Gene identification and characterisation for blinding inherited eye diseases

Shahrbanou Javadiyan (MSc)

Thesis is submitted for the degree of Doctor of Philosophy

Department of Ophthalmology

School of Medicine

Faculty of Medicine, Nursing and health Sciences

Flinders University of South Australia

July 2015

Table of contents

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Shahrbanou Javadiyan

Thesis Summary

Some devastating childhood diseases are inherited within families. Of potentially blinding inherited eye disorders, paediatric cataract and primary congenital glaucoma (PCG) are 2 major causes of childhood blindness.

Cataract is an opacity of the ocular lens that impairs vision. Paediatric cataract is a genetically and clinically heterogeneous condition with an incidence of 2.2 per 10,000 live births in Australia, with 8-25% of cases being hereditary. PCG is the most common type of childhood glaucoma. The condition is caused by a developmental defect of the trabecular meshwork and anterior chamber angle resulting in disruption in drainage of the aqueous humor and an increased intraocular pressure (IOP) resulting in damage to the optic nerve and subsequent visual loss. In Australia the incidence of PCG is 1 per 30000.

The overall aim of this project was to use Massively Parallel sequencing (MPS) technologies to mine our paediatric cataract and PCG DNA repository for genetic mutations in known candidate genes, and identify new paediatric cataract and PCG causing genes. Furthermore, this project aimed at providing a molecular genetic diagnosis to the affected family members in our repository.

Phase 1 aimed to screen our world leading repository of DNA samples from patients with familial or sporadic paediatric cataracts or PCG for mutations in candidate genes. We screened a novel PCG candidate gene, *TEK* (a recently identified gene), in our cohort of 53 Australian PCG cases, aiming to evaluate the association between variations in *TEK* and the disease in Australia. Five heterozygous protein changing variants in *TEK* were detected and this gene showed significant enrichment for mutations in PCG patients. In conjunction with functional data generated by our

collaborators, these data indicate that mutations in the *TEK* gene are likely associated with PCG.

A total of 98 samples with paediatric cataract (65 from Australia and 33 from Asia) were screened for mutations in 51 previously reported paediatric cataract genes. The study was able to detect the genetic cause behind 42% of Australian patients with familial paediatric cataracts, 40% of sporadic Australian cases and in 23% of the Asian cohort.

In phase 2 we undertook gene discovery in families (paediatric cataract or PCG) with no identifiable causes in known genes using next generation whole exome sequencing. Two novel candidate genes for paediatric cataract (*HTR1F* and *NOL9*) and one for PCG (*GREB1*) were identified. In addition, for the first time, we identified a copy number variant (CNV) of a crystallin gene in an Australian family with paediatric cataract, also identified through whole exome sequencing.

This study demonstrated the feasibility of using next generation sequencing technologies to screen genes panels in a heterogeneous condition like paediatric cataract and the potential for this technology in novel gene discovery. Furthermore, for the first time it showed the possibility of the involvement of copy number variation in isolated paediatric cataract pathogenesis. In addition a higher mutation rate in the *TEK* gene was detected in PCG cohort than in the general population which supports its association with PCG development. Further investigation is required to determine the role of the novel candidate genes identified in this study in PCG and paediatric cataract pathogenicity.

Acknowledgment

This thesis is dedicated to the precious memories of my late parents, my beloved mother leyli and my dearest father, Sohrab.

I would like to pay principal acknowledgement to my principal supervisor Assoc. Prof. Kathryn P. Burdon for her professional guidance and patience. It was such a great and inspirational experience to work under her supervision. To Kathryn, you have supported me all the time during my research in every single step trying to make this thrilling experience as stress-less as possible. Thank you for your great attitude, encouragements and care. It was not possible doing this research without your guidance. No words can describe how thankful I am.

My sincere thanks also go to my co-supervisor Dr Shiwani Sharma for her valuable guidance helping me widen my research from another perspective. Thanks for her helpful comments on my writing, the manuscripts for publications and this thesis, and for her patience for explaining the methods so thoroughly. I truly appreciate all of these.

I would like to thank my co-supervisor Prof. Jamie E. Craig. To Jamie, thanks for your support and care during my PhD, whenever I needed that and for making projects like this possible with your great passion for research.

Thanks also go to my co-supervisor Dr Karen M. Lower for her valuable comments on the writings of the thesis, her positive attitude and guidance.

My sincere gratitude goes to everyone in the Department of Ophthalmology, School of Medicine, Flinders University, beginning with Emmanuelle Souzeau for her effort in assisting with the recruitment of patients and data collection. Particular note needs to be made of the assistance of Dr Owen Siggs, with his great help in bioinformatics. I would

also like to thank Sara Martin, Sionne Lucas, Lefta Leonardos, Win Kee Beh, Kathleen Dowell for their valuable assistance in different parts of this research project.

My thanks go to Ms Nousha Chegini (Proteomic Facility, Flinders University) and Dr Renee Smith (Genomic Facility, Flinders University) for their assistance in proteomics and genomics laboratories.

The plural personal pronoun “we” was used instead of the singular form as the thesis is the result of a teamwork.

I would like to express my appreciation to my lovely family, Noushin, Amirkhosro, Khashayar and Golnoush for their endless emotional support and unfailing love during my life and this PhD. Finally, my heartfelt gratitude goes to my partner Behrang, for his unconditional love, understanding and encouragements.

Publications and conference abstracts arising from this thesis

Javadiyan S, Craig JE, Souzeau E, Sharma S, Lower KM, Pater J, Casey T, Hodson T, Burdon KP. 2016. **"Recurrent mutation in the crystallin alpha A gene associated with inherited paediatric cataract"**. *BMC Research Notes* 9:83.

Souma T, Tompson SW, Thomson BR, Siggs OM, Kizhatil K, Yamaguchi S, Feng L, Limviphuvadh V, Whisenhunt KN, Maurer-Stroh S, Yanovitch TL, Kalaydjieva L, Azmanov DN, Finzi S, Mauri L, **Javadiyan S**, Souzeau E, Zhou T, Hewitt AW, Kloss B, Burdon KP, Mackey DA, Allen KF, Ruddle JB, Lim SH, Rozen S, Tran-Viet KN, Liu X, John S, Wiggs JL, Pasutto F, Craig JE, Jin J, Quaggin SE, Young TL. 2016. **"Angiopoietin receptor TEK mutations underlie primary congenital glaucoma with variable expressivity"**. *Journal of Clinical Investigation.*

Siggs OM *, **Javadiyan S***, Sharma S, Souzeau E, Lower KM, Taranath DA, Black J, Pater J, G J, Willoughby and others. 2016. "Partial duplication of the *CRYBB1-CRYBA4* locus is associated with autosomal dominant congenital cataract". 2016. *European Journal of Human Genetics* (submitted).

Javadiyan S, Craig JE, Sharma S, Lower KM, Casey T, Haan E, Souzeau E, Burdon KP. 2016. **"Novel missense mutation in the bZIP transcription factor, MAF, associated with congenital cataract, developmental delay, seizures and hearing loss (Aymé-Gripp syndrome)"**. 2016. *BMC medical genetics* (submitted).

Javadiyan S, Craig JE, Souzeau E, Sharma S, Lower KM, Mackey DA, Staffieri SE, Elder JE, Pater J, Taranath D, Casey T, Hewitt AW, Burdon KP. 2016. **"High throughput genetic screening of 51 paediatric cataract genes identifies causative**

mutations in inherited paediatric cataract in South Eastern Australia.2016". *Human mutation* (Submitted).

*Contributed equally to the publication

Conference Abstract and talks

***Shari Javadiyan,** Shiwani Sharma, Karen Lower, Jamie E Craig and Kathryn P Burdon, "**Screening of 51 candidate genes in congenital cataract patients using a semiconductor based next generation sequencing method**", *Australian Society for Medical Research (ASMR),* Adelaide, South Australia, 5 thJune 2013.

***Shari Javadiyan**, Shiwani Sharma, Karen Lower, Jamie E Craig and Kathryn P Burdon, "**High throughput screening of 51 genes in families with congenital cataracts"**, *GeneMappers,* Barossa Valley, South Australia, 11-14 May 2014

***Shari Javadiyan**, Jamie Craig, Shiwani Sharma, Karen Lower, Kathryn Burdon, "High throughput screening of 51 known causative genes in families with congenital cataract", *American Society of Human Genetics (ASHG),* San Diego, California, United States, 18-22 October 2014

Owen M. Siggs, **Shari Javadiyan**, Shiwani Sharma, Kathryn P Burdon, Jamie E Craig "**Partial duplication of *CRYBB1* as a novel genetic mechanism for autosomal dominant congenital cataract"**, *The Association for Research in Vision and Ophthalmology (ARVO),* Denver, Colorado, United States, May 3-7 2015.

****Shari Javadiyan**, **"High throughput screening of 51 known causative genes in families with congenital cataract",** *Genetic Solutions World Tour,* Australia, Adelaide, 19-23 May 2014.

* Oral presentation
** Invited speaker

List of Abbreviations

α	Alpha
A	Ampers
ACG	Angle Closure Glaucoma
AD	Autosomal Dominant
AGRF	Australian Genomics Research Facility
AR	Autosomal Recessive
β	Beta
bp	Base Pair
CNV	Copy Number Variation
CSA	Cataract South Australia
DNA	Deoxy ribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DsDNA	Double-Strand DNA
ExAC	Exome Aggregation Consortium
γ	Gamma
hg19	Human genome build 19
IGV	Integrative Genomic Viewer
IOL	Intraocular Lens

IOP	Intraocular Pressure
kDA	Kilo daltons
LOD	logarithm of the Odds
MAF	Minor Allele Frequency
µg	Microgram
mg	Miligram
miRNA	Micro-RNA
ml	Millilitre
mM	Millimolar
µl	Microlitter
MPS	Massively Parallel Sequencing
nM	Nanomolar
PAGE	Polyacrylamide Gel electrophoresis
PBS	Phosphate Buffered Saline
PCG	Primary Congenital Glaucoma
PCR	Polymerase Chain Reaction
PGM	Personal Genome Machine
pM	Picomolar
POAG	Primary Congenital Glaucoma

PolyPhen	Polymorphism Phenotyping
RGC	Retinal Ganglion Cell
RNA	Ribonucleic Acid
RVEEH	Royal Victorian Eye and Ear Hospital
SAP	Shrimp Alkaline Phosphatase
SC	Schlemm's Canal
SDS	Sodium Dodcecyl Sulphate
SIFT	Sorting Intolerant from Tolerant
SNP	Single Nucleotide Polymorphism
TBE	Tris Borate-EDTA
TBST	Tris Buffered and Tween 20
TM	Trabecular Meshwork
U	Unit
UTR	Untranslated reagions
V	Voltage
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
XL	X Linked

Chapter1: Introduction

1.1 Overview

The focus of the work described in this thesis is the investigation of the genetics of two forms of Mendelian inherited eye diseases causing blindness in children: paediatric cataract and primary congenital glaucoma (PCG).

Cataract is defined as the opacity of the crystalline lens that prevents or disrupts the light entering the eye and causes the formation of a blurry image in the brain at the back of the eye (Francis, et al., 1999). PCG is related mainly to developmental defects of the trabecular meshwork (TM), which is a porous tissue located in the anterior segment of the eye (Yu Chan, et al., 2015). The introduction to this thesis will discuss the normal anatomy of the eye, with a focus on lens function and TM due to their involvement in paediatric cataract and PCG pathogenesis.

It is known that both paediatric cataract and PCG can be genetic. Some of the genes associated with these conditions have been identified, however with many affected families yet to have a causative mutation identified, it is clear that there are more genes yet to be identified. Traditional mapping methods like linkage analysis have found many genes, however this thesis aimed to find novel genes and mutations for PCG and paediatric cataract using massively parallel sequencing techniques.

The treatment of blinding congenital eye diseases is a significant economic cost on health care systems worldwide. Delay in diagnosis and treatment leads to life-long blindness, while early diagnosis and proper follow up treatment could be very beneficial for both patients and reduction of treatment cost. Understanding the genetic causes of congenital blinding eye diseases will aid developing genetic screening programs, which can lead to early diagnosis, decreasing the possibility of complete visual loss in patients.

1.2 Anatomy of the human eye

The human eye is a complex and delicate organ with many components (Figure.1.1). Three different layers can be detected in the human eye. The outer layer consists of the cornea and sclera. The cornea is located at the front part of the eye. It is transparent and allows the passage of light into the eye. It also has a protective role against infection and structural damage to the eye (Willoughby, et al., 2010). The sclera is the white outer layer that encompasses the eye. The middle layer of the eye, the uvea, consists of the choroid, ciliary body and iris. The choroid is responsible for the blood supply to the eye which provides nutritive and supportive function for the iris (Chalam, et al., 2009-2010). The iris is the coloured part of the eye. The pupil is the space through which light passes into the eye, and the size of the pupil is controlled by the contraction or relaxation of the iris in response to light intensity. The ciliary body is a thickened tissue which helps the lens to adjust its shape and also produces aqueous humor. The lens is a double convex transparent structure which, along with the cornea, is responsible for focusing the light to the retina. The cornea has four times more focusing power than the lens; however unlike the lens, it doesn't have the ability to change the shape to alter the focus point (Sebastian, 2010). The retina is the inner-most layer of the eye. It is photo sensitive and contains photoreceptors and neural elements which are responsible for the detection of light and processing of the visual information respectively (Chalam, et al., 2009-2010). The fovea is the centre of the field of vision. The electrical stimulus generated by light hitting the retina travels to the brain for vision processing through the optic nerve. Beside these layers, there are also two fluid chambers named the anterior chamber (the space between the lens and cornea) and the posterior/vitreous chamber (the space between the lens and retina). The anterior chamber contains aqueous humor. The aqueous humour provides nutrition for the central cornea and lens, which is required as they are avascular tissues (http://www.eyepedia.co.uk/). The liquid which

fills the posterior chamber is called vitreous humor and is also responsible for nourishment of the avascular lens (Sebastian, 2010). Moreover, it is important for maintaining the structural stability of the eye (http://www.eyepedia.co.uk/). The trabecular meshwork (TM), which is a tissue of an irregular network located in the angle between the cornea and iris, is responsible for draining the aqueous humor from the anterior chamber and is thus involve in maintaining the intraocular pressure (IOP) at its optimum level (Shentu, et al., 2013).

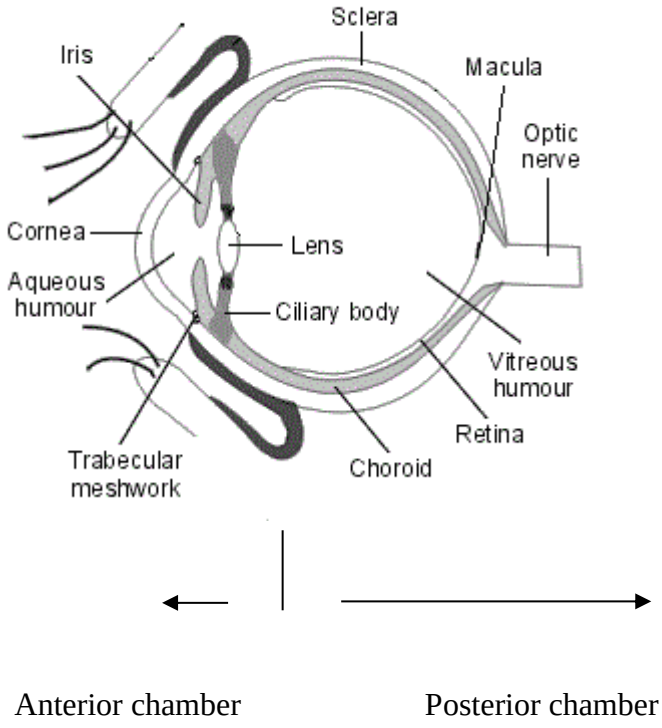


Figure 1.1. The diagram of human eye showing the location of its main components (adapted from http://www.patient.co.uk/).

Although all the components of the eye are important for its function, the following sections will focus on development and anatomy of the lens and trabecular meshwork as these are the tissues of most relevance to paediatric cataract and primary congenital glaucoma, respectively.

1.2.1 Lens

Lens growth and development

The main stages of human lens development are induction, morphogenesis, differentiation and growth (McAvoy, et al., 1999). Human lens induction and morphogenesis begin with the thickening of the surface ectoderm (around 4 weeks) to form the lens placode (Augusteyn, 2010; Francis, et al., 1999) (Figure. 1.2 a, b). The placode invaginates to form the lens pit which then closes over to form the lens vesicle (Figure 1.2 c, d). The lens vesicle formation is complete by day 56 of gestation. The vesicle is then filled with primary lens fibre cells (Augusteyn, 2010) which are differentiated from the single layer of epithelial cells on the posterior side of the lens vesicle. Epithelial cells just anterior to the equator then begin to undergo mitosis and migrate posterior to the equator where they start differentiation and form secondary fibre cells (Augusteyn, 2010) (Figure 1.2 e). During lens development and differentiation, fibre cells lose their nuclei, mitochondria and most of other organelles (for transparency) and become metabolically inactive (Wride, 2011).

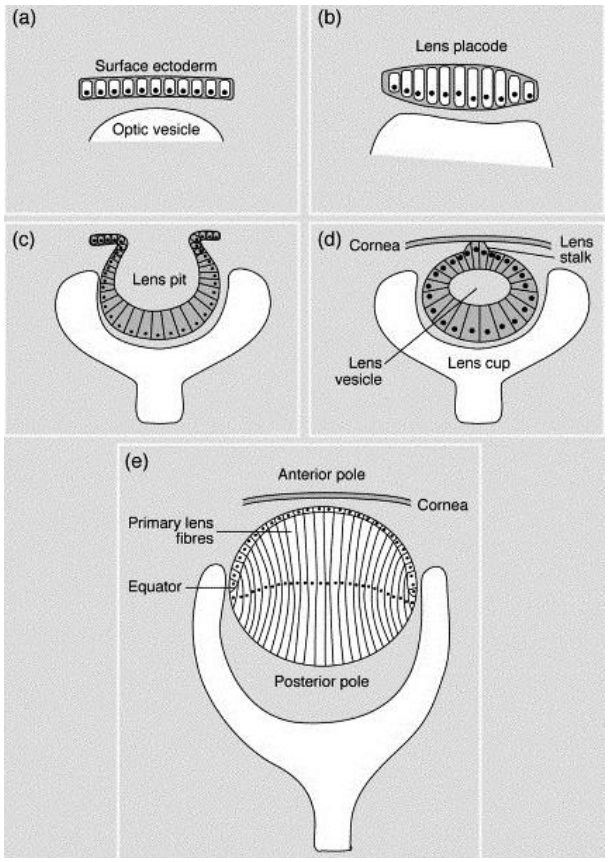


Figure 1.2. Diagram showing lens development (Francis, et al., 1999).

(a,b) formation of lens placode, (c) lens pit formation , (d) lens vesicle formation, (e) differentiation and formation of secondary fibre cells.

Anatomy and physiology of lens

The lens' main function is the focusing of light onto the retina (Rhode and Ginsberg, 1987). To perform this function, transparency of the lens is very important. This relies on the protein matrix of the lens-fibre cells (Chalam, et al., 2009-2010) which forms the bulk of the lens. The proteins involved in maintaining the structure of the lens are highly stable. The lens' ability to refract light is due to its different index of refraction compared with the surrounding aqueous and vitreous (Skuta, et al., 2009-2010).

The components of the mature lens are the capsule, lens epithelium, cortex, and nucleus (Figure 1.3). The lens capsule, a modified basement membrane, has a porous,

collagenous structure and surrounds the lens (Danysh and Duncan, 2009). Beneath the capsule, a layer of epithelial cells covers the anterior surface of the lens. (Donaldson, et al., 2010). The primary fibre cells form the centre of the lens, known as the embryonic nucleus. The adult nucleus forms during early childhood. The soft and newly generated fibres lying between the nucleus and epithelium form the cortex and are generated throughout life (Rhode and Ginsberg, 1987) .The ordered structure of the fibre cells and their lack of organelles leads to transparency of the lens, which is critical for clear vision (Harding, 1991; Wride, 2011).

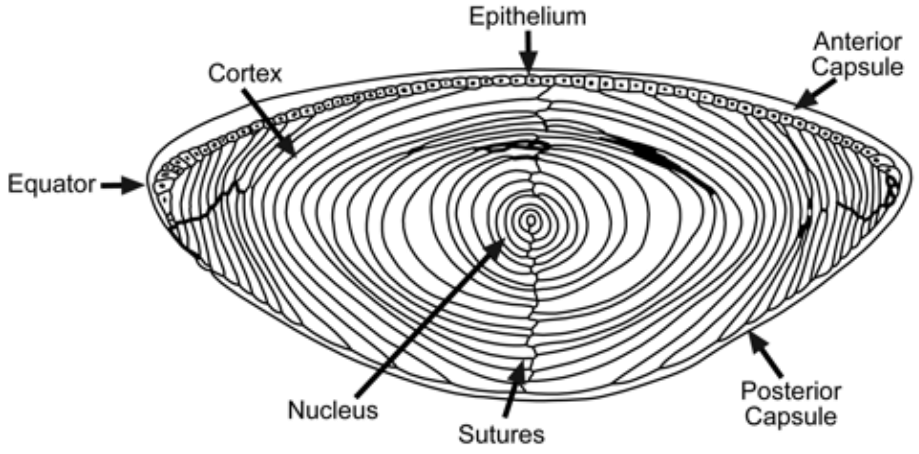


Figure 1.3. A diagrammatic structure of human lens (Harding, 1991)

The mechanism that controls water and electrolyte balance is an important aspect of lens physiology and is very important for lens transparency. Approximately 66% of a normal human lens is water, with protein accounting for only 33% of its mass.

1.2.2 Trabecular meshwork and circulation of aqueous humor

The trabecular meshwork is a porous filter–like structure which consists of the inner uveal meshwork, the deeper corneosceral meshwork and juxtacanalicular connective tissue (JCT) (Pébay, 2014).The trabecular meshwork is responsible for the drainage of

the majority of aqueous humor from the anterior chamber into Schlemm's canal (SC) (Figure 1.4). A smaller amount of aqueous humor leaves the eye through the ciliary body, just below the trabecular meshwork (uveoscleral route) (Kwon, et al., 2009). The circulation of aqueous humor in the anterior segment of the eye is very important for maintaining the optimum intraocular pressure (IOP). Elevated IOP leads to glaucoma through damage to the optic nerve and consequently leads to retinal ganglion cell (RGC) apoptosis (Cascella, et al., 2015).

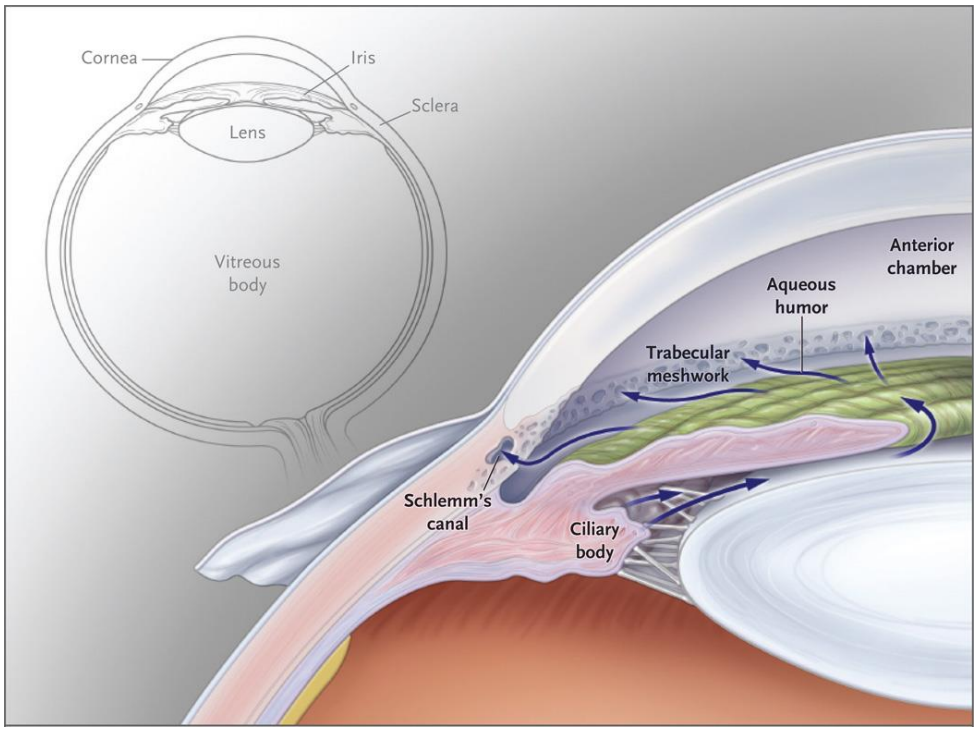


Figure 1.4: Circulation of aqueous humor and its drainage from the anterior chamber through trabecular meshwork. Arrows represent the flow of aqueous humor from the ciliary body and passage into the anterior chamber. The larger portion of aqueous humor drainage is through the trabecular meshwork into the Schlemm's canal while a smaller portion of aqueous humor passes the eye through the ciliary body below the trabecular meshwork. Source: (Kwon, et al., 2009).

1.3 Cataract

Cataract is defined as a partial or complete opacity of the lens. Several mechanisms of cataract formation are known. Lens micro structure breakdown which causes vacuole formation and disruption of the organised packaging of fibre cells is often seen in developmental cataract. Protein precipitation can also lead to opacity and light scattering (Graw, 2009b; Hejtmancik, 2008). It has been estimated that cataract is responsible for half of the blindness worldwide (Graw, 2009b) and it is predicted there will be 40 million people blind from cataract by 2020 (Brian and Taylor, 2001). Cataract can be age-related or paediatric (congenital). Many factors are thought to contribute to the onset of age-related cataractogenesis, including reduced chaperone function of lens proteins, higher lens membrane permeability (compared to a normal lens), low antioxidant defence capacity of the lens and oxidative stress which has a direct effect on the solubility of lens proteins (Kaur, et al., 2012). Genetic factors, malnutrition, the high ultraviolet component of sunlight, diabetes, medications and smoking are all considered to be risk factors for cataract and the main causes of high prevalence of cataract worldwide. (Harding, 1991; Wang, et al., 2010a).

1.3.1 Paediatric cataract

Paediatric cataract is often referred to as congenital or infantile cataract which presents at birth or in first year of life, or juvenile cataract which presents during childhood. For simplicity, the term paediatric cataract will be used throughout this thesis unless the symptoms indicate the congenital form. Several causes of childhood cataracts have been reported including infection, drug exposure, metabolic disorders, chromosomal mutations, malnutrition, dehydrating diseases and heredity (Churchill and Graw, 2011). Paediatric cataract is the leading cause of blindness in children and 200,000 children worldwide are blind from the disease (Chan, et al., 2012). It has been estimated that the

frequency is 1-6 per 10,000 live births in industrialized countries (Santana and Waiswo, 2011) with approximately 25% to 33% of these cases being hereditary. According to Wirth *et al*. (2002) the incidence is 2.2 per 10,000 live births in Australia (over a 25-year study with 421 cases of paediatric cataract). Hereditary paediatric cataract can be isolated (occurring as the only feature) or syndromic (cataracts are associated with other systemic or ocular features) (Hejtmancik, 2008).

Treatment of paediatric cataract is the same as for the age-related form, which is surgery, however there is a high rate of complications. Complications include early and late development of glaucoma (which occurs mostly in children under the age of 12 months (Chan, et al., 2012) and posterior capsule opacification. Intraocular lens (IOL) implants need revision as the eye grows. If not treated, dense opacities lead to amblyopia (lazy eye). Sometimes further surgery is required for those children who undergo lens implantation to solve the posterior capsule opacification (Chan, et al., 2012). A better understanding of the potential cause of cataract may be beneficial for the development of novel patient-specific treatments.

Paediatric Cataract Phenotypes

Several classification systems have been developed to categorise congenital and paediatric cataracts, generally based on anatomical location and morphology of the opacity, or the name of the family in which the phenotype was first observed (Huang and He, 2010). Some examples of different types of cataracts are shown in Figure 1.5.

Total cataract is usually bilateral (affecting both eyes). In lamellar (zonular), nuclear, oil droplet, cortical, or coronary cataract, distinct parts of the lens are affected (Amaya, et al., 2003). Nuclear cataract is central and causes more visual problems because of the location of the opacity. Lamellar cataract occurs in specific layer or layers of the cortex. Cortical cataract affects the outer layers of the cortex.

Cataracts can affect anterior sections of the lens (anterior polar, anterior subcapsular, or anterior lenticonus) or the posterior pole (Mittendorf's dot, posterior lenticonus, posterior cortical cataracts, or posterior subcapsular) (Amaya, et al., 2003).

Pulverulent cataracts have a dust-like appearance and can be restricted to a part of lens or affect the whole lens. Cerulean cataracts consist of blue and white dots in the lens.

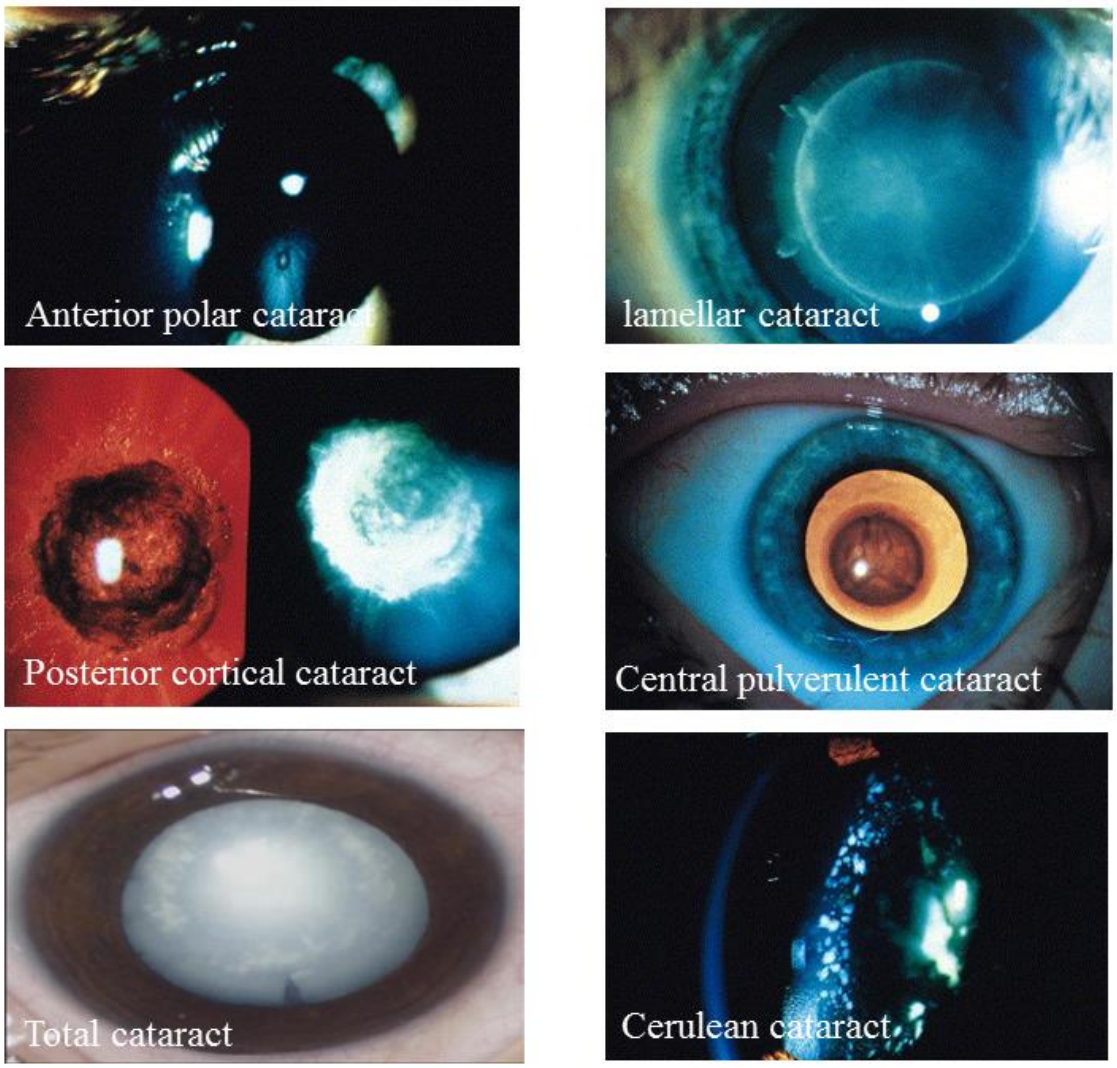


Figure 1.5. Examples of different types of paediatric cataracts. Taken from (Amaya, et al., 2003). The total cataract image is taken from (http://www.ophthalmologymanagement.com/).

1.3.2 Genetics of paediatric cataract

Inherited forms of paediatric cataract account for 8-25% of all cases. It has been reported that 28 % of bilateral paediatric cataract cases have a genetic basis while only 2% of unilateral cases are genetic (Rahi and Dezateux, 2001; Santana and Waiswo, 2011). Cataract can be transmitted as autosomal recessive, autosomal dominant or X-linked diseases, although autosomal dominant is the most common mode of inheritance.

Mutations in genes that encode structural proteins, membrane or cytoskeletal proteins, transcription factors and signalling molecules are associated with hereditary forms of paediatric cataracts (Churchill and Graw, 2011; Santana and Waiswo, 2011; Shiels A, 2007). Genetic heterogeneity (mutations in different genes resulting in a similar phenotype) and clinical heterogeneity (mutations in the same gene resulting in different phenotypes) have been observed in paediatric cataract, adding to the complexity of this disease (Lorenz, 2007).

Structural protein encoding genes and paediatric cataract

Crystallins are the main structural proteins in the lens. They were discovered more than 100 years ago by Morner and he named them crystallins because of their presence in the crystalline lens (Churchill and Graw, 2011; Graw, 2009a). Crystallins are divided to α-, β- and γ- crystallin according to their separation by gel chromatography. They have been also characterized according to their genetic characteristics, expression pattern and their contribution to cataract (Graw, 2009a; Santana and Waiswo, 2011).

α-crystallins are composed of αA and αB-crystallin which are encoded by *CRYAA* and *CRYAB* genes respectively. Evolutionary speaking, they belong to a family of heat shock proteins (which usually up-regulate in response to stress) which share a domain of 90 amino acids (Graw, 2009a). αB-crystallin is a stress-inducible protein which has been detected in lens epithelial cells (Andley, 2007). It has been demonstrated that αA-

crystallin has increased expression in epithelial cells at the elongation zone and also plays a role in the differentiation of lens fibre cells (Boyle and Takemoto, 2000).

β- and γ-crystallins have a native molecular mass of ~200 kDa and they form an antiparallel β sheet structure with a common feature named the "Greek key Motif". The functionality of this motif is to prevent light scattering by dense packing of proteins. β-crystallins are subdivided into acidic and basic subunits, encoded by *CRYBA* and *CRYBB* genes (Wistow, 2012) *CRYBA1, CRYBA2, CRYBA4, CRYBB1, CRYBB2* and *CRYBB3* belong to the family of β–crystallins (Churchill and Graw, 2011; Vanita, et al., 2001a). γ –crystallins (*CRYGA, CRYGB, CRYGC, CRYGD* and *CRYGS*) are another subgroup of crystallin proteins which are highly expressed in the human lens (Wistow, 2012). Mutations in these proteins decrease protein solubility, leading to a decrease in the transparency of the lens. Crystallin genes in which mutations are known to cause cataract (*CRYAA, CRYAB, CRYBA1, CRYBA4, CRYBB1, CRYBB2, CRYBB3, CRYGC, CRYGD, CRYGS*) with emphasis on paediatric cataract, are listed in Appendix 1. Mutations in these genes cause autosomal dominant and recessive paediatric cataract.

Membrane or cytoskeleton proteins encoding genes and paediatric cataract

Cytoskeleton proteins

The cytoskeleton is a network of cytoplasmic proteins which provides structural support and is important in the maintenance of cell volume and shape and protein homeostasis (Song, et al., 2009). The known cataract-causing genes that are involved in the cytoskeleton maintenance are *BSFP1, BFSP2, VIM, CHMP4B* and *FYCO1*. Appendix 2 shows mutations reported in these genes and their associations with cataract phenotypes. Lens fibre cells contain a characteristic beaded filament cytoskeletal structure. The beads are encoded by *BSFP1, BFSP2* and *VIM* (Song, et al., 2009) and are highly expressed in the lens. Mutations in *VIM* result in the formation of dominant,

pulverulent cataract (Muller, et al., 2009). Another filament coding gene implicated in cataract is chromatin modifying protein-4B (*CHMP4B*) with 2 mutations reported to date (Appendix 2). Twelve different mutations have been reported in families with paediatric cataract in Coiled-coil domain containing 1 (*FYCO1*) gene (Appendix 2) which encodes protein expressed in human lens epithelial cells (Chen, et al., 2011).

Connexin proteins

GJA1, *GJA3* and *GJA8* encode gap junction proteins, also are known as connexins (CX43, CX46 and CX50 respectively). Their impaired function has been shown to be involved in paediatric cataract formation. Forming inter-cellular channels, their function is to mediate the exchange of ions and small molecules (such as second messengers and metabolites) between the cytoplasm of adjacent cells (Pfenniger, et al., 2011) and they are involved in the transport of metabolites into and out of the avascular lens (Pfenniger, et al., 2011). It has been reported that the lens epithelial cells express *GJA1* and *GJA8* while specialized lens fibres express *GJA3* and *GJA8* (Beyer and Berthoud, 2014). *In vitro* expression studies of several mutants of *GJA3* and *GJA8* studied the mechanism behind the disease formation. These mechanisms include: reduction or modification in intercellular communication due to channel alterations, impaired cellular trafficking, hemichannel function (which leads to cell injury and death) and formation of cytoplasmic accumulations (light scattering particles) (Beyer, et al., 2013). Mutations in these genes related to paediatric cataract are listed in Appendix 2.

Membrane proteins

Mutations in several other membrane proteins have also been linked with paediatric cataract. Major intrinsic protein (MIP) (also known as Aquaporin-0 (AQP0)) is a member of the aquaporin superfamily. Its function is the regulation of water flow between cells. It also has a role in intracellular communication in ocular lens and is

important in proper lens function (Yi, et al., 2011). Another major intrinsic protein, Lens Intrinsic Membrane Protein 2 (LIM2) encodes an eye specific protein expressed in fibre cells. It acts as a receptor for calmodulin and is important in lens development (Yi, et al., 2011). The gene encoding trans-membrane protein 114 (TMEM114) is a sequence homolog of *LIM2* with unknown function. A segregating chromosomal translocation of *TMEM114* in a family with autosomal dominant cataract has been reported. Tmem114 is expressed in the lens epithelial cells at the fibre differentiation zone (Jamieson, et al., 2007).

Another gene which falls into this category is *PVRL3* (poliovirus receptor-like 3) belongs to the Nectin family, which are immunoglobulin-like adhesion molecules with roles in cytoskeleton structure. It has been identified that *PVRL3* is a critical gene involved in a Nectin-mediated cell-cell adhesion mechanism in human ocular development. The expression of *PVRL3* encoding the cell adhesion protein Nectin 3, is significantly reduced in patients with a balanced translocation, resulting in severe bilateral paediatric cataract and mild developmental delay (Lachke, et al., 2012) .

Transcription factor and signalling molecules encoding genes associated with paediatric cataract

A range of transcription factors have been associated with paediatric cataract (Appendix 3). *HSF4* is essential in lens development and differentiation (Bu, et al., 2002) and mutations in the gene cause various forms of isolated paediatric cataract. In addition to *HSF4*, mutations in a number of additional growth factors including *FOXE3, MAF,* and *PITX3* are also associated with isolated paediatric cataract. These genes are all involved in the development and organisation of the eye, primarily the anterior segment and lens. Mutations in these genes also cause syndromes where paediatric cataract are a feature. The FOX family of transcription factors play roles in regulating the expression of genes

involved in cell growth and differentiation (Tuteja and Kaestner, 2007). *FOXE3* in particular is involved in regulating the genes which are associated with epithelial cell proliferation and eye development. *MAF* is involved in the regulation of embryonic lens fibre cell development (Churchill and Graw, 2011; Yi, et al., 2011). *PITX3*, which is a member of *PITX* family, encodes a paired-like class of homeobox transcription factor. It has been demonstrated that *PITX2* (another member of this family) and *PITX3* are involved in eye development and their expression has been detected in cornea, lens, and retina (Berry, et al., 2011).

EPHA2 belongs to the subfamily of receptor tyrosine kinases and is enriched in epithelial cells. The extracelluar domains of Eph receptors interact with membrane-bound ligands known as ephrins and this signalling system is involved in multiple developmental processes such as forming of nervous, skeletal and vascular systems (Shentu, et al., 2013). The expression of *EPHA2* gene in human lens epithelial cell line and human anterior lens capsule tissues has been demonstrated (Zhang, et al., 2009) but its specific role in human lens development is unknown. Ephrin-A5 (EFNA5) is the main ligand in the lens for EPHA2 and although mutations in this gene cause cataract in mouse (Cooper, et al., 2008), there have not been any reports to date of mutations in human paediatric cataract patients; however three age-related cataract risk alleles have been reported in this gene (Lin, et al., 2014).

Tudor domain RNA binding protein (*TDRD7*) is expressed in lens fibre cells and has been shown to be related to paediatric cataract. It is required for post-transcriptional control of mRNAs which are necessary for normal lens development (Lachke, et al., 2011).

MicroRNA encoding gene and paediatric cataract

MicroRNAs are small non-coding molecule of about 22 nucleotide-long which are involve in post-transcriptional regulation of gene expression (Ambros, 2004). *MIR184* has reported to be a potential cause in rare cases of keratoconous (abnormal cone-shaped protrusion of the cornea of the eye) and cataract co-segregating in autosomal dominant fashion (Hughes, et al., 2011). *MIR184* is expressed in the cornea and lens epithelium.

Syndromic and enzymatic paediatric cataract

Syndromic cataract

Cataract is a feature of many genetic syndromes and metabolic disorders. Nance-Horan syndrome (NHS), hereditary hyperferritinemia cataract syndrome (HHCS), aniridia and Brachio-oto-renal syndrome-1 are examples of these syndromes. NHS is an X-linked disorder which is partially characterized by paediatric cataract along with other clinical features. The protein changing mutation was first identified in 5 families with NHS in a gene which was thus named *NHS* (Burdon, et al., 2003). NHS protein has a potential role as a key scaffold protein involved in regulation of cell adhesion by participating in multiple regulatory pathways. It has been shown that one of the isomers (NHS-A) is associated with the cell membrane and involved in cell-cell contact formation (Sharma, et al., 2009). Mutations in this gene related to paediatric cataract are listed in Appendix 4.The hereditary hyperferritinemia cataract syndrome (HHCS) is an autosomal dominant disorder characterized by high serum ferritin and early onset cataract which is caused by the intracellular accumulation of ferritin (the major intracellular iron storage protein in all organisms) in the lens. (Nonnenmacher, et al., 2011).

Paired box gene 6 (*PAX6*) is expressed in developing eye and is a highly evolutionarily conserved transcription factor important in ocular and neural development (Solomon, et

al., 2009). In a family with 3 distinct ocular phenotypes (classic aniridia, paediatric cataract and late-onset corneal dystrophy) two mutations (p.R103* and p.S353X) in the *PAX6* gene have been identified (Glaser, et al., 1994). The human homologue of the *Drosophila eyes absent* gene (*EYA1*) (Azuma, et al., 2000) is related to paediatric cataract and ocular anterior segment anomalies. The EYA protein family members have protein phosphatase function with enzymatic activity which is required for regulating genes encoding growth control and signalling molecules. This gene is involved in autosomal dominant forms of syndromic cataract (Brachio-oto-renal syndrome-1).

Mutation in genes encoding enzymes and paediatric cataract

Mutation(s) in some enzyme coding genes including *GALK1, SORD, AGK* and *GCNT2* are associated with paediatric cataract. GALK1 is involved in the first step of metabolism of galactose which is the conversion of galactose to galactose-1-phosphate. In the absence of GALK1, the accumulating galactose is converted to galactitol which leads to osmotic swelling and results in cataracts (Yasmeen, et al., 2010). It has been shown that pathogenic mutations in *GALK1* (Appendix 4) are responsible for autosomal recessive paediatric cataract in 2 families (Yasmeen, et al., 2010). The enzyme that converts sorbitol to fructose and sorbitol is encode by SORD which its deficiency has been reported in a family with paediatric cataract (Shin, et al., 1984). *AGK* (acylglycerol kinase) which is also known as *MULK* (multisubstrate lipid kinase), is a kinase capable of phosphorylating monoacylglycerol and diacylglycerol as well as ceramide. A mutation in this gene is reported to be involved in development of isolated paediatric cataract (Aldahmesh, et al., 2012a). Glucosaminyl (N-acetyl) transferase-2 (GCNT2) is a blood-group glycosylation enzyme. Mutations in this gene cause autosomal recessive cataract (Pras, et al., 2004) and associated with the adult i blood-group phenotype which

is linked to increased risk of early onset cataract, particularly in the Japanese population (Appendix 4).

Loci linked with the paediatric cataract

Multiple mapped loci have been detected in association with cataracts in which the causing gene has not been identified yet (Appendix 5). More than 60 genetic loci have been reported to be associated with paediatric cataract, of which over 50 have been linked to a specific gene as discussed above in detail. Therefore more than 10 genetic loci remain which have been associated with paediatric cataract without mutations having been identified in specific genes (Appendix 5). The majority of these genes and loci have been mapped using traditional approaches mainly linkage analysis. Some of these loci have been replicated in different studies indicating in increases likelihood of accuracy. For example, the *CTAA2* locus was initially mapped by Berry *et.al* (1996) in an English family with autosomal dominant anterior polar cataract. The finding was replicated in another study by Ionides, *et al* (1998).The *CPP3* locus located on 1p 34-p36 was reported in two different studies by Yamada, *et al* (2000) and Li *et al.* (2006) in Japanese and Chinese families respectively. Despite being identified by multiple studies, these traditional mapping approaches failed to find the paediatric cataract associated genes within these loci. This was due to detection of multiple candidate genes in the mapped regions and the high cost of Sanger sequencing of the whole region, illustrating the limitations of this approach and supporting the need for the application of high-throughput technologies in this area.

1.4 Glaucoma

Glaucoma is the world leading causes of irreversible blindness, and is predicted to affect approximately 80 million people by 2020 (Liu and Allingham, 2011; Quigley and Broman, 2006). According to the World Health Organization (WHO) estimation, in

2010 2% of visual impairment and 8% of blindness were caused by glaucoma (www.who.int/blindness/GLOBALDATAFINALforweb.pdf).

Since glaucoma consists of many diseases with overlapping phenotypic features, researchers have studied forms of glaucoma which are grouped according to specific clinical features (e.g. age of onset). There are many types of glaucoma. The major types are primary open-angle glaucoma (POAG) and angle-closure glaucoma (ACG), which are classified based on the anatomy and appearance of the irido-corneal angle, located at the junction between the iris and cornea in the anterior chamber (Quigley, 2011). POAG could be juvenile or old onset, depending on the age of diagnosis (old-onset POAG is typically diagnosed after age of 50 (Khan, 2011)). There are less common forms of glaucoma including primary congenital glaucoma (PCG), where genetics is a main causative factor.

Glaucoma is a heterogeneous group of disorders which are caused by the progressive loss of retinal ganglion cells. Ganglion cells are specialised neurons located near the inner surface of the retina. The axons of retinal ganglion cells (RGCs) make up the optic nerve, so these cells are critical in transferring visual signal from the eye to the brain (Gupta and Yucel, 2007).

Each type of glaucoma may have different triggers and only a little is known about these triggers. Imperfections of the trabecular meshwork and anterior chamber functions lead to derangement of intraocular fluid outflow, progressive optic nerve degeneration, loss of nerve cells and eventually blindness. Factors that are not affected by intraocular pressure also contribute to glaucoma degeneration. For example, apoptosis (programmed cell death) has been shown to be one of the mechanisms of retinal ganglion cell loss in glaucoma (Vasiliou and Gonzalez, 2008).

1.4.1 Primary congenital glaucoma

Congenital glaucoma is an important cause of blindness in children. Congenital glaucoma is a non-syndromic abnormality characterised by developmental defects of the trabecular meshwork which prevent adequate drainage of the aqueous humor and results in increased IOP, corneal clouding and buphthalmos (enlargement of the eyeball). One of the most common childhood glaucomas (both infantile and congenital) is primary congenital glaucoma (PCG). Also called primary infantile glaucoma, it accounts for 25% of paediatric glaucoma cases (Liu and Allingham, 2011).

Epidemiology and genetics of PCG

PCG is known as an autosomal recessive condition with higher frequency in consanguineous populations (Khan, 2011). In western countries the incidence is 1:30,000 (Khan, 2011) whereas in some populations such as Saudi Arabia (Bejjani, et al., 1998) and Slovakian Gypsies (Gencik, 1989; McGinnity, et al., 1987) the incidence is estimated to be 1:2,500. Europe has an average incidence of 1:10,000 (Sarfarazi, et al., 2003) where as in Australia it is reported to be 1 in 30,000 (MacKinnon, et al., 2004).

Four different loci associated with PCG have been identified through genetic linkage studies. Linkage analysis of 17 Turkish families identified a genetic region named *CLC3A* on chromosome region 2p21 in 11 families (Sarfarazi, et al., 1995). The second locus (*GLC3B*) was mapped to 1p36 chromosomal region in 4 Turkish families with PCG (Akarsu, et al., 1996). The third locus, *GLC3C*, is located on 14q24.3–14q31.1 (Sarfarazi, et al., 2003) and the fourth locus *GLC3D*, adjacent to the *GLC3C* locus was mapped to 14q24 in 2 consanguineous Pakistani families in which the genes associated with PCG in *GLC3A* and *GLC3D* loci were identified (Firasat, et al., 2008).

CYP1B1

The first gene reported to be involved in the pathogenicity of PCG, *CYP1B1*, was identified via positional cloning and is located within the *GLC3A* locus (Stoilov, et al., 1997). Over 60 mutations in *CYP1B1* are reported to be associated with PCG (Vasiliou and Gonzalez, 2008) and are listed in Appendix 6. The frequencies of *CYP1B1* mutations in familial cases were reported to be 100% in Slovakian Roma, 73 % in Saudi Arabia, 50% in Brazil and 21.6% in Australia (Dimasi, et al., 2007; Sarfarazi and Stoilov, 2000).

LTBP2

Autosomal-recessive mutations, associated with PCG, have been reported at the *GLC3D* locus within the *LTBP2* gene. The first mutations in this gene were reported in 4 Pakistani families (Ali, et al., 2009) and involvement of the gene was confirmed in a PCG case with Roma/Gypsy background (Azmanov, et al., 2011). Performing genome wide autozygosity mapping in 2 consanguineous Iranian families, 2 further novel segregating loss of function mutations were detected in this gene (Narooie-Nejad, et al., 2009). Mutations reported in *LTBP2* associated with PCG are listed in Appendix 6. Latent Transforming beta Binding Protein 2 (LTBP2) is a matrix protein involve in tissue repair and cell adhesion (Narooie-Nejad, et al., 2009). LTPB2 expression was detected in trabecular meshwork and ciliary body (Narooie-Nejad, et al., 2009) however its involvement in the mechanism of disease is not understood.

PXDN

Mutations in *PXDN* were reported to be associated with a PCG-like syndrome consisting of corneal vascularisation, and opacity in three families from Cambodia and Pakistan (Khan, et al., 2011). It has been suggested that PXDN is important for normal

development of the cornea and lens and possibly in structural support and it might have a functional role as an antioxidant to protect trabecular meshwork and the cornea from oxidative damage (Khan, et al., 2011).The reported mutations in *PXDN* associated with PCG are listed in Appendix 6.

MYOC

MYOC is located on chromosome 1 and codes for Myocilin protein (previously known as Trabecular Meshwork Inducible Glucocorticoid Response induced (TIGR) protein) (Tanwar, et al., 2010). Its expression has been detected in the trabecular meshwork (Karali, et al., 2000). This gene is the most common gene known to be involved in both primary open-angle glaucoma and adult-onset primary open-angle glaucoma. The possible role of *MYOC* in PCG was proposed by Kaur *et al.* (Kaur, et al., 2005). A digenic mode of inheritance of PCG was also proposed due to identification of a patient with heterozygous mutations in both *CYP1B1* and *MYOC*. *De novo* heterozygous mutations have been also reported in PCG cases (Kaur, et al., 2005).

FOXC1

FOXC1 (Forkhead-Related Transcription Factor 3) is a member of the winged helix/forkhead family of transcription factors and is located on 6p25. Its expression has been detected in trabecular meshwork (Wang, et al., 2001). Mutations in this gene also cause anterior segment dysgenesis which includes a wide variety of developmental conditions affecting the cornea, iris, and lens, with glaucoma being a main feature. Mutations in *FOXC1* can also cause isolated glaucoma like PCG (Kong, et al., 2015).The mutations reported in *FOXC1* reported to cause PCG are listed in Appendix 6. The reported mutations are located in a highly conserved 110-amino-acid DNA-binding domain, known as the forkhead domain (Chakrabarti, et al., 2009). The presence of double heterozygotes in *FOXC1* and *CYP1B1* in some probands suggests

the possibility of digenic inheritance in the pathogenicity of PCG, similar to *MYOC* and *CYPIB1* (Chakrabarti, et al., 2009).

1.5 Mendelian disease gene discovery and candidate gene screening methods

Traditional gene discovery methods for monogenic disorders use linkage analysis. Linkage analysis usually relies on shared heritage between people within a family. The genomic region(s) which carry the causative mutations (interval) will be shared between affected individuals and the size of this region will decrease with additional family members providing more recombination events between the markers and the causative mutation. The use of dense marker maps, such as modern single nucleotide polymorphism (SNP) arrays, can also help narrow down the region of interest (Majewski, et al., 2011). Subsequently genes at the linked locus are directly sequenced using either candidate gene or more systematic approaches to identify the segregating mutation. Almost all of the genes described in other sections of this chapter (apart from *AGK*) have been discovered using traditional gene discovery methods.

The two main problems with the gene mapping approach is that it does not always reduce the number of candidate genes to a level which makes Sanger sequencing feasible, and it is difficult, if not impossible, to be carried out in a small family with few affected members. Mendelian diseases are usually rare and monogenic, with high penetrance mutation(s). With the development of high-throughput massively parallel sequencing or "next generation sequencing", it is now possible to generate a vast amount of sequencing data at a much lower cost compared to traditional Sanger sequencing. About 2% of the human genome is protein coding. This portion of the genome is known as the exome as it contains all the exons in the genome. A high percentage of Mendelian disease causing mutations occur within coding regions of the

genome, and the effect of these mutations is interpretable, which makes whole exome sequencing a feasible method for disease gene/mutation discovery. Aldahmesh *et al.* (2012 a) has reported the identification of the *AGK* gene involved in paediatric cataract using a next generation exome sequencing method (Aldahmesh, et al., 2012a). Other paediatric cataract associated genes which were identified using next generation sequencing method are *CRYBA2* (Reis, et al., 2013) and *LSS* (Zhao, et al., 2015). These genes are not included in this study since their association to this disease was reported after the initiation of this research.

An important application of next generation sequencing is high throughput targeted gene screening. This approach can be applied for mutation screening of known causative genes when there are many genes involved in a heterogeneous condition such as in paediatric cataract. It is far more cost and time effective to screen many genes at once with this method compared to Sanger sequencing.

1.6 Aims

The childhood blindness rate is calculated to be 0.3 per 1000 children in developed countries while it is 1.5 per 1000 children in developing countries (Gilbert and Foster, 2001). A significant portion of these children have preventable blindness if their condition was detected and treated earlier. Blindness due to paediatric cataract or PCG can be caused by inherited genetic mutations. To date genes encoding structural proteins, membrane protein, enzymes, transcription factors and signalling molecules have been associated with isolated or non-isolated paediatric cataract. Six genes have been reported to be associated with PCG. Our group has a large and expanding congenital eye diseases repository consisting of DNA samples and clinical information of patients and their family members, with informed consent and ethics approval for genetic studies. The gene responsible for the phenotype in approximately 60 Australian

paediatric cataract families in this repository is still unknown. We also have DNA and relevant medical information of patients from Asian countries including Sri Lanka, Buthan and Cambodia. Furthermore there are over 40 PCG probands without mutations identified in our repository. In this thesis, I set out to achieve a better understanding of the genetic contribution to paediatric cataract and PCG within these cohorts.

1.6.1 Aim 1: Screening known paediatric cataract and PCG genes

Screening known paediatric cataract genes

We generated libraries with an Ion AmpliSeq Custom Primer Panel (Life Technologies) which were then sequenced on an Ion Torrent Personal Genome Machine (Life Technologies) in order to screen Australian and South East Asian probands with paediatric cataract for mutations in 51 previously reported disease-causing genes. With more than 50 genes already described for paediatric cataract and limited phenotype-genotype correlation, molecular diagnosis for this disease was not feasible before the emergence of next generation sequencing due to the difficulties with predicting which genes might be involved in any given patient. We hypothesized that a methodology to screen all known paediatric cataract causing genes would identify mutations in a large proportion of previously unsolved patients.

Screening the novel PCG gene (*TEK*)

In order to evaluate the contribution of a recently identified PCG associated gene to the disease, we aimed to screen this gene (*TEK*) in our cohort (Souma, et al., 2016). The Angiopoietin receptor TEK (tunica interna endothelial cell kinase) is a receptor tyrosine kinase highly expressed in the Schlemm's canal (SC) endothelium (Thomson, et al., 2014). It is required for SC development in mice (Kizhatil, et al., 2014). Mutations in the *TEK* gene were originally identified in 3 of 35 PCG patients from the USA by our

collaborators (Souma, et al., 2016). We aimed to assess the contribution of this gene to PCG pathogenicity in our Australian cohort.

1.6.2 Aim 2: Undertaking gene discovery in paediatric cataract and PCG families

Some families screened in Aim 1 (1.6.1) were found not to have mutations in the 51 known paediatric cataract genes. Of these, we selected 4 well-structured families and performed whole exome sequencing on selected individuals aiming to discover new genes for paediatric cataract. Furthermore, we selected 1 PCG family for this gene discovery project. We selected these families based on the following criteria:

1) A well-structured family with at least 2 affected individuals in the family
2) DNA available from unaffected family members
3) Sufficient phenotypic details

We sequenced whole exomes in multiple family members using "Next Generation Sequencing" or massively parallel sequencing. Afterwards, we aimed to identify the protein-changing mutations, which were assessed for segregation .The segregating variants were screened in a control population. The novel genes identified in families with paediatric cataract, were then sequenced in other probands with unidentified mutations.

Linkage analysis was also performed on those families in which we didn't find any segregating variants. Functional analysis aimed at further understanding the disease mechanism was performed with available lens material from one family.

Overall, this project characterised the genetics of congenital cataract in Australia and allowed us to take full advantage of this valuable repository for the identification of novel cataract and PCG genes. Direct feedback to patients and their families is possible

with this approach, and this project supports the utility of this technology for future molecular diagnostics of paediatric cataract.

Chapter 2: Materials and methods

This chapter contains the general methods used in the thesis and each chapter has its own specific methods section.

2.1 Ethics statement

All of the studies were conducted in accordance with the revised Declaration of Helsinki and following the National Health and Medical Research Council statement of ethical conduct in research involving humans.

Primary congenital glaucoma: Ethics approval was obtained from the Southern Adelaide Clinical Human Research Ethics Committee.

Australian paediatric cataract cohort: The cataract study was approved by the Southern Adelaide Clinical Human Research Ethics Committee, Adelaide, Australia, and the Royal Victorian Eye and Ear Hospital (RVEEH) Human Research and Ethics Committee, Melbourne, Australia. The ethic approval regarding eye lenses were used in this study (normal eye lens and CSA106.06) was obtained from Southern Adelaide Clinical Human Research Ethics Committee (SAC HREC).

Asian paediatric cataract cohort

Cambodian paediatric cataract cases: Permission to visit schools was granted by the Ministry of Health, Cambodia. The study was approval by the National Ethics Committee for Health Research in Cambodia and the Royal Adelaide Hospital Research Ethics committee (Sia, et al., 2010). The ethic approval for the genetic part was obtained from Southern Adelaide Clinical Human Research Ethics Committee.

