassing the region previously shown to be linked to NHS, were typed in these five pedigrees. The map order was tel-DXS1224-2.8cM-DXS1053-2.3cM-DXS1195-0.1cM-DXS418-0.2cM-DXS999-2.8cM-DXS365-3.2cM-DXS989-cen. Primer sequences for the PCR amplification of all markers are available from the Genome Database (http://www.gdb.org). Allele frequencies were estimated from 22 unrelated unaffected individuals from south-eastern Australia (frequencies are shown in Appendix 3). The disease gene frequency was set to 0.0001. A codominant model was used with penetrance set at 1.0 in hemizygous males and homozygous females and 0.5 in heterozygous females, to indicate incomplete penetrance as the exact level of penetrance is unknown. No other linkage studies for Nance-Horan Syndrome specify the model used. A report by Francis *et al* (2002) investigating isolated X-linked cataract used both a recessive model and a dominant model. The exact penetrance in females is unknown in the case of NHS, but a recessive model seems unlikely given the numbers of females with a detectable phenotype. A conservative

value of 0.5 for heterozygous females was chosen. Two-point LOD scores were generated for each marker in MLINK (Lathrop *et al.* 1984, with modification by Cottingham *et al.* 1993) using crch2551 alone as well as all the families combined. Multipoint parametric LOD scores and haplotypes were generated with X-linked GeneHunter Plus (Kong and Cox 1997) using the same model as for two-point analysis. Map distances were obtained from the Genetic Location Database (GLD, http://cedar.genetics.soton.ac.uk/public_html/ldb.html). When genetic distances are small they may be inaccurate due to low numbers of observed recombinations but the GLD maps integrate both physical and genetic information (Collins *et al.* 1996a; Collins *et al.* 1996b) and hence should be quite accurate.

3.3.4 Identification of novel STR markers

The sequences of clones making up contig NT_011586 were downloaded from Genbank (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide) and were searched for dinucleotide repeats of all combinations of greater than 15 repetitive units. The distance from published markers was determined in physical distance from the genomic sequence. Over such small distances, the genetic distances become innacurate and physical mapping is stronger. The sequence surrounding each dinucleotide repeat was then masked using Repeat Masker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker; A.F.A. Smit and P. Green, unpublished) to ensure it was not part of a larger complex repeat that would interfere with PCR amplification. Markers were named for the repetitive element followed by the clone name. Primers were designed for all remaining markers and sequences are given in Appendix 2. Allele frequencies for informative markers were estimated from 22 unrelated unaffected individuals and are shown in Appendix 3.

3.3.5 Sequencing of candidate genes

Coding exons from all genes in the critical region were directly sequenced in two affected females (V:3 and V:9), one affected male (V:5) and one unaffected female (III:6) from pedigree crch2551. Primer sequences used for the amplification of each exon are given in Appendix 2 and methods are given in Chapter 2.4 and 2.7.1.

3.4 Results and Discussion

3.4.1 Pedigrees

Investigation of the pedigrees with congenital cataract collected in south-eastern Australia revealed two families with NHS (crch25 and crch51). Crch51 was the pedigree originally described by Horan and Billson (1974). Genealogical research revealed that the two families were related and the combined pedigree was named crch2551. A total of six generations, with DNA available from 11 affected individuals and 9 unaffected individuals across four generations, were identified (Figure 3.2).

In addition to this family, four pedigrees (crch02, crch05, crch20 and crch24) appeared to have isolated X-linked cataract on the basis of the phenotype and a segregation pattern consistent with an X-linked trait (Figure 3.3). Carrier females presented with variable mild sutural cataract while the affected males had dense bilateral nuclear cataracts. The possibility of NHS was raised in pedigree crch02 but the dental phenotype was not typical and no mental retardation was noted (J.E. Craig, pers. comm.).

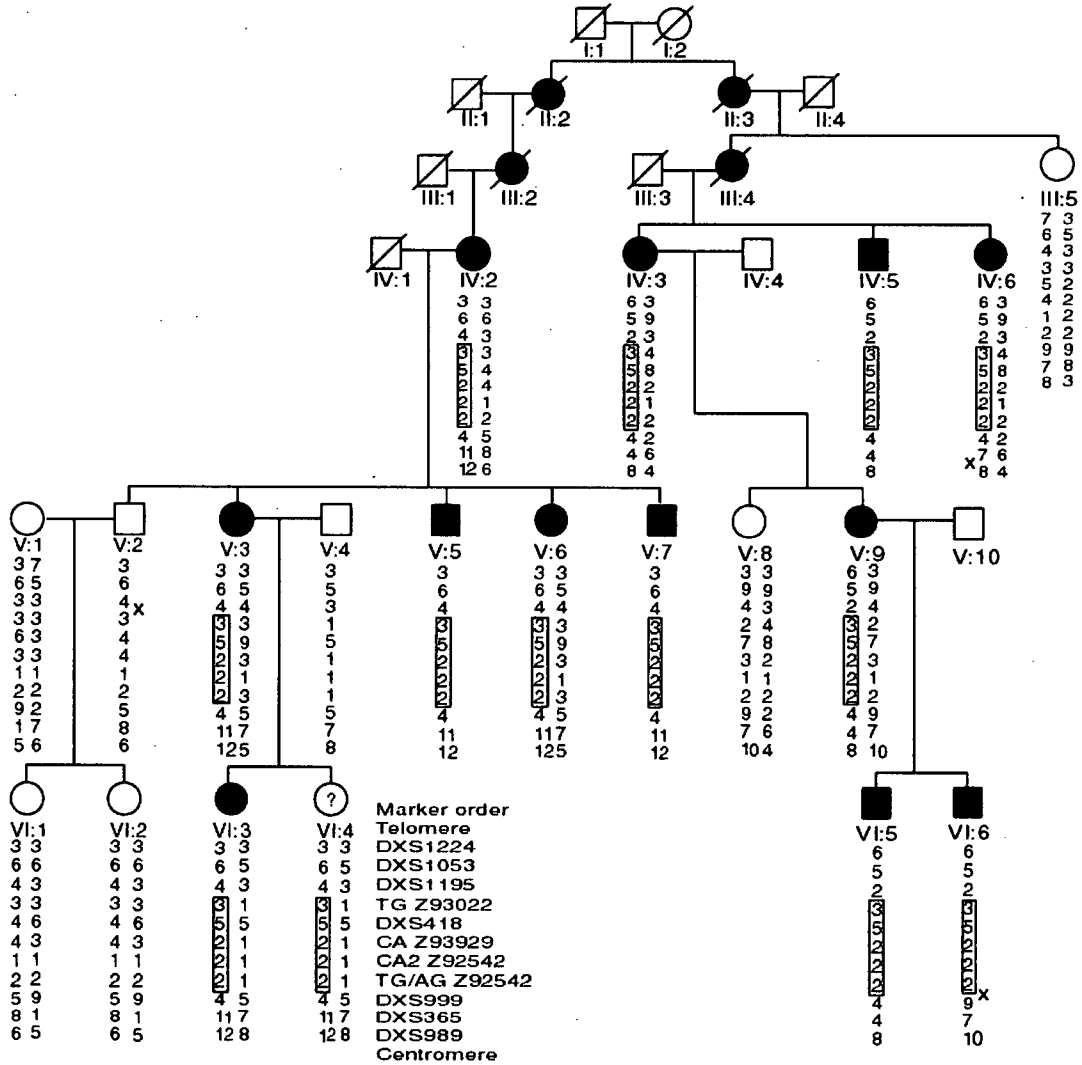


Figure 3.2. Pedigree crch2551. Shaded symbols indicate presence of ophthalmologist-confirmed Nance-Horan Syndrome. Individual VI:4 is a carrier but has not been examined. Shared haplotypes amongst affected individuals are boxed. Recombinations are marked with an 'x'.

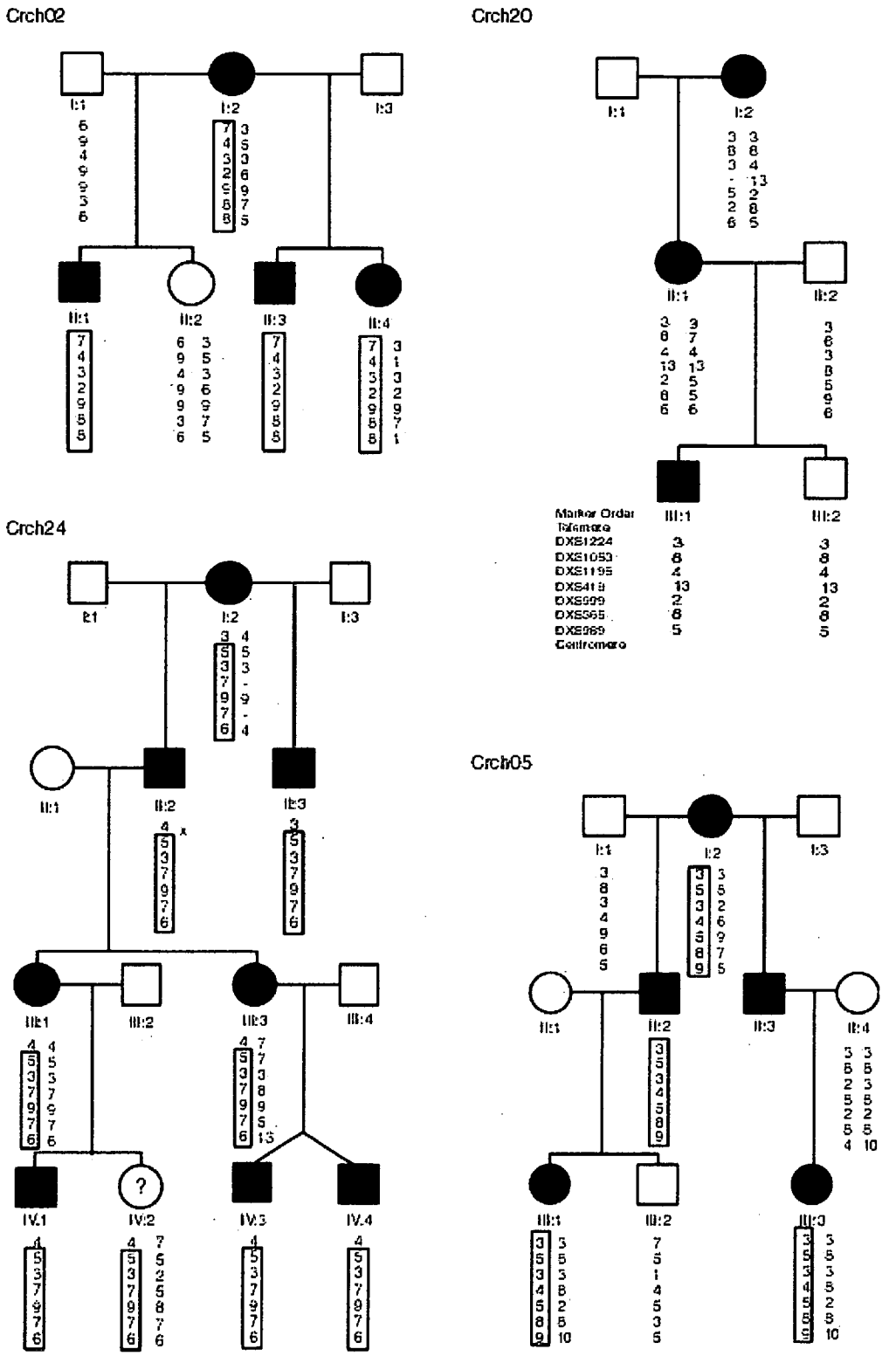


Figure 3.3. Pedigrees diagnosed with isolated X-linked cataract. Shaded symbols represent ophthalmologist confirmed cataract. Shared haplotypes amongst affected individuals within each pedigree are boxed. Recombinations are marked with an 'x'.

3.4.2 Simulations

Simulations using SLINK in these five pedigrees indicated a maximum LOD score of 4.68 using a single marker with four equally frequent alleles over 200 replicates. The average LOD score simulated in this group of pedigrees was 2.3 over 200 replicates for the study. For the X chromosome a LOD score of 2.0 is considered significant evidence for linkage. Whereas a LOD score of 3.0 is required for genome-wide significance, an X-linked disorder is already known to map to the X chromosome, so the threshold can be reduced to the same level as required for the confirmation of linkage to a single autosome (Lander and Kruglyak 1995). The analysis of crch2551 alone revealed a simulated maximum LOD score of 2.1 with an average of 1.1, indicating that this pedigree theoretically may have sufficient power to demonstrate linkage in its own right (Table 3.1).

Table 3.1: Simulated maximum and average LOD scores in X-linked pedigrees at θ=0.01.

Pedigree	Maximum LOD	Average LOD
CRCH2551	2.1	1.1
CRCH02	0.8	0.4
CRCH05	0.4	0.2
CRCH20	0.3	0.2
CRCH24	1.1	0.6
All Pedigrees	4.6	2.3

3.4.3 Linkage analysis

3.4.3.1 Pedigree crch2551

Two-point LOD scores were generated at a range of recombination fractions for all seven published markers by MLINK (Lathrop *et al.* 1984, with modification by Cottingham *et al.* 1993) for pedigree crch2551 (Table 3.2). The maximum LOD score obtained was 2.39, indicating significant linkage at this marker. This score is larger than the maximum simulated score, most likely due to the unlikely situation of four

equally frequent alleles employed in the simulations. Crch2551 displays negative LOD scores for DXS1053, DXS1195, DXS365 and DXS989, indicating no evidence of linkage to these markers (at θ=0).

Inspection of the haplotypes showed that all affected individuals share alleles at DXS418 and DXS999, the two markers that gave the highest two-point LOD scores (Figure 3.2). This suggested a common haplotype in this small region. Additional markers were required to confirm the haplotype and better define the recombinations observed in individuals VI:6 and V:2. No additional STR markers were available from the current databases, so novel dinucleotide repeats were identified from the draft human genome sequence (Appendix 2). Four of the eight typed markers were informative in pedigree crch2551. The LOD scores are presented in Table 3.2 and the haplotypes in Figure 3.2.

Table 3.2: Two-point LOD scores for pedigree crch2551.

Theta	0.00	0.001	0.01	0.02	0.03	0.04	0.05	0.10	0.20	0.30	0.40
DXS1224	-∞	-0.26	0.71	0.97	1.10	1.19	1.24	1.32	1.18	0.86	0.46
DXS1053	-∞	-1.33	-0.36	-0.08	0.07	0.17	0.24	0.42	0.45	0.34	0.18
DXS1195	-∞	-2.60	-0.66	-0.10	0.20	0.41	0.55	0.92	1.01	0.81	0.46
TG Z93022	1.19	1.18	1.16	1.14	1.11	1.09	1.07	0.95	0.72	0.50	0.27
DXS418	2.39	2.38	2.34	2.30	2.25	2.20	2.16	1.91	1.42	0.91	0.41
CA Z93929	1.49	1.48	1.45	1.42	1.38	1.34	1.31	1.13	0.76	0.40	0.12
CA2 Z92542	1.73	1.72	1.69	1.66	1.62	1.59	1.55	1.37	0.98	0.61	0.27
TG/AG Z92542	-∞	-1.91	-0.93	-0.66	-0.50	-0.40	-0.33	-0.13	0.00	0.02	0.02
DXS999	-∞	0.06	1.02	1.27	1.40	1.48	1.52	1.57	1.32	0.92	0.47
DXS365	-∞	-5.57	-3.12	-2.40	-1.94	-1.61	-1.36	-0.64	-0.10	0.05	0.05
DXS989	-∞	-3.62	-1.66	-1.09	-0.78	-0.56	-0.40	0.01	0.20	0.14	0.03

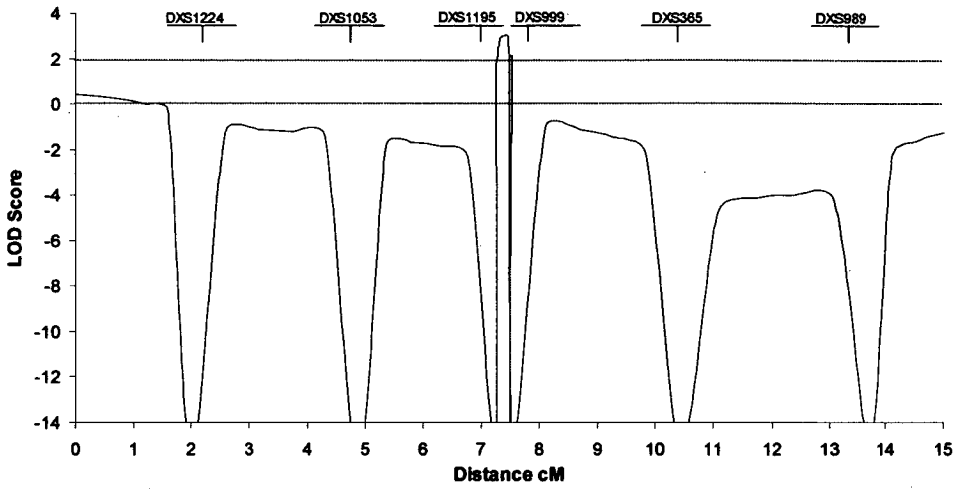
*novel markers are named for the repetitive element followed by the clone name in which it was identified

A shared haplotype was detected in the region of interest extending from TG Z93022 to DXS999 (Figure 3.2). A recombination was observed in individual VI:6 between CA2 Z32542 and DXS999, defining the proximal marker as DXS999. The distal

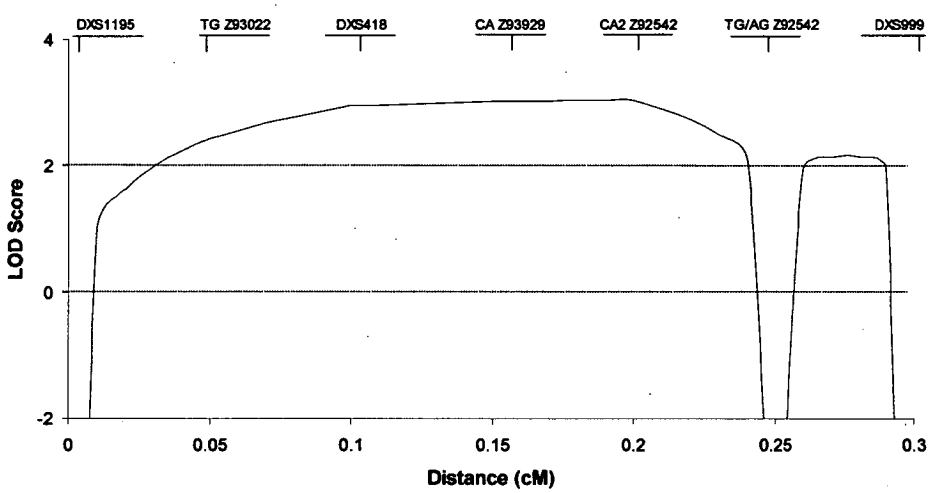
marker was also confirmed by a recombination in unaffected individual V:2 between DXS1195 and DXS418. The additional novel markers TG/AG Z92542 and TG Z93022 were not able to refine these recominations further as they were uninformative in the recombinant indivudual's mothers. A third recombination was observed between DXS999 and DXS365 in individual IV:6. This was not further investigated as it does not reduce the critical region for the NHS gene any more than the recombinations in individuals V:2 and VI:6.

A multipoint analysis using X-linked GeneHunter Plus in crch2551 with all 11 informative markers gave a maximum LOD score of 3.01 at CA2 Z92542, providing further significant evidence for linkage to this region (Figure 3.4A,B). Linkage is excluded by the large negative LOD scores distal to DXS1195 and proximal to DXS999, in agreement with the two-point LOD scores and observed recombinant events. TG/AG Z92542 gives negative LOD scores in both two-point and multipoint analyses, indicating no linkage to this marker even though it is within the region defined by the recombinants. The marker is homozygous in all affected individuals except V:3 and her daughters VI:3 and VI:4 so all linkage information comes from these individuals. Individual VI:4 has not been examined and therefore is coded as 'unknown phenotype' in the anaysis. If she is coded as 'affected' the LOD scores are less negative. This marker is generally uninformative leading to small negative LOD scores, although the scores do become positive at larger values of theta, consitent with the positive scores at marker DXS999. Additionally, the maker is homozygous in the mother or recombinant individual VI:6, making it impossible to determine if the recombination event is distal or proximal to TG/AG Z92542 in the multipoint analysis.

A



B



C

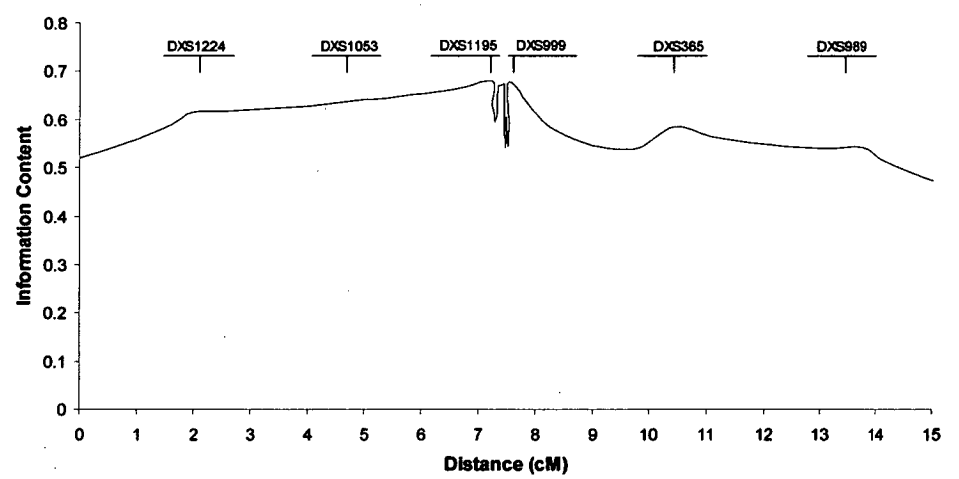


Figure 3.4. Multipoint parametric analysis of crch2551 calculated by X-linked GeneHunter Plus. **A.** LOD scores over 15cM between DXS1224 and DXS989. **B.** Expansion of the region between DXS1195 and DXS999 shown in A. **C.** Information content of the markers used across the whole region.

GeneHunter Plus is restricted in the size of the pedigrees it can analyse. When the pedigree exceeds the maximum size, the program excludes individuals from the analysis, beginning with the least informative. For example individual V:2 was excluded, as he was unaffected and had no affected offspring. However, haplotype analysis revealed that individual V:2 was a recombinant and informative. Alternative individuals were therefore manually removed from the analysis to force the inclusion of this individual. The maximum LOD score obtained without individual V:2 was 2.9 while a score of 3.0 was obtained with the inclusion of V:2. Both LOD scores provide significant evidence for linkage on the X chromosome, suggesting that in a fully penetrant single gene disorder where sufficient power is present, the overall significance of the results is unlikely to be greatly affected even when individuals who provide useful information are removed by the program. LINKMAP from the LINKAGE package (Lathrop *et al.* 1984) was also used to confirm these multipoint results with a subset of markers from the study and consistent results were obtained. LINKMAP can analyse pedigrees of arbitrary size, but is limited by the number of markers it can compute at one time.

The recombinations observed in this pedigree, along with the significant two-point and multipoint LOD scores, indicate that the critical region for the NHS gene is defined distally by DXS1195 and proximally by DXS999. Additional recombinations observed in this pedigree and crch24 do not add further information. Following this discovery refinement of the *NHS* locus to the same region was reported by Toutain *et al.* in a study of 11 pedigrees with NHS.

3.4.3.2 Isolated X-linked cataract pedigrees

Two point LOD scores were generated for each marker by MLINK. The values are presented in Table 3.3 for all 5 families combined and excluding crch2551. Although crch2551 provides most of the evidence for linkage, the smaller families are contributing positively to the signal (Table 3.3B). The negative LOD scores observed at most markers in Table 3.3A are considerably influenced by Crch2551, in which linkage is excluded to most markers by recombinant individuals.

Table 3.3: Two-point LOD scores at a range of recombination fractions. A. All pedigrees, B. Excluding crch2551

A.

Locus	0.00	0.001	0.01	0.02	0.03	0.04	0.05	0.10	0.20	0.30	0.40
DXS1224	-∞	-1.42	0.48	1.02	1.30	1.50	1.60	1.85	1.70	1.24	0.65
DXS1053	-∞	-2.41	-0.46	0.09	0.38	0.57	0.71	1.01	1.00	0.73	0.37
DXS1195	-∞	-2.30	-0.37	0.18	0.48	0.67	0.81	1.13	1.15	0.88	0.48
DXS418	4.29	4.28	4.20	4.11	4.02	3.94	3.85	3.39	2.47	1.53	0.67
DXS999	-∞	-0.56	0.43	0.71	0.86	0.96	1.02	1.16	1.05	0.76	0.40
DXS365	-∞	-6.43	-3.10	-2.40	-1.44	-1.02	-0.72	0.10	0.54	0.48	0.25
DXS989	-∞	-2.00	-0.06	0.47	0.75	0.93	1.05	1.29	1.13	0.71	0.27

B.

Locus	0.00	0.001	0.01	0.02	0.03	0.04	0.05	0.10	0.20	0.30	0.40
DXS1224	-2.20	-1.16	-0.22	0.05	0.20	0.29	0.36	0.52	0.52	0.38	0.19
DXS1053	-∞	-0.70	-0.10	0.17	0.31	0.40	0.46	0.60	0.55	0.38	0.19
DXS1195	0.30	0.30	0.29	0.28	0.27	0.27	0.26	0.21	0.13	0.06	0.02
DXS418	1.90	1.89	1.86	1.81	1.78	1.73	1.69	1.48	1.05	0.63	0.25
DXS999	-0.62	-0.61	-0.59	-0.56	-0.54	-0.51	-0.49	-0.40	-0.26	-0.16	-0.07
DXS365	-∞	-0.87	0.10	0.36	0.50	0.58	0.64	0.74	0.63	0.42	0.20
DXS989	1.62	1.62	1.59	1.56	1.53	1.49	1.46	1.29	0.93	0.57	0.24

Haplotypes were generated for each pedigree and are shown in Figure 3.3. Analysis of crch05 indicates that this pedigree does not demonstrate linkage to this region of the X chromosome because the discordant siblings in the third generation have the same haplotype passed down from their maternal grandmother. The phenotype of male individual III:3 was determined on clinical examination to be unaffected. This diagnosis can be accurately made in males as they appear to show complete

penetrance of the disease due to hemizygosity of the X chromosome. The genotyping was repeated and the same result was obtained, suggesting a different etiology for the cataracts in this pedigree. It is possible that there is a second X chromosome locus for isolated X-linked cataract, but it is also possible that this small family has an autosomal dominant form. Although the phenotype is much more severe in the one male present than in his female relatives who had mild sutural lens opacities, it is difficult to draw conclusions about sex-specific severity of the phenotype in this family. The other three small pedigrees (crch02, crch20 and crch24) have segregating haplotypes across the NHS region, consistent with linkage, but the pedigrees do not have sufficient power to give significant LOD scores individually. The phenotype in individual crch24 IV:2 is unkown as she was not examined by an ophthalmologist, although she appears to carry the risk haplotype. A recombination between DXS1224 and DXS1053 in individual II:2 of crch24 indicates exclusion of the disease gene from the region distal to DXS1224 and is consistent with the findings in crch2551.

While these smaller pedigrees appear consistent with linkage to this region of Xp22.13, supporting the hypothesis that NHS and X-linked cataract are allelic, they do not refine the critical region further as there are no recombinations within the region defined by pedigree crch2551.

3.4.4 Candidate genes

The critical region for NHS as defined by DXS1195 and DXS999 is 1.3Mb in length. There are 6 characterised genes in the region, all of which are candidates for NHS: *SCML1, SCML2, RAI2, STK9, RS1* and *PPEF1*.

The Sex-comb on midleg-like 1 and 2 (*SCML1* and *SCML2*) genes are transcription factors involved in the regulation of *HOX* genes. They have homology to the Sex-comb on midleg (*Scm*) gene in *Drosophila*, which is involved in the regulation of transcription of homeotic genes, the *Drosophila* equivalent to the human *HOX* genes. *Scm* is a member of the Polycomb (PcG) group of genes, which are transcriptional repressors. These genes (along with the trithorax group that act as activators) ensure the correct spatial expression of the homeotic genes that drive normal segmental development in *Drosophila* (van de Vosse *et al.* 1998; Montini *et al.* 1999). Mice with mutations in the murine equivalent of the PcG genes display axial skeletal malformations and other abnormalities. Both *SCML1* and *SCML2* are involved in development, are expressed in a range of both adult and foetal tissues and are located within the critical region for NHS on Xp22.13. As such they were considered to be candidate genes for this disorder.

RAI2 is the human homologue of the murine retinoic-acid induced gene. It was excluded as the causative gene for NHS by the direct sequencing of nine unrelated NHS patients (Walpole *et al.* 1999), however, we felt it important to confirm this finding in our pedigree.

STK9 is a serine-threonine kinase with homology to cell cycle proteins (Montini *et al.* 1998). This 20 exon gene is also expressed in a range of adult human and mouse tissues, including brain. Although its exact role has not yet been elucidated, it was considered to be a candidate for NHS.

Mutations of the *RS1* gene have been shown to cause X-linked retinoschisis (XLRS1), which involves degeneration of the retina (Huopaniemi *et al.* 1999). This small 6 exon gene is almost exclusively expressed in the retina (Huopaniemi *et al.* 2000). Given the syndromic features of NHS, this gene was not a promising candidate but was included in the study for completeness and because it causes an ocular phenotype.

PPEF1 encodes a novel serine-threonine protein phosphatase with EF hand motif. The function of this protein has not been fully elucidated, but it is expressed almost exclusively in brain, particularly in sensory neurons derived from the neural crest (Montini *et al.* 1997). This gene was considered as a candidate because of its location within the critical region and undetermined target protein.

All coding regions of these genes were sequenced in 65 amplimers (Chapter 2.7.1) with the primers listed in Appendix 2 in two affected females (V:9, V:3), one affected male (V:5) and one unaffected female (III:5) of pedigree 2551. The sequence obtained was compared to that in the Genbank. No segregating mutations were identified in the coding regions of these six genes, suggesting that they are not responsible for the phenotype in this family. Three non-segregating polymoprhisms were indentified and presented in Table 3.4. An A to G substitiution at nucleotide 796 was detected in the *RAI1* gene, coding for K226R variant. An E42K variant was detected in the *SCML1* gene an S7T in the *PPEF1* gene. The variants did not segregate with disease in the four individuals sequenced and were not further investigated. The presence of cryptic splice sites or promoter mutations was not ruled out. Based on the nature of the disease (ie pleiotropic, developmental and severe), it

was thought that a coding mutation was more likely to account for the disease rather than regulatory mutations. Also, although the genomic structure of the genes was known, their functions and regulation were not well defined, making it difficult to assess the key promoter regions. This approach did run the risk of missing a causative mutation. Investigation of predicted novel genes in the region was undertaken and is presented in Chapter 4.

Table 3.4: Polymorphisms identified in candidate genes

Gene‡	Nucleotide*	Exon	Variant	Amino Acid Change	Individual†
RAI1	796	Exon 1	A>G	K226R	V:3, V:9
SCML1	123	Exon 4	G>A	E42K	V:9
PPEF1	18	Exon 4	T>A	S7T	III:5

*Nucleotides counted from start of translation
†Only individuals V:3, V:5, V:9 and III:5 were investigated
‡No variants were confirmed in RS1, SCML2 or STK9 in these 4 individuals

Chapter 4

The Identification of the Nance-Horan Syndrome Gene

4.1 Background

The refinement of the Nance-Horan Syndrome (NHS) critical region and exclusion of the six candidate genes by direct sequencing in the large Australian pedigree described in Chapter 3 represents a significant advance in the knowledge of the genetics of NHS. However, the causative gene remained to be identified. We next undertook an investigation of predicted genes in the draft human sequence, both "*in silico*" and *in vitro*.

4.2 Methods

4.2.1 Prediction of novel genes

There are numerous predicted genes and EST clusters in the region. We began the search for novel candidate genes for NHS by comparing computer predicted genes with mRNA and EST information mapping between DXS1195 and DXS999 on the X chromosome. GeneScan (Burge and Karlin 1997) predicts 22 genes in the region. This is the largest number of genes predicted by any of the programs represented at the University of California at Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu), which at the time of analysis included Acembly (http://www.acedb.org/Cornell/acembly/), Twinscan (Korf *et al.* 2001), Fgenesh++ (Salamov and Solovyev 2000), Geneid (Parra *et al.* 2000) and Genescan (Burge and Karlin 1997). These predictions were used as the basis for initial comparisons, and evidence such as ESTs from other species was compared to these. Expression profiles for predicted genes were investigated by the examination of the tissues from which

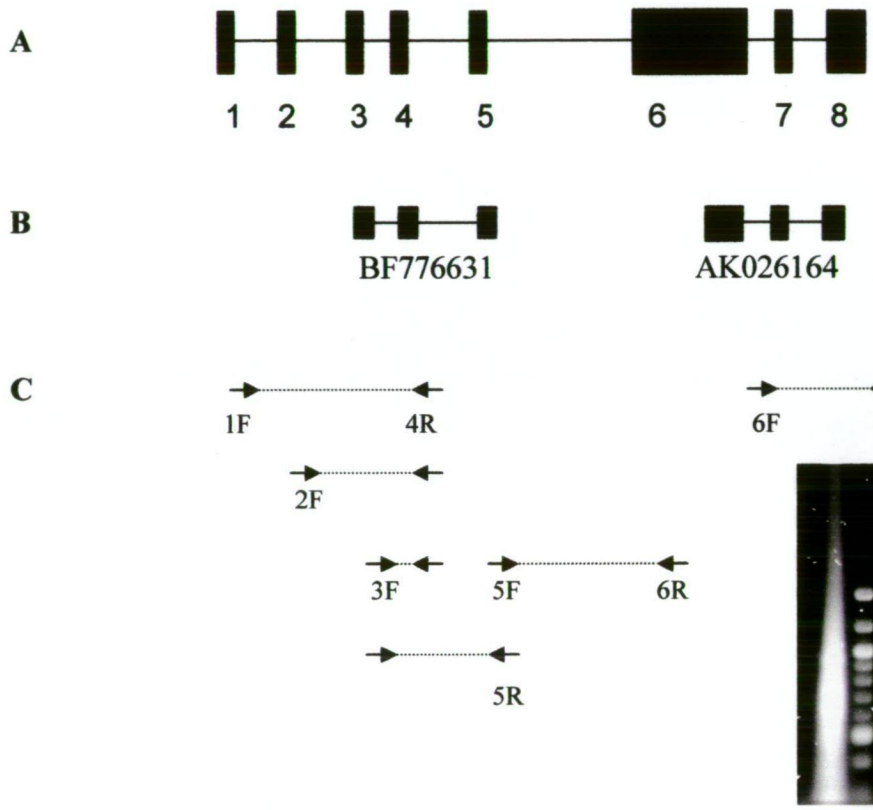
supporting ESTs were obtained. A candidate (similar to LOC90334) was chosen on the basis of its expression profile and conservation in the mouse and cow and the predicted structure is depicted in Figure 4.1A and B.

4.2.2 RT-PCR for the identification of transcripts

Primers were designed to capture the predicted exons of the chosen candidate by RT-PCR. These are presented in Appendix 2, with the strategy depicted in Figure 4.1C. The reverse primers in predicted exons 4, 5, 6 and 8 were used in separate reactions to reverse transcribe total RNA from normal adult human brain (described in Chapter 2.9) as the RNA used was known to have undergone some degradation (by visualisation on 1% agarose stained with ethidium bromide, Figure 4.1C). Therefore the predicted full-length transcript targeted by an oligo-dT primer would be relatively long and likely to be degraded. The four templates were then used in PCR reactions with appropriate combinations of primers targeted to the predicted exons. All RT-PCR products were sequenced using the method described in Chapter 2.7.2.

4.2.3 Mutation screening

All detected regions of the novel gene (later determined to be causative for NHS) were sequenced in affected individuals from all X-linked pedigrees using intronic primers given in Appendix 2. Remaining family members were then screened for the identified mutations. Greater than 200 control chromosomes from unaffected, unrelated individuals collected from nursing homes in Launceston, Tasmania were screened for the identified mutations.



Lane	Primers	Expected Size (bp)	Observed Size (bp)
1	1F-4R	411	
2	3F-4R	262	~250
3	3F-6R	533	~550
4	3F-5R	344	~350
5	blank		
6	5F-6R	194	smear
7	6F-8R	206	~200
RACE	5R	-	~200

Figure 4.1. A. Predicted gene structure. **B.** key ESTs involved in the prediction. **C.** RT-PCR strategy for detection of transcripts and degraded human brain RNA leading to reverse transcription with gene specific reverse primers. **D.** RT-PCR products amplified with annealing temperature of 60°C on 1.5% agarose gel stained with ethidium bromide with expected and observed product sizes. PCR with primers 5F and 6R was further optimised to obtain a single band. 5'RACE product was amplified from a nested reaction with primer 5R and nested primer supplied with kit (see section 2.10)

4.2.4 In Silico investigations

Nucleotide and protein database searches were carried out through the website of the National Centre for Biotechnology Information (NCBI) by BLAST algorithms (Altschul *et al.* 1997) blastn, blastx and tblastx against the non-redundant (nr) databases, EST database (dbEST), high-throughput genomic sequences (htgs) database and species-specific databases. The putative promoter was identified using the Neural Network Promoter Prediction tool (Reese 2001) and subcellular localisation of the putative NHS protein has been predicted using PSORTIII (http://psort.nibb.ac.jp/form2.html).

4.3 Results

4.3.1 In silico identification of a novel candidate gene

The expression profiles of predicted genes were investigated by examining the tissues from which ESTs supporting the prediction were obtained. Expression in eye was considered to be of importance, given the characteristic ocular features of NHS. One gene predicted by GeneScan matched ESTs extracted from human foetal eye and human brain, two tissues that are affected in NHS. This prediction was similar to LOC90334 in Genbank. Microsatellite marker DXS418, which gave the peak linkage signal, was located within a predicted intron. Further investigation of this predicted gene revealed identity to several spliced ESTs, from human, mouse and cow, providing evidence that the predicted exons are transcribed together and conserved amongst species. A reading frame was identified through all 8 predicted exons and the same gene was found to be predicted in the mouse genome in the syntenic region of chromosome X. There were many other predicted genes in the region, however, the lack of evidence for expression in eye and the lower level of conservation in mouse

led to them being given a lower priority. A BLAST search of the predicted protein against proteins in the databases did not reveal any significant homologies, indicating that this gene may encode a protein of novel class. The predicted gene structure and key ESTs are presented in Figure 4.1 along with the strategy for identifying transcripts.