Sri Lankan paediatric cataract cases: Ethics approval was obtained from the Faculty of Medicine, University Of Peradenya Ethical Review Committee in Sri Lanka and the Royal Adelaide Hospital Research Ethics Committee (Gao, et al., 2011). The

ethic approval for the genetic part was obtained from Southern Adelaide Clinical Human Research Ethics Committee.

Bhutanese paediatric cataract cases: ethic approval was obtained from the research committee from the National Referral Hospital, Thimphu, Buthan and the Royal Adelaide Hospital Research Ethics Committee (Farmer, et al., 2015). The ethic approval for the genetic part was obtained from Southern Adelaide Clinical Human Research Ethics Committee.

2.2 Participant recruitment

The cohorts in this study were recruited from two different populations: Australia and Asia (Cambodia, Sri Lanka and Bhutan).

Australian paediatric cataract cohort

In the Australian cohort, the probands in each family or sporadic paediatric cataract cases were recruited from the eye clinic at Flinders Medical Centre (Adelaide), the Women's and Children's Hospital (Adelaide), the Royal Children's Hospital (Melbourne) or the Royal Victorian Eye and Ear Hospital (Melbourne). Written informed consent was obtained from all participants or their guardians if they were under 18 years old. A detailed family history was obtained and additional affected and unaffected family members were invited to participate in the study. An ophthalmologist examined all available family members.

The majority of 65 screened Australian paediatric cataract probands were from Caucasian ethnicity (95%). Aboriginal and Middle Eastern patients were accounted for 3% and 2% of the total screened cases.

Cambodian paediatric cataract: Children under 16 years of age attending all four schools for the blind in Cambodia were examined and written consent was obtained

from parents or school principals (Sia, et al., 2010). Total of 5 available Cambodian cases were included in this study.

Sri Lankan paediatric cataract: Children under 16 years of age attending thirteen schools for the blind children were examined. Written consent was obtained from parents or the principal of each school (Gao, et al., 2011). Fourteen available Sri Lankan cases were included in this study.

Bhutanese paediatric cataract cases: Children under 16 years of age with visual impairment or blindness from schools in Bhutan and Laos were included in the study. The consent for each participant was obtained from parent, guardian or other authorized persons (Farmer, et al., 2015). Fourteen available Bhutanese cases were included in the study.

PCG samples

Individuals with PCG were recruited through the Australian and New Zealand Registry of Advanced Glaucoma by referral from their ophthalmologists (Souzeau, et al., 2015). Informed written consent and a blood sample for DNA extraction purposes were obtained. Clinical information was collected by the patient's clinical ophthalmologist (Souzeau, et al., 2015).

The majority of the total 64 screened PCG cases were from a Caucasian background (81%). Patients with African, Asian and Middle Eastern backgrounds were accounted for 6%, %5 and 6% accordingly. Two percent of the cases were of mixed ethnicity.

The diagnosis of PCG was based on combinations of corneal enlargement and buphthalmos, loss of corneal transparency, photophobia, raised IOP, or optic disc cupping. PCG was defined by the following characteristics: (1) age of onset less than 3 years; (2) increased corneal diameter greater than 10mm accompanied by corneal edema and/or Haab striae and (3) IOP greater than 21 mmHg and/or optic nerve cupping

greater than 0.4. Any patient with other ocular abnormalities or systemic conditions, was excluded from the study.

2.3 Control population

Above 300 normal Australian were recruited from Flinders Medical Centre, Adelaide, Australia and used as local controls to be screened for the candidate variants. These controls were recruited from the volunteer service at FMC and from retirement villages in Adelaide. They are all over 50 years old and had a full eye examination to rule out glaucoma and congenital cataract. They were not excluded for having age-related cataract. Furthermore, we had access to the exome sequences data of patients with POAG or PCG but not paediatric cataract. These exome data were also used to screen the observed variations in *BFSP2* and *HSF4* in family CSA92 (besides screening using genotyping technologies).

2.4 DNA extraction

Blood samples were kept at 4^{o}C in 10 millilitre DNA tubes (Vacuette) before extraction. Genomic DNA was extracted from peripheral whole blood using QiaAmp DNA Blood Maxi Kit (Qiagen, Hilden, Germany). The saliva is collected in the DNA saliva collection kit (Oragene DNA saliva collection kit) then, extracted using the using prepIT L2P (DNA Genotek Inc., Ottawa, Ontario, Canada), or buccal swab using Gentra PureGene Reagent (QIAGEN Pty Ltd, VIC, Australia) according to the manufacturers' protocols.

Whole genome amplification

DNA samples with low quantity were amplified using illustra GenomiPhi DNA amplification kit (GE Healthcare, NSW, Australia). One to five microliters of genomic DNA was mixed with sample buffer to make up a volume of 10 µl and denatured at

95°C for 3 minutes. The master mix with 9 µl reaction buffer and 1 µl of DNA polymerase enzyme was added to the existing mixture which was amplified at 30°C for 90 minutes. The enzyme was deactivated at 65°C for 10 minutes.

2.5 Primers design

Forward and reverse primers were designed using primer3 online tool (Boycott, et al., 2013; Kajiwara, et al., 1994) (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and were synthesized at GeneWorks (Adelaide, Australia). The list of primers used in this study is given at the appendices section of this thesis (Appendices 7, 10, 11, 12, 18, 19 and 20).

2.6 Validation of potential mutations by Sanger sequencing

Direct Sanger sequencing was used to confirm the detected protein changing mutations in probands and to evaluate the segregation of the mutation in families.

PCR reactions of 20µl final volume consisted of 1X Coraload PCR buffer (Qiagen) which gave a final concentration of 1.5 mM Mg^{2+}, 0.1mM dNTPs (Roche Diagnostics, Risch-Rotkreuz, Switzerland), 0.5 µM each primer, 0.5U Hot Star Plus Taq Polymerase (Qiagen) and 40 ng of gDNA. Five times Q Solution (Qiagen) was included at a final concentration of 1X as required and water volume adjusted accordingly. PCR was performed on a Palm Cycler (Corbett Life science, Qiagen) with 1 cycle at 95°C for 5 minutes, followed by 30 or 35 cycles at 95°C for 30 seconds, 57°C -65°C (annealing temperature, see the Appendices 7, 10, 11, 12, 18, 19 and 20 for each specific primer sets) for 30 seconds and 72°C for 30 seconds, and a final extension step at 72°C for 5 minutes. A 1% TBE agarose gel containing 1mg/ml ethidium bromide was used to confirm the DNA amplification. Five µl of PCR product was used to run the gel. Three micro-litres of 0.125µg/µl 100 base pair ladder (BioLabs Inc, New england) was loaded

and this was used to determine the size of the PCR products. The gel was electrophoresed at 110V for 35 to 45 minutes. To clean the PCR products for sequencing, 5µl of PCR product, 2µl Shrimp Alkaline Phosphatase (SAP; 1 units/µl) and 0.5 µl (20 units/µl) of Exonuclease 1 (Exo1) (New England Biolabs, Massachusetts, USA) were mixed. Reactions were incubated at 37°C for 1 hour, followed by incubation at 80°C for 20 minutes to inactivate the enzymes.

Sequencing reactions were prepared with the respective forward primer at 5 µM and purified PCR product at 10ng/100bp (i.e. 30ng for 300bp product) combined with BigDye Terminator v3.1 (Life Technologies), 5x Sequencing Buffer (Life Technologies) and made up to 20µl with water. Reactions were taken through a cycle sequencing PCR protocol on a MasterCycler thermal cycler (Eppendorf, Hamburg, Germany). PCR extension products were purified using Agencourt CleanSeq Magnetic Beads and a SPRI plate, according to the manufacturer's protocol (Beckman Coulter, California, USA). Purified extension products were then resolved using POP-7 polymer on the 3130xl Genetic Analyser (Life Technologies) in the IMVS Sequencing Facility (Flinders Medical Centre, Adelaide, Australia).

Sequence chromatograms of affected and unaffected individuals were compared to each other and the reference sequence using Sequencher v.5 (GeneCodes Corporation, Ann Arbor, MI, USA).

2.7 Genotyping methodology for control population screening

Sequenom MassARRAY

The majority of SNPs detected in the pediatric cataract and PCG samples were genotyped using the Sequenom MassARRAY® platform (Sequenom, San Diego, USA). The MassARRAY platform utilizes the iPLEX GOLD chemistry on a Sequenom

Autoflex Mass Spectrometer. This methodology was performed at the Australian Genome Research Facility, Brisbane, Australia.

TaqMan SNP Genotyping Assays

Variants identified in families CRCH139, CSA133, CSA95, CSA92 and PCG002 were screened in controls using custom TaqMan SNP genotyping assays (Life Technologies) following standard manufacturer's protocols. In summary, the 40× SNP assay was diluted with an equal volume of Tris-EDTA (TE) buffer to obtain a 20× concentration of the assay. Forty nanograms DNA was amplified using 5 µl of 2× Taqman genotyping mix (Applied Biosystems, VIC, Australia) and 0.5 µl of 20× SNP genotyping assay in a total 10 µl volume. AmpliTaq Gold Enzyme was activated at 95ºC for 10 minutes, followed by 40 cycles of denaturation for 15 seconds at 92ºC and annealing/extension for 1 minute at 60ºC. The assay was performed on a StepOne Plus RealTime PCR System (Applied Biosystems, VIC, Australia) and results were analyzed using StepOne software v2.1 (Applied Biosystems, VIC, Australia).

2.8 Variant Filtering strategy

Variants were filtered at different stages based on the criteria below considering the pattern of inheritance in the family:

1: Sequence quality: A quality and read depth of more than 19 was required (for exome data).

2: Protein changing: generally Nonsynonymous or presumed loss of function variants were selected for further analysis (missense, nonsense, the stop loss, stop gain, frameshift deletion, frameshift insertion and deleterious exonic/ splicing). The Sorting Tolerant From Intolerant (SIFT) (Kumar, et al., 2009)and/or Polyphen-2 (Adzhubei, et al., 2010) algorithms were used to assess the pathogenicity of missense variants.

3: Being novel, rare or known: the variants were considered for further analysis if they were absent from public databases (i.e. novel, not reported in dbSNP137 or ExAc (http://exac.broadinstitute.org/)) or were present at a Minor Allele Frequency (MAF) of less than 1% (rare) or have been previously reported to be associated with congenital cataract.

4: Validating: Variants meeting criteria 1 to 3 were validated by Sanger sequencing as described in sections 2.5 and 2.6 to filter out the false positive signals.

5: Segregating: Validated variants were assessed for segregation with the phenotype by Sanger sequencing in those additional family members who were not exome sequenced.

6: Ethnically matched control population screening: The segregating novel/rare protein changing variants were screened in local controls.

2.9 Definition of some terms

Novel gene: A gene related to a phenotype (here PCG and paediatric cataract) that hasn't previously been reported to be associated with the disease.

Known gene: A gene variants in which have been previously reported to be associated with the disease

Novel variant: A variant that hasn't previously been reported to be associated with the disease.

Known variant: A variant previously reported to be associated with the disease.

Rare variant: A variant that is present in public databases such as dbSNP and ExAC with a Minor Allele Frequency of less than 1%.

Phenotype: The collection of observable traits (in paediatric cataract or PCG conditions) of an affected individual.

Chapter 3: Investigating the genetics of primary congenital glaucoma in Australian cases

3.1 Introduction

Glaucoma is a leading cause of blindness worldwide and affects more than 60 million individuals (Quigley and Broman, 2006). The underlying cause of glaucoma is not clear, however ocular hypertension is a main risk factor (Quigley, 2011). Primary Congenital Glaucoma (PCG) is a type of childhood glaucoma. The symptoms, which mainly are high IOP, enlarged eye, optic neuropathy and cloudy cornea show in the first three to five years of life; however more than 80% of cases show symptoms within the first year of life (Vasiliou and Gonzalez, 2008). The disease incidence ranges from 1:1,250 to 1:30,000 in different ethnic groups and is higher in some populations with high levels of consanguinity (Gencik, 1989; Lim, et al., 2013).

A promising candidate gene involved in the pathogenicity of PCG was recently identified by our collaborators in the USA, led by Professor Terri Young at University of Wisconsin. They identified mutations in the angiopoietin receptor gene, TEK, in patients with PCG in their cohort. Of 35 patients screened by exome sequencing, 3 had likely pathogenic mutations in the *TEK* gene. TEK (tunica interna endothelial cell kinase, also known as Tie2) is a tyrosine kinase receptor and its expression has been detected in Schlemm's canal (SC) endothelium and is involved in SC development in mice (Thomson, et al., 2014).

In this part of the study, we evaluate PCG patients for mutations in *TEK*. In addition, we also used both whole exome sequencing and Sanger sequencing in an Australian family to identify additional novel candidate genes for PCG.

3.2 Aims

1: To determine involvement of mutations in recently identified gene, *TEK*, in PCG in Australian cases using whole exome and Sanger sequencing

2: To identify a novel candidate gene in an Australian family with no mutation in known PCG genes using whole exome sequencing.

3.3 Methods

3.3.1 Candidate gene screening

<u>*TEK* gene screening in PCG cohort</u>

A total of 64 PCG samples were sequenced for *TEK* gene using either Sanger (14 PCG cases were directly sequenced just for *TEK* gene), or whole exome sequencing (22 PCG cases) or a combination of both methods (28 PCG cases). Of the total of 64 sequenced samples, 11 were excluded from further analysis due to identification of causative variations in other PCG related genes (the analysis of this part was done by Dr. Owen Siggs and was not part of the current project). Therefore, the remaining 53 PCG cases were screened for the mutations in the *TEK* gene.

All 23 exons of the *TEK* gene (GenBank accession NM_000459.4 & NP_000450.2) were amplified by PCR and Sanger sequenced at the Australian Genome Research Facility (AGRF). All samples that failed to produce a sequence using automated systems at AGRF were sequenced using in-house primers. The primer sequences were provided by AGRF, and the in-house primer sequences (Exon5-V2 and Exon 18-V2) are listed in Appendix 7.

Fifty three PCG cases were subjected to exome capture (Agilent SureSelect v4) and paired-end sequencing on an Illumina HiSeq 2000 by an external contractor (Macrogen Inc, Seoul, South Korea).These PCG cases had been screened for variants in the known disease-causing genes, *CYP1B1*, *LTBP2*, *FOXC1* and *MYOC* and found to be negative (Souma, et al., 2016). The data were screened for potentially heterozygous and homozygous pathogenic variants in *TEK* gene. All non-synonymous changes were

examined for their presence in dbSNP build 137 and ExAC databases. Any potentially disease causing variants in an individual were further assessed for segregation with the disease in other available affected and unaffected family members by direct Sanger sequencing as described in Chapter 2, section 2.6. The sequence of the exons of the *TEK* gene in the Sanger sequenced probands and other candidate variants detected in exome data were visualized and compared to the reference sequence using Sequencher 4.1.0 (Gene Codes Corporation, MI, USA).

Statistical analysis

The two tailed Fisher's exact test (http://udel.edu/~mcdonald/statfishers.html) was used to compare the mutation frequencies in the PCG cohort with 60706 publicly available normal controls from the Exome Aggregation Consortium (ExAC) database (http://exac.broadinstitute.org/).

3.3.2 Whole exome sequencing in an affected family

Exome sequencing and filtering of the variants

Selected family members of PCG002 were exome sequenced through our collaborator at State Key Laboratory of Medical Genetics, Central South University, Changsha, China. Paired-end libraries were sequenced on the Illumina GAII platform and read summaries were provided (Appendix 8). Above 90% of reads in the target region had a minimum coverage of 8 fold. Variant call format (VCF) files were annotated using SeattleSeq Variation Annotation website against db SNP137 and db SNP 138 (http://snp.gs.washington.edu/SeattleSeqAnnotation137/; http://snp.gs.washington.edu/SeattleSeqAnnotation138/). The filtering strategy is described in section 2.8, chapter 2 (except that both detected variants in *GREB1* gene were synonymous but were included in the study).

Primer design, segregation analysis and control population screening

The primers for candidate variants were designed using primer 3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). The two candidate variants were 60 bp apart and were amplified in the same PCR fragment (Appendix 7) and sequenced in all members of family PCG002 as described in Chapter 2, section 2.6.The segregating mutations in *TEK* gene were assessed in using Sequenom MassArray on a Compact Spectrometer using iPLEX GOLD chemistry in 333 unrelated normal controls at AGRF, Brisbane, Australia. Both detected variants in *GREB1* were screened in 276 controls using custom Taqman assays as described in section 2.7 of Chapter2.

Predictions of the secondary structure and hydrophobicity of human wild type and mutated GREB1 protein

PSIPRED online web tool (http://bioinf.cs.ucl.ac.uk/psipred/) was used to predict the secondary protein structures for wild type and mutated *GREB1* protein. The peptide2 program was used to predict the protein hydrophobicity (http://www.peptide2.com/N_peptide_hydrophobicity_hydrophilicity.php).

3.4 Results

3.4.1 *TEK* gene screening in Australian PCG cohort

A total of 7 candidate mutations were considered for further investigation. Two 5' UTR variants were excluded from the study (one was a common SNP and the other was a Sanger sequencing error). Total of 5 variants were confirmed in the *TEK* in this Australian PCG cohort of 53 probands (Table 3.1). The variants were assessed for segregation in the appropriate family if other family members were available (Figure 3.1). Splice site variants were detected in PCG066 and PCG100; missense variants in PCG092 and PCG122 and a frameshift variant in PCG027. Patients PCG100, PCG027

and PCG066 were screened for *TEK* gene using both methods. The missense variants in PCG092 and PCG122 were predicted to be deleterious by SIFT, Polyphen-2 and mutation taster (http://www.mutationtaster.org/) and the frameshift and splice variants are assumed to be loss of function variants. Both variants in PCG066 and PCG100 were splice donor variants that change the 2 base pair region at the 5' end of an intron probably changing the 3' boundary of the upstream exon. All variants were heterozygous in the proband and inherited from an unaffected parent, except in PCG027 where the parents were unavailable for testing.

Table 3.1. Detected variants in *TEK* gene in an Australian PCG cohort. The table shows the detected variants chromosome position, Minor Allele Frequencies and the variant changes at DNA and protein level. Human genome build 19 was used for the chromosome position.

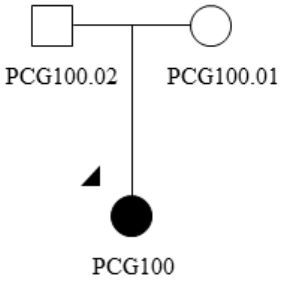
Family	Chromosome Position (hg19)	Minor Allele Frequency (ExAC)	DNA change	Protein change
PCG066	chr9:27172747	0	c.760+2T>C	Splicing
PCG092	chr9:27157924	5.7×10^{-5}	c.148C>T	p.Arg50Cys
PCG122	chr9:27213567	0	c.2963G>A	p.Gly988Asp
PCG100	chr9:27228305	0	c.3330+2delT	Splicing
PCG027	chr9:27204930	0	c.2673_2674insG	p.Lys745Glufs*76

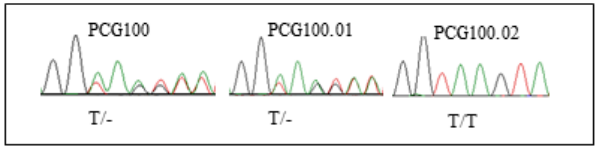
All the patients with mutation in *TEK* were from Caucasian background except PCG027 which is of unknown ethnicity.

3.4.2 Significant enrichment for loss of function mutations and missense variants in *TEK* in PCG patients

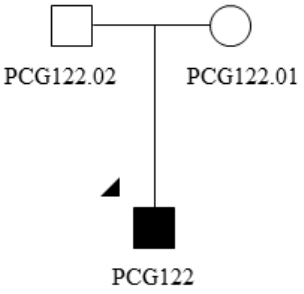
From screening the *TEK* gene in 53 Australian patients, five mutations were identified, consisting of two substitutions, two loss of functions and one frameshift mutation. Only two loss of function and 325 missense mutations have been reported in *TEK* in 60706 publicly available controls from the ExAC database. Performing a two-tailed Fisher's exact test, the extremely significant P values of 6.26×10^{-9} (when only loss of functions variants were considered) and 0.00001 (when all potentially pathogenic variants were included) demonstrate significant enrichment for protein changing mutations in PCG patients when compared to controls.

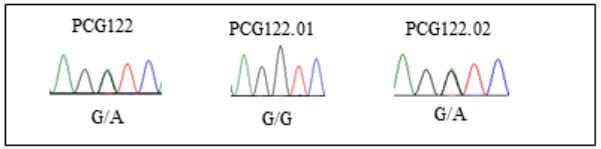
PCG100
***TEK*, c.3330+2delT**



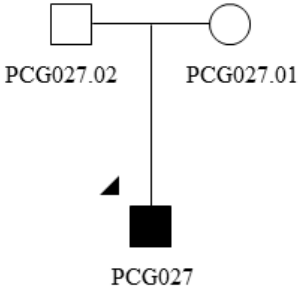


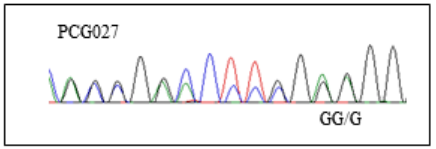
PCG122
***TEK*, c.2963G>A**
p.Gly988Asp



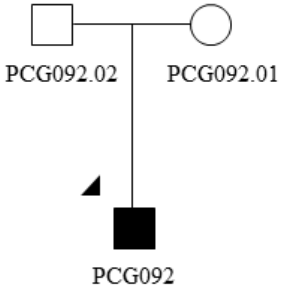


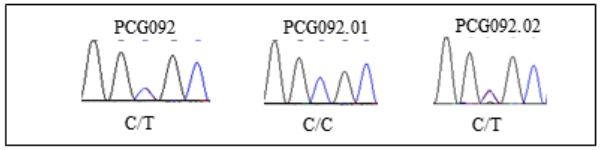
PCG027
***TEK*, c.2673_2674insG**
p.Lys745Glufs*76



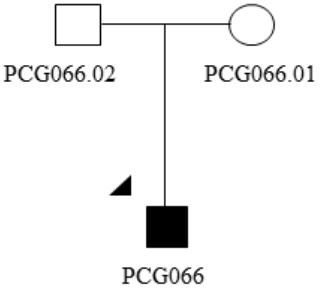


PCG092
***TEK*, c.148C>T**
p.Arg50Cys





PCG066
***TEK*, c.760+2T>C**



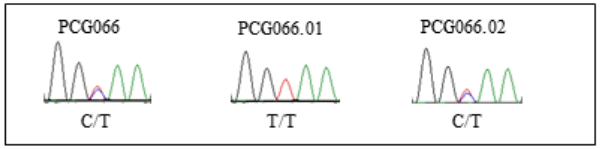
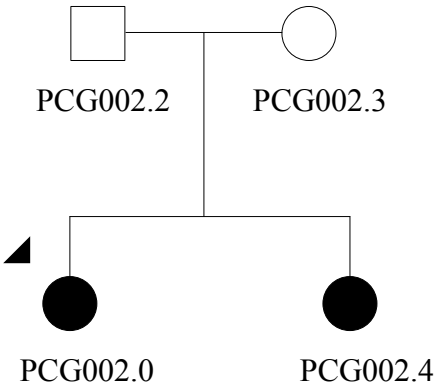


Figure 3.1. Pedigree and chromatogram sequences of the detected variants in *TEK* gene in probands and available family members. No other family members were available from family PCG027. The arrows indicate the proband. Solid circles indicate affected females and solid squares show the affected males.

3.4.2 Novel gene identification in an affected family

We sequenced 4 members of the family PCG002 (Figure 3.2 A), including two affected sisters (PCG002.0 and PCG 002.4) and the two unaffected parents (PCG002.2 and PCG002.3). Considering the hypothesis that PCG is a recessive disease, and due to the recessive inheritance pattern in this family, we expected a homozygous mutation in the two affected children and the parents were expected to be heterozygous for the change. No such variants were detected; therefore, compound heterozygous variants were investigated. We detected four heterozygous variations present in both affected children PCG002.0 and PCG002.4 in the *GREB1* and *RP1L1* genes which fit the compound heterozygous inheritance pattern. The variants in *RP1L1* were also detected in 12 controls and was excluded from the study (was present in exome sequence normal controls sequenced by the external provider in China). Therefore the only remaining candidate gene *GREB1* was considered for further analysis (Figure 3.2 B). The first variant in *GREB1* is heterozygous in the unaffected father (c.3577A>G, p.(S1193G)) and the second one is heterozygous in the unaffected mother c.3515C>G, (p.A1172G)). The mother's variant, c.3515C>G (rs140938943) is present in the ExAC database with MAF= 1.9%. This variant was present in our Australian control cohort with a similar MAF of 1.1% (6 of 552 control alleles). The father's variant is not present in public databases nor in our local control population. Both variants were predicted to be benign by both SIFT and Polyphen-2, however, this gene was the only candidate gene detected by exome sequencing and no bioinformatics tools are able to predict the effect of carrying both mutations simultaneously on the potential for disease development.

A



B

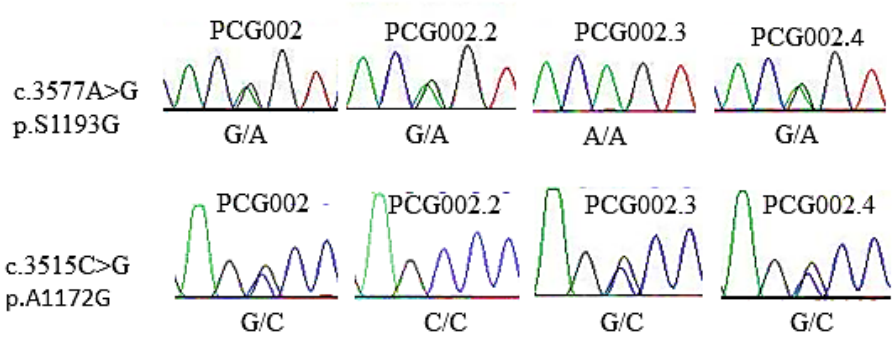


Figure 3.2. A: The pedigree of family PCG002. The square represents the unaffected father. Black circles represent the affected females (children) and the white circle shows the unaffected mother. All 4 members had full exome sequencing performed on their DNA. B: Chromatograms show the sequence of detected compound heterozygous variants in *GREB1* gene in all 4 family members revealed by Sanger sequencing.

PSIPRED did not predict any α-helix or β-sheet for the wild type proteins at the mutations location, and the prediction did not change for the mutant proteins (Figure 3.3). There was no difference in the predicted hydrophobicity of the wildtype and mutant protein hydrophobicity (based on PEPTIDE2 prediction).

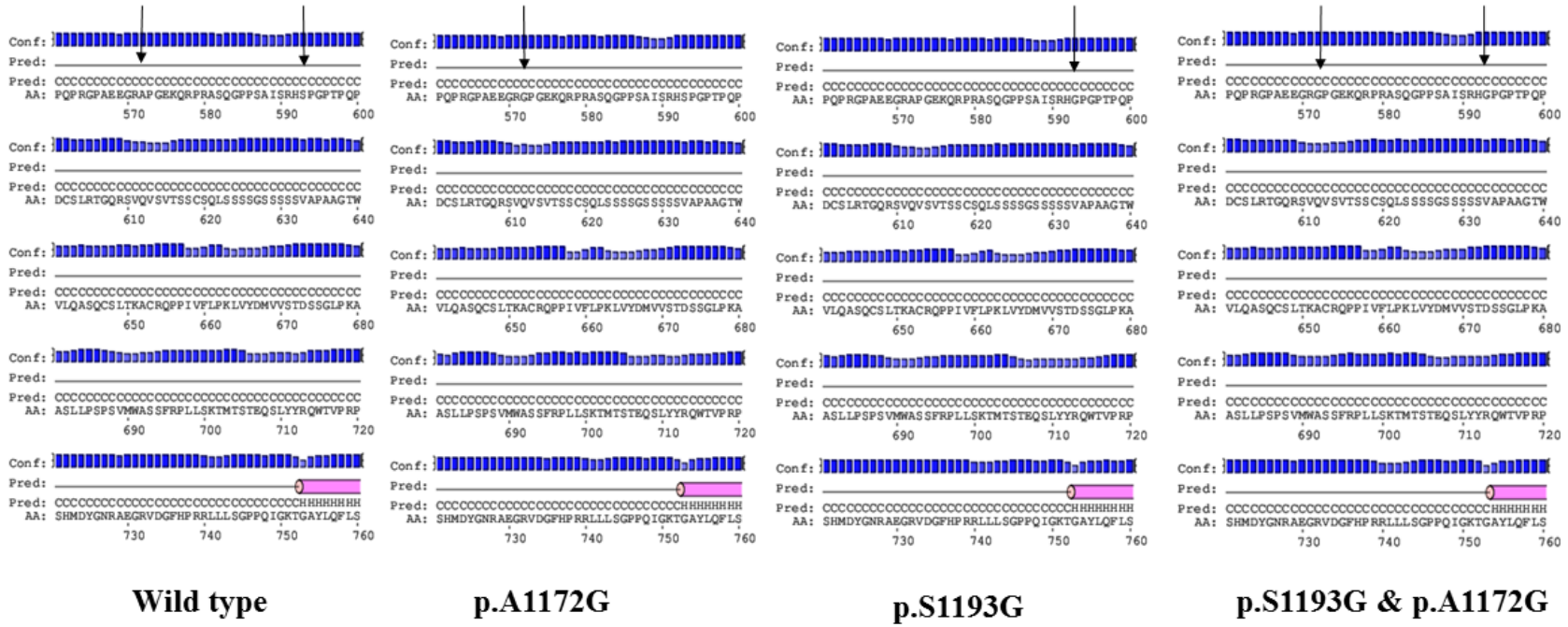


Figure 3.3. PSIPRED prediction of wild type and mutant proteins second structure. The softwear did not predict any α-helix or β-sheet for the wild type proteins at the mutations location, and the prediction did not change for the mutant proteins. The altered amino acid is marked with a black arrow. Yellow arrows represent β-sheet. Pink cyliders represent α-helix. The blue bars show the confidence in the prediction (on a scale of 1-10).

3.5 Discussion

3.5.1. Candidate gene screening: *TEK*

The study was able to detect mutations in *TEK* at a higher rate than in the general population. Variations in the *TEK* gene likely account for approximately 8% of PCG cases in Australia. Our collaborators demonstrate that TEK is highly expressed in the Schlemm's canal endothelium and is required for canal development in mice (Souma, et al., 2016). Confocal microscopy revealed normal development of SC in control mice as shown in Figure 3.4 (Souma, et al., 2016). A mouse model heterozygous for a Tek knockout mutation developed a severely abnormal canal with convolutions and focal narrowing (Figure 3.4 B). SC was completely absent in the Tek knockout mice (Figure 3.4 C). The study confirmed the absolute requirement for TEK during canal development. Moreover, elevated IOP in heterozygous mouse and null Tek mouse models was observed which is similar to the human PCG phenotype in carriers of heterozygous mutation in TEK (Souma, et al., 2016).

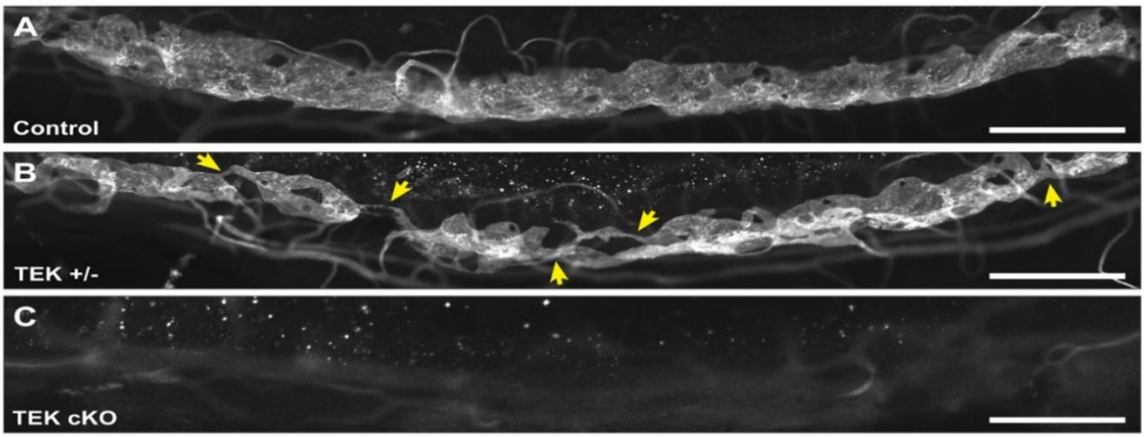


Figure 3.4. Evaluation of Schlemm's canal development in Tek mutant mice taken from (Souma, et al., 2016). Confocal microscopic images of (A) wild-type (control), (B) Tek heterozygous (Tek +/-), and (C) Tek null mice. The yellow arrows in panel B shows hypomorphic canal with convolutions. Panel C shows the absence of the Schlemm's canal in Tek knock out mice.

In the majority of PCG cases reported so far, the disease transmission is autosomal recessive, particularly in populations such as Turks, Romans and Saudis (Sarfarazi, et al., 2003), however autosomal dominant inheritance has also been considered, due to the disease heterogeneity (Sarfarazi, et al., 2003). The findings from the current research, including both the enrichment of mutations in PCG patients and the developmental abnormalities seen in heterozygous mice, demonstrate that PCG may not always be a recessive disease and dominant mutations with reduced penetrance also need to be considered. The inheritance patterns of all the 5 families in this study appeared to be recessive or *de novo*, but if *TEK* is the cause, this assumed inheritance pattern is incorrect. There are reports of PCG cases with heterozygous variations detected in *MYOC* or *FOXC1*, and also *CYP1B1* but with no second mutation detected (Kaur, et al., 2005; Medina-Trillo, et al., 2015). It has been thought the second mutation was yet to be identified, or maybe a second unknown gene was involved, but the current hypothesis indicates that dominant mutations with reduced penetrance should be

considered as well. There are also other forms of dominantly inherited congenital glaucoma (syndromic) like Riegers and Peters anomaly (Weisschuh, et al., 2008) and Axenfeld-Rieger syndrome (Seifi, et al., 2016). In a dominantly inherited disorder with reduced penetrance it's possible that the mutation doesn't lead to a phenotype in one generation despite its presence in one of the parents, but can be transmitted through unaffected parents to other generations where it starts to show itself. Examples of such generations skips are: mutation in *GLRA1* gene in hereditary hyperekplexia condition (Kwok, et al., 2001) and mutation in *SH3BP2* gene in Cherubism disorder (Preda, et al., 2010). It has been proposed that this variable penetrance might be due to involvement of modifier genes (Cooper, et al., 2013). An example of such modifiers is *CNOT3* which determines penetrance of *PRPF31* mutations (a gene which its malfunction is associate with Retinitis Pigmentosa), via a mechanism of transcriptional repression (Venturini, et al., 2012).

One of the main challenges of next generation sequencing data analysis is to find the pathogenic variants among the background of false positive signals. In order to narrow down the variants list, a multistep filtering approach is an essential strategy which has been widely used. One of the main risks of such strategies is losing the pathogenic variant during the filtering process. Interestingly we didn't detect the frameshift variants in PCG027 in the exome data despite the presence of the variant being detected by Sanger sequencing. In our selection criteria for finding novel *TEK* variants we set a cut-off threshold of 20 for SNP quality. The reason it wasn't picked up was the quality score of 4.42 which is far below the standard cut off of 20. We did have the whole *TEK* gene sequenced for the mentioned individual and we found the disease causing variants through direct sequencing. On the other hand the splice site variant detected in PCG066 was almost impossible to be detected by Sanger sequencing alone. The peak detected on the chromatogram for the minor allele was not the height one expects for a heterozygote

(probably due to some allele-biasing effect of the PCR). By initially detecting the variant through whole exome sequencing, we were able to confirm the variant via a second Sanger sequencing reaction with a fresh PCR product. It can be concluded that the combination of Sanger and next generation sequencing worked best for detecting novel variants in our PCG cohort. It is possible that additional mutations exist in the cohort that were missed because one method was used. Furthermore, the control frequency could be underestimated as some mutations could have been missed using whole exome sequencing technologies.

3.5.2 *GREB1*, a novel candidate gene: evaluation of its involvement in PCG

We detected compound heterozygous variants in a novel candidate gene, *GREB1*, in family PCG002. There are two major groups of estrogen receptors: ERα and ERβ. GREB1 is a chromatin-bound ERα coactivator and is essential for ER-mediated transcription due to its role in interactions between ER and additional cofactors (Mohammed, et al., 2013). Both receptors are present in human ocular tissues including lacrimal gland, lacrimal gland acinar (tissue involve in tear secretory system), palpebral and bulbar conjunctivae, cornea, iris, ciliary body, lens, retina, uvea, choroid and retinal pigment epithelial cells (Kobayashi, et al., 1998). They were also detected in the nerve fibre layer and ganglion cell layer in rat and bovine (Kobayashi, et al., 1998); however no report has been published showing their presence in Schlemm's canal or trabecular meshwork. The synthesis of collagen fibres is activated by estrogen. The decreased collagen and elastin density in several ocular tissues including sclera and cornea (Wei, et al., 2012) has been observed in glaucoma. Higher levels of collagen fibres might enhance flexibility of the whole eye which consequently can decrease IOP (Wei, et al., 2012). We hypothesize that being an ERα coactivator, *GREB1* may have an effect on

estrogen signalling regulation in the eye which subsequently could cause elevated IOP via misregulation of collagen synthesis. However this hypothesis needs to be tested by functional analysis in cell or animal models. Further, the presence of GREB1 in PCG relevant human ocular tissues need to be evaluated using techniques such as immunohistochemistry.

Both detected missense variations in *GREB1* were predicted to be non-pathogenic by both Polyphen-2 and SIFT, however how two proteins with different mutations will interact is unclear. More functional works need to be conducted to see the effect of these mutations at a same time on protein level. Although SIFT and PolyPhen-2 are useful tools in prioritizing variants (based on their potential pathogenicity) for further studies, their sensitivity was demonstrated to be approximately 70% which means predictions of non-pathogenic variants by these tools should be interpreted with caution and further support where it possible (Flanagan, et al., 2010).

Based on the assumption (made at the time of beginning this part of the research) of being PCG almost always being a recessive disorder, we looked for segregating recessive or compound heterozygous variations in family PCG002. However, with our current knowledge about the *TEK* gene, it is so important now to consider the possibility of autosomal dominant inheritance with reduced penetrance and not just recessive inheritance pattern when it comes to studying families such as PCG002. The reanalysis of family PCG002 using SeattleSeq Annotation 138 http://snp.gs.washington.edu/SeattleSeqAnnotation138/, we did not detect any potential homozygous causative variants shared between 2 affected children. Considering dominant mode of inheritance, when the variants were filtered for heterozygous novel protein changing variants (frameshift, missense, missense near splice, splice-acceptor, splice-donor, stop-loss, stop-gained), 42 missense variants (see Appendix 9) were shared between 2 affected children (PCG02 and PCG02.4). Of these shared variants 2

were present in both parents, 14 were present just in PCG02.2 and 20 were present just in PCG02.3. Six variants were not detected in any of parents. This illustrates the difficulty of finding new genes in small families when the hypothesis is low penetrance dominant inheritance pattern.

The InnateDB (http://www.innatedb.com/) database was used to see the possible interactions of *GREB1* with other genes. The InnateDB showed the interaction of this gene with *FOXA1* (Forkhead Box A1) and *ESR1* (Estrogen Receptor 1) (Meyer, et al., 2013). No such interactions were reported between *GREB1* and known PCG related genes.

3.6 Summary and conclusion

Five mutations including two substitutions, two loss of function and one frameshift insertion were identified in *TEK* gene in Australian PCG cohort (negative for mutations in other PCG associated genes including *CYP1B1*). The Fisher's exact test result demonstrates significant enrichment for loss of function mutations in PCG patients ($P=.6.26 \times 10^{-9}$). This highly significant enrichment suggests that this gene is highly likely to be involved in PCG pathogenesis, however the PCG numbers in the current study were small to be absolute confident about *TEK* role in PCG pathogenesis.

GREB1 transcripts have been detected in different ocular tissues above background estimation with the highest expression in optic nerve. Since no further candidate mutations in *GREB1* were detected in other PCG cases in Australian cohort (screened in 48 available PCG exomes), it could be concluded that *GREB1* may be a very rare cause of PCG. It is clear that the biological mechanisms need evaluation and other cohorts should be screened to determine if mutations are present.

Chapter 4: Massively parallel sequencing of known paediatric cataract genes identifies causative mutations in Australian and Asian cohorts

4.1 Introduction

The large number of genes known to cause paediatric cataract and the limited phenotype-genotype correlations reported makes clinical testing using traditional sequencing technologies challenging and expensive. Massively parallel (next generation) sequencing technologies are now feasible and cost effective tools to screen many candidate genes in parallel.

The Ion Torrent technology and Personal Genome Machine (PGM), which was released in 2010, provided a new approach in next generation sequencing technologies. Unlike other massively parallel sequencing platforms, it is based on semiconductor sequencing technology and does not rely on fluorescence and camera scanning (Liu, et al., 2012). During elongation of DNA molecules, when a new nucleotide is incorporated into the extending DNA strand by the polymerase enzyme, a proton is released which causes a pH change. The pH change can be detected via the Ion Torrent chip which acts as a pH meter (Merriman and Rothberg, 2012). By not requiring fluorescent labelling or chemiluminescent dNTPs, this technology is both faster and easier to use compared to other methods. In addition, Ion Torrent sequencing is less expensive (at the time of commencing the following research), and this combined with its other advantages make the PGM highly suited to targeted sequencing (Liu, et al., 2012; Merriman and Rothberg, 2012).

In this study, we screened our repositories of South Eastern Australian and Asian families/individuals with isolated paediatric cataract for mutations in known cataract-associated genes. This screening was carried out in a systematic fashion using the Ion Torrent Personal Genome Machine (PGM). We hypothesized that a significant proportion of congenital cataract cases would be accounted for by mutations in known

genes, and that screening genes in parallel would be a cost effective method for genetic testing in this heterogeneous disease.

4.2 Aim

The overall aim of the current study was to screen our repository of South Eastern Australian and Asian individuals with paediatric cataracts for mutations in known causative genes using the Ion Torrent Personal Genome Machine (PGM).

4.3 Methods

Patient recruitment and DNA extraction is described in section 2.2 of Chapter2. This study included probands from Australian and Asian paediatric cataract families, their family members where available and the control cohort as described in Chapter 2, section 2.3.

4.3.1 Known paediatric cataract genes selection for sequencing

Fifty-one genes known to cause paediatric cataract in human or mouse were selected through reviewing the literature (Table 4.1). The foundation of the list included all the genes covered in a review of paediatric cataract genes (Churchill and Graw, 2011) except the mouse gene *gjf1* which does not have a human analogue. Additional genes identified in recent publications (up until the beginning of the study in January 2013), were also included in the gene panel. We focused on those genes known to be associated with non-syndromic paediatric cataract, as this was the predominant phenotype in our cohort. In addition, we included some syndromic paediatric cataract genes known to cause an ocular syndrome (e.g. *PAX6* and *PITX3*) where paediatric cataract is a main diagnostic feature of the syndrome (*NHS, MAF*) as such genes may also contribute to non-syndromic cataract (Gillespie, et al., 2014). Some genes associate with cataract in mouse such as *SPARC* and *PCBD1* also were included.

Table 4.1. List of selected known congenital cataract genes for mutation screening

Gene	**Genbank Accession Number**	**Locus**	**Reference**
AGK	NM_023538.2	7q34	(Aldahmesh, et al., 2012a)
BFSP1	NM_001195.3	20p12.1	(Churchill and Graw, 2011; Ramachandran, et al., 2007)
BFSP2	NM_003571.2	3q22.1	(Churchill and Graw, 2011; Jakobs, et al., 2000)
CHMP4B	NM_176812.4	20q11.22	(Shiels, et al., 2007)
COL4A1	NM_001845.4	13q34	(Churchill and Graw, 2011; Sibon, et al., 2007)
CRYAA	NM_013501.2	21q22.3	(Churchill and Graw, 2011; Litt, et al., 1998)
CRYAB	NM_009964.2	11q23.1	(Berry, et al., 2001; Churchill and Graw, 2011)
CRYBA1	NM_009965.2	17q11.2	(Churchill and Graw, 2011; Kannabiran, et al., 1998)
CRYBA4	NM_021351.1	22q12.1	(Zhou, et al., 2010a)
CRYBB1	NM_023695.2	22q12.1	(Churchill and Graw, 2011; Willoughby, et al., 2005)
CRYBB2	NM_007773.3	22q11.23	(Churchill and Graw, 2011; Litt, et al., 1997)
CRYBB3	NM_021352.3	22q11.23	(Churchill and Graw, 2011; Riazuddin, et al., 2005)
CRYGA	NM_014617.3	2q34	(Churchill and Graw, 2011; Mehra, et al., 2011)
CRYGB	NM_005210.3	2q34	(AlFadhli, et al., 2012; Churchill and Graw, 2011)
CRYGC	NM_007775.2	2q33.3	(Churchill and Graw, 2011)
CRYGD	NM_007776.2	2q33.3	(Churchill and Graw, 2011; Heon, et al., 1999)
CRYGS	NM_017541.2	3q27.3	(Churchill and Graw, 2011; Sun, et al., 2005)
EFNA5	NM_001962.2	5q21.3	(Churchill and Graw, 2011)

EPHA2	NM_004431.3	1p36.13	(Churchill and Graw, 2011; Shiels, et al., 2008)
EYA1	NM_010164.2	8q13.3	(Azuma, et al., 2000)
FOXE3	NM_012186.2	1p33	(Bremond-Gignac, et al., 2010; Churchill and Graw, 2011)
FTL	NM_010240.2	19q13.33	(Nonnenmacher, et al., 2011)
FYCO1	NM_024513.3	3p21.31	(Chen, et al., 2011)
GALE	NM_000403.3	1p36.11	(Churchill and Graw, 2011; Schulpis, et al., 1993)
GALK1	NM_016905.2	17q25.1	(Churchill and Graw, 2011; Stambolian, et al., 1995)
GCNT2	NM_145649	6p24.3-p24.2	(Pras, et al., 2004); Yu, et al., 2001)
GJA1	NM_000165.3	6q22.31	(Churchill and Graw, 2011)
GJA3	NM_021954.3	13q12.11	(Churchill and Graw, 2011; Mackay, et al., 1999)
GJA8	NM_005267.4	1q21.2	(Churchill and Graw, 2011; Shiels, et al., 1998)
HSF4	NM_012186.2	16q22.1	(Churchill and Graw, 2011)
LIM2	NM_005267.4	19q13.41	(Churchill and Graw, 2011; Pras, et al., 2002)
MAF	NM_005360.4.	16q23.2	(Churchill and Graw, 2011; Jamieson, et al., 2002)
MIP	NM_012064.3	12q13.3	(Berry, et al., 2000; Churchill and Graw, 2011)
MIR184	NR_038997.1	Chr15	(Hughes, et al., 2011)
NHS	NM_001081052.1	Xp22.13	(Burdon, et al., 2003; Churchill and Graw, 2011)
NRCAM	NM_001193582.1	7q31.1	(Churchill and Graw, 2011; More, et al., 2001)
NSDHL	NM_015922.2	Xq28	(Churchill and Graw, 2011; Traupe and Has, 2000)
PAX6	NM_000280.4	11p13	(Churchill and Graw, 2011; Glaser, et al., 1994)
PCBD1	NM_000281.3	10q22.1	(Bayle, et al., 2002; Churchill and Graw, 2011)

PITX2	NM_011098.3	4q25	(Reis, et al., 2012)
PITX3	NM_005029.3	10q24.32	(Churchill and Graw, 2011; Semina, et al., 1998)
PVRL3	NM_001243288.1	3q13.13	(Churchill and Graw, 2011; Lachke, et al., 2012)
SIX5	NM_175875.4	19q13.32	(Churchill and Graw, 2011)
SORD	NM_003104.5	15q21.1	(Churchill and Graw, 2011; Vaca, et al., 1982)
SOX1	NM_005986.2	13q34	(Churchill and Graw, 2011; Nishiguchi, et al., 1998)
SOX2	NM_003106.3	3q26.33	(Churchill and Graw, 2011)
SPARC	NM_003118.3	5q33.1	(Churchill and Graw, 2011)
TDRD7	NM_014290.2	9q22.33	(Lachke, et al., 2011)
TMEM114	NM_001146336.1	16p13.2	(Jamieson, et al., 2007)
VIM	NM_203472.1	10p13	(Churchill and Graw, 2011; Muller, et al., 2009)
VSX2	NM_182894.2	14q24.3	(Narooie-Nejad, et al., 2009)

4.3.2 Primer design for targeted massively parallel sequencing

PCR primers to amplify coding, 3' and 5' untranslated regions of the 51 genes were designed with the Ion Ampliseq Designer tool v1.22 (Life Technologies, www.ampliseq.com). The final design consisted of a total of 1216 amplicons ranging from 125 bp to 225 bp, covering 94.26% of the target sequence. Primers were supplied in two 100 nM pools (Life Technologies, Carlsbad, CA, USA).

4.3.3 Library preparation

The concentration of genomic DNA of probands was determined using the dsDNA HS Assay Kit on a Qubit fluorometer (Life Technologies) and libraries prepared with the Ion AmpliSeq library kit version 2.0 according to manufacturer's protocols. Libraries

were prepared in 2 pools per individual and the amplified pools were combined before barcode adaptor ligation.

Libraries were quantified either with a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) using the High Sensitivity DNA Kit (Agilent Technologies) or qPCR using Ion library TaqMan quantification kit (Life Technologies) according to the manufacturers' protocols.

The amplified library was diluted to 10 pM, and 25 μL of the diluted library was used for template preparation using Ion PGM Template OT2 200 Kit and the manufacturer's protocol. The clonally amplified library was then enriched on an Ion OneTouch enrichment system. Samples were barcoded during library preparation using Ion Xpress barcode Adapters 1-16 kit (Life Technologies) and pooled in groups of 3-5 during template preparation on the Ion OneTouch.

4.3.4 Sequencing and analysis

Sequencing was performed on an Ion Torrent Personal Genome Machine using the Ion PGM Sequencing 200 Kit v2 and an Ion 318 chip (Life Technologies).

Torrent Suite (version 3.6) was used to align reads to the human genome reference sequence 19 (hg19). The Coverage Analysis plugin (v4.0-r77897) was used to calculate the number of mapped reads, the percentage of on-target reads, and the mean depth of reads. Variants were called using the Variant caller plugin (V4.0-r76860) with the germline algorithm (allele frequency of 0.15, minimum read quality of 10, and minimum coverage of 20 were set as cut-offs for both indels and single nucleotide polymorphisms (SNPs)). For annotation, variant call format (VCF) files were uploaded to Ion Reporter V4.0 (https://ionreporter.lifetechnologies.com/ir/) using Ion Reporter Uploader plugin for Torrent Suite (v4.1-r79929). In addition to filtering criteria

mentioned in section 2.8 of Chapter 2, identified variants were compared with an in-house list of common sequencing errors previously detected with this gene panel.

4.3.5 Validation, segregation analysis and evaluating potential functional effects of mutations

Integrative Genomics Viewer (IGV) was used to visualize the detected variants in the sequence data (https://www.broadinstitute.org/igv/). Primer design and Sanger sequencing were performed as described in sections 2.5 and 2.6 of Chapter 2 to confirm the detected mutations in the probands and to evaluate the segregation of the mutation in respective families. Forward and reverse primer sequences are listed in Appendices10-12 of this thesis.

The filtering strategy and segregation analysis of novel and previously reported variants were explained in section 2.8 of Chapter 2. The variants in families CSA117 and W1 were validated using sequenom MassArray platform and iPlEX chemistry due to technical difficulties in primer design for Sanger sequencing (see section 2.7 of Chapter 2).

4.3.6 Screening novel variants in population controls

The majority of novel variants were screened in at least 326 (and up to330) unrelated, unaffected Australian controls (Chapter 2, section 2.3). This control screening was carried out with the MassArray platform and iPlEX chemistry (Sequenom, USA). Variants identified in families CRCH139, CSA133 and CSA95 were screened in controls using custom TaqMan SNP genotyping assays (Life Technologies) as described in Chapter 2 section 2.7. The variants detected in Asian samples were screened in our control samples despite the fact they are not an ethnically matched population controls (no access to the reasonable number of Asian controls). As an alternative, the ExAC database was used to assess the frequency of these observed

mutations. A total number of 4327 sample from East Asian and 8256 samples from South Asian countries (total of 12583 samples) were included in ExAC database which makes it a reliable to use for this purpose.

The variant detected in CSA172 was only assessed in public databases and not in local controls because it was not feasible to design TaqMan or Sequenom assay.

4.3.7 Haplotype analysis of families CSA110 and CSA91 with a mutation in *CRYAA*

Haplotype analysis was performed in a family previously reported to carry the same missense mutation (CSA91) (Laurie, et al., 2013) and the new family reported here (CSA110). Three microsatellite markers (D21S1260, D21S1890 and D21S1912) on the long arm of chromosome 21 flanking the *CRYAA* gene were typed in all available members of the two families. Primer sequences for amplification of each marker are provided in Appendix 13. Forward primers were labelled on the 5' end with fluorescent dye FAM or HEX. PCR reactions of 20µl final volume consisting of 1X Coraload PCR buffer (Qiagen) which gave a final concentration of 1.5 mM Mg^{2+} , 0.1mM dNTPs (Roche Diagnostics), 0.5 µM each primer, 0.5U Hot Star Plus Taq Polymerase (Qiagen) and 80 ng of DNA. Q Solution (Qiagen) was included at a final concentration of 1X for markers D21S1890 and D21S1912. PCR was performed on a Palm Cycler (Corbett Life science, QIAGEN) with 1 cycle of 95°C for 5 minutes, followed by 35 cycles of 95°C for 30 seconds, 57°C for 30 seconds, 72°C for 30 seconds and a final extension step of 72°C for 15 minutes. PCR products were each diluted 1:15. Fragment analysis was carried out on a 3130xl Genetic Analyser (Life Technologies, NY, USA) using GS500LIZ_3130 size standard at the Flinders Sequencing Facility (Flinders Medical Centre, Adelaide, South Australia)**.** Fragment sizes were determined by comparing to

the size standard using Peak Scanner Software Version 1 (Life Technologies). Haplotypes were reconstructed using Merlin (Abecasis, et al., 2002).

4.3.8 Assessment of sequence conservation

The Mutation Taster tool (http://www.mutationtaster.org/) was used for protein sequence alignment and assessing amino acid conservation (Schwarz, et al., 2014) and assessment of nucleotide conservation by PhyloP. PhyloP values between -14 and +6 indicates conservation at individual nucleotides, ignoring the effects of their neighbours.

4.4 Results

We sequenced 51 known paediatric cataract genes in 65 unrelated Australian probands with familial history or sporadic paediatric cataract using Ion Torrent MPS technology (35 familial cases and 30 sporadic). Furthermore, we screened 33 probands with paediatric cataract from three Asian countries (Bhutan, Cambodia and Sri Lanka). Syndromic cataract was present in 6/65 (9%, five were familial and one was a sporadic case) Australian probands (syndromic features described in Table 4.2) while 59/65 (91%) had isolated paediatric cataract.

Table 4.2. Features observed in Australian probands with syndromic cataract

Proband	**Age at diagnosis**	**Systemic features**
CSA119.01	-	Retinitis Pigmentosa[1], coarctation of aorta[2]
CSA132.03	< 1Y	Hypospadias[3], vitamin D deficiency, behavioural issues
CSA128.01	8	Seizures, lower limb weakness episode
CSA161.01	5	Cleft palate, short stature
CSA158.01	5	Mild left hemiplegia[4], spastic diplegia[5], mild expressive language delay
CSA108	< 1Y	Aymé-Gripp Syndrome, nuclear and posterior polar congenital cataract, mild to moderate sensorineural hearing loss, autism spectrum disorder, borderline intellectual abilities, flat midface,

[1]Retinis pigmentosa is a group of genetic disorders that affect the retina and its respond to the light. [2]Coarctation of aorta is a congenital heart defect. [3]Hypospadias a birth defect of the urethra in the male. [4]Hemiparesis is weakness of one side of the body. [5]Spastic diplegia is a form of cerebral palsy which is neurological condition that usually appears in early childhood.

In total, 154.1 kb of target sequence was included in the design process with amplicons designed for 94.3% of the target sequence (8.8 kb not covered). The presence of repetitive sequence, unacceptable GC content and melting temperatures of the primers outside the optimal range were the main factors limiting primer design for the uncovered target regions.

The mean number of mapped reads per sample was 1435237, with 88% of reads on target. A mean of 1049 reads was achieved per amplicon, with a coverage uniformity of

88%. Of all the amplicons, 95% and 90% were covered at least 20 and 100 fold, respectively. The average coverage per gene is shown in Figure 4.1. The least covered gene in our panel was *SIX5* with > 400 fold coverage where as *MIR184* had the highest coverage of >1800 fold. Of the 1216 amplicons, 28 amplicons (2%) across 16 genes were covered less than 20 fold (Figure 4.2).

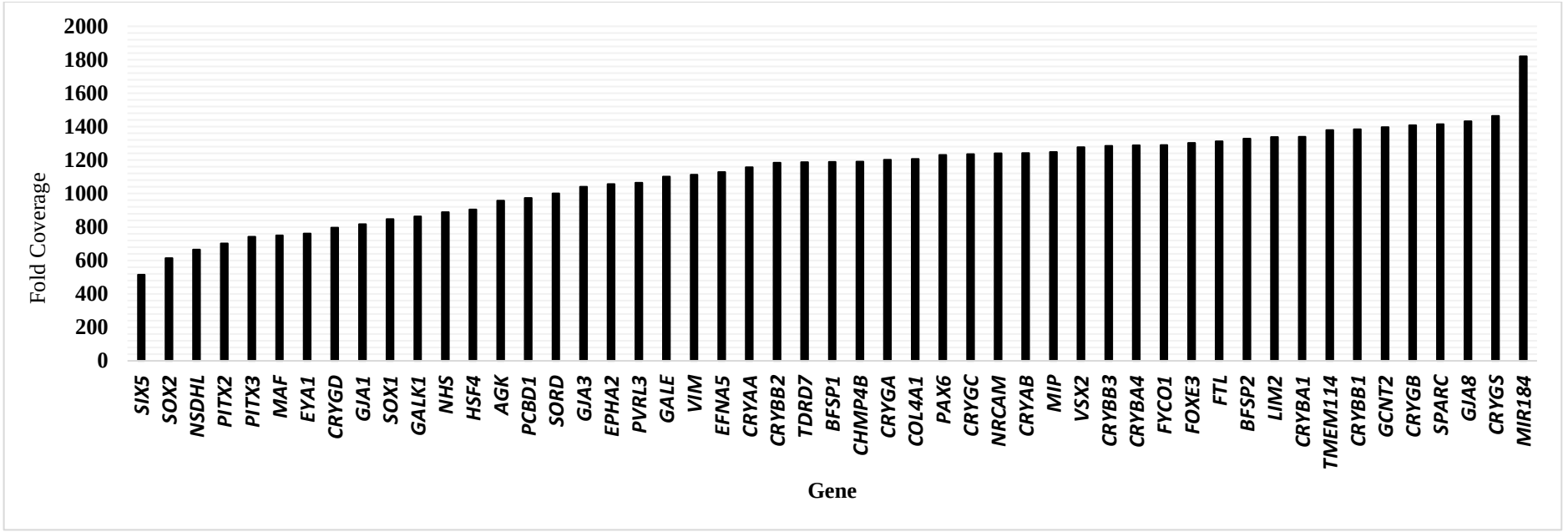


Figure 4.1. Average fold coverage of target genes sequenced from Ampliseq libraries in 98 paediatric cataract probands from Australia and South East Asia.