4.3.2 RT-PCR and sequence confirmation of fragments

Single PCR products were obtained with the primers designed to amplify predicted exons 3-4, 3-5, 5-6, and 6-8. No products were obtained by PCR with primers designed to capture exons 1 and 2. These two predicted exons are not predicted in the mouse sequence, although two alternative exons are postulated. The results of sequencing of RT-PCR products are presented in Figure 4.1D and 4.2A. The fragment amplified with primer 3F and 4R was of the predicted length (262 bp) and showed sequence identity with predicted exons 3 and 4. The 3' splice site of exon 3 was found to be identical to the EST BF776631, which covers predicted exons 3, 4 and 5 (Figure 4.1B), rather than the splice site predicted by GeneScan 60 bp further downstream. The fragment amplified with primers 3F and 5R was also of the predicted size (344 bp) and showed sequence identity with predicted exons 3, 4 and 5 as expected. The 3' splice site of exon 3 was confirmed. The fragment amplified with primers 5F and 6R was 326 bp long. This was larger than the predicted size of 194 bp. Sequencing of this product revealed sequence identity with predicted exons 5 and 6, and also with a 132bp region in the predicted intron separating these exons. A novel exon, 5a, was identified from these results as indicated in Figure 4.2A. The sequencing of the PCR products obtained using primers 6F and 8R showed sequence identity with exons 6, 7 and 8 as predicted. These results indicate that the predicted

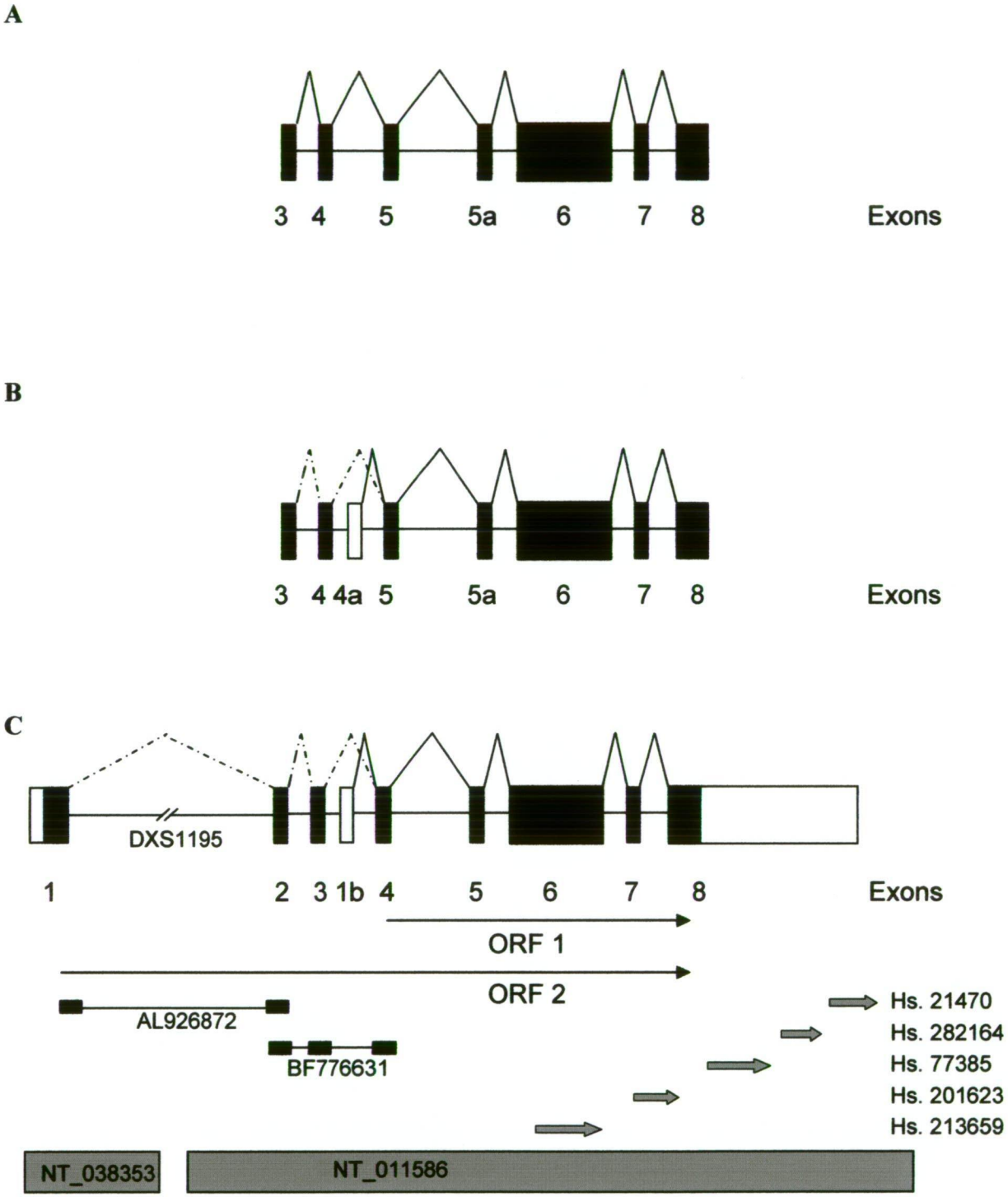


Figure 4.2. A. RT-PCR results; predicted exons plus exon 5a detected. No evidence for exons 1 and 2. **B.** 5'RACE results; exon 4a detected. No evidence for exon 4 splicing directly to exon 4a. Evidence for two isoforms. **C.** Final structure determined; exon 1 identified by RT-PCR, two isoforms and two open reading frames. Location of ESTs, Unigene clusters and human contigs indicated.

gene is transcribed in normal adult human brain, although its structure is not exactly as predicted.

4.3.3 Further investigations of genomic structure

4.3.3.1 5'RACE

In an attempt to identify the 5' region more accurately, 5' random amplification of cDNA ends (5'RACE) was used (Chapter 2.10). A smear containing products of varying molecular weight was obtained using primer 6R (Appendix 2) and the 5'RACE primer provided in the kit (Invitrogen). This product was then used as a template for nested PCR using the 5'nested primer provided and primer 5R. A small specific product was obtained and was sequenced which revealed identity to a region in the predicted intron between exons 4 and 5. This novel exon is labelled 4a in Figure 4.2B. Forward and reverse primers specific to this exon were designed (Appendix 2) and RT-PCR carried out on the original brain cDNA used for the previous RT-PCR experiments. A band of the predicted size (137 bp) was detected using primers 4aF and 5R and was shown by sequencing to have the same sequence as the fragment detected by 5'RACE. No fragments could be amplified using 3F and 4aR, suggesting that this novel exon does not splice to predicted exons 3 and 4 and raising the possibility of two isoforms. This is supported by the presence of only the one product amplified by primers 3F and 5R.

4.3.3.2 RT-PCR to investigate the 5' end

The work of Dr James McKay in this laboratory revealed that zebrafish (*Danio rerio*) EST AL926872 shows sequence similarity to predicted exon 3 at its 3' end and indicated the presence of additional sequence in the longer isoform. The 5' end of the EST was found to be homologous to human clone AL84533 that was unlocalised at

the time of the investigation. It has now been mapped distal to the identified gene in contig NT_038053. A forward primer was designed to capture this predicted 5' exon in conjunction with a reverse primer in predicted exon 3 (Appendix 2). A product of the predicted size was obtained and subsequent sequencing revealed the expected product. The acceptor splice site of exon 3 as predicted was confirmed, as was the donor splice site of the new exon, indicated as exon 1 in Figure 4.2C. There is an in frame methionine 565 nucleotides from the 3' splice site (Figure 4.2C). There are in frame stop codons upstream of this translation start site, indicating that there is no more coding sequence 5' to this ATG and this exon represents the first translated exon. Notably, the microsatallite marker defining the proximal end of the critical region, DXS1195, is within the large first intron of this gene (Figure 4.2C). Investigations are continuing into the structure and splicing alternatives of the 5' end of this gene.

4.3.3.3 Identification of the promoter

Work done by our collaborator Dr Jozef Gecz at the Women's and Children's Hospital, Adelaide, Australia, identified a *NotI* site indicating a CpG island upstream of exon 1. Analysis of the nearby sequence using Neural Network Promoter Prediction tool identified a transcription start site (score 1.0) at position -338 with respect to the putative ATG translation start site. This would make exon 1 903 bp in length. The sequence surrounding the transcription start site is:

GCCCATTTATATAGGCGGCGGCGACGGCAGTGGCCGGgGTGGCGACG,

with the first transcribed base (g) of the mRNA in lower case. This promoter has not been further investigated.

4.3.3.4 Identification of the 3'UTR

Work also carried out by the group of Dr Gecz aligns Unigene EST clusters Hs.77385, Hs.282164 and Hs.21470 to form the 3'UTR of this gene as part of exon 8 (Figure 4.2C). There is an in frame stop codon in exon 8, indicating the end of the coding region. A polyadenylation signal was identified in Hs.21470 3,538 bp past the stop codon, indicating the end of the transcript. Primers used for these RT-PCR experiments are given in Appendix 2.

4.3.3.5 Genomic structure

Overall, this novel gene, designated here as *NHS* was found to consist of 9 exons, which are possibly transcribed as two major isoforms A (8.7 kb), and B (7.7 kb). The longer isoform A consists of all exons except 1b and contains an open reading frame encoding a putative 1630 amino acid protein with translation start methionine in exon 1 (Figure 4.2C). The shorter isoform B consists of exons 1b through 8 and encodes a putative 1335 amino-acid protein with translation beginning in exon 4 (Figure 4.2C). Intron/exon boundaries are given in Table 4.1.

Table 4.1: Intron/Exon boundaries of the NHS gene.

Exon	3'acceptor splice site	5' donor splice site	Exon size (bp)	Intron size (bp)
1		GAGGCAGTGC**gt**gagtaccc	903	~350000
2	tgtcttgc**ag**CCGTCTCCAA	CTGCGCAGAG**gt**gacagatc	153	4440
3	tccttctc**ag**AACACCGGAG	CTTTTTAACG**gt**aagtttgg	134	27626
1b		CTCTGACAAG**gt**aagtaaat	138	1217
4	ttgtttgc**ag**TCCCATCCCC	AATGTTACTG**gt**atcgttct	193	2665
5	tgggttgc**ag**GAGTTGGCTT	CAAGAAATAG**gt**gtgatatc	132	916
6	tgtgccct**ag**ATTCTGATGA	TCATTGAAAG**gt**cagtcact	2982	320
7	ttattttta**ag**AATCATCACC	TCATTCACAG**gt**gaggcaac	127	3082
8	tttctcaaa**ag**ATCCAAGAGG	GCAAATAAACGTGACTGCA	4134	

* Exons are indicated in capital letters, and introns in lower case. The consensus splice sites are indicated in bold.

4.3.4 Mutation screening

4.3.4.1 Crch2551

Individuals V:9, V:3, V:5, III:5 of pedigree crch2551 (Figure 4.3A) were directly sequenced at all coding fragments using primers in Appendix 2. One sequence variant was detected in exon 6 fragment D and consisted of a single C nucelotide insertion at position 2387, with nucleotide 1 being the first base of codon 1 in exon 1 (Figure 4.3B, Figure 4.5). The insertion was identified in all affected individuals and no unaffected individuals. VI:4 is a female carrier of the mutation, but has not been examined clinically for NHS. The mutation was not detected in 202 control chromosomes from residents of nursing homes in Launceston, Tasmania, Australia who had been examined for eye disease. The insertion disrupts the predicted reading frame, resulting in a premature translation stop site, severely truncating the protein.

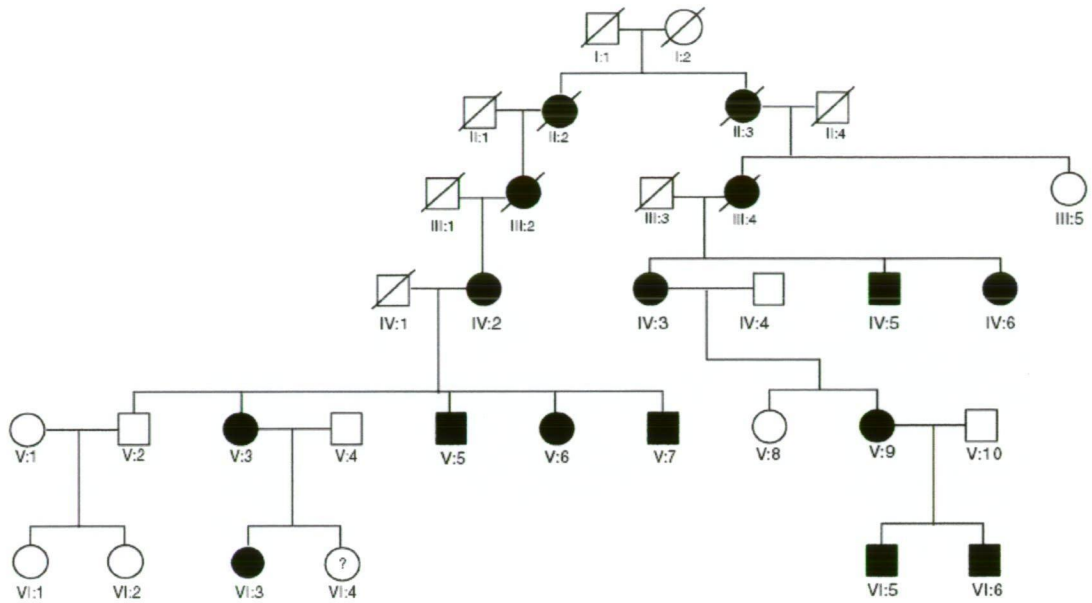
4.3.4.2 Screening of X-linked pedigrees

The other X-linked pedigrees in this collection were screened for mutations in this gene, including the entire 3'UTR. No mutations were identified in pedigrees crch02, crch20 or crch24.

4.3.4.3 Ascertainment and screening of additional NHS pedigrees

Additional NHS patients (and where possible their family members) were ascertained from Australia and the U.K. (Figure 4.4). Pedigree cwa70 is from Western Australia and was originally described in 1990 as having NHS (Walpole *et al.* 1990). Pedigrees UK1, UK4, UK5 and UK6 were diagnosed with NHS. Pedigrees UK2 and UK3 were believed to have isolated X-linked cataract (J.E. Craig and I.M. Russell-Eggitt, pers. comm.). Genotyping of the microsatellite markers previously used to detect linkage with this region revealed that pedigree UK2 contained no segregating haplotype and

A



B

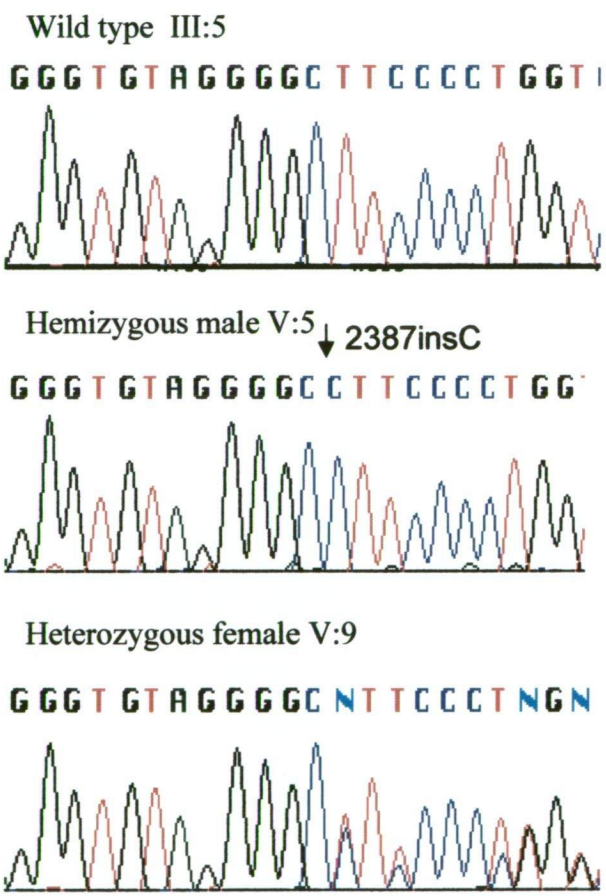


Figure 4.3. A. Pedigree crch2551. Shaded symbols indicate presence of ophthalmologist-confirmed Nance-Horan Syndrome. Individual VI:4 has not been examined. **B.** Sequence variant 2387insC detected in pedigree crch2551 in exon 6 of the novel gene.

the data were inconsistent with linkage to this region. Pedigree UK2 was not investigated further. The data from all other pedigrees studied were consistent with linkage, as all affected individuals within each pedigree displayed a common haplotype, but were not significant evidence for linkage due to the small size of the pedigrees.

Truncation mutations were detected in four of the five NHS families (Figures 4.4 and 4.5). 3459delC (Ala1153fs) was found in cwa70, 1117C>T (Arg378X) arising *de novo* was detected in a male infant from pedigree UK4, and pedigree UK1 revealed a 400delC (Arg134fs) mutation. Two sequence variants were identified in a female index case of UK5. This patient displays a C to G substitution in the 3' acceptor splice site of exon 3, as well as 718insG (Glu240fs) at the first base of the exon. The proband's mother displays the same genotype, indicating that the 2 mutations are in *cis*. All identified mutations resulted in a truncated form of the predicted protein and were absent in 200 control chromosomes. No coding mutations in this gene were detected in pedigree UK6 or the isolated X-linked cataract pedigree UK3.

4.3.5 Expression analyses

In addition to the RT-PCR carried out in human adult brain, work carried out in the laboratories of Dr Gecz and Dr Jamie Craig (Department of Ophthalmology, Flinders University, Adelaide, South Australia) have indicated that the *NHS* gene is expressed in all tissues tested. RT-PCR using the primers designed to amplify exons 6-8 was carried out in human adult and foetal brain, lens, retina, retinal pigment epithelium, lymphocytes, and fibroblasts. PCR amplimers were obtained from all tissues, although the levels in lymphocytes and fibroblasts were very low data not shown). Human Fetal Multiple Tissue Northern Blot II (cat# 7756-1, Clontech) was

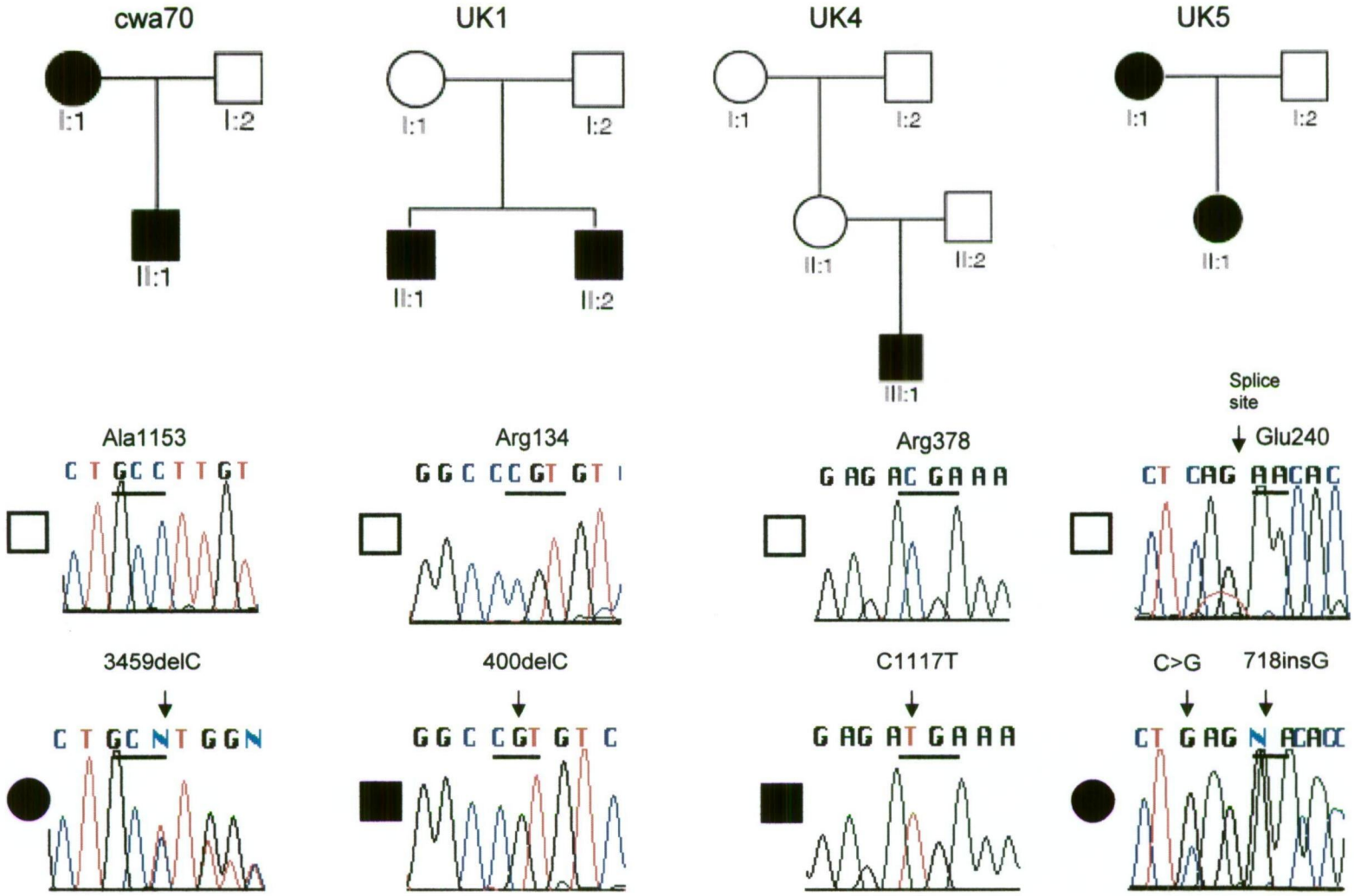


Figure 4.4. Additional NHS pedigrees and identified mutations. Pedigrees without identified mutations can be found in Appendix 1.

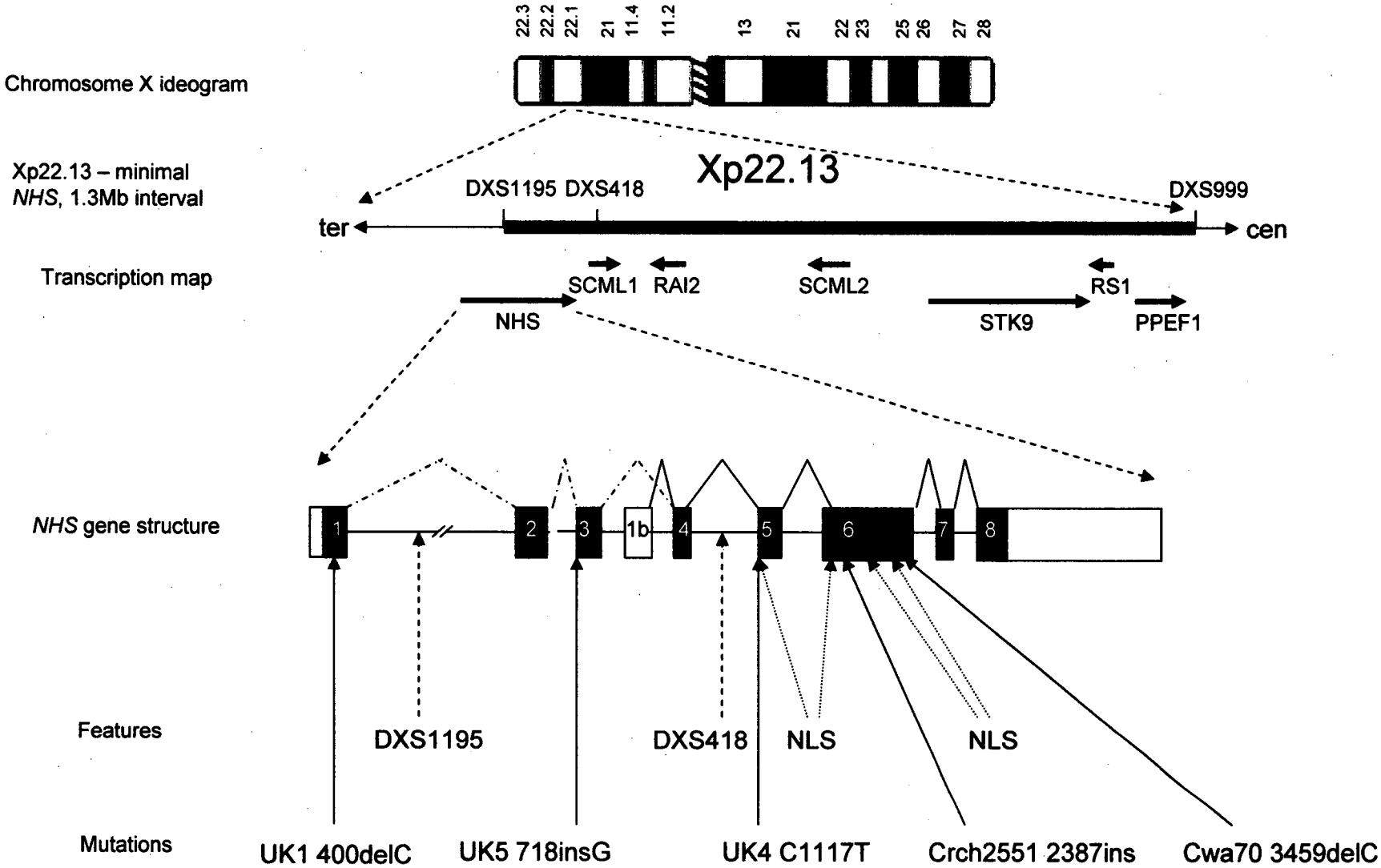


Figure 4.5: Transcription map of critical region and genomic structure of the *NHS* gene with location of identified mutations. Intragenic microsatellite markers DXS1195 and DXS418 are indicated. NLS = putative Nuclear Localisation Signal.

hybridised with 3 independent probes generated from exon 6, exons 6-8 and the 3'UTR of the novel gene. Transcripts of approximately 8.5 and 7.5 kb were detected in all tissues analysed, which included brain, heart, skeletal muscle, colon, thymus, spleen, kidney, liver, small intestine, placenta, lung and peripheral blood leukocyte, all at low levels (not shown). A mouse aging blot, (cat# MBAB 1009-1 and product# 1009-1-D103 Seegene) containing RNA from mouse brain tissues at different developmental stages indicated broad temporal expression in foetal and adult brain (Figure 4.6B).

4.3.6 Computational analyses

Work conducted by Dr McKay and Dr Gecz revealed orthologs in mouse, rat, chicken, cow, zebrafish and pufferfish with ESTs identified in all species. The genomic structure of the gene was shown to be conserved in mouse, rat and pufferfish through homology searches of genomic sequence and the GeneScan gene prediction program. The NHS protein shows no significant homology on BLAST search to any known protein or class of proteins and hence represents a novel class. Similarity between the NHS homologs is relatively high, with the human NHS protein 76%, 76% and 42% identical to the mouse, rat and pufferfish NHS proteins, respectively. Four putative monopartite nuclear localisation signals were detected at amino acid positions 371-379: RRRKLRRRK; 438-444: PSRRRIR; 822-825: RKPK; and 1026-1034: PGGSKRKPK. These are indicated on Figure 4.5. An NHS paralog (NHP) was identified on human chromosome 6. This predicted gene (similar to KIAA1357) has 24% identity at the amino acid level by BLAST. EST information suggests that NHP is expressed in a range of tissues including foetal eye, is likely to have two isoforms and homologs exist in other species.

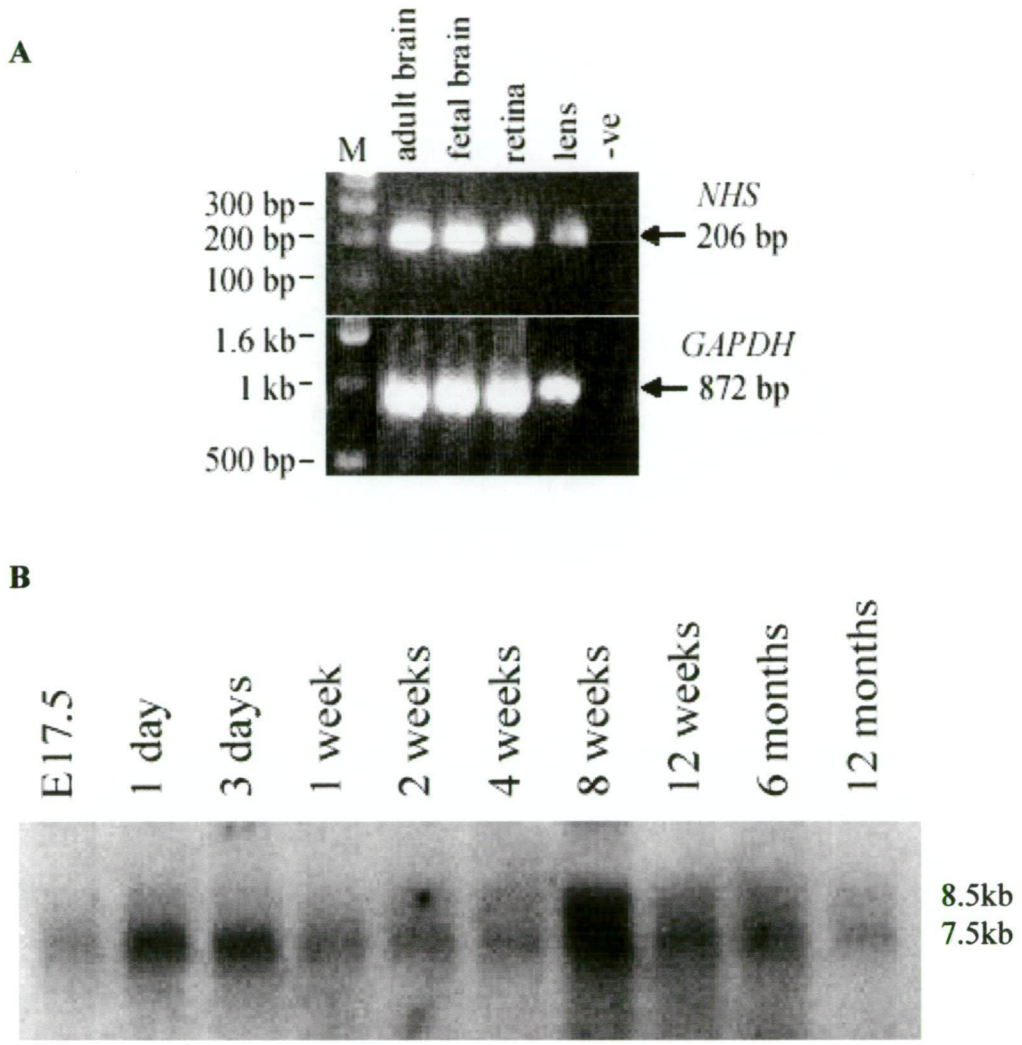


Figure 4.6. A. RT-PCR from human adult brain, fetal brain, retina and lens tissues. RNA was reverse transcribed and amplified using primers from exon 6 and 8. RT-minus controls were negative for each tissue. Expression was also documented from retinal pigment epithelium, placenta, lymphocytes, and fibroblasts. **B.** Mouse northern blot (Seegene, Korea) containing 20 mg of total RNA per lane from mouse brain at various developmental stages. The probe was generated from the 3' end of the mouse *nhs* gene. Hybridisation was carried out in ExpressHyb solution (Clontech) at 68°C as per manufacturer's instructions. (Experiments performed by Shiwani Sharma and Marie Shaw, Department of Ophthalmology, Flinders University, Adelaide, South Australia and Marie Shaw, Women's and Children's Hospital, Adelaide, South Australia)

4.4 Discussion

4.4.1 Evidence for the NHS gene

Mutations resulting in the truncation and likely loss of function of the predicted protein have been identified in a novel gene segregating with NHS in 5 pedigrees. We have directly documented *NHS* gene expression in normal human lens and brain and in most other human tissues (some at very low level). Mouse expression data indicate expression in brain over multiple developmental stages. Therefore, this gene has an expression profile consistent with the spectrum of clinical features of cataract, dental abnormalities and mental retardation seen in NHS.

4.4.2 The critical region

Marker DXS1195 was defined as the distal marker of the NHS critical region (Chapter 3). However, this marker is in the first intron of the *NHS* gene. Therefore individual V:2 of crch2551 with the recombination proximal to this marker received exon 1 from his mother's (IV:2) affected chromosome and the rest of the gene from the non-disease chromosome. While there was no variation detected in exon 1 that would allow for direct confirmation of this, the three distal microsatellites (DXS1195, DXS1053 and DXS1224), clearly indicate that a recombination has occurred. The mutation in pedigree crch2551 is in exon 6 (Figure 4.5). Individual V:2 does not carry the mutation, as determined by direct sequencing, consistent with his unaffected status. The mutation is within the critical region identified by this individual (also defined by Toutain *et al.* (2002)). However, the mutation in pedigree UK1 is distal to this critical region (Figure 4.5). Therefore, the critical region identified by any given pedigree does not limit the entire gene to that region. A second pedigree (eg UK1 if it was large enough for linkage analysis) may define a different region, based on the location of the pedigree specific mutation. In the case of a gene spread over such a

large distance as *NHS* (approximately 400 kb), these regions may not necessarily overlap. What was actually identified in this study was a critical region for the causative mutation in pedigree crch2551.

4.4.3 Analysis of the amino acid sequence

The *NHS* gene is conserved, with uncharacterised orthologs in pufferfish, zebrafish, mouse, rat and cow. Based on EST information and draft assemblies where available, all species appear to have both isoforms. There is no significant homology to any characterised protein and therefore the function of the *NHS* gene and its product is unknown. A BLAST search of the protein sequence reveals limited homology to dentin sialoprotein. This is a serine-rich structural protein found in dentin, the layer of tooth below the outer enamel. The homology is due to the high serine content rather than actual significant sequence alignment.

No known protein domains were detected by the blastp algorithm at the NCBI blastp program website. Four monopartite nuclear localisation signals (NLS) were detected (Figure 4.5). These consensus amino acid sequences target proteins to the nucleus. Whether or not these sequences are functional in this protein is under investigation in collaborating laboratories. The targeting of this protein to the nucleus would support a developmental role for the protein, consistent with the pleiotropic nature of NHS. The disorder is congenital, with multiple tissues affected from an early age. However diagnosis may be delayed as features are often difficult to recognise in early infancy because the teeth are not erupted.

4.4.4 The effect of the mutations

All mutations identified are predicted to result in truncation of the protein, removing the putative NLS regions. There does not appear to be any genotype-phenotype correlations with regard to the removal of these putative nuclear localisation sequences. The phenotype is equally severe in individuals whose mutations remove all four signals (eg. UK1) as in individuals for whom the mutation does not affect these sequences (Cwa70) (Figure 4.5). Therefore, it is likely that the phenotype is caused at least in part by a loss of function, rather than simply a loss of subcellular localisation. Alternatively, truncation mutations can result in nonsense-mediated mRNA degradation, resulting in insufficient protein levels and haploinsufficiency rather than disrupted function.

The mutations in pedigrees UK1 and UK5 in exons 1 and 3 respectively (Figure 4.5), indicate that loss of the long isoform is sufficient to cause disease. Additionally, in families bearing a mutation affecting only the longer isoform, there is no correlation between this and severity of disease. Therefore a disruption to the ratio of the two isoforms may play a role in etiology of the disease in these pedigrees.

The short arm of the X chromosome contains many genes that escape X inactivation, including genes on Xp22 (Carrel *et al.* 1999). This would explain the apparent intermediate phenotype in females giving the appearance of co-dominant inheritance. Also, the ratio of isoforms may be further disrupted in females with a mutation affecting only the longer isoform.

The significance of the two variants identified in UK5 is unknown. The G insertion results in a predicted truncated form of the protein, with the loss of the C-terminal end

of the protein. Alternatively, it is possible that the C/G substitution causes an alteration in the splicing of the *NHS* gene. Either mechanism is predicted to cause the loss of function of the NHS protein.

4.4.5 Genetic heterogeneity or unidentified mutations?

A mutation was not identified in NHS pedigree UK6. Although demonstrating typical NHS features, this pedigree is too small to demonstrate linkage to the NHS region. The NHS phenotype is quite variable, with features overlapping with other syndromes. Thus the absence of mutations in the pedigree UK6 index case could represent genetic heterogeneity. The phenotype may be caused by non-coding mutations undetected by the current study affecting splicing efficiency and/or relative amounts of the two isoforms, or there may remain additional exons to be identified, particularly at the 5'end. It is also plausible that mutations located some distance from the gene may produce expression of this phenotype. Such effects have recently been reported for another developmentally regulated gene, *PAX6*. Deletions approximately 20 kb 3' to the polyadenylation signal were shown to prevent transcription of the allele, presumably by interfering with a locus control region (Lauderdale *et al.* 2000).

4.4.6 Xcat mouse model

The *Xcat* mouse phenotype has been proposed as an animal model of NHS, based on the phenotype of X-linked cataract and the localisation of the *Xcat* mutation to the syntenic region of the mouse chromosome X (Favor and Pretsch 1990). This phenotype should be examined in relation to the murine homologue of NHS. The mutation was originally created by radiation of DBA/2 male mice and the phenotype identified in an F_1 female (Favor and Pretsch 1990). The phenotype consists of a

bilateral total opacity of the lens in hemizygous males, with varying severity of nuclear and cortical opacity in heterozygous females. The mice do not display the dental features characteristic of NHS in humans and may be better described as an animal model of isolated X-linked cataract rather than NHS.

4.4.7 X-linked cataract

Recently, Francis *et al.* (2002) described a large pedigree with isolated X-linked cataract consisting of severe congenital cataract requiring extraction in males and a milder progressive phenotype in females. This phenotype was shown to link to a region overlapping with, but larger than the critical region for NHS, suggesting the disorders may be allelic. We identified 4 small pedigrees (crch02, crch20, crch24 and UK3) with a phenotype and inheritance pattern suggesting X-linked inheritance. Haplotypes across Xp22 were consistent with linkage to this region, although there was insufficient power to generate a significant LOD score. No mutations were identified in the *NHS* gene in these pedigrees. It is possible that these pedigrees are not truly X-linked. This cannot be definitively confirmed with pedigrees of this size.

Further characterisation of the promoter and distal regulatory elements of the *NHS* gene and investigation in larger X-linked pedigrees identified in other studies such as that of Francis *et al.* (2002) is required before it can be definitively excluded as the cause of isolated X-linked cataract. This phenotype may be caused by mutations in uncharacterised regions of the gene. Alternatively, a second gene in the region may be the cause of isolated cataract. There are many other predicted genes within the region localised by Francis *et al.* (2002).

One Australian pedigree and one UK pedigree were shown not to link to Xp22 by investigation of the haplotypes. However, the phenotype and inheritance pattern indicate X-linked inheritance. This suggests that there may be a second locus for X-linked cataract on the X chromosome.

4.4.8 Summary

Our findings provide evidence that mutations in this novel gene, *NHS*, are causative of Nance-Horan Syndrome. This discovery will enable prenatal diagnosis in families known to have blindness and mental disability due to NHS, and facilitate early diagnosis in sporadic cases, followed by appropriate treatment. Further studies are required to establish the function of the *NHS* gene product, thereby identifing a novel molecular pathway for cataract development. In addition, delineation of *NHS* gene function will increase understanding of odontogenesis, craniofacial and neural development.

Chapter 5

The Investigation of Crystallin Genes in Autosomal Dominant Paediatric Cataract Pedigrees

5.1 Background

The refractive index of the lens is the key to its ability to focus light onto the retina. There is a gradient of refractive index from the centre to the periphery created by a gradient of protein concentration (Kannabiran and Balasubramanian 2000). Lens fibre cells have around double the protein content of other cell types to maintain the refractive index (Francis *et al.* 1999). The α-, β- and γ-crystallins make up 80-90% of the soluble protein in the lens fibre cell (Hejtmancik 1998). They are crucial in maintaining the transparency and refractive index of the lens to allow the focussing of images onto the retina.

The α-crystallins are members of the small heat shock family of proteins. They are expressed at high levels in the lens and act both as a structural component and as molecular chaperones, helping to maintain the stability of lens proteins and transparency of the lens throughout life (Horwitz *et al.* 1999). While αA-crystallin (CRYAA) is predominantly expressed in the lens, it can be found in trace quantities in other tissues such as brain, liver, spleen, retina and thymus (Vicart *et al.* 1998). αB-crystallin (CRYAB) is also abundantly expressed in lens, but in contrast to CRYAA is also expressed at relatively high levels in cardiac and skeletal muscle, brain, retina and lung (Vicart *et al.* 1998). Lens transparency is maintained by high levels of soluble protein remaining correctly folded and aggregated in solution. However, because proteins in the lens can be as old as the organism itself, a

mechanism is required to maintain their structural integrity and solubility. α-crystallins are believed to play this role. Subunits of both CRYAA and CRYAB aggregate into variable heterogeneous multimeric protein structures with a molecular weight of 300,000 to 1 million Daltons. Each molecule can contain between 15 and 50 subunits. This variability in quaternary structure has impeded attempts to elucidate the structure using common techniques (Horwitz *et al.* 1999). These large proteins are capable of binding irreversibly to unfolded or denatured protein, preventing inappropriate light scattering aggregation of all lens proteins. Whilst crystallins make up the majority of soluble lens fibre cell protein, the trace proteins such as glyceraldehyde-3-phosphate dehydrogenase, enolase, leucine aminopeptidase and aldehyde dehydrogenase are thought to be more susceptible to denaturation (Velasco *et al.* 1997).

Several mutations in the *CRYAA* (Litt *et al.* 1998) and *CRYAB* (Berry *et al.* 2001) genes act in a dominant fashion to cause congenital cataract. The mutations cause a decrease in chaperone activity, and affect the quaternary aggregation of the protein leading to progressive opacification (Kumar *et al.* 1999; Shroff *et al.* 2000). The R116C mutation of CRYAA has been associated with a congenital zonular central nuclear opacity in a large family. The opacity in this family is progressive, with the development of cortical and posterior subcapsular cataract during the 4th decade (Litt *et al.* 1998). Functional studies of this mutation have indicated that the addition of cysteine at this position reduces the chaperone activity of CRYAA. The mutation affected the quaternary structure, with the mutant protein forming an aggregate of twice the size of wildtype protein at physiological temperatures (Shroff *et al.* 2000).

The equivalent residue of cryαB is Arginine-120. The R120G mutation of CRYAB was found in a family with both desmin related myopathy (a condition with weakened cardiac and skeletal muscle) and cataract segregating together (Vicart *et al.* 1998). CRYAB has been shown to interact with cytoskeletal components and may be involved in the remodelling of the cytoskeleton that occurs during development, and also following stresses (Bova *et al.* 1999). The R120G mutation of CRYAB appears to affect is ability to interact with other proteins, in particular the desmin intermediate filaments, causing inappropriate aggregation of this protein (Bova *et al.* 1999). In addition to this, the 450delA mutation of CRYAB has been associated with isolated posterior polar cataract. Unlike the opacity caused by the mutation to CRYAA, this opacity was not progressive (Berry *et al.* 2001). It is hypothesised that this mutation, which results in a shortened protein with 35 novel amino acids at the C terminus may affect both the post-translational modification of the protein, and its chaperone function.

A mutation in *CRYAA* has also been reported to cause autosomal recessive congenital cataract. The W9STOP mutation was identified in the homozygous state in affected members of an inbred Persian family (Pras *et al.* 2000). This stop mutation would result in no expression of CRYAA protein in the cell. Heterozygosity was insufficient to cause disease, suggesting that the short abberant protein is unable to aggregate or precipitate to cause cataract in the way that dominant mutations do. It was shown that a targeted disruption of the *cryaa* gene in mouse caused cataracts with inclusion bodies consisting of CRYAB, suggesting that a mixture of CRYAA and CRYAB is necessary to maintain the soluble quaternary structure. This was supported by *in vitro*

experiments indicating the ratio 3:1 CRYAA to CRYAB gave the highest chaperone activity (Horwitz *et al.* 1999).

The β- and γ-crystallins do not share structural similarities with the α-crystallins, but are expressed at very high levels in the lens fibre cells. They share both tertiary and secondary structure, consisting of two globular domains, each of which contains two Greek-key motifs, named for their similarity to a common element in Greek pottery (Hejtmancik 1998; Slingsby and Clout 1999) (Figure 5.1). Each Greek key motif consists of four anti-parallel β-strands. The two Greek key motifs form a wedge shaped β-sheet sandwich, containing hydrophobic side chains known as the βγ domain fold (Slingsby and Clout 1999). This burying of the hydrophobic side chains contributes significantly to the solubility of these proteins and hence the transparency of the lens (Francis *et al.* 1999).

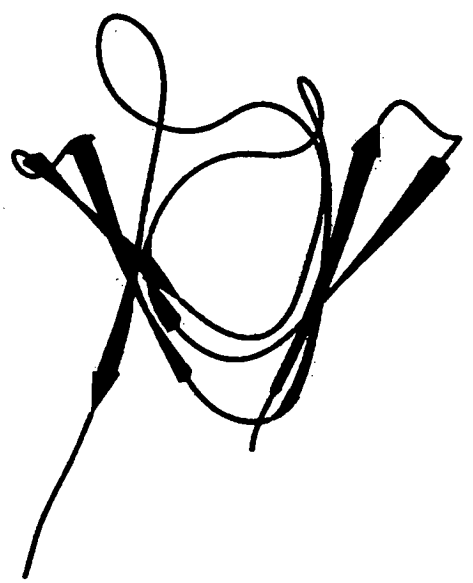


Figure 5.1. βγ domain fold (from Slingsby C, Clout N. J. (1999) Structure of the crystallins, *Eye* 13(3b): 395-402).

β-crystallins have N and C terminal extensions that allow the formation of oligomers not seen in γ-crystallins which remain as compact monomers (Slingsby and Clout 1999). The peptide sequence of the linker region between the two globular domains varies significantly between β and γ crystallins and is also thought to be important in the dimer formation of the β-crystallins (Kannabiran and Balasubramanian, 2000). The β-crystallins are divided into acidic (BA1, BA2, BA3 and BA4) and basic (BB1, BB2 and BB3) groups, based on their charge. The basic proteins have both N and C terminal extensions, whereas the acidic proteins only have an N terminal extension (Slingsby and Clout, 1999). A second copy of CRYBB2 is present in the human genome but is thought to be a pseudogene on the basis of sequence changes at the splice sites compared with functional CRYBB2 (Brakenhoff *et al*, 1992). Each protein is encoded by a separate gene except for BA1 and BA3 which are encoded by the same gene using alternative translation initiation codons. The BA3 polypeptide has an extra 17 residues on the amino terminus (Werten *et al*, 1996). The β-crystallins form homodimers as well as heterodimers of both basic and acidic forms. The largest oligomers are probably octamers (Kannabiran and Balasubramanian, 2000).

Truncation mutations have been reported in three β-crystallin genes. The shortened polypeptides lack the crucial fourth Greek-key motif necessary for correct folding, exposing hydrophobic residues (Slingsby and Clout 1999). The phenotypes associated with β-crystallin gene mutations include the Cerulean and Coppock-like cataract. Both are caused by the same truncating mutation, Q155X, of CRYBB2, which may result in the protein being unable to aggregate into higher order structures (Litt *et al.* 1997; Gill *et al.* 2000). A pulverulent phenotype is also associated with a truncation

mutation, G220X, of CRYBB1 which results in insoluble protein and is probably due to the disruption of the fourth Greek-key motif. This may also be the mechanism involved in the truncation mutation of *CRYBB2* (Mackay *et al.* 2002). A splice site mutation of *CRYBA1/A3* is associated with a zonular sutural cataract (Kannabiran *et al.* 1998). This mutation is predicted to either cause the formation of unstable mRNA or truncate the protein inappropriately.

The lack of terminal extensions and the different linker region in γ-crystallins allow the two domains to interact with each other, rather than with another molecule as the β-crystallins do. This gives the γ-crystallins a very compact structure as monomers (Kannabiran and Balasubramanian, 2000). Seven of the γ-crystallin genes are clustered on chromosome 2q33-q35, and the γS gene is located on chromosome 3 (Santhiya *et al.* 2002). Within the chromosome 2 cluster, *CRYGA* and *CRYGB* are expressed early, and only at low levels. *CRYGC* and *CRYGD* are the most abundant γ-crystallins in the mature lens. *CRYGE, CRYGF* and *CRYGG* are pseudogenes in humans and are not expressed (Francis *et al.* 2000b). Pulverulent phenotypes are associated with two reported mutations of *CRYGC*, T5P and 235insGCGGC, both of which are likely to affect the folding of the protein (Héon *et al.* 1999; Ren *et al.* 2000). The R168W mutation is associated with a lamellar cataract (Santhiya *et al.* 2002). Mutations of *CRYGD* are associated with an aculeiform cataract (R58H) (Héon *et al.* 1999) and a progressive opacity of the embryonic nucleus (R14C) (Stephan *et al.* 1999). The R58H mutation probably introduces hydrogen bonds that affect protein folding (Héon *et al.* 1999), whereas the R14C mutation increases the surface hydrophobicity of the protein and is predicted to allow the formation of

intermolecular disulfide bonds. The mechanism of opacification may be similar to that in age-related cataract (Stephan *et al.* 1999).

The phenotypes resulting from mutations of the crystallin genes indicate their importance in normal vision. Disruption to their structures results in lens opacities leading to severe visual impairment. This is the largest group of proteins shown to cause congenital cataract. At the outset of this study, mutations had only been described in five of these genes: *CRYAA* (Litt *et al.* 1998), *CRYBA1/A3* (Kannabiran *et al.* 1998), *CRYBB2* (Litt *et al.* 1997), *CRYGC* (Héon *et al.* 1999) and *CRYGD* (Héon *et al.* 1999), and were subsequently detected in *CRYAB* (Berry *et al.* 2001) and *CRYBB1* (Mackay *et al.* 2002). Mutations in each of these identified cataract genes have so far been described in only a few pedigrees each and only 16 pedigrees have been identified with crystallin gene mutations. This is indicative of the heterogeneity of the disorder. In order to detect causative mutations in our south-eastern Australian familial paediatric cataract population, large families were assessed by linkage analysis, while the genes known to cause cataract at the beginning of this study were screened in smaller families by SSCP. We investigated crystallin genes known to be associated with non-syndromic congenital or paediatric cataract in 40 pedigrees with autosomal dominant or recessive inheritance.

5.2 Methods

5.2.1 Primer Extension Preamplification

Genomic DNA from all individuals was amplified using the Primer Extension Preamplification method described in Chapter 2.4.

5.2.2 Linkage analysis simulations

Simulations were carried out using the SLINK program (Ott 1989; Weeks *et al.* 1990). A single marker with four equally frequent alleles was simulated with 200 replicates at a recombination fraction of 0.05 for all extended pedigrees to determine those that were suitable for linkage analysis.

5.2.3 Genotyping

Microsatellite markers representing the seven crystallin genes were chosen on the basis of their proximity to the gene and their reported heterozygosities. The closest marker with a heterozygosity >0.7 was chosen (Table 5.1). *CRYAB* was not reported

Table 5.1: Location of markers for linkage analysis of crystallin genes

Gene	Marker	Chromosome	Genetic Position of Marker (cM)[1]	Distance from Gene (Mb)[2]	Distance from Gene (cM)[3]
CRYGD	D2S2358	2	203.4	1.6	1.12
CRYGC	D2S2358	2	203.4	1.6	1.12
CRYAB	D11S1347	11	105.74	0.4	0.56
CRYBA1	D17S841	17	50.74	0.04	0.04
CRYAA	D21S1890	21	52.5	0.2	0.34
CRYBB2	D22S926	22	21.47	0.2	0.28
CRYBB1	D22S926	22	21.47	1.2	1.67

[1] Genetic position obtained from the Marshfield map
[2] Physical distances obtained from the UCSC Human Genome Browser
[3] Distance from gene in cM calculated from region specific recombination rates (Payseur and Nachman 2000)

in the literature as a cause of cataract until after the completion of the screening phase of the study (Berry *et al.* 2001), however a marker from chromosome 11 was included to represent this locus. Mutations of the *CRYBB1* gene (Mackay *et al.* 2002) were also only reported recently however this locus is represented by the same marker as *CRYBB2* as these genes map to the same region. Primer sequences were obtained

from the Genome Database (http://www.gdb.org). The four pedigrees were genotyped at all markers as described in Chapter 2.5. Allele frequencies were estimated from all founders in the pedigrees and are given in Appendix 3.

5.2.4 Linkage analysis

Two-point linkage analysis was carried out in MLINK, part of the FASTLINK package (Lathrop *et al.* 1984, with modification by Cottingham *et al.* 1993). Disease gene frequency was set to 0.0001. Penetrance was set at 0.0 in wild type homozygotes and 0.95 in heterozygotes and mutant homozygotes, indicating highly penetrant dominant inheritance, reflecting the pattern of segregation observed in the pedigrees. LOD scores were calculated at a range of recombination fractions between 0 and 0.5 for each marker. Affected individuals of linked pedigrees were then sequenced at the appropriate candidate genes as described in Chapter 2.7.2.

5.2.5 Single Stranded Conformational Polymorphism Analysis

The proband from each pedigree was screened by SSCP analysis, as described in Chapter 2.6, at coding exons of *CRYAA, CRYBA1/A3, CRYBB2, CRYGC* and *CRYGD*. *CRYAB* and *CRYBB1* had not been linked to congenital cataract at the beginning of this study and were therefore not included. Primer sequences for PCR amplification of all exons are given in Appendix 2. Samples showing altered mobility, together with other family members, were further investigated by direct sequence analysis of genomic DNA using methods described in Chapter 2.7.2. As congenital cataract is a heterogeneous disorder and it was expected that each pedigree would be likely to have a different mutation, unaffected controls were not included.

5.2.6 Restriction digest

Exon 3 of *CRYBA1/A3* was amplified by PCR, as described in Chapter 2.3, from all individuals of pedigree crch08. Ten μl of PCR product were digested with 1U of *NlaIII* (New England Biolabs) at 37°C for 1 hour. Products were electrophoresed on 2% agarose gel and stained with ethidium bromide.

5.2.7 Denaturing High Performance Liquid Chromatography (dHPLC)

Family members of pedigree ctas17 and 50 unaffected unrelated residents of Tasmanian nursing homes were screened for *CRYGD* mutations using dHPLC as described in Chapter 2.8. Individuals from pedigree ctas17 (IV:3. IV:7, V:15 and VI:8) who had been sequenced were used as positive controls. The fragment was injected onto a Varian Helix dHPLC column at 64°C.

5.3 Results

5.3.1 Linkage analysis simulations

Results of the simulations are given in Table 5.2. A recombination fraction of 0.05 was chosen as the markers to be used in the analysis would not be within the candidate gene. Four pedigrees were determined as suitable for linkage analysis; crch13, crch30, crch32 and ctas17 (Figure 5.2). Pedigree crch08 would also have provided sufficient power, however a causative mutation was identified in this pedigree by SSCP prior to this pedigree being typed at the described microsatellite markers. Pedigree ctas16 appears to be quite large (Appendix 1) but only ten samples were available thereby reducing the power to detect linkage.

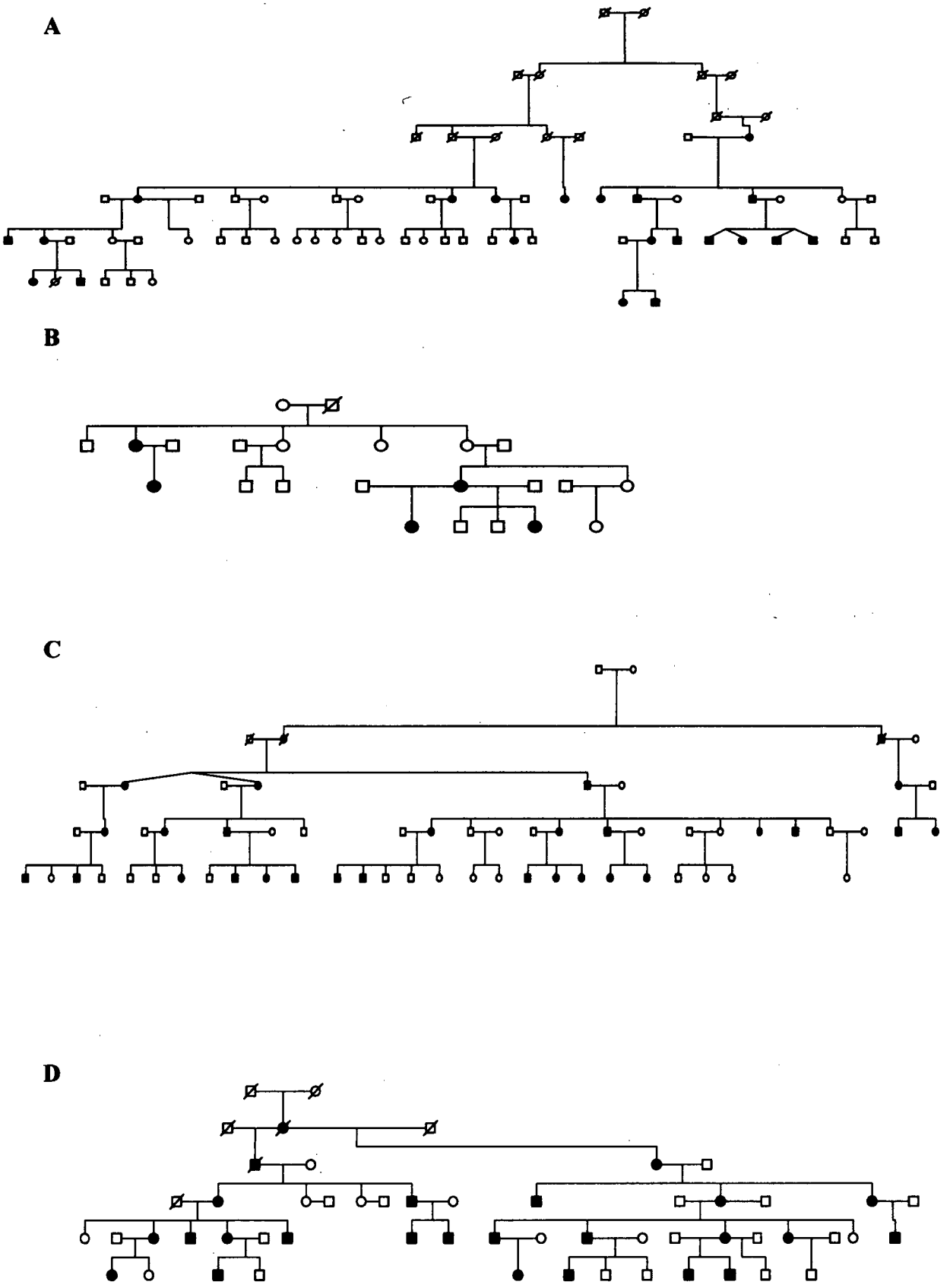


Figure 5.2. Pedigrees used in linkage analysis: **A.** Crch13, **B.** Crch30, **C.** Crch32, **D.** Ctas17.

Table 5.2: Simulated maximum and average LOD scores in pedigrees with autosomal inheritance.

Pedigree	Maximum LOD	Average LOD
CEEH42	0.74	0.26
CRCH08	7.80	4.30
CRCH10	0.65	0.32
CRCH13*	5.07	2.11
CRCH19	1.27	0.42
CRCH23	1.39	0.53
CRCH27	0.95	0.36
CRCH28	0.89	0.33
CRCH30*	2.54	0.91
CRCH32*	9.32	5.89
CTAS16	1.02	0.22
CTAS17*	6.33	3.55

* Pedigrees selected for linkage analysis.

5.3.2 Linkage analysis

Table 5.3 shows the results of the linkage analyses: LOD scores at a range of recombination fractions and the maximum LOD scores. A significant result was obtained for pedigree crch13 at a marker representing the *Connexin 46* (*CX46)* gene and is discussed in Chapter 6. This pedigree was not typed at the remaining loci.

Pedigree crch30 did not provide any evidence for linkage at the γ-crystallin locus (*CRYGC* and *CRYGD*), *CRYAA* nor *CRYBB2*. Results at the two remaining loci were equivocal. This pedigree is relatively small and of borderline power for detecting linkage as indicated by the SLINK simulations (Figure 5.2, Table 5.1). To obtain a significant result at the remaining loci would require typing additional markers that may be more informative in this particular pedigree. Additionally, multipoint analysis and inspection of the haplotypes could provide more information. Due to time

constraints and the identification of other significant results to investigate, this has not been done at this time.

Linkage to the five crystallin loci could be excluded for pedigree crch32. Marker D11S1347 is only excluded to a recombination fraction of 0.002, however, since the *CRYAB* gene is only about 250 kb from this marker (Nov. 2002 assembly), linkage to the gene of interest can be excluded.

A significant result of 3.72 at θ=0.06 from D2S2358 (representing the γ-crystallin locus) was obtained for pedigree ctas17. Following this result, this pedigree was not typed at the remaining loci.

Table 5.3: Two-point LOD scores calculated with founder frequencies.

Pedigree	Gene	Marker	0.0	0.1	0.2	0.3	0.4	max LOD	theta
crch13	CRYGC/GD	D2S2358	-11.53	-2.64	-1.15	-0.44	-0.10	0.00	0.50
	CRYBA1	D17S841	-8.32	-2.89	-1.43	-0.64	-2.0	0.00	0.50
crch30	CRYGC/GD	D2S2358	-8.62	-3.39	-1.78	-0.89	-0.33	0.00	0.50
	CRYAB	D11S1347	-5.25	-1.48	-0.84	-0.46	-0.19	0.00	0.50
	CRYBA1	D17S841	0.10	0.08	0.07	0.05	0.02	0.10	0.00
	CRYAA	D21S1890	-0.69	0.30	0.29	0.18	0.06	0.32	0.14
	CRYBB2	D22S926	-4.10	-1.23	-0.64	-0.35	-0.15	0.00	0.50
crch32	CRYGC/GD	D2S2358	-20.87	-3.39	-1.64	-0.50	-0.02	0.04	0.45
	CRYAB	D11S1347	-5.21	0.53	0.65	0.52	0.29	0.66	0.18
	CRYBA1	D17S841	-7.33	-2.00	-0.72	-0.14	0.06	0.07	0.42
	CRYAA	D21S1890	-6.62	-1.38	-0.22	0.16	0.17	0.20	0.37
	CRYBB2	D22S926	-16.44	-7.70	-1.11	-2.05	-0.78	0.00	0.50
ctas17	CRYGC/GD	D2S2358	2.94	3.61	2.89	1.88	0.80	3.72	0.06
	CRYBA1	D17S841	-11.27	-4.06	-2.10	-1.03	-0.39	0.00	0.50

5.3.3 Mutation identification in pedigree ctas17

The two crystallin genes previously shown to segregate with congenital cataract at this locus, *CRYGC* and *CRYGD*, were sequenced in three affected and one unaffected

individual in pedigree ctas17. Two sequence variants in *CRYGD* were identified. The first was a synonymous C→T substitution at nucleotide 51, the third base of codon 17 encoding tyrosine (Figure 5.3A). This polymorphism does not affect the protein and was previously described by Héon *et al.* (1999). The second was a non-synonymous C→A transversion at the first base of codon 24 (nucleotide 70), resulting in an amino acid substitution from proline to threonine (Figure 5.3B). Both these mutations were then screened in the entire pedigree and 100 normal chromosomes by dHPLC. Four profiles were observed, corresponding to wild type at both loci, heterozygous at base 51, heterozygous at base 70, or heterozygous at both, as determined by sequencing a subset of individuals (Figure 5.4A). The 51C→T polymorphism did not segregate with disease in the pedigree and was found at a prevalence of 25% in controls (Figure 5.4B). The P24T mutation segregated with disease (Figure 5.3C) and was not detected in controls or any other pedigree in the collection (Figure 5.4B).

The phenotype in pedigree ctas17 is quite severe. All affected individuals had surgery prior to age 3 years, with the exception of two patients who had surgery at approximately 7 years of age. Thus the cataract phenotype was only well documented for a few patients. It was described as possessing a flaky nuclear cataract, which looked like silica and was associated with a poor red reflex. Aphakic acuities varied from 6/6 to no perception of light. Thirteen of the 20 patients with documented acuities had at least 6/15 vision in one eye. Nystagmus was rare and only two family members were legally blind (<6/60) (D.A. Mackey, pers. comm.).

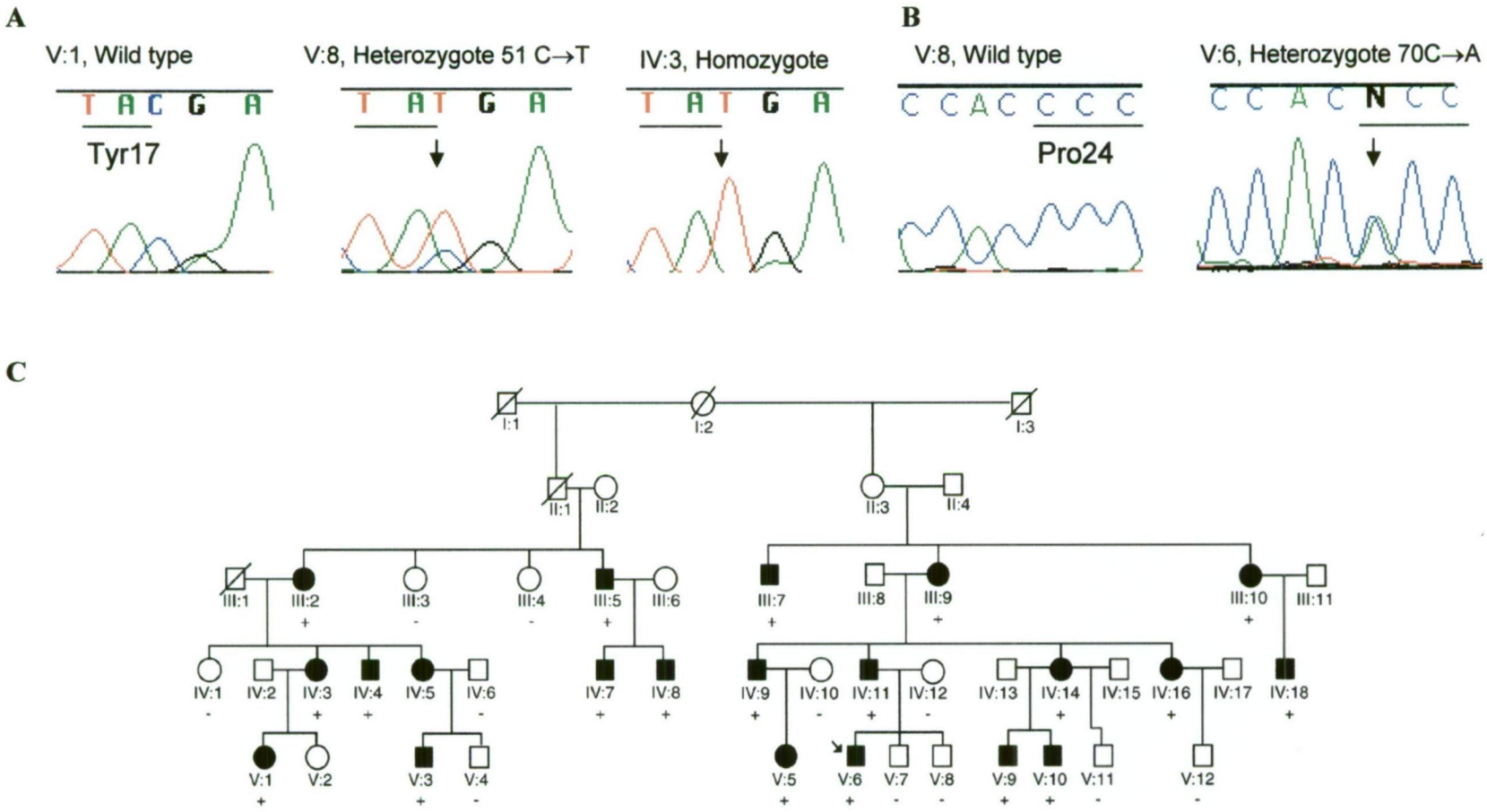
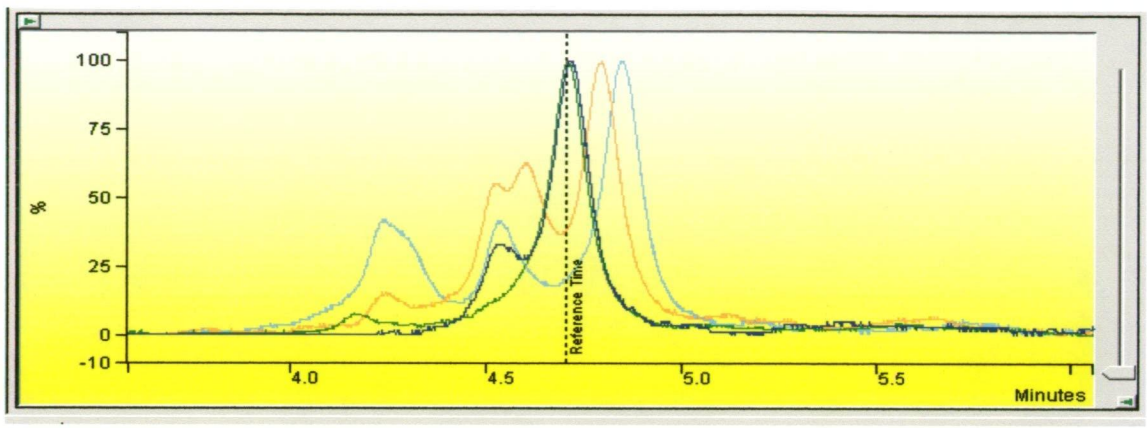


Figure 5.3. A. Sequence variant detected in *CRYGD* at codon 17, nucleotide 51C→T. **B.** Sequence variant detected in *CRYGD* at codon 24 resulting in P24T mutation, nucleotide 70C→A. **C.** Pedigree diagram of ctas17, indicating presence of *CRYGD* P24T mutation. Shaded symbols indicate presence of ophthalmologist-confirmed cataract. '+' indicates heterozygote, '-' indicates wild type. Individuals with no +/- symbol have not been typed. The proband is indicated by an arrow.

A



B

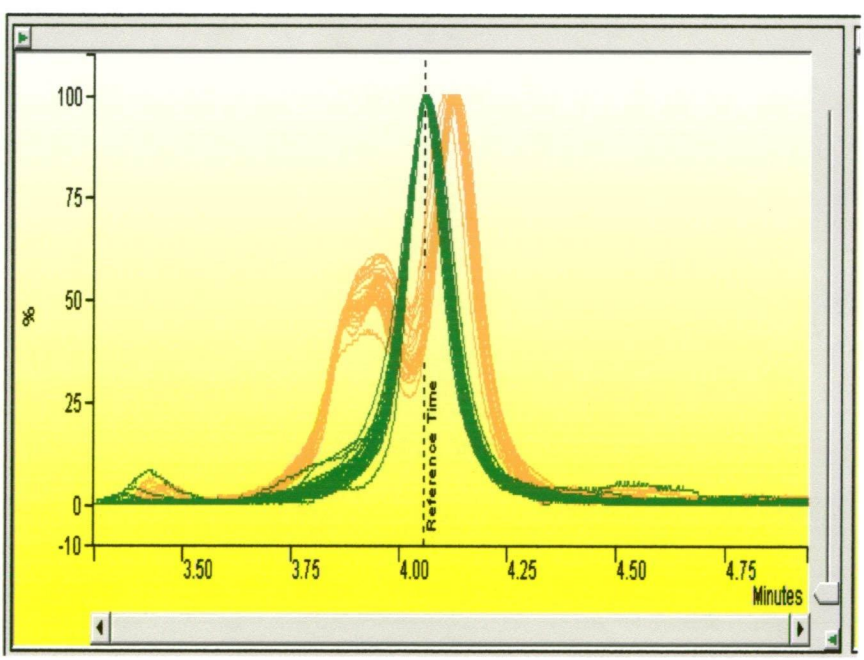


Figure 5.4. dHPLC of *CRYGD* exon 2. **A.** Profiles observed in pedigree ctas17. **B.** Chromatograms of unrelated, unaffected controls. Green = wild type; blue = 70C→A which causes the P24T mutation; orange = C→T polymorphism at nucleotide 51; light blue = C70A mutation and C51T polymorphism.

5.3.4 Single-Stranded Conformational Polymorphism Analysis

Since many of the pedigrees are too small to detect linkage, SSCP was used to screen them for mutations in the five crystallin genes previously found to cause non-syndromic inherited cataract. Probands from the large pedigrees studied by linkage analysis were also included, in case mutations could be detected by this methodology where linkage results were inconclusive. SSCP screening revealed one causative mutation and four polymorphisms (Table 5.4, Figure 5.5). All mutations identified in PEP DNA through SSCP were subsequently confirmed by direct sequencing of genomic DNA, indicating that the use of PEP DNA did not introduce spurious SSCP-positive PCR artifacts. Interestingly, the *CRYGD* P24T mutation detected in pedigree ctas17 following linkage analysis was not detected in the proband by SSCP.

Table 5.4: Variants of crystallin genes detected by SSCP in familial cataract families.

Gene	Exon/ Intron	Pedigree	Position	Base Change	Predicted Effect on Protein
Crystallin BA1/A3	Exon 3	Crch08	+1	G→A	Donor Splice Site
Crystallin BA1/A3	Exon 5	Multiple	Codon 148	C→T	None (Gly)
Crystallin BA1/A3	Intron 2	Crch10	+73	T→G	None
Crystallin BA1/A3	Intron 3	36, 39, 41, 45	+16	C→T	None
Crystallin BB2**	Intron 3	Crch40	+1	G→A	Donor Splice Site

* Intron position numbers refer to the number of nucleotides past the 3' end of the exon.