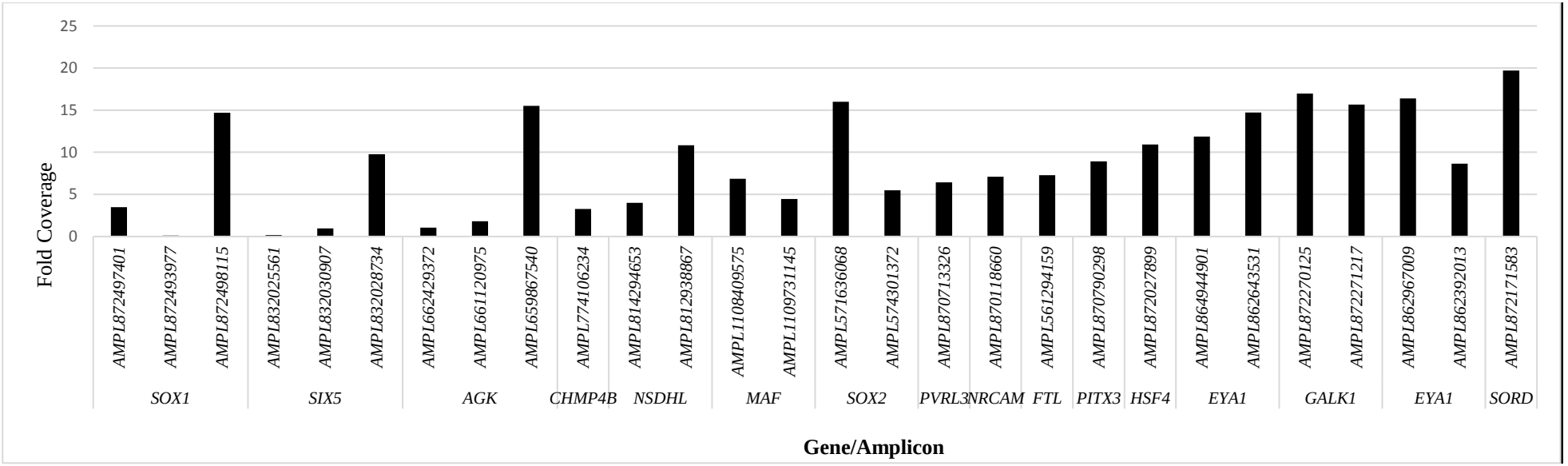


Figure 4.2. Amplicons (names are given by Ampliseq designer) with less than 20 fold coverage. The bars show the average amplicon coverage of 98 Australian and Asian individuals with paediatric cataract screened for 51 known paediatric cataract genes.

4.4.1 Screening Australian paediatric cataract samples with family history of paediatric cataract

A total of 4835 variants were annotated, giving an average of 138 variants per individual. In total, 188 variants were absent/rare (Minor Allele Frequency <1%) in publically referenced databases. Of these, 63 were non-synonymous exonic variants. Twenty-five variants were selected for validation using Sanger sequencing after filtering out the variants in an in-house list of sequencing artefacts (38 variants were on the in-house list of artefacts). Nineteen variants validated and 6 were found to be false positives.

Of the 19 validated variants (Table 4.3), 16 appeared to be the likely cause of cataract in the respective families, accounting for 45% of the 35 screened probands. The validated mutations were considered to be causative of the disease if the protein change was predicted to be pathogenic by SIFT (Kumar, et al., 2009) and/or Polyphen-2 (Adzhubei, et al., 2010), the variant was segregating with the phenotype in the family, and was absent from screened local controls. Two of the 19 validated variants did not segregate with the phenotype, and one was considered benign by both SIFT and Polyphen-2. All segregating mutations were highly conserved across species (see Table 4.3 and Appendix 14).

In total, as shown in Table 4.3, we detected 11 novel mutations in 9 different genes (*GJA3, GJA8, CRYAA, CRYBB2, CRYGS, CRYGA, GCNT2, MIP* and *MAF*), 4 previously reported known cataract causing mutations in 3 different genes (*CRYAA, CRYBB2* and *GJA8*) and 3 rare (which are not previously reported to be associated with this condition and present in

public databases with MAF<1%) potentially pathogenic variants in *CRYGA*, *PVRL3* and *BFSP2*. The phenotype in each of the 16 families is given in Table 4.5 and representative clinical photos where available are shown in Figure 4.3. Protein sequence alignments demonstrating the conservation of the altered amino acid detected in these families are given in Appendix 14, A, B and C.

Table 4.3. List of mutations detected in Australian families with Paediatric cataract.

Family ID	Minor Allele Frequencies in ExAC	Novel/Known	Gene	Chr. Position	Nucleotide change	Protein change	Phylop score	Polyphen-2 prediction	SIFT prediction	Inheritance/ Segregation/ penetrance
CSA95	0	Novel	*GJA3*	chr13:20717372	c.56C>T	p.(Thr19Met)	6.141	Probably Damaging	Deleterious	AR/Yes/ Full
CSA109	0	Novel	*GJA3*	chr13:20716962	c.466A>C	p.(Lys156Gln)	3.268	Probably Damaging	Deleterious	AD/Yes/ Full
CRCH20	0	Novel	*GJA8*	chr1:147380155	c.73T>C	p.(Trp25Arg)	4.833	Probably Damaging	Deleterious	AD/Yes/ Incomplete
CSA125	0	Novel	*GJA8*	chr1:147380566	c.484G>A	p.(Glu162Lys)	5.784	Probably Damaging	Deleterious	AD/Yes/ Full
CSA162	0	Known	*GJA8*	chr1:147380216	c.134G>C	p.(Trp45Ser)	5.786	Probably Damaging	Deleterious	AD/Yes/ Full
CSA159	0	Novel	*CRYAA*	chr21:44592307	c.440delA	p.(Gln147 Argfs*48)	0.333	NA	NA	AD/Yes/ Full
CRVEEH111	0	Known	*CRYAA*	chr21:44589369	c.160C>T	p.(Arg54Cys)	4.982	Probably Damaging	Tolerated	AD/Yes/ Full
CSA110	1.648×10^{-5}	Known	*CRYAA*	chr21:44589271	c.62G>A	p.(Arg21Gln)	5.35	probably damaging		AD/Yes/ incomplete
CSA94	0	Novel	*CRYGS*	chr3:186257377-78	c.30CT>AA	p.(Phe10Tyr11delinsLeuAsn)	1.266, 4.858	Probably Damaging	Deleterious	AD/Yes/ Full
CRCH139	4.28×10^{-3}	Rare	*CRYGA*	chr2:209027941	c.239G>A	p.(Arg80His)	0.799	Probably Damaging	Deleterious	AD/Yes/ Incomplete
	1.816×10^{-3}	Rare	*PVRL3*	chr3:110841054	c.886A>C	p.(Asn296His)	4.027	Probably damaging	Deleterious	AD/No
CSA133	0	Known	*CRYBB2*	chr22:25627584	c.463C>T	p.(Gln155*)	5.23	NA	Deleterious	AD/Yes/ Full
CRVEEH85	0	Novel	*CRYBB2*	chr22:25627684	c.563G>T	p.(Arg188Leu)	5.11	Probably Damaging	Deleterious	AD/Yes/ Full
	8.489×10^{-6}	Novel	*BFSP2*	chr3:133191301	c.1136C>A	p.(Ala379Glu)	0.366	Benign	Tolerated	AD/Yes/ Full

Table 4.3 continued

Family ID	Minor Allele Frequencies in ExAC	Novel/Known	Gene	Chr Position	Nucleotide change	Protein change	Phylop score	Polyphen-2 prediction	SIFT prediction	Segregation/ penetrance/ Inheritance
CRCH136	0	Novel	*GCNT2*	chr6:10626796	c.1166_1169delTCAA	p.(Asn388Arg*20)	4.217, -1.272	NA	NA	AR/Yes/ heterozygous in cases
CRCH89	0	Novel	*GCNT2*	chr6:10626722	c.1091T>C	p.(Phe364S)	4.256	Possibly damaging	Deleterious	AR/Yes/ homozygous in cases
CSA131	0	Novel	*MIP*	chr12:56845225	c.631G>T	p.(Gly211*)	5.081	NA	NA	AD/Yes
	2.473×10^{-6}	Novel	*FYCO1*	chr3:46009288	c.1538G>A	p.(Arg513Gln)		Benign	Tolerated	AD/No
CSA108	0	Novel	*MAF*	chr16:79633624	c.176C>G	p.(Pro59Arg)	5.32	Probably damaging	Deleterious	AD/Yes/ Full

AD: autosomal dominant; AR: autosomal recessive. Genbank Accession numbers are shown in Table 4.1. "-"indicates no information is available.

Table 4.4. Observed phenotypes in families with segregating causative mutations identified in paediatric cataract associated genes.

Family	**Gene**	**Affected members**	**Phenotype**	**Age at diagnosis**	**Age at surgery right eye**	**Age at surgery left eye**
CSA95	*GJA3*	CSA95.01	-	< 1Y	< 1Y	< 1Y
		CSA95.02	-	20 Y	-	-
CSA109	*GJA3*	CSA109.01	Fetal nuclear	3 Y	-	-
		CSA109.02	Fetal nuclear / lamellar	5Y	16 Y	17 Y
CRCH20	*GJA8*	CRCH20.02	Bilateral congenital nuclear	-	35 Y	-
		CRCH20.07	Bilateral minor lens opacities	-	-	-
CSA125	*GJA8*	CSA125.01	Nuclear	10 Y	-	-
		CSA125.02	Posterior polar	-	6 Y	-
CSA162	*GJA8*	CSA162.01	-	-	-	-
		CSA162.02	-	-	-	-
CSA159	*CRYAA*	CSA159.01	Severe congenital	< 1Y	1 M	2 M
		CSA159.02	Nuclear & cortical, blue-dot component: mild	19 Y	25 Y	25 Y
		CSA159.04	Lamellar: mild	4 Y	NA	NA
CSA110	*CRYAA*	CSA110.01	Bilateral lamellar	7 Y	16 Y	16 Y
		CSA110.03	Bilateral paediatric	12 Y	46 Y	46 Y
		CSA110.04	Paediatric cataract		76 Y	76 Y
		CSA110.05	Bilateral mild lamellar with anterior cortical spokes	-	61 Y	61 Y
		CSA110.07	Severe bilateral cataracts	-	4 Y	4 Y
CRVEEH111	*CRYAA*	CRVEEH111.01	Bilateral	-	-	-
		CRVEEH111.04	Bilateral	-	17 M	17 M
		CRVEEH111.05	Central, Anterior polar rider faint nuclear opacity only	-	-	-
		CRVEEH111.06	central nuclear opacity	-	-	-
CSA94	*CRYGS*	CSA94.01	Lamellar cortical-nuclear clear	6 Y	6 Y	5 Y
		CSA94.02	Cortical	4 Y	6 Y	5 Y
		CSA94.03	Lamellar	2 Y	3 Y	4 Y
		CSA94.04	Lamellar	2 Y	5 Y	5 Y
CRCH139	*CRYGA*	CRCH139.02	-	-	-	-

Table 4.4 continued

CSA133	*CRYBB2*	CSA133.01	-	-	-	-
		CSA133.03	-	-	-	-
CRVEEH85	*CRYBB2*	CRVEEH85.01	-	-	-	-
		CRVEEH85.02	-	2-3 Y	3 Y	3 Y
		CRVEEH85.03	-	-	-	-
CRCH89	*GCNT2*	CRCH89.01	Bilateral	-	3 W	3 W
		CRCH89.02	Bilateral	-	1 Y	1 Y
		CRCH89.05	Bilateral	-	-	-
		CRCH89.07	Bilateral	-	-	-
*CRCH136	*GCNT2*	CRCH136.01	Bilateral dense central opacity	-	-	-
		CRCH136.02	Bilateral dense central opacity	-	-	-
CSA131	*MIP*	CSA131.01	White dots	20 Y	NA	NA
		CSA131.02	white dots	22 Y	NA	NA
		CSA131.04	Cortical and nuclear sclerotic multiple cortical dots as well anterior cortical spokes	45 Y	46 Y	46 Y
CSA108	*MAF*	CSA108.01	Aymé-Gripp Syndrome Nuclear and posterior polar Mild to moderate sensorineural hearing loss Autism spectrum disorder Borderline intellectual abilities Flat midface	< 1Y	4 M	
		CSA108.02	Aymé-Gripp Syndrome Mild learning disability Hearing impairment Widely spaced lower teeth Flat midface	-	40 Y	-

Missing data is indicated by" –". NA indicates the individual has not had surgery to date, Y = years, M = months, W = weeks. * One heterozygous deletion detected in affected members of this family with autosomal recessive inheritance pattern.

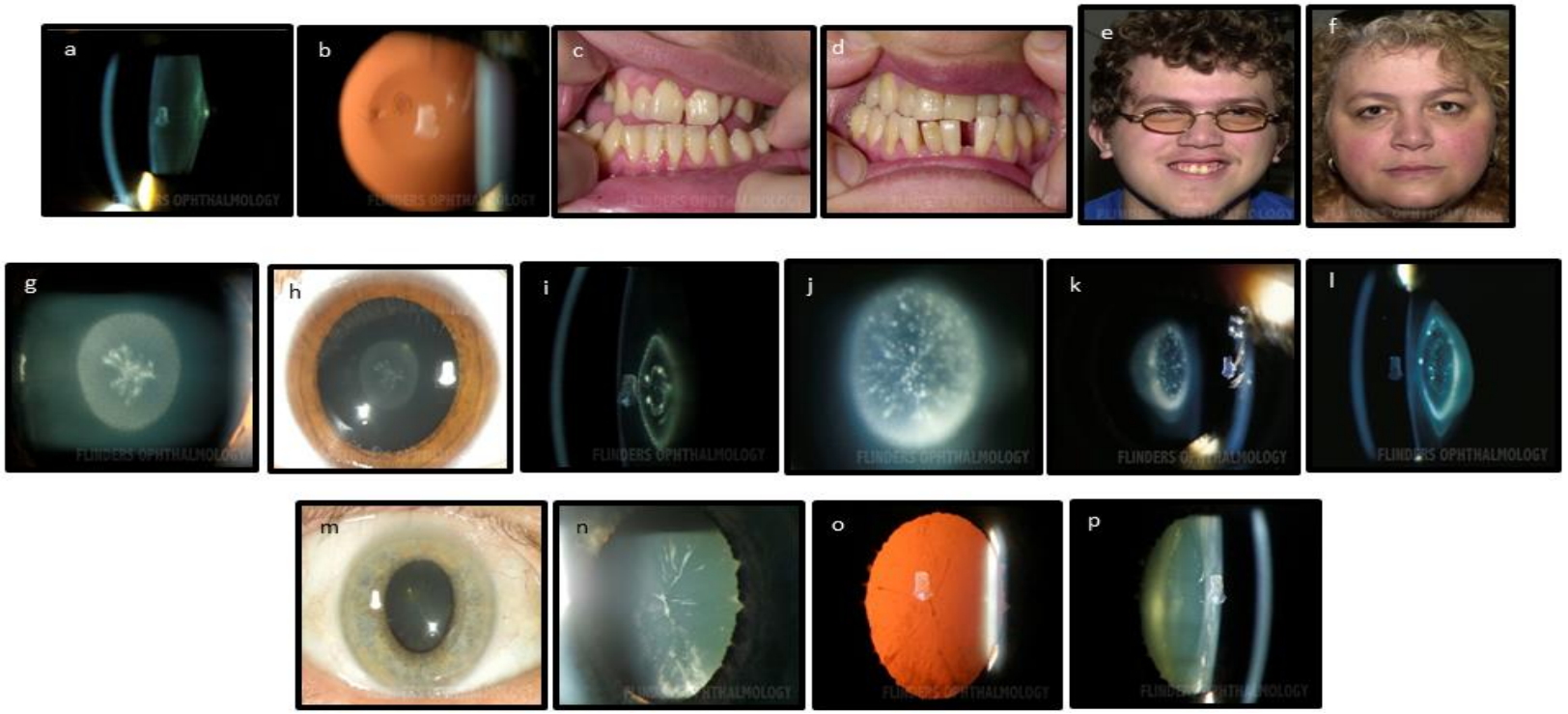
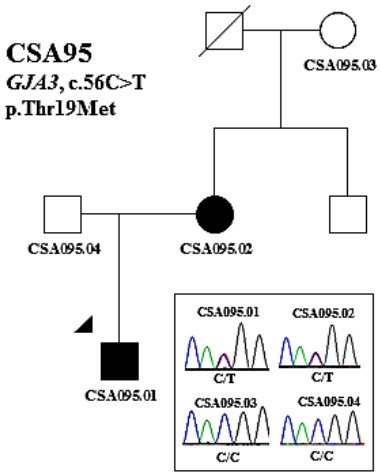

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p

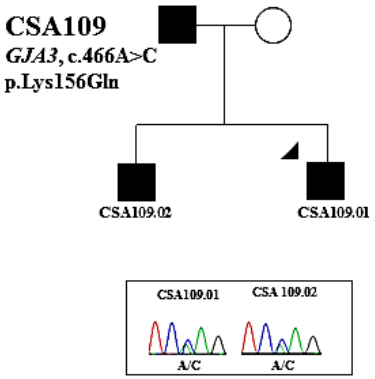
Figure 4.3. Phenotype of Paediatric cataracts in family CSA108, CSA109 and CSA110. Panels a-f show the phenotype of syndromic cataract in CSA108 with a novel mutation in *MAF*. Slit-lamp photographs (a and b) show posterior polar oil droplet cataract and posterior lenticonus. Dental abnormalities in CSA108.01 (c) and CSA108.02 (d). Facial features in CSA108.01 (e) and CSA108.02 (f). In particular, note flat mid-face in both, and short philtrum, long/narrow chin and upturned ear lobules in CSA108.01. Photographs of individual CSA109.01 (g, h and i) show fetal nuclear cataract. Panels j, k and l show fetal nuclear / lamellar cataract in individual CSA109.02. Panels m-p demonstrate the phenotype of cataract in CSA110.05 which is lamellar and cortical cataract with white spoke.

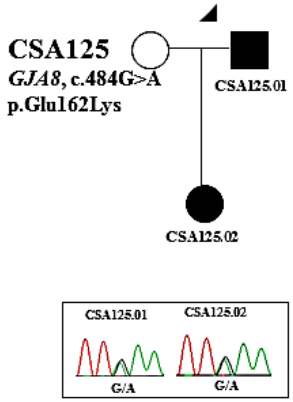
Mutations detected in gap junctions genes in Australian families with paediatric cataract

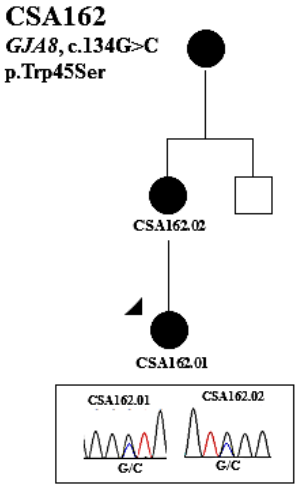
We identified novel mutations in two gap junction genes (*GJA3* and *GJA8*) in five families (Figure 4.4). Of the two families with mutations in *GJA3* (CSA95 and CSA109), phenotypic information was not available for family CSA95. However, the variant p.(Thr19Met) was predicted to be pathogenic and was present in two affected individuals (Figure 4.4). Both tested individuals in family CSA109 had fetal nuclear lamellar cataracts (Figure 4.3, Table 4.4) and the variant p.(Lys156Gln) was predicted to be damaging and segregated in 2 affected siblings (Figure 4.4). The affected father was not available for testing.

Two of the three families with *GJA8* mutations (CSA125 and CRCH20) have cataracts described as nuclear, with no information available for family CRCH162 (Table 4.4). In family CRCH20 the damaging mutation (p.(Trp25Arg)) segregated in two generations and appeared to have incomplete penetrance as individual CRCH20.04 carries the mutation but as yet does not have cataract. Mutations in families CSA162 (p.(Trp45Ser)) and CSA125 (p.(Glu162Lys)) were inherited from the affected mother and the affected father, respectively (Figure 4.4). The mutation in Family CSA162 also was reported in a recent study (Ma, et al., 2016) in a proband with sporadic paediatric cataract.


CSA95
GJA3, c.56C>T
p.Thr19Met
CSA095.03
CSA095.04
CSA095.02
CSA095.01
CSA095.01
CSA095.02
C/T
C/T
CSA095.03
CSA095.04
C/C
C/C


CSA109
GJA3, c.466A>C
p.Lys156Gln
CSA109.02
CSA109.01
CSA109.01
CSA 109.02
A/C
A/C


CSA125
GJA8, c.484G>A
p.Glu162Lys
CSA125.01
CSA125.02
CSA125.01
CSA125.02
G/A
G/A


CSA162
GJA8, c.134G>C
p.Trp45Ser
CSA162.02
CSA162.01
CSA162.01
CSA162.02
G/C
G/C

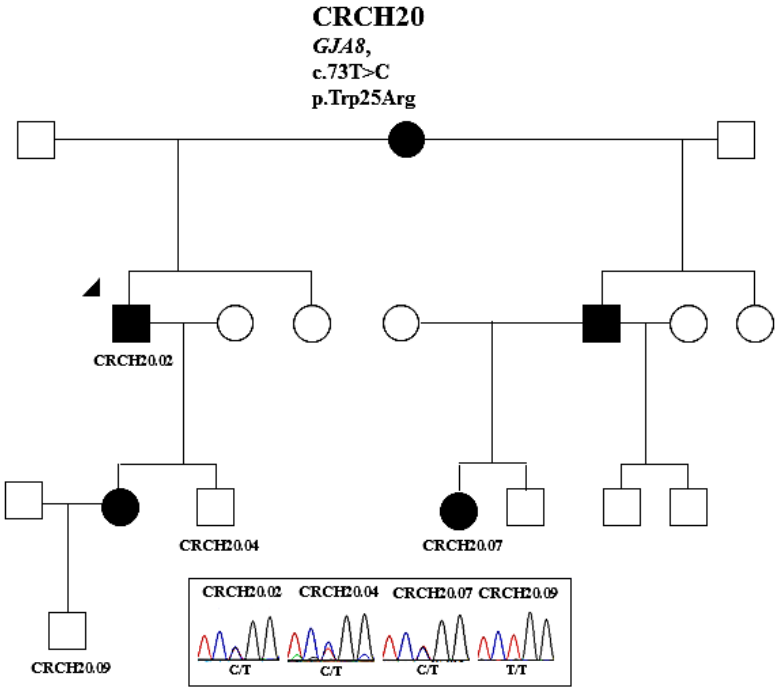

CRCH20
GJA8,
c.73T>C
p.Trp25Arg
CRCH20.02
CRCH20.04
CRCH20.07
CRCH20.09
CRCH20.02
CRCH20.04
CRCH20.07
CRCH20.09
C/T
C/T
C/T
T/T

Figure 4.4. Pedigree and Sanger sequencing analysis of families with variations in Gap junction genes (*GJA3* and *GJA8*). The chromatograms below each pedigree show the sequence detected via Sanger sequencing for each variant in families and the gene names and mutation at cDNA and protein level are shown on each pedigree. The penetrance of mutations in family CRCH20 (*GJA8*, c.73T>C) is incomplete. The arrows indicate the proband sequenced on the gene panel by Ampliseq. Solid circles indicate affected females and solid squares indicate affected males.

Mutations detected in crystallin genes in Australian families with paediatric cataract

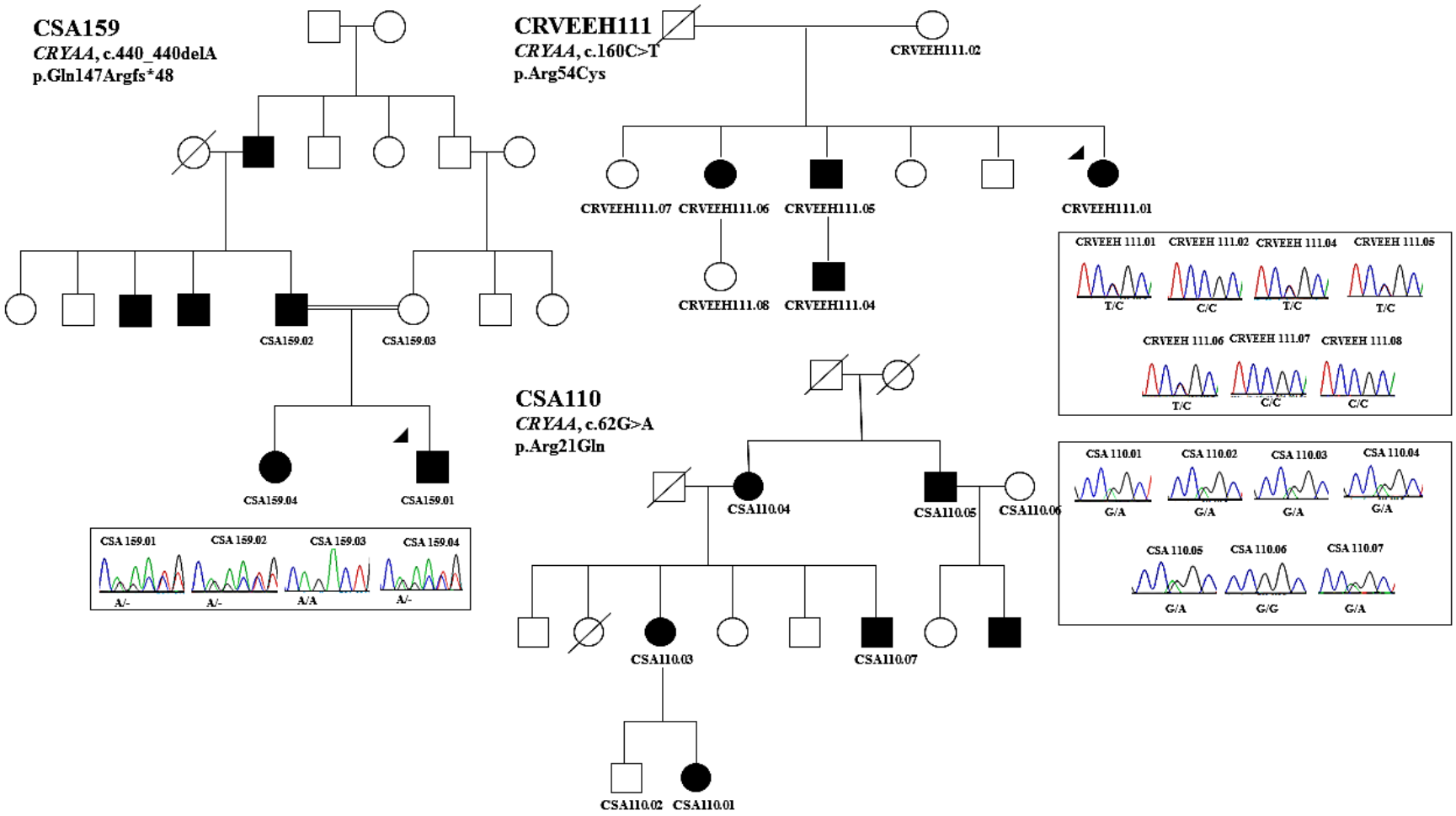
Seven mutations were identified in crystallin genes. A previously reported mutation in *CRYAA* (p.(Arg54Cys)) (Devi, et al., 2008; Khan, et al., 2007) was detected in family CRVEEH111 that segregated with the disease (Figure 4.5 A). SIFT predicted this variant to be tolerated, but Polyphen-2 predicted it to be pathogenic (Table 4.3). Family CRVEEH111 had central nuclear cataract with varying severity in affected family members. Another previously reported mutation (p.Arg21Gln) was detected in family CSA110. The mutation was previously reported to be associated with paediatric cataract in another family from our cohort (Laurie, et al., 2013) and is described in more detail in the next section. The third mutation detected in *CRYAA* is a novel frameshift deletion (p.(Gln147Argfs*48)) in a consanguineous family (CSA159) displaying autosomal dominant inheritance (Figure 4.5A). The father and both children carried the mutation; however the severity of the phenotype varied between affected members. The proband (CSA159.01) was diagnosed at birth and underwent cataract surgery at 1 month of age. His sister (CSA159.04) was diagnosed with a milder lamellar cataract with a similar appearance to that in the father.

One novel and one previously reported mutation were detected in *CRYBB2*. Three affected individuals from family CRVEEH85 carried a novel mutation (p.(Arg188Leu)). These individuals also carried a variant in *BFSP2* (p.(Ala379Glu)). This variant was reported to be benign by both SIFT and Polyphen-2 and the altered amino acid was less conserved, therefore considered most likely not to be pathogenic. A previously reported truncating mutation, Gln155* (rs74315489) in the *CRYBB2* gene was identified in two affected individuals in family CSA133. No information was available regarding the phenotype in this family, however this mutation was reported in numerous studies associated with a variety of phenotypes including Cerulean (Litt, et al., 1997), Central

zonular pulverulent (Gill, et al., 2000) and cortical (Devi, et al., 2008). This variant has not been reported in normal populations and was not detected in our local controls, thus is likely pathogenic.

Three mutations were detected in two different γ-crystallin genes. Family CSA94 (Figure 4.5 B) had a novel dinucleotide substitution (c.30_31CC>TT) resulting in the substitution of 2 amino acids (p.(Phe10_Tyr11delinsLeuAsn)) in *CRYGS* which was predicted to be damaging (Table 4.3). Affected members of this family had a juvenile onset cortical lamellar cataract and all the members had surgery by 6 years of age (Table 4.3). Family CRCH139 had a missense variant (p.(Arg80His), rs139353014) in *CRYGA* segregating with the phenotype in 3 individuals. This variant was predicted to be damaging by both SIFT and Polyphen-2 and the residue was conserved across species. However, it had a minor allele frequency (MAF) of 0.2% in dbSNP, 0.4% in the ExAC database and 0.1% in our Australian controls. The variant was also present in unaffected individual CRCH139.03 (Figure 4.5 B). The reduced penetrance is consistent with this variant being present in the population at lower frequency. A second variant was detected in this family in *PVRL3* however it did not segregate with the phenotype. Although it is not clear if the rare *CRYGA* variant is responsible for the disease in this family, it remains the best candidate mutation observed to date in family CRCH139.

A


CSA159
CRYAA, c.440_440delA
p.Gln147Argfs*48
CSA159.02
CSA159.03
CSA159.04
CSA159.01
CSA 159.01
CSA 159.02
CSA 159.03
CSA 159.04
A/-
A/-
A/A
A/-
CRVEEH111
CRYAA, c.160C>T
p.Arg54Cys
CRVEEH111.02
CRVEEH111.07
CRVEEH111.06
CRVEEH111.05
CRVEEH111.01
CRVEEH111.08
CRVEEH111.04
CRVEEH 111.01
CRVEEH 111.02
CRVEEH 111.04
CRVEEH 111.05
T/C
C/C
T/C
T/C
CRVEEH 111.06
CRVEEH 111.07
CRVEEH 111.08
T/C
C/C
C/C
CSA110
CRYAA, c.62G>A
p.Arg21Gln
CSA110.04
CSA110.05
CSA110.06
CSA110.03
CSA110.07
CSA110.02
CSA110.01
CSA 110.01
CSA 110.02
CSA 110.03
CSA 110.04
G/A
G/A
G/A
G/A
CSA 110.05
CSA 110.06
CSA 110.07
G/A
G/G
G/A

B

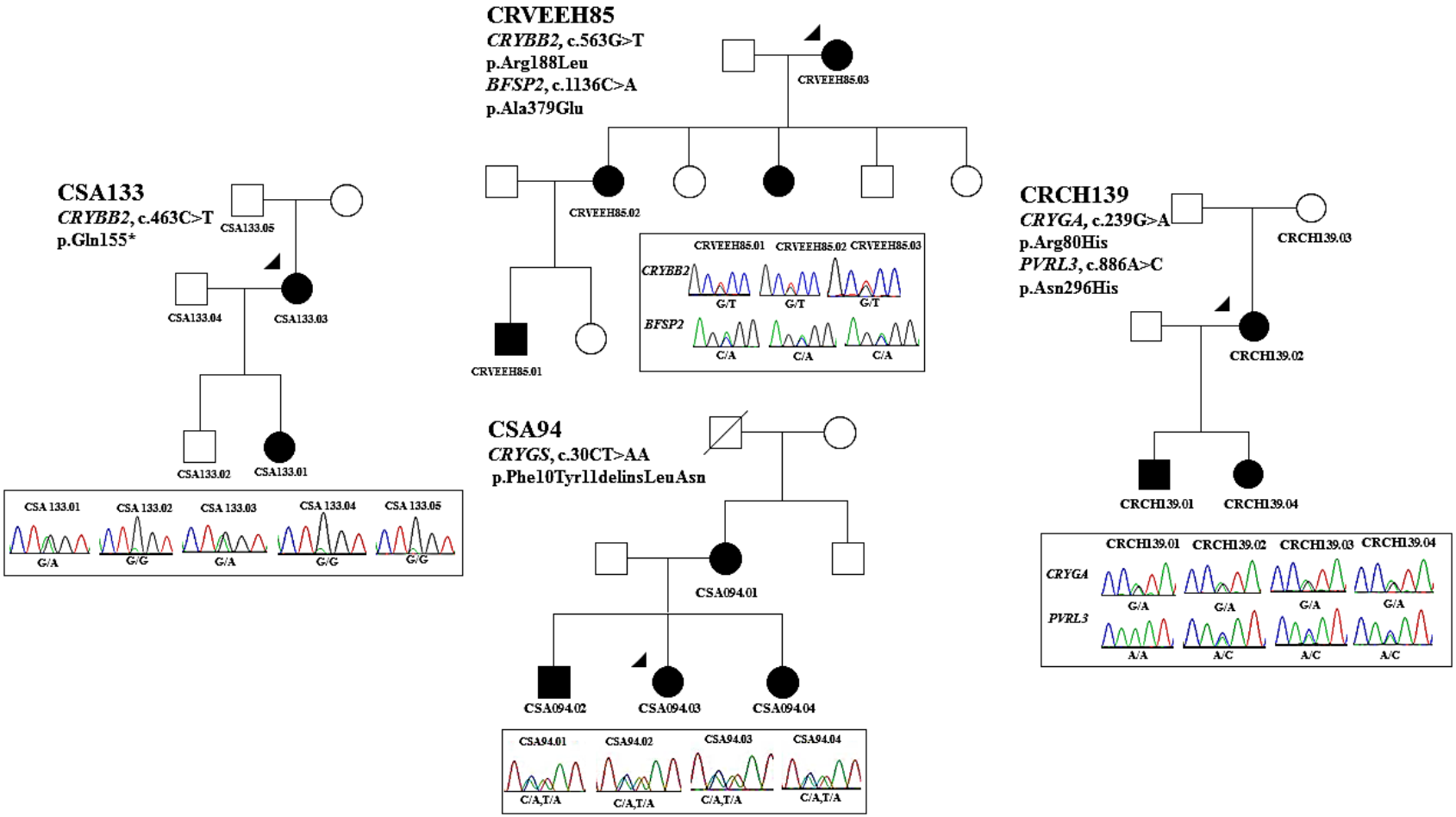

CSA133
CRYBB2, c.463C>T
p.Gln155*
CSA133.05
CSA133.04
CSA133.03
CSA133.02
CSA133.01
CSA 133.01
CSA 133.02
CSA 133.03
CSA 133.04
CSA 133.05
G/A
G/G
G/A
G/G
G/G
CRVEEH85
CRYBB2, c.563G>T
p.Arg188Leu
BFSP2, c.1136C>A
p.Ala379Glu
CRVEEH85.03
CRVEEH85.02
CRVEEH85.01
CRVEEH85.01
CRVEEH85.02
CRVEEH85.03
CRYBB2
G/T
G/T
G/T
BFSP2
C/A
C/A
C/A
CSA94
CRYGS, c.30CT>AA
p.Phe10Tyr11delinsLeuAsn
CSA094.01
CSA094.02
CSA094.03
CSA094.04
CSA94.01
CSA94.02
CSA94.03
CSA94.04
C/A,T/A
C/A,T/A
C/A,T/A
C/A,T/A
CRCH139
CRYGA, c.239G>A
p.Arg80His
PVRL3, c.886A>C
p.Asn296His
CRCH139.03
CRCH139.02
CRCH139.01
CRCH139.04
CRCH139.01
CRCH139.02
CRCH139.03
CRCH139.04
CRYGA
G/A
G/A
G/A
G/A
PVRL3
A/A
A/C
A/C
A/C

Figure 4.5. Pedigree and Sanger sequencing analysis of families with variations in A: Crystallin α (*CRYAA*); B: β and γ (*CRYBB2*, *CRYGA* and *CRYGS*). The penetrance of mutations in family CRCH139 (*CRYGA*, c.239G>A) is incomplete. The variants in *PVRL3* in CRCH139 do not segregate with the phenotype. The segregating variation in *BFSP2* in CRVEEH85 is predicted to be non-pathogenic by both SIFT and Polyphen-2. The arrows indicate the proband sequenced on the gene panel by Ampliseq. Solid circles indicate affected females and solid squares indicate affected males. Diagonal lines indicate the person is deceased. The chromatograms below each pedigree show the Sanger sequencing result of each detected variant in family members. The gene names and mutation at cDNA and protein level have been mentioned on each pedigree.

A recurrent mutation in *CRYAA* gene demonstrates a founder effect in South Australia

We identified a previously reported oligomerization disrupting mutation, c.62G>A (p.R21Q), in the *CRYAA* gene segregating in this three generation family of CSA110 (Javadiyan, et al., 2016). No other novel coding mutations were detected in known cataract genes.

Six members of the three generation Family CSA110 (Figure 4.5 A) were diagnosed with paediatric cataract (Table 4.3). The proband CSA110.01 presented with bilateral lamellar cataracts diagnosed at 7 years of age and had bilateral cataract surgery at age 16. Her mother, CSA110.03 also had bilateral cataracts which were diagnosed at 12 years of age and had bilateral cataract surgery at the age of 46. CSA110.07, the proband's maternal uncle, had severe bilateral cataracts which were surgically removed at age 4 years. The proband's grandmother, CSA110.04 had bilateral cataract surgery at the age of 76 but no information was available on the cataract phenotype. Her brother, CSA110.05, had bilateral mild lamellar cataracts, with anterior cortical spokes (Figure 4.3 panels m-p) and had cataract surgery at the age 61. These siblings had surgery significantly later than their children and grandchildren; however the cataract of CSA110.05 is distinctly congenital, despite the late intervention. The proband's brother, CSA110.02 shows no evidence of cataracts at the last examination in 2005 when he was 20 years old.

One affected member of the family, CSA110.03, (Figure 4.5 A), was sequenced for 51 congenital cataract genes. Interestingly, the detected mutation (c.62G>A, p.(Arg21Gln)) in *CRYAA* was previously shown to be a oligomerization disrupting mutation in another South Australian family independently recruited to our congenital cataract research program (family CSA91) (Laurie, et al., 2013). All the affected family members as well

as one unaffected member (CSA 110.02) carry the mutation as determined by Sanger sequencing (Figure 4.5 A).

Haplotype analysis around the *CRYAA* gene was conducted to determine if the mutation has arisen in both families independently, or if a founder effect may be responsible. A recombination event was detected in CSA91.05 between *CRYAA* and D21S2160. All of the mutation carriers in family CSA110 share a haplotype across this region of chromosome 21 with the affected family members of CSA91, indicating that the mutation has likely arisen only once on this genetic background (Figure 4.6 A and B). The haplotype carrying the mutant allele was identical in the two families, suggesting that the recurrence of the mutation was possibly due to a founder effect in South Australia.

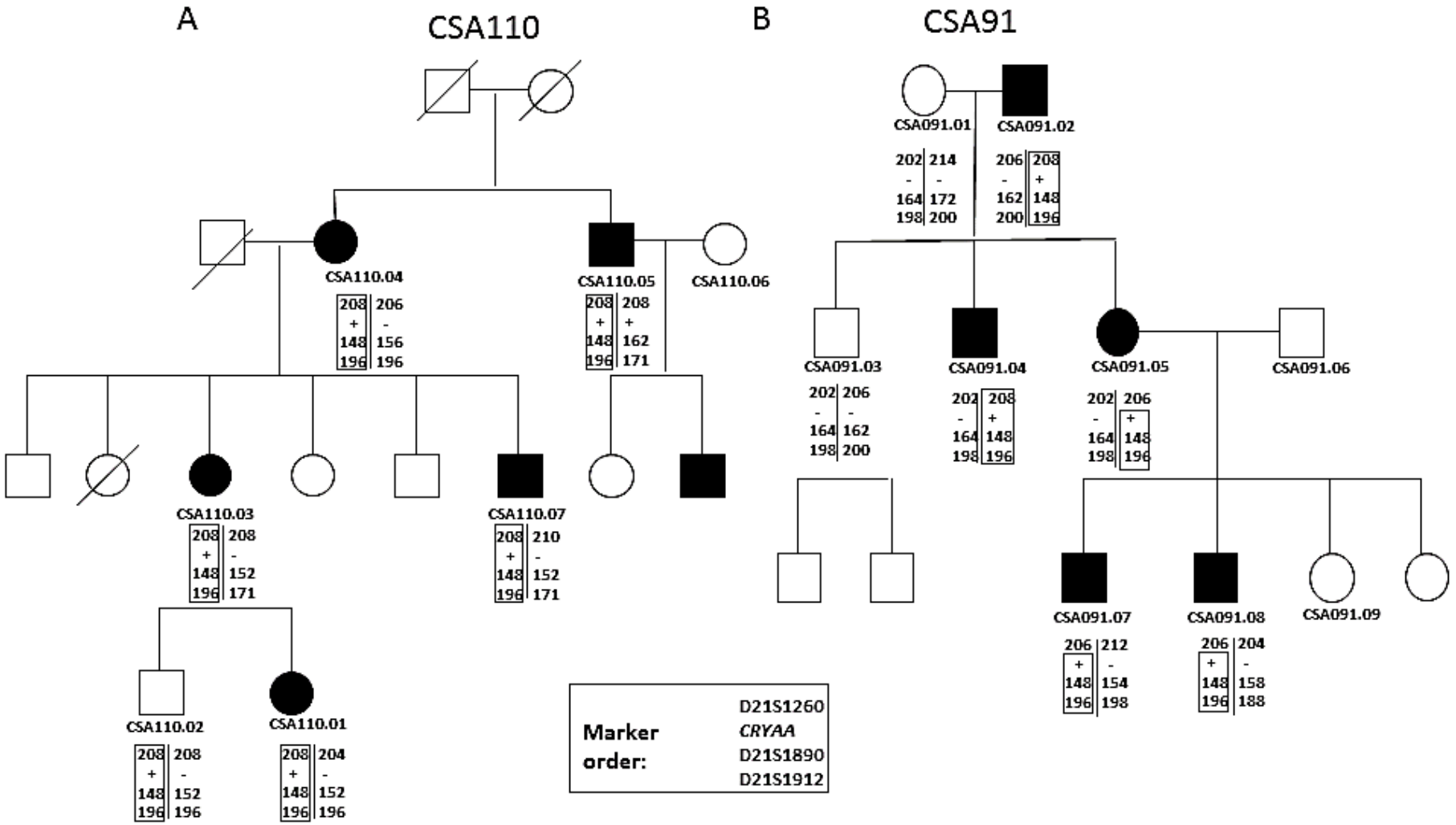
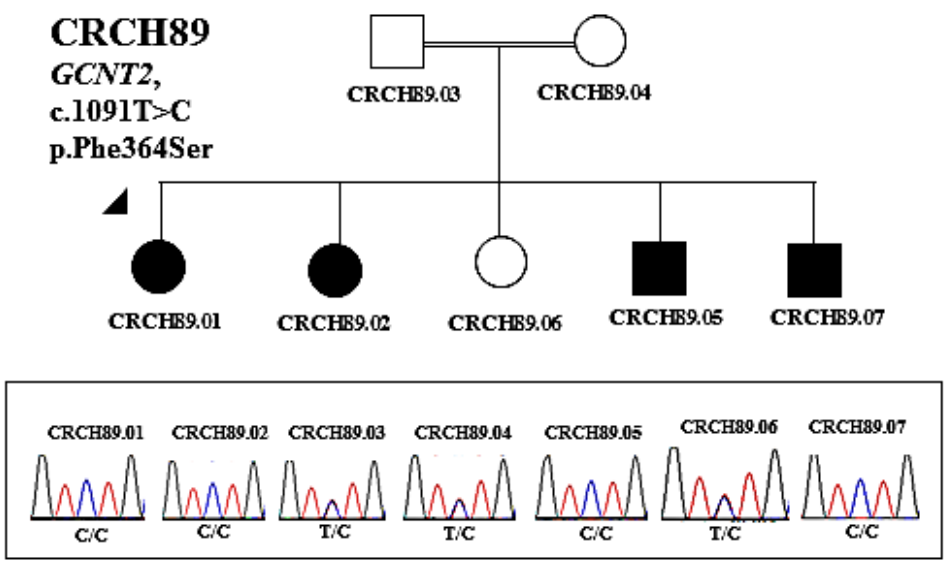

A
CSA110
CSA110.04
208 + 148 196 | 206 - 156 196
CSA110.05
208 + 148 196 | 208 + 162 171
CSA110.06
CSA110.03
208 + 148 196 | 208 - 152 171
CSA110.07
208 + 148 196 | 210 - 152 171
CSA110.02
208 + 148 196 | 208 - 152 196
CSA110.01
208 + 148 196 | 204 - 152 196
Marker order:
D21S1260
CRYAA
D21S1890
D21S1912
B
CSA91
CSA091.01
202 - 164 198 | 214 - 172 200
CSA091.02
206 - 162 200 | 208 + 148 196
CSA091.03
202 - 164 198 | 206 - 162 200
CSA091.04
202 - 164 198 | 208 + 148 196
CSA091.05
202 - 164 198 | 206 + 148 196
CSA091.06
CSA091.07
206 + 148 196 | 212 - 154 198
CSA091.08
206 + 148 196 | 204 - 158 188
CSA091.09

Figure 4.6. Haplotype analysis in 2 families with congenital cataract and the same mutation in *CRYAA* A: CSA91 (previously reported family) and B: CSA110 (reported in this study). Solid circles indicate affected females and solid squares indicate affected males. Only individuals with IDs were available for study. Marker and gene order is D21S1260, *CRYAA*, D21S1890 and D21S1912. Alleles at each marker are presented as the size of the PCR product detected. + indicates *CRYAA* mutation carrier. The segregating haplotype is boxed and is the same in both families. A recombination event between D21S1260 and *CRYAA* was observed in CSA91.05.

Mutations detected in *GCNT2* gene in Australian families with recessive paediatric cataract

Families CRCH89 and CRCH136 displayed an autosomal recessive inheritance pattern of cataract. Affected members of the consanguineous family, CRCH89, were homozygous for a novel variant (p.(Phe364Ser)) in *GCNT2* which is predicted to be pathogenic (Figure 4.7). The four affected siblings all had bilateral paediatric cataracts with surgery in the first few weeks to 1 year of age (Table 4.3). A single heterozygous variant in *GCNT2* resulting in a premature stop codon (p.(Asn388Argfs*20)) was detected in family CRCH136 which was inherited from the unaffected mother. No other variant was identified in *GCNT2* in this family.


CRCH89
GCNT2,
c.1091T>C
p.Phe364Ser
CRCH89.03
CRCH89.04
CRCH89.01
CRCH89.02
CRCH89.06
CRCH89.05
CRCH89.07
CRCH89.01
CRCH89.02
CRCH89.03
CRCH89.04
CRCH89.05
CRCH89.06
CRCH89.07
C/C
C/C
T/C
T/C
C/C
T/C
C/C

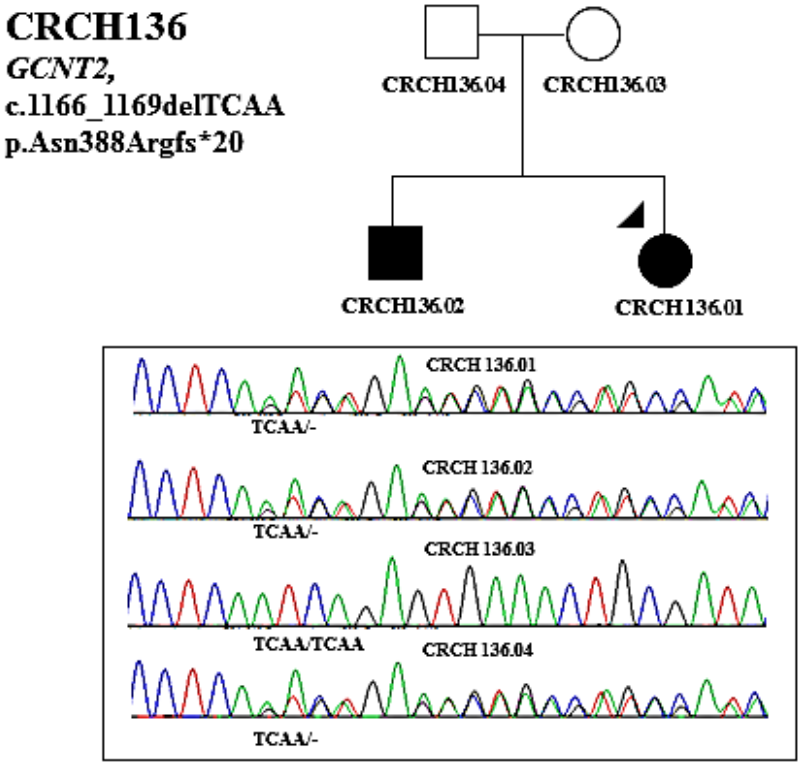

CRCH136
GCNT2,
c.1166_1169delTCAA
p.Asn388Argfs*20
CRCH136.04
CRCH136.03
CRCH136.02
CRCH136.01
CRCH 136.01
TCAA/-
CRCH 136.02
TCAA/-
CRCH 136.03
TCAA/TCAA
CRCH 136.04
TCAA/-

Figure 4.7. Pedigree and Sanger sequencing analysis of families with variations in *GCNT2*. The arrows indicate the proband sequenced on the gene panel by Ampliseq. Solid circles indicate affected females and solid squares shows the affected males. Double line in CRCH89 shows consanguinity. The chromatograms below each pedigree show the segregation analysis of the variants in families. The gene names and mutation at cDNA and protein level have been mentioned on each pedigree.

Mutations detected in membrane protein MIP in an Australian family with autosomal dominant paediatric cataract

A segregating stop mutation (p.(Gly211Ter)) in *MIP* was detected in family CSA131 (Figure 4.8). The affected individual (CSA131.04) had cortical and nuclear sclerotic cataracts with multiple cortical dots while two other affected members (CSA131.01 and CSA131.02) had anterior cortical spokes, and white dots like cataracts (Table 4.4). This family also carries a non-segregating variation in *FYCO1*; however this variant is predicted to be benign and does not segregate with the disease.

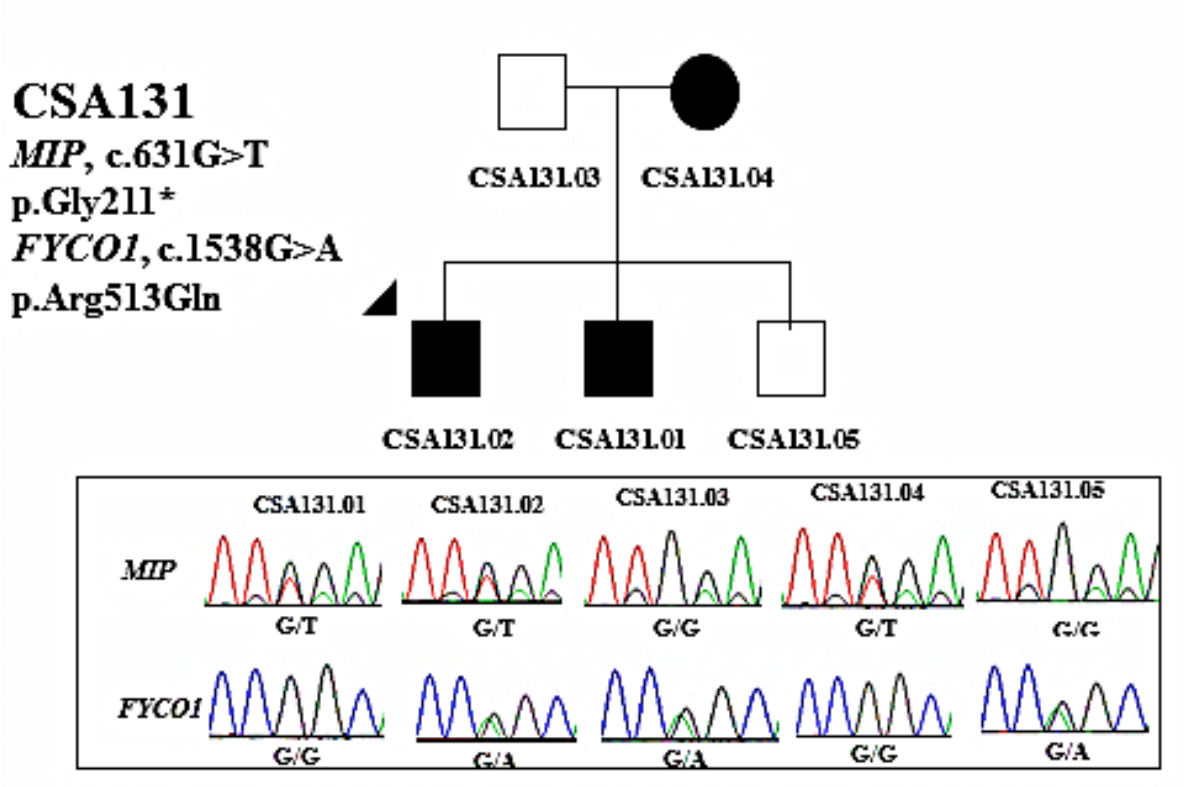


Figure 4.8. Pedigree and Sanger sequencing analysis of family CSA131 with variations in *MIP*. The arrows indicate the proband sequenced on the gene panel by Ampliseq. Solid circles indicate affected females and solid squares shows the affected males. The chromatograms below the pedigree show the segregation analysis of the variants in *MIP* and *FYCO1*. The gene names and mutation at DNA and protein level (hg19) have been mentioned on each pedigree.

A novel missense mutation in the bZIP transcription factor, *MAF*, is associated with paediatric cataract, developmental delay, seizures and hearing loss (Aymé-Gripp Syndrome)

One rare, non-synonymous variant which was predicted to be pathogenic was detected in family CSA108. This novel coding variant is a missense mutation in the *MAF* gene (c.176C>G, p.(Pro59Arg)) (Figure 4.9). It was predicted to be pathogenic by SIFT and Polyphen-2 and is in a highly conserved region of the protein (Appendix 14 C). The variant was also present in the affected mother and absent in 326 screened unrelated Caucasian controls.

The proband (CS108.01) was diagnosed with bilateral congenital cataract described as nuclear and posterior polar in the right eye, and milder posterior polar oil droplet cataract in the left eye at birth (Figure 4.3). Cataract in the right eye was removed at five months of age and the patient subsequently developed aphakic glaucoma. The proband also had mild to moderate sensorineural hearing loss (he did not appear to have a hearing impairment in early childhood). The proband was diagnosed with autism spectrum disorder and borderline intellectual abilities at childhood. His only seizures were at 13.5 years (two, two weeks apart).

The mother of the proband, CSA108.02, had cataract extraction at the age of 40. She also has a mild learning disability, hearing impairment, and widely spaced lower teeth (Figure 4.3). She has not been diagnosed to have an autism spectrum disorder, nor has she had seizures. Although the proband was not significantly dysmorphic, both he and his mother have a flat midface. CSA108.01 had a distinctive facial appearance with narrow posteriorly rotated ears with upturned ear lobules, downslanting palpebral fissures, flat mid-face, short philtrum, prominent narrow chin and dental maloccusion;

his mother also had mildly downslanting palpebral fissures, flat mid-face and relatively prominent chin (Figure 4.3).

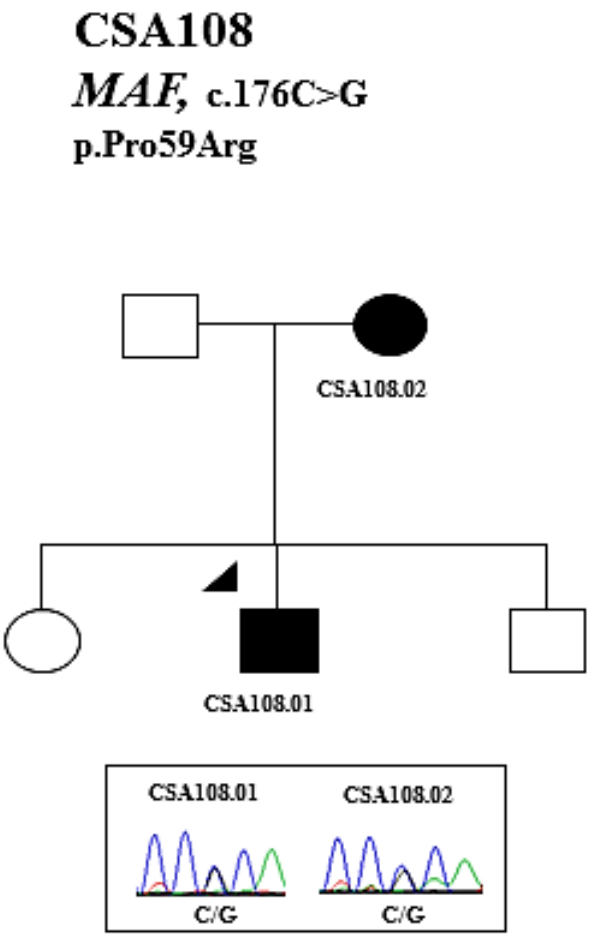


Figure 4.9. Pedigree of family CSA108 and Sequence chromatogram of two examined individuals at the mutation c.176C>G in *MAF*. Both sequenced affected members are heterozygous for this mutation. Solid circles indicate affected females and solid squares indicate affected males.

4.4.2 Screening sporadic paediatric cataract Australian cohort

We sequenced 30 probands with sporadic paediatric cataract. A total of 4127 variants were annotated with an average of 137 variants per individual). In total 144 variants were absent or rare (Minor Allele Frequency <1%) in public databases, of which 30 were non-synonymous exonic variants (Table 4.5). Sixteen variants (including the one predicted benign variant) were selected for validation using Sanger sequencing after filtering out common sequencing artefacts, and all have been validated by Sanger sequencing (Table 4.5). Protein sequence alignments demonstrating the conservation of the altered amino acid detected in these families are given in Appendix 15 A, B and C.

Table 4.5. List of mutations detected in Australian probands with sporadic paediatric cataract

Proband	Minor Allele Frequencies in ExAC	Novel/Known	Gene	Position in hg19	Nucleotide change	Protein change	Phylop score	Polyphen-2 prediction	SIFT prediction
CSA150	0	Novel	*GJA8*	chr1:147381190	c.1108G>T	p.(Glu370*)	3.82	-	-
CSA175	9.88×10^{-5}	Novel	*GJA8*	chr1:147380639	c.557G>A	p.(Trp186*)	5.78	-	-
*CSA172	0	Novel	*CRYGC*	chr2:208993028	c.423_424insG	p.(Arg142Alafs22)	0.9	-	-
CSA155	0	Novel	*CRYGS*	chr3:186257168-69	c.239_240delinsGT	p.(Leu80Arg)	1.2, 4.8	Probably damaging	Deleterious
CSA164	5.3×10^{-5}	Rare	*CRYGB*	chr2:209007382	c.508C>T	p.(Arg170*)	0.81	-	-
	0	Novel	*CRYAA*	chr21:44590718	c.281T>G	p.(Val94Gly)	4.35	Probably damaging	Tolerated
CSA154	0	Novel	*BFSP2*	chr3:133119349	c.422A>C	p.(Glu141Ala)	1.59	Probably damaging	Deleterious
CSA107	0	Novel	*BFSP2*	chr3:133167461	c.701T>C	p.(Leu234Pro)	4.5	Probably Damaging	Deleterious
CSA146	5.08×10^{-5}	Rare	*BFSP2*	chr3:133191276	c.1111G>A	p.(Glu371Lys)	5.46	Probably damaging	Deleterious
	0	Novel	*MIP*	chr12:56845224	c.632G>A	p.(Gly211Glu)	5.08	Probably damaging	Deleterious
JM	0	Novel	*VIM*	chr10:17272686	c.601A>T	p.(Asn201Tyrp)	2.09	Probably damaging	Deleterious
	0	Novel	*NHS*	chrX:17743566	c.1277_1278insTG	p.(Glu427Val*9)	-	-	-
CSA147	0	Novel	*NHS*	chrX:17394202	c.322G>T	p.(Glu108*)	0.99	-	-
CSA117	0	Novel	*MAF*	chr16:79632962	c.838G>A	p.(Gly280Arg)	5.3	Probably damaging	Deleterious
CSA163	1.95×10^{-4}	Rare	*HSF4*	chr16:67201032	c.636G>T	p.(Met212Ile)	3.73	Possibly Damaging	Tolerated
JP	0	Novel	*FYCO1*	chr3:46009379	c.1447T>C	p.(Trp483Arg)	0.13	Benign	Tolerated

Genbank Accession numbers are shown in Table 4.1. *The detected variant in this proband was only screened in ExAC and dbSNP 137. AD: autosomal dominant; AR: autosomal recessive. Genbank Accession numbers are shown in Table 4.1. "-"indicates no information was available.