**The variant observed in *CRYBB2* is most likely in the pseudogene *CRYBB2-2.*

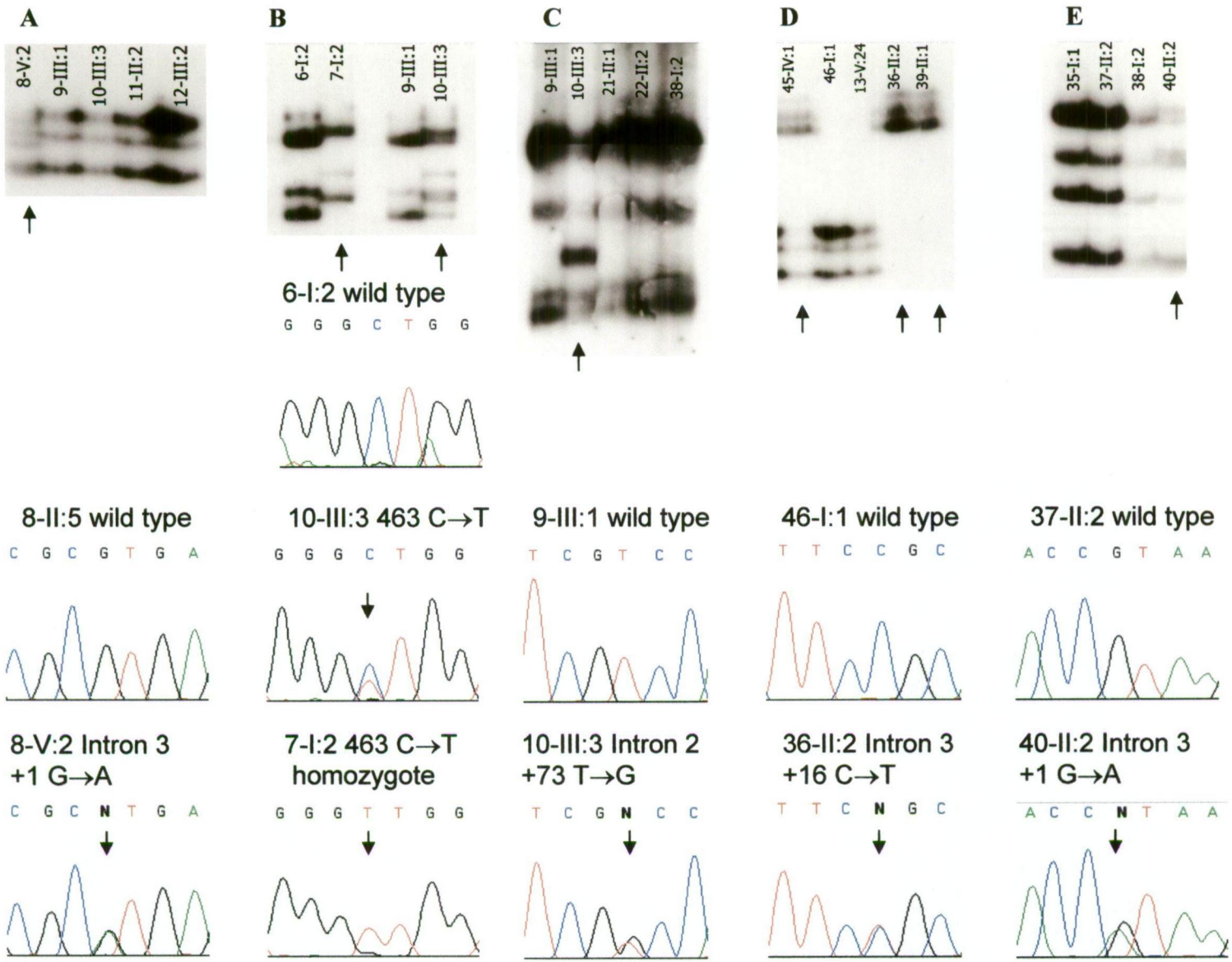


Figure 5.5. SSCP shifts observed (marked with an arrow) and sequencing of appropriate individuals. **A.** *CRYBA1/A3*, exon 3 shift 1. **B.** *CRYBA1/A3* exon 5. **C.** *CRYBA1/A3* exon 2. **D.** *CRYBA1/A3* exon 3 shift 2. **E.** *CRYBB2* exon 3.

A gel shift was observed by SSCP in the *CRYBA1/A3* gene in the large pedigree crch08 prior to linkage analysis being carried out. The clinical features of the phakic patients in this Australian family of European origin showed a consistent spectrum of: 1) Y-sutural opacities, 2) mild opacification throughout the region of the foetal nucleus, and 3) peripheral cortical dot opacities, all with variable severity. Some patients were operated on in childhood, whilst some older mildly affected individuals with good acuity were unaware of the diagnosis prior to our study (D.A. Mackey, pers. comm.). The phenotype is depicted in Figure 5.6A.

The *CRYBA1/A3* gene was sequenced from genomic DNA and a G→A transversion was detected at the intron 3 5' donor splice site (Figure 5.5A). This mutation cosegregates in pedigree crch08 with autosomal dominant congenital cataract (Figure 5.6B) as demonstrated by RFLP analysis. The mutation introduces a restriction site for *NlaIII,* resulting in three fragments in affected heterozygotes although the smallest band was not clear on the gel. Individual III:3 of pedigree crch08 does not have clinically relevant cataract. He was initially diagnosed as unaffected but was reassessed once his status as carrier of the mutation was identified. Mild opacities consistent with his age were noted, but it is not conclusive whether he has the same phenotype as the remainder of the pedigree. This may represent reduced penetrance of the mutation, or a broader spectrum of the clinical features than originally thought.

A silent polymorphism in the third base of codon 148 (C→T) was detected in exon 5 of the *CRYBA1/A3* gene in 5 pedigrees (Table 5.4, Figure 5.5B). Polymorphisms were also detected in introns 2 and 3 of this gene in 1 pedigree each. None segregated with

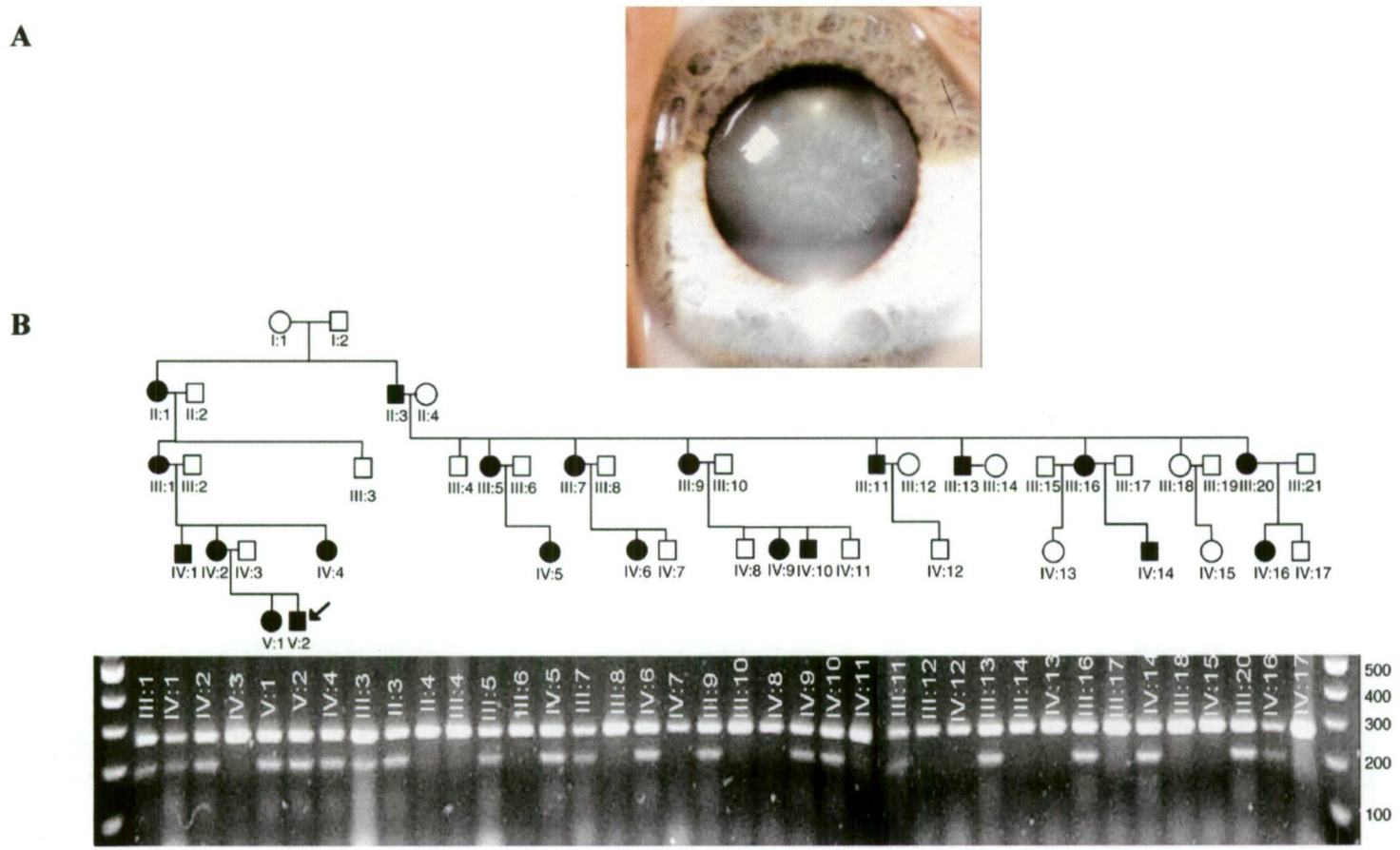


Figure 5.6. A. Photograph of the lens of individual IV:6 of pedigree crch08 showing the opacities of the nucleus and cortex. **B.** Restriction Fragment Length Polymorphism analysis of the mutation detected at the donor splice site of exon 3 of the *CRYBA1/A3* gene in pedigree crch08. Shaded symbols indicate presence of ophthalmologist-confirmed cataract. The proband is indicated with an arrow. Wild type individuals display only the 263bp band, representing the undigested product. Individuals heterozygous for the mutation display the indigested 263bp band and the 172bp representing the digestion product. The remaining 91bp band is not clear on the gel.

disease (Table 5.4, Figure 5.5C & D) as determined by direct sequencing of additional family members.

A G→A transversion was identified at the 5' donor splice site of intron 3 of the *CRYBB2* gene (Figure 5.5E) in a small nuclear pedigree (crch40) with a posterior subcapsular autosomal dominant congenital cataract. However this mutation did not segregate with disease as it was present in the unaffected father and only one of his two affected offspring. The affected mother was wildtype at this nucleotide position (Figure 5.7). This mutation may in fact be in the *CRYBB2* pseudogene fragment *CRYBB2-2*.

5.4 Discussion

5.4.1 Patient ascertainment

The pedigrees studied represent a comprehensive collection of almost all inherited cataract in south-eastern Australia. All patients with cataract requiring surgery in the Australian states of Victoria, Tasmania and southern New South Wales are referred to the Royal Children's Hospital and the Royal Victorian Eye and Ear Hospital, Melbourne, Australia. Only those cases of inherited cataract where none of the family members required specialist medical care would have been overlooked (Wirth *et al.* 2002) although the study did not include sporadic cases that may later be shown to be familial (eg. *de novo* mutations or recessive inheritance). Participation rates were high, and consequently the pedigrees are likely to be representative of familial cataract in south-eastern Australia. This collection is valuable for the determination of the relative contribution of each congenital cataract gene to the overall causes of congenital cataract as well as for the identification of cataract genes.

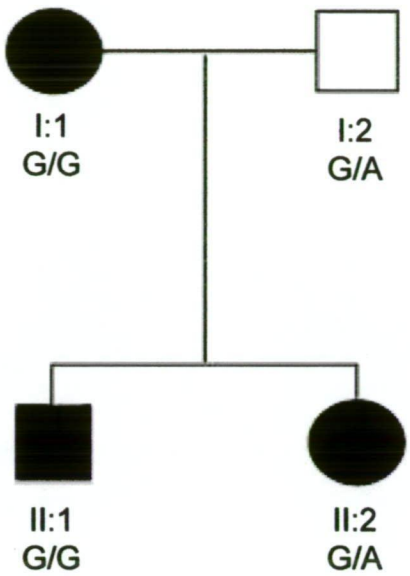


Figure 5.7. Crch40 with genotypes observed at the exon 3 donor splice site of CRYBB2, thought to reflect genotypes of putative pseudogene *CRYBB2-2*.

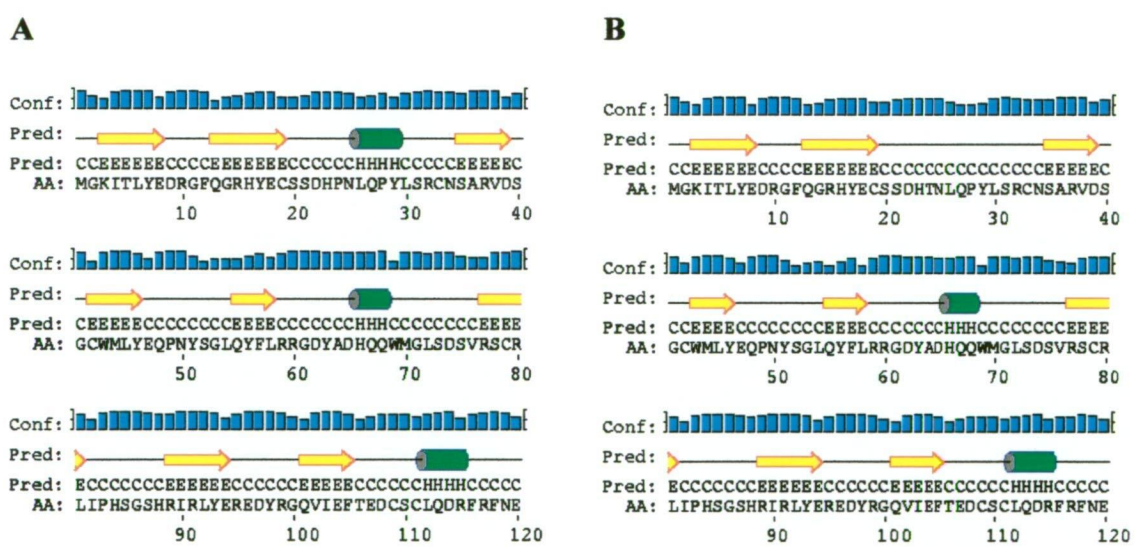


Figure 5.8. PSIPRED predictions of the structure of human CRYGD protein **A.** Wild type **B.** P24T mutation. Yellow arrows represent β-sheet. Green cylinders represent α-helix. The confidence in the prediction is indicated by the height of the blue bars at each amino acid on a scale of 1-10.

5.4.2 Crystallin gene mutations

Mutations in seven crystallin genes (Litt *et al.* 1997; Kannabiran *et al.* 1998; Litt *et al.* 1998; Héon *et al.* 1999; Ren *et al.* 2000; Berry *et al.* 2001; Mackay *et al.* 2002) have been shown to cause monogenic paediatric cataract, although only five (*CRYAA, CRYBA1/A3, CRYBB2, CRYGC, CRYGD*) had been reported at the commencement of this study. Mutations in crystallin genes represent 45% of the reported paediatric cataract mutations to date. We screened all genes by linkage and the five latter genes by SSCP in a familial cataract population from south-eastern Australia, and detected two causative mutations: *CRYGD* P24T and a *CRYBA1/A3* splice site mutation, as well as five non-disease causing polymorphisms.

5.4.3 Mutations of CRYGD

Several other mutations of *CRYGD* have been reported. An R58H mutation has been associated with aculeiform cataract (Héon *et al.* 1999), an R14C mutation with a juvenile onset punctate progressive form (Stephan *et al.* 1999), an R156X mutation with a central nuclear phenotype (Santhiya *et al.* 2002) and R37S was reported as a *de novo* mutation in a boy with a crystalline cataract (Kmoch *et al.* 1999). The P24T mutation identified in ctas17 is associated with a nuclear phenotype and was previously shown to cause lamellar cataract in an Indian pedigree (Santhiya *et al.* 2002). Although *CRYGD* is expressed throughout the lens from the early stages of lens fibre cell differentiation, different mutations of this gene can have remarkably different effects on lens phenotype, and the one mutation (P24T) can cause different phenotypes. Such phenotypic heterogeneity has been reported for the Q155X mutation of *CRYBB2* which has been detected in three pedigrees with different phenotypes. The first family had Cerulean cataract (Litt *et al.* 1997), the second

possessed a Coppock-like phenotype (Gill *et al.* 2000) and the third had sutural cataract with punctate and Cerulean opacities (Vanita *et al.* 2001a).

The Indian study of Santhiya *et al.* (2002) investigated seven small pedigrees with varying types of autosomal dominant congenital cataract and found mutations in the *CRYGD* gene in two of them. In contrast, our study involves 34 autosomal dominant congenital cataract pedigrees, with only one *CRYGD* gene mutation identified. This difference in the prevalence of γ-crystallin mutations may reflect ethnic differences between the south-eastern Australian population, which is predominantly of European descent, and the Indian population, or may have occurred by chance considering the relatively small number of families studied.

5.4.4 Predicted effects of P24T mutation on CRYGD structure and function

P24 is in the third β-sheet of the first Greek-key motif. The structure of the γ-crystallins has been solved by X-ray diffraction, but structural prediction programs such as PSI-PRED (http://bioinf.cs.ucl.ac.uk/psipred) do not predict a β-sheet at this location. PSI-PRED predicts codon 24 to be in a linking region between the second β-sheet and a short α-helix. The substitution of the proline at this location for the mutant threonine removes the prediction of the helix (Figure 5.8). X-ray diffraction of bovine CRYGD has indicated that this region is in fact a short β-sheet (Chirgadze *et al.* 1991). However, most species, including the cow, have a serine at residue 24 of CRYGD whereas all human γ-crystallins have a proline that is the site of the mutation in this pedigree.

Like human CRYGD, bovine γB-crystallin has a proline at this position and X-ray diffraction of this protein shows β-sheet in this location indicating that a wild type proline is compatible with the Greek-key motif formation (Wistow *et al.* 1983). The substitution identified in ctas17 replaces this proline with threonine, a polar amino acid. The wildtype serine in bovine and murine CRYGD forms a hydrogen bond with Asn50, whereas the proline in humans does not. It is possible that the introduction of the polar amino acid in the human CRYGD forms a similar hydrogen bond, which may affect the structure and stability of the protein in the human lens. Similarly the R58H mutation of CRYGD, reported in a pedigree with aculeiform cataract, is predicted to introduce a hydrogen bond that could affect the three-dimensional structure of the protein (Héon *et al.* 1999).

5.4.5 Possible effect of P24T on splicing efficiency

Splicing of exons is a process that is not well understood. It is widely accepted that there are well conserved consensus sequences that indicate exons to the splicing machinery, such as the 5' and 3' splice sites and the intronic branch site. However, while these signals are necessary for correct splicing, they are not sufficient. Recently a group of proteins responsible for the enhancement of exonic splicing have been reported (Cartegni *et al.* 2002). These proteins are rich in serine and arginine and hence are called SR proteins. They bind to consensus sequences within exons known as exonic splicing enhancers (ESEs). Additionally, there are similar proteins that exert a silencing effect although these are even less well characterized (Cartegni *et al.* 2002). The *CRYGD* exon 2 sequence was searched for these ESEs around the position of the mutation identified in ctas17 using the ESE finder (http://exon.cshl.edu/ESE/). Scores above a threshold indicate the presence of a putative ESE. Recognition sites

with scores above the threshold values were identified for proteins SF2/ASF, SC35 and SRp40 in this region of exon 2 of *CRYGD* (Table 5.5). When the P24T mutation was introduced, the two overlapping binding sites for SC35 were lost.

Table 5.5: Scores obtained for SR protein binding sites around the P24T mutation of CRYGD.

Protein Threshold score*	SF2/ASF 1.956	SC35 2.383	Srp40 2.670
Wildtype	3.018	2.714, 2.589**	3.082
Mutant	4.011		4.819

* The threshold score is the score required before a site is deemed to be a putative ESE (Cartegni *et al.* 2002).
** Two overlapping sites were detected for SC35.

The presence of a score above the threshold indicating a consensus binding site does not necessarily indicate an active ESE and the highest score is not necessarily the best ESE (http://exon.cshl.edu/ESE/). The mechanisms of ESE action have not yet been fully elucidated and certain combinations of sites in a particular location may be required. There could also be a silencer that negates the effect of the ESE. Nevertheless, it is possible that the P24T mutation has its effect on the phenotype not by destabilising the protein, but by affecting splicing of the pre-mRNA, resulting in unstable messenger or aberrant protein. Interestingly, the mouse *CRYGD* sequence displays similar consensus sites to the human gene in this area, despite having a serine at codon 24 where humans have a proline. The SRp40 site is in exactly the same position, although the SC35 site is shifted 3bp upstream and the SF2/ASF site is shifted 5bp upstream. Again, it is difficult to determine which of these, if any, are actually having an effect without further studies.

5.4.6 Predicted effects of CRYBA1/A3 splice site mutation

The splice site mutation in intron 3 of the *CRYBA1/A3* gene described here segregates with disease and is likely to be causative. A mutation identical to that found in the present study was previously reported by Kannabiran *et al.* (1998) in a three-generation Indian family with posterior and anterior sutural cataract and not found in a control population of 140 chromosomes. The phenotype in pedigree crch08 appears to match the description of the phenotype in the Indian family. The mutation would affect both the CRYBA1 and CRYBA3 proteins as both are translated from the same mRNA (Werten *et al.* 1996). The same base was reported as substituted for a C in another pedigree with a similar phenotype (Bateman *et al.* 2000).

Mutations of this base are predicted to disrupt the splicing of the mRNA of this gene. There are several possible outcomes. It could cause skipping of exon 3 altogether, with exon 2 splicing directly onto exon 4. Alternatively, the mutation could cause intron 3 to remain in the mRNA or the recruitment of a cryptic splice site. All these possibilities would result in a premature termination of the polypeptide and incomplete folding of the Greek-key motifs or degradation of the unstable mRNA (Kannabiran *et al.* 1998).

The analysis of this region with the ESE finder revealed that the mutation introduces a binding site for the SR protein SF2/ASF. ESEs are by definition in the exons themselves and this splice site mutation is in the intron. SR proteins are thought to act by promoting the binding of proteins involved in splicing, early in spliceosome assembly (Liu *et al.* 1998). Hence, if SF2/ASF were to bind to the newly created recognition site, it may not be able to promote the binding of the splicing machinery to the splice site. ESTs representing the murine homologue of this gene were

expressed in the embryonic eye (Unigene database, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene) hence this protein may be involved in the normal splicing of *CRYBA1/A3* mRNA.

5.4.7 Silent polymorphisms

The silent C→T transition in exon 5 of the *CRYBA1/A3* gene was detected in 13% of the families studied. Population frequencies of this allele were not assessed as it did not appear to cause disease. A silent C→G transversion at this same base was described by Bateman *et al.* (Bateman *et al.* 2000) in a large Brazilian family with congenital cataract. Recently it has been found that many so called 'silent mutations' may actually have an effect on mRNA splicing (Cartegni *et al.* 2002). This region of exon 5 of *CRYBA1/A3* was investigated for ESEs but none were identified in the region. The other two *CRYBA1/A3* polymorphisms detected in our study were intronic, and not predicted to have an effect on the amino acid sequence. All these polymorphisms are likely to occur in more of the pedigrees than the results here suggest, as only one individual from each family was investigated by SSCP.

5.4.8 CRYBB2 splice site mutation likely to be in the pseudogene

The G→A substitution identified at the exon 3 donor splice site of the *CRYBB2* gene in pedigree crch40 would be expected to have a significant effect on the protein in a similar way to the mutation at the same position of the *CRYBA1/A3* gene. However, the mutation does not segregate with disease in this pedigree, indicating that it has no effect on the phenotype (Figure 5.7). Repeat sequencing of this pedigree from independent amplification reactions confirmed this result. Unfortunately independent re-collection of samples from family members was not possible.

It is probable that the mutation observed is actually in a likely pseudogene *CRYBB2-2*. Evidence that this is a pseudogene is based on the lack of signal detected in hybridisation experiments using oligonucleotides specific for both genes, and splice site variants that would severely disrupt the protein sequence (Brakenhoff *et al.* 1992). Since sequence variants have not previously been reported in the *CRYBB2-2* pseudogene in the region of exon 3, it is not possible to determine the extent of amplification of the pseudogene relative to *CRYBB2,* or to design primers specific for the pseudogene. It is expected that the level of amplification of *CRYBB2* and the pseudogene using current primers would be similar. The sequences of the two copies of this gene are almost identical with nucleotide changes observed only in exon 4, following a splice site variant that would inactivate the putative *CRYBB2* product (Brakenhoff *et al.* 1992).

Vanita *et al.* (2001) described mutations of the *CRYBB2* gene in a congenital cataract pedigree that are believed to be due to gene conversion between the gene and the pseudogene, as the mutant *CRYBB2* sequence is identical to that of the pseudogene for a segment of between 9 and 104 bp in the pedigree. In this case the two sequences were distinguished by the use of PCR primers specific for the gene and pseudogene. As mentioned, this approach was not possible in our study due to sequence identity between the gene and the pseudogene in the region of exon 3.

5.4.9 Limitations of SSCP analysis

SSCP analysis does not detect 100% of sequence variants (Hayashi and Yandell 1993). The method relies on the sequence variant causing a change in the three-dimensional conformation of the single stranded DNA fragment that affects the

mobility of the fragment through the gel. Many factors such as the size of the fragment and the 'G+C' content affect its mobility (Hayashi and Yandell 1993; Hayashi 1999). It is difficult to predict the mobility change of a nucleotide substitution. SSCP analysis is most appropriate for small fragments (200-400 bp in length) with a low 'G+C' content. As the length and complexity increase, detection of variants becomes more difficult. Also, variants may only affect the mobility under certain conditions, such as a specific temperature or in the presence of glycerol. SSCP gels must be electrophoresed at cool temperatures to avoid denaturing the single-stranded conformation. All gels were electrophoresed at room temperature with a low power over an extended time period (~16 hours) to prevent overheating above room temperature, but lower temperatures may be required to detect some variants. It is possible that some mutations were undetected by this methodology. In fact this was observed with the P24T variant of *CRYGD* which not detected by SSCP. Conditions were not varied in order to detect it, as it was readily detectable by denaturing High Performance Liquid Chromotography for high throughput screening of the pedigree and controls. The only way to detect 100% of sequence variants is through direct sequencing. However, SSCP was considered an appropriately rapid, relatively efficient, and less expensive method for initial screening of these pedigrees. It was decided to undertake linkage analysis in large pedigrees at these loci in addition to SSCP because of the known limitations of the method. It is possible that mutations in the smaller pedigrees not investigated by linkage analysis were missed, however, linkage is not a valid approach in these small heterogeneous pedigrees. Linkage analysis is valuable when sufficient power is present, but otherwise it is necessary to rely on direct methods of mutation detection such as SSCP or direct sequencing in order to conclusively identify disease loci.

5.4.10 Other congenital cataract loci

The other eight known cataract genes (*CX46, CX50, MIP, BFSP2, PITX3, CRYAB, CRYBB1* and the recently reported *HSF4*) remain to be screened in the smaller pedigrees from this population to evaluate their contribution to this phenotype. The large pedigrees used here for linkage analysis were investigated at these loci and the analyses are described in Chapter 6. The gene for the membrane protein LIM2 has recently been associated with recessive congenital cataract (Pras *et al.* 2002) and two of our families displayed recessive inheritance. In addition, eight loci (at 1pter-p36.1 (Eiberg *et al.* 1995; Ionides *et al.* 1997), 3p (Pras *et al.* 2001), 9q13-q22 (Héon *et al.* 2001), 15q21-22 (Vanita *et al.* 2001b), 17p13 (Berry *et al.* 1996; Ionides *et al.* 1998), 17q24 (Armitage *et al.* 1995) and 20p12-q12 (Yamada *et al.* 2000) have shown linkage to autosomal dominant and recessive congenital cataract, but the genes at these loci have not yet been elucidated. It is likely that there are additional as-yet unmapped cataract loci.

5.4.11 Summary

The high degree of genetic and phenotypic heterogeneity observed in congenital and paediatric cataract significantly complicates investigations of this disorder. In Australian pedigrees of European ancestry, we have detected two causative crystallin mutations, both described previously in Indian pedigrees. Considering the different ethnic origins of the two populations, these mutations are likely to have arisen independently (although this is not confirmed). This is only the third report describing cataract mutations common to more than one family, indicating the critical nature of these residues to lens stability. Other than a non-disease causing splice site mutation

believed to be present in a pseudogene, we did not detect a single protein-altering polymorphism in the five crystallin genes, highlighting the conserved nature of these highly stable proteins. The low prevalence of causative mutations in the five crystallin genes studied here – only two pedigrees of 38 (5.3%) – indicates that these genes are not a common cause of paediatric cataract in this population.

Chapter 6

Investigation of Known Cataract Loci in Four Large Pedigrees

6.1 Background

The investigation of crystallin genes in this population revealed only two causative mutations in 38 pedigrees. While it is possible that mutations were missed in the smaller pedigrees due to the limitations of SSCP, this number seems particularly low given the prevalence of mutations within these genes in the cataract literature, and suggests that other loci play an important role in cataract susceptibility in this population. The other loci previously associated with congenital cataract were investigated. As discussed, SSCP is time consuming, so investigations began with linkage analysis in a group of four pedigrees large enough to provide a chance of detecting significant linkage: crch13, crch30, crch32 and ctas17. All reported ADCC loci were investigated using linkage analysis, apart from *PAX6* as there were no anterior segment anomalies in any of the four large pedigrees. However, anterior segment anomalies were detected in the smaller pedigrees crch01 (aniridia) and crch06 and crch22 (anterior cataract). These pedigrees are undergoing analysis at the laboratory of Professor Veronica van Heyningen (Western General Hospital, Edinburgh) at the *PAX6* locus although they were still screened for crystallin genes mutations using SSCP as discussed in the previous chapter.

6.2 Known genes

The genes studied include the two connexin genes, Connexin 46 (*CX46)* and Connexin 50 (*CX50*), Major Intrinsic Protein of the lens (*MIP*), Beaded Filament

Structural Protein 2 (*BFSP2*), Heat Shock Factor 4 (*HSF4*) and *PITX3*. Mutations in each of these genes have been found in congenital cataract pedigrees and are all listed in Table 1.1.

6.2.1 Connexins

Two connexins, *CX46* and *CX50,* are expressed in lens fibre cells and mutations in both have been linked to zonular pulverulent congenital cataract. Connexin subunits associate into hexamers to form connexons which bind to a connexon in the neighbouring cell membrane to form a channel for energy independent transport of small molecules (Francis *et al.* 1999; Kannabiran and Balasubramanian 2000). All vertebrate connexins have an asparagine at codon 61, 62 or 63 which is within the first extracellular loop of the protein. A family with zonular pulverulent cataract was shown to have a mutation resulting in the substitution of N63 with serine in *CX46*. While this is considered a conservative substitution since both residues are neutral under physiological conditions, the lack of this important asparagine is thought to interrupt the docking of the connexon hexamer with its counterpart on the membrane of the neighbouring cell (Mackay *et al.* 1999). Similarly, the P187L mutation of *CX46*, identified in another pedigree (Rees *et al.* 2000), is situated in the highly conserved Pro-Cys-Pro motif of the second extracellular loop. Again this mutation appears to be reasonably conservative and is predicted to affect protein docking. A third pedigree has been identified with a mutation in the *CX46* gene. The insertion of a cytosine after coding nucleotide 1137 (1137insC) was found to segregate with disease (Mackay *et al.* 1999). This insertion causes a frame shift after codon 379, possibly resulting in replacement of the final 56 amino acids and addition of 31 amino acids also incorporated into the chain, severely affecting the structure of the protein.

Similar mutations have been identified in *CX50* in pedigrees with zonular pulverulent congenital cataract. The E48K mutation is in the conserved first extracellular loop and is also predicted to affect protein docking (Berry *et al.* 1999). The P88S mutation is in the second transmembrane domain (Shiels *et al.* 1998) and the I247M mutation is in the intracellular C-terminal tail of the protein (Polyakov *et al.* 2001).

6.2.2 MIP

Another protein involved in cell homeostasis is the Major Intrinsic Protein of the lens (MIP). This protein is a member of the aquaporin family (AQ0) of transmembrane water channels (Berry *et al.* 2000). Normal MIP proteins form homo-tetramers in the cell membrane and allow the selective transport of water between cells in an energy independent manner. A second role as a cell to cell adhesion molecule has also been proposed (Francis *et al.* 2000b). MIP represents greater than 50% of total membrane protein, which suggests it has a structural role (King and Yasui 2002). Two mutations have been identified in this gene in families with congenital cataract (Berry *et al.* 2000). Both the E134G and T138R mutations have been shown to affect water transport through cell membranes by preventing trafficking of the protein to the plasma membrane (Francis *et al.* 2000b). Additionally, when expressed with wild type protein, the mutants interfere with the function of the normal protein, in a dominant negative effect (Francis *et al.* 2000b). The phenotypes in the two families are quite different, even though the effects on the protein are similar. The T138R mutation causes a polymorphic progressive phenotype with punctate opacities in the mid and peripheral lamellae as well as anterior and posterior polar opacities, while the

E134G mutation causes a stable lamellar cataract in fibres developing during the late foetal period (Francis *et al.* 2000a).

6.2.3 BFSP2

Some inherited cataracts are caused by mutations of the lens specific cytoskeletal protein Beaded Filament Structural Protein-2 (BFSP2) (Conley *et al.* 2000; Jakobs *et al.* 2000). This protein is also known as phakinin, CP47 or CP49 and assembles with a second cytoskeletal protein, filensin, to form the beaded filament of the lens. An R287W mutation segregates with disease in a large family with juvenile-onset progressive cataract (Conley *et al.* 2000). The phenotype in older members of this family was a lamellar cataract. The positive charge of the arginine at residue 287 is well conserved amongst proteins of this class, therefore the authors believe this mutation to be causative of the progressive cataract in this family. They also speculate about the possible effect of mild mutations in this gene on later onset age-related cataract. The age of onset in this family ranged from 9 years to the late twenties. A second mutation, ΔE233 was found in a family with a congenital cataract consisting of spoke-like anterior and posterior sub-capsular cortical opacities (Jakobs *et al.* 2000). The deletion is predicted to disrupt a hydrophobic stripe along an α-helix that is involved in an interaction with filensin. This hydrophobic stripe is highly conserved amongst other members of this protein family. From these two identified mutations, it appears that *BFSP2* plays an important structural role both during early development (ΔE233) and later growth of the lens (R287W).

6.2.4 HSF4

Marner's cataract has been known to map to 16q22 for some time (Eiberg *et al.* 1988), however the gene was only described recently. Following the refinement of the

critical region through the analysis of a Chinese pedigree with lamellar cataract, a mutation was discovered in the gene encoding Heat Shock Factor 4 (*HSF4*) (Bu *et al.* 2002). This gene controls the expression of heat shock proteins via both activating and inhibitory activities. Subsequently, a mutation was identified in a pedigree with Marner's cataract which is characterised by zonular stellate and anterior polar opacities (Bu *et al.* 2002). Several other mutations have been described in other cases of lamellar cataract (Bu *et al.* 2002). The effects of these mutations on HSF4 function are currently unknown.

6.2.5 PITX3

PITX3 is a homeobox protein involved in eye development. Several mutations in *PITX3* have been found to cause cataracts. The first – a missense mutation S13N – was found at a highly conserved serine in a patient with congenital cataract and her affected mother (Semina *et al.* 1998). The second – a 17bp insertion – has been identified in a four-generation pedigree with congenital posterior polar cataract (Finzi *et al.* 2002). This mutation was originally described in a pedigree with both anterior segment mesenchymal dysgenesis and congenital cataract (Semina *et al.* 1998). It is possible that this mutation is responsible for only the congenital cataract part of the phenotype in this family and that a second modifying gene also segregating in this family is the cause of the anterior segment mesenchymal dysgenesis. This is further supported by a mutation, 550delG, in another family with isolated congenital cataract (Yang *et al.* 2002). The 17bp insertion results in a frameshift causing the alteration of the 82 C-terminal amino acids, while the 550delG single base deletion results in 86 novel amino acids.

6.3 Other mapped loci

Several other loci at which the genes have not been identified have also been reported. Two phenotypes have been reported to map to the short arm of chromosome 1. The Volkmann cataract consists of a progressive central opacity of both the embryonic and foetal nuclei, as well as the anterior and posterior Y-sutures (Eiberg *et al.* 1995). The gene has been localised to between 1pter and marker D1S243 following a genome-wide scan in the Volkmann cataract pedigree. A posterior polar cataract maps to an overlapping region (Ionides *et al.* 1997). The clinical characteristics of these two phenotypes are quite distinct. The posterior polar cataract is not progressive and does not involve the nucleus or sutures. While mapping information suggests the two phenotypes may be allelic, there may also be two distinct genes in the region responsible for cataract.

A unique phenotype of central pouch-like cataract with sutural opacities has been found to link to 15q21-q22 in an Indian pedigree (Vanita *et al.* 2001a). This phenotype was identified in a single seven-generation Indian pedigree, and localised by recombinations to a small region of 1.2 cM between markers D15S117 and D15S1033.

Anterior polar cataract has been linked to 17p13 in a four-generation British pedigree (Berry *et al.* 1996). The disease gene was found to map to a 13cM region between D17S849 and D17S796. There are no other reports of cataract mapping to this region.

The distinctive Cerulean cataract has been shown to be caused by a mutation in *CRYBB2* on chromosome 22. However, this phenotype also maps to 17q24 (Armitage *et al.* 1995). A genome-wide scan in a five-generation pedigree in which candidate

loci had been excluded revealed linkage to a 6 cM region between markers D17S802 and D17S836. This locus was actually identified prior to the *CRYBB2* locus, however the gene remains elusive.

Posterior polar cataract has been mapped to 20p12-q12 in a genome-wide scan of a Japanese pedigree (Yamada *et al.* 2000). The region was localised by recombination events to between D20S851 and D20S96. The *BFSP1* gene encoding the lens cytoskeletal protein filensin is located in this region as was considered to be an appropriate candidate gene, however no coding mutations were identified in the pedigree (Yamada *et al.* 2000).

6.4 Aims

We aimed to identify linkage to these known loci through the analysis of the four large pedigrees described in Chapter 5, and determine the causative mutations at any linked loci where cataract genes have been identified.

6.5 Methods

6.5.1 Genotyping

Markers were chosen on the basis of their proximity to the target gene or region using physical maps and their reported heterogeneity. The marker closest to the gene with a heterozygosity of >70% was chosen (Table 6.1A). Markers reported in the literature as the peak or flanking markers for reported loci were chosen (Table 6.1B). Primer sequences for all markers were obtained from the Genome Database (http://www.gdb.org). The four large pedigrees described in Chapter 5 were genotyped and analysed by two-point linkage analysis carried out in MLINK, part of

the FASTLINK package (Lathrop *et al.* 1984, with modification by Cottingham *et al.* 1993). Disease gene frequency was set to 0.0001. Penetrance was set at 0.0 in wild type homozygotes and 0.95 in heterozygotes and mutant homozygotes, indicating highly penetrant dominant inheritance, reflecting the pattern of segregation in the pedigrees. LOD scores were calculated at a range of recombination fractions between 0 and 0.5 for each marker. Allele frequencies were estimated from all founders in the pedigrees and are given in Appendix 3. Where less than four pedigrees were typed, 24 unrelated unaffected individuals collected from nursing homes in Launceston, Tasmania were typed to estimate allele frequencies.

Table 6.1: Location of markers for linkage analysis. Genetic positions are from the Marshfield map. Physical distances are calculated from the centre of the gene.