Table 4.6. Phenotypic information of Australian probands with sporadic paediatric cataract and potential candidate causative gene identified.

Proband	**Gene**	**Phenotype**	**Age at diagnosis**	**Age at surgery right eye**	**Age at surgery left eye**
CSA150	*GJA8*	Nuclear cataract	25 Year	-	-
CSA175	*GJA8*	Lamellar, central	-	69 years	69 years
CSA172	*CRYGC*	-	Birth	1 month	-
CSA155	*CRYGS*	-	Birth	1 month	1 month
CSA164	*CRYGB* and *CRYAA*	Nuclear lamellar	2 years	No	no
CSA154	*BFSP2*	Lamellar	Birth	-	5 year
CSA107.01	*BFSP2*	Cortical dots, rather prominent posterior opacification in the sutural region	Birth	26	26
CSA146	*BFSP2* and *MIP*	-	-	2 year	-
JM	*VIM* and *NHS*	Dense central cataract, sever glaucoma, thick cornea, developmental delay	2 months	2 months	2 months?
CSA147	*NHS*	-	1 year	2	-
CSA163	*HSF4*	Posterior subcapsular congenital cataracts (PSCC)	49	49 year	49 years
CSA117	*MAF*	Lamellar	4 years	10 years	10 years

"-"indicates no information is available.

Connexin encoding genes

We identified 2 different nonsense mutations in *GJA8* in 2 probands with nuclear and lamellar paediatric cataract (CSA150 and CSA175). Both of the mutations were predicted to be pathogenic and are located within a protein conserved region (Appendix 15 A). CSA150 was diagnosed at the age of 25 but there was no information available regarding the age of surgery (Table 4.6). The nature of the cataract was identified as paediatric despite the late diagnosis. The surgery at the age of 65 demonstrates mild nature of cataract in CSA175.

Cytoskeleton encoding genes

We identified 2 mutations in these two probands, CSA154 and CSA107, (Figure 4.10 B) in the *BFSP2* gene which encodes a beaded filament structural protein, important in maintenance of cell volume and shape and protein homeostasis. The heterozygous mutation in CSA107.01 was inherited from the patient's unaffected father with no history of paediatric cataract. Both of the variants (p.(Glu141Ala) and p.(Leu234Pro)) detected in these 2 individuals were predicted to be pathogenic by both SIFT and Polyphen-2 (Table 4.5). The altered amino acid residues were located in a conserved region of the protein (Appendix 15 C). CSA154 was diagnosed with lamellar cataract at birth and had cataract surgery at the age of 4 (Table 4.6). The cataract phenotype was described as cortical dots, and rather prominent posterior opacification in the sutural region in CSA107 (Table 4.6, Figure 4.10 B). The cataract phenotype was less severe compared to CSA154 as CSA107 had a cataract surgery at the age of 26.

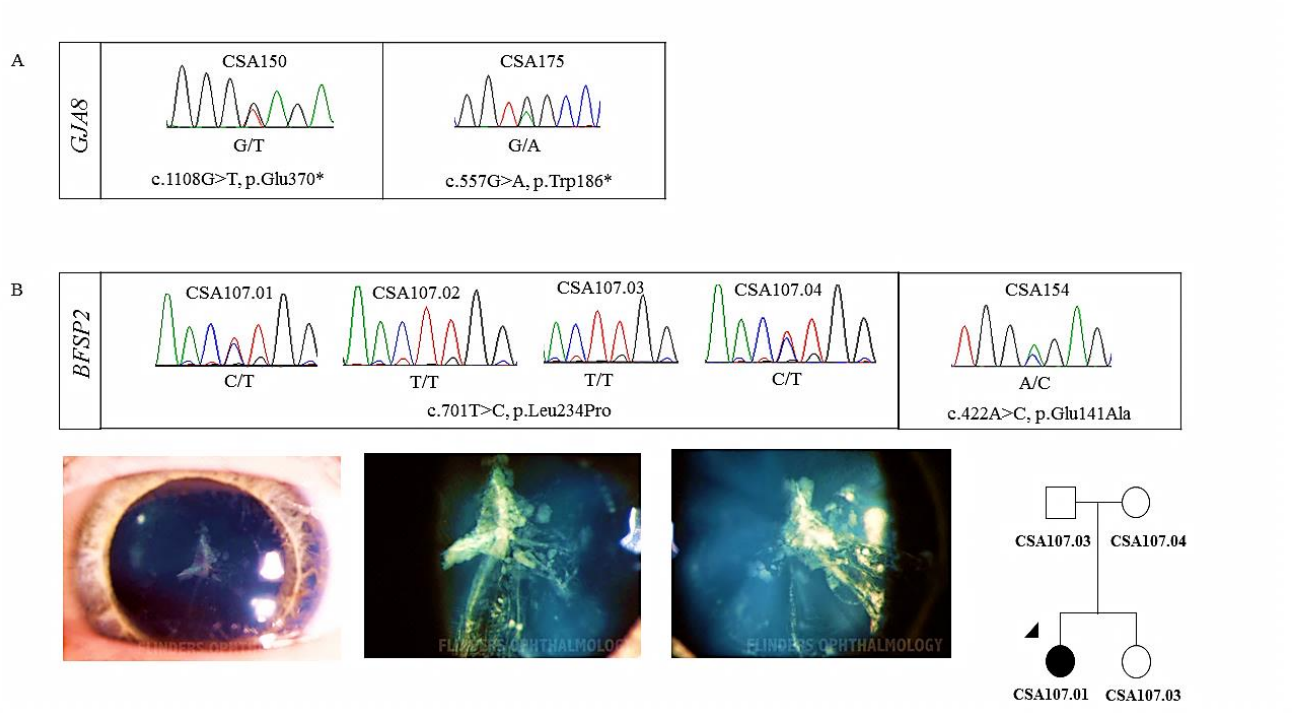


Figure 4.10. Chromatogram sequences of variations detected in cytoskeleton encoding genes. A: *GJA8* variations in CSA150 and CSA175. B: *BFSP2* variations in CSA107 and CSA154. Photographs of the cataract of individual CSA107.01 show a prominent cortical posterior opacification in the sutural region (B). The pedigree of family CSA107 is shown in section B.

Crystallin genes

We detected mutations in *CRYGC* and *CRYGS* in CSA172 and CSA155 respectively (Table 4.5, Figure 4.11). Both the heterozygous insertion in CSA172 (c.423_424insG) and dinucleotide substitution in CSA155 (c.239TC>GT) were predicted to alter the protein sequence (Table 4.5). The pathogenicity of the variant in CSA172 couldn't be evaluated with SIFT or Polyphen-2 which only evaluate missense mutations, however the insertion causes the formation of 12 amino acids shorter protein which potentially have high impact on its function. The altered amino acids residues are located within a conserved protein region (Appendix 15 B).

Both of CSA172 and CSA155 had a very severe cataract which were diagnosed at birth and underwent cataract surgery at 1 month of age (Table 4.6). There is no phenotypic information available regarding the type of the paediatric cataract in these two individuals.

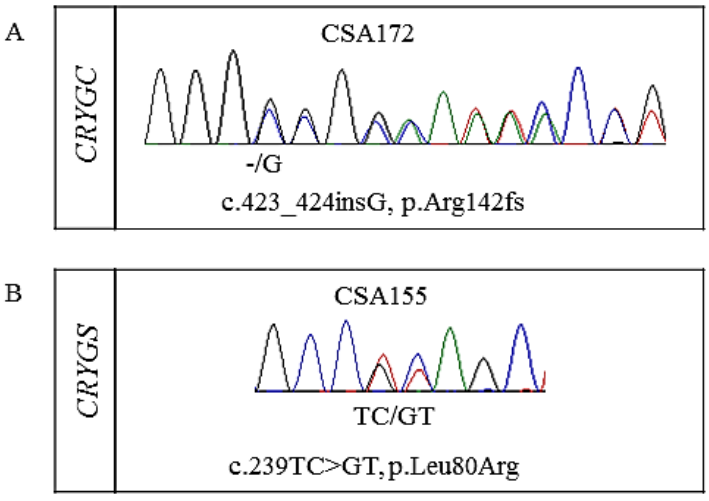


Figure 4.11. Chromatogram sequences of variations detected in *CRYGC* and *CRYGS* genes in A: CSA172 and B: CSA155.

Transcription factor, signalling molecules and syndromic genes

One variant detected in *HSF4* (a transcription factor essential in lens development and differentiation (Bu, et al., 2002) in CSA163 (Figure 4.12 A). The detected variant also was present in screened local controls with Minor Allele Frequency of 0.005 which is higher than has been reported in ExAC (0.001) database.

The detected variant in CSA117.01in *MAF* was confirmed in the patient using a Sequenom assay (in conjunction with screening of unaffected controls) due to difficulty in the primer design for Sanger sequencing. Therefore no sequence chromatogram is presented, however the presence of the variant can be seen in the Integrative Genomics Viewer (IGV) view from the Ion Torrent sequence data (Figure 4.12 B).

One variant was detected in CSA147.01 in *NHS*. The variant was not present in his mother, indicating a de *novo mutation* (Figure 4.12 C).

CSA163 was diagnosed with posterior subcapsular congenital cataracts phenotype at the age of 49 and underwent surgery at this age. CSA117.01 was diagnosed at the age of 4 with lamellar isolated paediatric and had cataract surgery at the age of 10 (Table 4.5 and Figure 4.12 B). There is no phenotypic information available on the type of paediatric cataract in family CSA147.

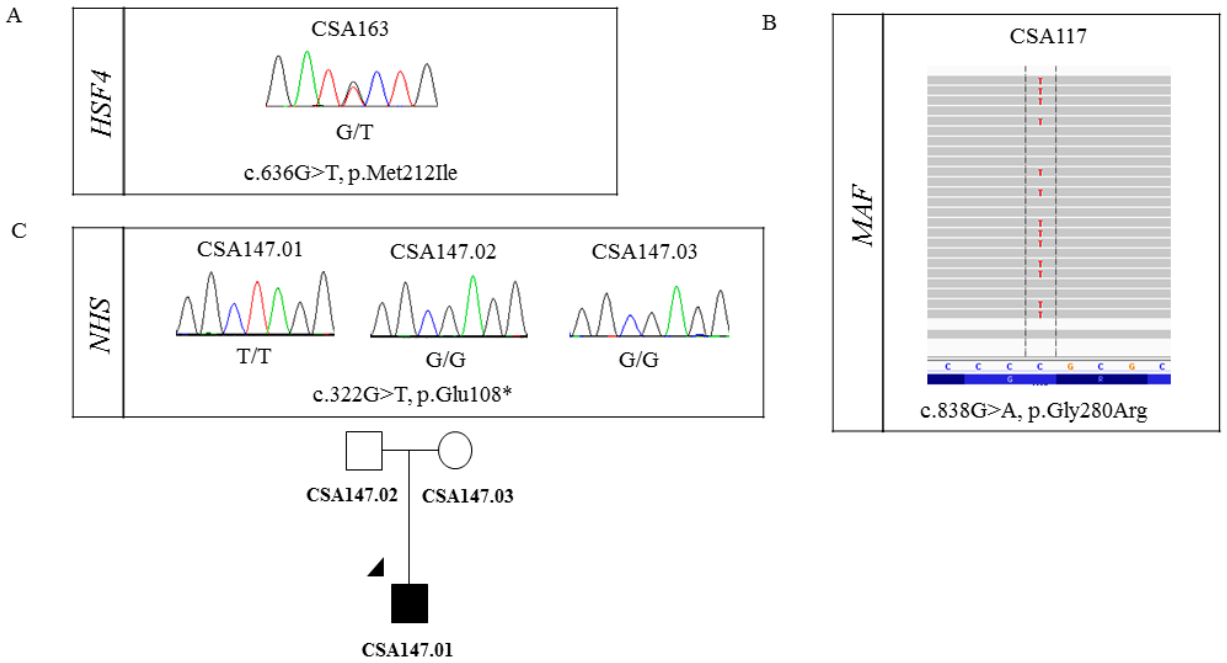


Figure 4.12. Chromatogram sequences of the variants detected in A: *HSF4* in CSA163, B: *MAF* in CSA117 and C: *NHS* in CSA147. The pedigree of family CSA147 is shown is section C. The altered nucleotide was visualised in CSA117 using The Integrative Genomics Viewer (IGV) and was confirmed using Sequenom.

Probands with mutations in more than one candidate gene

There were probands in the Australian cohort with sporadic paediatric cataract in which we have detected variants in two different candidate genes (Table 4.5, Figure 4.13). As these are sporadic cataract cases, there were no other family members available to access the segregation of the variant with cataract.

Mutations were detected in both *VIM* (p.(Asn201Tyrp)) and *NHS* (p.(Glu427Val*9)) in JM (Figure 4.13 A). Both detected variants were predicted to be pathogenic by both SIFT and polyphen-2 and are located within a conserved region of the protein (Table 4.5 and Appendix 15 C). Both of the variants are novel and neither has been previously reported to be associated with this condition. Two mutations were detected in crystallin genes (Table 4.5) in CSA164. One nonsense mutation was detected in *CRYGB* (p.(Arg170*)) and the other was a missense variation (p.(Val94Gly)) in *CRYAA* (Figure 4.13 B). The missense variation in *CRYAA* predicted to be pathogenic by Polyphen-2 but not SIFT and none of the programs utilised are able to predict the potential pathogenicity effect of a nonsense variation. Two potentially pathogenic missense variants (Table 4.5, Figure 4.13 C) were detected in *BFSP2* (p.(Glu371Lys)) and *MIP* (p.(Gly211Glu)) in CSA146.

The proband JM was diagnosed with dense central cataract at infancy and went through cataract surgery. JM also had severe glaucoma and developmental delay. The cataract phenotype was described as nuclear lamellar in CSA164 (Table 4.6). CSA146 had a cataract surgery at the age of two and there is no information available regarding his cataract phenotype (Table 4.6).

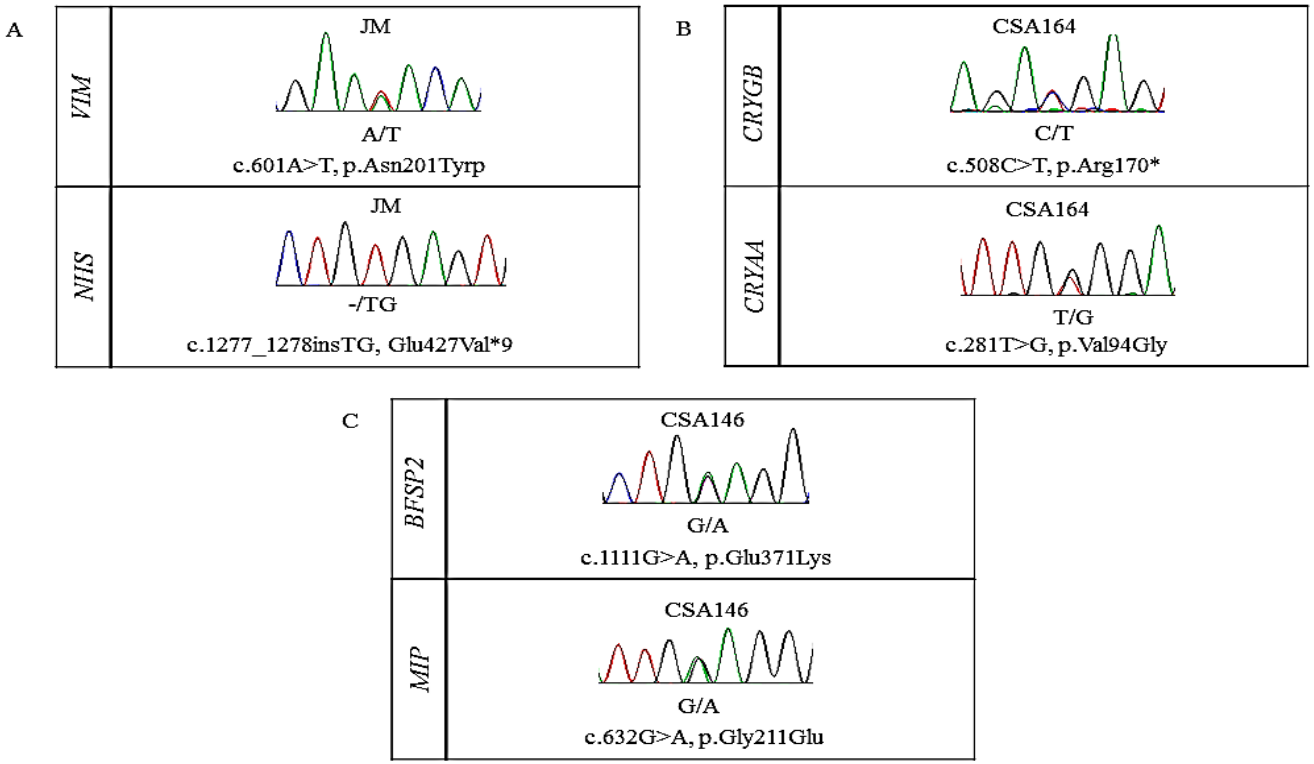


Figure 4.13. Chromatogram sequences of the variants detected in three probands with two potential causative candidate genes. Variants were detected in *VIM* and *NHS* in JM, in *CRYGB* and *CRYAA* in CSA164, in *BFSP2* and *MIP* in CSA146.

4.4.3 Screening Asian paediatric cataract cases

Thirty three paediatric cataract cases were screened for mutations in the 51 genes using the PGM. A total of 4844 variants were annotated with an average of 146 variants per individual. In total 188 variants were absent/rare (Minor Allele Frequency <1%) in publically referenced databases, of which 90 were non-synonymous exonic variants. Sixty five of these variants were filtered out due to their presence in our in-house list of sequencing artefacts, leaving 25 variants to be validated using Sanger sequencing. Twenty four variants successfully validated and one was a false positive signal.

All of the variants were screened in public databases (ExAC and dbSNP137) in order to determine their novelty. Our Asian paediatric cataract cohort is from Cambodia, Sri Lanka and Bhutan. A total of 4,327 individuals from East Asia and 8,256 from South Asia are included in the ExAC project, which makes it a suitable reference population for this cohort. Protein sequence alignments demonstrating the conservation of the altered amino acid detected in these families are given in Appendix 16 A, B, C and D. The information regarding the cataract phenotype in families with variations is given in Table 4.8.

Table 4.7. List of mutations detected in Asian probands with paediatric cataract.

Proband	Minor Allele Frequencies in ExAC	Novel/ Known	Country Of Origin	Gene	Position in hg19	Nucleotide change	Protein change	Phylop score	Polyphen-2 prediction	SIFT prediction	Segregation Penetrance Inheritance
P1	0	Novel	Bhutan	*GJA8*	chr1:147380921	c.839C>G	p.(Pro280Arg)	5.36	Probably damaging	Deleterious	Yes-AD
BB 16 cat	0	Novel	Cambodia	*GJA8*	chr1:147380102	c.20T>C	p.(Leu7Pro)	3.3	probably damaging	Deleterious	-
PP-50 cat	0	Known	Cambodia	*MIP*	chr12:56848301	c.97C>T	p.(Arg33Cys)	3.6	probably damaging	Deleterious	Yes Full/AD
	0	Novel		*COL4A1*	chr13:110864795	c.356C>G	p.(Pro119Arg)	5.27	probably damaging	Tolerated	No/AD
	0	Novel		*NSDHL*	chrX:152037470	c.932T>C	p.(Val311Ala)	0.9	probably damaging	Tolerated	No/AD
PCC 02-105	0	Novel	Sri Lanka	*CRYGD*	chr2:208986444	c.477_477delC	p.(Thr160Argfs*8)	0.75	-	-	Yes Full/AD
	0	Novel		*CRYGD*	chr2:208986623	c.299G>A	p.(Gly100Asp)	4.96	Probably damaging	Deleterious	No/-
	1.648×10^{-5}	Rare		*VIM*	chr10:17275680	c.719A>T	p.(Glu240Val)	5.2	probably damaging	Deleterious	No/-
E1	3.623×10^{-5}	Rare	Bhutan	*GALE*	chr1:24123216	c.766A>G	p.(Arg256Gly)	2.4	Benign	Deleterious	Yes Full/AD
PCC 10-183	0	Novel	Sri Lanka	*GCNT2*	chr6:10626784	c.1153C>T (homozygous)	p.(Arg385Cys)	1.51	probably damaging		Yes/AR
SR 11 cat	0	Novel	Cambodia	*PAX6*	chr11:31823289	c.177G>C	p.(Arg59Ser)	1.79	probably damaging	Tolerated	-/-
PCC 10-188	0	Novel	Sri Lanka	*EPHA2*	chr1:16464671	c.987_988insT	p.(Ser330Phe)	-	-	-	-/-
	0	Novel		*GCNT2*	chr6:10626784	c.1153C>T	p.(Arg385Cys)	1.51	probably damaging		
	1.824×10^{-4}	Rare		*NHS*	chrX:17743727	c.1438C>T	p.(Arg480Cys)	5,33	probably damaging	Tolerated	
**W1	9.946×10^{-5}	Rare	Bhutan	*GALK1*	chr17:73758836	c.742C>T	p.(Arg248Trp)	0.25	probably damaging	Deleterious	Yes Reduced/-
	4.157×10^{-5}	Rare		*GALK1*	chr17:73759221	c.485C>G	p.(Thr162Arg)		Benign	Deleterious	

Table 4.7 continued

Family ID	Minor Allele Frequencies in ExAC	Novel/Rare Known	Country of Origin	Gene	Position in hg19	Nucleotide change	Protein change	Phylop score	Polyphen-2 prediction	SIFT prediction	Segregation Penetrance Inheritance
PCC01-97A	1.105×10^{-3}	Rare	Sri Lanka	*AGK*	chr7:141255292	c.26G>A	p.(Arg9Gln)	2.47	Possibly damaging	Deleterious	-/-
	2.474×10^{-5}	Rare		*TDRD7*	chr9:100234592	c.1759G>T	p.(Asp587Tyr)	1.38	Probably damaging	Deleterious	
	0	Novel		*PAX6*	chr11:31815036	c.982G>T	p.(Ala328Ser)	6.22	Benign	Tolerated	
	1.483×10^{-4}	Rare		*BFSP1*	chr20:17479645	c.776G>C	p.(Cys259Ser)	3.99	probably damaging	Deleterious	
	0	Novel		*CRYBB1*	chr22:27008146	c.186_188delGGT	p.(Val62_Phe64del)	-	-	-	
PCC 01-34	1.53×10^{-4}	Rare	Sri Lanka	*HSF4*	chr16:67201678	c.910G>A	p.(Glu304Lys)	0.9	Benign	Tolerated	-/-
SR 12 cat	1.649×10^{-5}	Rare	Cambodia	*TDRD7*	chr9:100245251	c.2533C>G	p.(Gln845Glu)	3.5	probably damaging	Deleterious	No/-
	1.647×10^{-5}	Rare		*COL4A1*	chr13:110857844	c.900T>A	p.(Ser300Arg)				No/-

Genbank Accession numbers are shown in Table 4.1. *This variant is only covered in 43866 individuals in ExAC database (adjusted allele number = 87732). ** The variants in W1 were validated using sequenom platform. "-"indicates no information is available.

Table 4.8. Observed phenotypes in Asian probands with segregating causative mutations identified in paediatric cataract associated genes.

Proband	**Gene**	**Phenotype**	**Age at diagnosis**	**Surgery Detail**	
				right eye	**left eye**
P1	*GJA8*	Congenital cataract with posterior capsule opacification	-	Yes	Yes
BB 16 cat	*GJA8*	Paediatric cataract	-	-	-
PP-50 cat	*MIP,COL4A1, NSDHL*	Paediatric cataract	-	-	-
PCC 02-105	*CRYGD,CRYGD, VIM*	Bilateral congenital cataract (familial)	-	No	No
E1	*GALE*	Congenital cataract, amblyopia, retinal dystrophy	-	Yes	Yes
PCC 10-183	*GCNT2*	Congenital cataract (familial) with nystagmus	Birth	No	No
SR 11 cat	*PAX6*	Paediatric cataract	-	-	-
PCC 10-188	*EPHA2, GCNT2, NHS*	Paediatric cataract (familial)	Birth	Yes	Yes
W1	*GALK1*	Congenital cataract of unknown aetiology	-	Yes	Yes
PCC01-97A	*AGK, TDRD7, PAX6 BFSP1, CRYBB1*	Paediatric cataract, microphthalmos and pseudophakia (familial)	Birth	Yes	Yes
PCC 01-34	*HSF4*	Paediatric cataract (sporadic)	Birth	-	-
SR 12 cat	*TDRD7, COL4A1*	Paediatric cataract	-	-	-

"-"indicates no information is available.

Mutations detected in connexin encoding genes

Two different mutations in *GJA8* were detected in BB16cat (p.(Leu7Pro)) and P1 (p.(Pro280Arg)). Both mutations are located in a conserved protein region, and are predicted to be pathogenic by both SIFT and Polyphen-2 analysis (Table 4.7) (Appendix 16 A). The mutation detected in P1 was present in the affected sister (P2) as well (Figure 4.14 A).

The phenotype in family P1 was described as congenital cataract with posterior capsule opacification. The proband had cataract removal surgery which demonstrates the severity of the disease nature (Table 4.8). No information is available regarding the disease status of the parents in this family.

Membrane proteins

One previously reported (Ma, et al., 2016) segregating mutation (p.(Arg33Cys)) in *MIP* was detected in the family PP50 Cat. All affected individuals in the family (the proband's affected sister, brother and father) carry the mutation (Figure 4.14 B). The other two variants in *COL4A1* (p.(Pro119Arg)) and *NSDHL* (p.(Val311Ala)) didn't segregate with the disease as some of the affected members did not carry these variants (Figure 4.14 B).

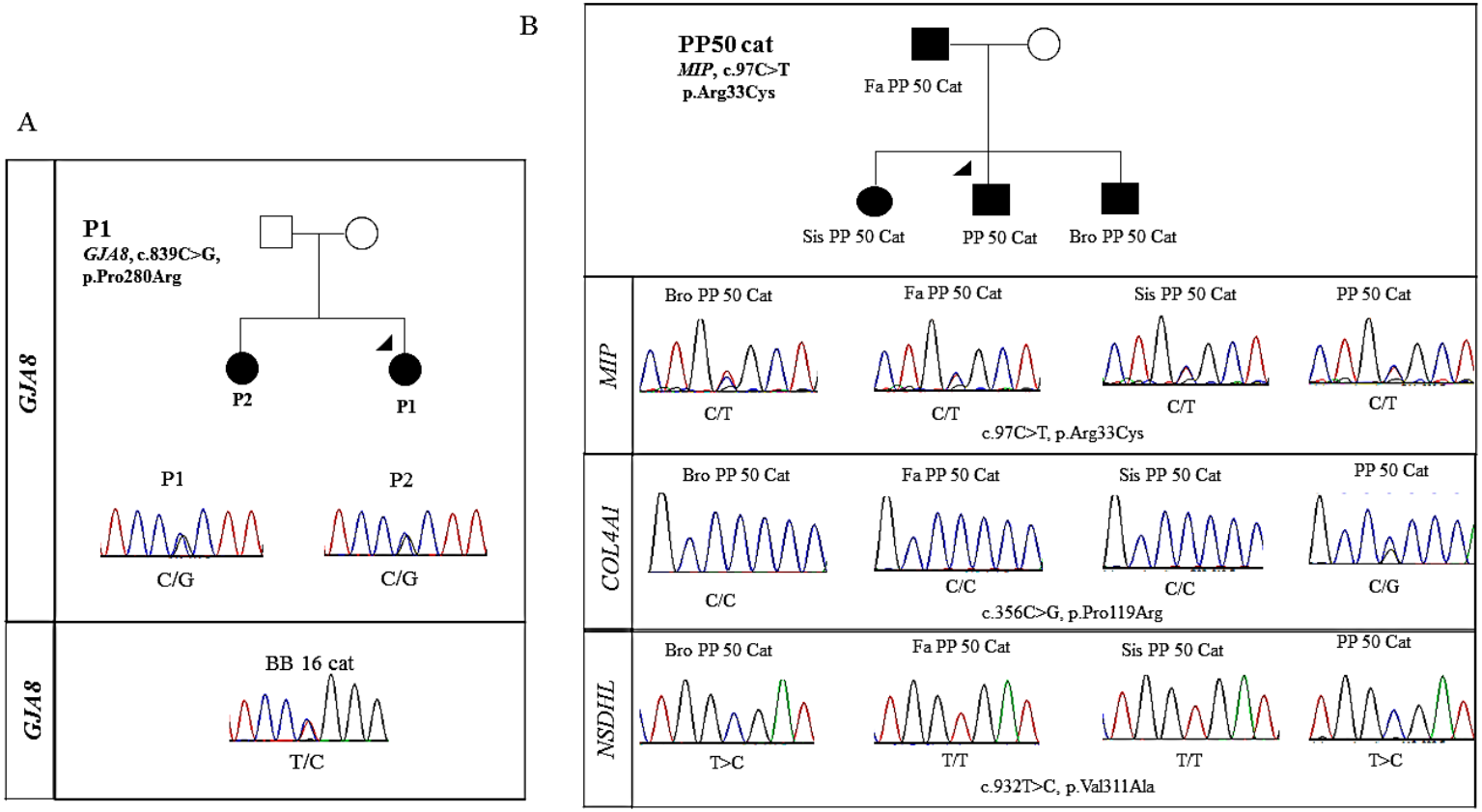


Figure 4.14. The mutations detected in *GJA8* (A) and *MIP* (B) genes in BB 16 cat (A), P1 (A) and PP 50 cat (B). The mutation in P1 (*GJA8*, p.(Pro280Arg)) is segregating with the phenotype as the affected sister P2 also carries the mutation. The only mutation that segregates with the phenotype in PP50 Cat is the mutation detected in *MIP* (p.(Arg33Cys)).

Variations detected in crystallin genes

One frameshift mutation in *CRYGD* (p.(Thr160Argfs*8)) was detected in PC 02-105 (Figure 4.15) and segregates with the phenotype, as the affected cousins (PCC 02-103, PCC 02-107 and PCC 02-109) also carry the mutation. The detected variant is not present in ExAC database and therefore is classified as novel. The other two missense variants, p.(Gly100Asp) in *CRYGD* and p.(Glu240Val) in *VIM*, did not segregate with the phenotype in this family.

The cataract phenotype in family PCC02 is classified as familial bilateral paediatric cataract.

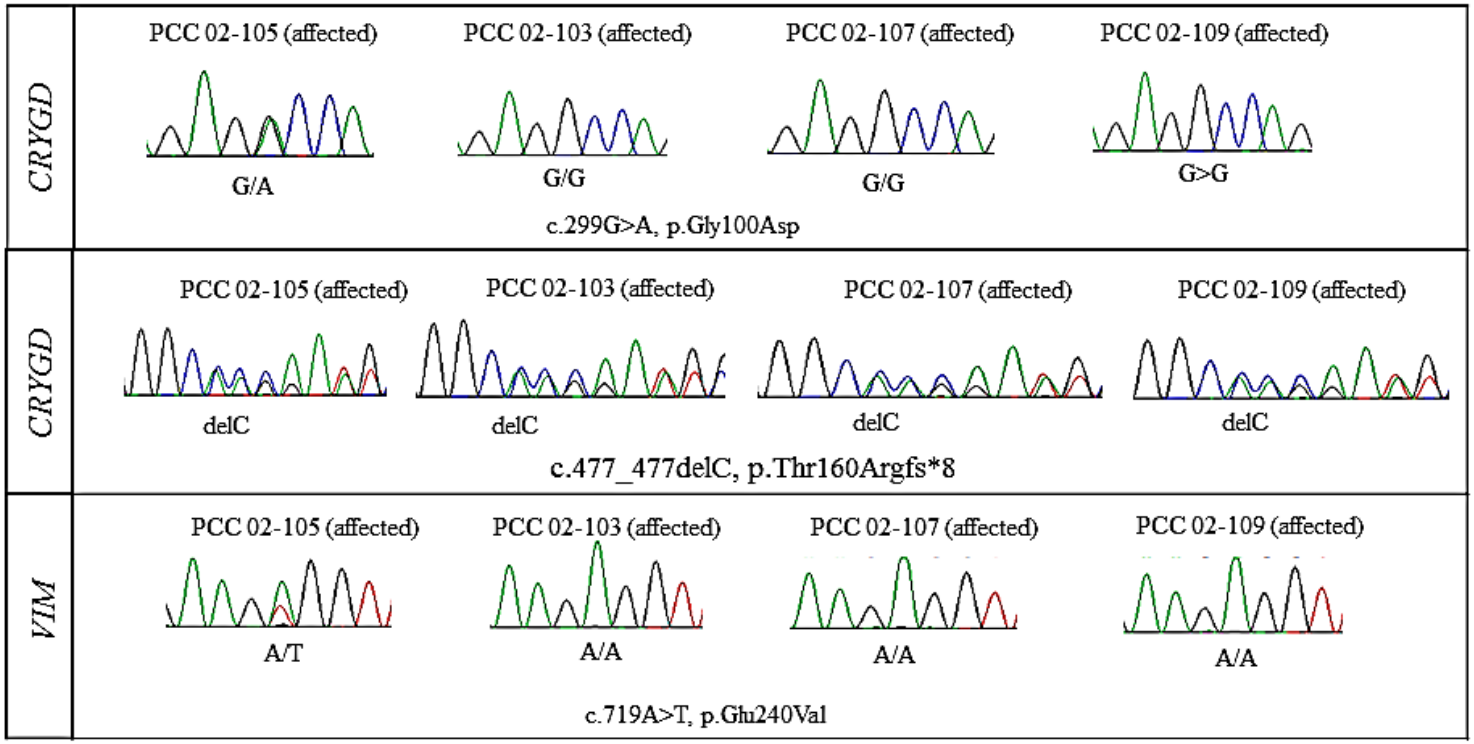


Figure 4.15. The variations detected in family PCC 02. The only segregating variant which is present in all the affected members (the proband PCC 02-105 and the affected cousins PCC 02-103, PCC 02-107 and PCC 02-109) is p.(Thr160Argfs*8) in *CRYGD*. The other 2 variants in 2 *VIM* (p.(Glu240Val)) and *CRYGD* (p.(Gly100Asp)) were not segregating.

Mutations in encoding enzymes and syndromic cataract associated genes

Mutation p.(Arg256Gly) in UDP-galactose-4-prime-epimerase encoding gene, *GALE*, was detected in patient E1 from Bhutan (Figure 4.16). Amblyopia, retinal dystrophy was other phenotypes beside paediatric cataract in E1 indicating the syndromic nature of the disease.

Two mutations in *GALK1* were detected in proband W1 from Bhutan. Both of these variants were validated in the proband and W2 (unaffected brother) using the Sequenom platform (due to difficulty in primer design for sequencing).

Both PCC10-183 and PCC10-189 (affected brother) carry the homozygous mutation (p.(Arg385Cys)) in *GCNT2* (Figure 4.16). The parents are unaffected and there were no DNA samples available from them.

One missense variant was detected in SR11 cat in *PAX6* (p.(Arg59Ser), Figure 4.16). The detected variant was predicted to be tolerated by SIFT and pathogenic by Polyphen-2 (Table 4.7). There is no information available regarding the phenotype of cataract of the proband. The altered amino acid is located within a conserved region of the protein (Appendix 16 C).

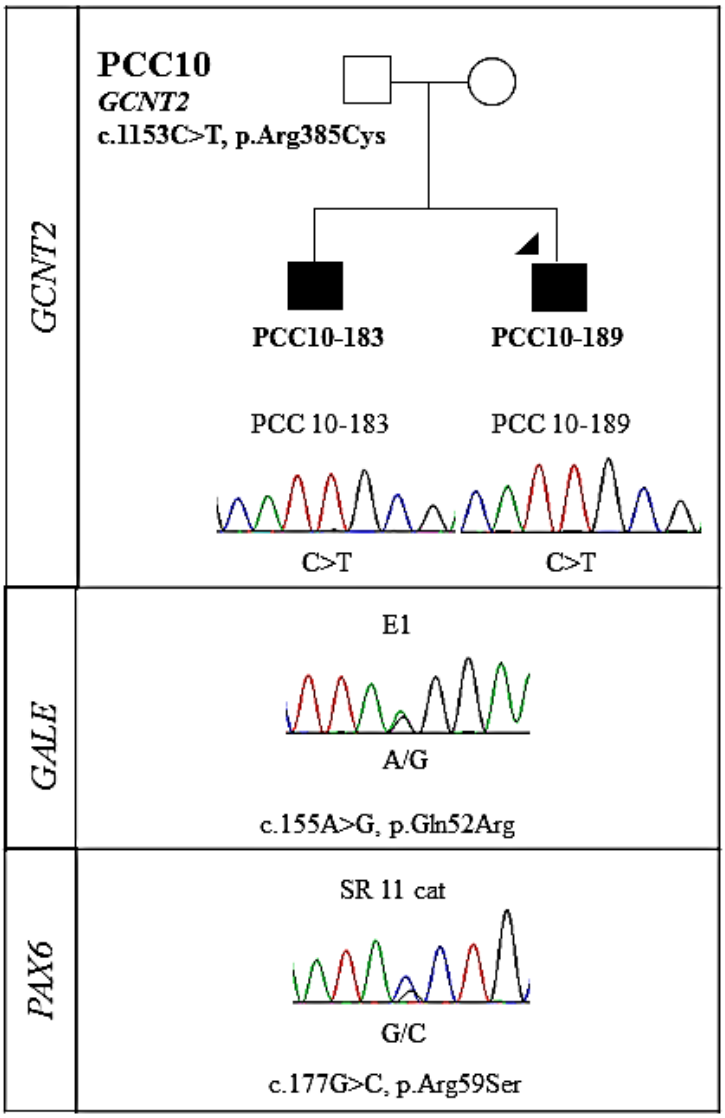


Figure 4.16. Mutations in syndromic and enzymatic paediatric cataract associated genes in Asian cohort. E1 with the mutation in *GALE* (p.(Gln52Arg)). Homozygous mutation in *GCNT2* detected in 2 affected members of PCC 10. Detected variant in SR11 cat in *PAX6* (p.(Arg59Ser)).

Probands with mutations in multiple paediatric cataract genes

There were 3 probands from the Asian cohort (W1, PCC01-97A and PCC10-188) in which mutations in more than 1 candidate gene were identified. The variants in *GALK1* which were detected in W1 are not present in his unaffected brother (validated using sequenom technologies). Of the 5 detected variants in PCC01-97A, 3 were rare (in *AGK*, *TDRD7*and *BFSP1*) and 2 (in *PAX6* and *CRYBB1)* were novel (Figure 4.17 A and Table 4.7). One rare variant in *NHS* and 2 novel variants in *EPHA2* and *GCNT2* were detected in PCC 10-188 (Figure 4.17 B and Table 4.7).

Unfortunately no information is available regarding the phenotype for 2 of these probands and cataract of PCC01-97A was described as syndromic (paediatric cataract with microphthalmos and pseudophakia).

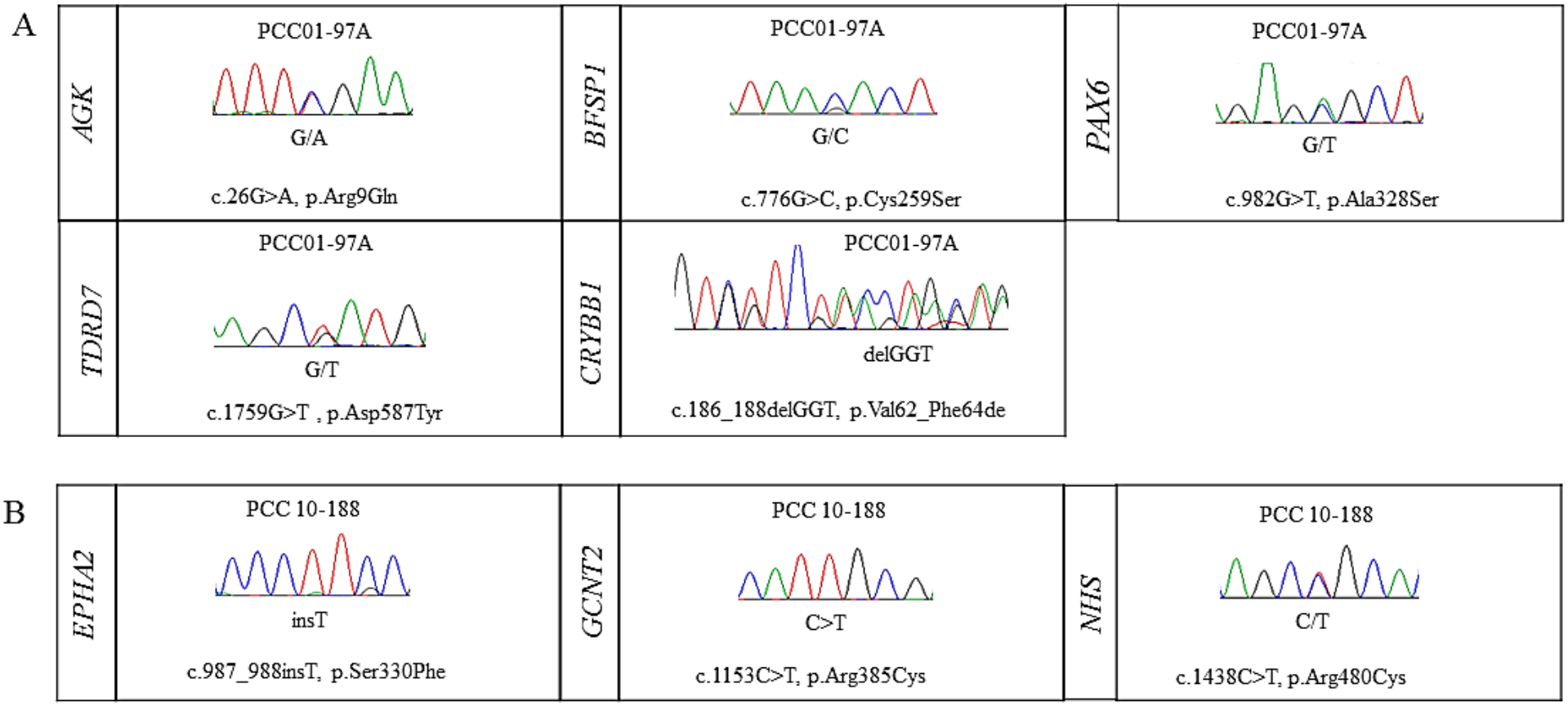


Figure 4.17. Detected variations in individuals PCC01-97A 9A (A) and PCC 10-188 (B) in multiple candidate genes.

Non Segregating or benign variations

The missense variant detected in PCC 01-34 in *HSF4* (p.(Glu304Lys)) is predicted to be benign by both SIFT and Polyphen-2 analysis and is located in a poorly conserved region of the protein (Table 4.6, Appendix 16 C).

None of the rare variants detected in SR 12cat in *COLA41* (p.(Ser300Arg)) and *TDRD7* (p.(Gln845Glu)) segregated with the phenotype. There are 3 affected children (SR9, SR12 and Bro SR9 +SR12) in the family and both of the parents are unaffected (Fa SR9 + SR 12 and Mo SR 9 + SR 12), however it is notable that the parents are first cousins. The affected SR 9 cat does not carry the detected variants in *COL4A1* or *TDRD7* (Figure 4.18).

As mentioned before, two missense variants, p.(Gly100Asp) in *CRYGD* and p.(Glu240Val) in *VIM* in family PCC 02 and 2 variants in *COL4A1* (p.(Pro119Arg)) and *NSDHL* (p.(Val311Ala)) in PP-50 did not segregate with the phenotype (Table 4.7).

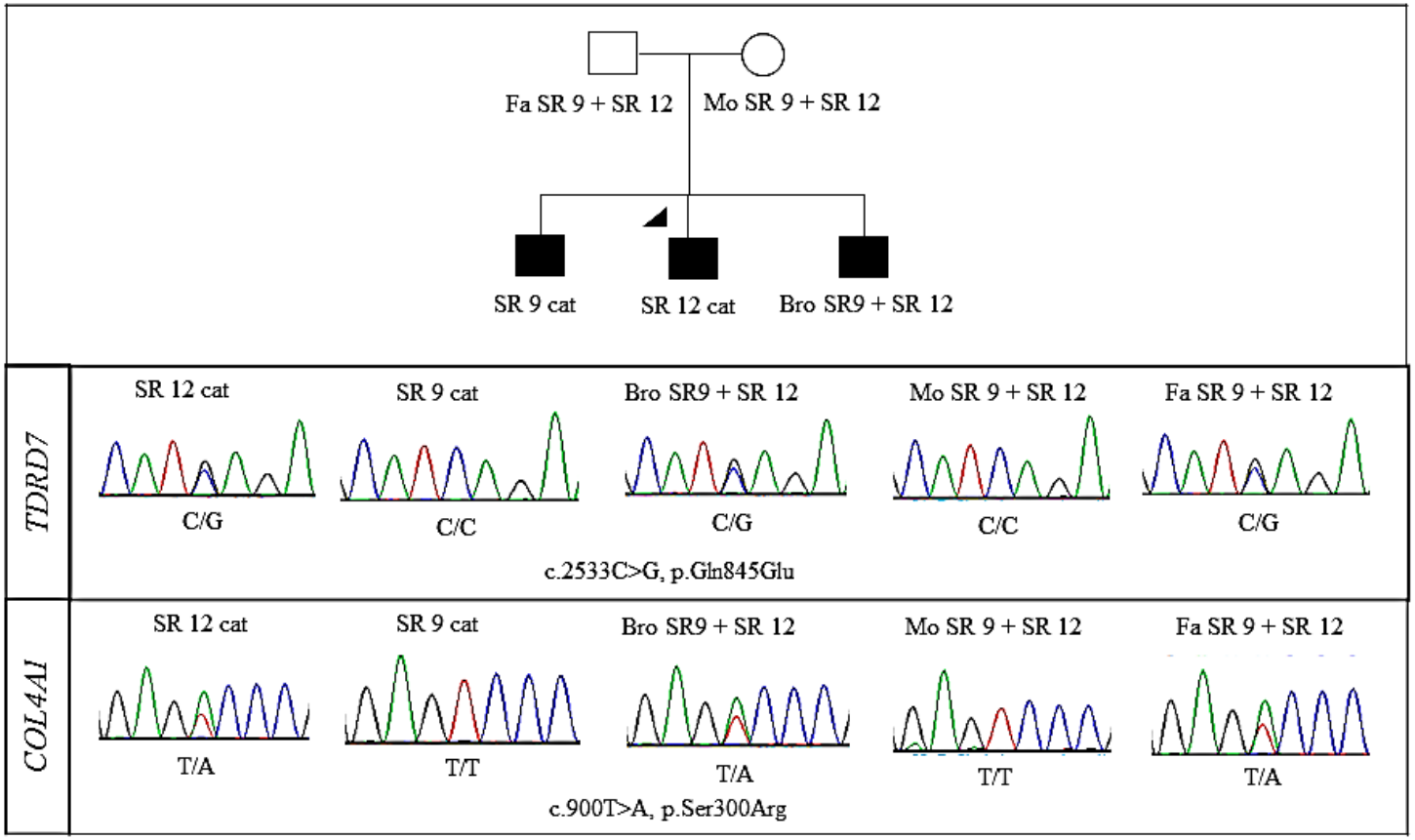


Figure 4.18. The chromatograms show the detected variants in family SR 12. None of the detected variants in *TDRD7* and *COL4A1* is segregating with the phenotype as the affected SR 9 cat doesn't carry any of the variations. Three children (SR9, SR12 and Bro SR9 +SR12) are affected and both of the parents (Fa SR9 + SR 12 and Mo SR 9 + SR 12) are unaffected.

4.5 Discussion

We hypothesized that a significant proportion of paediatric cataract cases would be accounted for by mutations in known genes, and that screening genes in parallel would be an effective method for genetic testing in this heterogeneous disease.

In this study we used targeted massively parallel sequencing to identify genetic variants associated with inherited or sporadic paediatric cataract cases from Australia and South East Asia.

The phenotypes observed in these families or individuals, where detailed information is available, are similar to phenotypes which have been described in previous reports of mutations in these genes. For example, we describe predominantly nuclear phenotypes related to mutations in the gap junction genes *GJA3* and *GJA8* (Shiels, et al., 2010). Similarly, nuclear or total cataracts are observed in *CRYAA* mutation carriers (Devi, et al., 2008; Khan, et al., 2007) while mutations in *CRYBB2* give rise to cortical and lamellar cataracts (Devi, et al., 2008; Faletra, et al., 2013). This provides further evidence that the identified mutations are indeed causative of the condition in these individuals.

4.5.1. Mutations detected in membrane or cytoskeleton proteins encoding genes

Over 60 different mutations in *GJA8* and *GJA3* have been reported to cause paediatric cataract, with around a third of them causing nuclear cataract (Shiels, et al., 2010). We previously described an Australian family with a "faint lamellar nuclear opacity (surrounding pulverulent nuclear opacities)" with a mutation in *GJA3* (Burdon, et al., 2004b) which is similar to the phenotype in individual CSA109.01 in this study. Moreover, multiple studies have reported a link between nuclear congenital cataract (the

majority of them being familial paediatric cataract) and mutations in *GJA8* including p.(Arg23Thr) (Iran (Willoughby, et al., 2003)); p.(Asp47Asn) (China (He, et al., 2011b)) and p.(Val64Gly) (China (Zheng, et al., 2005)). We also detected a mutation in *GJA8* p.(Glu370*) in CSA150 with sporadic nuclear paediatric cataract. Previously, one homozygous mutation (p.(Arg425Ter)) in *GJA8* has been reported to be associated with sporadic paediatric cataract; however no information was given regarding the cataract phenotype (Gillespie, et al., 2014). Furthermore, approximately 25% of the detected mutations were in *GJA8* in our screened Asian cohort. To our knowledge this is the first report of mutations detected in a gap junction gene in families from Cambodia (p.(Leu7Pro) in BB 16 cat) and Bhutan (p.(Pro280Arg) in P1). Being membrane proteins, connexins have four transmembrane domains (TM1-TM4), two extracellular regions, and three intracellular regions containing the NH2 terminal, a cytoplasmic region, and the COOH terminal (Chen, et al., 2015) The mutations in families CRCH20 and CSA162 lie within TM1 and the mutation in family CSA125 lies within TM3. The TM1-TM4 domains of the connexins are important for oligomerization into connexin hemichannels and for the proper transportation of the protein into the plasma membrane (Devi and Vijayalakshmi, 2006). It has been proposed that pore lining residues lie in the TM1 domain are essential for the pore formation and channel permeability (Kronengold, et al., 2003). Therefore this supports our hypothesis that this mutation causes the phenotype in this family.

One novel missense mutation was detected in *BFSP2* in family CSA107 (p.(Leu234Pro)). The phenotype described in the proband (CSA107.01) is cortical with prominent posterior opacification in the sutural region (Figure 4.10). The proband was diagnosed at birth with mild cataract and had surgery in both eyes at age 24, indicating a slowly progressive disease (Table 4.5). As no other affected family members were examined, it is unclear whether this family represents autosomal recessive inheritance or

autosomal dominant with reduced penetrance in the parental generation. The unaffected mother carries the *BFSP2* mutation inherited by her affected daughter (Figure 4.10). The possibility of the autosomal dominant transmission with reduced penetrance in the parental generation could not be excluded.

The phenotype in individual CSA154 with a mutation in *BFSP2* (p.(Glu141Ala)) was described as lamellar. One study demonstrated an association between a mutation in *BFSP2* in a Chinese family with lamellar cataract (Ma, et al., 2008). The proband of this four-generation family was diagnosed at birth which is similar to CSA154, however the cataract in CSA154 was not described as familial.

Family CSA131 has a novel truncating mutation in the *MIP* gene (p.(Gly211*)). The proband from this family had cortical and nuclear sclerotic cataracts with multiple cortical dots whereas two other affected members (CSA 131.01 and CSA 131.02) had anterior cortical spoke and white dots. This phenotypic spectrum is similar to some extent to what has been described in another autosomal dominant cataract family with variable severity ranging from fine punctate opacities in the posterior cortex and Y sutures, few anterior cortex punctuate opacities or white punctuate opacities in the cortex with larger posterior cortical region opacities (Geyer, et al., 2006). The phenotype in family CSA131 was relatively moderate as the age of diagnosis was over 20 and only one of the affected family members has had surgery to date (at the age of 46). There is no information available regarding the phenotype of the other 2 *MIP* variants detected in CSA146 (p.(Gly211Glu)) with sporadic paediatric cataract and PP 50 cat from Cambodia (p.(Arg33Cys)). Despite the segregation of the variation with the phenotype, the altered amino acid residue, Arg, in PP 50 cat is not that well conserved among different species (Appendix 16). The altered amino acid in CSA146, Gly, is well conserved among different species as well as the other altered amino acid (Glu) detected

in this proband in *BFSP2* (p.(Glu371Lys)) which makes it difficult to determine which is the causative variant or if both are required to cause cataract.

4.5.2 Mutations detected in crystallin genes

Nuclear or total cataract phenotypes were observed in family CRVEEH111, in which a mutation was identified in *CRYAA*. Others have reported similar nuclear phenotypes due to mutations in this gene (Devi, et al., 2008; Khan, et al., 2007) (Mackay, et al., 2003), although a variety of other phenotypes have also been reported (Laurie, et al., 2013; Sun, et al., 2011a). This gene is expressed early in lens development and is involved in fibre cell differentiation (Boyle and Takemoto, 2000). The phenotype in family CRVEEH111 is a central nuclear opacity with various levels of severity in affected members, very similar to that described by Mackay et al (Mackay, et al., 2003). Despite the SIFT prediction that the p.(Arg54Cys) variant in this family would be tolerated functionally, the similarity of the phenotypes between this family and others, combined with a Polyphen-2 prediction of pathogenicity and previous reports of this mutation in paediatric cataract patients in other populations, such as Saudi Arabian (Khan, et al., 2007) and Indian (Devi, et al., 2008), seems to suggest that this variant is likely pathogenic. The novel frameshift deletion in *CRYAA* (p.(Gln147Argfs*48)) detected in consanguineous family CSA159 (Figure 4A) is also associated with nuclear & cortical cataract, strengthening this genotype/phenotype correlation.

Only one potential disease causing mutation was identified in family CSA110 (Table 4.4). This mutation is located in *CRYAA* and segregates with the phenotype in the family. The phenotype in the previously reported family CSA91 is described as lamellar with variable severity (Laurie, et al., 2013) which is similar to the observed phenotype in CSA110, who also display variability in both severity and age at diagnosis (Figure 4.3 and table 4.4). Of 7 available family members, all five affected members carried the

mutation (Figure 4.5 A), whilst one of the unaffected members also carried the mutation, indicating probable reduced penetrance. Previously, the same oligomerization disrupting mutation in *CRYAA*, p.(Arg21Gln), was reported in an apparently unrelated Australian family with lamellar congenital cataract of variable severity (Laurie, et al., 2013). The detection of the same mutation in *CRYAA* in two families from the same state of Australia and with such a similar phenotype led us to question whether these families were related. Genetic haplotype analysis suggests that the mutation has arisen only once on the same ancestral chromosome, and therefore these families are likely to be distantly related (Figure 4.6 A and B). No genealogical link was readily identified on specific questioning of the family members. The mutation was not detected previously in Australian controls (Laurie, et al., 2013), nor was it present in commonly accessed databases at the time of sequencing, however, it has subsequently been reported in the ExAC database (http://exac.broadinstitute.org/) with a minor allele frequency of 0.000016 (i.e. 1.6 per 100,000 chromosomes). Thus, this variant is extremely rare, which is consistent with a role in a rare disease such as paediatric cataract.

There have been several *CRYAA* mutations reported in multiple families, although the ancestral origins have not been assessed for most. Hansen *et al.* (2009) investigated the founder effect of a recurrent mutation (p.(Arg21Trp)) which was found in 3 families by utilising a SNP (rs872331) located 55 nucleotides upstream of the mutant codon. They detected an identical haplotype c. [6T; 61C] in 2 of the families, suggesting a common ancestral founder which was then confirmed with subsequent genealogical studies. The detection of the c. [6C; 61C] haplotype in the third family suggested that the mutation arose independently in that family (Hansen, et al., 2009).

Similar to many heat shock proteins, CRYAA has three distinct domains: the N-terminal region encoded by exon 1, a α-crystallin domain (ACD) and a short C-terminal encoded by exon 2 and 3 (Hansen, et al., 2007b). The p.(Arg54Cys) and p (Arg21Gln)

recurrent mutations identified here in CRVEEH111 and in CSA110 lie within the N-terminal domain, which has been demonstrated to be important for subunit exchange and oligomerization. It has been suggested that the deletion of this highly conserved amino acid which is located in a conserved motif (SRLFDQFFG), leads to disruption of the correct assembly of the quaternary alpha-crystallin structure (Pasta, et al., 2003). Western blotting of lens protein from a mutation carrier also showed a decrease in the ability of the protein to form higher order oligomers essential for its function (Laurie, et al., 2013). There have been multiple reports of mutations altering the highly conserved arginine residue at position 21 (p.(R21W) and p.(R21L)) associated with a variety of congenital cataract phenotypes, indicating its importance for proper CRYAA function.

A dinucleotide substitution in the *CRYGS* gene (p.(Phe10_Tyr11delinsLeuAsn)) segregated with the phenotype in family CSA94. This family had lamellar and cortical cataracts. Two novel mutations in *CRYGC* (p.(Arg142fs)) in CSA172 and *CRYGS* (p.(Leu80Arg) in CSA155 were detected in the Australian cohort. Both of these individuals were diagnosed at birth with severe congenital cataract and went through cataract surgery at infancy (no family history of paediatric cataract). Mutations in *CRYGS* have also been reported in families with progressive juvenile onset sutural, lamellar cataract (p.(Ser39Cys)) (Devi, et al., 2008) or progressive polymorphic cortical cataract (p.(Gly18Val)) (Sun, et al., 2005). There has also been one report of a sporadic paediatric cataract case from China with a mutation in *CRYGS* (p.(D26G)) (Sun, et al., 2011a). We detected one segregating mutation in PCC 02 from Sri Lanka (p.(Thr160Argfs*8)) in *CRYGD*. There have been reports of over 60 paediatric/congenital cataract cases with mutations detected in *CRYGD*, however none of these mutations were detected in any Asian cohort similar to this study.

Mutations in the *CRYBB2* gene have been reported with cortical and lamellar cataracts (Devi, et al., 2008; Faletra, et al., 2013), similar to those seen in the present study in

families CRVEEH85 (p.(Arg188Leu)) and CSA133 (p.(Gln155*)). Variant p.(Gln155*) has been previously reported with various paediatric cataract phenotypes in different populations including cerulean in China (Litt, et al., 1997; Wang, et al., 2009b); central zonular pulverulent in Switzerland (Gill, et al., 2000) and cortical, pulverulent in India (Devi, et al., 2008). One more variant (p.(Val62_Phe64del)) in *CRYBB2*, along with 4 more variants in different genes (Table 4.7) were detected in PCC01-97A, a proband from Sri Lanka, however due to limited available members from the family, the contribution of the detected variants to their disease couldn't be evaluated.