A. Candidate genes

Gene	Marker	Chromosome	Genetic Position of Marker (cM)[1]	Distance from gene (Mbp)[2]	Distance from gene (cM)[3]
CX50	D1S2635	1	165.62	11.5	12.40
BFSP2	D3S1349	3	Not mapped genetically	0.1	0.12
PITX3	D10S1268	10	160.11	1.6	1.52
MIP	D12S83	12	75.17	4.0	3.32
CX46	D13S1236	13	2.77	1.9	4.1
HSF4	D16S496	16	85.94	1.7	1.45

B. Reported loci

Locus	Marker	Chromosome	Genetic Position of Marker (cM)[1]	Distance between markers (Mbp)[2]	Distance between markers (cM)[1]	Reference
1pter-p36	D1S243	1	0.00	1.4	4.22	Eiberg *et al.* 1995 Ionides *et al.* 1997
	D1S468	1	4.22			
15q21-q22	D15S117	15	51.21	0.3	1.12	Vanita *et al.* 2001a
	D15S1033	15	52.33			
17p13	D17S849	17	0.63	6.0	14.06	Berry *et al.* 1996
	D17S796	17	14.69			
17q24	D17S802	17	106.80	1.1	6.12	Armitage *et al. 1995*
	D17S836	17	112.92			
20p12-q12	D20S894	20	30.56			Yamada *et al.* 2000

[1] Genetic positions obtained from the Marshfield map
[2] Physical distances obtained from the UCSC Human Genome Browser
[3] Distance between marker and candidate gene in cM calculated from region specific recombination rates (Payseur and Nachman 2000)

Allele frequencies reported by the Centre d'Etude Polymorphisme Humain (CEPH) database (http://www.cephb.fr/) were also used and are given in Appendix 3. These frequencies are calculated from around 50 chromosomes or less (as are the founder frequencies calculated here). However, they are generally accepted as representative of the Caucasian population. LOD scores were also calculated using these frequencies to help assess possible inaccuracies in the founder frequencies. Selected CEPH individuals reported to be heterozygous (http://www.ceph.fr) were included during genotyping as controls in order to directly compare CEPH frequencies and allele sizes with observed sizes in this data set, as the size of fragments detected can vary between systems. LOD scores were calculated at a range of recombination fractions between 0 and 0.5 for each marker with both sets of frequencies. Haplotypes for pedigree crch30 at chromosome 15 were estimated using Genehunter Plus.

6.5.2 Sequencing

The *CX46* gene was sequenced in individuals from pedigree crch13 as described in Chapter 2.7.2. Primers were designed to sequence the one large exon in three overlapping fragments and are presented in Appendix 2. Fragment 1b was designed as a smaller product to screen 50 unaffected unrelated controls collected from nursing homes in Launceston, Tasmania for the identified mutation. Probands from all other pedigrees in the collection were also sequenced for this mutation using fragment 1b.

6.5.3 Denaturing High Performance Liquid Chromatography

Fragment 1b was amplified and analysed by dHPLC in pedigree crch13 as described in Chapter 2.8. Samples were injected at 64°C.

6.6 Results

6.6.1 Linkage analysis

Maximum LOD scores obtained with both sets of allele frequencies are presented in Tables 6.2 (founders) and 6.3 (CEPH). Using the CEPH frequencies, D13S1236, the marker representing *CX46,* produced a significant LOD score of 4.37 (θ=0.03) in pedigree crch13. The founder allele frequencies gave a LOD score of 2.96 (θ=0.04) at this locus. Several individuals failed to genotype at this marker, however, as a significant LOD score was obtained (followed up by positive mutation identification described below), they were not repeated. A LOD score of 1.36 (θ=0.08) was also obtained at D16S496 using founder allele frequencies. This pedigree was not typed at additional markers following the significant result at *CX46.*

Marker D15S1033 within the 15q21-q22 locus gave LOD scores of 1.5 and 1.7 (θ=0) for pedigree crch30 using founder and CEPH frequencies respectively. All other results were equivocal, although generally not indicative of linkage. While the majority of LOD scores at θ=0 were negative, the distances over which markers are excuded from linkage to the disease are small as would be expected.

Most markers were excluded from linkage in crch32. D15S1033 gave a maximum LOD score of 1.0 (θ=0.26) with both frequency sets and D10S1268 gave a maximum LOD score of 0.49 (θ=0) with founder frequencies.

Pedigree ctas17 was typed at several markers prior to the identification of the *CRYGD* mutation discussed in Chapter 5. Once the *CRYGD* mutation was identified, no

Table 6.2. Two-point LOD scores calculated with founder frequencies.

Pedigree	Locus	Marker	0.0	0.05	0.1	0.2	0.3	0.4	max LOD	theta
Crch13	1pter-p36.1	D1S243	-10.28	-3.85	-2.55	-1.29	-0.63	-0.24	0.00	0.50
		D1S468	-21.72	-9.01	-6.04	-3.06	-1.48	-0.54	0.00	0.50
	CX50	D1S2635	-7.47	-2.62	-1.42	-0.37	0.00	0.04	0.09	0.39
	CX46	D13S1236	2.74	2.94	2.74	2.11	1.37	0.57	**2.96**	**0.04**
	15q21-q22	D15S117	-18.03	-6.73	-5.26	-2.82	-1.35	-0.52	0.00	0.50
		D15S1033	-14.74	-6.26	-4.00	-1.83	-0.81	-2.28	0.00	0.50
	HSF4	D16S496	1.04	1.33	1.36	1.13	0.72	0.24	**1.36**	**0.08**
	17p13	D17S796	-12.94	-4.37	-2.80	-1.22	-0.44	-0.08	0.00	0.50
		D17S849	-18.57	-8.92	-6.08	-3.01	-1.37	-0.45	0.00	0.50
	17q24	D17S836	-5.54	-1.96	-1.28	-0.61	-0.29	-0.15	0.00	0.50
		D17S802	-4.33	-1.78	-1.04	-0.36	-0.09	0.00	0.00	0.45
Crch30	1pter-p36.1	D1S243	-2.05	-0.63	-0.18	0.17	0.22	0.14	0.22	0.27
		D1S468	-1.65	-1.10	-0.79	-0.42	-0.20	-0.07	0.00	0.50
	CX50	D1S2635	-2.46	-1.12	-0.73	-0.31	-0.09	0.00	0.01	0.44
	BFSP2	D3S1349	-7.80	-3.89	-2.60	-1.33	-0.70	-0.28	0.00	0.50
	PITX3	D10S1268	-6.56	-1.54	-0.89	-0.29	-0.04	0.04	0.04	0.42
	MIP	D12S83	-7.23	-1.65	-1.02	-0.49	-0.29	-0.17	0.00	0.50
	CX46	D13S1236	0.02	0.03	0.04	0.04	0.04	0.02	0.04	0.20
	15q21-q22	D15S117	-1.72	-0.07	-0.42	-0.86	-0.02	0.02	0.02	0.42
		D15S1033	1.54	1.39	1.24	0.91	0.59	0.28	**1.54**	**0.00**
	HSF4	D16S496	-0.91	-0.36	0.18	0.33	0.31	0.20	0.33	0.23
	17p13	D17S796	-6.55	-2.36	-1.51	-0.69	-0.28	-0.22	0.00	0.50
		D17S849	-3.97	-2.33	-1.69	-0.96	-0.52	-0.22	0.00	0.50
	17q24	D17S836	-3.50	-1.59	-1.04	-0.48	-0.19	-0.05	0.00	0.50
		D17S802	-5.54	-1.87	-0.96	-0.18	0.11	0.15	0.15	0.36
	20p12-q12	D20S894	-6.17	-2.05	-0.17	-0.48	-0.11	0.03	0.04	0.43
Crch32	1pter-p36.1	D1S243	-27.82	-10.76	-6.99	-3.38	-1.55	-0.52	0.00	0.50
		D1S468	-19.30	-5.13	-2.79	-0.83	-0.11	-0.08	0.08	0.41
	CX50	D1S2635	-13.24	-5.01	-3.08	-1.27	-0.46	-0.09	0.00	0.49
	BFSP2	D3S1349	-15.46	-5.91	-2.60	-0.19	0.35	0.37	0.41	0.35
	PITX3	D10S1268	0.49	0.41	0.37	0.20	0.09	0.04	0.49	0.00
	MIP	D12S83	-16.65	-6.00	-3.51	-1.32	-0.42	-0.08	0.00	0.50
	CX46	D13S1236	-5.71	-1.27	-0.48	0.17	0.34	0.26	0.34	0.31
	15q21-q22	D15S117	-10.92	-3.10	-1.95	-0.86	-0.35	-0.92	0.00	0.50
		D15S1033	-12.53	-2.07	-0.38	0.83	0.98	0.59	**1.00**	**0.26**
	HSF4	D16S496	-7.14	-3.83	-2.59	-1.28	-0.59	-0.19	0.00	0.50
	17p13	D17S796	-14.55	-5.77	-3.68	-1.69	-0.73	-0.22	0.00	0.50
		D17S849	-15.33	-6.81	-4.63	-2.43	-1.24	-0.49	0.00	0.50
	17q24	D17S836	-12.99	-5.59	-3.28	-1.15	-0.28	0.02	0.04	0.44
		D17S802	-19.83	-10.23	-7.06	-3.81	-2.02	-0.81	0.00	0.50
	20p12-q12	D20S894	-21.23	-8.54	-5.71	-2.89	-1.48	-0.64	0.00	0.50
Ctas17	1pter-p36.1	D1S243	-6.51	-1.42	-0.69	-0.15	-0.26	-0.01	0.00	0.50
		D1S468	-17.07	-7.50	-5.07	-2.48	-1.13	-0.37	0.00	0.50
	CX50	D1S2635	-12.81	-4.87	-3.10	-1.45	-0.64	-0.21	0.00	0.50
	CX46	D13S1236	-3.62	0.09	0.51	0.65	0.49	0.24	0.66	0.18
	15q21-q22	D15S117	-17.56	-5.58	-3.53	-1.59	-0.66	-0.19	0.00	0.50
		D15S1033	-9.78	-2.67	-1.49	-0.48	-0.08	0.05	0.05	0.42
	HSF4	D16S496	-4.97	-1.94	-1.08	-0.27	0.06	0.13	0.14	0.38
	17p13	D17S796	-8.24	-2.82	-1.73	-0.71	-0.26	-0.05	0.00	0.50
		D17S849	-12.68	-5.44	-3.39	-1.48	-0.59	-0.16	0.00	0.50
	17q24	D17S836	-3.42	-1.98	-1.10	-0.30	0.24	0.11	0.11	0.39
		D17S802	-10.88	-4.41	-2.75	-0.23	-0.54	-0.19	0.00	0.50

Table 6.3. Two-point LOD scores calculated with CEPH frequencies.

Pedigree	Locus	Marker	0.0	0.05	0.1	0.2	0.3	0.4	max LOD	theta
Crch13	1pter-p36.1	D1S243	-13.72	-4.29	-2.77	-1.37	-0.66	-0.25	0.00	0.50
		D1S468	-21.72	-8.97	-5.99	-3.03	-1.47	-0.54	0.00	0.50
	CX50	D1S2635	-8.03	-2.38	-1.22	-0.23	0.09	0.12	0.13	0.36
	CX46	D13S1236	4.21	4.31	3.99	3.06	1.97	0.85	**4.37**	**0.03**
	15q21-q22	D15S117	-17.90	-6.71	-5.12	-2.64	-1.16	-0.40	0.00	0.50
		D15S1033	-22.43	-7.14	-4.65	-2.22	-1.06	-0.40	0.00	0.50
	HSF4	D16S496	0.73	1.09	1.12	0.89	0.53	0.17	**1.12**	0.08
	17p13	D17S796	-19.88	-4.75	-2.65	-0.82	-0.12	0.07	0.07	0.42
		D17S849	-18.57	-8.92	-6.08	-3.01	-1.37	-0.45	0.00	0.50
	17q24	D17S836	-12.24	-3.76	-2.43	-1.23	-0.68	-0.33	0.00	0.50
		D17S802	-5.35	-0.83	-0.16	0.36	0.43	0.28	0.43	0.27
Crch30	1pter-p36.1	D1S243	-7.67	-0.74	-0.27	0.11	0.19	0.13	0.19	0.30
		D1S468	-1.65	-1.10	-0.79	-0.42	-0.20	-0.07	0.00	0.50
	CX50	D1S2635	-3.96	-1.21	-0.80	-0.37	-0.13	-0.02	0.00	0.48
	BFSP2	D3S1349	-7.80	-3.88	-2.59	-1.33	-0.68	-0.28	0.00	0.50
	PITX3	D10S1268	-6.68	-1.76	-1.09	-0.44	-0.14	-0.01	0.01	0.46
	MIP	D12S83	-7.29	-1.64	-0.98	-0.42	-0.20	-0.11	0.00	0.50
	CX46	D13S1236	0.02	0.03	0.04	0.04	0.04	0.02	0.04	0.20
	15q21-q22	D15S117	-2.43	-1.11	-0.69	-0.28	-0.09	-0.01	0.00	0.46
		D15S1033	1.71	1.54	0.38	1.03	0.67	0.32	**1.71**	**0.00**
	HSF4	D16S496	0.48	0.52	0.53	0.49	0.38	0.22	0.53	0.10
	17p13	D17S796	-11.42	-2.70	-1.66	-0.72	-0.27	-0.06	0.00	0.49
		D17S849	-3.86	-2.27	-1.65	-0.94	-0.51	-0.22	0.00	0.50
	17q24	D17S836	-3.80	-1.76	-1.14	-0.52	-0.21	-0.05	0.00	0.50
		D17S802	-10.76	-1.94	-0.99	-0.18	0.11	0.15	0.16	0.37
	20p12-q12	D20S894	-5.93	-1.86	-1.12	-0.38	-0.06	0.05	0.05	0.41
Crch32	1pter-p36.1	D1S243	-33.84	-11.19	-7.23	-3.49	-1.60	-0.55	0.00	0.50
		D1S468	-19.46	-5.15	-2.81	-0.85	-0.12	0.07	0.07	0.41
	CX50	D1S2635	-22.18	-5.16	-3.15	-1.31	-0.13	-0.10	0.00	0.49
	BFSP2	D3S1349	-15.53	-3.94	-1.88	-0.21	0.34	0.37	0.40	0.35
	PITX3	D10S1268	0.40	0.33	0.27	0.16	0.07	0.02	0.40	0.00
	MIP	D12S83	-16.51	-5.87	-3.39	-1.23	-0.35	-0.04	0.00	0.47
	CX46	D13S1236	-8.52	-3.95	-2.39	-0.74	-0.03	0.17	0.17	0.40
	15q21-q22	D15S117	-11.33	-3.17	-1.94	-0.84	-0.35	-0.10	0.00	0.50
		D15S1033	-24.02	-2.17	-0.42	0.83	0.98	0.59	**1.00**	**0.27**
	HSF4	D16S496	-4.67	-3.17	-2.23	-1.14	-0.54	-0.19	0.00	0.50
	17p13	D17S796	-22.82	-6.54	-4.31	-2.13	-1.00	-0.35	0.00	0.50
		D17S849	-15.40	-6.81	-4.62	-2.41	-1.23	-0.48	0.00	0.50
	17q24	D17S836	-19.56	-5.85	-3.41	-1.21	-0.32	0.00	0.02	0.47
		D17S802	-25.80	-10.45	-7.15	-3.85	-2.03	-0.82	0.00	0.50
	20p12-q12	D20S894	-21.30	-8.62	-5.79	-2.99	-1.57	-0.69	0.00	0.50
Ctas17	1pter-p36.1	D1S243	-8.58	-1.49	-0.78	-0.22	-0.06	-0.02	0.00	0.50
		D1S468	-17.17	-7.69	-5.22	-2.58	-1.18	-0.39	0.00	0.50
	CX50	D1S2635	-18.99	-4.79	-3.07	-1.44	-0.64	-0.21	0.00	0.50
	CX46	D13S1236	-3.62	-0.78	0.51	0.65	0.49	0.24	0.66	0.18
	15q21-q22	D15S117	-16.99	-5.69	-3.61	-1.63	-0.68	-0.19	0.00	0.50
		D15S1033	-13.70	-2.91	-0.70	-0.68	-0.21	-0.03	0.01	0.47
	HSF4	D16S496	-3.41	-1.52	-0.83	-0.15	0.11	0.14	0.15	0.37
	17p13	D17S796	-10.45	-3.13	-1.92	-0.77	-0.25	-0.03	0.01	0.47
		D17S849	-13.56	-6.22	-4.02	-1.81	-0.74	-0.22	0.00	0.50
	17q24	D17S836	-3.78	-2.31	-1.37	-0.46	-0.06	0.07	0.08	0.41
		D17S802	-10.74	-3.91	-2.40	-1.03	-0.44	-0.15	0.00	0.50

further markers were genotyped in this pedigree. Results of interest for each pedigree are summarised in Table 6.4.

Table 6.4 Maximum LOD scores of interest

Pedigree	Locus	Marker	Founders		CEPH	
			Max LOD	Theta	Max LOD	Theta
Crch13	CX46	D13S1236	2.96	0.04	4.37	0.03
	HSF4	D16S496	1.36	0.08	1.12	0.08
Crch30	15q21-q22	D15S1033	1.54	0.00	1.71	0.00
Crch32	15q21-q22	D15S1033	1.00	0.26	1.00	0.27
	PITX3	D10S1268	0.49	0.00	0.40	0.00

6.6.2 Mutation analysis of the Connexin 46 gene in pedigree crch13

6.6.2.1 Sequence Analysis

After obtaining a significant LOD score of 3.14 at D13S1236 in pedigree crch13, the *CX46* gene was investigated by sequencing three overlapping fragments covering the whole coding exon. A mutation was identified in several affected individuals in fragment 1 (Figure 6.1, 6.3). The G→A substitution at nucleotide 218 of the mRNA results in an Arg→His substitution at codon 76. As this fragment was quite large (496bp) and difficult to sequence due to a high 'G+C' content, a smaller fragment was amplified for confirmation of this mutation (fragment 1b, Appendix 3).

6.6.2.2 Denaturing High Performance Liquid Chromatography

Denaturing HPLC was used to screen all individuals in the pedigree. The optimal temperature for the detection of the mutation was determined to be 64°C, however, the profiles obtained were not always readily identified as wild type or mutant (Figure 6.2). The 12 individuals displaying these ambiguous profiles were directly sequenced at fragment 1b to determine their genotype. All affected individuals were found to be

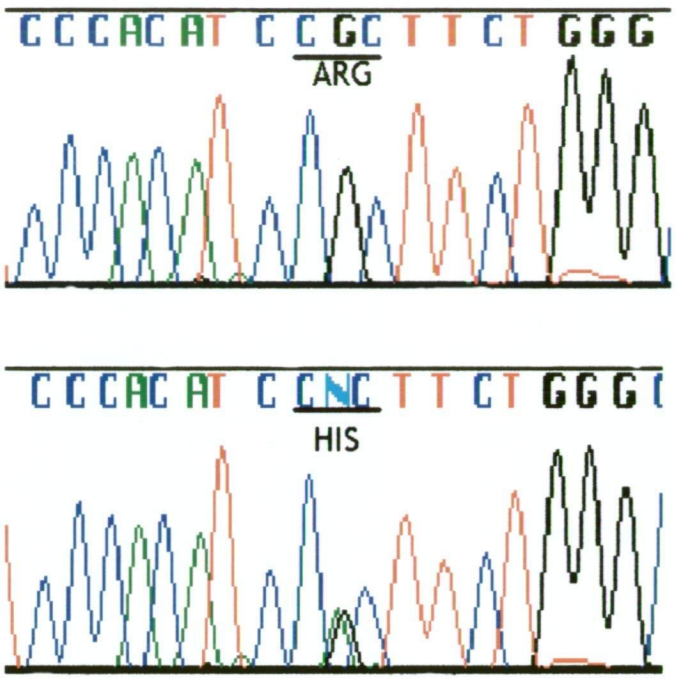


Figure 6.1. Sequence variant 218G→A of the *CX46* gene in pedigree crch13, resulting in R76H.

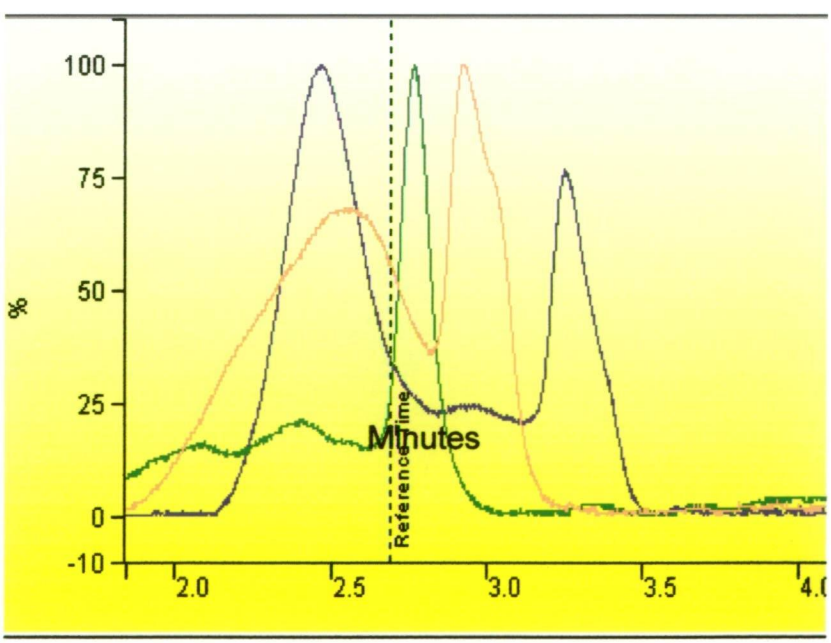


Figure 6.2. Representative dHPLC profiles of pedigree crch13 at fragment 1b of *CX46*. Green is wild type sequence; blue represents the R76H mutation; orange is an example of an ambiguous profile that was followed up by sequencing.

heterozygous for the mutation, and the majority of unaffected individuals were homozygous wild type. However, as indicated in Figure 6.3, unaffected individuals IV:4, V:7, V:8, V:9, V:27 and VI:15 also carry the mutation, resulting in a penetrance of 20/26 or 77%. The genotype for all these individuals was confirmed by direct sequencing. For individuals IV:4, V:7, V:8 and V:9, independent samples were obtained, which confirmed these results.

6.6.3 Reduced penetrance model

The LOD score was recalculated using an autosomal dominant model with 77% penetrance in heterozygotes as observed in the pedigree. The maximum LOD score was increased from 2.96 under the high penetrance model to 3.42 (θ=0.04) under the reduced-penetrance model.

6.6.4 The phenotype

The phenotype in this pedigree is a faint lamellar nuclear opacity with fine gold dots or haze and in some cases with needle like peripheral riders (Figure 6.4). Half of those with cataract had not had surgery and had good or minimally reduced vision. The mean age for surgery was 27 years (age range 10-67 years).

6.6.5 Analysis of controls and other pedigrees

Fifty unrelated unaffected individuals (100 chromosomes) and a proband from all pedigrees in the study were screened for the R76H mutation by direct sequencing, as this appeared significantly more reliable than the dHPLC. The mutation was not detected in any of these individuals.

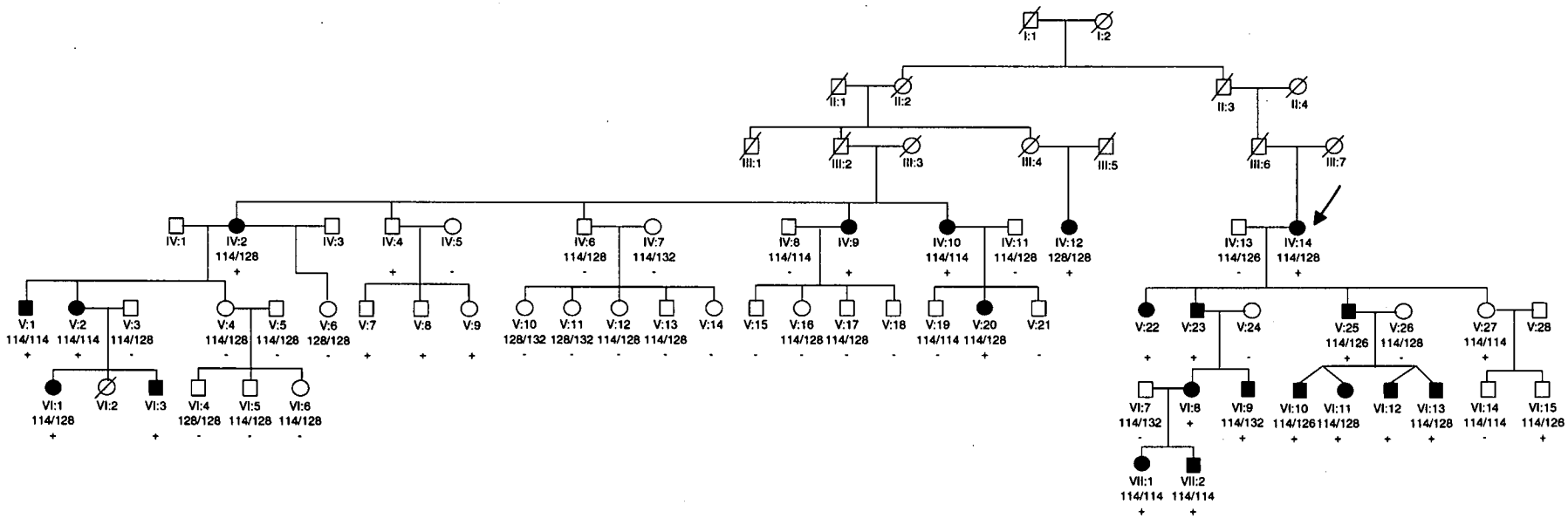


Figure 6.3. Pedigree diagram of Crch13. Genotypes for marker D13S1236 and the G218A (R76H) mutation of *CX46* are indicated. Shaded symbols indicate presence of ophthalmologist-confirmed cataract. '+' indicates heterozygous mutation carrier, '-' indicates wild type. Individuals with no genotypes have not been typed. The proband is marked with an arrow.

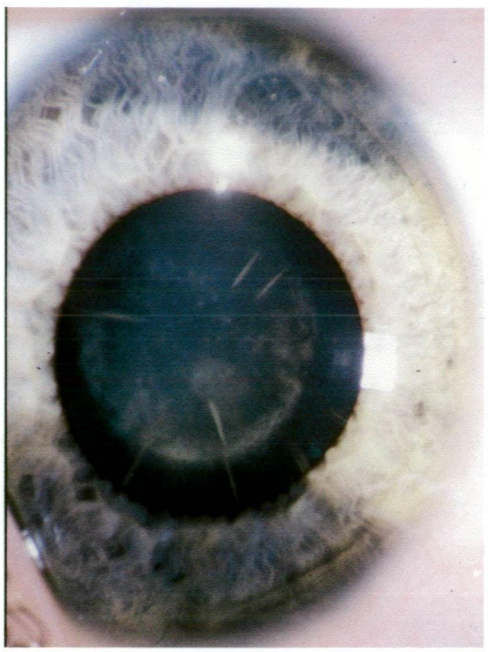


Figure 6.4. Photograph of the lens of individual VI:3 of pedigree crch13 showing the pulverulent phenotype in the nucleus with cortical riders.

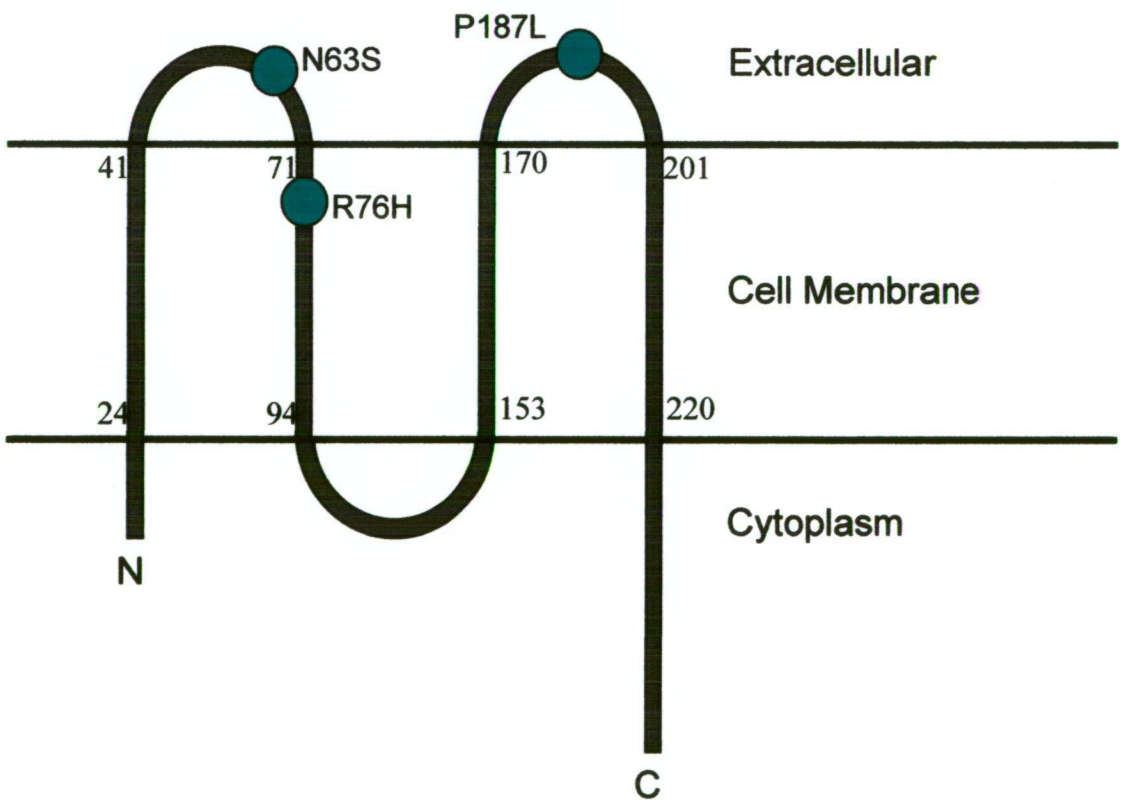


Figure 6.5. Predicted topology of Connexin 46 by HMMTOP. The location of the R76H mutation and other reported mutations are marked. The first and last residues of each transmembrane domain are numbered.

6.7 Discussion

6.7.1 Comparison of different allele frequency sets

The majority of results were similar between both sets of allele frequencies used, and the trends agree in every case. However, the analysis of D13S1236 in pedigree crch13 with the CEPH frequencies gave a significant LOD score of 4.37, whereas the result with the founder frequencies was a LOD of only 2.96. Inspection of the allele frequencies (Appendix 3) revealed the origin of the difference to be frequencies for alleles 114bp and 132bp. The pedigree contains many individuals with allele 114. This allele is not present in the CEPH set and was therefore given a low frequency of 0.01 for the LOD score calculations. Thus, to have so many affected individuals carrying this allele becomes unlikely by chance, and the significance of the LOD score increases. In the frequencies calculated from the founders, many individuals have this allele, indicating that it is common in the population being studied here. Thus the sharing becomes more likely by chance, and the LOD score is no longer significant. Both frequency estimates are calculated from 28 chromosomes. There are only 25 founders in the 4 pedigrees. Therefore neither frequency set is likely to be very accurate. This raises the question of the usefulness of using CEPH allele frequency if the common alleles are different between the populations. The most accurate linkage results would be obtained by the typing of additional individuals from the south-eastern Australian population to give more accurate allele frequencies. However because the *CX46* gene is small, it was investigated in this pedigree by direct sequencing to remove speculation. The usefulness of the CEPH frequencies needs to be evaluated on a case by case basis. In general they can be a useful resource for Caucasian populations, especially if there is insufficient data from the population under investigation. However, they should be used with caution and the limitations of the population recognised.

6.7.2 The Connexin 46 R76H mutation in pedigree crch13

6.7.2.1 Predicted effects

The *CX46* R76H mutation identified in pedigree crch13 is the fourth amino acid of the second transmembrane domain. A schematic of the structure of the CX46 protein as predicted by the HMMTOP program (www.enzim.hu/hmmtop) is given in Figure 6.5. All species represented in Genbank have an arginine at this point. Histadine appears to be a relatively conservative substitution as it does not change the structural predictions by HMMTOP (which asseses transmembrane domains) or PSIPRED (which searches for all secondary structure). However, this residue is highly conserved across species suggesting that this mutation is likely to be causative of the phenotype in this pedigree. The lack of gross effects on protein structure is implied by the incomplete penetrance observed in the pedigree. The mRNA for CX46 is encoded in one large exon so there is no requirement for the splicing out of introns. Therefore, it is unlikely that putative exon splice enhancers would be functional and they were not investigated. Other reported mutations of this gene include N63S and P187L. These two mutations, as well as the E48K mutation of *CX50* are in the extracellular loops and are believed to affect docking of the protein with the neighbouring cell (Berry *et al.* 1999; Mackay *et al.* 1999; Rees *et al.* 2000). The P88S mutation of *CX50* is in a transmembrane domain and causes a zonular pulverulent phenotype (Shiels *et al.* 1998). *In vitro* mutagenesis of the phylogenetically conserved P88 suggests that it plays a role in the voltage gating of gap junctions. It is yet to be determined if this is functionally significant in the non-innervated lens (Shiels *et al.* 1998). The functional effect of the novel R76H mutation detected in this study also remains to be determined.

6.7.2.2 Modification of the phenotype

Several individuals of pedigree crch13 have been classified as unaffected but have been clearly shown by direct sequencing to carry the R76H mutation identified in all affected individuals in the pedigree. Individuals V:27 and VI:15 have been examined and are clearly unaffected, however, carrier individuals IV:4, V:7, V:8 and V:9 have not been seen by an ophthalmologist and it is possible that they are affected sub-clinically, as DNA was collected from family members by mail-out of buccal swab kits. These individuals are currently being contacted for clinical assessment.

It appears from the presence of unaffected mutation carriers V:27 and VI:15 that this mutation is not fully penetrant. Recalculation of the two-point LOD score with a penetrance in carriers of 77% (as described in section 6.6.3) resulted in an increased LOD score compared with the fully penetrant model as would be expected. This indicates that the model used for linkage analysis more accurately fits reality and that the presence of the mutation correlates well with the genotype of marker D13S1236, despite missing values due to failed PCR. The failure rate for this marker suggests that a different marker should have been typed to represent this locus, however, as discussed, it was chosen to directly sequence the candidate gene to overcome all the problems with this marker definitively.

The only other report of a mutation with incomplete penetrance causing ADCC is a 5bp insertion in the *CRYGC* gene causing a zonular pulverulent phenotype (Ren *et al.* 2000). The phenotype of the reported pedigree is quite variable. All 26 affected patients carried the mutation, however two displayed only minimal cataract. One mutation carrier displayed no visible opacity and one obligate carrier was not examined. The variability was suggested to be due to modification by environmental

factors, although what these factors may be was not hypothesised. This may also be the case with pedigree crch13.

Ren *et al.* (2000) also suggest the possibility of a modifying gene acting in the pedigree with incomplete penetrance. The phenotype in their pedigree showed variability both between individuals and between the two eyes of patients. However, the authors suggest that a modifying gene would act evenly on both eyes and propose that environmental factors are more likely to be the cause of the variation. It is difficult to separate the effect of environment from genes, however, the presence of a modifier gene in pedigree crch13 is supported by individuals IV:4 and V:27. These unaffected individuals passed on the R76H mutation to their offspring, who also remain unaffected. Environmental factors are unlikely to show this type of pattern, unless there are significant household effects. As the majority of mutation carriers in the pedigree are affected, it appears that the R76H mutation does cause the phenotype. However, these carriers may also carry a protective allele of a different gene, which has been passed onto their offspring. Alternatively the affected individuals may carry a second harmful allele, however this may be less likely considering that the phenotype is expressed in multiple distant branches of the pedigree.

The *Lop10* mutation in mice may provide some insight into gene interactions in the development of congenital cataract. This phenotype was recently shown to be caused by a substitution mutation of the murine *cx50* gene (G22R) (Chang *et al.* 2002), however, the phenotype is variable and dependent on the genetic background of the mouse (Runge *et al.* 1992). Functional analyses of this mutation revealed that at least

part of the phenotype is caused by a reduction in the levels of phosphorylated cx46, although the mechanism of interaction between cx50 and the phosphorylation of cx46 is still unknown (Chang *et al.* 2002).

6.7.3 The remaining pedigrees

6.7.3.1 Crch32

The range of recombination values over which linkage is excluded is an important consideration. All markers (except D10S1268) are excluded from linkage at $\theta=0$ in pedigree crch32. In most cases, this exclusion extends to at least $\theta=0.1$. This recombination fraction equates to 9.8 Kosambi centiMorgans (as calculated by mapfun from the linkage utilities package, available from http://linkage.rockefeller.edu/soft/), which is a considerable distance. Because the markers were chosen to be as close as possible to the candidate gene, the range over which linkage is excluded includes the candidate gene when known. Although positive LOD scores are reached in some instances, they are not significant and they are a considerable distance from the candidate gene of interest. For D13S1236, linkage is only excluded to $\theta=0.02$ which is approximately 2 cM. The *CX46* gene is located 1.88 Mb from marker D13S1236. In this region of chromosome 13, comparisons of physical and genetic maps have shown there to be 2.18 cM/Mb (Payseur and Nachman 2000). Therefore, 1.88 Mb equates to 4.1 cM and linkage to the *CX46* gene cannot be excluded in this pedigree.

For one marker linkage could not be excluded at any recombination fraction: D10S1268, representing *PITX3*. The linkage simulations carried out on this pedigree and presented in Chapter 5 indicate that a LOD score of 9.32 at $\theta=0.05$ is theoretically possible in this pedigree, under the conditions used in the simulations.

The marker used in simulation had four equally frequent alleles which is rarely the case with real markers. Simulation with actual marker frequencies observed in the founders for D10S1268 gave a maximum LOD score at θ=0.05 of 10.22, and an average LOD of 6.66. This indicates that the pedigree should have sufficient power to detect significant linkage at this locus. The equivocal LOD score at the *PITX3* locus may be caused by a lack of information at this marker. The heterozygosity of this marker was observed to be 80% in the founders. The CEPH database reports an observed heterozygosity of 74% in its population, hence the marker should contain sufficient information. However, allele 139 is common in both populations (0.37 in founders and 0.46 in CEPH population) and many individuals in the pedigree are homozygous for this allele. This reduces the information content of this marker. The typing of additional markers in this region is therefore required to evaluate the importance of this result.