CRYBB2 protein consists of 4 'Greek Key' domains, N-terminal region, and C-terminal region. Both mutations reported here in CSA133 and CRVEEH85 fall within the 4th Greek Key domain. The Greek key motif is one of the most stable structures in proteins, and has been suggested to have a role in intermolecular associations in the β-crystallin (Yao, et al., 2011a). The mutation in CRYBB2 in family CSA133 which we report here creates a premature stop codon. This is predicted to produce a severely truncated protein, lacking the final 51 amino acids (more than 10% of the protein length from the COOH-terminus) and disrupting the domain structure of betaB2-crystallin and its stability (Liu and Liang, 2005).

4.5.3 Mutations detected in transcription factors, signalling molecules or in genes associated with syndromic and enzymatic paediatric cataract

Among the many genes associated with congenital cataracts is the transcription factor gene *MAF*. The MAF family of transcription factors are divided into two subgroups, large and small. The large subgroup (*MAFA*, *MAFB*, *c-MAF* or *v-MAF*, and retina-specific leucine zipper (*NRL*)) is characterized by a bZip structure, a motif for DNA binding, a protein dimerization domain and a transactivation domain (Tsuchiya, et al.,

2015). The small MAF proteins (MAFF, MAFG, and MAFK) lack the transactivation domain (Kataoka, 2007; Tsuchiya, et al., 2015). Mutations in *MAF* has been shown to be associated with both syndromic and isolated paediatric cataract as seen in our Australian cohort in the current project.

Two missense mutations in *MAF* gene were identified in this Australian cohort (Table 4.3 and Table 4.5). The features described in CSA108.01 are consistent with the condition previously reported independently by Aymé and Phillip (Ayme and Philip, 1996) and Gripp et al. (Gripp, et al., 1996) (MIM 601088). There also have been reports of a similar syndrome by Fine and Lubinsky (Fine and Lubinsky, 1983) and Preus *et al.* (Preus, et al., 1984). A recent study by Niceta *et al*. (Niceta, et al., 2015) reported a narrow spectrum of amino-acid substitutions within the MAF protein (Figure 4.19), causing cataract, deafness, intellectual disability, seizures, a distinctive flat facial appearance, skeletal anomalies and reduced growth. The authors proposed the eponym Aymé-Gripp for this multisystem disorder. The reported *de novo* amino acid substitutions in *MAF* associated with this syndrome are p.(S54L), p.(T58A), p.(T58I), p.(P59H), p.(P59L), p.(T2R) and p.(P69R). Interestingly, all these mutations are located within the N-terminal transactivation domain of *MAF,* as is the p.P59R substitution reported here (Figure 4.19). Unlike other reported mutations in *MAF* associated with Aymé-Gripp syndrome (Niceta, et al., 2015), the mutation described here is inherited and transmitted from the mother to the affected child. This study also shows that mutations causing Aymé-Gripp syndrome can display intra-familial variability since the mother had a substantially milder phenotype than the proband.

The second mutation was a *de novo* mutation identified in CSA117. There have been multiple reports of mutations in *MAF* associated with various forms of isolated or syndromic paediatric cataracts (Figure 4.19):, (p.(R294W)) has been linked with nuclear congenital cataract, (Sun, et al., 2014), (p.(K297R)) with cerulean congenital cataract

and microcornea (Vanita, et al., 2006b), (p.(R299S)) with lamellar cataract with microcornea and iris coloboma, (Hansen, et al., 2007a) and (p.(K320E)) has been linked with nuclear, punctate, stromal cataract with microcornea (Hansen, et al., 2009). Jamieson *et al*. described a mutation in the DNA-binding domain of *MAF* (p.(R288P)) in a three generation family with lamellar cortical and nuclear pulverulent cataract, microcornea, and iris coloboma. Narumi et al. (Narumi, et al., 2014) identified a *MAF* mutation (p.(E303L)) through whole exome sequencing in a family with phenotypically variable congenital cataract (lamellar or anterior polar with microcornea and iris coloboma). The affected proband was diagnosed with lamellar cataract without any other eye malformation (similar to the case presented here: CSA117) with language development delay and autism. All of these mutations are located in the C-terminal DNA binding domain of the MAF protein (Figure 4.19).

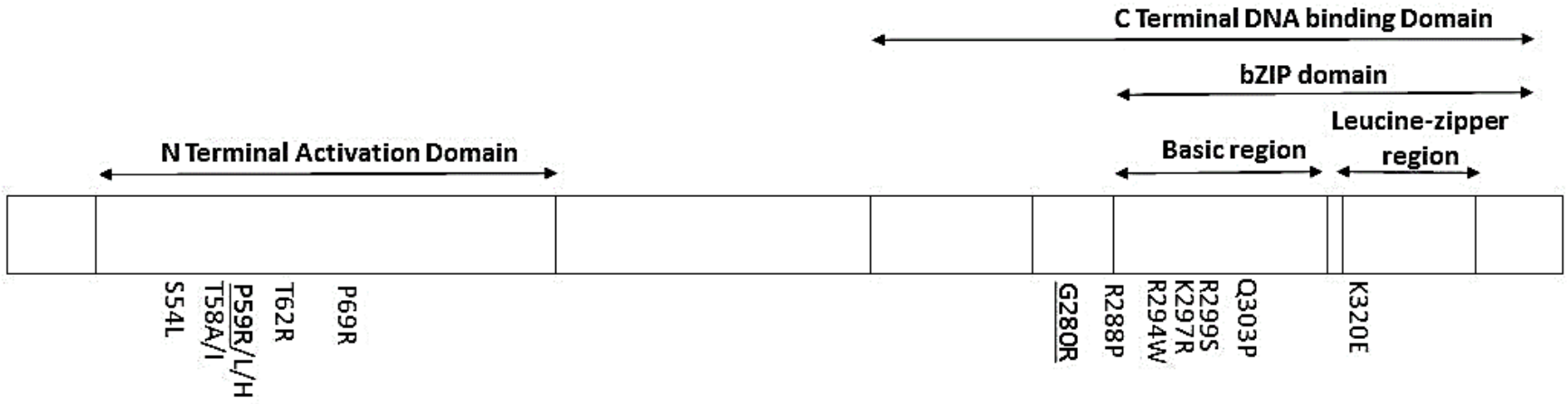


Figure 4.19. Schematic of the human MAF protein indicating the positions of reported mutations (Adapted from (Niceta, et al., 2015)). The protein contains an N-terminal transactivation domain and a C-terminal DNA binding domain. The C-terminal domain consists of an extended homology region, basic region (aa288-313) and leucine-zipper region (aa316- aa337). The mutations associated with Aymé-Gripp syndrome are located in the N-terminal transactivation domain including the mutation (p.(P59R)) reported here (bolded and underlined). Other mutations are located within the C-terminal DNA-binding including the one detected in CSA117 (p (G280R), bolded and underlined) domain and are associated with other forms of congenital cataract, either isolated or syndromic.

One rare missense mutation (p.(Met212Ile)) in Heat Shock Transcription Factor 4 (HSF4) was detected in CSA163, a family affected with posterior subcapsular congenital cataracts (PSCC). To our knowledge, this is the first report of such an association with this phenotype. As shown in Appendix 3, a variety of paediatric cataract phenotypes are associated with the mutations in *HSF4*, however there isn't any correlation between a particular phenotype and mutation in this gene. The altered methionine residue is not perfectly conserved among different species, however the mutation was predicted to be possibly damaging by Polyphen-2 (Figure 4.12). The variant (p.(Glu304Lys)) detected in PCC 01-34 in *HSF4* was predicted to be non-functional by both SIFT and Polyphen-2 and the altered amino acid is not well conserved (Appendix 16 D), therefore most likely it's not the cause of the observed phenotype.

Homozygous or compound heterozygous mutations in *GCNT2* have been reported in families with autosomal recessive cataract (Borck, et al., 2012). Affected individuals in family CRCH136 carry a single copy of a 4bp frameshift deletion (p.(Asn388Argfs*20)) in this gene. Although inheritance in this family is consistent with autosomal recessive (Figure 4.7), a second mutation in *GCNT2* or in any other gene in the panel could not be identified in the proband. The affected individuals had bilateral dense central opacities, similar to those reported by Borck *et al*. (2012) in other families with a homozygous mutation in this gene and are also similar to those seen in family CRCH89 in which a homozygous *GCNT2* mutation was identified (Borck, et al., 2012). It is possible that the second mutation in family CRCH136 was not detectable by the methods employed in this study, which may include partial gene deletions or mutations affecting non-coding regions. The possibility that this mutation does not contribute to the disease in this family cannot be excluded. Furthermore one homozygous (p.(Arg385Cys)) mutation in this gene was detected in PCC10-183 from

Sri Lanka. The mutation is also present in the affected brother (Figure 4.17). Parents were reported to be unaffected which demonstrates an autosomal recessive pattern of inheritance as expected.

GALE or UDP (uracil-diphosphate)-galactose-4-prime-epimerase catalyses interconversion of UDP-galactose and UDP-glucose. There has been just one report of association between the deficiency of *GALE* and paediatric cataract in a 5.5 year old girl with autosomal recessive paediatric cataract (Schulpis, et al., 1993). Here for the first time we report an association between a missense variation (p.(Arg256Gly)) in this gene and paediatric cataract in a proband from Bhutan with paediatric cataract, amblyopia, and retinal dystrophy (patient E1).

The variant detected in *PAX6* in SR 11 cat was novel and the altered protein residue is conserved among different species (Appendix 16 C). There is high chance that the observed phenotype is associated with the variation in this proband, however obtaining more phenotypic information would be beneficial, as PAX6 is associated with syndromic forms of paediatric cataract.

Multiple variants detected in 2 screened populations were predicted to be non-pathogenic by both SIFT and Polyphen-2: p.(Ala379Glu) in *BFSP2* in CRVEEH85; p.(Arg513Gln) in CSA131 in *FYCO1*; p.(Glu304Lys) in *HSF4* in PCC 01-34; p.(Ala328Ser) in *PAX6* in PCC01-97A. In total, there were 7 variants detected in the familial probands which met the filtering criteria but did not segregate with the phenotype: p.(Asn296His) in *PVRL3* in CRCH139; p.(Pro119Arg) in *COL4A1* and p.(Val311Ala) in *NSDHL* both in PP-50 cat; p.(Gln845Glu) in *TDRD7* and p.(Ser300Arg) in *COL4A1* both in SR 12 cat; p.(Gly100Asp) in *CRYGD* and p.(Glu240Val) in *VIM* both in PCC 02-105. Consequently, none of these variants are considered to be the cause of cataract in mentioned families and alternative mutations

were identified in some families (CRCH139, CSA131, PP-50 cat and PCC 02-105) as described above.

For some of the probands included in this study such as CSA164, CSA146, and JM from the Australian cohort (Table 4.5), and PCC10-188, PCC01-97A and SR12 cat from South East Asia (Table 4.7), we identified more than one candidate gene associated with the phenotype. Accessing other affected and unaffected members from the mentioned proband's families would definitely help with identification of the potential causative variants (for example families CSA131 and CRCH139), however the possibility of paediatric cataract not being a monogenic condition couldn't be excluded.

4.6 Summary and conclusion

We identified likely causative variants in 42% of previously unsolved Australian familial cases and 40% of sporadic cases. Furthermore, we were able to determine the genetic causes in approximately 23% of paediatric cataract cases screened from South East Asia. It is likely that the percentage detected in Asian cohort would have been higher if more individuals from different families were available for segregation assessment. We have previously identified mutations in genes included in this panel in other Australian families in our repository (Burdon, et al., 2007; Burdon, et al., 2004a; Burdon, et al., 2004b; Craig, et al., 2003; Dave, et al., 2013; McLeod, et al., 2002; Reches, et al., 2007; Sharma, et al., 2008). When considered together with our earlier published work, these 51 genes account for more than 60% of familial paediatric cataract cases, a proportion comparable to that reported in a similar study of patients from the UK (Gillespie, et al., 2014) and another Australian cohort (Ma, et al., 2016). In total, in two Australian probands we identified 2 potential candidate genes associated with the observed phenotype. In Asian probands, six (20%) of families carried mutations in more than two candidate genes. Although recruiting more family members

and assessing them might eliminate some of these variants, here for the first time we propose the possibility of paediatric cataract being a polygenic disease. It is clear that this hypothesis needs to be evaluated using functional studies.

Although some degree of genotype/phenotype correlation is beginning to emerge for some paediatric cataract genes, the clinical evaluation of a patient is rarely sufficient to establish which genes are most likely involved in order to initiate specific genetic testing. Thus it can be concluded that gene panel testing as it has been shown in previous studies (Gillespie, et al., 2014) and here in an Australian cohort, is an efficient way forward for rapidly determining the genetic cause of heterogeneous diseases such as paediatric cataract.

Chapter 5: Identification of novel genes and mechanisms underlying paediatric cataract using whole exome sequencing

5.1 Introduction

Mendelian disease study has been one of the earliest and most important successes of next generation sequencing and it has been successfully applied in identifying disease causing genes (Bamshad, et al., 2011) and mutations for paediatric cataract (Zhao, et al., 2015) (Reis, et al., 2013). Above 900 novel genes were identified to be associated with Mendelian phenotypes during 2011-2015 majority via WES technologies (Chong, et al., 2015) which have accelerated understanding mechanisms underlying these disorders.

As shown in Chapter 4, we were able to detect the genetic causes for above 60% (including previously solved and published cases from this cohort) of familial cases despite very limited numbers of affected and unaffected individuals available from these families. Therefore, we hypothesized that more congenital cataract associated genes are yet to be discovered. We aimed to undertake gene discovery in congenital cataract families without mutations in known congenital cataract causing genes using exome sequencing. The genetic causes of cataract in many of the large families with multiple affected individuals in our repository have been identified using traditional linkage analysis. The remaining families were screened for known congenital cataract genes as described. Their discovery eventually could provide new insight into the understanding of normal lens biology and cataract pathophysiology. The whole exome sequencing approach was taken to detect any potential candidate gene or mechanism responsible for cataract in selected families which do not have a mutation identified in known genes.

5.2 Aim

1: To perform exome sequencing and discover novel paediatric associated genes in four families with no previously identified mutations in known paediatric cataract associated genes.

5.3 Methods

5.3.1 Exome sequencing

Genomic DNA was extracted from blood or saliva samples of selected individuals of families CSA106, CRCH11, CSA92 and CRCH26 as described in Chapter 2, section 2.4. All the available individuals (at the time when the research was conducted) from families CRCH26 and CRCH11 were exome sequenced. A subset of informative individuals were chosen from family CSA106 for exome sequencing. All selected samples were subjected to exome capture (Agilent SureSelect v4) and paired-end sequencing on an Illumina HiSeq 2000 by an external contractor (Macrogen Inc, Seoul, South Korea). The quality control and read summaries were provided by the contractor and inspected to ensure data reached minimum standards (Appendix 17).

Reads were mapped to the human reference genome (hg19) using the Burrows-Wheeler Aligner (v0.7.10). Variants were called using SAMtools (v1.0) (Li, et al., 2009) and annotated against RefSeq transcripts using ANNOVAR (2014Nov12) using in-house pipelines at Macrogen.

Selected family members of CSA92 were exome sequenced through our collaborator at State Key Laboratory of Medical Genetics, Central South University, Changsha, China. Paired-end libraries were sequenced on the Illumina GAII platform and read summaries were provided (Appendix 17). Variant call format (VCF) files were annotated using SeattleSeq Variation Annotation website against dbSNP137 (http://snp.gs.washington.edu/SeattleSeqAnnotation137/).

5.3.2 Variant filtering strategy

The filtering strategy described in Chapter 2, section 2.8. The SNPs with the quality of 20 and above were selected.

5.3.3 Processing exome sequencing data for linkage analysis

The exome sequencing variant dataset was prepared to perform linkage analysis using Linkdatagen scripts (Smith, et al., 2011) and MERLIN (Abecasis, et al., 2002). Linkdatagen is a PERL script that generates linkage style files for programs like MERLIN from massively parallel sequencing data like whole exome sequence data. Merlin is a linkage analysis tool suitable for parametric and non-parametric linkage analysis of small to moderately large families. The Tizard High-performance computing (HPC) system at eResearch SA was used for temporary data processing, storage and analysis. The SSH client, PuTTy (http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html) was used for connection to eResearch SA through windows computers.

The HapMap reference population (Annotation for up to 4,071,899 SNPs for the four HapMap Phase II populations (CHB, CEU, JPT, and YRI)) was used to obtain population allele frequencies for linkage analysis using the recommended (by Linkdatagen) annotation HapMap II file (annotHapMap2.txt) downloaded from linkdatagen website. (http://bioinf.wehi.edu.au/software/linkdatagen/). Human reference genome (hg19) was obtained via the UCSC database (hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/chromFa.tar.gz) and extracted using the following commands:

tar -zxvf chromFa.tar.gz

cat chr*.fa > hg19.fa

Variants were re-called from the BAM file using SamTools (Li, et al., 2009) (v0.1.19) and output as Variant Call Format (VCF) files. The genotypes at the location of HapMap SNPs were detected in each sample sequenced using the following commands:

samtools mpileup -d10000 -q13 -Q13 -gf hg19.fa -l annotHapMap2L.txt CatarctSample.bam | bcftools view -cg -t0.5 - > sample.HM.vcf

Perl (5.16.3) was used to run vcf2linkdatagen.pl script (http://bioinf.wehi.edu.au/software/linkdatagen/) to modify variant calls from VCF to BRLMM format which contains both ID and SNPs columns for each individual. A text file listing the path to each VCF file on a separate line was created for each family and the name of file *(MyVCFlist.txt)* was specified using the -idlist argument. This allowed obtaining the BRLMM format for multiple VCF files in parallel. The BRLMM file was generated using the following command:

vcf2linkdatagen.pl -annotfile annotHapMap2.txt -pop CEU -mindepth 10 -missingness 0 -idlist MyVCFlist.txt > MySNPs.brlmm

5.3.4 Linkage analysis using exome sequence data

We performed linkage analysis on families CRCH26 and CSA106 to identify the potential linkage region. The linkage analysis of family CSA106 was conducted by Dr.Owen Siggs (Department of Ophthalmology, Flinders University). Parametric linkage analysis was then performed using MERLIN (v1.1.2) under a rare dominant model for CSA106 and under a rare recessive model for CRCH26 using the following parameters:

0.001 0.0001, 1.0, 1.0 Rare_dominant

0.001 0.0001, 0.0001, 1.0 Rare_recessive

The first parameter of 0.001 is the disease allele frequency in the general population. The next set of values are the penetrance of the disease for carrying 0 (homozygous-wild type), 1 (heterozygous) or 2 (homozygous disease allele) copies of the disease allele.

5.3.5 Copy number variation (CNV) analysis

This analysis was conducted by Dr. Owen Siggs (Department of Ophthalmology, Flinders University). In brief, coverage depth across the critical region was extracted from exome BAM files using SamTools (0.1.19) (Li, et al., 2009). For copy number variant analysis using CoNIFER (v0.2.2), the same interval was analysed in 343 population-matched control exomes (including 11 from family CSA106) using the following parameters: SVD 5, ZRPKM 1.5.

5.3.6 Quantitative PCR to confirm *CRYBB1* partial duplication

Commercial TaqMan Copy Number Assays for duplicated (Hs04088405_cn - chr22:27006444 (hg19), within intron 3) and non-duplicated (Hs00054226_cn - chr22:26995522 (hg19), within exon 6) regions of *CRYBB1* were utilized (Life Technologies, Carlsbad, California, USA). All available CSA106 family members were tested for partial duplication in *CRYBB1* using genomic DNA according to the manufacturer's protocol. Briefly, the segment of *CRYBB1* gene was amplified in 4 replicates for each CSA106 DNA sample. The experiment was performed using the StepOne Plus real-time polymerase chain reaction instrument using an endogenous reference gene known to be present in 2 copies in a diploid genome (TaqMan Copy Number Reference Assay, human, RNase P) along with the *CRYBB1* assay. The CopyCaller 2.0 software (Life Technologies) was used to predict the copy number of the target genomic DNA. A total of 118 controls were screened using both assays for duplicated and non-duplicated regions (in duplicates). We also screened 47 congenital cataract probands with unknown/unidentified genetic cause for the duplicated region.

5.3.7 Protein extraction from surgical lens specimen

At the time of cataract surgery at the age of 13 years, lens material of CSA106.06 was collected in balanced salt solution and stored at −80°C with 1mM EDTA. Normal human lens was obtained from an 18 year-old deceased donor (the Eye Bank of South Australia, FMC) for use as control. Lenses were homogenized in 2 mL extraction buffer containing 50 mM imidazole (pH 7), 50 mM NaCl, 2 mM 6-aminohexanoic acid, 1 mM EDTA, and protease inhibitor cocktail (Roche Diagnostics, NSW, Australia) and ultracentrifuged at 150,000×g (5810 R centrifuge, Eppendorf) for 30 min at 4°C to collect cleared supernatants as previously described (Laurie, et al., 2013) .The EZQ Protein Quantitation method (Life Technologies) was used to determine the concentration of protein. Due to low protein concentration of lens sample, the soluble protein fraction was acetone precipitated according to the Thermo Scientific protocol (The Thermo Fisher Scientific, Australia). Four times the sample volume of cold (-20°C) acetone was added to the sample then vortexed and incubated for 60 minutes at -20°C. The sample was centrifuged for 10 minutes at 13,000-15,000 × g. The pellet was left at room temperature for 30 minutes for drying and resuspended in the extraction buffer.

5.3.8 Denaturing protein gel electrophoresis and western blotting

Twenty microgram of total soluble and insoluble protein fraction from each lens was size fractionated by SDS-PAGE using a 12% polyacrylamide gel. The precision plus protein standards (Biorad, Australia) were used for size detection and comparison. The gel was prepared according to the Laemelli method (Laemmli, 1970) and electrophresis performed at 250 V. For Western blotting, after SDS-PAGE, the proteins were transferred on to Hybond-C Extra nylon membrane (GE Healthcare Australia Pty Ltd., NSW, Australia) using a Trans-Blot Turbo Transfer Starter System (BioRad

Laboratories, NSW, Australia) at 30 V and 1A overnight in Western transfer buffer (supplier or components). The membrane was blocked in blocking buffer containing 5% skim milk made in 1× TBST (Tris Buffered Saline and Tween 20) for one hour and hybridised with the mouse anti-CRYBB1 (1:400, Sigma-Aldrich, USA) primary antibody diluted in blocking buffer, for 1 hour. After three washes in 1× TBST for 10 minutes each, the blot was hybridised with the anti-mouse IgG conjugated with horse-radish peroxidase (HRP) (1:1000, Jackson ImmunoResearch, USA) secondary antibody diluted in blocking buffer, for 1 hour. The blots were washed three times in 1× TBST for 10 minutes each and were developed using Clarity Western ECL Blotting Substrate (Biorad, Australia) or Amersham ECL Prime western blotting reagent (GE Healthcare Australia Pty Ltd, NSW, Australia). The signal was imaged using ImageQuant LAS 4000 Imager (GE Healthcare Australia, NSW). The images were edited using Microsoft PowerPoint. The same membrane was striped in stripping solution (100mM β-Mercaptoethanol, 2% SDS and 62.5 mM Tris.HCH (PH:7)) at 50°C for 30 minutes and was washed twice, 10 minutes each, with 1x TBST and blocked in 5% skin milk prepared in TBST. Following the above described procedure, the stripped membrane was hybridized with the rabbit anti-CRYBA4 (1:200, abcam, Australia) or sheep anti-CRYAA (1:1,000; Flinders University Antibody Production Facility, South Australia) primary antibody and then hybridized with the anti-rabbit IgG conjugated with horse-radish peroxidase (HRP) (1:1000, Jackson ImmunoResearch, USA) or anti-sheep IgG conjugated with horse-radish peroxidase (HRP)(1:1000, Jackson ImmunoResearch, USA) secondary antibody, respectively.

5.3.9 Validation of the variants detected through exome sequencing

The general procedures for validation by PCR and Sanger sequencing are given in Chapter 2, sections 2.5 and 2.6. Specific primer set used for validating variants in families CSA92, CRCH11 are given in Appendix 18.

5.3.10 Screening of normal population controls for identified potential causative mutations in CSA92, CRCH11 and CSA106

The variants in *HTR1F* detected in family CSA92 was screened in 282 normal controls using a Custom TaqMan SNP Genotyping Assay, CSA92.02 as positive heterozygous control and water as negative control (see Chapter 2, section 2.7 for detailed procedures). Other variants detected in family CSA92 (in *HSF4* and *BFSP1*) and in CRCH11 (in *NOL9*) were screened in our local normal cohort of 332 and 326 individuals using Sequenom MassArray as described in Chapter 2, section 2.7. A total of 118 controls were screened using both assays for duplicated and non-duplicated regions (in duplicates) detected in CSA106.

Different numbers of controls were used in each analysis as the genotyping for different variants was performed at different stages of the research, altering the number of available control individuals. The genotyping using sequenom massArray was performed at two different stages resulting in different numbers of controls available. For the CNV analysis, it was decided to directly screen only 100 normal controls by Taqman assay as data were also available from exome sequencing data of 343 population-matched controls (Chapter 2, section 2.3).

5.3.11 Sequencing *NOL9* and *HTR1F* gene in unsolved congenital cataract cohort

The coding region and 5' untranslated region of *HTR1F* gene were screened using Sanger sequencing in 43 previously unsolved congenital cataract patients following the procedure described in Chapter 2, sections 2.5 and 2.6. The list of primers is given in Appendix 19. The coding (except exon 2 which failed to amplify even after trying some optimization of the PCR by the external contractor) and 5' untranslated region of *NOL9* gene was screened in 47 previously unsolved congenital cataract probands by an external contractor (AGRF, Brisbane, Australia). The PCR primer list is given in Appendix 20.

5.3.12 Predicted effects of identified mutations on protein structure

Effects of R289W mutation on HTR1F and p.(*703R5*) on NOL9 structures were predicted using PSIPRED (Buchan, et al., 2013) (http://bioinf.cs.ucl.ac.uk/psipred/).

Peptide hydrophobicity analysis was conducted using peptide-2 tool (http://www.peptide2.com/N_peptide_hydrophobicity_hydrophilicity.php). The 3D protein structure evaluation in mutant and wild type was performed using SWISS-MODEL (http://swissmodel.expasy.org/) (Arnold, et al., 2006).

5.4 Results

5.4.1 Family CSA92

There are 4 unaffected individuals and 4 affected individuals in the family with no report of consanguinity. The cataract phenotype in this family is variable (as shown in Table 5.1) and ranges from very faint fetal lamellar (undiagnosed) cataract to severe

cataract with intraocular lens (IOL) implantation surgery in two affected individuals. We performed whole exome sequencing on CSA92.01, CSA92.02, CSA92.03, CSA92.04 and CSA92.05 (Figure 5.1). The coverage and mapping data are shown in Appendix 17. Over 93 % of the target bases had at least 8 fold coverage. Average throughput depth of target regions ranged from 64 to 108 fold in sequenced individuals. The family history was most consistent with an autosomal dominant pattern of inheritance (Figure 5.1 A), therefore we looked for novel protein changing heterozygous variants shared among the mother and the 2 affected sons and absent in the unaffected CSA92.05. The cataract phenotype of CSA92.07 is shown in Figure 5.1 B. Seven filtered variants were selected for sequencing in the affected proband for validation of the exome sequence results (Table 5.2). Two variants in *KIR2DS4* and *ZEBD3* appeared to be false positive signal as they were not confirmed by Sanger sequencing. One variant (in *ZNF717*) was excluded from the study as it was detected in a highly variable gene and is included in a list of recommended genes to exclude from exome sequencing analysis (Fuentes Fajardo, et al., 2012). The remaining four variants were sequenced in the other family members to assess segregation of the variant with the phenotype. Only one variant in *HTR1F (*c.865C>T) showed consistent segregation with the phenotype (Figure 5.1, C) being present in all affected individuals. The mutation in *HTR1F* is not fully penetrant as CSA92.06, the unaffected member of the family , also carries the mutation (Figure 5.1 C).

Furthermore we detected 2 rare SNPs in *BFSP1* (c.736A>G, p.(Thr246Ala); rs143865632; MAF of 0.2% in dbSNP and 0.1% in ExAC) and *HSF4* (c.636G>T, p.(Met212Ile); rs199742128; MAF of 0.2 in dbSNP 138 and 0.1 in ExAC) in the Ion Torrent PGM targeted sequencing data (described in Chapter 4). These variants were confirmed in the exome data and were sequenced directly in remaining family members to fully assess segregation; however neither segregated completely (Figure 5.1). The

affected CSA92.04 does not carry the variants in *BFSP1* and the affected mother (CSA92.02) does not carry the mutation in *HSF4*.

The variant in *HTR1F* was absent in 282 normal Australian controls (Chapter 2) however it is present in the ExAC database as an extremely low frequency allele (MAF= 0.0000083) in 1 out of 120,792 chromosomes. The variants in *BFSP1* and *HSF4* were assessed in exome sequencing data from 250 normal and non-congenital cataract samples. The *HSF4* variant was present with MAF of 0.9% in 250 exome sequences available, while the variation in *BFSP1* was absent in those controls. Furthermore, screening our local control population revealed that the *BFSP1* variant was not present in those controls, however the *HSF4* variant was present with a MAF of 0.4 %.

Table 5.1. Cataract phenotype in family CSA92

Proband	Cataract phenotype	Teeth condition/other symptoms	Diagnosis Age	Current Age Age recruited
92.01	Normal eye exam	-	-	57-47
92.02	Very faint fetal lamellar (undiagnosed)	Top incisors slight rotation outward , diastema but healthy looking	42	52-42
92.03	Bilateral IOL	Very poor enamel but shape ok/ epilepsy	4	32-22
92.04	Fetal nuclear lamellar some central opacity	Slightly crowded but not bad	-	30-20
92.05	Couple of salt grains, slight cortical streaking but probably normal	Very crowded, small pushed back lateral incisor	-	29-19
92.06	Normal eye exam	Good condition	-	27-17
92.07	Right eye: tiny dots no suture; left eye IOL,	Central incisors rotated outward	4	24-14
92.08	Trace, probably normal	Slight diastema , tiny serrations but probably normal	-	20-10

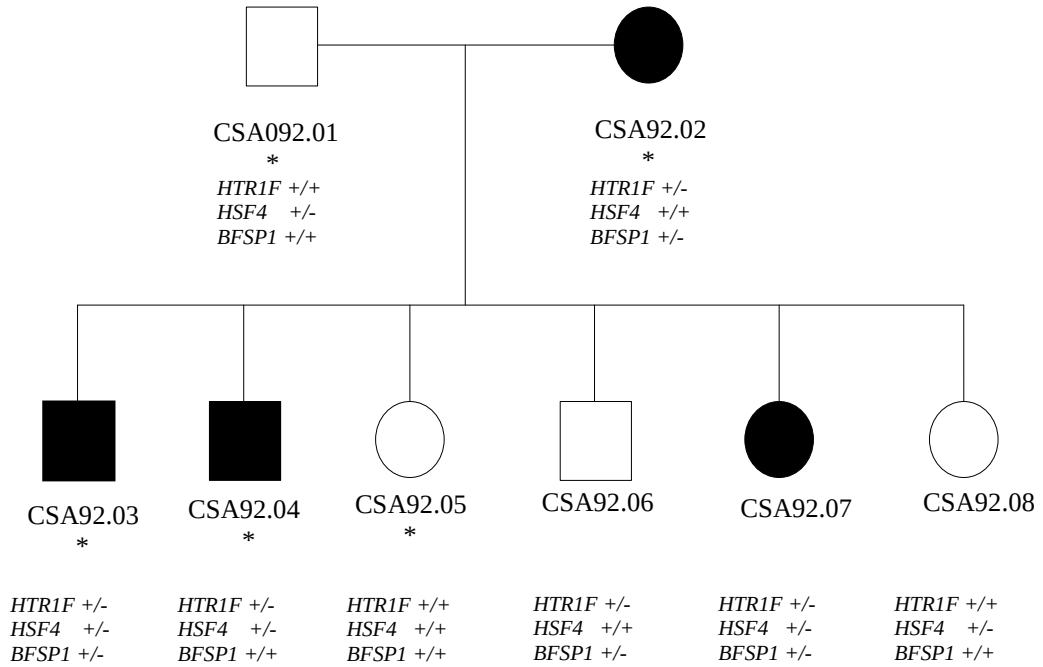
IOL: intraocular lens implant. “-“indicates no information is available.

Table 5.2. List of heterozygous variants in CSA92 obtained through exome sequencing and their segregation with cataract in the family.

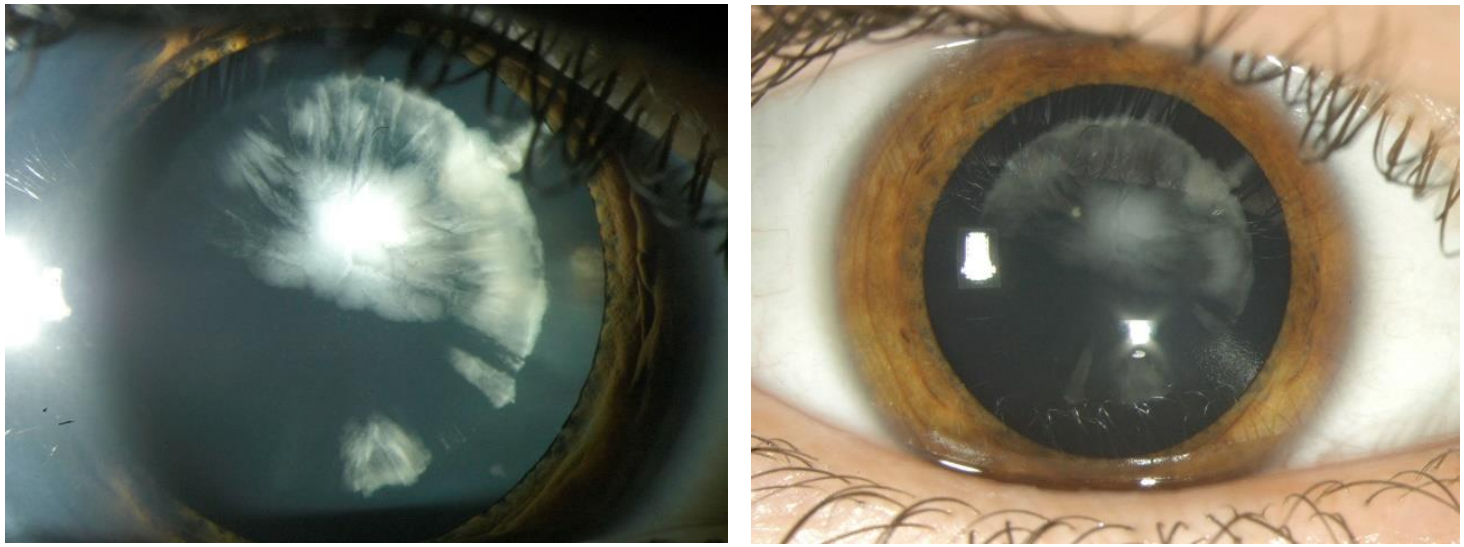
Position	Reference	Sample	Accession	Function	DNA Change	Protein Change	Poly-Phen Prediction	Gene	Validated in proband	Segregated
chr3:75786933	G	A/G	NM_001128223.1	Missense	c.935G>A	T614I	probably-damaging	ZNF717	Yes	-
chr3:88040764	C	C/T	NM_000866.3	Missense	c.865C>T	R289W	probably-damaging	HTR1F	Yes	Yes
chr5:76373168	G	C/G	NM_032367.2	Missense	c.526C>G	R176G	possibly-damaging	ZBED3	No	-
chr11:83674028	T	C/T	NM_001142699.1	Missense	c.925A>G	N309D	probably-damaging	DLG2	Yes	No
chr11:88330507	T	A/T	NM_000842.3	Missense	c.1408A>T	N470Y	probably-damaging	GRM5	Yes	No
chr11:116649768	T	A/T	NM_003904.3	Missense	c.1253A>T	Y418F	probably-damaging	ZNF259	Yes	No
chr19:55358681	T	A/T	NM_012314.3	Missense	c.736A>T	V253D	possibly-damaging	KIR2DS4	No	-

The variants shared between affected mother (CSA92.02) and affected sons (CSA92.03, CSA92.04) and not present in unaffected CSA92.05 were selected through exome sequence data analysis. The filtered variants were assessed for segregation in other family members. The *HTR1F* variation was the only variant segregating with the phenotype in the family (presents in all affected individuals and unaffected CSA92.06).

A


CSA092.01
*
HTR1F +/+
HSF4 +/-
BFSP1 +/+
CSA92.02
*
HTR1F +/-
HSF4 +/+
BFSP1 +/-
CSA92.03
*
HTR1F +/-
HSF4 +/-
BFSP1 +/-
CSA92.04
*
HTR1F +/-
HSF4 +/-
BFSP1 +/+
CSA92.05
*
HTR1F +/+
HSF4 +/+
BFSP1 +/+
CSA92.06
HTR1F +/-
HSF4 +/+
BFSP1 +/-
CSA92.07
HTR1F +/-
HSF4 +/-
BFSP1 +/-
CSA92.08
HTR1F +/+
HSF4 +/-
BFSP1 +/+

B



C

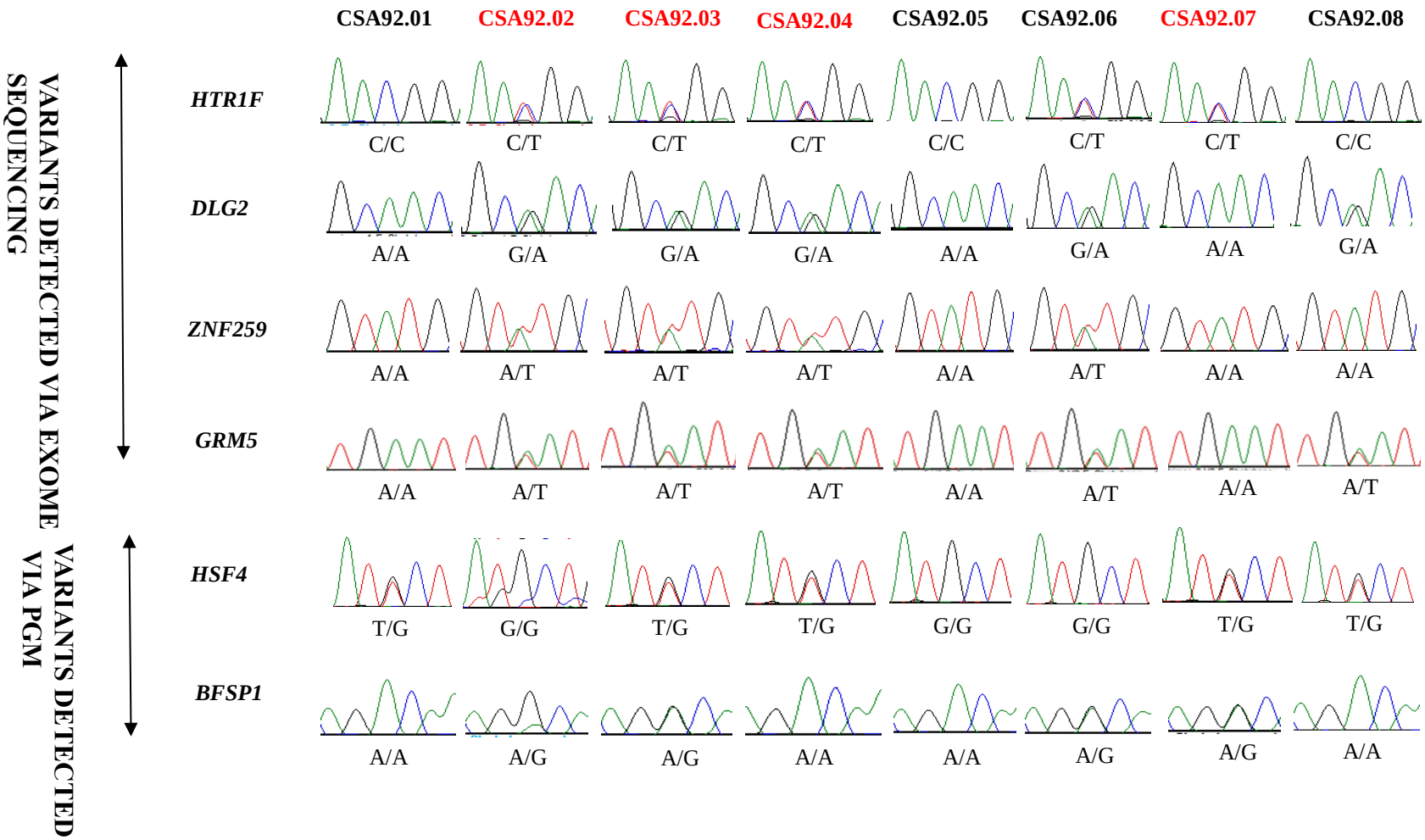

CSA92.01
CSA92.02
CSA92.03
CSA92.04
CSA92.05
CSA92.06
CSA92.07
CSA92.08
VARIANTS DETECTED VIA EXOME SEQUENCING
VARIANTS DETECTED VIA PGM
HTR1F
C/C C/T C/T C/T C/C C/T C/T C/C
DLG2
A/A G/A G/A G/A A/A G/A A/A G/A
ZNF259
A/A A/T A/T A/T A/A A/T A/A A/A
GRM5
A/A A/T A/T A/T A/A A/T A/A A/T
HSF4
T/G G/G T/G T/G G/G G/G T/G T/G
BFSP1
A/A A/G A/G A/A A/A A/G A/G A/A

Figure 5.1. Variants detected in family CSA92. Mutations in 4 novel genes and 2 previously reported congenital cataract associated genes were assessed for segregation. A: Pedigree of family CSA92 showing the segregation of detected variants in *HTR1F*, *BFSP1* and *HSF4*. "+"indicates the wild type allele while "–"indicated the mutated allele. The family members with exome sequence data are indicated by the asterisk. Affected members are shown by black symbols and unaffected members by white symbols. Circles indicate female and squares indicate males. B: The cataract phenotype in CSA92.07.C: chromatograms show the sequence of novel variations in novel genes. *HTR1F* (c.865C>T); *DLG2* (c.925A>G); *ZNF259* (c.1253A>T) and *GRM5* (c.1408A>T). The chromatogram also shows the segregation analysis of two variants detected via PGM: *HSF4* (c.636G>T) and *BFSP1* (c.736A>G). The only segregating variant (reduced penetrance) is in *HTR1F*. The affected members are indicated in red text.

PSIPRED predicted codon 289 is located in an α-helix of the protein. The substitution of the Arginine at this location for the mutant tryptophan doesn't change the prediction of the helix, however it slightly changes its location (Figure 5.2 A and B). Furthermore, the mutation changes the size of α-helix and its adjacent β-sheet. The protein is conserved among different species which is indicative of its evolutionary significance (Figure 5.3). Performing peptide hydrophobicity analysis using peptide-2 tool, the hydrophobicity of mutant protein (46.4%) is ot significantly different from wildtype (46.1%). According to Uniprot domain prediction the mutated residue is close to the junction of cytoplasmic and transmembrane domains and so may affect membrane topology of the protein and hence its function.

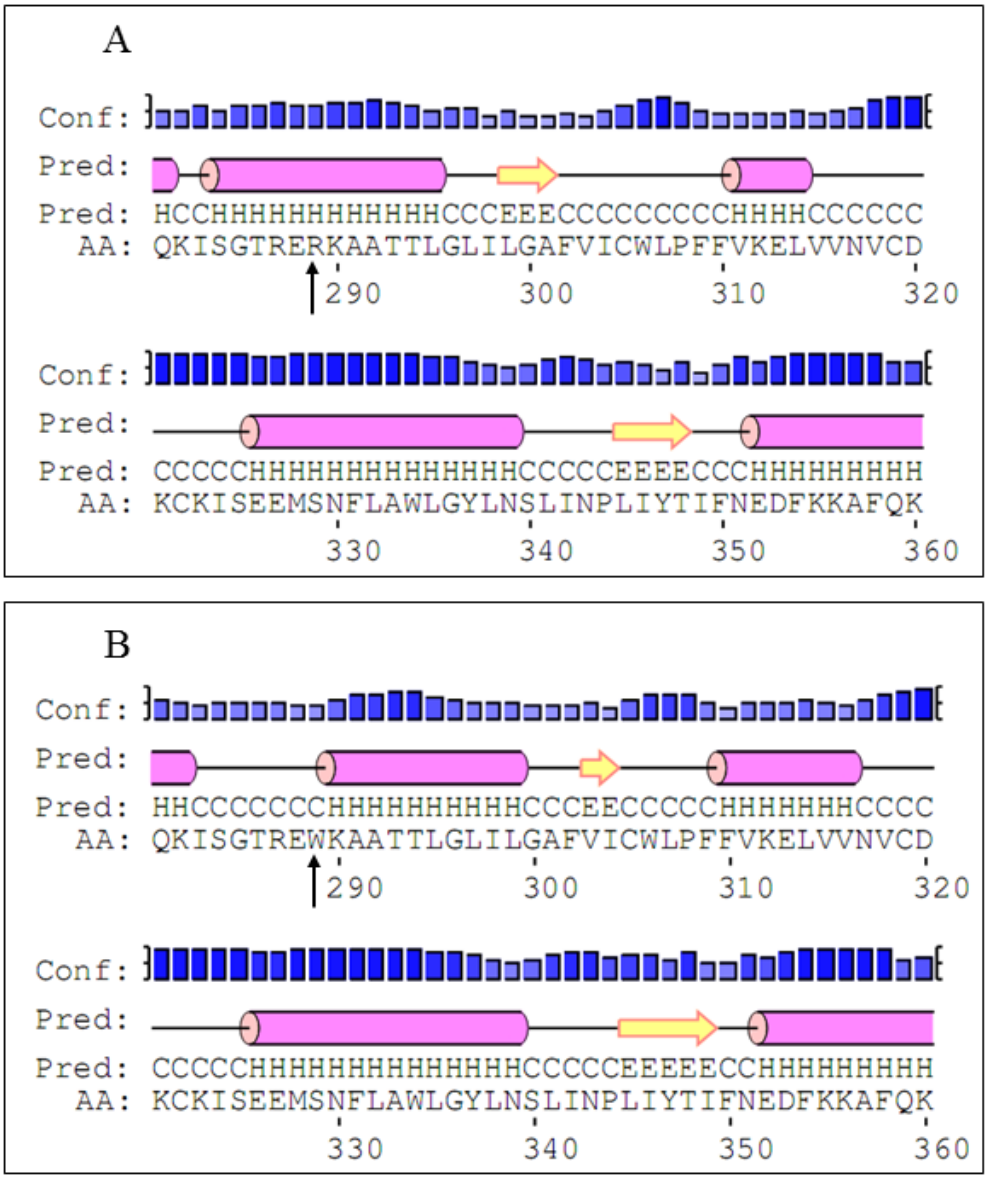


Figure 5.2. Prediction of the secondary structure of human HTR1F protein with PSIRED. A. Wild type B. R289W mutation. The altered amino acid is marked with a black arrow. Yellow arrows represent β-sheet. Pink cyliders represent α-helix. The blue bars show the confidence in the prediction (on a scale of 1-10).

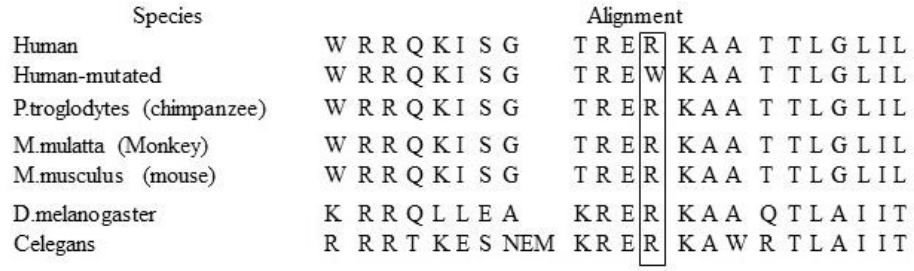


Figure 5.3. The conservation of residue 289 (Arginine) in HTR1F protein across different species. The altered amino acid at position 289 is boxed.

5.4.2 Family CSA106

We identified a 6-generation autosomal dominant congenital cataract pedigree (Figure 5.4 A), with affected family members diagnosed between birth and 10 years of age (Table 5.3). The cataract phenotype is very variable in the family and is unique (Figure 5.4 B). The phenotype in CSA 106.06 is described as sutural cataract involving the nucleus. CSA106.02, CSA106.03 and CSA106.07 were diagnosed at birth. CSA 106.02 and CSA106.03 went through eye surgery at the age of 44 and 3 respectively. CSA106.01, CSA106.04, CSA106.06 and CSA106.14 were diagnosed at the age of 5, 3, 4, and 10 respectively. CSA106.01 and CSA106.14 both had surgery in the right eye at an older age (76 and 51) where as CSA106.04 and CSA106.06 had surgery at the ages of 18 and 7 respectively.

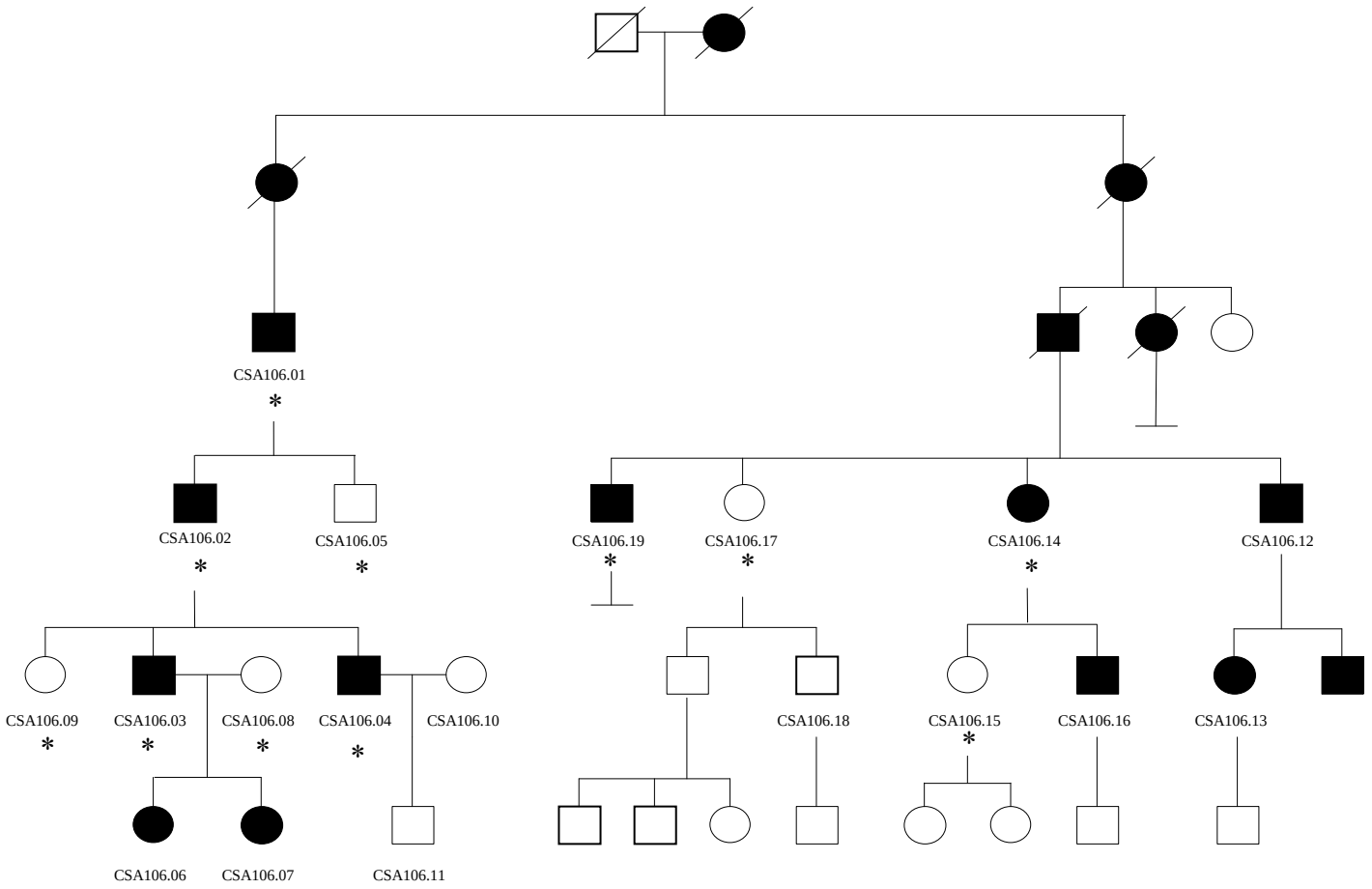
Exome data processing and Linkage analysis

We sequenced the exomes of 11 family members (6 affected, 5 unaffected as indicated by an asterisks in Figure 5.4). Based on the autosomal dominant pattern of inheritance we looked for the heterozygous variants shared between the affected members and not present in unaffected ones, however no segregating variant was detected.

Parametric linkage analysis was performed under a rare dominant inheritance model and revealed a peak LOD score of 3.3 on chromosome 22 (hg19 chr22:26422980-

29414001) (Figure 5.5), within which lay 17 protein-coding genes, 9 non coding RNA and 7 pseudogenes including the known congenital cataract genes *CRYBB1* and *CRYBA4*. As both PGM and exome sequencing technologies failed to detect any potential variants in these genes, we investigated copy number variation (CNV) as an alternative mechanism underlying the observed phenotype.

A


CSA106.01
CSA106.02
CSA106.05
CSA106.19
CSA106.17
CSA106.14
CSA106.12
CSA106.09
CSA106.03
CSA106.08
CSA106.04
CSA106.10
CSA106.18
CSA106.15
CSA106.16
CSA106.13
CSA106.06
CSA106.07
CSA106.11

B

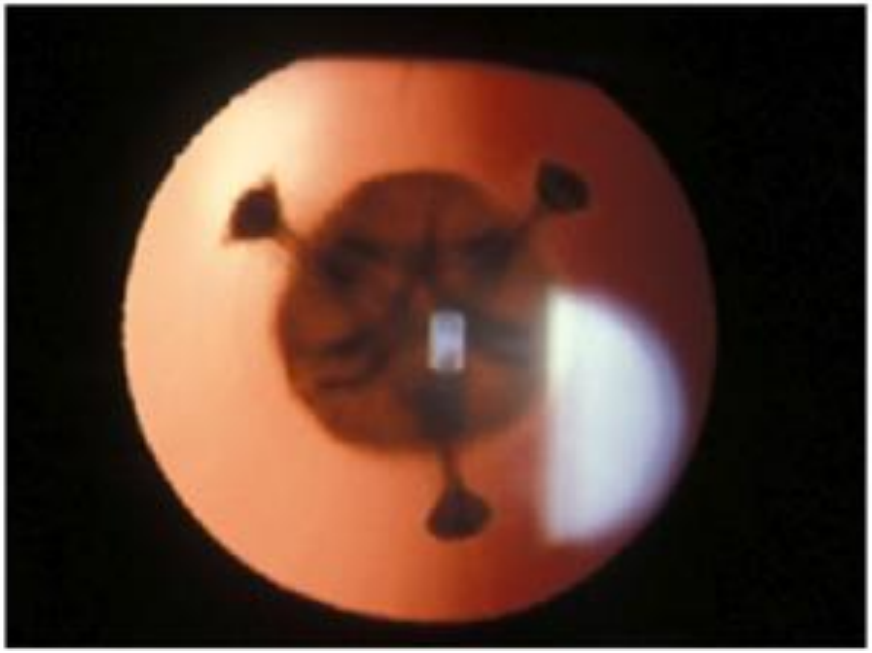


Figure 5.4. The pedigree and the cataract phenotype in family CSA106. (A) Family CSA106 pedigree with an autosomal dominant congenital cataract, indicating affected (black) and unaffected (white) members. Circles indicate females and squares indicate males. The exome sequenced individuals are marked with an asterisk. (B) Photograph of the lens of CSA106.06, indicating a sutural cataract involving the nucleus.

Table 5.3. Clinical details of CSA106 family affected members.

ID	Age diagnosed	Age of surgery RE	Age of surgery LE	Cataract phenotype
CSA106.01	5	76	76	
CSA106.02	0	44	44	
CSA106.03	0	3	22	
CSA106.04	3	18	19	
CSA106.06	4	7	13	Nuclear, sutural
CSA106.07	0	NA	NA	Faint nuclear with sutural involvement
CSA106.12	NA	NA	NA	
CSA106.13	NA	NA	NA	
CSA106.14	10	51	10	
CSA106.16	NA	24	NA	
CSA106.19	NA	NA	NA	

LE, left eye; RE, right eye; NA, not available.

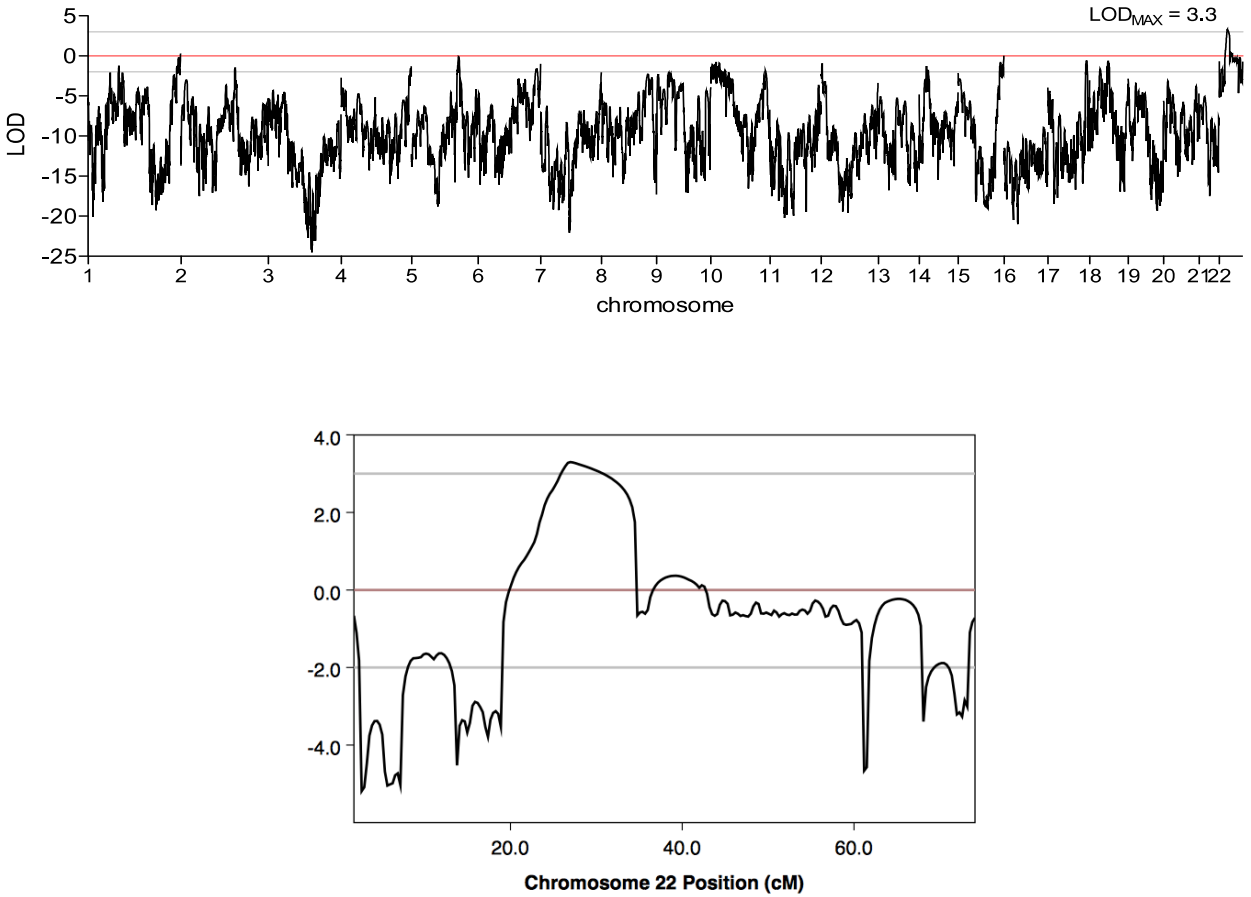


Figure 5.5. Parametric linkage analysis of family CSA106 performed under a rare dominant inheritance model. The linkage analysis revealed a peak with LOD score of 3.3 on chromosome 22 with linkage interval between rs2236005 and rs2347790 (22:26422980-29414001)

Partial duplication of the *CRYBB1-CRYBA4* locus

The coverage depth across the linkage interval was investigated using SAMtools which revealed an increase in coverage depth at the *CRYBB1* and *CRYBA4* locus in all affected individuals (Figure 5.6). This indicated the presence of a copy number variation. Mean coverage depth analysis showed that all five exons of *CRYBA4* appeared to be duplicated and only the first five exons of *CRYBB1* (*CRYBB1* has a total of 6 exons) had been duplicated (Figure 5.6).

CRYBB1 partial copy number variation was validated in the family using quantitative polymerase chain reaction assays (Figure 5.7). All the family members have 2 copies of the non-duplicated exon 6 region of *CRYBB1* (Figure 5.7A) while the affected members had 3 copies of *CRYBB1* in the duplicated region (Figure 5.7B). No CNV was detected in *CRYBB1* or *CRYBA4* in any screened controls (total of 118) or congenital cataracts probands with unknown genetic cause (Appendix 21). While complete duplication of *CRYBB1* (or *CRYBA4*) could be expected to be benign, a partially duplicated *CRYBB1* may act as a gain of function allele.

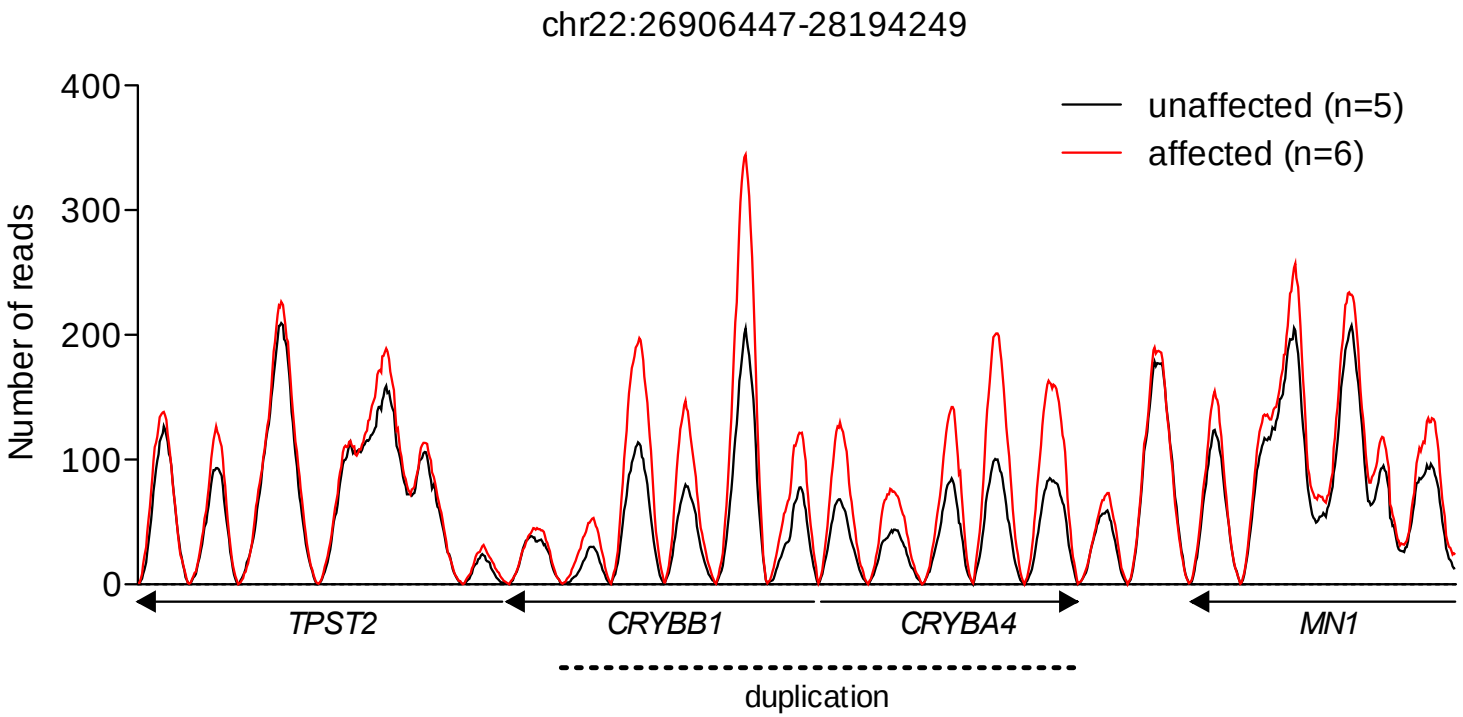


Figure 5.6. Partial and complete duplication of *CRYBB1* and *CRYBA4* in CSA106.The black line represents the coverage of unaffected individuals while the red represents the mean coverage of affected individuals.

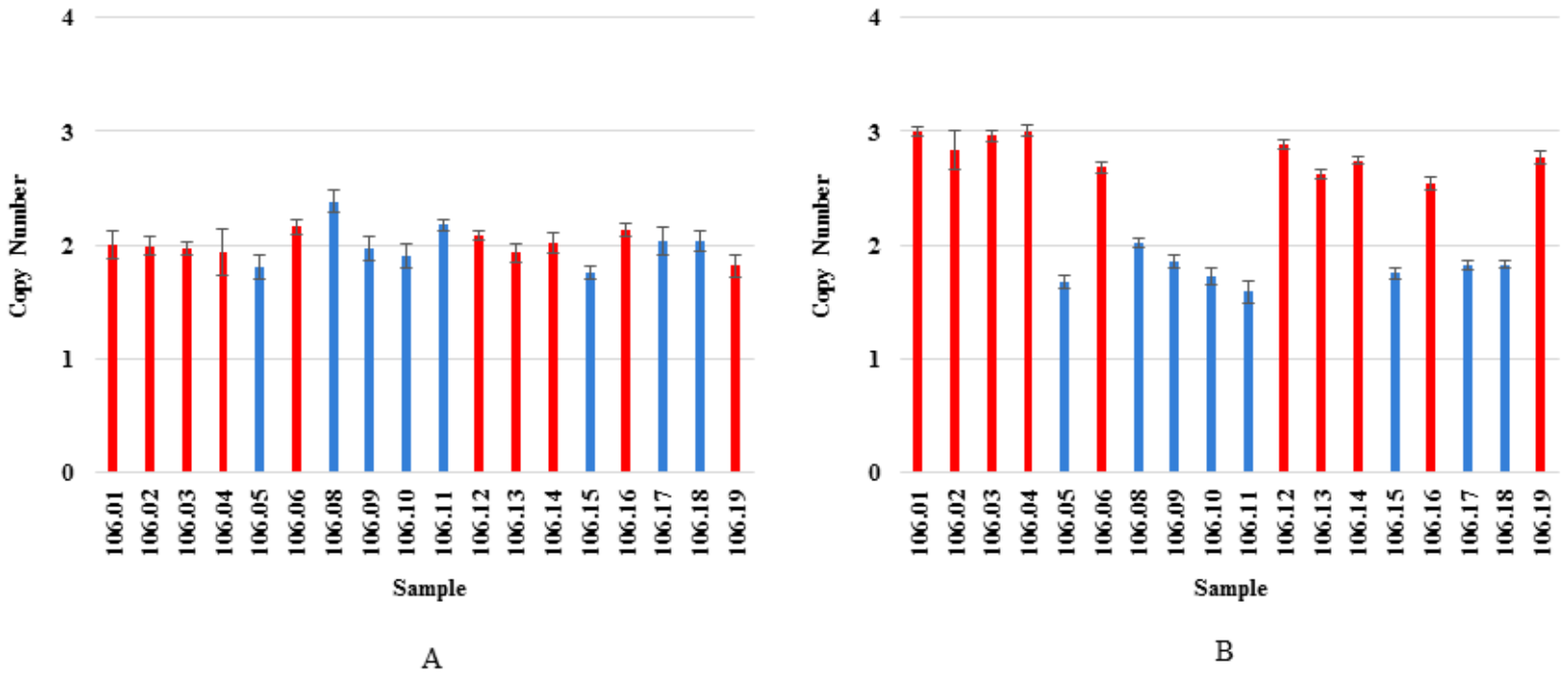


Figure 5.7. Assessment of *CRYBB1* gene dosage in non-duplicated (A) and duplicated (B) regions by quantitative polymerase chain reaction in family CSA106. The Y-axis shows the number of copies of the *CRYBB1* gene that were detected in each sample. The x-axis shows the family member ID. The normal dosage of 2 copies of *CRYBB1* was detected in the unaffected family members (blue bars) while the affected members have 3 copies in duplicated region (B) of the gene (red bars). In non-duplicated region (A) of *CRYBB1* (exon 6), all the members have 2 copies. The segment of the *CRYBB1* gene was amplified in 4 replicates for each DNA sample. The error bars on the graph indicate the minimum and maximum copy number (CN) calculated for the sample replicate group.

Western blot analysis

Western blotting was conducted to confirm the detected CNV variation in family CSA106 at protein level. Whole lens soluble proteins of a cataractous lens from CSA106.06 and an aged match normal lens were subjected to semi-quantitative western blotting under denaturing conditions for detecting CRYBB1 and CRYBA4 proteins. Equal amount of 20 micrograms proteins were loaded in each lane. CRYAA antibody was used as loading control. The membrane was hybridized with CRYBB1 antibody followed by CRYBA4 and CRYAA hybridization. Western blotting with the anti-CRYAA antibody indicated equivalent loading between the cataract and control samples. The anti-CRYBA4 and anti-CRYBB1 antibodies detected bands of 22 kDa and 28 kDa, respectively corresponding to the expected size of a monomer of each protein in both the samples (Figure 5.8). An additional band was detected with the anti-CRYBA4 antibody corresponding to a CRYBA4 dimer (44kDa), although no other bands were apparent. We did not detect any additional anti-CRYBB1reactive bands of the expected size of the truncated protein (~21 kD) in the cataract lens sample.

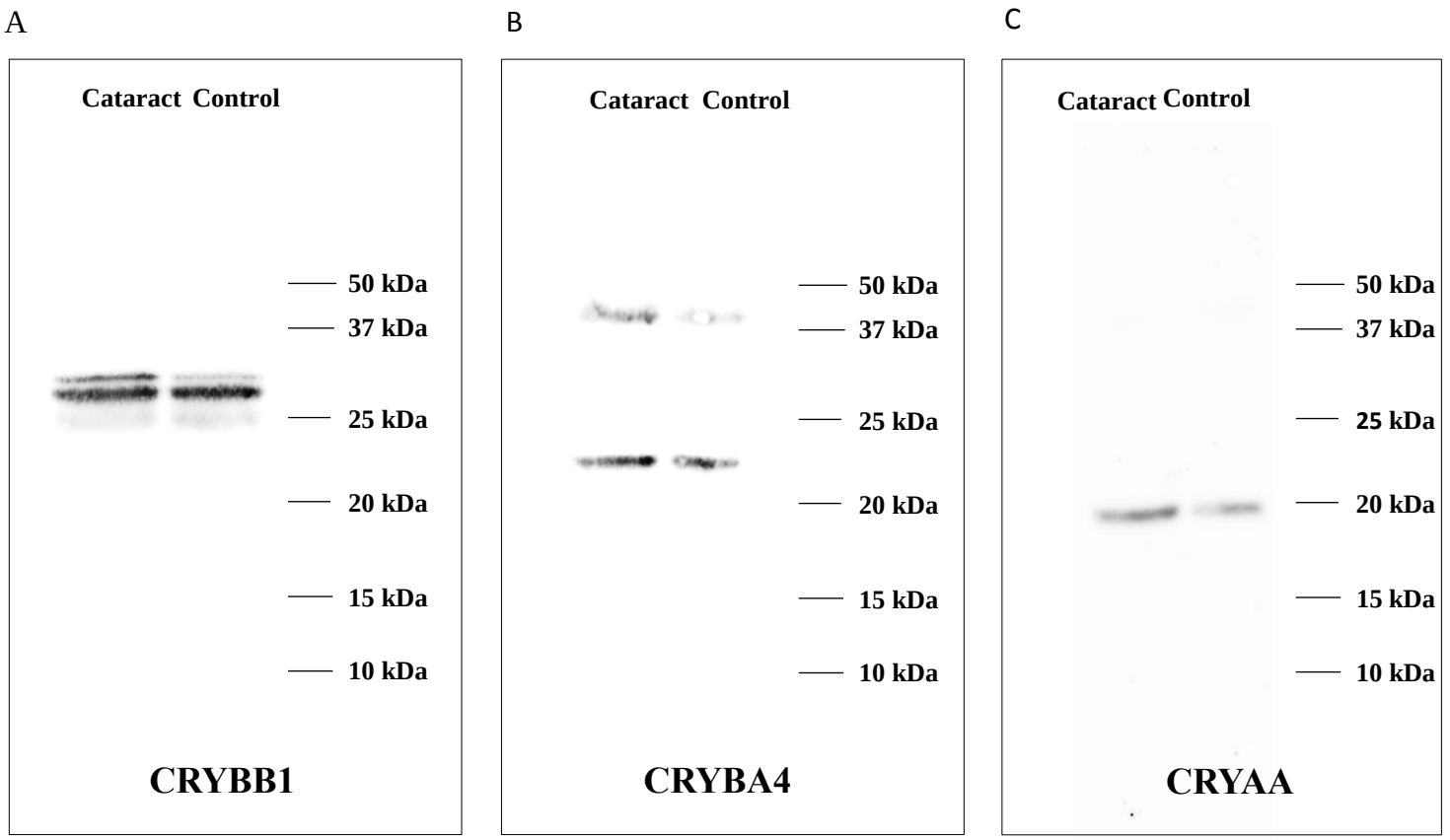

A
Cataract Control
50 kDa
37 kDa
25 kDa
20 kDa
15 kDa
10 kDa
CRYBB1
B
Cataract Control
50 kDa
37 kDa
25 kDa
20 kDa
15 kDa
10 kDa
CRYBA4
C
Cataract Control
50 kDa
37 kDa
25 kDa
20 kDa
15 kDa
10 kDa
CRYAA

Figure 5.8. Western blot analysis of whole lens soluble proteins of a normal and the cataractous CSA106.06 lens carrying the *CRYBB1* partial and *CRYBA4* complete duplication. Equal amounts of soluble whole lens proteins were analysed under denaturing conditions (20 micrograms). The CRYBB1 (A), CRYBA4 (B) and CRYAA (C) proteins were detected with the anti-CRYBB1, anti-CRYBA4 and anti-CRYAA antibodies respectively (the same blot was hybridized with the mentioned antibodies after stripping). The CRYAA Western blot image (Panel C) shows equal loading of protein from cataract and normal lens. The molecular masses of protein standards in kilo Daltons are indicated. The protein size of 28 kD and 21 kD for CRYBB1 and truncated CRYBB1, 22 kD for CRYBA4 and 19 kD for CRYAA were expected. The western analysis failed to detect the truncated CRYBB1 protein product (~21 kD). The additional band for CRYBA4 is possibly represents CRYBA4 dimer.

5.4.3 Family CRCH11

We identified a 4-generation autosomal dominant congenital cataract pedigree (Figure 5.9), with affected members diagnosed between birth and 10 years of age (Table 5.4). The cataract phenotype of the family CRCH11 was classified as congenital dense nuclear. The phenotypic information of the affected family members is given in Table 5.4. CRCH11.01 was diagnosed with an unknown congenital cataract phenotype and cataract was removed from the left eye at the age of 6 months and from the right eye at the age of 9 years. CRCH11.02 was diagnosed with congenital bilateral cataract and had delayed cataract surgery at 12 and 15 months of age due to low birth weight. CRCH11.03 was diagnosed with bilateral posterior lenticonus and cataract at birth and went through cataract surgery at 18 months of age for left eye and 17 years for right eye. CRCH11.04 was diagnosed with congenital, bilateral dense nuclear cataract and had bilateral cataract surgery at 3 months of age. Regarding patient CRCH11.07, both cataracts were removed when he was very young which left him with limited sight in both eyes which has persisted to this day. Family members at the left hand side of the pedigree (CRCH11.01, CRCH11.02, CRCH11.03, CRCH11.04, CRCH11.05 and CRCH11.06) were exome sequenced as we did not have DNA samples available from CRCH11.07 or CRCH11.08 at the time of initiating this research.

We analysed heterozygous protein changing novel/rare (MAF<1%) variants (stop loss, stop gain, frameshift insertion or deletion, missense and splicing) based on dominant inheritance of the disease in the family. We filtered the variants to those shared between affected exome sequenced members (CRCH11.01, CRCH11.02, CRCH11.03 and CRCH11.04) and not present in unaffected members (CRCH11.05 and CRCH11.06). We detected 3 segregating, protein-changing variants in the left hand side of the pedigree in *SAG*, *NOL9* and *ID3* genes (Table 5.5). The variants detected in *SAG* and

NOL9 were absent in the ExAC database. The variant detected in *ID3* was detected in ExAC database at very low minor allele frequency of 0.000008432. None of these 3 variants were present in our screened Australian control population. To be able to narrow down the number of candidate genes, we contacted 2 more affected family members (CRCH11.07 and CRCH11.08) and obtained saliva samples from those individuals now living in London. Following DNA extraction and Sanger sequencing, the variant in *NOL9* (c.2107T>A) gene remained to be the only segregating candidate variant in the family (Figure 5.10). We screened 47 unsolved affected cataract cases in our cohort for mutations in *NOL9,* however we did not detect any mutations in this gene in other congenital cataract cases. The stoploss variant (p.(*703R5*)) results in an extension of five amino acids which may affect its function.

The carboxyl terminus of NOL9 protein is not well conserved among different species and the mutation does not change the secondary structural predictions of the protein. In addition, PSIPRED doesn't predict any β-sheet or α-helix at this location for wild type protein (Figure 5.11 A, B). Three-dimension prediction of the wild type and mutant protein is slightly different (Figure 5.11 C, D).

Table 5.4. Clinical details of CRCH11 family members.

ID	Age diagnosed	Disease Status	Age of surgery RE	Age of surgery LE	Cataract phenotype
CRCH11.01	0	Affected	9 years	6 months	congenital cataract
CRCH11.02	0	Affected	15 months	11 months	congenital cataract
CRCH11.03	0	Affected	17 years	18 months	congenital dense nuclear
CRCH11.04	0	affected	3 months	3 months	congenital dense nuclear
CRCH11.05	-	unaffected	-	-	-
CRCH11.06	-	unaffected	-	-	-
CRCH11.07	NA	affected	Very young	Very young	NA
CRCH11.08	0	affected	6-8 months	6-8 months	NA

LE, left eye; RE, right eye; NA, not available. The cataract phenotype was bilateral in this family. Missing data are indicated with "-".

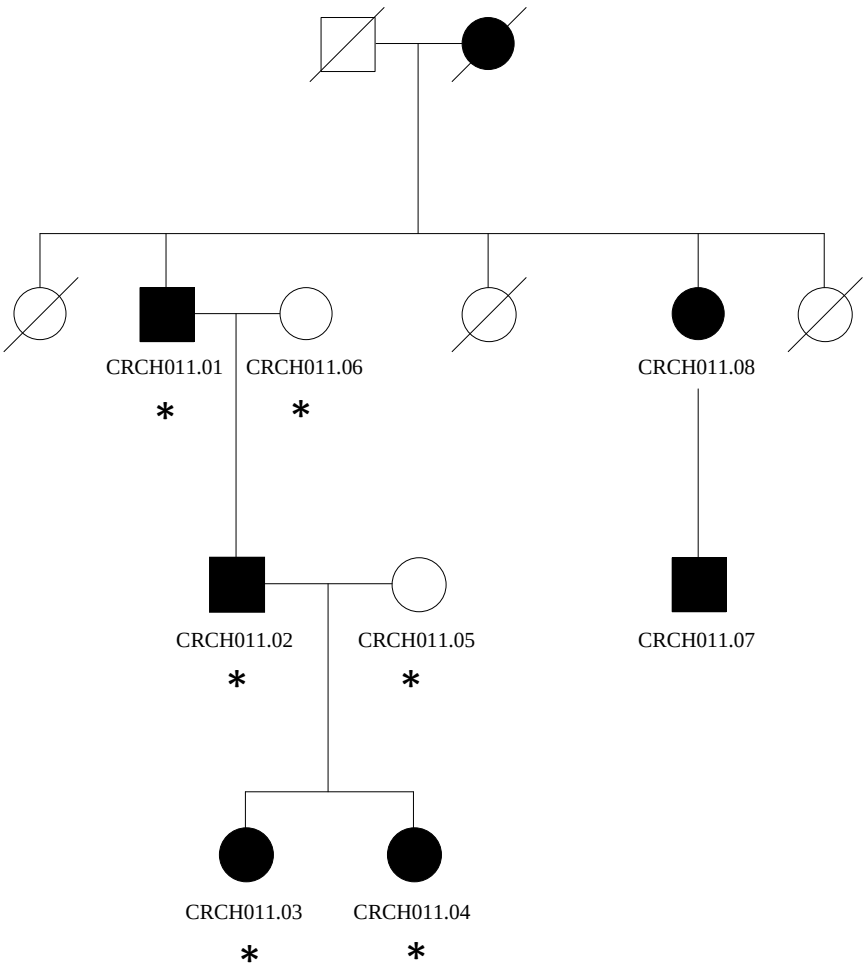


Figure 5.9. Family CRCH11 pedigree with an autosomal dominant congenital cataract, indicating affected (black) and unaffected (white) members. Circles indicate females and squares indicate males. The exome sequenced individuals are marked with an asterisk. Diagonal lines indicates deceased family member.

.

Table 5.5.Segregation analysis of detected variations in exome sequenced samples

Position	**Reference**	**Sample**	**Accession**	**Change**	**DNA change**	**Protein Change**	**Poly-Phen 2 Prediction**	**Gene**	**Validated in proband**	**Segregated**
chr1:6585916	A	A/T	NM_024654	Stop loss	c.2107T>A	P.(*703R5*)	No prediction	*NOL9*	Yes	Yes
chr1:23885474	C	C/T	NM_002167	nonsynonymous	c.337G>A	p.(D113N)	Possibly-damaging	*ID3*	Yes	No
chr2:234231596	C	C/G	NM_000541	nonsynonymous	c.380C>G	p.(P127R)	probably-damaging	*SAG*	No	No

Variant initially were considered for further consideration if present in affected CRCH11.01, CRCH11.02, CRCH11.03 and CRCH11.04 and not present in unaffected CRCH11.05 and CRCH11.06. Only the *NOL9* variant segregated in the newly recruited right hand side of the pedigree as well (in CRCH11.07 and CRCH11.08).

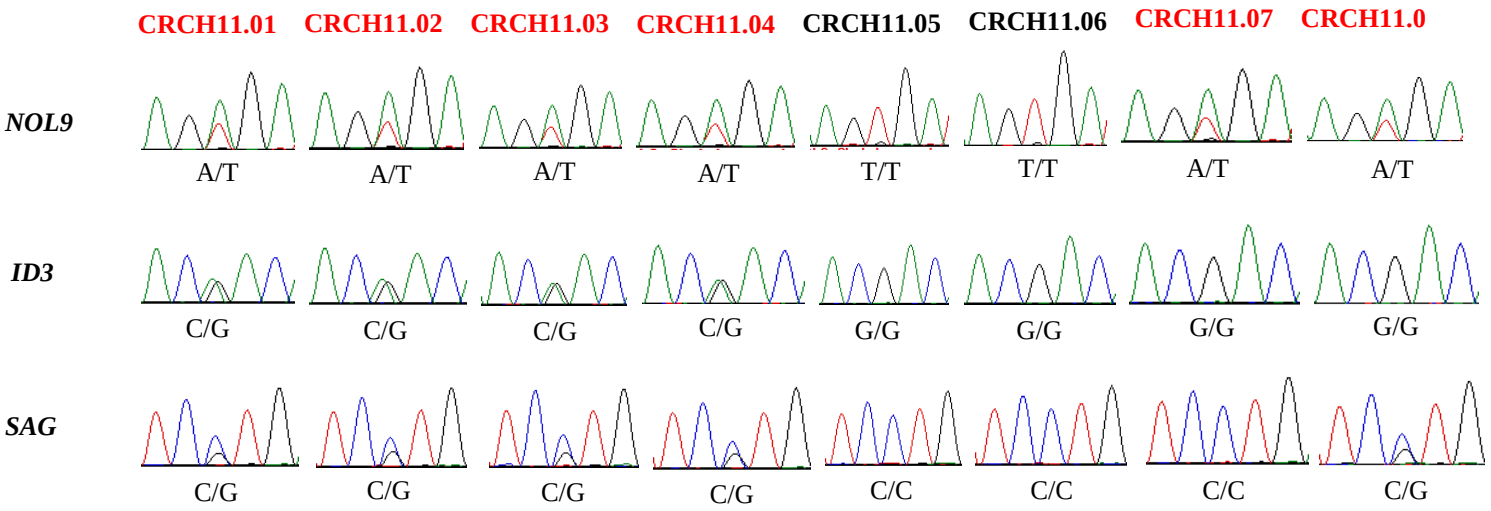


Figure 5.10. The chromatograms showing the sequence of detected heterozygous variations in the family CRCH11. Only *NOL9* variation segregates with the phenotype in the family (the affected individuals in the family indicated in red text).

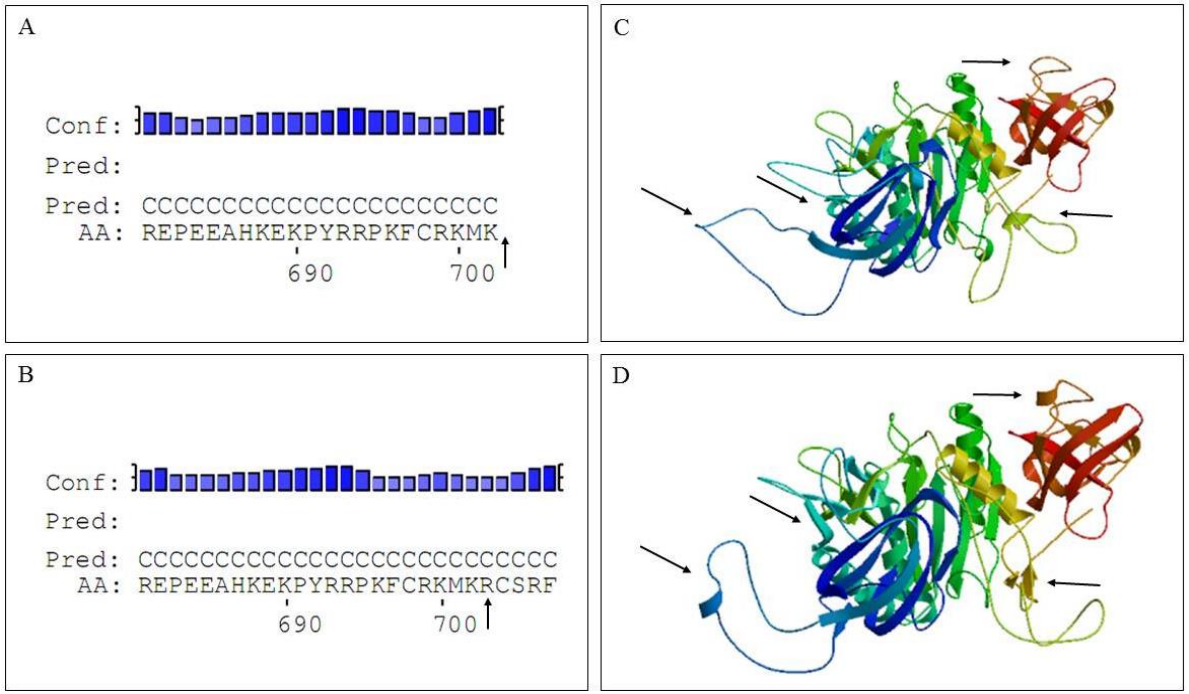


Figure 5.11. Predictions of the secondary and 3D structure of human NOL9 protein with PSIRED and SWISS-MODEL. A and C. represent Wild type where B and D represent P.(*703R5*) mutant protein. The altered amino acid and differences in the structures of wildtype and mutant NOL9 is marked with black arrows in sections A, B, C and D. There is no prediction of β-sheet or α-helix at the carboxyl end in wild type or in mutant protein with PSIPRED. The blue bars show the confidence in the prediction (on a scale of 1-10).

5.4.4 Family CRCH26

The cataract phenotype in the 2 generation autosomal recessive family CRCH26 was classified as nuclear congenital cataract (Figure 5.12). CRCH26.01 was diagnosed at the age of 4 and two other affected members; CRCH26.02 and CRCH26.03 were diagnosed at birth. No further information is available regarding their cataract surgery. The available family members were exome sequenced. Based on the recessive inheritance pattern of the family, initially we looked for potential homozygous mutations shared between the affected members and absent in unaffected CRCH26.04, CRCH26.05 and CRCH26.06. In total, 25 potential causative homozygous variants were detected in all affected members which did not qualify to be causative (all were present in whole exome sequence data from normal controls as homozygous, heterozygous or combination of both status). Furthermore we analysed the sequence for compound heterozygous variants (a condition when affected individuals have 2 unrelated heterozygous alleles in the same gene, with each mutation inherited from one parent). We did not detect any potential candidate gene fitting these criteria.

Next, linkage analysis was performed under rare recessive mode of inheritance (parameters: 0.001; 0.0001, 0.0001, 1) or low penetrance dominant mode of inheritance (Parameters: 0.001; 0.007, 0.7, 1.0). No significant linkage was detected under either of these models (Figure 5.13 A, B).

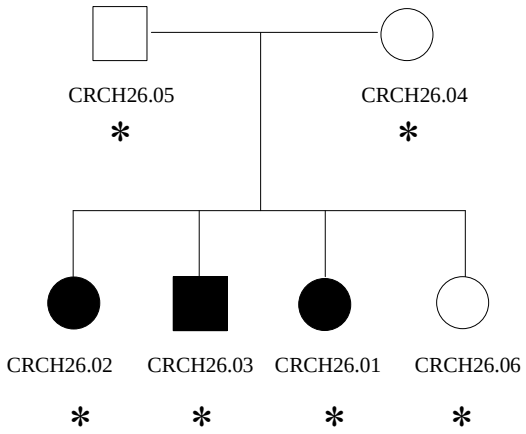
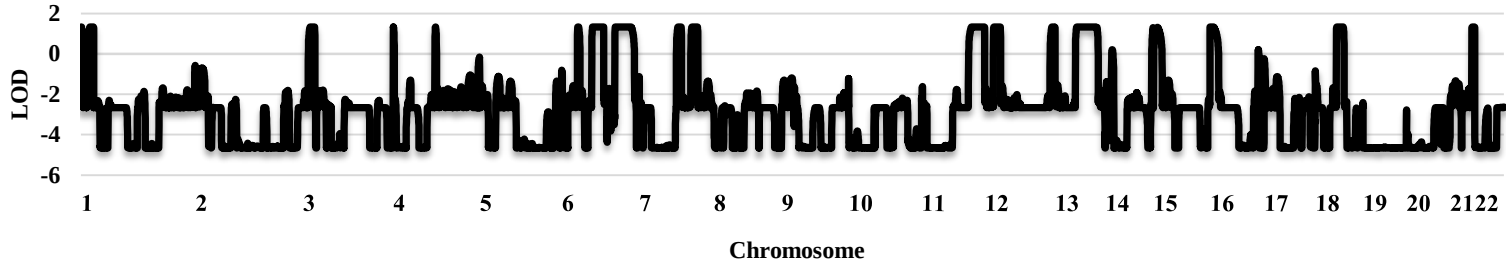


Figure 5.12. An autosomal recessive congenital cataract pedigree of CRCH11. Affected (black) and unaffected (white) members which were exome sequenced are marked with an asterisk. Circles indicate female and squared indicate male members.

A



B

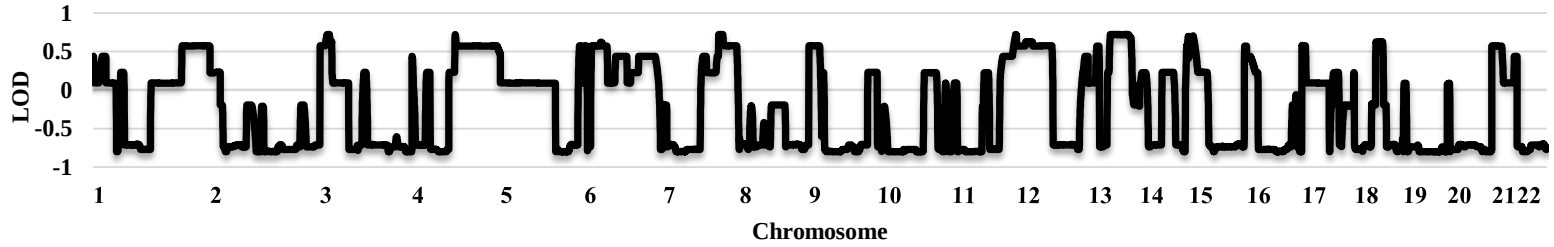


Figure 5.13. The parametric linkage analysis result of CRCH26 under rare recessive (A) and low penetrance dominant (B) mode of inheritance. No linkage region was detected (max LOD of 1.32).

5.5 Discussion

Finding the correlation between variations/mutations in genes that cause a particular phenotypic feature has been a key research focus for a long time. Whole exome sequencing has revolutionised the field of clinical genetics and is now regularly used in the study of diseases with Mendelian inheritance patterns.

Multiple affected and unaffected individuals from families CSA106, CSA92, CRCH11 and CRCH26 were selected for whole exome sequencing. We detected potential novel congenital cataract causing genes in 2 families; *NOL9* in CRCH11 and *HTR1F* in CSA92 (with potential involvement of modifier genes). Furthermore we identified a mechanism which had not previously been reported to be associated with isolated congenital cataract; copy number variation in *CRYBB1* in CSA106.

Variations in HTR1F, *BFSP1* and *HSF4* genes in CSA92

We have detected one segregating variant with reduced penetrance in *HTR1F* in CSA92. This variant is extremely rare in the general population. The variant is present in CSA92.06 who had a normal eye exam at the age of 17. We also detected 2 more rare SNPs in *BFSP1* and *HSF4* genes with the data generated from custom amplicon sequencing. Neither of these variants segregated appropriately with the phenotype; however they may act as modifier variants to account for the variable phenotype in this family. Both of these variants were predicted to be pathogenic by polyphen-2, SIFT tools.

Here we propose the possibility of involvement of a novel gene in cataract *HTR1F*, where two other variants in *BFSP1* and *HSF4* act as modifiers to change the penetrance of the variation. The phenotype severity trend in the family is: CSA92.03=CSA92.07> CSA92.04> CSA92.02. The affected individuals with the severe phenotypes (IOL

implants), CSA92.03 and CSA92.07, characterized by diagnosis at age 4 and cataract surgeries, carry all three variants. Carrying the *HSF4* variant from the unaffected father in addition the novel *HTR1F* variant appears to make the phenotype more severe as CSA92.04 carries this variant while CSA92.02 does not. The variant in *BFSP1* is either of no significance or it only contributes to the severity if the other two variants are present. Interestingly, the unaffected individual CSA92.06 carries the variants in both *HTR1F* and *BFSP1* as does his affected mother CSA92.02. CSA92.02 was diagnosed at the age of 42 with a very faint fetal lamellar cataract. The mother was not diagnosed until recruitment in the study following enrolment of her two affected sons. We initially hypothesised this family showed recessive inheritance, but on later examination of the mother, hypothesised it was more likely dominant with variable penetrance. It is possible that CSA92.06 who was only 17 at the time of examination could develop a mild cataract at later age, as his mother.

There have been previous suggestions of modifier genes involved in the development of congenital cataract (Burdon, et al., 2004b; Devi, et al., 2005; Maeda, et al., 2001), although such genes have not yet been identified. Of note, mutations in *GJA3* and *CRYBB2* have been reported with reduced penetrance in congenital and paediatric cataract (Burdon, et al., 2004b; Reis, et al., 2013).

HTR1F (5-hydroxytryptamine (serotonin) receptor 1F, G protein-coupled) or 5-HT_{1F} is one of several different receptors for 5-hydroxytryptamine (serotonin). Seven distinct families of serotonin receptors, $5HT_1$ to $5HT_7$, have been identified to date (Costagliola, et al., 2004). Although the role of serotonin in lens metabolism is unclear, its increased levels have been shown to cause cataract in rats (Boerrigter, et al., 1992; Erie, et al., 2014). It also has been demonstrated that administration of a selective 5-HT_3 antagonist, SDZICT3220 causes posterior subcapsular lens opacities in rat (Langle, et al., 1993).

Serotonin is a melatonin precursor and is a substrate for AANAT (arylalkylamine N-acetyltransferase), an enzyme involve in production of melatonin by serotonin modification. Immunohistochemistry studies revealed both AANAT and serotonin are localized in the lens cortical fibre cells of rat (Itoh, et al., 2007). It also has been proposed that melatonin synthesized in lens may prevent cataract formation in lens by acting as an antioxidant (Itoh, et al., 2007).

HTR1F RNA expression in different components of the eye including lens has been detected above background estimates (https://genome.uiowa.edu/otdb/). The expression is much lower compared to some of the well-known congenital cataract associated genes like crystallins, however it has a similar expression level to other congenital associated genes such as *GCNT2* and *MAF*. Therefore the gene still can be a candidate despite its low level of expression in lens.

CRYBB1 partial duplication in CSA106

We identified a partial duplication of the *CRYBB1-CRYBA4* locus in family CSA106. There are examples of point mutations in both *CRYBB1* and *CRYBA4* causing dominant congenital cataract (Appendix 1). Mutations in *CRYBA4*, for example, have been described in autosomal dominant congenital cataract (Billingsley, et al., 2006; Zhou, et al., 2010a). All three reported variants are missense mutations (G64W, L69P, and F94S) which presumably promote cataract formation by creating a less soluble protein. However, the *CRYBA4* duplication described here covered the complete gene, did not contain any missense variants, and did not lead to any obvious change in protein expression levels. *CRYBB1* mutations appear to cause both dominant and recessive congenital cataracts. The recessive *CRYBB1* variants p.(G57GfsX107) or M1K presumably cause cataracts by removing an important structural component altogether

(Cohen, et al., 2007; Meyer, et al., 2009). On the other hand, the dominant variants like Q223X (Yang, et al., 2008) and X253R (Willoughby, et al., 2005) are predicted to cause cataract by disrupting the coding sequence of the final exon (exon 6), and creating a protein with reduced solubility.

We hypothesised that the partial duplication in *CRYBB1* causing the truncated *CRYBB1* duplication product might also be the reason we see the dominant phenotype in this family, however we did not detect the truncated CRYBB1 protein product by western blotting (expected additional ≈21 kD product).Thus, we hypothesise other potential additional mechanism such as paralogous recombination. For this mechanism to be in play, two genes must be sufficiently homologous to recombine with each other. *CRYBA4* and *CRYBB1* are located head to head on chromosome 22, transcribed in opposite directions and share the same exon/intron structure and significant sequence homology (Figure 5.14). (Wistow, 2012). It is possible that exon 6 of *CRYBB4* was inverted and replaced the partially duplicated region in *CRYBB1*. The inverted exon 6 of *CRYBA4* would still map under *CRYBA4* in the sequencing data but the breakpoint would not be detectable through exome data. The resultant hybrid protein would likely not be functional due to its sequence disruption but would also not be detectable by western blot due to the products having the same expected size. The hybrid protein could act in a dominant negative fashion leading to the phenotype in family CSA106. Further analysis such as whole genome sequencing (or targeted sequencing not limited to exons) is required to map the sequence surrounding the breakpoint and validate this hypothesis (need to perform whole genome sequencing to map the breakpoints).

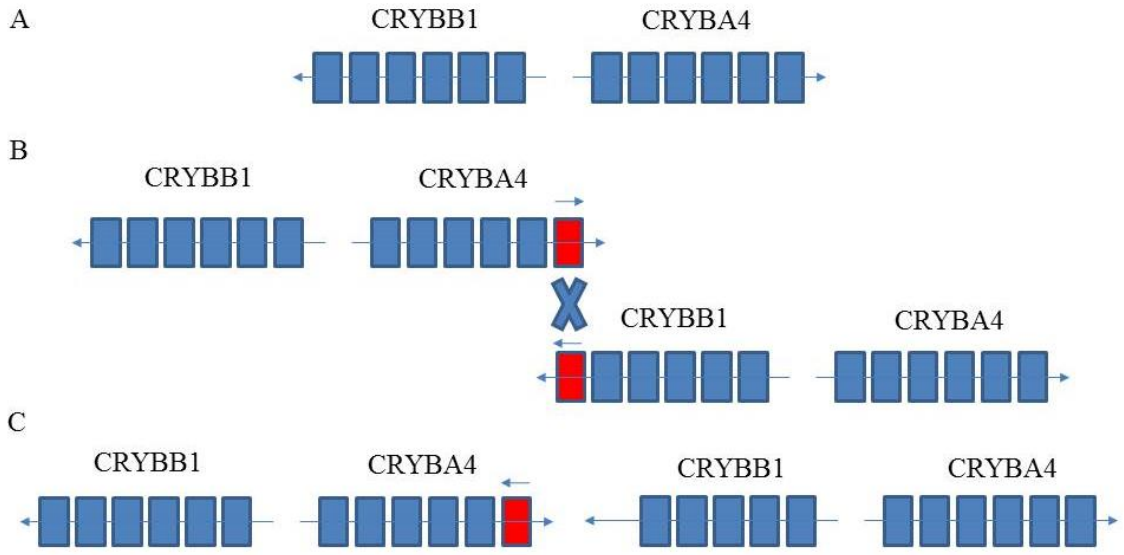


Figure 5.14. Paralogous recombination hypothesis in CSA106. A: *CRYBA4* is located head-to-head with *CRYBB1* and transcribed in the opposite direction. Both genes have 6 exons and similar structure. B: The possible complete inversion event may happen at the last exon of *CRYBA4* and replace the last exon of *CRYBB1*. C. The resultant hybrid gene would produce a hybrid protein, but the exon still would map to *CRYBA4* resulting the detection of partial duplication in CRYBB1.

Here for the first time we propose an association between autosomal dominant isolated congenital cataract with a unique duplication of the CRYBB1-CRYBA4 locus, however the mechanism of disease requires further investigation. A complex *NHS* gene CNV was reported to be associated with syndromic form of congenital cataract (a ~5.1 kb deletion located between ChrX: 17461256 and ChrX: 17466720), however to our knowledge there is no report of any CNVs associated with isolated congenital cataract. The CNV mechanism should be considered more broadly in cataract gene mapping.

***NOL9* in CRCH11**

Considering the mode of inheritance in the family after a multi-step filtering approach, there were 3 potential segregating candidate variants in 3 genes. By recruiting additional affected family members we succeeded in narrowing down the number of potential candidates to one variant in *NOL9* (c.2107T>A, p.(*703R5*)). This demonstrates the importance of having data from the most distantly related individuals within a family, who share fewer of alleles that are identical by decent (a segment of DNA shared by two or more people that has been inherited from a recent common ancestor). This concept has been well recognised through traditional linkage analysis, but is still often limited by the practicalities of identifying and recruiting individuals to a study.

NOL9 is a polynucleotide 5′-kinase with both RNA and DNA 5’-kinase activities. It is located within the nucleus of the cell and plays a role in ribosomal RNA (rRNA) gene processing. The primary transcripts produced from pre-rRNA genes undergo broad processing (a series of ordered exonucleolytic events). The 41S precursors is one of the first intermediates generated during this process which generates 21S and 32S rRNA (Lodish, et al., 2000). NOL9 is involved in rRNA processing by its polynucleotide kinase activity having a role in generation of 5.8S and 28S rRNA from 32S precursor.

The production of ribosomes from rRNA is essential in all cells as ribosomes are required for protein synthesis. The complex process of ribosome production requires the coordinated actions of ribosomal and non-ribosomal proteins. A recent study has demonstrated that a Zebrafish *nol9* mutant shows a defect in 28S rRNA processing (Bielczyk-Maczynska, et al., 2015). The *nol9* mutant zebrafish larvae demonstrated a complex phenotype of developmental issues in pancreas, liver and intestine and signs of pathological processes occurring in endothelial cells of the caudal vein (Bielczyk-

Maczynska, et al., 2015). The study also demonstrated that pancreatic and hematopoietic deficiencies in *nol9* mutant embryos were due to impaired cell proliferation of progenitor cells (Bielczyk-Maczynska, et al., 2015). Failing in ribosome biogenesis and function is associated with a group of diseases called ribosomopathies. There is no report of any association between congenital cataract and impaired ribosome synthesis; however it has been demonstrated that decreased expression of ribosomal proteins is associated with human age-related cataract (Zhang, et al., 2002). The segregation of a variation in *NOL9* which is a gene involve in rRNA processing with an early onset nuclear congenital cataract phenotype, suggesting changes in ribosome biogenesis may play an important role in lens transparency, possibly during early stages of lens development.

The stoploss variant in *NOL9* causes the addition of 5 amino acids to the carboxy-terminal end of the polypeptide, is the only detected segregating variant in this family thus is the only candidate for the observed congenital cataract phenotype at this time. Functional work is essential to evaluate the pathogenicity of the variant, especially when it is not possible to assess the pathogenicity using the prediction tools.

Family CRCH26

We were not able to detect any potential candidate mutations through exome sequencing for recessive congenital cataract in family CRCH26. No clear linkage region was detected through parametric linkage analysis under either recessive or low penetrance dominant mode of inheritance. This might be due to limited available family members for exome sequencing which decreases the linkage analysis power. Performing linkage analysis using exome data, several peaks with LOD>1 were detected. The gene is likely within one of those regions, but possibly not coding or not captured by the exome

sequencing protocol. As with the CSA106 and CRCH11 families, linkage analysis in more family members would likely reduce the number of these peaks and increase the LOD at the true location. The other issue that needs to be considered is the tightness of filtering strategy that has been used. SIFT uses sequence homology (Kumar, et al., 2009), while PolyPhen2 predictions is based on conservation, protein folding and crystal structure (Adzhubei, et al., 2010). One study demonstrated that the SIFT and Polyphen-2 prediction of a variant being benign may not be always correct. For example one particular cystic fibrosis causing mutation which is among the 13 most frequent missense mutations identified in a Canadian cystic fibrosis cohort, was predicted to be benign by both SIFT and polyphen (Dorfman, et al., 2010). It is also possible that the disease causing variation or mechanism underlying the observed phenotype is in non-coding regions of genome. It has been demonstrated that congenital cataracts facial dysmorphism neuropathy (CCFDN) syndrome (OMIM 604168) is caused by a single-nucleotide substitution in an antisense Alu element in intron 6 of CTDP1(RNA polymerase II subunit A C-terminal domain phosphatase) which is important for proper eukaryotic transcription (Varon, et al., 2003). The mutated Alu element is located on the antisense strand of CTDP1 in intron 6. Performing RT−PCR and sequencing analysis of one of the flanking introns in the critical region, the study identified a rare mechanism of aberrant splicing in which the donor site created by the C>T variation and results in the insertion of 95 nucleotides of the Alu sequence in CTDP1 mRNA. There is no information regarding the replication of these findings (Varon, et al., 2003) but it may illustrate a mechanism through which non-coding mutations can lead to disease including cataract.

5.6 Summary and conclusion

We identified 2 potential novel genes (*HTR1F* and *NOL9*) and propose a novel mechanism (CNV) responsible for paediatric cataract phenotype in human. Screening our unsolved paediatric cataract cohort for variations in these genes, we were not able to detect any second mutations in these genes in other affected probands in our repository; however they might cause cataract at higher frequencies in other populations. Screening other populations with different genetic background and functional analysis to investigate the effect of these mutations would be beneficial to further support the potential role of these genes in cataract formation.

The next step of the research would be performing functional analysis to see the effects of these variants on the protein function. Moreover, replications of these finding in other genetic populations would further support establishing *HTR1F* and *NOL9* as novel congenital cataract genes.

Chapter 6: General Discussion

During the last 25 years, approximately 50% of the genes associated with the approximately 7000 known monogenic disorders have been identified, and it is predicted that the majority of the remaining ones will be discovered by 2020 (Boycott, et al., 2013).

As discussed in previous chapters, paediatric cataract and PCG are Mendelian traits that are typically caused by inherited mutations in a single gene. Paediatric cataract can be due to autosomal dominant, autosomal recessive or X-linked inheritance, whereas mainly autosomal recessive inheritance has been reported for PCG. The development of massively parallel sequencing technologies has dramatically changed the field of gene identification in these and other Mendelian diseases. MPS also has been shown to be an efficient diagnostic tool when it comes to the clinical application. Understanding the genetics of such disease is important to provide a better understanding of the mechanisms and pathways involved in the development of them.

6.1 Application of Massively Parallel Sequencing (MPS) technologies in Mendelian disorders: Known gene screening and novel gene discoveries

Enormous progress has been made in recent years regarding speed, read length, and throughput in MPS platforms (van Dijk, et al., 2014). When it comes to the implementation of a new diagnostic test, the evaluation of the accuracy, cost and time efficiency of the test is critical. The cost of sequencing using MPS is much lower compared to conventional Sanger sequencing, and as time goes on it is going to be much more economically efficient as the price for MPS continues to decrease. However, Sanger sequencing is still the gold standard for validating novel findings. The challenge

now is the identification of which platform is best to use for a specific aim. In this study, we investigated the genetic contribution of the known paediatric cataract genes to inherited congenital cataract in Australian and Asian cohorts by utilising Ion Torrent Personal Genome Machine (PGM).

Application of PGM for genetic diagnosis in Mendelian disorders

The utility of the PGM as a diagnostic tool has been examined for several diseases, including congenital hearing loss (Nishio, et al., 2015), cystic fibrosis (Abou Tayoun, et al., 2013), long QT syndrome (Millat, et al., 2014a) and cardiomyopathies (Millat, et al., 2014b). For example Nishio *et al.* used the PGM system to determine the contribution of known genes toward diagnosis of known genes behind deafness and concluded it has sufficient efficiency to be used as a diagnostic tool. Another study by Tayoun *et al.*, demonstrated that the reproducibility of PGM platform was 100 % for mutation detection in independent runs for a common autosomal recessive disorder like Cystic fibrosis (Abou Tayoun, et al., 2013). Thus the PGM is a suitable technology for genetic testing of gene panels in Mendelian disorders.

One of the limitations of the Ion Torrent PGM, which is common to all MPS platforms, is the presence of sequencing errors associated with homopolymer stretches (Bragg, et al., 2013). This limitation makes Ion Torrent PGM not very appropriate for detecting variants such as poly (T) tract polymorphisms (Abou Tayoun, et al., 2013; Bragg, et al., 2013). However the false positive or recurrent mutations we detected in this study were not located in polynucleotide regions.

In the current study, 2% of amplicons were covered less than 20 fold, potentially limiting the ability to detect heterozygous mutations in this amplicons. By chance, the majority of the 16 genes containing these low coverage amplicons are involved in

syndromic forms of congenital cataract. In this study, 13 % of our screened probands have syndromic paediatric cataracts, which means sequencing the majority of the isolated paediatric cataract samples was not affected by this limitation. Evaluating the effect of such limitation is important for future studies. For example a trial sequencing run to check the coverage, and then supplement with extra primers (if the region is not located within GC rich or repetitive areas) to try to increase depth in such regions would be a good start for other studies.

One of the main advantages of targeted sequencing technologies like Ion Torrent PGM is that they require low amount of DNA (only needs 10-40 ng of DNA in the case of PGM) which makes it possible to do sequencing from limited samples. In our experience, in many cases, the quality of DNA was not a crucial factor in obtaining good sequencing results, however, we would require a further detailed investigation to conclude the minimum quality of DNA for successful sequencing.

Another advantage is that targeted sequencing requires smaller storage space than WES or WGS and the interpretation of the data is more straight forward and less time consuming.

There are some controversies regarding the cost efficiency and error rate regarding PGM and similar technologies such as the Illumina MiSeq. Some argue thatdespite Ion Torrent PGM requiring more hands-on time, it is stillcheaper and faster compared to the MiSeq (Li, et al., 2013) and has better GC depth distribution compared to Illumina's HiSeq 2000 (Liu, et al., 2012). On the other hand, another study demonstrated that the MiSeq had the higher throughput per run and lower error rates (Loman, et al., 2012). Quail *et al*. detected a strong bias when sequencing the extremely AT-rich genome of Plasmodium falciparum on the PGM, resulting in no coverage for approximately 30%

of the genome when compared to other similar platforms like MiSeq and Pacific Bioscience (Quail, et al., 2012). However, this isn't particularly relevant to this sequencing project and other studies which are looking at human samples. They also analysed the ability to correctly call variants from each platform and reported that despite more variants being called from Ion Torrent data compared to MiSeq data, a higher false positive rate also exists (Quail, et al., 2012). In our experience, the false positive calls rate decreased significantly when Ion Torrent released the latest version of their analysis program, Ion Reporter, which we used in this study (we also performed some analysis with earlier versions at the beginning of the study). However, these findings are from personal experience and we have not performed a dedicated comparison study between PGM and similar platforms.

Application of targeted MPS in paediatric cataract

To our knowledge, only two studies have used targeted MPS to assess the contribution of known paediatric cataract genes to the disease. The first study used Miseq, Hiseq and SOLiD 5500 platforms to screen 36 patients with isolated paediatric cataract for 115 isolated and syndromic paediatric cataract associated genes (Gillespie, et al., 2014). The study had a 75% detection rate for causative variants, of which 48% were classified as "clearly pathogenic" based on Polyphen-2 and SIFT pathogenicity predictions. These are the same criteria which we used in this study.

A recent study used the Illumina TruSeq Custom Amplicon platform to screen 32 selected paediatric cataract-associated genes in their dominantly Caucasian Australian cohort of 46 cases of isolated familial or sporadic paediatric cases (Ma, et al., 2016). They detected likely causative variants in 73% of familial and 68% of sporadic paediatric cataract cases.

It is important to note that through Sanger sequencing we have previously identified mutations in genes included in our panel in other families in our Australian repository (Burdon, et al., 2007; Burdon, et al., 2004a; Burdon, et al., 2004b; Craig, et al., 2003; Dave, et al., 2013; McLeod, et al., 2002; Reches, et al., 2007; Sharma, et al., 2008). When considered together with our earlier published work, mutations in the 51 genes on our gene panel account for 62% of familial paediatric cataract, a proportion comparable to that reported in a similar study of patients from the UK (Gillespie, et al., 2014) and Australia (Ma, et al., 2016). Furthermore, our screening of 51 genes provided similar overall results to those from the whole exome sequencing approach.

Application of whole exome sequencing in paediatric cataract

Uncovering genetic defects causing monogenic inherited diseases is one of the most important applications of whole exome sequencing. While protein-coding regions account for about 1% of the human genome, the majority of Mendelian diseases phenotypes known so far, result from altered function of a particular protein (Chong, et al., 2015). However, this percentage will change as studies start to look more (and understand more) about the non-coding regions of our genome.

Using exome sequencing, multiple novel mutations in known paediatric cataract-associated genes including *GJA3* (Li, et al., 2016; Yuan, et al., 2015), *GJA8* (Chen, et al., 2014), *CRYGD* (Zhuang, et al., 2015), *MAF* (Narumi, et al., 2014; Niceta, et al., 2015), *HSF4* (Behnam, et al., 2016), *EPHA2* (Reis, et al., 2014) and *BFSP1* (Wang, et al., 2013) have been identified.

Recently, novel genes associated with paediatric cataracts have been reported in the literature in which whole exome sequencing was involved in their identification.

The very first congenital cataract associated gene that was identified through a combination of traditional linkage analysis and whole exome sequencing technology was *AGK*, which is a lipid metabolism gene.

The other paediatric cataract associated gene which was identified through exome sequencing is *CRYBA2*. *CRYBA2* is another gene from βγ-crystallin superfamily in which a mutation has been reported to cause autosomal dominant paediatric cataract in humans. The protein expression studies showed the presence of *cryba2* transcripts during early lens development in zebrafish which supports its potential association with cataract (Reis, et al., 2013).

A recent study mapped an interval on chromosome 19p13.11–q13.2 for autosomal recessive paediatric cataract using whole exome sequence data of family members (Evers, et al., 2015). They identified a homozygous nonsense variant (c.4489C>T, p.(R1497*)) in the gene *SIPA1L* which causes the production of a shorter protein. The gene encodes a GTPase-activating protein (GAP) which is predicted to contain a Rap GTPase-activating protein (Rap-GAP) domain. The protein interacts with small GTPases of the Rap family via its Rap-GAP-domain (Evers, et al., 2015). GTPases are suggested to have important regulatory functions in the developing lens such as epithelial cell proliferation, differentiation of the fibre cells and cytoskeletal structure (Evers, et al., 2015).

Another study identified two homozygous missense mutations in *LSS* gene in two families with congenital cataract. *LSS* encodes lanesterol synthase; an enzyme involve in cholesterol synthesis (Zhao, et al., 2015).

WES has led to discovery of four novel paediatric cataract causing genes with or without linkage information. In this study we also have identified 2 novel candidate

genes associated with paediatric cataract (*HTR1F* and *NOL9*). Further investigation using functional studies is required to show how they may be involved in disease pathogenesis and this may help with developing therapeutic methods. However the first step toward this goal will be to prove pathogenicity of the mutations to confirm diagnosis in the two families and then they would be included for screening for genetic testing.

Which platform to use?

One of the main challenges of scientists and clinicians is to decide whether to use targeted sequencing, WES or WGS. As the cost of sequencing continues to decrease, WGS or in some cases WES appears to be a more cost-effective approach. However, there are certain issues that need to be considered before choosing a method.

The use of targeted gene panels now allows the analysis of all the genes known to cause a disease in a single test, however in a condition like paediatric cataract there are many genes yet to be identified. Adding the new causative genes to the panel requires changing the primer designs and primer pool(s) resynthesis which is a great cost and a main limitation in a targeted gene screening system like PGM. With enormous decrease in the cost of MPS, WES or even WGS are very appropriate and efficient alternatives or complimentary to the targeted sequencing approach. By using a WGS approach, if no candidate coding mutation is identified, then pathogenic non-coding variants could be considered. However, sometimes the biological interpretation of non-coding variants remains a difficulty.

In addition, as yet WGS is more expensive compared to WES, and WGS data files are much larger compared to WES, which raises the issue of data storage and management.

One issue regarding WES is the coverage, which is approximately 85%-95%. This means there is a chance of missing a particular disease related candidate variant due to missing or poor coverage (Jamuar and Tan, 2015). Coverage can be much higher (even complete) when a targeted approach is used. One interesting example of the potential effect of this issue is a study which found that four hearing loss candidate genes were covered 0-40% by WES, whereas the coverage by targeted sequencing was 100% (Jamuar and Tan, 2015).

Currently, finding mutations underlying Mendelian disorders is focused on exome sequencing. We were able to find potential protein changing causative variants in two (CSA92 and CRCH11) out of four paediatric cataract families using exome sequencing. Moreover, using WES data we were able to perform linkage analysis in the other two families, which led to the identification of a linkage region in family CSA106. Performing CNV analysis in this family, we were able to identify a putative mutational mechanism (Partial and complete duplication of *CRYBB1* and *CRYBA4*) which has not previously reported to be associated with isolated paediatric cataract. In the unsolved family of CRCH26, no linkage region or a CNV were detected. These same issues would have occurred if we had used WGS instead of WES. In fact, the situation would have been significantly more complicated as there would be even more variants to assess without having a linkage region to reduce the search region.

In our experience, the overall results we obtained from targeted sequencing of 51 genes is comparable to the studies where the whole exome approach (Ma, et al., 2016) was applied, which shows PGM is a suitable platform when it comes to clinical application in paediatric cataract. In conclusion, targeted approach could be used at the first instance for clinical application and if no mutation is found in any of the known disease

genes, WES/WGS approaches could be considered depend on the requirements and available funds. The current study is not able to compare targeted and whole exome sequencing platforms, however it would be interesting to see if the same sample subjected to both platforms had a similar output in terms of quality control parameters, raw data coverage and more importantly the spectrum of observed variations.