For adjacent markers on 15q, D15S1033 and D15S117, the range of excluded values of θ is only 0.05 and 0.09 respectively. As the markers at 15q21-22 are separated by only 280 kb, the exclusion results indicate that there is unlikely to be linkage to a gene between these markers. However, linkage cannot be excluded at distances beyond 0.09 on either side of the markers, and results are more inconclusive further away. The gene at this locus, while unknown, was found to be located between D15S117 and D15S1033 by recombinations in informative indivuals in the pedigree of Vanita *et al.* (2001), hence the distances over which exclusion of linkage was observed should encompass the disease gene. For other loci where the candidate gene has not been identified, the typing of additional markers may be required to exclude or identify linkage over greater distances.

Although further work is required, it is possible that the phenotype in pedigree crch32 is caused by a gene at a novel locus. If typing of additional markers around *PITX3* and *CX46* does not reveal linkage, pedigree crch32 should be the subject of a genome-wide scan to identify a possible novel cataract locus.

6.7.3.2 Crch30

There is less power to detect linkage with pedigree crch30 than with the other large pedigrees in the study, with a maximum simulated LOD score of 2.5 and average LOD of 0.91 (Table 5.2). Marker D15S1033 gave a LOD score of 1.5 (θ=0) using founder frequencies and a LOD score of 1.7 using CEPH frequencies. The p-value for this LOD score can be calculated from the χ^2 distribution (Nyholt 2000):

$$\chi^2 = \text{observed LOD} \times 2 \times \log_e 10 = 1.7 \times 2 \times \log_e 10 = 7.82.$$

The corresponding p-value is 0.003, which is significant for confirmation of linkage, as a LOD score of genome-wide significance has previously been identified at this locus (Vanita *et al.* 2001b). The empirical p-value, calculated by simulating genotypes in SLINK over 1000 replicates under the null hypothesis of no linkage, with the allele frequencies observed for D15S1033, was also 0.003 (0.3% of LOD scores were greater than a threshold of 1.7). Since congenital cataract is so heterogeneous and there was no prior hypothesis that pedigree crch30 would link to this locus, it may be more appropriate to employ a genome-wide significance level of 0.000049 (Lander and Kruglyak 1995). However, simulations indicate this would not be possible in this small pedigree. In summary, this locus should be the first to be followed up in this pedigree, but the result is not statistically significant. The LOD score at the second marker at this locus - D15S117 - is -1.7 (θ=0), which is approaching the criterion for exclusion. Using the CEPH frequencies, linkage to

D15S117 is excluded in this pedigree at θ=0 with a LOD score of -2.43. Haplotypes for the two markers were estimated with GeneHunter Plus. There did not appear to be a segregating haplotype, however, allele 1 of D15S1033 is carried by all affected individuals (Figure 6.6). This is contributing to the positive LOD scores obtained at this marker. In contrast, there does not appear to be any segregation of D15S117 alleles, resulting in the negative LOD score observed. The typing of additional markers and haplotype analyses are required to further investigate this locus. As the gene at this locus has not yet been determined, we could not check for mutations by direct sequencing as was done with the *CX46* gene in crch13. Markers representing other mapped loci with as yet unknown genes also give inconclusive results. The observed LOD scores do not approach significance and the possibility that the gene lies nearby could not be excluded.

Three loci (*CX50, PITX3* and *MIP*) are only excluded over small recombination values. *CX50* is approximately 11 Mb from the marker D1S2635, hence this gene is not excluded from linkage in this pedigree as the LOD score is -0.73 at θ=0.1. The CX50 gene has not been placed on any genetic maps, so it is difficult to know its genetic distance from the marker, however, using the average recombination rate in that region of chromosome 1, it appears to be around 12.4cM from the marker (Table 6.1). A similar situation exists for *PITX3* and *MIP*. *BFSP2* is excluded over a larger distance (θ=0.13), with the marker being only 27kb from the gene.

The results at all loci in pedigree crch30 are equivocal, neither including nor excluding linkage, due mainly to the lack of power. The pedigree would benefit from the typing of additional markers at all loci so that detailed haplotype analyses can be

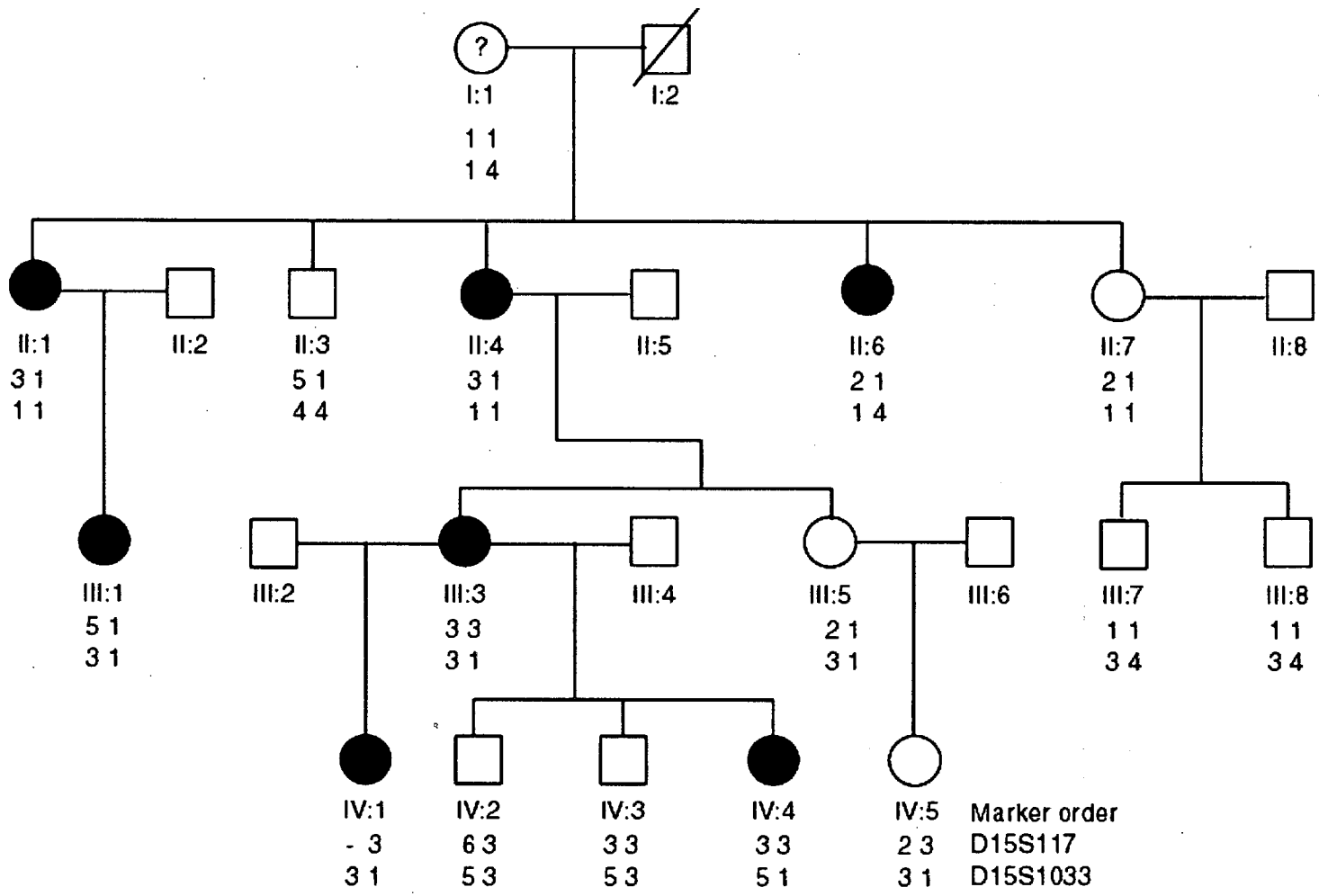


Figure 6.6: Pedigree Crch30 with 2 marker haplotyes for chromosome 15 markers showing no segregating haplotype but segregation of allele 1 of D15S1033 with paediatric cataract. Shaded symbols indicate presence of ophthalmologist-confirmed cataract

carried out to reduce the number of candidates to be investigated by immediately excluding those loci where there is no segregating haplotype.

6.8 Summary of the investigation of congenital cataract genes

In summary, the analysis of all known cataract genes and loci discussed in this Chapter and Chapter 5 revealed three causative mutations in large pedigrees. Two of these (P24T of *CRYGD* and R76H of *CX46*) were identified by linkage analysis and the third (splice site mutation of *CRYBA1/A3*) was detected by SSCP analysis. There is no strong evidence that the congenital cataract phenotype in pedigree crch32 links to any known congenital cataract locus, although further work is required to confirm exclusion or linkage at a few loci. Additional investigations are also required at all loci for the smallest of the pedigrees investigated by linkage analysis. The loci previously linked to recessive forms of congenital cataract have not been investigated in this population. These loci may also play a role in autosomal dominant congenital cataract as *CRYAA* does (Litt *et al.* 1998; Pras *et al.* 2000). The molecular cause of the congenital cataract observed in the remaining 41 pedigrees is currently unknown. While these pedigrees are too small to be useful individually in gene identification studies, they will be important for the determination of the relative contributions of cataract genes to the overall causes of familial cataract.

Chapter 7

Discussion

Research into the molecular causes of congenital cataract has given insights into lens biology through the study of the effect of the mutations on the protein functions. This study aimed to determine the causative mutations in a collection of congenital cataract pedigrees and, where possible, identify novel genes involved in cataractogenesis. A novel gene was identified on the X chromosome and shown to be associated with Nance-Horan Syndrome (NHS), an X-linked pleiotropic syndrome involving congenital cataract and dental anomalies. Mutations were also identified in three large pedigrees from south-eastern Australia with autosomal dominant congenital or familial cataract (ADCC).

Congenital syndromes contribute significantly to health care costs because sufferers require long-term care and management. NHS is one such disorder, with most affected males suffering severe visual defects from an early age and also requiring extensive dental care. Learning disabilities and mental retardation associated with the syndrome place added difficulties onto patients and their families. The gene for Nance-Horan Syndrome (*NHS*) was identified in our study in the pedigree in which the syndrome was first described. This may lead to the development of pre-natal testing in families known to carry the gene, and earlier molecular diagnosis and appropriate treatment in sporadic cases. Current clinical diagnoses are often not confirmed until tooth eruption. If isolated X-linked cataract is in fact a genetically distinct disorder from NHS, management of the disorder, particularly in sporadic female cases, can be based on the molecular diagnosis or exclusion of NHS rather

than on variable clinical findings, even in the absence of knowledge of the causative gene for isolated X-linked cataract. The protein encoded by this novel gene does not belong to any known family of proteins and its function is yet to be determined. However, it appears to be involved in the development of a range of tissues. The determination of the function of this gene will provide insights into the developmental biology of many tissues, in particular the eye, tooth and brain.

There are many approaches that can be taken in order to elucidate the function of the protein. Homology searches did not reveal any significant similarity to known protein families. However, putative nuclear localisation sequences were identified, suggesting the protein is targeted to the cell nucleus. This can be tested through tagging of expressed protein with green flourescent protein and investigating the cellular localisation of the protein (Gomes *et al.* 2003). Additionally, the truncating mutations identified in NHS pedigrees can be introduced to the expression contruct and the effect these have on cellular localisation can be investigated.

X-inactivation can play an important role in the expression of a phenotype in females. The phenotype may only be expressed in cells where the disease chromosome is the active allele, leaving other cells healthy. This may result in a "mosaic" pattern of disease expression. There are many X-linked genes that escape inactivation in females and appear not to be subject to the dose correction that chromosomal inactivation allows. It would be interesting to understand the effect of X-inactivation on the mild and variable phenotype observed in females with NHS. A chromosome wide study of inactivation indicated that nearby genes *RAI2* and *SCML2* are inactivated (Carrel *et al.* 1999). However, the process appears to act in a gene specific

manner and the physical proximity of inactivated genes does not necessarily imply the inactivation of *NHS*. There are several ways this can be investigated. Methylation of CpG dinucleotides is thought to be a key mechanism in gene inactivation. A simple method for investigation methylation patterns involves treatment of the genomic DNA with sodium bisulphite which converts unmethylated cytosines to uracil, while leaving methylated bases intact. The target region is then amplified by PCR and sequenced (Uhlmann *et al* 2002). Methylated cytosines remain as cytosines while unmethylated nucleotides are detected as thymines (uracil is amplified as thymine in PCR). This technique would require genomic DNA from the tissue of interest at a time when the gene would normally be expressed, as methylation may be used to silence genes at various stages in development. More sophisticated techniques also exist. Rodent/human somatic cell hybrids provide such a method. Hybrids are generated containing an inactivated human X chromosome. Transcripts of the gene of interest are detected through RT-PCR. If a transcript is detected, the conclusion is that the gene has escaped inactivation (Carrel *et al.* 1999). This type of technique may be useful in this case to determine if the gene escapes inactivation and therefore the mutant alleles in females with NHS tend to be expressed. If this were the case, the mild phenotype could be attributed to a co-dominant mode of action.

A putative paralogue of *NHS* was identified on chromosome 6. While this predicted gene requires confirmation in the laboratory and extensive functional analyses to determine its role, it may also be involved in the development of the eye and other tissues.

While the discovery of novel genes is important for understanding the biology, it is also important to society and the affected families to understand the cause of their disabilities. As the genetics community moves towards the study of complex multifactorial traits, it is important to continue the study of Mendelian traits in order to identify novel genes and determine their functions. The sequencing of the human genome assisted in the discovery of the *NHS* gene, and greatly reduced the time and resources required. *In silico* analysis is an important tool in the overall process. However, genes and their functions cannot be definitively predicted by computer programs, as evidenced by the process of gene identification undergone in this project. While the *in silico* analyses indicated a promising candidate for a novel gene, the predictions were not 100% correct, with unpredicted exons identified in the laboratory. As the human sequence improves and algorithms for gene prediction become more powerful, gene prediction *in silico* should become more accurate. An exon was identified approximately 350kb upstream from the rest of the gene but was not predicted and was only found following the investigation of ESTs in other species. Comparative genomics is an increasingly powerful tool as the genomes of more organisms are completely sequenced.

Familial cataract is genetically heterogeneous, with mutations in 14 genes identified as causing ADCC and mutations of two genes responsible for autosomal recessive forms of the disease. Several other loci have also been mapped, although the genes are yet to be characterised. Two of the three mutations identified in known cataract genes in this study (*CRYBA1/A3* and *CRYGD*) were previously reported in pedigrees from India, while the third (*CX46*) is novel. Until now, only the Q155Stop mutation of *CRYBB2* had been identified in more than 1 pedigree although Kannabiran *et al.*

(1998) and Bateman *et al.* (2000) report different mutations of the same base in *CRYBA1/A3*.

The three mutations identfied here are in genes well known to contribute to cataract formation from studies of both mice and humans. However, to definitively assign cause to these mutations, functional studies should be conducted. For example, in order to study the effect of the P24T mutation of CRYGD, site directed mutagenesis could be used to introduce the mutation to an expression vector containing the *CRYGD* coding sequence. The mutant protein can then be expressed in *E. coli* and isolated and its properties investigated and compared to those of wildtype CRYGD. These types of experiments have been conducted for other mutations of this gene including the R58H and R14C variants (Pande *et al.* 2000, Pande *et al.* 2001). These experiments show that the R58H mutant protein is less soluble than wildtype and that R14C protein forms disulfide linked oligomers where wildtype protein does not. Similar experiments have been carried out for a Connexin50 mutation (P88S). Xenopus oocytes injected with the mutant protein failed to form functional gap junctions compared to wildtype protein (Pal *et al.* 1999). A similar system could be used to test the R76H mutation of CX46 identified here.

There remain 31 pedigrees in this collection with ADCC for which no mutation has been identified, although only five of the 14 known genes have been thoroughly investigated in all pedigrees. In addition, one large pedigree does not show strong evidence of involvement of any of the known cataract loci by linkage analysis, indicating that an as yet unknown locus may exist for this disease. Following further work to confirm exclusion of known loci, this pedigree would be an ideal candidate

for a genome-wide scan. Other monogenic disorders also display extreme heterogeneity. For example there are 19 known genes for the retinal degeneration disease retinitis pigmentosa and at least 17 additional loci (Phelan and Bok 2000).

The remaining pedigrees in this collection are currently too small to contribute to linkage analyses. They cannot be used in a pooled analysis due to the heterogeneity of the disorder. There are likely to be too many loci involved in such a large collection of pedigrees, so any effect would be too dilute to detect. It is not feasible to stratify the pedigrees on phenotype to improve the power, as similar phenotypes have been shown to be caused by different genes, such as zonular pulverulent cataracts being caused by mutations of *CX46, CX50* and *CRYGC* (Shiels *et al.* 1998; Héon *et al.* 1999; Mackay *et al.* 1999). Additionally the same mutation, such as the Q155X mutation of *CRYBB2* can cause distinct phenotypes (Litt *et al.* 1997; Gill *et al.* 2000; Vanita *et al.* 2001a) as the P24T mutation of *CRYGD* detected here also does. Mutations would therefore be missed by the exclusion of a pedigree from a particular locus on the basis of phenotype alone. The best method for the identifying mutations in known genes in these small pedigrees is direct sequencing of the genes, however this is an expensive and time-consuming process.

Linkage analysis in a few suitably large pedigrees revealed two cataract mutations, including one that was missed by SSCP. While SSCP is cheaper than sequencing it is still labour-intensive. Therefore a faster method of screening known genes in these pedigrees is required. Denaturing HPLC was investigated as an alternative. While less labour-intensive than gel-based techniques such as SSCP, dHPLC was also found to be problematic. The temperature at which a variant will be identified may be very

specific. It is not always the same as that calculated as the melting temperature of the strand and must be determined empirically for the most accurate result. As with SSCP, some variants are easier to detect than others. Retention time is related mainly to melting temperature of the strand which is dependent on the 'G+C' content and length of the fragment. The overall success rates for dHPLC are similar to or better than SSCP (Benit *et al.* 2000; Buyse *et al.* 2000; Ellis *et al.* 2000; Jones *et al.* 2000), however, in our experience it is necessary to employ multiple temperatures to be confident of having captured the majority of variants as the melting temperature of the strand can vary in different regions of the fragment. Also, it is still necessary to screen each individual exon, or even multiple fragments per exon for large exons. The limitation of SSCP are discussed in Chapter 5. It is obvious that neither method will detect all variants.

Although not presented here, the splice site mutation in *CRYBA1/A3* detected in pedigree crch08 by SSCP was not detected by dHPLC during trials of the method in our laboratory. This lack of confidence in the screening methods is what promted the linkage analysis of crystallin genes in the larger pedigrees. It is not clear if one method is "better" than the other. In fact they are complementary. In order to be sure of identifying all the variation, it would be necessary to conduct both SSCP and dHPLC at a range of conditions for both methods. In reality, this is not feasible and researchers choose one or the other and conduct experiments at a limited number of conditions. It is accepted that some variation will likely be missed this way, but the significant reduction in costs offset the risk. However, with the current state of high-throughput sequencing technology and the accompanying reduction in cost of this technique, direct sequencing of the patient samples is a superior method for searching

for genetic variation. Even within the time frame of this project this decision was taken with the candidate genes for NHS to be directly sequenced in affected individuals, whereas the early investigations of crystallin genes were carried out by SSCP analysis.

Although traditional linkage analysis is problematic in this data set, a haplotype approach may be useful. The typing of several microsatellite markers across candidate gene regions to establish haplotypes would be relatively cheap and fast. The non-segregation of haplotypes through a pedigree would immediately exclude that pedigree from further analysis at that locus. Therefore, only genes where a haplotype consistent with linkage was observed would need to be investigated in detail. Recombination events may also be useful in excluding genes. As congenital cataract is a rare and highly penetrant single gene disorder, phenocopies are not expected in these pedigrees. Only one affected individual with a different haplotype would be required to exclude a locus. This methodology was employed in the analysis of putative X-linked pedigrees. Two pedigrees were shown to not have segregating haplotypes and were excluded from further analysis. Mutations were not identified in the remaining pedigrees however, and it is unknown if the correct candidate gene for isolated X-linked cataract was sequenced, or indeed if there are uncharacterised regions of the *NHS* gene such as distal regulatory elements or additional 5'UTR sequence. In the case of the ADCC genes, the candidate genes are already well known, and the absence of mutations in these genes would be conclusive. Haplotype analysis at congenital cataract loci is underway in the largest of the small pedigrees, ctas16.

A potential problem with a haplotype approach still relates to the size of the pedigrees. In small pedigrees with only a few offspring (eg ctas34 and crch37, amongst others) it will still be difficult to determine if a haplotype is segregating. With only a few meioses in these pedigrees it is likely that a haplotype will segregate by chance and thus a large number of genes would still need to be investigated. As well, if there are no unaffected offspring it cannot be established that an alternative haplotype segregates with unaffected status. Also, as illustrated by the shared haplotype observed in the large NHS pedigree crch2551, the shared region may be very small and could be missed if the markers selected are too few or uninformative. It is also feasible that there will be recombinations between the gene of interest and the typed markers, further reducing the power.

An alternative to this traditional genotyping approach may be to use single nucleotide polymorphisms (SNPs) that are within the genes themselves and therefore more likely to be in linkage disequilibrium with the causative mutation. In theory, a SNP map could be typed across the region of the candidate gene which should determine if the same chromosome is inherited by all affected individuals. In practice this requires informative SNPs to be located in appropriate places and high-throughput SNP genotyping to be cost effective. In smaller pedigrees, SNPs may be uninformative and thus the size of the pedigree will still also be a limiting factor in determining SNP segregation with disease phenotype.

These small pedigrees are of limited value for the identification of novel loci. However, they remain an important resource for the investigation of known genes. Genotype-phenotype correlations have been difficult to establish, due to the large

numbers of genes and mutations and relatively small numbers of pedigrees with a wide range of phenotypes. Some generalisations can be made such as all mutations identified to date in connexin genes give rise to pulverulent phenotypes (Shiels *et al.* 1998; Berry *et al.* 1999; Mackay *et al.* 1999; Rees *et al.* 2000; Polyakov *et al.* 2001). The pedigree collection in this study represents a significant contribution to the overall numbers of congenital cataract pedigrees in the literature, as only around 40 pedigrees with ADCC have previously been described in detail. Collection of ADCC pedigrees from south-eastern Australia is ongoing, and several more pedigrees have been identified since this study began. If all cataract genes can be evaluated in these pedigrees, genotype-phenotype correlations may emerge based on greater functional understanding, allowing for the preferential investigation of candidate genes in the diagnosis of congenital cataract pedigrees.

A genome-wide scan in pedigree crch32 may reveal a novel cataract locus. The identification of a potentially novel cataract-causing gene in this pedigree is likely to reveal information about lens biology and cataract formation. It is currently unknown to what extent congenital cataract genes contribute to age-related cataract. An understanding of lens biology has been enhanced through the study of congenital cataract pedigrees. This knowledge may be useful in treating age-related cataract which is a significant cause of blindness in older individuals. A study of the prevalence of cataract in Australia revealed that 6% of individuals over 50 years of age had undergone surgery for age-related cataract, and that 50% of the study group had significant nuclear opacities and 25% had moderate cortical opacities (Mitchell *et al.* 1997). The basic mechanism of cataract formation is believed to be the degradation of lens proteins over time in response to stresses such as oxidative

damage by UV light (Ottonello *et al.* 2000). As with congenital cataract, a range of phenotypes have been described. The nuclear and cortical forms are thought to have a strong genetic component (Heiba *et al.* 1995). Heritability has been estimated at 48% for nuclear cataract (Hammond *et al.* 2000) and 53-58% for cortical cataract (Hammond *et al.* 2001). Surprisingly, the congenital cataract genes have not been investigated in a cohort of late-onset cataract patients. This type of investigation may reveal similarities between the two forms with particular variants possibly contributing to genetic susceptibility for late-onset cataract. Surgery to remove late-onset cataracts is one of the most frequent operations in Australia, leading to a significant cost burden. Estimates range from $US2020 to $US4500 per quality-adjusted life year in Western countries (Busbee *et al.* 2002; Kobelt *et al.* 2002). A clearer understanding of the genetics and molecular mechanisms of this multifactorial trait could lead to improved diagnostic and preventative techniques as well as alternative treatments.

ADCC is referred to as a monogenic disorder. However, the results presented here and in the literature suggest that it is more complex than mutations in any given gene giving rise to the phenotypic spectrum, even within a single pedigree. This is illustrated by the difficulty in assigning genotype-phenotype correlations. For example, the *CRYBA1/A3* splice site mutation identified in crch08 gives rise to a similar phenotype as described in an Indian pedigree with the same mutation (Kannabiran *et al.* 1998). This fits with the 'monogenic' descriptions. However, the P24T mutation of *CRYGD* in pedigree ctas17 gives rise to a nuclear cataract, whereas the identical mutation in an Indian pedigree causes a lamellar opacity. This suggests that either environmental factors or differences in the genetic backgrounds between

the Australian and Indian families are affecting the phenotype, more in line with a 'complex' disorder. Phenotypic variation can also be seen within families as illustrated by pedigree crch13 in which all affected individuals carry the R76H mutation of *CX46*. This mutation appears to cause an opacity, however it is not fully penetrant and it is likely that other factors such as modifier genes or environmental effects are acting within this pedigree. There is only one other report of incomplete penetrance in a pedigree with congenital cataract. This is possibly due to ascertainment bias as, to date, only large pedigrees with multiple affected individuals have been thoroughly examined. Pedigrees with apparently unaffected individuals may be due to incomplete penetrance of the mutation. This will only be resolved as mutations are identified in smaller pedigrees.

Animal models are being investigated for the effect of modifying genes. As mentioned in Chapter 6, the *Lop10* mutation in mice was found to be a *cx50* mutation, however, the phenotype was dependent on the genetic background of the mouse and interactions with *cx46* were believed to be involved. A similar animal model is the Upjohn Pharmaceuticals Limited (UPL) rat. This rat expresses a semi-dominant cataract. The phenotype is believed to be caused by a point mutation in *CX50*, however, a second locus acting as a quantitative trait locus (QTL) has been mapped in these rats. This second locus is thought to affect the age at which the cataract matures, but the gene has not yet been isolated (Yamashita *et al.* 2002). These are examples of animal models of a 'monogenic' disorder that display evidence for multilocus involvement. As laboratory animals are more homogeneous than humans due to inbreeding and close control of their environment, it is possible to detect the effects of modifying genes. Linkage analysis and genome wide scans may be useful for

detecting modifier loci in some large human pedigrees, however, in most cases, the effects are probably too small to be found in this way. Therefore animal models become very important for investigation of gene-gene interactions. Mutations with the major effect can be introduced onto different genetic backgrounds and the differences in phenotype (caused by modifier genes) mapped through linkage analysis. Identified genes can then be tested in human pedigrees. Candidates for modifier genes may also be identified through an understanding of biochemical pathways and cell biology.

Modifying genes are also being found to play a role in other human 'monogenic' disorders such as the β-thalassemias which are caused by mutations of the β-globin genes. However, the genotype-phenotype correlations are weak. There are now two well defined secondary modifing loci and several tertiary modifiers that are less well characterised (Weatherall 2001). It is becoming evident that many diseases previously classified as 'monogenic' are in fact multifactorial and very few genes actually act in isolation. Experience from the research into β-thalassemia has suggested that even though genome scans may be able to identify these modifier loci, it is still neccesary to understand how the gene interacts at a molecular level with the cellular machinery in order to fully disect the medical significance of mutations in such a gene.

This investigation into the genetic causes of familial cataract in south-eastern Australia has revealed two mutations previously reported in Indian pedigrees with interesting implications for genotype-phenotype correlations (or lack of them), and a novel mutation of a connexin gene whose functional effect is yet to be determined. A novel developmental gene responsible for Nance-Horan Syndrome has also been

identified. This represents a significant advance in cataract genetics and the investigations of *NHS* gene function will provide novel insights into developmental biology.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**:3389-3402

Ariturk N, Oge I, Mohajery F, Erkan D, Turkoglu S. (1998) Secondary glaucoma after congenital cataract surgery. *Int Ophthalmol* **22**:175-180

Armitage MM, Kivlin JD, Ferrell RE. (1995) A progressive early onset cataract gene maps to human chromosome 17q24. *Nat Genet* **9**:37-40

Bateman JB, Geyer DD, Flodman P, Johannes M, Sikela J, Walter N, Moreira AT, Clancy K, Spence MA. (2000) A new betaA1-crystallin splice junction mutation in autosomal dominant cataract. *Invest Ophthalmol Vis Sci* **41**:3278-3285

Benit P, Kara-Mostefa A, Berthelon M, Sengmany K, Munnich A, Bonnefont JP. (2000) Mutation analysis of the hamartin gene using denaturing high performance liquid chromatography. *Hum Mutat* **16**:417-421

Bergen AA, ten Brink J, Schuurman EJ, Bleeker-Wagemakers EM. (1994) Nance-Horan syndrome: linkage analysis in a family from The Netherlands. *Genomics* **21**:238-240

Berry V, Francis P, Kaushal S, Moore A, Bhattacharya S. (2000) Missense mutations in MIP underlie autosomal dominant 'polymorphic' and lamellar cataracts linked to 12q. *Nat Genet* **25**:15-17

Berry V, Francis P, Reddy MA, Collyer D, Vithana E, MacKay I, Dawson G, Carey AH, Moore A, Bhattacharya SS, Quinlan RA. (2001) Alpha-B crystallin gene (CRYAB) mutation causes dominant congenital posterior polar cataract in humans. *Am J Hum Genet* **69**:1141-1145

Berry V, Ionides AC, Moore AT, Plant C, Bhattacharya SS, Shiels A. (1996) A locus for autosomal dominant anterior polar cataract on chromosome 17p. *Hum Mol Genet* **5**:415-419

Berry V, Mackay D, Khaliq S, Francis PJ, Hameed A, Anwar K, Mehdi SQ, Newbold RJ, Ionides A, Shiels A, Moore T, Bhattacharya SS. (1999) Connexin 50 mutation in a family with congenital "zonular nuclear" pulverulent cataract of Pakistani origin. *Hum Genet* **105**:168-170

Bova MP, Yaron O, Huang Q, Ding L, Haley DA, Stewart PL, Horwitz J (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 U S A* **96**:6137-42

Boyle DL, Takemoto L. (2000) A possible role for alpha-crystallins in lens epithelial cell differentiation. *Mol Vis* **6**:63-71

Brakenhoff RH, Aarts HJ, Schuren F, Lubsen NH, Schoenmakers JG. (1992) The second human beta B2-crystallin gene is a pseudogene. *Exp Eye Res* **54**:803-806

Bu L, Jin Y, Shi Y, Chu R, Ban A, Eiberg H, Andres L, Jiang H, Zheng G, Qian M, Cui B, Xia Y, Liu J, Hu L, Zhao G, Hayden MR, Kong X. (2002) Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. *Nat Genet* **31**:276-278

Burge C, Karlin S. (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**:78-94

Busbee BG, Brown MM, Brown GC, Sharma S. (2002) Incremental cost-effectiveness of initial cataract surgery. *Ophthalmology* **109**:606-612

Buyse IM, Fang P, Hoon KT, Amir RE, Zoghbi HY, Roa BB. (2000) Diagnostic testing for Rett syndrome by DHPLC and direct sequencing analysis of the MECP2 gene: identification of several novel mutations and polymorphisms. *Am J Hum Genet* **67**:1428-1436

Carrel L, Cottle AA, Goglin KC, Willard HF. (1999) A first-generation X-inactivation profile of the human X chromosome. *Proc Natl Acad Sci U S A* **96**:14440-14444

Cartegni L, Chew SL, Krainer AR. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**:285-298

Chang B, Wang X, Hawes NL, Ojakian R, Davisson MT, Lo WK, Gong X. (2002) A Gja8 (Cx50) point mutation causes an alteration of alpha 3 connexin (Cx46) in semi-dominant cataracts of Lop10 mice. *Hum Mol Genet* **11**:507-513

Chirgadze Y, Nevskaya N, Vernoslova E, Nikonov S, Sergeev Y, Brazhnikov E, Fomenkova N, Lunin V, Urzhumtsev A. (1991) Crystal structure of calf eye lens gamma crystallin IIIb at 2.5A resolution: its relation to function. *Exp Eye Res* **53**:295-304

Collins A, Frezal J, Teague J, Morton NE. (1996a) A metric map of humans: 23,500 loci in 850 bands. *Proc Natl Acad Sci U S A* **93**:14771-14775

Collins A, Teague J, Keats BJ, Morton NE. (1996b) Linkage map integration. *Genomics* **36**:157-162

Conley YP, Erturk D, Keverline A, Mah TS, Keravala A, Barnes LR, Bruchis A, Hess JF, Fitzgerald PG, Weeks DE, Ferrell RE, Gorin MB. (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**:1426-1431

Cottingham RW, Jr., Idury RM, Schaffer AA. (1993) Faster sequential genetic linkage computations. *Am J Hum Genet* **53**:252-263

Eckstein M, Vijayalakshmi P, Gilbert C, Foster A. (1999) Randomised clinical trial of lensectomy versus lens aspiration and primary capsulotomy for children with bilateral cataract in south India. *Br J Ophthalmol* **83**:524-529

Eiberg H, Lund AM, Warburg M, Rosenberg T. (1995) Assignment of congenital cataract Volkmann type (CCV) to chromosome 1p36. *Hum Genet* **96**:33-38

Eiberg H, Marner E, Rosenberg T, Mohr J. (1988) Marner's cataract (CAM) assigned to chromosome 16: linkage to haptoglobin. *Clin Genet* **34**:272-275

Ellis LA, Taylor CF, Taylor GR. (2000) A comparison of fluorescent SSCP and denaturing HPLC for high throughput mutation scanning. *Hum Mutat* **15**:556-564

Favor J, Pretsch W. (1990) Genetic localization and phenotypic expression of X-linked cataract (Xcat) in Mus musculus. *Genet Res* **56**:157-162

Finzi S, Li YY, Mitchell TN, Farr A, Sundin O, Maumenee IH. (2002) Posterior polar cataract: clinical spectrum and genetic analysis in a large family. *Am J Hum Genet* **71**:453

Fraccaro M, Morone G, Manfredini U, Sanger R. (1967) X-linked cataract. *Ann Hum Genet* **31**:45-50

Francis P, Berry V, Bhattacharya S, Moore A. (2000a) Congenital progressive polymorphic cataract caused by a mutation in the major intrinsic protein of the lens, MIP (AQP0). *Br J Ophthalmol* **84**:1376-1379

Francis P, Chung JJ, Yasui M, Berry V, Moore A, Wyatt MK, Wistow G, Bhattacharya SS, Agre P. (2000b) Functional impairment of lens aquaporin in

two families with dominantly inherited cataracts. *Hum Mol Genet* **9**:2329-2334

Francis PJ, Berry V, Bhattacharya SS, Moore AT. (2000c) The genetics of childhood cataract. *J Med Genet* **37**:481-488

Francis PJ, Berry V, Hardcastle AJ, Maher ER, Moore AT, Bhattacharya SS. (2002) A locus for isolated cataract on human Xp. *J Med Genet* **39**:105-109

Francis PJ, Berry V, Moore AT, Bhattacharya S. (1999) Lens biology: development and human cataractogenesis. *Trends Genet* **15**:191-196

Francis PJ, Ionides A, Berry V, Bhattacharya S, Moore AT. (2001) Visual outcome in patients with isolated autosomal dominant congenital cataract. *Ophthalmology* **108**:1104-1108

Gill D, Klose R, Munier FL, McFadden M, Priston M, Billingsley G, Ducrey N, Schorderet DF, Héon E. (2000) Genetic heterogeneity of the Coppock-like cataract: a mutation in CRYBB2 on chromosome 22q11.2. *Invest Ophthalmol Vis Sci* **41**:159-165

Glaser T, Jepeal L, Edwards JG, Young SR, Favor J, Maas RL. (1994) PAX6 gene dosage effect in a family with congenital cataract, aniridia, anophthalmia and central nervous systme defects. *Nat Genet* **7**:463-471

Gomes I, Aumuller G, Wennemuth G, Bette M, Albrecht M (2003) Independent signals determine the subcellular localization of NEP in prostate cancer cells. *Biochem Biophys Res Commun* **310**:919-26

Hammond CJ, Duncan DD, Snieder H, de Lange M, West SK, Spector TD, Gilbert CE. (2001) The heritability of age-related cortical cataract: the twin eye study. *Invest Ophthalmol Vis Sci* **42**:601-605

Hammond CJ, Snieder H, Spector TD, Gilbert CE. (2000) Genetic and environmental factors in age-related nuclear cataracts in monozygotic and dizygotic twins. *N Engl J Med* **342**:1786-1790

Hayashi K. (1999) Recent enhancements in SSCP. *Genet Anal* **14**:193-196

Hayashi K, Yandell DW. (1993) How sensitive is PCR-SSCP? *Hum Mutat* **2**:338-346

Head MW, Hurwitz L, Kegel K, Goldman JE. (2000) αB-crystallin regulates intermediate filament organization *in situ*. *NeuroReport* **11**:361-365

Heiba IM, Elston RC, Klein BE, Klein R. (1995) Evidence for a major gene for cortical cataract. *Invest Ophthalmol Vis Sci* **36**:227-235

Hejtmancik JF. (1998) The genetics of cataract: our vision becomes clearer [editorial]. *Am J Hum Genet* **62**:520-525

Tsifildis C. (1998) Gene localization for aculeiform cataract, on chromosome 2q33-35 [published erratum appears in Am J Hum Genet 1999 Jan;64(1):334]. *Am J Hum Genet* **63**:921-926

Héon E, Paterson AD, Fraser M, Billingsley G, Priston M, Balmer A, Schorderet DF, Verner A, Hudson TJ, Munier FL. (2001) A progressive autosomal recessive cataract locus maps to chromosome 9q13-q22. *Am J Hum Genet* **68**:772-777.