6.2 Potential downstream applications of finding novel paediatric cataract associated mutations/genes

In our experience, both ophthalmologists and their patients are eager to know the genetic basis of their disease. As a research group working in this area, we have been asked on multiple occasions to provide our genetic research findings to the individual's family's ophthalmologist.

In one particular case (CSA159), we were asked to deliver our findings to the affected child's family in order to help them evaluate their options during an ongoing pregnancy. In another case (CSA108.01), the clinical diagnosis of this proband with a rare syndromic congenital cataract phenotype was complicated and protracted, largely due to the use of traditional diagnostic assessment procedures. Past investigations on this individual included brain MRI, EEG, karyotyping, subtelomere FISH, FISH for Smith-Magenis syndrome, TORCH serology, and a urine metabolic screen; all were normal apart from a diffusely abnormal EEG. Sequencing of the *NHS* gene associated with Nance-Horan syndrome (congenital cataract, dental anomalies and developmental delay) did not detect any pathogenic variants. The implementation of our targeted next generation sequencing, which provided a more convenient molecular diagnostic process, meant that we were able to detect a mutation in the *MAF* gene, leading to the diagnosis of a rare condition called Ayme-Gripp syndrome in this family (CSA108).

In addition, we also have received requests from other ophthalmologists asking that we include their patient samples on our sequencing panel.

These examples of prenatal screening, efficient diagnosis, and ongoing requests strongly support the very real need for developing diagnostic testing for paediatric cataracts in Australia.

As far as PCG is concerned, clinical diagnoses are often difficult and it is sometimes impossible to separate patients with similar molecular causes of disease. This is largely because this condition is a cluster of diseases with overlapping molecular causes, and there is no strong phenotype-genotype correlation. As an example, PCG cases with *TEK* mutations appear clinically identical to regular PCG cases which contain recessive *CYP1B1* mutations. This again emphasises the advantages of using MPS as a diagnostic tool.

Novel genes discovery

Finding novel genes and mechanisms underlying congenital cataract would contribute to our knowledge of gene function and regulation which potentially could lead to developing new therapeutics. New therapeutics for this condition are keenly needed, as the current method of treating congenital cataract is surgery, which often comes with secondary complications (like glaucoma) and is not suitable for newborns.

An example of the identification of a gene which then had a clear therapeutic application is *LSS* (Zhao, et al., 2015). The study demonstrated that lanosterol could dissolve the cataract causing precipitate and amyloid-like fibril in mutant lens. This study demonstrated the *in vivo* therapeutic effect of lanosterol in treating cataract in rabbits and dogs (Hejtmancik, 2015).

Mackley *et al*. investigated another therapeutic approach for both age- related and hereditary cataract. This study identified a pharmacological chaperone (a small molecule that binds to the protein and stabilizes the native state of the protein) which partially reversed protein aggregation by binding and stabilizing the soluble form of cryAA and cryAB (Makley, et al., 2015).

These examples clearly demonstrate that finding novel candidate genes (such as *HTR1F* and *NOL9* in this study) and determining their role in the development of cataract can greatly assist in finding targeted therapeutic approaches for this condition.

6.3 PCG not always a recessive disease

In the majority of familial PCG cases, the mode of inheritance is reported as autosomal recessive (Cascella, et al., 2015; Sarfarazi and Stoilov, 2000) with a higher rate of incidence among populations with consanguinity (Sarfarazi, et al., 2003). The assumption that PCG is a recessive disorder determines the mutation filtering strategy when analysing large scale sequence data, and defines the research plan when choosing to investigate the genetics of this disease. Mutations in *CYP1B1* account for the majority of the PCG cases reported so far (Appendix 6) and are usually recessive. There are reports of PCG cases with just one *CYP1B1* mutation identified (Pasutto, et al., 2010). Interestingly, variable severity in the phenotype and incomplete penetrance has been observed in PCG cases with *CYP1B1* mutations (Cascella, et al., 2015; Khan, 2011). An autosomal dominant mode of inheritance has been proposed by some authors based on reports of several pedigrees in which the disease was transmitted in successive generations (Sarfarazi and Stoilov, 2000). There are also reports of PCG cases with heterozygous variations detected in two other genes, *MYOC* or *FOXC1*, but no second mutation identified in these genes or *CYP1B1* (Kaur, et al., 2005; Medina-Trillo, et al.,

2015). It is possible that the second mutation has just not been identified in the known genes with the technologies used, or maybe a second as yet unknown gene is involved. However, it is important to consider an autosomal dominant form of inheritance for PCG as well. We identified five heterozygous mutations in our PCG cohort in a novel PCG associated gene: *TEK*. The penetrance of the mutations is variable in four of these families (the DNA from unaffected parents were available from these families for further analysis). The molecular studies demonstrated that a mouse model heterozygous for a *Tek* knockout mutation developed an abnormal Schlemm's canal and had higher IOP, similar to human PCG cases (Souma, et al., 2016). Through our collaboration with another group led by Professor Terri Young at the University of Wisconsin, we were able to support the hypothesis that PCG is not necessarily always a recessive disease as it has been previously considered (Ma, et al., 2016).

6.4 Paediatric cataract not always monogenic?

Before the emergence of the next generation sequencing technologies, finding a potentially pathogenic variant in one candidate gene in a family with Mendelian genetic disorder may have been considered the final molecular diagnosis. When using Sanger sequencing as a diagnostic or research tool, one would proceed one gene at a time (if at all) and stop when a mutation was found. However, with next generation sequencing, there is a higher chance of detecting relevant mutations in more than one candidate gene. Here in the current study, in some sporadic paediatric cataract individuals from Australia (see sections 4.4.1 and 4.4.2 from Chapter 4), and in some cases with either familial or sporadic paediatric cataract from Asian countries (see section 4.4.3 from Chapter 4), we detected more than one potential candidate gene associated with paediatric cataract. In familial cases, recruiting additional affected and unaffected

family members would be beneficial for segregation analyses, in order to narrow down the number of candidate genes to the causative one (as was achieved for *NOL9* in family CRCH11: see chapter 5, section 5.4.3). However, it's also very tempting to consider that paediatric cataract may not always be a monogenic disorder. It is also possible that the phenotype in the sporadic paediatric cases in this study are actually due to having recessive mutations within two separate genes (digenic inheritance). Examining the parents and collecting DNA samples from them is the first step in examining this hypothesis. It is clear that the theory of paediatric cataract being a polygenic or recessive digenic disorder needs official investigation in the future via functional studies.

Despite recent advances in identifying genes causing monogenic disorders, the knowledge about the genes involved in potentially digenic and polygenic disease is limited. Traditionally, it has been thought that Mendelian inherited diseases are caused by a mutation in a single gene that segregates with the phenotype in the family. However, there have been multiple reports of a digenic inheritance pattern being associated with some so-called monogenic Mendelian inherited disorders.

One example of studying a digenic disease was shown by Kajiwara et al. (1994), who demonstrated that mutations in two unlinked photoreceptor genes *ROM1* and *RDS* in three unrelated families causes retinis pigmentosa, which was classified as a monogenic (only) disorder prior to the study (Kajiwara, et al., 1994). ROM1 and RDS are homologous proteins that form noncovalent tetramers and higher order oligomers (Loewen, et al., 2001). It has been demonstrated that the level of these oligomers and the two protein interaction is critical for proper photoreceptor disc formation (Loewen, et al., 2001).

Another disorder in which digenic inheritance has been investigated, this time using whole exome sequencing technology, is facioscapulohumeral muscular dystrophy type 2 (FSHD2) (Lemmers, et al., 2012). A mutation in *SMCHD1* and an allele of the D4Z4 microsatellite array permissive for the expression of *DUX4* gene, have been shown to together result in the FSHD2 phenotype (Lemmers, et al., 2012). This study firstly demonstrated that reducing SMCHD1 levels in skeletal muscle results in D4Z4 contraction-independent *DUX4* expression, and secondly demonstrated that *SMCHD1* is an epigenetic modifier of the D4Z4 metastable epiallele (Lemmers, et al., 2012).

One of the main challenges of next generation sequencing panels is the ability to be able to provide a molecular diagnosis for a disease when this requires finding multiple potential causative variants with uncertainty about their involvement in the disease development or the level of their contributions to the observed phenotype. This highlights the necessity of additional functional studies to either confirm the involvement of the detected variants or to eliminate them as non-pathogenic variants.

6.5 Involvement of CNV in paediatric cataract pathogenesis: a novel mechanism

Here for the first time, we propose a new genetic mechanism for the development of isolated pediatric cataract in family CSA106. This mechanism, being a copy number variant containing *CRYBB1-CRYBA4* locus wasn't detected by either conventional candidate gene sequencing or by exome variant calling. This highlights the importance of investigating copy number variation in an inherited eye disease, and in particular, the significance of partial gene duplications.

Duplication or triplication of genes has been observed in other ocular diseases. A dosage change in the *TBK1* gene in normal tension glaucoma is one example of a gene duplication being associated with an ocular disease (Awadalla, et al., 2015; Fingert, et al., 2011). Another example is patients with CNV in *NHS* gene have X-linked congenital cataract, an allelic disorder of Nance Horan Syndrome (Coccia, et al., 2009).

Finding a novel mechanism such as CNVs, shows that the research approach must be flexible in gene and mutation discovery. The involvement of such mechanisms in disease pathogenesis may be rare (or underestimated), but they should be considered in families with no potential missense mutation identified.

6.6 Non-coding regions or synonymous variations

We have not been able to detect potential candidate mutations in approximately 40% of the screened paediatric cataract cohort. Performing whole exome sequencing on four selected paediatric cataract families when looking for new candidate genes, we were not able to find any potential candidate variants (including CNVs) in one family. Furthermore, there is a huge uncertainty about the compound heterozygous variants detected in *GREB1* in PCG002, both of which were synonymous. All of these indicate that either the causative variants are located in regions of the exome that were not well covered or were missed during the primer design process, or other disease-causing mechanisms are involved which need to be considered.

It is very interesting to notice that there are reports of Mendelian phenotypes associated with mutations outside the coding regions. Families in which the disease-associated mutation hasn't been identified in coding regions are ideal candidates to be considered for such non-coding mutations. Approximately one-third of the human genome is known to be biochemically active or conserved. This so-called "Medical genome" is

where the focus of finding pathogenic non-coding variants should be (Makrythanasis and Antonarakis, 2013). Examples of such pathogenic variants related to a variety of Mendelian and non-Mendelian phenotypes in microRNAs (miRNAs), long non-coding RNA (lncRNA), small non-coding RNA (ncRNA) and enhancers have all been reported (Makrythanasis and Antonarakis, 2013).

MiRNAs (about 22 nucleotides long) are involved in post-transcriptional regulation of hundreds of target genes, leading either to mRNA degradation or suppression of translation (Hughes, et al., 2011). An example of miRNA related to a Mendelian condition is *MIR184*. A mutation in the seed region of *MIR184* has been reported to be responsible for severe familial keratoconus combined with early-onset anterior polar cataract (Hughes, et al., 2011). There is a high chance that this gene doesn't contribute to paediatric cataract in the screened Australian and Asian cohorts, as this gene was included in our PGM gene panel, as well having the highest coverage of above 1800 fold, which indicates there is a very slim chance of missing variants in this gene (Figure 4.1).

Mutations in RNA genes have been reported to cause a variety of Mendelian diseases. More than 30 pathogenic variants in *TERC* which is a long ncRNA (longer than 200 nucleotides) have been reported to be associated with a variety of phenotypes such as Dyskeratosis congenital (Makrythanasis and Antonarakis, 2013), a progressive bone-marrow failure which could be autosomal recessive or dominant (Vulliamy, et al., 2001).

Microcephalic osteodysplastic primordial dwarfism type I is a severe autosomal recessive skeletal disorder (MOPD I). The disease is caused by mutations in *RNU4ATAC* on chromosome 2q14.2 which encodes U4atac snRNA (He, et al., 2011a).

U4atac snRNA is a component of the minor spliceosome (a molecular machine to remove introns from a transcribed pre-mRNA), and is involve in proper removal of the U12-dependent class of introns (He, et al., 2011a) (two spliceosomes, designated U2- or U12-dependent are involve in catalysing pre-mRNA splicing). There is no report of any association between mutations in RNA genes and paediatric cataract yet, however, these genes should not be excluded from future gene discovery studies.

One interesting concept that needs to be considered is the effect of synonymous variations on protein folding and production. There are 61 triplets or codons corresponding to 20 amino acids thus some of them are redundant. Synonymous codons encode the same amino acid; therefore they are assumed to have no effect on protein when one codon is substituted for another (Hershberg and Petrov, 2008). However, it has been suggested that the presence of synonymous codons in open reading frames (ORFs) of genes is not random and codon choice is under an evolutionary pressure (Buhr, et al., 2016). Buhr *et al*. showed that synonymous codon variants in *CRYGB* would alter the rate of the translation and increased protein misfolding and degradation (Figure 6.1) (Buhr, et al., 2016).

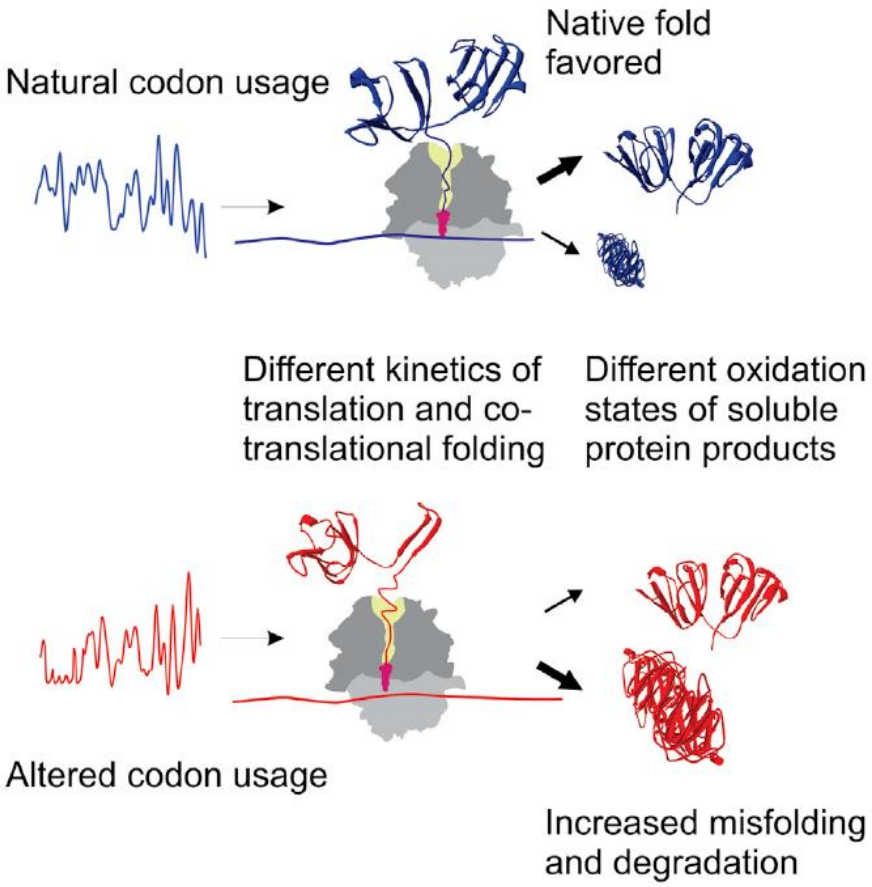


Figure 6.1. The illustration taken from (Buhr, et al., 2016) shows how synonymous codon usage alters the real-time kinetics of translation and CRYGB protein folding.

The effect of synonymous or so-called "silent" mutations on mRNA secondary structure and produced protein function (Nackley, et al., 2006), mRNA splicing (Parmley, et al., 2006) or protein activity have been demonstrated. More than 50 genetic disorders such as cystic fibrosis, schizophrenia and cancers are shown to be associated with synonymous mutations (Katsnelson, 2011).

On average about 20,000 variants are present in an individual exome data (Stitziel, et al., 2011). The focus of the Mendelian disorder studies is to narrow down this number to preferably one in a single gene. The variants detected via PGM or exome sequencing

were filtered out in different stages with the main focus of identifying the protein changing variants in paediatric cataract or PCG probands. This means that around 50-75% of variants (which were synonymous or UTR (in PGM data) were excluded from the analysis. It is not impossible that a synonymous variant causes the phenotype in some of the families in which we didn't find any obviously pathogenic mutations. This possibility should be noted and the data re-analysed with this in mind for unsolved cases, however, there are a large number of variants to be considered, making the analysis challenging. Multiple filtering steps used in most studies (including this one) and the assumptions made at each stage are given in Figure 6.2 (modified from (Stitziel, et al., 2011)).

Apart from non-protein coding regions of the genome or synonymous variants, UTRs of protein coding genes should be considered as well. An example of disease causing mutations in UTRs is the mutation in 5' UTR region of *DLG3* gene (associated with intellectual disability) which predicted to disrupt the folding of mRNA (Kumar, et al., 2016).

It needs to be emphasized that it is not possible to consider all the variants without taking any filtering strategy, however the possibility of losing variants needs to be considered.

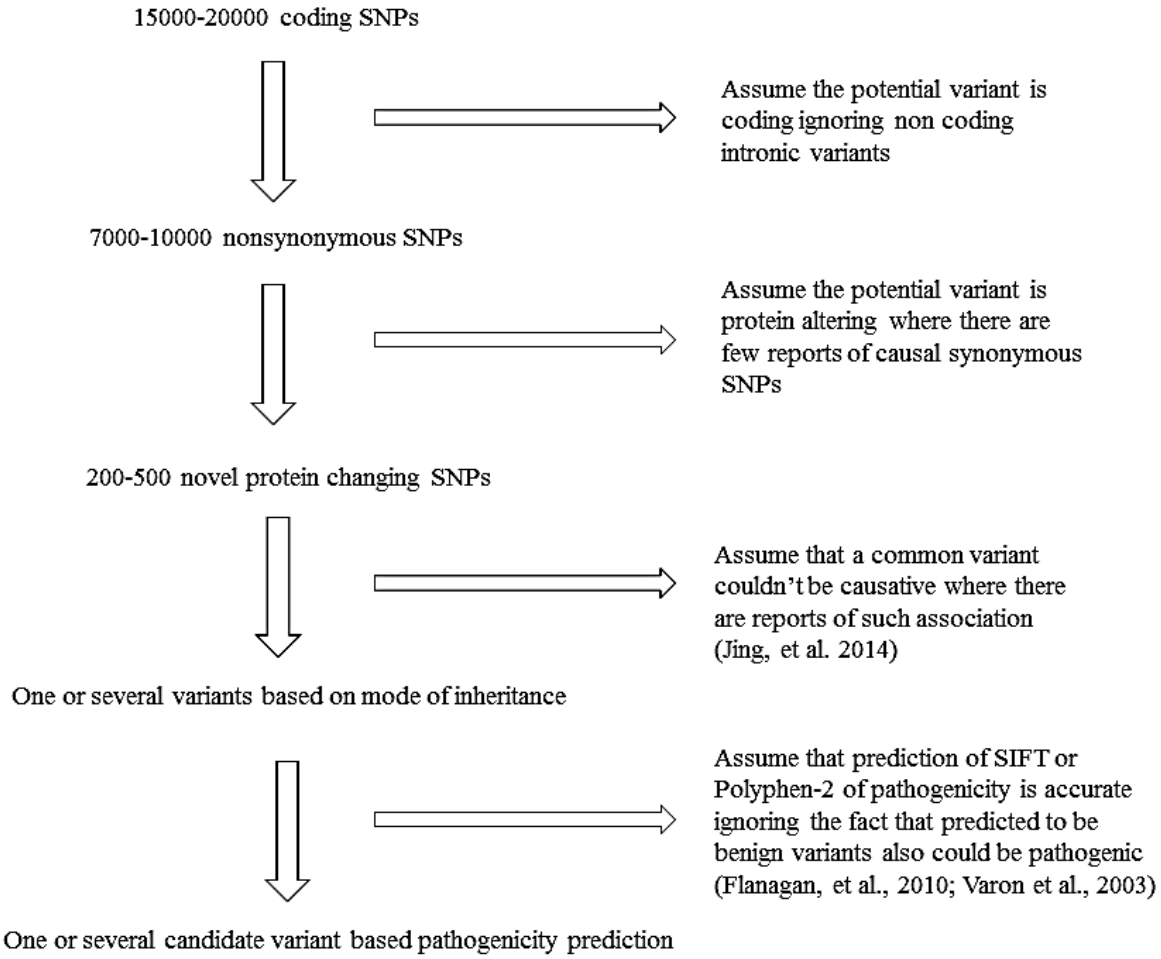


Figure 6.2. Classic filtering strategy applied to exome sequence data in this project. Modified from (Stitziel, et al., 2011).

6.7 Modifier genes in paediatric cataract

There are some familial paediatric cataract cases in which we noticed the variable severity of phenotypes ranging from no detectable cataract to a severe cataract among individuals with the same mutation (e.g., CSA110 and CSA92). In family CSA92, we identified a segregating variant in a novel gene *HTR1F*. Two non-segregating variants were also detected in known paediatric cataract-associated genes *HSF4* and *BFSP1*. Interestingly, patients carrying all three variants developed a more severe phenotype. We postulate the presence of modifier variants, possibly in *HSF4* and *BFSP1* altering

the penetrance of the mutation in *HTR1F* and the severity of cataract. There have been previous suggestions of modifier genes involved in the development of paediatric cataract (Burdon, et al., 2004b; Devi, et al., 2005; Maeda, et al., 2001), although such genes have not yet been identified. Of note, mutations in *GJA3* and *CRYBB2* have been reported with reduced penetrance in paediatric cataract in humans (Maeda, et al., 2001).

Reduced penetrance is a challenge in genetic counselling, as this means that the phenotype in a dominant trait might skip a generation and then express in the following generation. Discussing reduced penetrance inheritance with the patient and determining the disease risk for the patient's offspring is complicated (Cooper, et al., 2013).

6.8 Final statement and future work

The current study was able to apply MPS successfully both as candidate gene screening and a novel gene discovery tool. Finding novel heterozygous mutations in a novel PCG candidate gene, *TEK*, in a second cohort of Australian PCG patients demonstrated the enrichment of loss of function mutations in the PCG cohort compared to a normal population. Furthermore, it shows that PCG might not always be a recessive disorder (supported by the functional studies performed by (Souma, et al., 2016)).

We were able to detect potential novel and previously reported variations in known paediatric cataract associated genes in 40% and 23% of screened Australian and Asian cohorts respectively using targeted Ion Torrent PGM MPS. Moreover, we used WES to find novel candidate genes in PCG and paediatric cataract patients. We also reported a genetic mechanism not previously reported in isolated paediatric cataract before, by performing linkage and CNV analysis with WES data. All of these identified mutations demonstrate how successful and flexible MPS is in the genetic study of heterogeneous Mendelian disorders.

Having found no variants in a proportion of our screened paediatric cataract cohorts, it is possible that mutations may have been missed due to low sequencing coverage, or that these patients have mutations in genes not targeted by this panel. As discussed above, a few genes were identified after the start of our paediatric cataract screening project and therefore were not included in our panel. It's necessary to screen these genes in the remaining unsolved paediatric cataract probands. Moreover, UTR regions of the selected genes in PGM panel also have been sequenced. These regions need to be investigated to identify potential variants (if any), since the priority of the current study was finding variants in coding regions.

It is vital to develop screening tests in the future for extremely heterogeneous conditions like paediatric cataract which could help to determine the molecular genetic cause of these conditions more efficiently and inexpensively. It is also beneficial for families with the disease history who may be planning (now or in the future) to have children. The current study supports the implementation of MPS as a frontline diagnostic tool for congenital and paediatric diseases.

Functional studies need to be designed to prove the involvement of novel candidate genes in PCG or paediatric cataract development. Demonstrating tissue-specific expression of RNA and protein encoded by these genes in different sections of the eye is a basic essential. This can be achieved via reverse transcript PCR and immunohistochemistry in RNA and protein levels. Furthermore, the pathogenicity of such variants could be investigated using PCR-based mutagenesis. In summary, using this method, the mutant cDNA carrying the desired mutation will be created and the recombinant DNA clone will be generated. The proper cell line will be transfected using

the mutant and wild-type constructs. Western blotting and immunohistochemistry could be used to demonstrate the effect of the mutation on protein expression and localisation.

If the protein of interest is expressing in blood, Lymphoblastoid cell lines (LCLs) could be generated from affected and healthy individuals to investigate the effect of the detected mutations. A similar study was performed in our laboratories to investigate the effect of a 25bp deletion in FTL gene in hereditary cataract syndrome (HHCS) which is characterised by distinctive cataracts and high serum ferritin in the absence of iron overload (et al. 2007). As it is discussed at page 126, Western blotting of lens protein from a mutation carrier Burdon in family CSA91 showed a decrease in the ability of the protein to form higher order oligomers essential for its function (Laurie et al. 2013). It is possible to perform similar functional studies on tissues from affected and unaffected individuals (upon the availability of required tissues).

Appendices

Appendix 1

List of reported mutations in crystallin genes associated with paediatric cataract. Inh, inheritance; AD, autosomal dominant; AR, autosomal recessive; XL; X linked. "-"indicates no information was available.

Gene	Locus	Accessions	OMIM #	DNA Change	Protein Change	Inh	Phenotype	Reference
CRYAA	21q22.3	NM_013501.2	123580	c.161G>C	p.R54P	AD	Y-suture Paediatric	(Su, et al., 2012)
				c.142T>G	p.Y48D	AD	-	(Ma, et al., 2016)
				c.246_248delCGC	p.117delR	AD	Perinuclear	(Kong, et al., 2015)
				c.246_248delCGC	p.117DelR	AD	Punctate, Perinuclear	(Kong, et al., 2015)
				c.350_352delGCT	p.R117H; Y118del	AD	Total nuclear	(Su, et al., 2012)
				c.27G>A	p.W9X	AR	Paediatric	(Pras, et al., 2000)
				c.61C>T	p.R21W	-	Paediatric and microcornea	(Kondo, et al., 2013)
				c.292G>A	p.G98R	-	-	(Santhiya, et al., 2006)
				c.34C>T	R12C	AD	Posterior polar, microcornea,	(Hansen, et al., 2007b)
				c.347G>A	p.R116H	AD	Nuclear, microcornea,	(Hansen, et al., 2007b)
				c.346C>T	p.R116C	AD	Nuclear disc-like paediatric	(Li, et al., 2010)
				c.160C>T	p.R54C	AR	Total, Microcornea	(Khan, et al., 2007)
				c.62G>T	p.R21L	-	Central posterior	(Graw, et al., 2006)
				c.145C>T	p.R49C	-	-	(Hansen, et al., 2009)
CRYAB	11q22.1-q23.2	NM_009964.2	123590	c.527A>G	p.X176Trp	-	-	(van der Smagt, et al., 2014)
				c.32G>A	p.R11H	AD	Nuclear cataract	(Chen, et al., 2009)
				c.59C>G	p.P20R	AD	Posterior polar	(Xia, et al., 2014)
				c.58C>T	p.P20S	AD	Posterior polar	(Liu, et al., 2006a)
				c.166C>T	p.R56W	AR	Juvenile	(Safieh, et al., 2009)
				c.205C>T	p.R69C	AD	-	(Sun, et al., 2011a)
				c.320G>T	p.R107L	AD	AD	(Ma, et al., 2016)
				c.418G>A	p.D140N	AD	Lamellar	(Liu, et al., 2006b)
				c.450delA	p.K150NfsX35	AD	Posterior polar	(Berry, et al., 2001)

				c.511G>A	p.A171T		Lamellar	(Devi, et al., 2008)
CRYBA1/A3	17q11.2-q12	NM_009965.2	123610	c.215+1G>A		AD	Zonular	(Kannabiran, et al., 1998)
				c.215+2T>G		AD	Nuclear	(Yang, et al., 2011c)
				c.272_274delGAG	p.G91del	AD	Nuclear	(Qi, et al., 2004)
				c.626C>G	p.S209W	-	-	(Gillespie, et al., 2014)
				c.590-591delAG	p.E197VfsX22	AD	-	(Zhang, et al., 2014)
CRYBA4	22q12.1	NM_021351.1	123631	c.190G>T	p.G64W		Nuclear, Microcornea	(Zhou, et al., 2010a)
				c.242C>T	p.L69P	AD	Lamellar, Microphthalmia	(Billingsley, et al., 2006)
				c.317T>C	p.F94S	AD	Lamellar, Microphthalmia	(Billingsley, et al., 2006)
CRYBB1	22q12.1	NM_023695.2	600929	c.2T>A	p.M1K	AR	Pulverulent	(Meyer, et al., 2009)
				c.171delG	p.G57GfsX107	AR	-	(Cohen, et al., 2007)
				c.368G>A	p.R123H	AR	-	(Ma, et al., 2016)
				c.658G>T	p.G220X	AD	Lamellar	
				c.667C>T	p.Q223X	AD	Nuclear	(Yang, et al., 2008)
				c.682T>C	p.S228P	AD	-	(Wang, et al., 2007)
				c.698G>A	p.R233H	AD	Nuclear	(Wang, et al., 2011b)
				c.757T>C	p.X253RextX27	AD	Paediatric, Microcornea	(Willoughby, et al., 2005)
CRYBB2	22q11.23	NM_007773.3	123620	c.5C>T	p.A2V	AD	Posterior subcapsular	(Yao, et al., 2011a)
				c.326T>A	p.I109N	AD	-	(Sun, et al., 2014)
				c.455T>G	p.V152G	-	-	(Gillespie, et al., 2014)
				c.54G>A	p.K18KfsX17	-	Zonular	(Santhiya, et al., 2010)
				c.62T>A	p.I21N	-	Nuclear	(Wang, et al., 2011b)
				c.92C>G	p.S31W	AD	Coronary paediatric	(Lou, et al., 2009)
				c.177G>C	p.W59C	AD	Total	(Santhiya, et al., 2010)
				c.383A>T	p.D128V	AD	-	(Pauli, et al., 2007)
				c.343C>A	p.P115T	AD	-	(Ma, et al., 2016)
				c.355G>A	p.G119R	AD	-	(Ma, et al., 2016)
				c.433C>T	p.R145W	-	-	(Hansen, et al., 2009)
				c.440A>G	p.Q147R	AD	-	(Hansen, et al., 2009)
				c.436G>A	p.V146M	AD	Nuclear, Microcornea	(Wang, et al., 2011b)
				c.449C>T	p.T150M	AD	-	(Hansen, et al., 2009)
				c.493G>C	p.G165R	AR	-	(Ma, et al., 2016)
				c.452G>C	p.W151C	AD	-	(Pauli, et al., 2007)
				c.556T>C	p.S186P	-	-	(Ma, et al., 2016)
				c.463C>T	p.Q155X	AD	Cerulean	(Litt, et al., 1997)
				c.477C>A	p.Y159X	-	-	(Hansen, et al., 2009)

				c.583T>G	p.W195G	AD	-	(Ma, et al., 2016)
				c.607G>A	p.V187M	AD	Nuclear	(Mothobi, et al., 2009)
				c.563G>A	p.R188H	-	-	(Weisschuh, et al., 2012)
CRYBB3	22q11.23	NM_021352.3	123630	c.224G>A	p.R75H	AD	-	(Hansen, et al., 2009)
				c.493G>C	p.G165R	AR	-	(Ma, et al., 2016)
				c.581T>A	p.V194E	AD	-	(Reis, et al., 2013)
				c.634T>C	p.*212Rext*40	AD	-	(Ma, et al., 2016)
CRYGC	2q33-q35	NM_007775.2	123680	c.13A>C	p.T5P	-	Central zonular pulverulent	(Heon, et al., 1999)
				c.119-123dup5bp	p.C42AfsX63	AD	Zonular pulverulent	(Ren, et al., 2000)
				c.143G>A	p.R48H	AD	-	(Kumar, et al., 2011)
				c.328_329delinsT	p.Pro110Serfs*37	AD	-	(Ma, et al., 2016)
				c.470G>A	p.W157X	AD	Nuclear, Microcornea	(Guo, et al., 2012)
				c.502C>T	p.R168W	AD	-	(Gonzalez-Huerta, et al., 2007)
				c.124delT	p.C42AfsX60	AD	-	(Kondo, et al., 2013)
				c.134T>C	p.L45P	-	-	(Gillespie, et al., 2014)
				c.417C>G	p.Y139X	AD	-	(Reis, et al., 2013)
				c.497C>T	p.S166F	AD	-	(Prokudin, et al., 2014)
CRYGD	2q33-q35	NM_007776.2	123690	c.43C>T	p.R15C	AD	Progressive	(Stephan, et al., 1999)
				c.70C>A	p.P24T	AD	lamellar	(Santhiya, et al., 2002)
				c.106G>C	p.A36P	AD	Nuclear	(Sun, et al., 2011a)
				c.109C>A	p.R36S	AD	Nuclear	(Gu, et al., 2006)
				c.110G>C	p.R37P	AD	Nuclear	(Wang, et al., 2011c)
				c.127T>C	p.W43R	AD	Nuclear	(Wang, et al., 2011c)
				c.168C>G	p.Y56X	AD	Nuclear	(Santana, et al., 2009)
				c.176G>A	p.R59H	AD	-	(Heon, et al., 1999)
				c.181G>C	p.G61C	AD	Coralliform	(Li, et al., 2008)
				c.229C>A	p.R77S	AD	Juvenile cataract	(Roshan, et al., 2010)
				c.320A>C	p.E107A	AD	Nuclear	(Messina-Baas, et al., 2006)
				c.402C>A	p.Y134X	AD	Microcornea	(Hansen, et al., 2007b)
				c.418C>T	p.R140X	AD	Nuclear	(Devi, et al., 2008)
				c.448dup	p.D150Gfs*3	AD	-	(Ma, et al., 2016)
				c.470G>A	p.W157X	AD	-	(Santhiya, et al., 2002)
				c.494delG	p.G165AfsX3	AD	Nuclear	(Zhang, et al., 2007)
				c.451_452insGACT	p.Y151X	AD	Nuclear	(Zhuang, et al., 2015)
				c.453T>G	p.Y151X	-	-	(Gillespie, et al., 2014)
CRYGS	3q25-qter	NM_017541.2	123730	c.77A>G	p.D26G	-	-	(Sun, et al., 2011a)

c.116C>G	p.S39C	AD	-	(Devi, et al., 2008)
c.124G>A	p.V42M	AD	-	(Vanita, et al., 2006b)
c.169G>T	p.G57W	AD	Pulverulent	(Yang, et al., 2015b)

Appendix 2

Mutations reported in membrane or cytoskeleton proteins encoding genes associated with paediatric cataracts. Inh, inheritance; AD, autosomal dominant; AR, autosomal recessive; XL; X linked. "-"indicates no information was available.

Gene	Locus	Accessions	OMIM #	DNA Change	Protein Change	Inh	Phenotype	Reference
BFSP1	20p11.23-p12.1	NM_001195.3	603307	c.736-957del	p.T246del74fsX6	AR	Juvenile-onset cortical	(Ramachandran, et al., 2007)
				c.812T>C	p.I271T	AR	-	(Ma, et al., 2016)
				c.1042G>A	p.D348N		Nuclear	(Wang, et al., 2013)
				c.1492del	p.S498Lfs*24	-	-	(Ma, et al., 2016)
BFSP2	3q21-q22	NM_003571.2	603212	c.598_599dupAA	p.A201RfsX19	AR	Juvenile-onset	(Aldahmesh, et al., 2011)
				c.335G>A	p.G112E	AD	Pulverulent	(Liu, et al., 2014)
				c.697_699delGAA	p.E233del		-	(Gillespie, et al., 2014)
				c.697_699delGAA	p.E233del	AD	-	(Jakobs, et al., 2000)
				c.859C>T	p.R287W	AD	Nuclear lamellar	(Conley, et al., 2000)
				c.1016G>A	p.R339H	AD	Lamellar	(Ma, et al., 2008)
GJA1	6q21-q23.2	NM_000165.3	121014	c.559C>T	p.P187S		Nuclear pulverulent	(Ding, et al., 2011)
GJA3	13q11-q12	NM_021954.3	121015	c.5G>A	p.G2D	AD	Nuclear pulverulent	(Yao, et al., 2011b)
				c.7G>T	p.D3Y	AD	Zonular pulverulent	(Addison, et al., 2006)
				c.7G>C	p.D3H	AD	-	(Ma, et al., 2016)
				c.32T>C	p.L11S	AD	Ant-egg	(Hansen, et al., 2006)
				c.56C>T	p.T19M	AD	Posterior polar	(Santhiya, et al., 2010)
				c.82G>A	p.V28M	AD	Total cortical	(Devi, et al., 2005)
				c.96C>A	p.F32L	AD	Nuclear pulverulent	(Jiang, et al., 2003)
				c.98G>T	p.R33L	AD	Granular embryonal	(Guleria, et al., 2007a)
				c.130G>A	p.V44M	-	Nuclear	(Zhou, et al., 2010b)
				c.148T>C	p.S50P	-	-	(Gillespie, et al., 2014)
				c.176C>T	p.P59L	AD	-	(Ma, et al., 2016)
				c.260C>T	p.T87M	AD	-	(Guleria, et al., 2007b)
				c.578T>C	p.F193S	-	-	(Gillespie, et al., 2014)
				c.596A>C	p.E199A	-	-	(Gillespie, et al., 2014)
				c.268C>T	p.L90F	-	-	(Yang, et al., 2015a)

				c.428G>A	p.G143E	AD	Nuclear	(Yuan, et al., 2015)
				c.1361insC	p.A397GfsX71	AD	-	(Zhou, et al., 2013)
				c.134G>C	p.W45S	AD	Nuclear	(Ma, et al., 2005)
				c.139G>A	p.D47N	AD	Nuclear	(Yang, et al., 2011a)
				c.176C>T	p.P59L	AD	Nuclear punctate	(Bennett, et al., 2004)
				c.188A>G	p.N63S	AD	-	(Mackay, et al., 1999)
				c.1137insC	p.S380QfsX87	AD	-	(Mackay, et al., 1999)
				c.226C>G	p.R76G	AD	Total	(Devi, et al., 2005)
				c.227G>A	p.R76H	AD	Nuclear lamellar	(Burdon, et al., 2004b)
				c.260C>T	p.T87M	AD	Pearl box	(Guleria, et al., 2007b)
				c.427G>A	p.G143R	AD	Coppock-like	(Zhang, et al., 2012)
				c.560C>T	p.P187L	AD	Zonular pulverulent	(Rees, et al., 2000)
				c.563A>C	p.N188	AD	Nuclear	(Li, et al., 2004)
				c.1143_1165del23	p.381fs*48	AD	-	(Sun, et al., 2011a)
				c.1361insC	p.Ala397Glyfs×71	AD	-	(Zhou, et al., 2013)
GJA8	1q21.1	NM_005267.4	600897	c.68G>C	p.R23T	AD	Nuclear	(Willoughby, et al., 2003)
				c.89dupT	p.I31Hfs*18	-	-	(Ma, et al., 2016)
				c.131T>C	p.V44A	AD	Nuclear	(Zhu, et al., 2014)
				c.92T>C	p.I31T	AD	Nuclear	(Wang, et al., 2009a)
				c.116C>G	p.T39R	AD	Microcornea	(Sun, et al., 2011b)
				c.131T>A	p.V44E	AD	Microcornea	(Devi and Vijayalakshmi, 2006)
				c.134G>C	p.W45S	AD	Jellyfish-like	(Vanita, et al., 2008)
				c.137G>T	p.G46V	AD	Total	(Minogue, et al., 2009)
				c.136G>A	p.G46R	AD	Microcornea	(Sun, et al., 2011b)
				c.139G>T	p.D47Y	AD	-	(Lin, et al., 2008)
				c.142G>A	p.E48K	AD	Zonular nuclear	(Berry, et al., 1999)
				c.151G>A	p.D51N	-	-	(Ma, et al., 2016)
				c.191T>G	p.V64G	AD	Nuclear	(Zheng, et al., 2005)
				c.218C>T	p.S73F	AD	-	(Hansen, et al., 2009)
				c.235G>C	p.V79L	AD	Sutural opacities	(Vanita, et al., 2006a)
				c.262C>T	p.P88S	AD	Zonular pulverulent	(Shiels, et al., 1998)
				c.566C>T	p.P189L	AD	Nuclear, Microcornea	(Hansen, et al., 2007b)
				c.592C>T	p.R198W	AD	Microcornea	(Hu, et al., 2010)
				c.593G>A	p.R198Q	AD	Posterior subcapsular and Microcornea	(Devi and Vijayalakshmi, 2006)
				c.608insA	p.T203NfsX47	AR	-	(Ponnam, et al., 2007)

				c.741T>G	p.I247M	AD	Bilateral	(Graw, et al., 2009)
				c.773C>T	p.S258F	AD	Nuclear	(Gao, et al., 2010)
				c.836C>A	p.S259Y	AD	-	(Hansen, et al., 2009)
				c.827C>T	p.S276F	AD	Pulverulent nuclear	(Yan, et al., 2008)
				c.842T>C	p.L281C	AD	Lamellar, Nystagmus	(Kumar, et al., 2011)
				c.200A>G	p.D67G	AD	-	(Reis, et al., 2013)
				c.226A>G	p.R76C	AD	-	(Reis, et al., 2013)
				c.566C>T	p.P189L	AD	Total	(Gillespie, et al., 2014)
				c.649G>A	p.V196M	AR	-	(Ponnam, et al., 2009)
				c.829C>T	p.H277Y	AD	Nuclear	(Chen, et al., 2015)
				c.1273C>T	p.R425X		-	(Gillespie, et al., 2014)
LIM2	19q13.4	NM_005267.4	154045	c.313T>G	p.F105V	AR	-	(Pras, et al., 2002)
				c.462G>A	p.G154E	AR	Congenital or juvenile-onset cataract, Nystagmus, amblyopia	(Ponnam, et al., 2008)
MIP	12q13	NM_012064.3	154050	c.2T>C	p.M1T	AD	Cerulean	(Xiao, et al., 2011)
				c.97C>T	p.R33C	AD	Total	(Gu, et al., 2007)
				c.319G>A	p.V107I	AD	Nuclear pulverulent	(Wang, et al., 2010b)
				c.401A>G	p.E134G	AD	lamellar	(Berry, et al., 2000)
				c.413C>G	p.T138R	AD	Progressive punctuate	(Berry, et al., 2000)
				c.530A>G	p.Y177C	AD	Nuclear cataract	(Yang, et al., 2011b)
				c.559C>T	p.R187C	-	Nuclear cataract	(Wang, et al., 2011a)
				g.IVS3-1G>A	p.V203fs	AD	Snail-like	(Jiang, et al., 2009)
				c.638delG	p.G213VfsX46	AD	Nuclear	(Geyer, et al., 2006)
				c.702G>A	p.R233K	AD	Binocular polymorphic	(Lin, et al., 2007)
				c.448G>C	p.D150H	AD	Cortical punctate	(Shentu, et al., 2015)
				c.494G>A	p.G165D	AD	Lamellar cataract	(Senthil Kumar, et al., 2013)
				c.597_598ins GGGAACATTCCACT	p.N200Gfs*12	AD		(Ma, et al., 2016)
				c.605G>A	p.W202X	AD	-	(Reis, et al., 2013)
				c.606+1G>A	-	AD	-	(Zeng, et al., 2013)
				c.657C>G	p.Y219X	AD	Posterior polar	(Song, et al., 2015)
				c.644G>A	p.G215D	AD	Progressive punctate	(Ding, et al., 2014)
VIM	10p13	NM_203472.1	193060	c.451G>A	p.E151K	AD	-	(Muller, et al., 2009)
				c.15del	p.Val6Cysfs*26	AD	-	(Ma, et al., 2016)
PVRL3	3q13	NM_001243288.1	607147	t(1;3)(q31.3;q13.13)	-	-	Developmental delay	(Lachke, et al., 2012)

CHMP4B	20q11.22	NM_176812.4	610897	c.386A>T	p.D129V	AD	Posterior subcapsular	(Shiels, et al., 2007)
				c.481G>A	p.E161K	AD	Posterior polar	(Shiels, et al., 2007)
FYCO1	3p21.31	NM_024513.3	607182	c.1045C>T	p.Q349X	AR	-	(Chen, et al., 2011)
				c.1546C>T	p.Q516X	AR	-	(Chen, et al., 2011)
				c.2206C>T	p.Q736X	AR	-	(Chen, et al., 2011)
				c.2761C>T	p.R921X	AR	-	(Chen, et al., 2011)
				c.2830C>T	p.R944X	AR	-	(Chen, et al., 2011)
				c.3150+1G>T	-	AR	-	(Chen, et al., 2011)
				c.3755delC	p.A1252DfsX71	AR	-	(Chen, et al., 2011)
				c.3858_3862dupGGAA	p.L1288WfsX37	AR	-	(Chen, et al., 2011)
				c.4127T>C	p.L1376P	AR	-	(Chen, et al., 2011)
				c.2505del	p.A836PfsX80	AR	-	(Aldahmesh, et al., 2012b)
				c.3670C>T	p.R1224X		-	(Gillespie, et al., 2014)
				c.3945-1G>C			-	(Gillespie, et al., 2014)
TMEM114	16p13.2	NM_001146336.1	611579	t(16;22)(p13.3;q11.2)		AD	-	(Jamieson, et al., 2007)
				c.104T>C	p.I35T	AD	-	(Jamieson, et al., 2007)
				c.318T>G	p.F106L	AD	-	(Jamieson, et al., 2007)

Appendix 3

Genes encoding transcription factor and signalling molecules associated with paediatric. Inh, inheritance; AD, autosomal dominant; AR, utosomal recessive; XL; X linked. "-"indicates no information was available.

Gene	Locus	accessions	OMIM #	DNA change	Protein change	Inh	Phenotype	Reference
FOXE3	1p32	NM_012186.2	601094	c.942dupG	p.L315AfsX117	-	Posterior embryotoxin	(Semina, et al., 2001)
			602438	c.958T>C	p.X320RextX72	AD	Paediatric cataract, aphakia, sclerocornea	(Iseri, et al., 2009)
				c.959G>C	p.X320SextX72	AD	-	(Bremond-Gignac, et al., 2010)
				c.959G>T	p.X320LextX72	AD	Paediatric cataract , Anterior segment dysgenesis	(Doucette, et al., 2011)
HSF4	16q21	NM_012186.2	602438	c.56C>A	p.A19D	AD	-	(Bu, et al., 2002)
				c.218G>A	p.R73H	AD	Total cataract	(Ke, et al., 2006)
				c.256A>G	p.I86V	AD	Cortical lamellar	(Bu, et al., 2002)
				c.341T>C	p.L114P	AD	Lamellar	(Bu, et al., 2002)
				c.355C>T	p.R119C	AD	Zonular stellate & anterior polar	(Bu, et al., 2002)
				c.595-599del5bp	p.G199EfsX15	AR	-	(Forshew, et al., 2005)
				c.524G>C	p.R175P	AR	Nuclear & cortical	(Forshew, et al., 2005)
				c.1213C>T	p.R405X	AR	-	(Sajjad, et al., 2008)
				c.1327+4A>G	p.M419GfsX29	AR	Total	(Smaoui, et al., 2004)
				c.69G>T	p.K23N	AD	-	(Lv, et al., 2014)
				c.103C>T	p.H35Y		-	(Gillespie, et al., 2014)
MAF	16q22-q23	NM_005360.4.	177075	t(5;16)(p15.3;q23.2)	-	AD	Progressive posterior subcapsular, Peters anomaly, myopia, developmental delay	(Jamieson, et al., 2002)
				c.863G>C	p.R288P	AD	Cortical nuclear pulverulent	(Jamieson, et al., 2002)
				c.880C>T	p.R294W	AD	Nuclear	(Sun, et al., 2014)
				c.890A>G	p.K297R	AD	Cerulean, Microcornea	(Vanita, et al., 2006b)

				c.895C>A	p.R299S	AD	Lamellar, Microcornea	(Hansen, et al., 2007a)
				c.958A>G	p.K320E	-	Nuclear peadiatric , Microcornea	(Hansen, et al., 2007a)
				c.161C>T	p.S54L	-	Aymé-Gripp Syndrome	(Niceta, et al., 2015)
				c.172A>G	p.T58A	-	Aymé-Gripp Syndrome-	(Niceta, et al., 2015)
				c.173C>T	p.T58I	-	Aymé-Gripp Syndrome	(Niceta, et al., 2015)
				c.176C>A	p.P59H	-	Aymé-Gripp Syndrome	(Niceta, et al., 2015)
				c.176C>T	p.P59L	-	Aymé-Gripp Syndrome	(Niceta, et al., 2015)
				c.185C>G	p.T62R	-	Aymé-Gripp Syndrome	(Niceta, et al., 2015)
				c.206C>G	p.P69R	-	Aymé-Gripp Syndrome	(Niceta, et al., 2015)
				c.908A>C	p.Q303L	AD	-	(Narumi, et al., 2014)
				c.915C>T	p.C305W	AD	-	(Ma, et al., 2016)
PITX3	10q25	NM_005029.3	602669	c.650delG	p.G217AfsX91	AD	Posterior polar, Neurodevelopmental abnormalities	(Bidinost, et al., 2006)
				c.640_656dup17bp	p.G220PfsX95	AD	Anterior cortical, Anterior segment mesynchymal dysgenesis (ASMD)	(Semina, et al., 1998)
				c.573delC	p.S192AfsX117	AD	-	(Verdin, et al., 2014)
PITX2	4q25	NM_011098.3	601542	19.2 mb deletion			Paediatric cataract & aniridia, ARS,CGL	(Reis, et al., 2012)
EPHA2	1p36	NM_004431.3	602756	c.1405T>C	p.Y469H	AR	Paediatric cataract , Persisatent fetal vasculature	(Aldahmesh, et al., 2012b)
				c.2353G>A	p.A785T	AR	Nuclear	(Kaul, et al., 2010a)
				c.2819C>T	p.T940I	AD	Posterior polar	(Zhang, et al., 2009)
				g.IVS16-9G>A	p.D943PfsX71	AD	Total paediatric	(Zhang, et al., 2009)
				c.2842G>T	p.G948W	AD	Posterior polar	(Shiels, et al., 2008)
				c.2915-2916delTG	V972GfsX39	AD	Posterior polar	(Zhang, et al., 2009)
				c.1046C>T	p.T349M	AD	-	(Sun, et al., 2014)
				c.1059_1060dupCA	p.S354MfsX40	-	-	(Gillespie, et al., 2014)
				c.1751C>T	p.P584L	AD	Nuclear	(Dave, et al., 2013)
				c.2875G>A	p.A959T	AD	Subcapsular and cortical	(Dave, et al., 2013)
				c.2925dupC	p.I976HfsX37	AD	-	(Reis, et al., 2014)
VSX2	14q24.3	NM_182894.2	142993			AD	Paediatric cataract , Microphthalmia, Anophthalmia, Iris	(Ferda Percin, et al., 2000)

TDRD7	9q22.33	NM_014290.2	611258	inv(9)(q22.3-q34.11) c.1852_1854 del	coloboma, Juvenile cataract , hypospadias, de novo balanced paracentric inversion of chromosome 9	(Lachke, et al., 2011)
MIR184	Chr15	NR_038997.1	613146	r.57c>u	Anterior polar cataract, severe keratoconus	(Hughes, et al., 2011)

Appendix 4

Mutations reported in enzymatic and syndromic paediatric cataracts cases. Inh, inheritance; AD, autosomal dominant; AR, autosomal recessive; XL; X linked. "-"indicates no information was available.

Gene	Locus	Accessions	OMIM #	DNA change	Protein change	Inh	Phenotype	reference
NHS	Xp22.13	NM_001081052.1	302350	c.400delC	p.R134VfsX62	XL	Paediatric cataract , Dental anomalies	(Burdon, et al., 2003)
				c.2387insC	p.A796AfsX36	XL	Fetal nuclear	(Burdon, et al., 2003)
				c.718insG	p.G240fs	XL	-	(Burdon, et al., 2003)
				c.3459delC	p.A1153AfsX29	XL	-	(Burdon, et al., 2003)
				c.3624C>A	p.C1208*	XL	-	(Ma, et al., 2016)
				c.472C>T	p.Q158X	XL	Cortical wedge(in female);nuclear(in male), microcornea	(Coccia, et al., 2009)
				c.614delC	p.P206fsX282	XL	-	(Coccia, et al., 2009)
				c.742C>T	p.R248X	XL	Sutural cataract, microcornea, dental anomalies	(Sharma, et al., 2008)
				c.792delA	p.P264fs	XL	Posterior, microphthalmia, dental and ear anomalies	(Sharma, et al., 2008)
				c.1117C>T	p.R373X	XL	Dental anomalies	(Burdon, et al., 2003)
				c.2635C>T	p.R879X	XL	-	(Sharma, et al., 2008)
				c.3596insA	p.K1198fs	XL	Nuclear, Microcornea,secondry glaucoma	(Sharma, et al., 2008)
				3908del11bp	p.I1302fs	XL	Total, dental/craniofacial anomalies...	(Sharma, et al., 2008)
				c.1108C>T	p.Q370X	XL	Paediatric cataract ,	(Huang, et al., 2007)

							Microcornea, dental/ear anomalies	
				c.2550-2553del4bp	p.K850fsX852	XL	Bilateral paediatric cataract, dental anomalies	(Coccia, et al., 2009)
				0.92Mb deletion	-	XL	Paediatric cataract, dental anomalies, facial dysmorphisms, mental retardation	(Liao, et al., 2011)
				2.8Mb deletion	-	XL	Paediatric cataract, microcornea	(Van Esch, et al., 2007)
				c.558insA	p.E186Efs11X	XL	Paediatric cataract, mental retardation, dental anomalies	(Tug, et al., 2013)
				c.322G>T	p.E108X	XL	Cortical, dental anomalies, facial dysmorphism	(Hong, et al., 2014)
				c.556G>T	p.E186X	XL	Nuclear	(Sun, et al., 2014)
				c.852delG	p.S285PfsX13	XL	-	(Li, et al., 2015)
GALK1	17q24	NM_016905.2	604313	c.410delG	p.G137VfsX27	AR	Nuclear	(Yasmeen, et al., 2010)
				c.727T>C	p.C243R	-	-	(Gillespie, et al., 2014)
				c.416T>C	p.L139P	AR	-	(Yasmeen, et al., 2010)
				c.1144C>T	p.Q382X	AR	-	(Chacon-Camacho, et al., 2014)
AGK	7q34	NM_023538.2	610345	c.424_518del	p.A142TfsX4	AR	Paediatric cataract, Sengers syndrome	(Aldahmesh, et al., 2012b)
GCNT2	6p24.2	NM_008105.3	600429	Complete Del	-	-	Paediatric cataract, Adult i blood group	(Yu, et al., 2001)
				del (93kb)	-	AR	Nuclear	(Borck, et al., 2012)
				c.935G>A	p.G312D	AR	Paediatric cataract, Adult i blood group, leukocoria	(Wussuki-Lior, et al., 2011)
				c.983G>A	p.W328X	AR	Paediatric cataract Adult i blood group	(Pras, et al., 2004)
				c.1040A>G	p.Y347C	AR	-	(Aldahmesh, et al., 2012b)
				c.1049G>A	p.G350E	AR	Paediatric cataract, Adult i blood group	(Yu, et al., 2001)
				c.1154G>A	p.R385H	AR	Paediatric cataract, Adult i blood group	(Yu, et al., 2001)
FTL	19q13.33	NM_010240.2	134790	c.-220_-196del25		AD	Cataract, Hyperferritinemia	(Burdon, et al., 2007)

				c.-168_165delGCTT		-	-	(Garber and Pudek, 2014)
				c.-190_162del29		AD	Bilateral congenital, hyperferritinemia	(Girelli, et al., 1997)
				c.-182C>T		AD	Nuclear, hyperferritinemia	(Cazzola, et al., 1997)
				c.-171C>G	-	AD	Early onset of bilateral cataract, hyperferritinemia	(Bosio, et al., 2004)
				c.-168G>A	-	AD	Nuclear, hyperferritinemia	(Cazzola, et al., 1997)
				c.-167C>T	-	AD	Ferritin crystals, hyperferritinemia	(Brooks, et al., 2002)
				c.-164C>A	-	AD	Early onset cataracts, hyperferritinemia	(Brooks, et al., 2002)
				c.-159G>C	-	AD	Bilateral nuclear, hyperferritinemia	(Girelli, et al., 1997)
EYA1	8q13.3	NM_010164.2	601653	c.1177G>A	p.G393S	AD	Nuclear, brachio-oto-renal syndrome-1	(Azuma, et al., 2000)
				c.1320G>A	p.R407Q	AD	Nuclear-type brachio-oto-renal syndrome-1	(Azuma, et al., 2000)
PAX6	11p13	NM_0001604.4	607108	c.1119del	p.T374Pfs*5	AD		(Ma, et al., 2016)
				c.239T>A	p.I80N	AD		(Ma, et al., 2016)

Appendix 5

Loci linked with paediatric cataract. Inh, inheritance; AD, autosomal dominant; AR, autosomal recessive; XL; X linked.

Name	locus	OMIM #	Inh	Phenotype	Reference
CCV	1p36	115665	AD	Progressive central and zonoular nuclear	(Eiberg, et al., 1995)
CTTP	1p 34-p36	116600	AD	Posterior polar	(Ionides, et al., 1998; McKay, et al., 2005)
CCNP	2p 12	607304	AD	Nuclear progressive	(Khaliq, et al., 2002)
CCSSO	15q21-q22	212500	AR	Central pounchlike with sutural opacities	(Vanita, et al., 2001b)
CTAA2	17p13	601202	AD	Anterior polar	(Berry, et al., 1996; Ionides, et al., 1998)
CCA1	17q24	115660	AD	Nuclear and cortical	(Armitage, et al., 1995)
CPP3	20p12-q12	605387	AD	Congenital zonular nuclear , posterior subcapsular opacity	(Li, et al., 2006; Yamada, et al., 2000)
CXN	Xp22.13	300457	XL	Nuclear	(Francis, et al., 2002)
	7q21.11		AR	Nuclear	(Kaul, et al., 2010b)
	19q12- q13.12		AR	Nuclear	(Hejtmancik, 2008)
	19q13.42-qter		AD	Nuclear	(Zhao, et al., 2011)
	Xq24		XL	Nuclear and lamellar; subtle dysmorphic features	(Craig, et al., 2008; Zhao, et al., 2011)

Appendix 6

Mutations reported to be associated with primary congenital glaucoma (PCG).