Héon E, Priston M, Schorderet DF, Billingsley GD, Girard PO, Lubsen N, Munier FL. (1999) The gamma-crystallins and human cataracts: a puzzle made clearer. *Am J Hum Genet* **65**:1261-1267

Horan MB, Billson FA. (1974) X-linked cataract and hutchinsonian teeth. *Aust Paediat J* **10**:98-102

Horwitz J, Bova MP, Ding LL, Haley DA, Stewart PL. (1999) Lens alpha-crystallin: function and structure. *Eye* **13**:403-408

Huopaniemi L, Rantala A, Forsius H, Somer M, de la Chapelle A, Alitalo T. (1999) Three widespread founder mutations contribute to high incidence of X-linked juvenile retinoschisis in Finland. *Eur J Hum Genet* **7**:368-376

Huopaniemi L, Tyynismaa H, Rantala A, Rosenberg T, Alitalo T. (2000) Characterization of two unusual RS1 gene deletions segregating in Danish retinoschisis families. *Hum Mutat* **16**:307-314

Ionides A, Berry V, Mackay D, Shiels A, Bhattacharya S, Moore A. (1998) Anterior polar cataract: clinical spectrum and genetic linkage in a single family. *Eye* **12**:224-226

Ionides A, Francis P, Berry V, Mackay D, Bhattacharya S, Shiels A, Moore A. (1999) Clinical and genetic heterogeneity in autosomal dominant cataract. *Br J Ophthalmol* **83**:802-808

Ionides AC, Berry V, Mackay DS, Moore AT, Bhattacharya SS, Shiels A. (1997) A locus for autosomal dominant posterior polar cataract on chromosome 1p. *Hum Mol Genet* **6**:47-51

Jakobs PM, Hess JF, Fitzgerald PG, Kramer P, Weleber RG, Litt M. (2000) Autosomal-dominant congenital cataract associated with a deletion mutation in the human beaded filament protein gene BFSP2. *Am J Hum Genet* **66**:1432-1436

Jamieson RV, Perveen R, Kerr B, Carette M, Yardley J, Héon E, Wirth MG, van Heyningen V, Donnai D, Munier F, Black GC. (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.

Jones AC, Sampson JR, Hoogendoorn B, Cohen D, Cheadle JP. (2000) Application and evaluation of denaturing HPLC for molecular genetic analysis in tuberous sclerosis. *Hum Genet* **106**:663-668

Kannabiran C, Balasubramanian D. (2000) Molecular genetics of cataract. *Indian J Ophthalmol* **48**:5-13

Kannabiran C, Rogan PK, Olmos L, Basti S, Rao GN, Kaiser-Kupfer M, Hejtmancik JF. (1998) Autosomal dominant zonular cataract with sutural opacities is associated with a splice mutation in the betaA3/A1-crystallin gene. *Mol Vis* **4**:21

King LS, Yasui M. (2002) Aquaporins and idsease: lessons from mice to humans. *Trends Endocrinol Metab* **13**:355-360

Kmoch B, Asfaw K, Bezouska J, Brynda J, Sedlacek J, Filipec M, Elleder M. (1999) Cataract due to crystal deposits of 37RS mutated gamma-D-crystalline. *Am J Hum Genet* **65**:A305

Kmoch S, Brynda J, Asfaw B, Bezouska K, Novak P, Rezacova P, Ondrova L, Filipec M, Sedlacek J, Elleder M. (2000) Link between a novel human gammaD-crystallin allele and a unique cataract phenotype explained by protein crystallography. *Hum Mol Genet* **9**:1779-1786

Kobelt G, Lundstrom M, Stenevi U. (2002) Cost-effectiveness of cataract surgery. Method to assess cost-effectiveness using registry data. *J Cataract Refract Surg* **28**:1742-1749

Kong A, Cox NJ. (1997) Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* **61**:1179-1188

Korf I, Flicek P, Duan D, Brent MR. (2001) Integrating genomic homology into gene structure prediction. *Bioinformatics* **17**:S140-148

Krill AE, Woodbury G, Bowman JE. (1969) X-chromosomal-linked sutural cataracts. *Am J Ophthalmol* **68**:867-872

Kumar LV, Ramakrishna T, Rao CM. (1999) Structural and functional consequences of the mutation of a conserved arginine residue in alphaA and alphaB crystallins. *J Biol Chem* **274**:24137-24141

Lander E, Kruglyak L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**:241-247

Larsen WJ (2001) Human Embryology, 3rd ed. Churchill Livingstone, Philadelphia, USA

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci U S A 81:3443-6

Lauderdale JD, Wilensky JS, Oliver ER, Walton DS, Glaser T. (2000) 3' deletions cause aniridia by preventing PAX6 gene expression. *Proc Natl Acad Sci U S A* **97**:13755-13759

Lewis RA, Nussbaum RL, Stambolian D. (1990) Mapping X-linked ophthalmic diseases. IV. Provisional assignment of the locus for X-linked congenital cataracts and microcornea (the Nance- Horan syndrome) to Xp22.2-p22.3. *Ophthalmology* **97**:110-120

Litt M, Carrero-Valenzuela R, LaMorticella DM, Schultz DW, Mitchell TN, Kramer P, Maumenee IH. (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, Murphey W, Lovrien E, Weleber R. (1998) Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Hum Mol Genet* **7**:471-474

Liu HX, Zhang M, Krainer AR. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev* **12**:1998-2012

Mackay D, Ionides A, Kibar Z, Rouleau G, Berry V, Moore A, Shiels A, Bhattacharya S. (1999) Connexin46 mutations in autosomal dominant congenital cataract. *Am J Hum Genet* **64**:1357-1364

Mackay DS, Boskovska OB, Knopf HL, Lampi KJ, 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

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

Marner E, Rosenberg T, Eiberg H. (1989) Autosomal dominant congenital cataract. Morphology and genetic mapping. *Acta Ophthalmol* **67**:151-158

Mitchell P, Cumming RG, Attebo K, Panchapakesan J. (1997) Prevalence of cataract in Australia: the Blue Mountains eye study. *Ophthalmology* **104**:581-588

Montini E, Andolfi G, Caruso A, Buchner G, Walpole SM, Mariani M, Consalez G, Trump D, Ballabio A, Franco B. (1998) Identification and characterization of a novel serine-threonine kinase gene from the Xp22 region. *Genomics* **51**:427-433

Montini E, Buchner G, Spalluto C, Andolfi G, Caruso A, den Dunnen JT, Trump D, Rocchi M, Ballabio A, Franco B. (1999) Identification of SCML2, a second human gene homologous to the Drosophila sex comb on midleg (Scm): A new gene cluster on Xp22. *Genomics* **58**:65-72.

Montini E, Rugarli EI, Van de Vosse E, Andolfi G, Mariani M, Puca AA, Consalez GG, den Dunnen JT, Ballabio A, Franco B. (1997) A novel human serine-threonine phosphatase related to the Drosophila retinal degeneration C (rdgC) gene is selectively expressed in sensory neurons of neural crest origin. *Hum Mol Genet* **6**:1137-1145

Nance WE, Warburg M, Bixler D, Helveston EM. (1974) Congenital X-linked cataract, dental anomalies and brachymetacarpalia. *Birth Defects Orig Artic Ser* **10**:285-291

Nyholt DR. (2000) All LODs are not created equal. *Am J Hum Genet* **67**:282-288

Ott J. (1989) Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci U.S.A.* **86**:4175-4178

Ottonello S, Foroni C, Carta A, Petrucco S, Maraini G. (2000) Oxidative stress and age-related cataract. *Ophthalmologica* **214**:78-85

Pal JD, Berthoud VM, Beyer EC, Mackay D, Shiels A, Ebihara L (1999) Molecular mechanism underlying a Cx50-linked congenital cataract [published erratum appears in Am J Physiol 1999 Dec;277(6 Pt 1):section C following table of contents]. *Am J Physiol* **276**:C1443-6

Pande A, Pande J, Asherie N, Lomakin A, Ogun O, King JA, Lubsen NH, Walton D, Benedek GB (2000) Molecular basis of a progressive juvenile-onset hereditary cataract. *Proc Natl Acad Sci U S A* **97**:1993-8

Pande A, Pande J, Asherie N, Lomakin A, Ogun O, King J, Benedek GB (2001) Crystal cataracts: human genetic cataract caused by protein crystallization. *Proc Natl Acad Sci U S A* **98**:6116-20

Parra G, Blanco E, Guigo R. (2000) GeneID in Drosophila. *Genome Res* **10**:511-515.

Payseur BA, Nachman MW. (2000) Microsatellite variation and recombination rate in the human genome. *Genetics* **156**:1285-1298

Phelan JK, Bok D. (2000) A brief review of retinitis pigmentosa and the identified retinitis pigmentosa genes. *Mol Vis* **6**:116-124

Polyakov AV, Shagina IA, Khlebnikova OV, Evgrafov OV. (2001) Mutation in the connexin 50 gene (GJA8) in a Russian family with zonular pulverulent cataract. *Clin Genet* **60**:476-478

Pras E, Bakhan T, Levy-Nissenbaum E, Lahat H, Assia EI, Garzozi HJ, Kastner DL, Goldman B, Frydman M. (2001) A gene causing autosomal recessive cataract maps to the short arm of chromosome 3. *Isr Med Assoc J* **3**:559-562

Pras E, Frydman M, Levy-Nissenbaum E, Bakhan T, Raz J, Assia EI, Goldman B. (2000) A nonsense mutation (W9X) in CRYAA causes autosomal recessive cataract in an inbred Jewish Persian family. *Invest Ophthalmol Vis Sci* **41**:3511-3515

Pras E, Levy-Nissenbaum E, Bakhan T, Lahat H, Assia E, Geffen-Carmi N, Frydman M, Goldman B. (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**:1363-1367

Rahi JS, Dezateaux C. (2001) Measuring and interpreting the incidence of congenital ocular anomalies: lessons from a national study of congenital cataract in the UK. *Invest Ophthalmol Vis Sci* **42**:1444-1448

Rees MI, Watts P, Fenton I, Clarke A, Snell RG, Owen MJ, Gray J. (2000) Further evidence of autosomal dominant congenital zonular pulverulent cataracts linked to 13q11 (CZP3) and a novel mutation in connexin 46 (GJA3). *Hum Genet* **106**:206-209

Reese MG. (2001) Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Comput Chem* **26**:51-56

Ren Z, Li A, Shastry BS, Padma T, Ayyagari R, Scott MH, Parks MM, Kaiser-Kupfer MI, Hejtmancik JF. (2000) A 5-base insertion in the gammaC-crystallin gene is associated with autosomal dominant variable zonular pulverulent cataract. *Hum Genet* **106**:531-537

Runge PE, Hawes NL, Heckenlively JR, Langley SH, Roderick TH. (1992) Autosomal dominant mouse cataract (Lop-10). Consistent differences of expression in heterozygotes. *Invest Ophthalmol Vis Sci* **33**:3202-3208

Salamov AA, Solovyev VV. (2000) Ab initio gene finding in Drosophila genomic DNA. *Genome Res* **10**:516-522

Santhiya ST, Shyam Manohar M, Rawlley D, Vijayalakshmi P, Namperumalsamy P, Gopinath PM, Loster J, Graw J. (2002) Novel mutations in the gamma-crystallin genes cause autosomal dominant congenital cataracts. *J Med Genet* **39**:352-358

Semina EV, Ferrell RE, Mintz-Hittner HA, Bitoun P, Alward WL, Reiter RS, Funkhauser C, Daack-Hirsch S, Murray JC. (1998) A novel homeobox gene PITX3 is mutated in families with autosomal-dominant cataracts and ASMD. *Nat Genet* **19**:167-170

Shiels A, Mackay D, Ionides A, Berry V, Moore A, Bhattacharya S. (1998) A missense mutation in the human connexin50 gene (GJA8) underlies autosomal dominant "zonular pulverulent" cataract, on chromosome 1q. *Am J Hum Genet* **62**:526-532

Shroff NP, Cherian-Shaw M, Bera S, Abraham EC. (2000) Mutation of R116C results in highly oligomerized alpha A-crystallin with modified structure and defective chaperone-like function. *Biochemistry* **39**:1420-1426

Slingsby C, Clout NJ. (1999) Structure of the crystallins. *Eye* **13**:395-402

Stambolian D, Lewis RA, Buetow K, Bond A, Nussbaum R. (1990) Nance-Horan syndrome: localization within the region Xp21.1-Xp22.3 by linkage analysis. *Am J Hum Genet* **47**:13-19

Stephan DA, Gillanders E, Vanderveen D, Freas-Lutz D, Wistow G, Baxevanis AD, Robbins CM, VanAuken A, Quesenberry MI, Bailey-Wilson J, Juo SH, Trent JM, Smith L, Brownstein MJ. (1999) Progressive juvenile-onset punctate cataracts caused by mutation of the gammaD-crystallin gene. *Proc Natl Acad Sci U.S.A.* **96**:1008-1012

Thylefors B, Negrel A-D, Pararajasegaram R, Dadzie KY. (1995) Global data on blindness. *Bull World Health Organ* **73**:115-121

Toutain A, Ayrault AD, Moraine C. (1997a) Mental retardation in Nance-Horan syndrome: clinical and neuropsychological assessment in four families. *Am J Med Genet* **71**:305-314

Toutain A, Dessay B, Ronce N, Ferrante MI, Tranchemontagne J, Newbury-Ecob R, Wallgren-Pettersson C, Burn J, Kaplan J, Rossi A, Russo S, Walpole I, Hartsfield JK, Oyen N, Nemeth A, Bitoun P, Trump D, Moraine C, Franco B. (2002) Refinement of the NHS locus on chromosome Xp22.13 and analysis of five candidate genes. *Eur J Hum Genet* **10**:516-520

Toutain A, Ronce N, Dessay B, Robb L, Francannet C, Le Merrer M, Briard ML, Kaplan J, Moraine C. (1997b) Nance-Horan syndrome: linkage analysis in 4 families refines localization in Xp22.31-p22.13 region. *Hum Genet* **99**:256-261

Uhlmann K, Brinckmann A, Toliat MR, Ritter H, Nurnberg P. (2002) Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* **23**:4072-9

van de Vosse E, Walpole SM, Nicolaou A, van der Bent P, Cahn A, Vaudin M, Ross MT, Durham J, Pavitt R, Wilkinson J, Grafham D, Bergen AA, van Ommen GJ, Yates JR, den Dunnen JT, Trump D. (1998) Characterization of SCML1, a new gene in Xp22, with homology to developmental polycomb genes. *Genomics* **49**:96-102

Vanita, Sarhadi V, Reis A, Jung M, Singh D, Sperling K, Singh JR, Burger J. (2001a) A unique form of autosomal dominant cataract explained by gene conversion between beta-crystallin B2 and its pseudogene. *J Med Genet* **38**:392-396.

Vanita, Singh JR, Sarhadi VK, Singh D, Reis A, Rueschendorf F, Becker-Follmann J, Jung M, Sperling K. (2001b) A novel form of "central pouchlike" cataract, with sutural opacities, maps to chromosome 15q21-22. *Am J Hum Genet* **68**:509-514

Velasco PT, Lukas TJ, Murthy SN, Duglas-Tabor Y, Garland DL, Lorand L. (1997) Hierarchy of lens proteins requiring protection against heat-induced precipitation by the alpha crystallin chaperone. *Exp Eye Res* **65**:497-505.

Vicart P, Caron A, Guicheney P, Li Z, Prevost MC, Faure A, Chateau D, Chapon F, Tome F, Dupret JM, Paulin D, Fardeau M. (1998) A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy. *Nat Genet* **20**:92-95

Walpole IR, Hockey A, Nicoll A. (1990) The Nance-Horan syndrome. *J Med Genet* **27**:632-634

Walpole SM, Ronce N, Grayson C, Dessay B, Yates JR, Trump D, Toutain A. (1999) Exclusion of RAI2 as the causative gene for Nance-Horan syndrome. *Hum Genet* **104**:410-411

Watts P, Rees M, Clarke A, Beck L, Lane C, Owen MJ, Gray J. (2000) Linkage analysis in an autosomal dominant 'zonular nuclear pulverulent' congenital cataract, mapped to chromosome 13q11-13. *Eye* **14**:172-175

Weatherall DJ. (2001) Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet* **2**:245-255

Weeks DE, Ott J, Lathrop GM. (1990) SLINK: a general simulation program for linkage analysis. *Am J Hum Genet* **47**:A204 (Supplement)

Weinreb O, van Rijk AF, Dovrat A, Bloemendal H. (2000) In vitro filament-like formation upon interaction between lens α-crystallin and βL-crystallin promoted by stress. *Invest Ophthalmol Vis Sci* **41**:3893-3897

Werten PJ, Carver JA, Jaenicke R, de Jong WW. (1996) The elusive role of the N-terminal extension of beta A3- and beta A1-crystallin. *Protein Eng* **9**:1021-1028

Wirth MG, Russell-Eggitt IM, Craig JE, Elder JE, Mackey DA. (2002) Aetiology of congenital and paediatric cataract in an Australian population. *Br J Ophthalmol* **86**:782-786

Wistow G, Turnell B, Summers L, Slingsby C, Moss D, Miller L, Lindley P, Blundell T. (1983) X-ray analysis of the eye lens protein gamma-II crystallin at 1.9A resolution. *J Mol Biol*; **170**:175-202

Wolf MT, Lorenz B, Winterpacht A, Drechsler M, Schumacher V, Royer-Pokora B, Blankenagel A, Zabel B, Wildhardt G. (1998) Ten novel mutations found in Aniridia. *Hum Mutat* **12**:304-313

Yamada K, Tomita H, Yoshiura K, Kondo S, Wakui K, Fukushima Y, Ikegawa S, Nakamura Y, Amemiya T, Niikawa N. (2000) An autosomal dominant posterior polar cataract locus maps to human chromosome 20p12-q12. *Eur J Hum Genet* **8**:535-539

Yamashita S, Furumoto K, Nobukiyo A, Kamohara M, Ushijima T, Furukawa T. (2002) Mapping of a gene responsible for cataract formation and its modifier in the UPL rat. *Invest Ophthalmol Vis Sci* **43**:3153-3159

Yang Z, Lin W, Pellerano G, Valdez-Guerrero ME, Pellarono-Noboa G, Travera D, Caraballo R, Jiang L, Thirumalaichary S, Pan J, Zhang K. (2002) A novel

mutation in PITX3 associated with an autosomal dominant form of congenital cataract. *Am J Hum Genet* **71**:525

Zhu D, Alcorn DM, Antonarakis SE, Levin LS, Huang PC, Mitchell TN, Warren AC, Maumenee IH. (1990) Assignment of the Nance-Horan syndrome to the distal short arm of the X chromosome. *Hum Genet* **86**:54-58

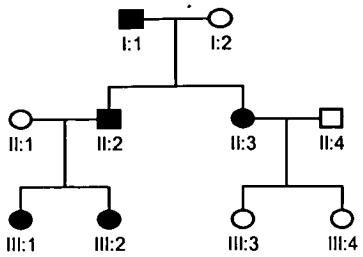
Zhu D, Li Y, Traboulsi EI, Mitchell TN, Maumenee IH. (1998) Refined mapping of the Nance-Horan syndrome to a 2cM region on Xp22.2. *Am J Med Genet* **63**:A317

Appendix 1 - Pedigrees with Familial Cataract

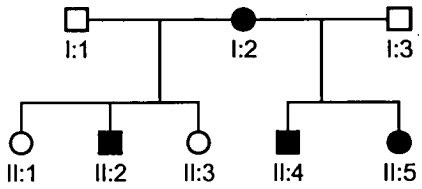
The pedigree's number in the study is given along with the type of inheritance observed:

ADCC – Autosomal Dominant Congenital Cataract (includes all dominant inheritance)
ARCC – Autosomal Recessive Congenital Cataract
NHS – Nance-Horan Syndrome

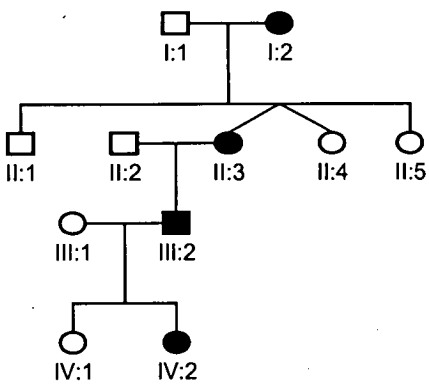
Crch01-Aniridia



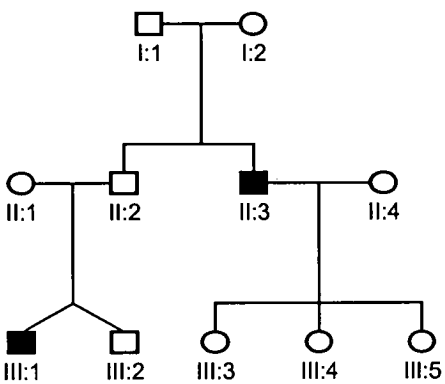
Crch02-X-linked



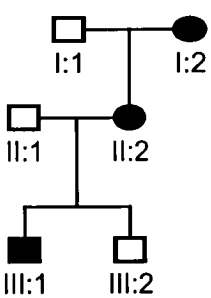
Crch03-ADCC



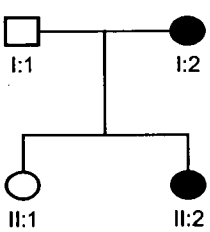
Crch04-uncertain



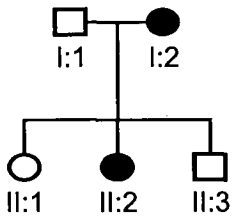
Crch05-X-linked



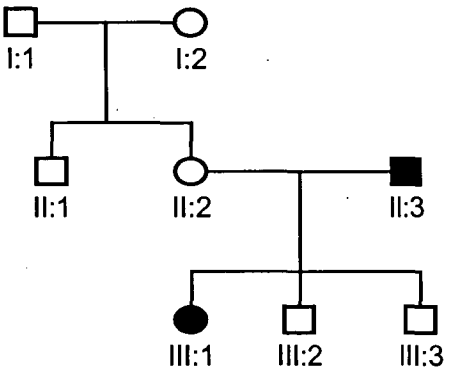
Crch06-ADCC



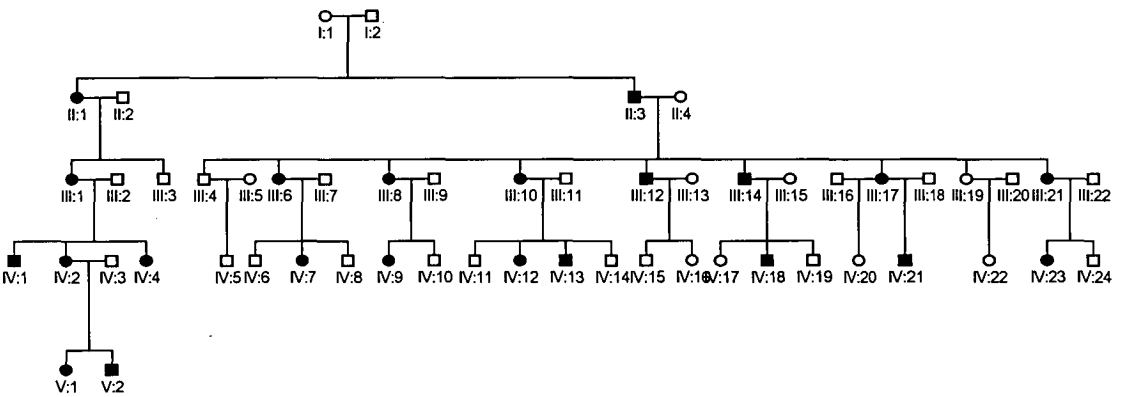
Crch07-ADCC



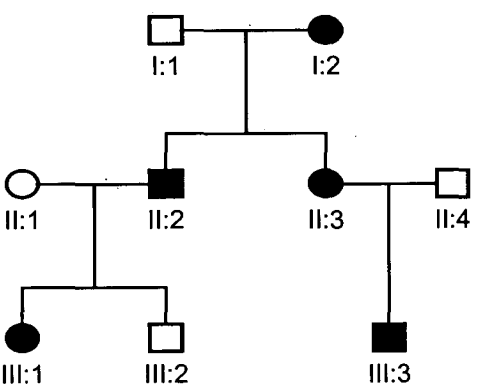
Crch09-ADCC.



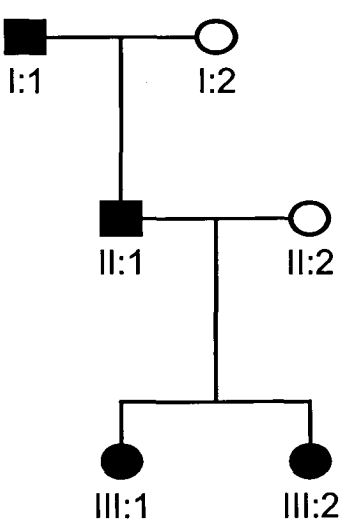
Crch08-ADCC



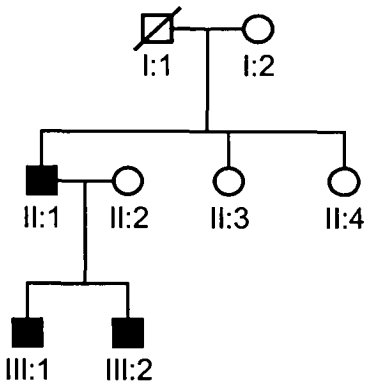
Crch10-ADCC



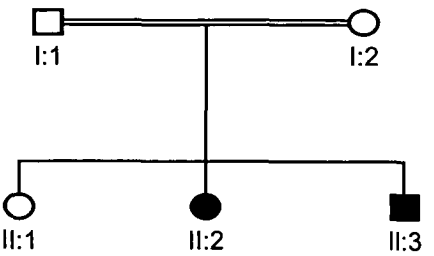
Crch11-ADCC



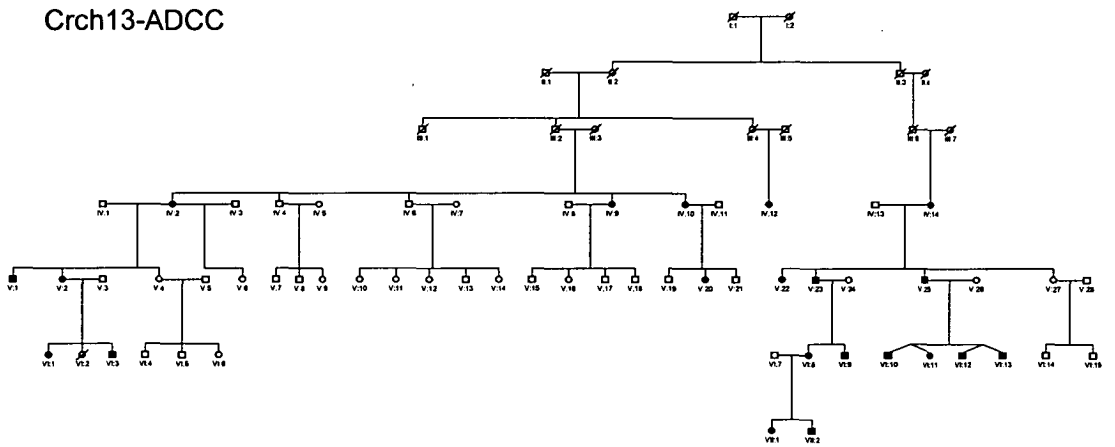
Crch12-ADCC



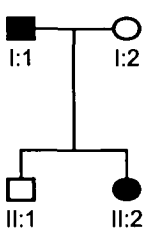
Crch14-ARCC



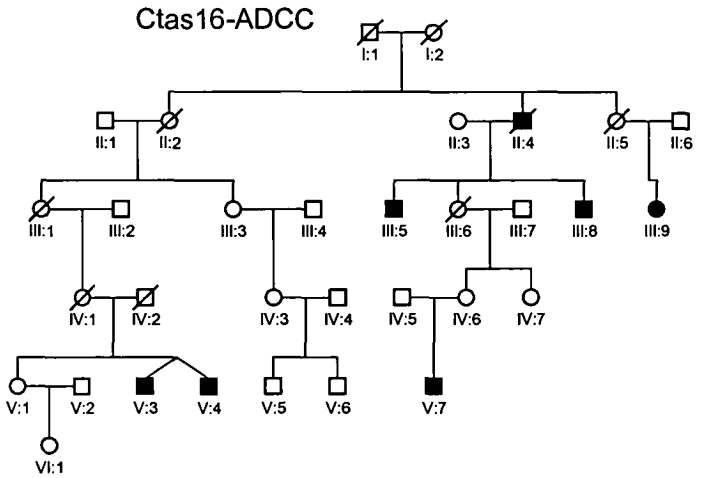
Crch13-ADCC



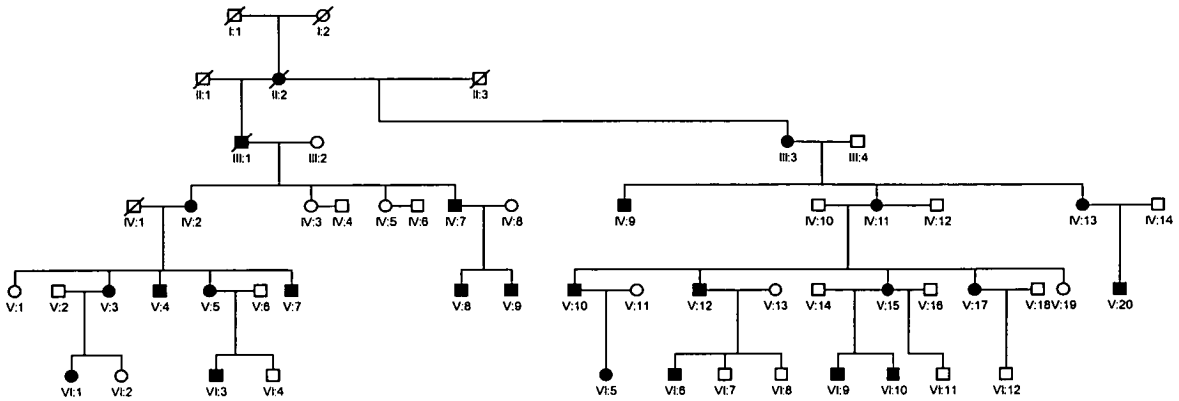
Crch15-ADCC



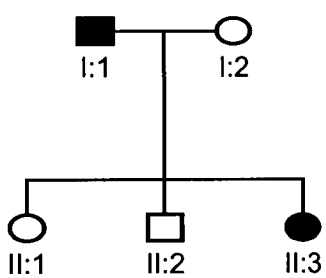
Ctas16-ADCC



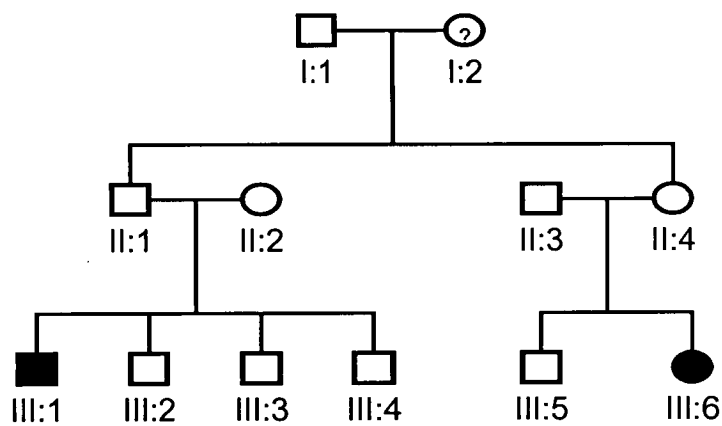
Ctas17-ADCC



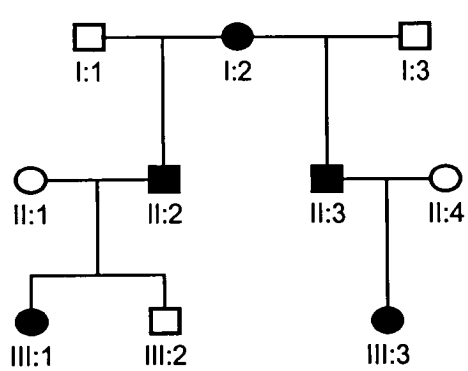
Crch18-ADCC



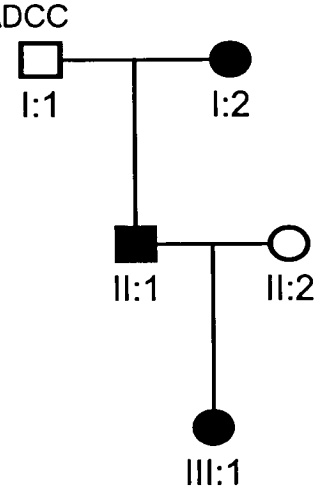
Crch19-uncertain



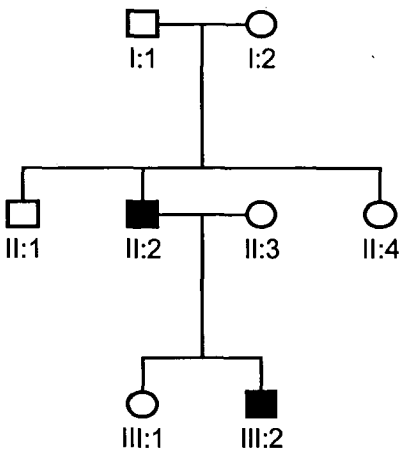
Crch20-X-linked



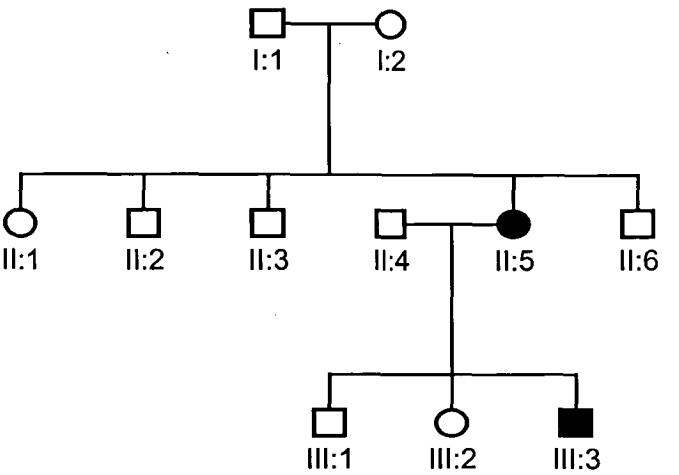
Crch21-ADCC



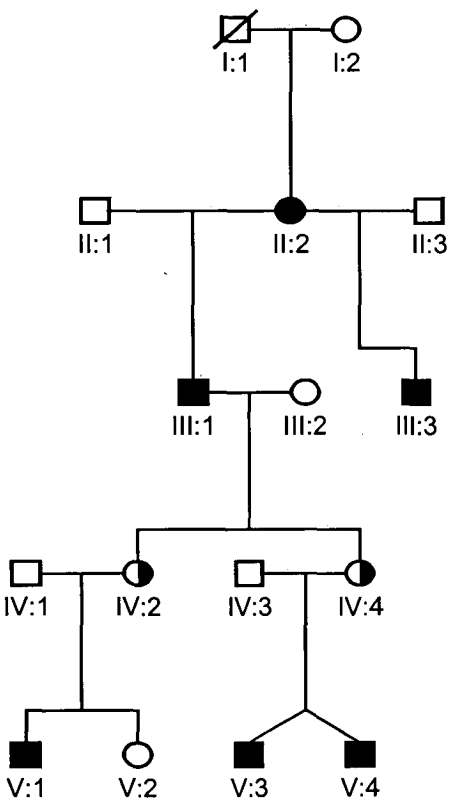
Crch22-ADCC



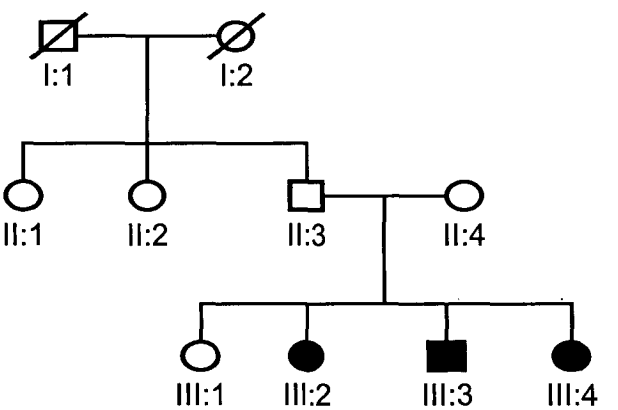
Crch23-ADCC



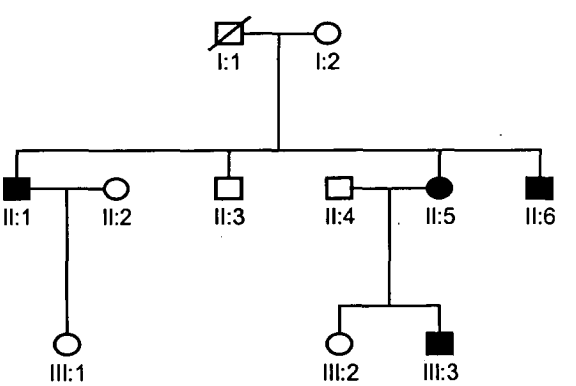
Crch24-X-linked



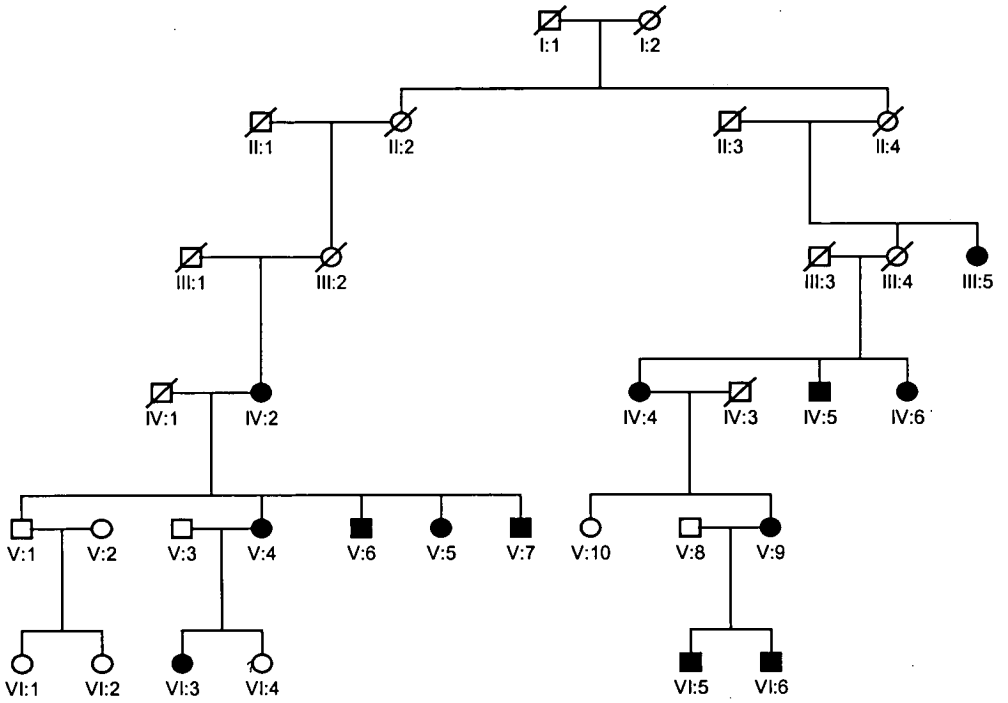
Crch26-ARCC



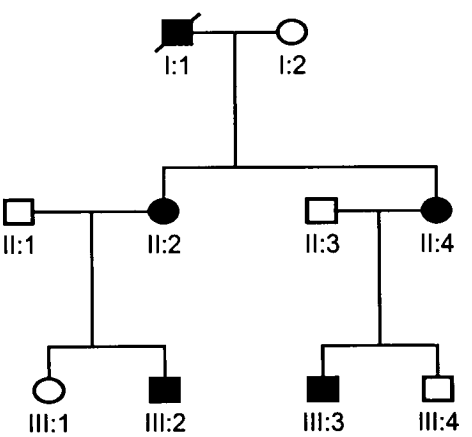
Crch27-ADCC



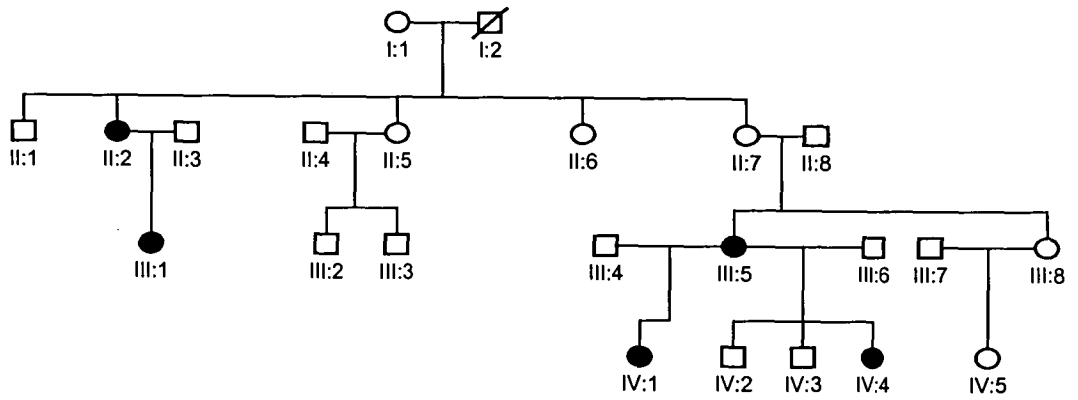
Crch2551-NHS



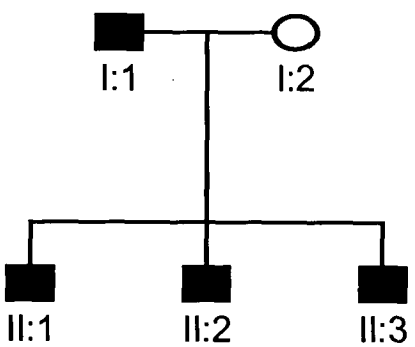
Crch28-ADCC



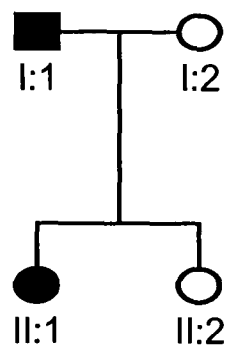
Crch30-ADCC



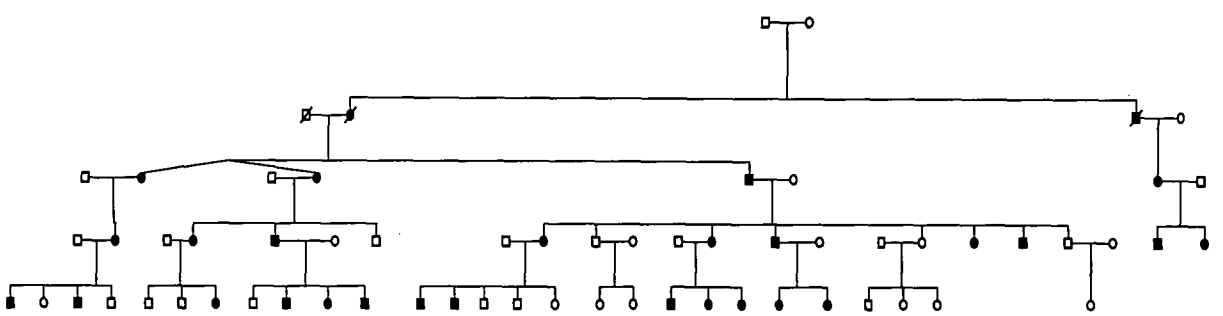
Crch31-ADCC



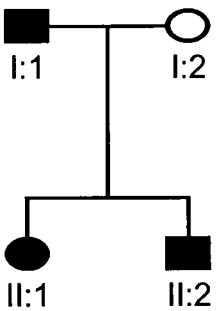
Crch33-ADCC



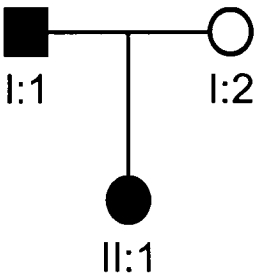
Crch32-ADCC



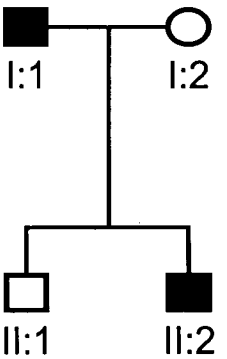
Ctas34-ADCC



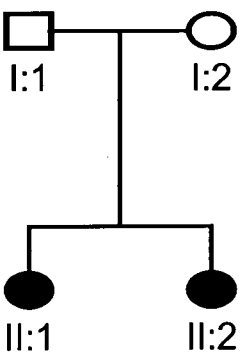
Ctas35-ADCC



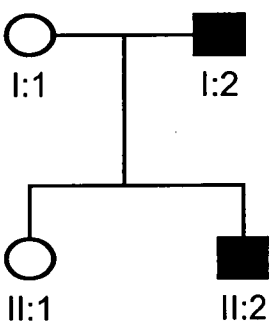
Ctas36-ADCC



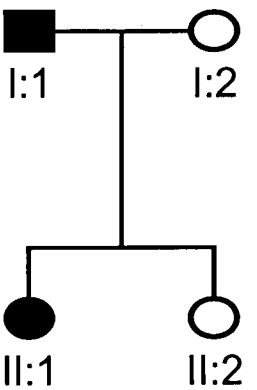
Crch37-uncertain



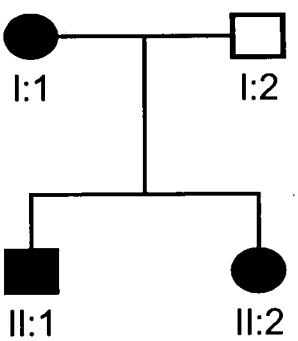
Crch38-ADCC



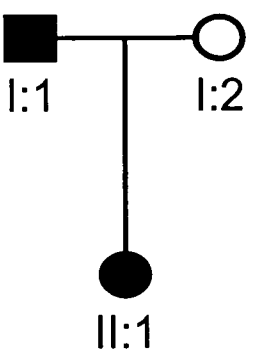
Crch39-ADCC



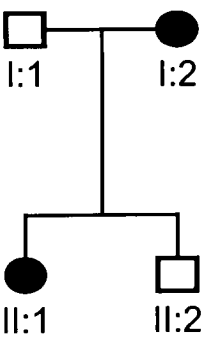
Crch40-ADCC



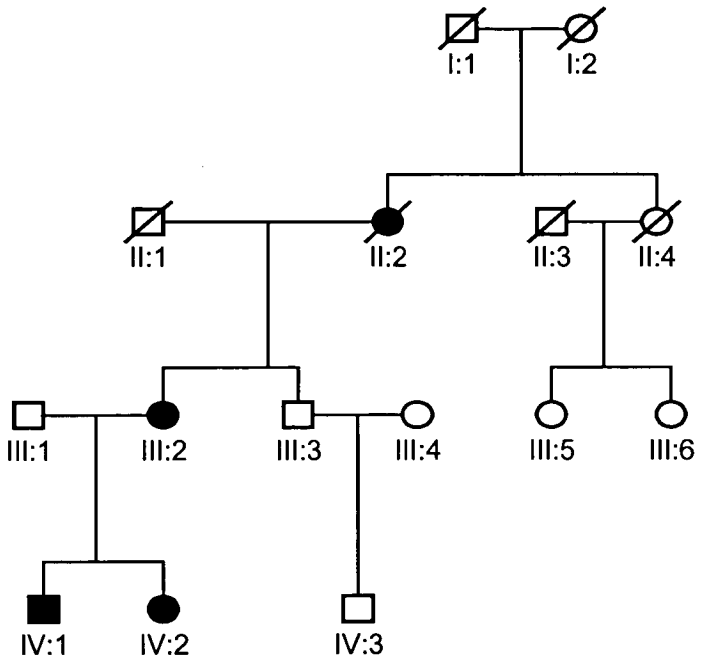
Crch41-ADCC



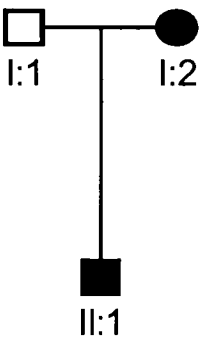
Crch43-ADCC



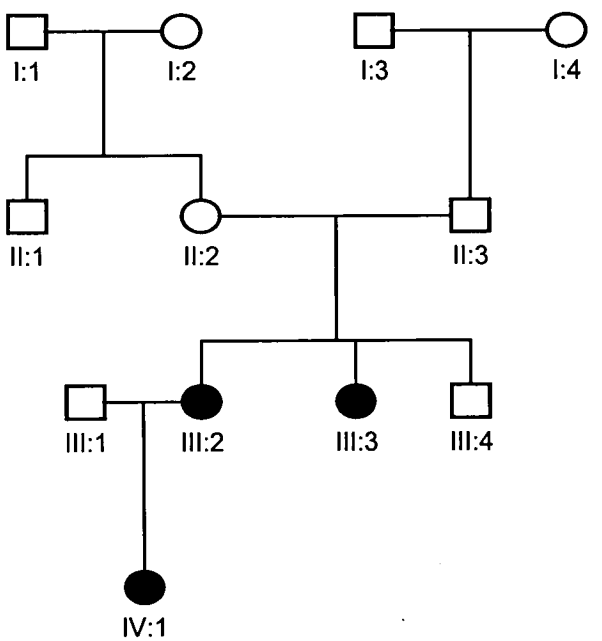
Ceeh42-ADCC



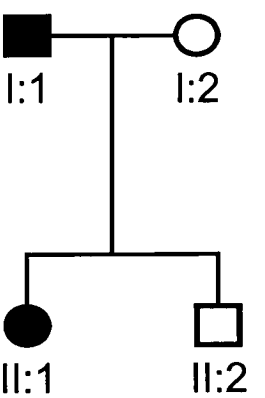
Crch44-ADCC



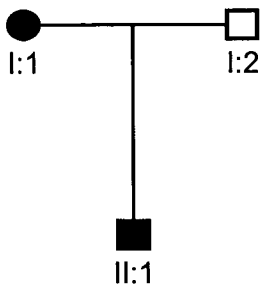
Crch45-uncertain



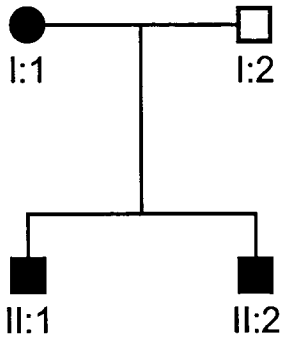
Crch46-ADCC



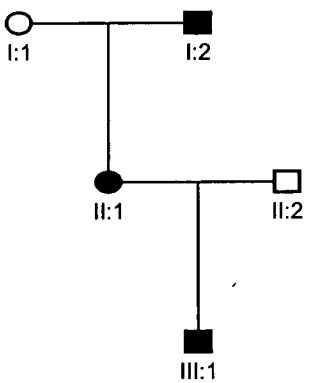
Cwa70-NHS



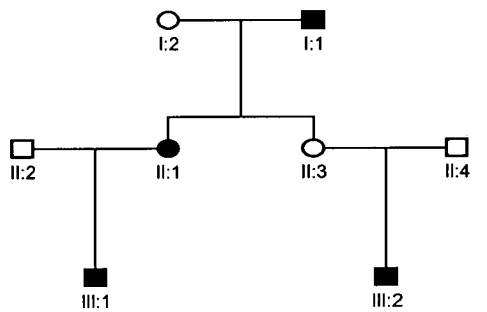
UK1-NHS



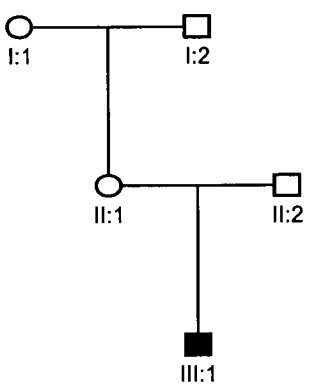
UK2-X-linked



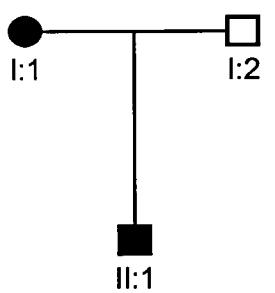
UK3-X-linked



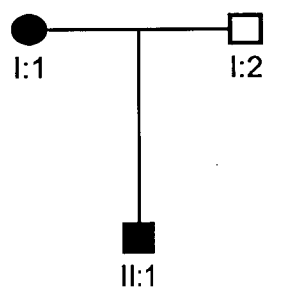
UK4-NHS



UK5-NHS



UK6-NHS



Appendix 2 – Primer Sequences and PCR Conditions

All primers were designed using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi).

Primer sequences for microsatellite markers were obtained from the Genome DataBase (http://www.gdb.org/) and are not listed here.

Section 3.3.4 Primer sequences for novel STR markers on the X chromosome. STRs are named after the repeated sequence followed by the clone name. Optimal Mg^{2+} concentration and annealing temperature (°C) is also given.

Marker	Forward	Reverse	Mg2+	Temp
TG Z93022	TGAGGAAAATATCTCATTGCAGAA	TGGCTCTTGGAGGGTATTTG	1.5	60
GA Z93242	TGCAGCAAGAGATGAAGACAA	TATTTATTGCAGCCCCCTTG	1.5	52
CA Z93929	GCAGTGGAAGAGTCCCTCAG	CCCGATTTCCTCTAAAACAGTG	1.5	60
GT Z92542	TCACTGACTTGGCTTGAAGG	TGAGACAGCAGGTAGCCAAA	1.5	60
CA1 Z92542	TCCTAATTTCATTGATTTCCAGGT	TGAGGCCTATTTGGATTTTGTT	1.5	60
CA2 Z92542	CCTGCTCTTGGGGAAGATAA	TGCTTCTGCTCACAATGACC	1.5	60
TG/AG Z92542	AATGAATGGTTTCGGCTGTG	ATGCGCTAATTCCTGGTGAG	1.5	60
GT Z94056	AGCCTGGCCCTGTTCTAGTT	TTCTCTCTGGGGAAATGCAG	1.5	60

Section 3.3.5 Primers for the amplification of coding exons of candidate genes for NHS. Optimal Mg^{2+} concentration and annealing temperature (°C) is also given.

Gene	Exon	Forward Primer	Reverse Primer	Mg2+	Temp
PPEF1	4	CCCTCCCTCCTGTGTATTCC	TCATTCATAACCACAGGCACA	1.5	60
	5	AATCCAGGTGGACTGCTTTTT	GAAGAGAAAAGCTTGCAAAACA	1.5	60
	6	GTGGAACACCAACAGCATTT	ACCGCTCTTGATGAAGACAA	1.5	60
	7	AGTGCCTTACATGGGCTAGG	TTCAACAGTTATTGGGCATCTG	1.5	61
	8	CGTGCTTGCTCCACCTTTAC	AGGCAAACGATGTAGGACCA	1.5	60
	9	AATGCCCTGAGAACACTTGG	CCTCCCCCAAGAAAAATGTTA	1.5	60
	10	TTTTCCCTTCTCCTCCTTC	ATGGCAGCATTTTCCAGAGT	1.5	57
	11	GCTTCCCTTTTTGCATGACT	CAATCTGGTCTTTCTTGGCTCT	1.5	60
	12	CTGAACCTGTCAGAAAAACAAAAC	TCCCTGAGAGAAGATTCATGC	1.5	60
	13	GCAGAGGGTTGGACTAAAGC	AGTGTCATTGTTGCCACTGC	1.5	61
	14	CAGCGAAATCTGGTTTGTTGT	TGGGAATGTCACCAACAATTT	1.5	60
	15	CGTGAACCTGACCCTCTTTT	TCCGTGATACCAGTGCTCAG	1.5	60
	16	AAAATGAAACACAACAGGATGA	GGAAGCAAGGAAAATGAAAAG	1.5	60
	17	CCAAGGAGGTTGCATTCTTT	CCCTGGCTAGGTTTTAGCTG	1.5	60
	18	TGCTAAAAGCGATCAGTGTCA	CCAATGAAGGTGTCAAACCA	1.5	60
	19	CCTTGCCTTAGGTGGGGTCTT	ATTTGGGCCTAGCTGTTTCA	1.5	60
STK9	2	CTTCCTTCAGACGGTTTTGG	GGAGAGTGAGACCCAAATTGTT	1.5	60
	3	CCACTGAGAAGCAATGTCAGTATAG	TACATGCCCACACGCAAAG	1.5	60
	4	GAATCCCCAGTCGGAAAAAC	TGACCAGCTAGATCCCACTTC	1.5	60
	5	TGGAATTCTTTGAATAGTAGCTTGAA	AAATAAAAGAATCGGGCAAATG	1.5	60
	6	GGCCTACCTAATTTGGGAAA	GCAATGCCAATGCTTATGAA	1.5	60
	7	TTTTTGCTGCCACAGTTTTC	TGTGACTCAAAAGAATGTTCCTC	1.5	60
	8	GCTGAAATATTTTGCCCACA	TCAGCAGATGTGGAAATGTCA	1.5	60
	9	TTTTTCAGTTGCCAAAATAATCT	TGTGCAATGAACAATGACTCAA	1.5	60
	10	CATGTCCTTCCCCAAATGTT	AAAAGCAGGCTATGGTCACA	1.5	60

	11	TTTGACTTTGTAATGTTCTTAACGAT	AGCCACCTCCTCCACCTACT	1.5	60
	12A	TGTCAGCTATTGAGGGAAACTG	AAGGTTGCTCACAGACTTGGA	1.5	60
	12B	GAGTCCCTCCTACAGGACCA	GCCACCAGATTCAGTCAAGG	1.5	60
	13	CATCCCTAGGAACAGGATGC	GGCAGTAATGATGTTGAACAAA	1.5	60
	14	GCAATATTGTCATCAATGTGTGG	AGGAGAGGGGAACCTGTTGA	1.5	60
	15	CGTGAAAACTGCTTTGACCA	AAGGCCTAGCAGGAGAAAGG	1.5	60
	16	TTTGTCACACAATGGCAAGAA	TTTGATTGCCAAGTGCAAAG	1.5	55
	17	CTCTTGGGTGTGGTTGCATA	CCGCTCCTCAGGACAGTTAC	1.5	60
	18	CCAGCCTTATGGTCGCTCTA	CACAGAGGACACATGCCAAC	1.5	55
	19	GTGGGCAGAAGTGGCCAATA	GTCTAGGGTCGTTATGGCAGC	1.5	58
	20	CCCCAAGCTTCATTCCACT	CAGGTGTCATTGGCTGGAC	1.5	55
	21	TCCTGGCAGCTCTGAGTGA	AGAGGTGATTCAGGGCATTG	1.5	60
SCML2	2	TCATTTTTGTTTAAGGGAAACAG	AAAGGCTTTATTTCAAAAGTTCAGT	1.5	60
	3	TGGTAAGTCCTACATAGAAAATGAAGA	CACTAGTGAGGAAGAAAACTATGCTG	1.5	60
	4	GGCATGGAACTAAAACCCATT	TCACACAGACGTCCTCCAAC	1.5	58
	5	TTCTGGTGACATAGCTGTCCA	AAGGCCACTTAAAGCCATTTC	1.5	58
	6	TTGCCCATGTTCAAAGTTCT	TTTGCTTTTCCATGTCACAGA	1.5	58
	7	GCAACTTTTCACTTCGAGCA	AATCACAATGCCACCAATGA	1.5	58
	8	TGAGGAGACAATGCCTTAATTG	CATTTCTAAAGACCCTCATCATACT	1.5	58
	9	TCATGACACACACGGGAGTT	AGGACTTCGAAGACACTGAAAA	1.5	62
	10	GGTTTTCTGTGAGTGAAGTGGA	AGGGGGAAGGGACTGAATTA	1.5	60
	11	TCTCTCTGAGGTTACTGACATCCTT	TTTGAGGAGAACACAAGGAAAA	1.5	58
	12	CAGAAGTGTTTCATGCTGATGAT	CCAGTCCTGGGAGGTGTAAA	1.5	58
	13	TGCTAGGGATGAGATTTTTAACCT	AAAGACAAACACATTGAGAGTCAC	1.5	58
	14	AGGTTCCAGGAATGAGACCA	CCTGCACAGAGGTCATTCAA	1.5	58
	15	CAGCTGCTGTATTTTACATGTTCT	GGTTGGTGACTCCAGGAAAA	1.5	58
RS1	1	AGGCCAGGGCTCAACTTAAT	TGCAATGAATGTCAATGGTT	1.5	62
	2	AGGGGTGTTGGTAGCTTTCAT	TCCTCTATGTTATTTTTGGCTAGGA	1.5	62
	3	TTGGCCATTGTAGCAAAGC	CAGCTGAGTGGGAAAAAGGA	1.5	62
	4	GCAAAGCAGATGGGTTTGTT	GCCACGCTGGTAGAGAGG	1.5	62
	5	GGCTTTTTGCAGACATGCAC	GTGCGAGCTGAAGTTGGTTT	1.5	62
	6	GGAACCCTTCTTCCTTTTCC	GCGAAATATAGCCCTGTCCA	1.5	62
SCML1	4	AAAGCATTCTTACCGGCTTG	TGTACCAAATTGGGCTCTCC	1.5	58
	5	TTCCAACAAAGAGAGCATTCC	TCAAAGGCATTACAGCAGGA	1.5	58
	6	TCTGCAATGAGCTGCTCTCTA	TCCACACTAAGAAGGCATTGG	1.5	58
RAI2	1	TGCCATTTTCCATAGTCATAGC	CTCAGCTCTGATGCCACTTG	1.5	60
	2	TGGGTTATGCTCCCTGGTTA	TCATGACGTAGGTGGCATTG	1.5	60
	3	AGAGCCCAGTGGTGATGC	GGGGATGGGAATAGGGACT	2.0	60
	4	CTTTTAGCTCCCCCTTGTCC	CATGCCACTGGGAAGTTTCT	1.5	60
	5	AGCCTCCCCCTGAGACAC	TTTTTGATTGGGAGCATGTG	2.0	58
	6	AGAAGACTCCGTGCTTCAGG	CAAAGCACTTTGTGGCAATG	2.0	62

Section 4.2.2 Primers for amplification of predicted exons from adult human brain cDNA.

Exon numbers are those given to the exon predictions by Genescan. The Unigene clusters are those that were linked to form the 3'UTR.

Predicted exon	Forward	Reverse
5'end	GCTGCTCATGCTGGACCTA	
exon 1	CTGTGGCTTTGGAAAATGCT	
exon 2	GGAGTAGCAGGGGAGAAAGC	
exon 3	CCAACCTGGACATAGAGAGTAAGC	CGCGGTAGTACACACTCAGCTTAC
exon 4		TAAACTGACGCCTCCTCTGG
exon 4a	GACGAACATATTTTTCCCAAG	CCAACACACCCAACAGGATA
exon 5	TTCCACACTGGACAAGCAGA	TATTGGGTTTTTCGGCCTCT
exon 6	GAGACCCAAGGAAATGTGGA	TGACAGTGTTGGAGGGGTCT
exon 8		ATGTCCCCGGAATCTTTTCT
Hs.77385	TACGCAATCTCCTTCCTTGTG	
Hs.282164	GAAGGAGGAAGCCTGGTCTAAG	CAATAGCTCTGTGGCATGCTTAG
Hs.21470		CTGGAAATGACGGTTTCTTATCG

Section 4.2.3 Primers for amplification and sequencing of all detected regions of *NHS* gene.

Exon 1 was sequenced in 3 overlapping fragments (A-C), exon 6 in 8 overlapping fragments (A-H) and exon 8 as 2 overlapping coding fragments (A & B) and 7 untranslated fragments (UTR A- UTR G). Optimal Mg^{2+} concentration and annealing temperature (°C) is also given.

Exon	Forward Primer	Reverse Primer	Mg2+	Temp
1 – A*	GCGCACCCCTAAATTTCT	CTGCCTGCACAGCCATTG	2.0	60
1 – B*	GCTTGGAGGAGACCAGAAAGT	CCTGGAAGAGGCTGCAAG	2.0	62
1 – C*	GCTGCTCATGCTGGACCTA	GGTGGGAAGGCGAGAGTAGT	1.5	60
2	TGCAGTAGTCTGGACTTCCACT	CCCATGGATTTCATTTCAGG	1.5	62
3	ACTCCCAAGGGGAAAAGAGA	CTGATGTTTCCTCAGCAGCA	1.5	62
1b	TCTCTCTCTCTGATGCACCAAA	TCATGCCCTGTTGACTCAGA	1.5	60
4	GGGCCTATTTCTGACCTCATT	AGCAGCACAGATTTTGAACA	1.5	62
5	TCTTTCCTTACTTCCCGTCAAA	CATCTGTACTAGGCGGAGGAA	1.5	62
6 - A	CTGCCAGCCCACAGATCTAC	GCTTGGAGCCTCACTGAAAT	1.5	62
6 - B	TCCCCGGGAAGGTAATAGAG	TGAGGGGCTGTGTTTAGTGA	1.5	62
6 - C	TGACCACCAGTGATCCAGTG	ACCACAGTCAAAGTGCATGG	1.5	62
6 - D	CATCAGACAAAGCGGACACT	TGGTGGAATATCCGAAGCAT	1.5	62
6 - E	AATGGAAAACGCCAATCTTC	AGAGGGTTGCTGTAGTGAGGA	1.5	62
6 - F	CTTGGCATCTCCATCAAGTG	TGCGGCCTAATCTTACTTGG	1.5	62
6 - G	CCTTGCAATTACACCAACGA	GAGGGTGCCTTCTGTACCTG	1.5	62
6 - H	AAAACACGCCAACCAAAAAC	ATTCCAGGAAGTGCCATGAG	1.5	62
7	CAGTAGCGTGCTGGGTAACTT	GGGCAAAACCTTTGTTGTA	1.5	62
8 - A	TTTCATAAAAACGTGAACTGAGTGA	GCAAAGCTCTTCGAGGAAAA	1.5	62
8 - B	CCAGAGGTCTCCTGGTCTCA	GTAAGGGTTTTGGCCTTTGC	1.5	62
8 UTR A	ACCGTTCCTCCCAGAGTTCA	TGGCATTTGATATGAAAATGGTAT	1.5	60
8 UTR B	GATGAGTGGACATGTGGTCAA	GCTCTGTGGCATGCTTAGGT	1.5	60
8 UTR C	ACAGTTTAGAAATTAGCTCCAGGTTC	AAGTCAGATCCCTTCTTAATTCTCT	1.5	60
8 UTR D	AGGTTTTGAATGGGAGAGGT	ATGTTCCCATGTGGTGTCAA	1.5	60
8 UTR E	GGACCACCATTTCTCATTGG	AGGGAACTTCCTGCTTTTTCTT	1.5	57
8 UTR F	GGATTTAAAATTCACTGTGGAAAA	ACTGCAGCAAATGCACTCAC	1.5	57
8 UTR G	CAGAACATGCTGTTTGCTCTT	GCCCCAAGAAATCAATTACG	1.5	55

*' Required the addition of Q solution (Qiagen) for optimal results.

Section 5.2.5 Primer sequences for PCR amplification of coding exons of five crystallin genes.
Optimal Mg^{2+} concentration and annealing temperature (°C) is also given.

Gene	Exon	Forward Primer	Reverse Primer	Mg2+	Temp
CRYAA	1	CTCCAGGTCCCCGTGGTA	AGGAGAGGCCAGCACCAC	1.0	63
	2	CTGTCTCTGCCAACCCCAG	CTGTCCCACCTCTCAGTGCC	1.0	63
	3*	GGCAGCTTCTCTGGCATG	GAGCCAGCCGAGGCAATG	1.5	65
CRYBA1/A3	1*	GGTCTTAGGAAGATCCCAAG	AAGGAGAGGAAGGGCAAGGG	2.5	60
	2	CGTGTGTGCTCTGTCTTCC	GGTCAGTCACTGCCTTATGG	1.5	63
	3	CCTCTGTTTTTCCTGATGT	AACATCCTTCTTCCCCTAT	3.0	60
	4	CAAACACTACATGTCTTTGG	CTTGCTACCCTCATATGC	1.5	58
	5	TGCTTCCTTGTATAATCC	ACTATTGATGCAACCTCAGG	1.5	55
	6	GGTTTGCTACCATTATCTTGG	TCATCCACTTAGGTGTGATTT	1.5	60
CRYBB2	2	TGCTCTCTTTCTTTGAGTAGACCTC	CCCATTTTACAGAAGGGCAAC	2.5	63
	3	ACCCTTCAGCATCCTTTGG	GCAGACAGGAGCAAGGGTAG	1.5	63
	4	GCTTGGAGTGGAACTGACCTG	GGCAGAGAGAGAAAGTAGGATGATG	2.0	65
	5	GCCCCCTCACCCATACTC	CCCCAGAGTCTCAGTTTCCTG	2.5	65
	6	CCTAGTGGCTTATGGATGCTC	TCTTCACTTGGAGGTCTGGAG	1.5	63
CRYGD	1*	CTGCTATATAGCCCGCCGC	AGTAGGGCTGCAGGTTGGG	1.0	63
	2	GATCACCCTCTACGAGGAC	GTGAGTGTCCTGAGGACCT	3.0	60
	3	CTTTTTATTTCTGGGTCCG	CCAGGAACACACAGAAAAT	3.0	60
CRYGC	1+2	TCACACTGAACTCGCATCA	AACTCACTTGGGGGATGAG	2.5	63
	3	CGAGTAGACAGTCTCCCAC	TATTAGGTTCCAAAATGGG	1.5	56

* Required the addition of Q solution (Qiagen) for optimal results.

Section 6.5.2 Primers for amplification and sequencing of Connexin 46 gene.
Optimal Mg^{2+} concentration and annealing temperature (°C) is also given.

Fragment	Forward	Reverse	Size bp)	Mg2+	Temp
1	CGGTGTTCATGAGCATTTTC	GACGTAGGTCCGCAGCAG	496	1.5	60
2*	GCAGGACAATCCCTCGTC	GGTCAGGGCTAGCAGTTTGA	532	1.5	56
3	TCGGGTTCCCACCCTACTAT	TGCACTTTGGTTTTGGTTTC	579	2.0	56
1b	CAGACTTCACCTGCAACAAC	CTCTTCAGCTGCTCCTCCTC	202	1.5	60

* Required the addition of Q solution (Qiagen) for optimal results

Appendix 3 - Allele Frequencies for Microsatellite Markers

Sections 3.3.3, 3.3.4 Allele frequencies observed for markers on chromosome X
Calculated from 22 unrelated unaffected individuals from the south-eastern Australian population.

Marker	Allele	Frequency observed
DXS1224	153	0.08
	155	0.52
	157	0.04
	159	0.04
	161	0.14
	163	0.18
DXS1053	192	0.04
	194	0.38
	196	0.25
	200	0.25
	202	0.04
	204	0.04
DXS1195	230	0.02
	232	0.07
	234	0.50
	236	0.41
DXS418	131	0.04
	133	0.07
	135	0.04
	137	0.16
	139	0.13
	141	0.30
	143	0.07
	145	0.07
	147	0.04
	149	0.04
	153	0.04
TG Z93022	217	0.08
	231	0.15
	233	0.08
	235	0.42
	237	0.19
	239	0.08
TG/AG Z92542	211	0.01
	213	0.06
	215	0.56
	217	0.09
	219	0.24
	221	0.04

Marker	Allele	Frequency observed
DXS999	257	0.04
	263	0.07
	265	0.29
	269	0.04
	271	0.04
	273	0.52
DXS365	199	0.02
	200	0.02
	201	0.02
	202	0.02
	203	0.04
	205	0.20
	207	0.24
	209	0.13
	211	0.07
	213	0.04
	215	0.02
DXS989	169	0.10
	175	0.04
	177	0.13
	179	0.35
	181	0.07
	183	0.04
	185	0.07
	187	0.04
	189	0.04
	191	0.04
	193	0.04
	195	0.04
CA2 Z92542	442	0.04
	444	0.70
	446	0.26
CA Z93929	144	0.01
	153	0.03
	155	0.06
	157	0.34
	159	0.34
	161	0.13
	163	0.06
	165	0.03

Section 5.2.3 Allele frequencies observed at markers representing crystallin genes
Calculated from 22 unrelated unaffected individuals from the south-eastern Australian population

Marker	Allele	Frequency observed	Marker	Allele	Frequency observed
D2S2358	199	0.09	**D21S1890**	147	0.03
	200	0.13		151	0.08
	202	0.07		153	0.20
	204	0.27		155	0.08
	206	0.33		157	0.03
	208	0.02		159	0.08
	210	0.09		161	0.03
D11S1347	175	0.03		163	0.08
	177	0.11		165	0.20
	185	0.03		167	0.08
	189	0.12		169	0.05
	191	0.13		171	0.03
	193	0.24		173	0.03
	195	0.07	**D22S926**	294	0.05
	197	0.06		296	0.05
	199	0.07		298	0.10
	201	0.11		300	0.22
D17S841	258	0.02		302	0.40
	260	0.02		304	0.13
	262	0.13		306	0.05
	264	0.70			
	166	0.09			
	270	0.04			

Section 6.5.1 Allele frequencies at markers representing known cataract loci.
Calculated from founders within the pedigrees and reported by the CEPH database

Marker	Allele	Frequencies Founders	CEPH	Marker	Allele	Frequencies Founders	CEPH
D1S2635	140	0.18	0.00	**D1S468**	172	0.07	0.02
	142	0.09	0.00		177	0.05	0.02
	144	0.02	0.07		179	0.00	0.02
	145	0.14	0.07		181	0.21	0.29
	146	0.00	0.09		183	0.14	0.09
	148	0.24	0.20		185	0.05	0.05
	150	0.20	0.00		187	0.38	0.37
	151	0.02	0.14		189	0.10	0.14
	152	0.07	0.16	**D15S1033**	154	0.18	0.05
	155	0.02	0.11		159	0.05	0.02
	157	0.00	0.11		161	0.29	0.45
	161	0.02	0.02		163	0.23	0.23
D3S1349	114	0.13	0.09		165	0.23	0.21
	118	0.06	0.11		167	0.02	0.02
	124	0.10	0.12	**D15S117**	128	0.16	0.20
	128	0.23	0.22		132	0.31	0.33
	132	0.40	0.32		134	0.30	0.14
	136	0.04	0.09		140	0.09	0.16
	140	0.04	0.04		142	0.14	0.12

D10S1268	131	0.03	0.04
	137	0.03	0.04
	139	0.37	0.46
	141	0.06	0.11
	143	0.14	0.04
	147	0.00	0.02
	149	0.03	0.09
	151	0.16	0.05
	153	0.09	0.12
	155	0.03	0.04
	157	0.06	0.02
D12S83	80	0.18	0.20
	82	0.05	0.10
	84	0.08	0.17
	86	0.11	0.10
	88	0.48	0.30
	90	0.01	0.09
	92	0.01	0.02
	94	0.01	0.02
D13S1236	112	0.04	0.00
	114	0.56	0.00
	124	0.00	0.06
	126	0.04	0.11
	128	0.29	0.43
	130	0.00	0.35
	132	0.07	0.07
D16S496	208	0.17	0.16
	212	0.03	0.02
	214	0.00	0.04
	216	0.17	0.18
	218	0.63	0.42
	220	0.00	0.11
	224	0.00	0.05
D1S243	143	0.00	0.20
	145	0.03	0.00
	147	0.28	0.01
	149	0.28	0.14
	151	0.10	0.20
	154	0.00	0.05
	156	0.03	0.01
	158	0.06	0.09
	160	0.22	0.05
	162	0.00	0.12
	164	0.00	0.05
	166	0.00	0.03
	144	0.00	0.01
D17S796	142	0.00	0.14
	158	0.03	0.01
	160	0.17	0.16
	162	0.33	0.29
	164	0.08	0.01
	166	0.08	0.16
	168	0.28	0.14
	170	0.03	0.03
D17S849	249	0.07	0.02
	251	0.40	0.50
	253	0.14	0.20
	255	0.14	0.16
	256	0.03	0.10
	257	0.12	0.02
	270	0.10	0.00
D17S802	162	0.00	0.01
	168	0.03	0.08
	170	0.03	0.01
	172	0.06	0.05
	174	0.12	0.28
	177	0.38	0.10
	178	0.26	0.25
	180	0.03	0.08
	182	0.03	0.01
	184	0.06	0.07
D17S836	197	0.26	0.32
	199	0.08	0.02
	201	0.50	0.50
	203	0.13	0.11
	205	0.03	0.05
D20S894	259	0.06	0.04
	261	0.09	0.20
	263	0.03	0.01
	265	0.34	0.30
	266	0.03	0.01
	268	0.09	0.09
	270	0.03	0.07
	272	0.15	0.17
	274	0.18	0.09
	276	0.00	0.02