Gene	Locus	Accessions	OMIM #	DNA Change	Protein Change	Inh	Reference
CYP1B1	2p22.2	NM_000104.3	231300	g.3976G>C	p.W57C	AR	(Stoilov, et al., 1998)
				g.3947C>G	p.R48G	AR	(Bejjani, et al., 2000)
				g.3987G>A	p.G61E	AR	(Stoilov, et al., 1998)
				g.4035T>C	p.L77P	AR	(Bejjani, et al., 2000)
				g.4155G>C	p.R117P	AR	(Hollander, et al., 2006)
				g.4157C>A	p.P118T	AR	(Vincent, et al., 2006)
				g.4160G>T	p.A119S	AR	(Bejjani, et al., 2000)
				g.4380A>T	p.D192V	AR	(Mashima, et al., 2001)
				g.4383C>T	p.P193L	AR	(Panicker, et al., 2002)
				g.4397G>A	p.V198I	AR	(Mashima, et al., 2001)
				g.4430T>C	p.C209R	AR	(Hollander, et al., 2006)
				g.4449G>T	p.S215I	AR	(Sitorus, et al., 2003)
				g.4490G>A	p.E229K	AR	(Michels-Rautenstrauss, et al., 2001)
				g.4763G>T	p.V320L	AR	(Mashima, et al., 2001)
				g.4793G>T	p.A330F	AR	(Mashima, et al., 2001)
				g.4793G>T	p.A330F	AR	(Mashima, et al., 2001)
				g.4838C>T	p.L345F	AR	(Vincent, et al., 2002)
				g.7927G>A	p.V364M	AR	(Mashima, et al., 2001)
				g.7930G>T	p.G365W	AR	(Stoilov, et al., 1998)
				g.7940G>A	p.R368H	AR	(Bejjani, et al., 2000)
				g.7983C>T	p.P379L	AR	(Stoilov, et al., 1998)
				g.7996G>A	p.E387K	AR	(Plasilova, et al., 1999)
				g.7999G>A	p.A388T	AR	(Alfadhli, et al., 2006)
				g.8005C>T	p.R390C	AR	(Curry, et al., 2004)
				g.8005C>A	p.R390S	AR	(Bejjani, et al., 2000)
				g.8006G>A	p.R390H	AR	(Stoilov, et al., 1998)
				g.8033T>G	p.I399S	AR	(Colomb, et al., 2003)
				g.8104A>T	p.N423Y	AR	(Colomb, et al., 2003)
				g.8131G>C	p.L432V	AR	(Stoilov, et al., 1997)

g.8147C>T	p.P437L	AR	(Stoilov, et al., 2002)
g.8165C>G	p.A443G	AR	(Stoilov, et al., 2002)
g.8168G>A	p.R444Q	AR	(Mashima, et al., 2001)
g.8242C>T	p.R469W	AR	(Bejjani, et al., 1998)
g.3860C>T	p.Q19X	AR	(Stoilov, et al., 2002)
g.3976G>A	p.W57X	AR	(Vincent, et al., 2001)
g.4645C>A	p.C280X	AR	(Mashima, et al., 2001)
g.4646G>T	p.G281X	AR	(Stoilov, et al., 1998)
g.7900C>T	p.R355X	AR	(Michels-Rautenstrauss, et al., 2001)
g8104A>T	p.N423Y	AR	(Melki, et al., 2004)
g.8139G>A	p.W434X	AR	(Chavarria-Soley, et al., 2006)
g.8167C>T	p.R444X	AR	(Colomb, et al., 2003)
g.3964delC	Frameshift	AR	(Mashima, et al., 2001)
g.3979delA	Frameshift + stop at 59	AR	(Colomb, et al., 2003)
g.4238_4247del	Frameshift	AR	(Bejjani, et al., 2000)
g.4081delC	-	AR	(Michels-Rautenstrauss, et al., 2001)
g.4339delG	-	AR	(Belmouden, et al., 2002)
g.4340delG	-	AR	(Stoilov, et al., 2002)
g.4356delG	A179R/X 17 aa downstream	AR	(Messina-Baas, et al., 2007)
g.4611_4619del	S268_F270del	AR	(Bejjani, et al., 2000)
g.4635delT	p.L277X	AR	(Messina-Baas, et al., 2007)
g.7901_7913del	fs and 422X	AR	(Stoilov, et al., 1997)
G7945delC	P370L / X 57 aa downstream	AR	(Messina-Baas, et al., 2007)
g.8182delG	-	AR	(Stoilov, et al., 1998)
g.8214_8215del	Frameshift	AR	(Stoilov, et al., 2002)
g.3956insC	Frameshift	AR	(Chavarria-Soley, et al., 2006)
g.4306insT	Frameshift	AR	(Stoilov, et al., 1998)
g.4673insC	Frameshift	AR	(Stoilov, et al., 1998)
g.4776insAT	-	AR	(Mashima, et al., 2001)
g.8037_8046dup	-	AR	(Stoilov, et al., 2002)
g.8039_8048	A179R/X 17 aa downstream	AR	(Messina-Baas, et al., 2007)
g.8240_8266dup	frameshift	AR	(Stoilov, et al., 1998)
g.3834insA	frameshift	AR	(Li, et al., 2011)
g. 3876T>G	p.L24R	AR	(Li, et al., 2011)
g. 3905del23	frameshift	AR	(Li, et al., 2011)
g.3913C>T	P.Q37X	AR	(Li, et al., 2011)

g.3960C>T	p. P52L	AR	(Li, et al., 2011)
g. 3972delC	frameshift	AR	(Li, et al., 2011)
g. 3985C>G	p. I60M	AR	(Li, et al., 2011)
g. 3988delA	frameshift	AR	(Li, et al., 2011)
g. 4004del8	frameshift	AR	(Li, et al., 2011)
g. 4046T>A	p.Y81N	AR	(Li, et al., 2011)
g. 4048C>A	p.Y81X	AR	(Li, et al., 2011)
g. 4052delG	frameshift	AR	(Li, et al., 2011)
g.4089T>C	p.V95A	AR	(Li, et al., 2011)
g.4122C>A	p.A106D	AR	(Li, et al., 2011)
g.4124C>G	p.L107V	AR	(Li, et al., 2011)
g.4168ins18a	Frameshift	AR	(Li, et al., 2011)
g.4196del5	Frameshift	AR	(Li, et al., 2011)
g.4200T>G	p.M132R	AR	(Li, et al., 2011)
g.4206T>C	p.F134S	AR	(Li, et al., 2011)
g.4236A>C	p.Q144P	AR	(Li, et al., 2011)
g.4236A>G	p.Q144R	AR	(Li, et al., 2011)
g.4259delAT	Frameshift	AR	(Li, et al., 2011)
g.4280C>T	p.Q159X	AR	(Li, et al., 2011)
g.4292C>T	p.R163C	AR	(Li, et al., 2011)
g.4322G>A	p.E173K	AR	(Li, et al., 2011)
4322G>T	p.E173X	AR	(Li, et al., 2011)
g.4330delTG	Frameshift	AR	(Li, et al., 2011)
g.4335T>G	p.L177R	AR	(Li, et al., 2011)
g.4335T>C	p.L177P	AR	(Li, et al., 2011)
g.4342delG	Frameshift	AR	(Li, et al., 2011)
g.4373T>C	p.F190L	AR	(Li, et al., 2011)
g.4375C>A	p.F190L	AR	(Li, et al., 2011)
g.4379G>T	p.D192Y	AR	(Li, et al., 2011)
g.4410C>A	p.A202D	AR	(Li, et al., 2011)
4413A>G	p.N203S	AR	(Li, et al., 2011)
g.4490G>A	p.E229K	AR	(Li, et al., 2011)
g.4499G>C	p.G232R	AR	(Li, et al., 2011)
g.4520A>C	p.S239R	AR	(Li, et al., 2011)
g.4523delC	p.S239R	AR	(Li, et al., 2011)
g.4530dup16/del6	Frameshift	AR	(Li, et al., 2011)

g.4531del22	Frameshift	AR	(Li, et al., 2011)
g.4547C>T	p.Q248X	AR	(Li, et al., 2011)
g.4578C>A	p.F261L	AR	(Li, et al., 2011)
g.4589G>T	p.E262X	AR	(Li, et al., 2011)
g.4611dup9	Frameshift	AR	(Li, et al., 2011)
g.4602del9	In-frame deletion	AR	(Li, et al., 2011)
g.4633delC	Frameshift	AR	(Li, et al., 2011)
g.4635delT	Frameshift	AR	(Li, et al., 2011)
g.4640C>G	p.H279D	AR	(Li, et al., 2011)
g.4650G>A	p.S282N	AR	(Li, et al., 2011)
g.4664G>A	p.A287S	AR	(Li, et al., 2011)
g.4668insC	Frameshift	AR	(Li, et al., 2011)
g.4677A>G	p.D291G	AR	(Li, et al., 2011)
g.4680–4681TG>AA	p.M292K	AR	(Li, et al., 2011)
g.4761A>G	p.N319S	AR	(Li, et al., 2011)
g.4763G>T	p.V320L	AR	(Li, et al., 2011)
g.4791G>T	p.G329V	AR	(Li, et al., 2011)
g.4791G>A	p.G329D	-	(Li, et al., 2011)
g.4793–4794GC>TT	p.A330F	AR	(Li, et al., 2011)
g.4812C>A	p.S336Y	AR	(Li, et al., 2011)
g.4825G>T	p.Q340H	AR	(Li, et al., 2011)
g.4828G>A	p.W341X	AR	(Li, et al., 2011)
g.4849delb	Frameshift	AR	(Li, et al., 2011)
g.7899del12	In-frame deletion	AR	(Li, et al., 2011)
g.7925T>A	p.V363D	AR	(Li, et al., 2011)
g.7930G>T	p.G365W	AR	(Li, et al., 2011)
g.7934delG	Frameshift	AR	(Li, et al., 2011)
g.7939C>T	p.R368C	AR	(Li, et al., 2011)
g.7940G>T	p.R368L	AR	(Li, et al., 2011)
g.7945delC	frameshift	AR	(Li, et al., 2011)
g.7957G>A	p.D374N	AR	(Li, et al., 2011)
g.7959C>G	p.D374E	AR	(Li, et al., 2011)
g.7970T>A	p.L378Q	AR	(Li, et al., 2011)
g.7990C>T	p.L385F	AR	(Li, et al., 2011)
g.7996G>A	p.E387K	AR	(Li, et al., 2011)
g.7999G>A	p.A388T	AR	(Li, et al., 2011)

g.8005C>T	p.R390C	AR	(Li, et al., 2011)
g.8005C>A	p.R390S	AR	(Li, et al., 2011)
g.8006G>A	p.R390H	AR	(Li, et al., 2011)
g.8033T>G	p.I399S	AR	(Li, et al., 2011)
g.8035C>T	p.P400S	AR	(Li, et al., 2011)
g.8037dup10	Frameshift	AR	(Li, et al., 2011)
g.8047dup10	Frameshift	AR	(Li, et al., 2011)
g.8104A>T	p.N423Y	AR	(Li, et al., 2011)
g.8111insG	Frameshift	AR	(Li, et al., 2011)
g.8127C>G	p.D430E	AR	(Li, et al., 2011)
g.8131C>G	p.L432V	AR	(Li, et al., 2011)
g.8139G>A	p.W434X	AR	(Li, et al., 2011)
g.8167C>T	p.R444X	AR	(Li, et al., 2011)
g.8147C>T	p.P437L	AR	(Li, et al., 2011)
g.8162C>G	p.P442R	AR	(Li, et al., 2011)
g.8165C>G	p.A443G	AR	(Li, et al., 2011)
g.8168G>A	p.R444Q	AR	(Li, et al., 2011)
g.8170T>A	p.F445I	AR	(Li, et al., 2011)
g.8171T>G	p.F445C	AR	(Li, et al., 2011)
g.8171T>C	p.F445S	AR	(Li, et al., 2011)
g.8182delG	Frameshift	AR	(Li, et al., 2011)
g.8214dup27	Frameshift	AR	(Li, et al., 2011)
g.8214delAG	Frameshift	AR	(Li, et al., 2011)
g.8234G>A	p.G466D	AR	(Li, et al., 2011)
g.8240dup27	Frameshift	AR	(Li, et al., 2011)
g.8242C>T	p.R469W	AR	(Li, et al., 2011)
g.8246G>A	p.C470Y	AR	(Li, et al., 2011)
g.8249T>G	p.I471S	AR	(Li, et al., 2011)
g.8297T>C	p.L487P	-	(Li, et al., 2011)
g.8329A>G	p.N498D	AR	(Li, et al., 2011)
g.8333A>G	p.E499G	AR	(Li, et al., 2011)
g.8341delA	Frameshift	-	(Li, et al., 2011)
g.8405G>A	p.R523K	-	(Li, et al., 2011)
g.4148G>C	p.A115P	AR	(Li, et al., 2011)
g.8209del5ins11c	Frameshift	AR	(Li, et al., 2011)
g.1793delC	Frameshift	AR	(Al-Haddad, et al., 2015)

LTBP2	14q24	NM_000428.2	602091	c.412 delG	p.A138PfsX278	AR	(Ali, et al., 2009)
				c.895C >T	p.R299X	AR	(Ali, et al., 2009)
				c.1243-1256 del	p.E415RfsX596	AR	(Ali, et al., 2009)
				c.331C >T	p.Q111X	AR	(Ali, et al., 2009)
				c.1415delC	p.Ser472fsX3	AR	(Narooie-Nejad, et al., 2009)
				c.5376delC	p.Tyr1793fsX55	AR	(Narooie-Nejad, et al., 2009)
PXDN	2p25.3	NM_012293.2	605158	c.2568delC	p.Cys857AlafsX5	AR	(Khan, et al., 2011)
				c.2638C>T	p.Arg880Cys	AR	(Khan, et al., 2011)
				c.1021C>T	p.Arg341X	AR	(Khan, et al., 2011)
MYOC	1q25		601652	c.144G>T	Q48H	AR	(Kaur, et al., 2005)
FOXC1				g.1457A>G	H128R	AR	(Chakrabarti, et al., 2009)
				g.2713G.A	C135Y	AR	(Chakrabarti, et al., 2009)
				g.1086delC	Frameshift	AR	(Chakrabarti, et al., 2009)
				g.1155del9bp	Frameshift	AR	(Chakrabarti, et al., 2009)
				g.1947dup25bp	Frameshift	AR	(Chakrabarti, et al., 2009)
				c.889C>T	P297S	-	(Medina-Trillo, et al., 2015)
				c.1134-1144del	G380Rfs*144	-	(Medina-Trillo, et al., 2015)

Appendix 7

List of primers used for sequencing of the translated regions of *TEK* gene in 42 unsolved PCG cases.

Gene	Region	Forward Primer	Reverse Primer	Product Size
TEK	Exon 1	CCGGGAGAGCTGTTAGAAGTC	CACAAATGTGCATGAGGTCC	491
TEK	Exon 1	TACAGCCTGCTTCTGTGCTG	TTCTAACACAGGGGCCAATC	443
TEK	Exon 2	TTTGTGAGCACCAGTTTACCC	AGCTGCCAAGACAAAAGGTG	458
TEK	Exon 3	CACAAAGTCCTCATCTCCCC	GCCCACAAGACCACAATAGG	496
TEK	Exon 4	AGCATTTTCATTCTTCTCCCAC	TGCTTTGCAGCATAAGAACAG	450
TEK	Exon 5 Exon 5-v2	TTGTTTCTGTCGCGATTTCC CCATTGTCCACTGAATGACTG	CAGTGGTTGGCATTAAAGAACC AATCATGCAGGTTTGTGCAG	505 500
TEK	Exon 6	TGGGCCATAAGGGTATGTTC	GTGGGTAGCTAAGCAGTCCAG	404
TEK	Exon 7	CAAATCAGCAAAGTTGATGGC	AAGAGCACGTGGTTTATGCC	469
TEK	Exon 8	GACTGTTCCTCCCTGGTCC	AAAGCCAACAACACACTAGCC	568
TEK	Exon 9	ATCTGACAGGGCATCAATCC	CCTGGAAATTACCCCAAAGG	466
TEK	Exon 10	AAGAGGACTTTGTTGGACATGTAA	AGAGCGCACAGTGACTTCAG	475
TEK	Exon 11	ATCGCAATAACAACAACCCC	CCTTCCTGAAGGCTCATCTG	495
TEK	Exon 12	CCTCCAACACTGGGGATTAC	GGGGGCACTAATTAGCCTTC	526
TEK	Exon 13	CATTTGTTTGCCTTATATGAGCTG	TCATAGGCTTTCATCTCACAGG	500
TEK	Exon 14	TGCTGTTAAGTTCCCATTACACTG	AATGAAAGCCAAAGAGAAGATGAG	400
TEK	Exon 15	GTGGATGCCAACCAGAAGAC	TTAAATGGCTCCTGCCTTTG	493
TEK	Exon 16	AGCTGAAGGTTCTTAGGGGG	CCATGAATACCTTTGGGCAG	451
TEK	Exon 17	GCCCCTTTTGAGTATTGCAG	AAAGAAGAGAGGTTGACAGACCC	475
TEK	Exon 18 Exon 18-v2	TGTTCCCCAAAGTTTTCAGC TGTGTCATTTGGCAGAATCC	TTGGGGTCTGGACTGGATAG GCTTCAGTCACCACAGAGCA	470 465
TEK	Exon 19	GGAATGACCGACTACCATGC	TTGTTCCCGAGAGCTACAGG	432
TEK	Exon 20	TGTTTTGGACAGTACCACCTTG	CACATACAAACGGCATCCTG	404
TEK	Exon 21	TCACCCTCTCTTGCCATACC	CTTTTCTGCATTTGTTGCCC	481
TEK	Exon 22	CCAAGGTCCTGCAGAAACAG	AATACTTAGGGCCATGCCAAG	444
TEK	Exon 23	TATTGCTGTAACTGCCGCTG	CGCCTTCCTATGAAGTCCAC	436
TEK	Exon 23	GCTGTACACCTGGGACCTTC	CTAAATGAAACGGGACTGGC	454
TEK	Exon 23	TGCATAACTCATTGTTGTCCTAGA	ATGTCTCCCAAATGTCACCC	483
TEK	Exon 23	TGTCTTGTGTTTCCACAGCC	TCTGATGAGCTTCAGATTCCG	497
GREB1		TGAGATGGGCCACTCCTG	GAGCTTGGGCAAGAAGACAA	542

Primers were designed by AGRF (Brisbane, Australia. The Exon 5-V2 and Exon 18-V2 primer sets were designed in house. The primers set used to validate the variations in *GREB1* are also given

Appendix 8

The coverage/read data statistics on PCG002, PCG002.2, PCG002.3 and PCG002.4.

Sample	Total reads (Mb)	Average throughput depth of target regions (X)	% Coverage of target regions (more than 8X)
PCG002	118.47	62.12	92
PCG002.2	118.35	59.53	93
PCG002.3	116.93	58.85	92
PCG002.4	126.13	67.76	93

Table shows total reads per sample, average read depth of target region and coverage percentage of the target region covered more than 8 fold for sequenced members (provided by our collaborator in State Key Laboratory of Medical Genetics, Central South University, Changsha, China).

Appendix 9

Heterozygous novel protein changing variants shared between affected children in family PCG002.

Chr	Position	Reference	Sample	Accession	Function	Amino Acids	Poly Phen	Gene
1	63282320	T	C/T	XM_005271288.1	missense	PHE,LEU	probably-damaging	ATG4C
1	160063544	G	A/G	XM_005245615.1	missense	ALA,VAL	possibly-damaging	IGSF8
1	237870348	G	G/T	XM_005273224.1	missense	ARG,LEU	probably-damaging	RYR2
2	179438146	G	A/G	NM_001256850.1	missense	SER,LEU	probably-damaging	TTN
3	13667963	C	C/T	NM_001004019.1	missense	THR,MET	probably-damaging	FBLN2
5	83360657	G	A/G	NM_001278642.1	missense	ARG,CYS	probably-damaging	EDIL3
7	27134277	C	C/T	NM_005522.4	missense	ALA,THR	probably-damaging	HOXA1
7	101925160	C	C/T	NM_001202544.1	missense	ALA,VAL	probably-damaging	CUX1
7	102453899	C	C/G	XM_005250213.1	missense	ASP,HIS	probably-damaging	FBXL13
9	74361146	G	A/G	NM_001135820.1	missense	THR,ILE	probably-damaging	TMEM2
11	7530763	A	A/G	XM_005252878.1	missense	THR,ALA	probably-damaging	OLFML1
11	48285509	T	C/T	NM_001004726.1	missense	TYR,HIS	probably-damaging	OR4X1
12	109717540	G	A/G	XM_005253839.1	missense	SER,LEU	possibly-damaging	FOXN4
12	121206820	C	C/G	NM_139015.4	missense	ARG,PRO	probably-damaging	SPPL3
13	21549115	C	C/T	XM_005266342.1	missense	ARG,GLN	possibly-damaging	LATS2
14	20916116	C	C/T	NM_017807.3	missense	ARG,GLN	probably-damaging	OSGEP
14	23374622	G	A/G	NM_001077351.1	missense	ARG,CYS	probably-damaging	RBM23
14	78285409	G	A/G	NM_001142545.1	missense	ASP,ASN	probably-damaging	ADCK1
15	23686207	T	G/T	XM_005268282.1	missense	LYS,THR	possibly-damaging	GOLGA6L2
15	41988287	G	G/T	NM_001080541.2	missense	SER,ILE	possibly-damaging	MGA
15	49575841	A	A/G	XM_005254284.1	missense	SER,GLY	probably-damaging	GALK2
15	56395801	C	C/T	NM_022841.5	missense	VAL,ILE	probably-damaging	RFX7
15	59373219	C	C/T	XM_005254479.1	missense	PRO,LEU	probably-damaging	RNF111
16	58073898	T	A/T	NM_002428.2	missense	ILE,ASN	probably-damaging	MMP15
17	7188435	G	A/G	NM_001042.2	missense	ARG,GLN	probably-damaging	SLC2A4
17	19451386	C	C/T	NM_018242.2	missense	ALA,VAL	probably-damaging	SLC47A1
17	19559796	G	G/T	NM_001031806.1	missense	ALA,SER	probably-damaging	ALDH3A2
17	25909947	C	C/T	NM_014238.1	missense	HIS,TYR	possibly-damaging	KSR1
18	74091680	G	A/G	NM_014643.3	missense	PRO,LEU	possibly-damaging	ZNF516
19	8008506	A	A/C	NM_006351.3	missense	CYS,GLY	possibly-damaging	TIMM44
19	34832733	G	A/G	NM_014686.3	missense	GLY,SER	possibly-damaging	KIAA0355
19	40424186	G	G/T	XM_005259365.1	missense	LEU,ILE	possibly-damaging	FCGBP
19	40719777	C	C/T	NM_002446.3	missense	ARG,CYS	probably-damaging	MAP3K10
19	46997074	G	A/G	NM_020709.1	missense	ALA,VAL	probably-damaging	PNMAL2
19	49102494	A	A/G	NM_004605.2	missense	ASP,GLY	possibly-damaging	SULT2B1
19	58600105	C	C/G	XM_005259175.1	missense	SER,THR	possibly-damaging	ZSCAN18
20	40714415	C	C/T	NM_007050.5	missense	ASP,ASN	probably-damaging	PTPRT
20	49509492	G	G/T	XM_005260354.1	missense	PRO,THR	possibly-damaging	ADNP
21	15561443	C	C/T	NM_198996.2	missense	ARG,GLN	possibly-damaging	LIPI
21	39671265	C	C/T	XM_005260977.1	missense	ARG,CYS	probably-damaging	KCNJ15
X	39923735	T	C/T	NM_001123383.1	missense	GLN,ARG	possibly-damaging	BCOR
X	55172521	G	G/T	NM_001166703.1	missense	THR,LYS	possibly-damaging	FAM104B

Colour code	Description
	Variants present in PCG02 and PCG02.4 and not present in PCG02.2 and PCG02.3
	Variants present in all 4 sequenced members
	Variants present in PCG02, PCG02.4, PCG02.3 and not in PCG02.2
	Variants present in PCG02, PCG02.4, PCG02.2 and not in PCG02.3

Appendix 10

PCR primers used to validate novel, rare or known coding mutations detected by next generation sequencing in Australian paediatric cataract cases with familial history.

Family	Gene	Mutation position	Forward primer sequence	Reverse primer sequence	Annealing temperatures
CSA94	CRYGS	chr3:186257377-78	TGCCTCTCAAAATTTAATGTGAA	TGCTTTGTCCAAGGACCTAC	57
CSA159	*CRYAA*	chr21:44592307	GGCAGCTTCTCTGGCATG	GAGCCAGCCGAGGCCAATG	65
CSA109	*GJA3*	chr13:20716962	CGAGAACGTCTGCTACGACA	ATGAAGCAGTCCACCGTGTT	57
CRVEEH85	*CRYBB2*	chr22:25627684	CCCCTCGTTCACCCTCCCATCA	CACTGTGTCCAAGGTCACACAGCTAAGC	57
CSA125	*GJA8*	chr1:147380566	GTGCTGCAGATCATCTTCGT	GCTGCTCTACAGGCCTCTTC	57 (35×)
CRVEEH111	*CRYAA*	chr21:44589369	GCTGACTGAGCAGCCTTCTT	GACGGAGCAAGACCAGAGTC	57
CSA131	*MIP*	chr12:56845225	GAAAGCAACATACAAACTAGTGCAA	CCCCTCCACGTAAACTCAGA	57 (35×)
CSA162	*GJA8*	chr1:147380216	CCGCGTTAGCAAAAACAGAT	CAGCCGGAACTTCTTAGTGC	57
CRCH136	*GCNT2*	chr6:10626796	GGCTGAGACTGCACAATCAT	TTACGTAGCCAGGTCCTGAAG	57
CRCH89	*GCNT2*	chr6:10626722	GGCTGAGACTGCACAATCAT	TTACGTAGCCAGGTCCTGAAG	57
CRCH20	*GJA8*	chr1:147380155	TCTGCACAAAGGAAGCACTG	CTTTTGCGCTTCTCCTCCAT	57
CSA95	*GJA3*	chr13:20717372	GAGAAGCTGCCCATCAGC	GCGTGGACACGAAGATGA	57
CSA133	*CRYBB2*	chr22:25627584	AGAAAGCAGAGGCTCAGTGC	CAAAGACCCACAGCAGACAA	60
CRCH139	*CRYGA*	chr2:209027941	GGATGTTCCTTCCAGCTGAC	TGAACACTCATCCTGTGTTGG	57 (35×)
CSA110	*CRYAA*	chr21:44589271	GCTGACTGAGCAGCCTTCTT	GACGGAGCAAGACCAGAGTC	57
CSA108	*MAF*	chr16:79633624	GGGGGTGTGTGTGTGAGC	CTGGAGCTGGTGGCTGTT	57 (35x)

Appendix 11

PCR primers used to validate novel, rare or known coding mutations detected by next generation sequencing in Australian cohort with sporadic paediatric cataract.

Family	Gene	Mutation Position	Forward primer sequence	Reverse primer sequence	Annealing temperature
CSA150	*GJA8*	chr1:147381190	GGGCTACCAAGAGACACTGC	TCATCTGTTGTCAGCTCTGGA	57
CSA175	*GJA8*	chr1:147380639	CAGCAGGCGGGGACTAAC	TGGAGTGGAGGGATTTCTCA	57
CSA172	*CRYGC*	chr2:208993028	TGGTTGGACAAATTCTGGAAG	CCCACCCCATTCACTTCTTA	57
CSA155	*CRYGS*	chr3:186257168-69	CTCAAAATTGAGGTGAAAGGAA	AGCAGCCAACAAGCAGCTA	57
CSA164	*CRYGB*	chr2:209007382	GAGGGGACAAATGTCAGAGC	TGCTTCCCATCATGAAAACAT	57
CSA164	*CRYAA*	chr21:44590718	GGTGACCGAAGCATCTCTGT	GTCCCTCTCCCAGGGTTG	57
CSA154	*BFSP2*	chr3:133119349	TGTAGGAACAGCACCCAGTG	AGTGGGTAGTGCACGTATGG	57
CSA107	*BFSP2*	chr3:133167461	CAGTGACTTTTACCATTCATTGTC	TCTGTACCACACTGGTTCCTG	57
CSA146	*BFSP2*	chr3:133191276	CTCTGCACACCTCCCTCTTC	GAGGAAGTCTGGGGTGATTTC	57
Csa146	*MIP*	chr12:56845224	GAAAGCAACATACAAACTAGTGCAA	CCCCTCCACGTAAACTCAGA	57
JM	*VIM*	chr10:17272686	GAAGTCCCGCTGAAACCTG	GAAGCTCCAACCTGTGGTCT	57
JM	*NHS*	chrX:17743566	CTGCCAGCCCACAGATCTACA	ATTTCAGTGAGGCTCCAAGCAGC	57
CSA147	*NHS*	chrX:17394202	GCTTGGAGGAGACCAGAAAGT	CCTGGAAGAGGCTGCAAG	57
CSA163	*HSF4*	chr16:67201032	TAGCAACAGGCCTCAGCTCT	GGAGAAGGTTTGAGGGAGGA	57
JP	*FYCO1*	chr3:46009379	AGACAGCAGAGTGCCCAACT	TTCTGCTCCTCCAGGTCACT	57

Appendix 12

List of primers used for validating variants detected in Asian paediatric cataract cohort.

Family	Gene	Mutation Position	Forward primer sequence	Reverse primer sequence	Annealing temperature
P1	*GJA8*	chr1:147380921	TGGCCTCTGTGTCCCTATTC	CTCGCCCTCCACTTCTTGT	57
BB 16 cat	*GJA8*	chr1:147380102	TCTGCACAAAGGAAGCACTG	CTTTTGCGCTTCTCCTCCAT	57
PP-50 cat	*MIP*	chr12:56848301	GACTGTCCACCCAGACAAGG	AGCAGGTGGGGTAACGCTAT	57
PP-50 cat	*COL4A1*	chr13:110864795	GACCTCCGGGAGCATCTG	AAGGAACAATTATATCTTTCTGGTACA	57
PP-50 cat	*NSDHL*	chrX:152037470	TGGGGGTGGTGTTTCTAACT	GAGTCACAGGCTCAGAGCAA	57
PCC 02-105	*CRYGD*	chr2:208986444	GCTGGACTGCCTAACAATGC	CACATCTTGGTTGCCATTTG	57
PCC 02-105	*CRYGD*	chr2:208986623	GCTGGACTGCCTAACAATGC	CACATCTTGGTTGCCATTTG	57
PCC 02-105	*VIM*	chr10:17275680	TCCAGGGTCATAAAATGTGTCA	GTGAGGTCAGGCTTGGAAAC	57
E1	*GALE*	chr1:24123216	CTGCTCATTGACTGCAGGTT	GGCCTACCTTCTTCCCAGA	57
PCC 10-183	*GCNT2*	chr6:10626784	GCACAGTTGTAGTTAGTCGGAGA	GTGCCCAGATTGCTCTACC	57
SR 11 cat	*PAX6*	chr11:31823289	CAGGGAGGGCAGATGTTCT	TAGACCCATGCAGATGCAAA	57
PCC 10-188	*EPHA2*	chr1:16464671	TCCTGCGAGTGTGAGGAAG	CAGGTCGCTCACTGTCACAC	57
PCC 10-188	*GCNT2*	chr6:10626784	GGCTGAGACTGCACAATCAT	AGGTGAGCATCAAGAAACAGC	57
PCC 10-188	*NHS*	chrX:17743727	CTGCCAGCCCACAGATCTACA	ATTTCAGTGAGGCTCCAAGCAGC	57
W1	*GALK1*	chr17:73758836	CTGCTCATTGACTGCAGGTT	GCAGTGTTTGAAGGGACTGG	57
W1	*GALK1*	chr17:73759221	AAGTGGCCACGTACACCTTC	GACCTGGGGTGGAGTTACAA	57
PCC01-97A	*AGK*	chr7:141255292	TGTGGGTGGATGAGAAGATG	TTTTAACCTGAGGCCCTTTG	57 (35X)
PCC01-97A	*TDRD7*	chr9:100234592	TGGTCAAAACAGACACAATCC	CAAGCAGGTGGCATTGAT	57
PCC01-97A	*PAX6*	chr11:31815036	CTCTCAAGGGTGCAGACACA	GCCAGCAACACACCTAGTCA	57
PCC01-97A	*BFSP1*	chr20:17479645	GCTCATTCTTCAGGGTTTGC	ACATTCACCGAGTCCAGGTT	57 (35X)
PCC01-97A	*CRYBB1*	chr22:27008146	CCTGCACTGCTGGCTTTTAT	ATGTGCCAGGAGTACGAACG	57
PCC 01-34	*HSF4*	chr16:67201678	CCAGATGGCTGTAGGGGTAG	TTCAGGCTGTTGGGCATT	57 (35X)
SR 12 cat	*TDRD7*	chr9:100245251	GGATTTTAGCAAGGGTTTTGG	GCCACAGGCACATACACATC	57
SR 12 cat	*COL4A1*	chr13:110857844	TTAGTGGAGACGGGATTTCG	TTGTGTTTTTACCCAGAAGAAGC	57

Appendix 13

Primer sequences for the three microsatellite markers used for haplotype analysis.

Marker	Fluorescent Label	Forward primer 5' to 3'	Reverse primer 5' to 3	PCR product size (bp)	Position on chr21 (cM)
D21S1260	FAM	TCCAAGGGGTTCATCC	CCCAAGGCACTGTTCC	200-214	42.7
D21S1890	FAM	GGTCTGACCACAGATTTCC	AAAAACACTCTGAACGATTAAGG	143-173	44.8
D21S1912	HEX	CCCTCATACAGATTTAAAACACAC	GAGCCCACCCTGGTAAC	173-205	45.5

CRYAA is located between D21S1260 and D21S189.

Appendix 14

Protein sequence alignments demonstrating the conservation of the altered amino acid (boxed) in Australian families with causative mutations. Families with protein alterations in A: Gap junction (GJA3 and GJA8); B: Crystallin (CRYAA, CRYBB2, CRYGA and CRYGS) and C: GCNT2, MIP and MAF. The alignments are generated using Mutation Taster (http://www.mutationtaster.org/) and show alignments of 9 protein sequences to the human protein sequence. Both mutated and wild type human protein sequences are given beside sequences from other species. Empty rows mean that there is no homologue for a gene in that particular species. Family references for each mutation are shown in brackets under the alignments.

A

species	alignment
Human	G R L L E N A Q E H S T V I G K V W L T V L F
Human-mutated	G R L L E N A Q E H S M V I G K V W L T V L F
Ptroglodytes (chimpanzee)	G R L L E N A Q E H S T V I G K V W L T V L F
Mmulatta (Monkey)	G R L L E N A Q E H S T V I G K V W L T V L F
Fcatus	
Mmusculus (mouse)	G R L L E N A Q E H S T V I G K V W L T V L F
Ggallus (rooster)	G R L L E N A Q E H S T V I G K V W L T V L F
Trubripes (fish)	G R L L E N A Q E H S T V I G K V W L T V L F
Drerio (zebrafish)	G R L L E N A Q E H S T V I G K V W L T V L F
Xtropicalis (Frog)	G R L L E N A Q E H S T V I G K V W L T V L F

GJA3, p.(Thr19Met) (CSA95)

species	alignment
Human	L L R T Y V F N I I F K T L F E V G F I A G Q
Human-mutated	L L R T Y V F N I I F Q T L F E V G F I A G Q
Ptroglodytes (chimpanzee)	L L R T Y V F N I I F K T L F E V G F I A G Q
Mmulatta (Monkey)	L L R T Y V F N I I F K T L L E V G F I A G Q
Fcatus	
Mmusculus (mouse)	L L R T Y V F N I I F K T L F E V G F I A G Q
Ggallus (rooster)	L L R T Y I F N I I F K T L
Trubripes (fish)	L L R T Y V F N I I F K T L F E V G F I L G Q
Drerio (zebrafish)	L L R T Y V F N I I F K T L F E V G F
Xtropicalis (Frog)	L L R T Y V F N I I F K T L F E I G F I V G Q

GJA3, p.(Lys156Gln) (CSA109)

species	alignment
Human	I L I L G T A A E F V W G D E Q S D F V C N T
Human-mutated	I L I L G T A A E F V S G D E Q S D F V C N T
Ptroglodytes (chimpanzee)	I L I L G T A A E F V W G D E Q S D F V C N T
Mmulatta (Monkey)	I L I L G T A A E F V W G D E Q S D F V C N T
Fcatus	
Mmusculus (mouse)	I L I L G T A A E F V W G D E Q S D F V C N T
Ggallus (rooster)	I L I L G T A A E L V W G D E Q S D F V C N T
Trubripes (fish)	
Drerio (zebrafish)	I L I L G T A A E F V W G D E Q S D Y V C N T
Xtropicalis (Frog)	I L I L G T A A E F V W G D E Q S D F V C N T

GJA8, p.(Trp45Ser) (CSA162)

species	alignment
Human	V N E H S T V I G R V W L T V L F I F R I L I
Human-mutated	V N E H S T V I G R V R L T V L F I F R I L I
Ptroglodytes (chimpanzee)	V N E H S T V I G R V W L T V L F I F R I L I
Mmulatta (Monkey)	V N E H S T V I G R V W L T V L F I F R I L I
Fcatus	
Mmusculus (mouse)	V N E H S T V I G R V W L T V L F I F R I L I
Ggallus (rooster)	V N E Q S T V I G R V W L T V L F I F R I L I
Trubripes (fish)	
Drerio (zebrafish)	V N E H S T V I G R V W L T V L F I F R I L I
Xtropicalis (Frog)	V N E H S T V I G R V W L T V L F I F R I L I

GJA8, p.(Trp25Arg) (CRCH20)

species	alignment
Human	Y I C H I I F K T L F E V G F I V G H Y F L Y
Human-mutated	Y I C H I I F K T L F K V G F I V G H Y F L Y
Ptroglodytes (chimpanzee)	Y I C H I I F K T L F E V G F I V G H Y F L Y
Mmulatta (Monkey)	Y I C H I I F K T L F E V G F I V G H Y F L Y
Fcatus	
Mmusculus (mouse)	Y V C H I I F K T L F E V G F I V G H Y F L Y
Ggallus (rooster)	Y I L H I I F K T L F E V G F I V G Q Y F L Y
Trubripes (fish)	
Drerio (zebrafish)	Y I C H I I F K T L F E I G F V V G Q Y Y L Y
Xtropicalis (Frog)	Y I C H I I F K T L F E V G F V V G Q Y F L Y

GJA8, p.(Glu162Lys) (CSA125)

B

species	alignment
Human	T I S P Y Y R Q S L F R T V L D S G
Human-mutated	T I S P Y Y R Q S L F C T V L D S G
Ptroglodytes (chimpanzee)	T I S P Y Y R Q S L F R T V L D S G
Mmulatta (Monkey)	
Fcatus	T I S P Y Y R Q S L F R T V L D S G
Mmusculus (mouse)	T I S P Y Y R Q S L F R T V L D S G
Ggallus (rooster)	T I S P Y Y R Q S L F R S V L E S G
Trubripes (fish)	T I S P Y Y R Q S L F R S F L D S
Drerio (zebrafish)	T V S P Y Y R H S L F R N I L D S
Xtropicalis (Frog)	T L G V I Y K E S M C R C L M

CRYAA, p.(Arg54Cys) (CRVEEH111)

species	Alignment
Human	D G M L T F C G P K I Q T G L D A T H A E R A I P V S R E E K P T S
Human-mutated	D G M L T F C G P K I R L A W M P P T P S E P S P C R G R R S P P R
Ptroglodytes (chimpanzee)	D G M L T F C G P K I Q T G L D A T H A E R A I P V S R E E K P T S
Mmulatta (Monkey)	
Fcatus	D G M L T F S G P K V P S G V D A G H S E R A I P V S R E E K P S S
Mmusculus (mouse)	D G M L T F S G P K V Q S G L D A G H S E R A I P V S R E E K P S S
Ggallus (rooster)	D G M L T F S G P K V P S N M D P S H S E R P I P V S R E E K P T S
Trubripes (fish)	D G L L T L S G A K V Q G G S E S G R S E R S I P V T R D D K P N A
Drerio (zebrafish)	D G L L T L C G P K - T S G I D A G R G D R T I P V T R E D K S N S
Xtropicalis (Frog)	D G I L S F S G P K L Q P N V D S S H S D R T I P V S R E E K S G S

CRYAA, p.(Gln147 Argfs*48) (CSA159)

species	alignment
Human	F K R T L G P F Y P S R L F D Q F F G E G L F
Human-mutated	F K R T L G P F Y P S Q L F D Q F F G E G L F
Ptroglodytes (chimpanzee)	F K R T L G P F Y P S R L F D Q F F G E G L F
Mmulatta (Monkey)	
Fcatus	F K R A L G P F Y P S R L F D Q F F G E G L F
Mmusculus (mouse)	F K R A L G P F Y P S R L F D Q F F G E G L F
Ggallus (rooster)	F K R A L G P L I P S R L F D Q F F G E G L L
Trubripes (fish)	F R R A L G S A Y P T R L F D Q F F G E G M F
Drerio (zebrafish)	F R R T L G - - Y P T R L F D Q F F G E G L F
Xtropicalis (Frog)	F K R S L G P L Y P N R L F D Q V F G E G M F

CRYAA, p.(Arg21Gln) (CSA110)

species	alignment
Human	V R V Q S G T W W G Y Q Y P G Y R G L Q Y L L E K G D Y K D S S D F G A P H P Q V Q S V R R I R D M Q W H Q R G A F H P S N*
Human-mutated	
Ptroglodytes (chimpanzee)	V R V Q S G T W W G Y Q Y P G Y R G L Q Y L L E K G D Y K D S S D F G A P H P Q V Q S V R R I R D M Q W H Q R G A F H P S N
Mmulatta (Monkey)	V R V Q S G T W W G Y Q Y P G Y R G L Q Y L L E K G D Y K E S S D F G A P H P Q V Q S V R R I R D M Q W H Q R G A F H P S N
Fcatus	V R V Q S G T W W G Y Q Y P G Y R G L Q Y L L E K G D Y K D S G D F G A P H P Q V Q S V R R I R D M Q W H Q R G A F H P S N
Mmusculus (mouse)	V R V Q S G T W W G Y Q Y P G Y R G L Q Y L L E K G D Y K D N S D F G A P H P Q V Q S V R R I R D M Q W H Q R G A F H P S S
Ggallus (rooster)	V R V Q S G T W W G Y Q Y P G Y R G Y Q Y L F E K G D Y K D S S E F G A Q H P Q I Q S V R R I R D M Q W H Q R G A Y H P S N
Trubripes (fish)	V K A L N G T W W G Y L Y P G Y R G R Q F I F E Q G D F K H W N D W E A P D P Q I Q S L R R V R D M Q W H K K G C F
Drerio (zebrafish)	V R V Q S G T W W G Y Q Y P G Y R G F Q Y L F E K G E F K E C S E F G A A L P Q I Q S V R R I R D M Q Y H P R G A F Q A T S
Xtropicalis (Frog)	V R V L S G T W W G Y Q Y P G Y R G Y Q Y L F E K G D Y K D S S D F G A K H P H I Q S V R R I R D M Q W H Q R G T F H P T N

CRYBB2, p.(Gln155*) (CSA133)

species	alignment
Human	F G A P H P Q V Q S V R R I R D M Q W H Q
Human-mutated	H P Q V Q S V L R I R D M Q W H Q
Ptroglodytes (chimpanzee)	H P Q V Q S V R R I R D M Q W H Q
Mmulatta (Monkey)	H P Q V Q S V R R I R D M Q W H Q
Fcatus	H P Q V Q S V R R I R D M Q W H Q
Mmusculus (mouse)	H P Q V Q S V R R I R D M Q W H Q
Ggallus (rooster)	Q I Q S V R R I R D M Q W H Q
Trubripes (fish)	A P D P Q I Q S L R R V R D M Q W H K
Drerio (zebrafish)	G A A L P Q I Q S V R R I R D M Q Y H P
Xtropicalis (Frog)	K H P H I Q S V R R I R D M Q W H Q

CRYBB2, p.(Arg188Leu) (CRVEEH85)

species	alignment
Human	W M G L S D S V Q S C R I I P H T S S H K L R
Human-mutated	W M G L S D S V Q S C H I I P H T S S H K L R
Ptroglodytes (chimpanzee)	W M G L S D S V Q S C R I I P H T S S H K L R
Mmulatta (Monkey)	W M G L S D S V Q S C R I I P H T R S H K L R
Fcatus	W M G L S D S V R S C R A I P Y T S S H R I R
Mmusculus (mouse)	W M G F S D S I R S C R S I P Y T S S H R I R
Ggallus (rooster)	
Trubripes (fish)	
Drerio (zebrafish)	W M G Y N D T I R S C R M V R N H T G S F R I R
Xtropicalis (Frog)	W M G F N D C I K S C R A I P Q Q R G A Y K I K

CRYGA, p.(Arg80His) (CRCH139)

species	alignment
Human	M S K T G T K I T F Y E D K N F Q G R R Y
Human-mutated	M S K T G T K I T L Y E D K N F Q G R R Y
Ptroglodytes (chimpanzee)	M S K T G P K I T F Y E D K N F Q G R R Y
Mmulatta (Monkey)	
Fcatus	I T F Y E D K N F Q G R H Y
Mmusculus (mouse)	M S K T G G K I S F Y E D R N F Q G R R Y
Ggallus (rooster)	M S R A G P K V T F Y E D K N F L G R R Y
Trubripes (fish)	
Drerio (zebrafish)	M G R V S T K I I F F E D K N F Q G R R H
Xtropicalis	

CRYGS, p.(Phe10_Tyr11delinsLeuAsn) (CSA94)

C

species	Alignment
Human	N H WV Y WV G P I I G G G L G S L L Y D F L L F P R L K S I S E R L S V L K G A K P D V S N G Q P E V T G E P V E L N T Q A L
Human-mutated	
Ptroglodytes (chimpanzee)	N H WV Y WV G P I I G G G L G S L L Y D F L L F P R L K S I S E R L S V L K G A K P D V S N G Q P E V T G E P V E L N T Q A L
Mmulatta (Monkey)	N H WV Y WV G P I I G G G L G S L L Y D F L L F P R L K S I S E R L S V L R G V K P E D A N G Q P E V T G E P V E L N T Q A L
Fcatus	
Mmusculus (mouse)	N H WV Y WV G P I I G G G L G S L L Y D F L L F P R L K S V S E R L S I L K G A R P S D S N G Q P E G T G E P V E L K T Q A L
Ggallus (rooster)	
Trubripes (fish)	N H WV Y WV G P MI G G A MG A L L Y D F ML F P R MR G L S E R L A T L K G T R P P E A E G Q Q E T R G E P I E L K T Q A L
Drerio (zebrafish)	N H WV Y WV G P MI G A A MG A L L Y D F ML F P R V R G L S E R L A V L K G N K P T E P E A Q Q E T R G E P I E L K T Q A L
Xtropicalis (Frog)	N H WV Y WV G P I I G G A V G G L V Y D F I L F P R MR G L N E R L S I L K G A R P A E P E G Q R E T I R D P I E L K T Q S L

MIP, p.(Gly211*) (CSA131)

species	Alignment
Human	C L E L R H R E R T L N Q S E T A I Q P S W Y F *
Human-mutated	C L E L R H R E R T L R V K L R Y N P A G I F E I
Ptroglodytes (chimpanzee)	C L E L R H R E R T L N Q S E T A I Q P S W Y F
Mmulatta (Monkey)	
Fcatus	
Mmusculus (mouse)	C L E L R L R E R T L N Q S E I A I Q P S W Y F
Ggallus (rooster)	C L E L R H R K R T L A Q S E V Q V E P N W Y F
Trubripes (fish)	
Drerio (zebrafish)	
Xtropicalis (Frog)	C L E L K V R E R T L N Q S E V T V P P E W Y

GCNT2, p.(Asn388Arg*20) (CRCH136)

species	alignment
Human	D L K W L V N S P S L F A N K F E L N T Y
Human-mutated	P S L S A N K F E L N T Y
Ptroglodytes (chimpanzee)	P S L F A N K F E L N T Y
Mmulatta (Monkey)	
Fcatus	
Mmusculus (mouse)	D L Q W L I N S Q S L F A N K F
Ggallus (rooster)	D L K W L F N S T C M F A N K F E L
Trubripes (fish)	
Drerio (zebrafish)	
Xtropicalis (Frog)	S R S I F A N K F E A K S Y

GCNT2, p.(Phe364S) (CRCH89)

species	alignment
Human	R L I A G G S L S S T P M S T P C S S V P P S F
mutated	R L I A G G S L S S T R M S T P C S S V P P S
Ptroglodytes (chimpanzee)	R L I A G G S L S S T P M S T P C S S V P P S
Mmulatta (Monkey)	
Fcatus	
Mmusculus (Mouse)	R L I A G G S L S S T P M S T P C S S V P P S
Ggallus (rooster)	
Trubripes	
Dmelanogaster	
Xtropicalis	R L I A G G S L S S T P M S T P C S S V P P S

MAF, p.(Pro59Arg) (CSA108)

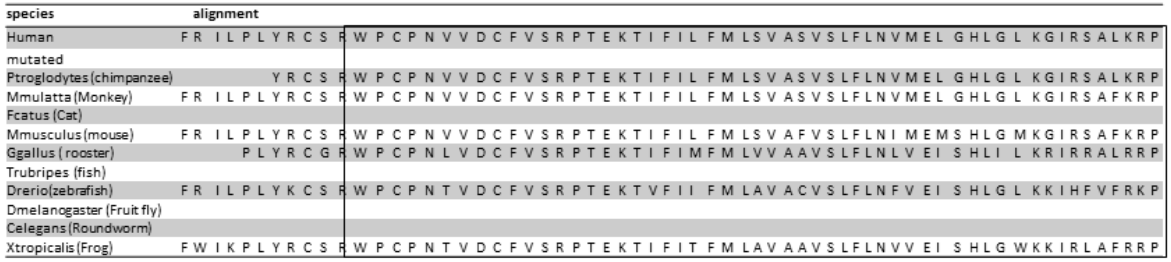
Appendix 15

Protein sequence alignments demonstrating the conservation of the altered amino acid (boxed) in sporadic paediatric cases (Australian cohort). Families with protein alterations in A: Gap junction (GJA8); B: Crystallin (CRYAA, CRYGC and CRYGB) and C: BFSP2, MIP, VIM, NHS, MAF and HSF4. The alignments are generated using Mutation Taster (http://www.mutationtaster.org/) and show alignments of 11 protein sequences to the human protein sequence. Both mutated and wild type human protein sequences are given beside sequences from other species. Empty rows mean that there is no homologue for a gene in that particular species. Family references for each mutation are shown in brackets under the alignments

A

species	alignment
Human	EEQEKVAVPEGEKVETPGVDKEGEKEPQSEKVKQGLPAEKTPLCPELTTDDARPLSRLSKASSRARSDDLT
mutated	
Ptroglodytes (chimpanzee)	QEKVAVPEGEKVETPRVDKEGEKELQSEKVKQGLPAEKTPLCPELTTDDARPLSRLSKASSRARSDDLT
Mmulatta (Monkey)	EEQEKVAVPEGEEVETPGVGKEGEKELQSEKVKQGLPAEKTPLCQELTTDEARPLSRLSKASSRARSDDLT
Fcatus (Cat)	
Mmusculus (mouse)	EGQETVAVPDRERVETPGVGKEDEKELQAEKVKQGLSAEKAPLCPELTTDDNRPLSRLSKASSRARSDDLT
Ggallus (rooster)	EE-------------PGVKK--AEEVVSDEVGPSAPAE--------LATDVRSLSRLSKASSRARSDDLT
Trubripes (fish)	
Drerio(zebrafish)	E----------ADMEASE----------TIEDTRPLSSLSKASSRARSDDLT
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	RAPSDRAPSSERAPTRAASEHAGSETPSVRTSNSATPERVPRAASEHAPYSRSLSRLSKASSRARSDDLT

GJA8, p.(Glu370*), CSA150



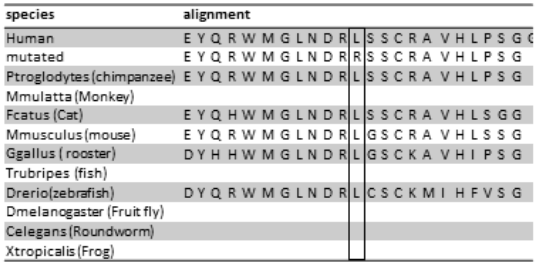
species	alignment
Human	FRILPLYRCSRWPCPNVVDCFVSRPTEKTIFILFMLSVASVSLFLNVMELGHLGLKGIRSALKRP
mutated	
Ptroglodytes (chimpanzee)	YRCSRWPCPNVVDCFVSRPTEKTIFILFMLSVASVSLFLNVMELGHLGLKGIRSALKRP
Mmulatta (Monkey)	FRILPLYRCSRWPCPNVVDCFVSRPTEKTIFILFMLSVASVSLFLNVMELGHLGLKGIRSAFKRP
Fcatus (Cat)	
Mmusculus (mouse)	FRILPLYRCSRWPCPNVVDCFVSRPTEKTIFILFMLSVAFVSLFLNIMEMSHLGMKGIRSAFKRP
Ggallus (rooster)	PLYRCGRWPCPNLVDCFVSRPTEKTIFIMFMLVVAAVSLFLNLVEISHLILKRIRRALRRP
Trubripes (fish)	
Drerio(zebrafish)	FRILPLYKCSRWPCPNTVDCFVSRPTEKTVFIIFMLAVACVSLFLNFVEISHLGLKKIHFVFRKP
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	FWIKPLYRCSRWPCPNTVDCFVSRPTEKTIFITFMLAVAAVSLFLNVVEISHLGWKKIRLAFRRP

GJA8, p.(Trp186*), CSA175

B

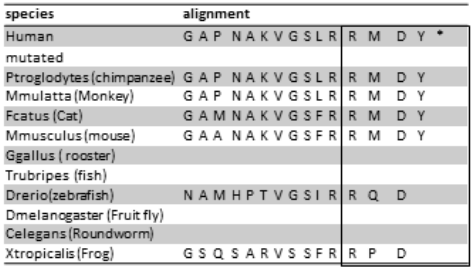
species	alignment
Human	V L Y E L P N Y R G R Q Y L L R P Q E Y R R C Q D W G A M D A K A G S L R R V V D L Y *
mutated	M L Y E R P N Y Q G Q Q Y L L R R G E Y P D Y Q Q W
Ptroglodytes (chimpanzee)	V L Y E L P N Y R G R Q Y L L R P Q E Y R R C Q D W G A M D A K A G S L R R V V D L Y
Mmulatta (Monkey)	V L Y E L P N Y R G R Q Y L L R P Q E Y R R C Q D W G A M D A K A G S L R R V V D L Y
Fcatus (Cat)	
Mmusculus (mouse)	V L Y E M P N Y R G R Q Y L L R P Q E Y R R F Q D W G S V D A K A G S L R R V V D L Y
Ggallus (rooster)	
Trubripes (fish)	
Drerio(zebrafish)	I F F E H P N Y R G R Q Y L L E K G E Y R C F T D W N A M H P T V G S I R R I Q D
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	I F Y E Q P N Y R G R Q Y Y L K P G E Y K R F S D W G S Q S A R V S S F R R V P D V

CRYGC, p.(Arg142Alafs22), CSA172



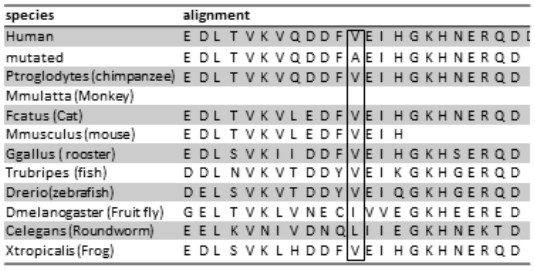
species	alignment
Human	E Y Q R W M G L N D R L S S C R A V H L P S G C
mutated	E Y Q R W M G L N D R R S S C R A V H L P S G
Ptroglodytes (chimpanzee)	E Y Q R W M G L N D R L S S C R A V H L P S G
Mmulatta (Monkey)	
Fcatus (Cat)	E Y Q H W M G L N D R L S S C R A V H L S G G
Mmusculus (mouse)	E Y Q R W M G L N D R L G S C R A V H L S S G
Ggallus (rooster)	D Y H H W M G L N D R L G S C K A V H I P S G
Trubripes (fish)	
Drerio(zebrafish)	D Y Q R W M G L N D R L C S C K M I H F V S G
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	

CRYGC, p.(Leu80Arg), CSA155



species	alignment
Human	G A P N A K V G S L R R M D Y *
mutated	
Ptroglodytes (chimpanzee)	G A P N A K V G S L R R M D Y
Mmulatta (Monkey)	G A P N A K V G S L R R M D Y
Fcatus (Cat)	G A M N A K V G S F R R M D Y
Mmusculus (mouse)	G A A N A K V G S F R R M D Y
Ggallus (rooster)	
Trubripes (fish)	
Drerio(zebrafish)	N A M H P T V G S I R R Q D
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	G S Q S A R V S S F R R P D

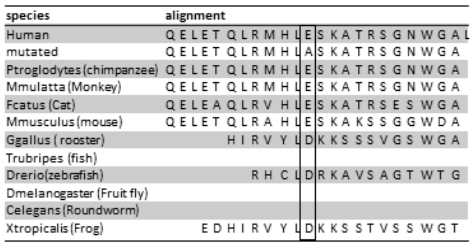
CRYGB, p.(Arg170*) , CSA164



species	alignment
Human	E D L T V K V Q D D F V E I H G K H N E R Q D
mutated	E D L T V K V Q D D F A E I H G K H N E R Q D
Ptroglodytes (chimpanzee)	E D L T V K V Q D D F V E I H G K H N E R Q D
Mmulatta (Monkey)	
Fcatus (Cat)	E D L T V K V L E D F V E I H G K H N E R Q D
Mmusculus (mouse)	E D L T V K V L E D F V E I H
Ggallus (rooster)	E D L S V K I I D D F V E I H G K H S E R Q D
Trubripes (fish)	D D L N V K V T D D Y V E I K G K H G E R Q D
Drerio(zebrafish)	D E L S V K V T D D Y V E I Q G K H G E R Q D
Dmelanogaster (Fruit fly)	G E L T V K L V N E C I V V E G K H E E R E D
Celegans (Roundworm)	E E L K V N I V D N Q L I I E G K H N E K T D
Xtropicalis (Frog)	E D L S V K L H D D F V E I H G K H N E R Q D

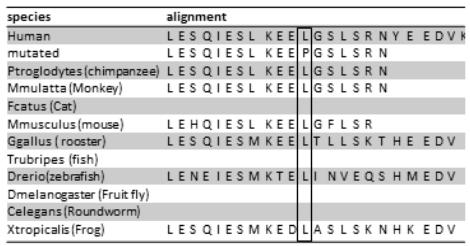
CRYAA, p.(Val94Gly) , CSA164

C



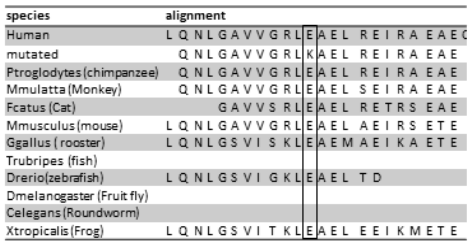
species	alignment
Human	Q E L E T Q L R M H L E S K A T R S G N W G A L
mutated	Q E L E T Q L R M H L A S K A T R S G N W G A
Ptroglodytes (chimpanzee)	Q E L E T Q L R M H L E S K A T R S G N W G A
Mmulatta (Monkey)	Q E L E T Q L R M H L E S K A T R S G N W G A
Fcatus (Cat)	Q E L E A Q L R V H L E S K A T R S E S W G A
Mmusculus (mouse)	Q E L E T Q L R A H L E S K A K S S G G W D A
Ggallus (rooster)	H I R V Y L D K K S S S V G S W G A
Trubripes (fish)	
Drerio(zebrafish)	R H C L D R K A V S A G T W T G
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	E D H I R V Y L D K K S S T V S S W G T

BFSP2, p.(Glu141Ala) , CSA154



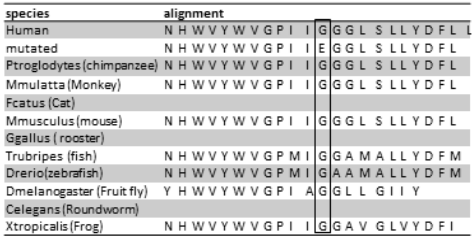
species	alignment
Human	L E S Q I E S L K E E L G S L S R N Y E E D V K
mutated	L E S Q I E S L K E E P G S L S R N
Ptroglodytes (chimpanzee)	L E S Q I E S L K E E L G S L S R N
Mmulatta (Monkey)	L E S Q I E S L K E E L G S L S R N
Fcatus (Cat)	
Mmusculus (mouse)	L E H Q I E S L K E E L G F L S R
Ggallus (rooster)	L E S Q I E S M K E E L T L L S K T H E E D V
Trubripes (fish)	
Drerio(zebrafish)	L E N E I E S M K T E L I N V E Q S H M E D V
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	L E S Q I E S M K E D L A S L S K N H K E D V

BFSP2, p.(Leu234Pro), CSA107



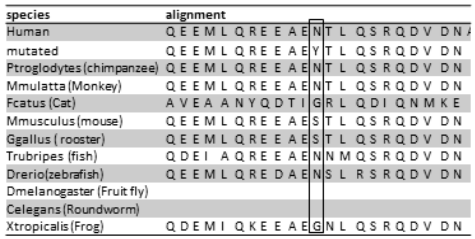
species	alignment
Human	L Q N L G A V V G R L E A E L R E I R A E A E C
mutated	Q N L G A V V G R L K A E L R E I R A E A E
Ptroglodytes (chimpanzee)	Q N L G A V V G R L E A E L R E I R A E A E
Mmulatta (Monkey)	Q N L G A V V G R L E A E L S E I R A E A E
Fcatus (Cat)	G A V V S R L E A E L R E T R S E A E
Mmusculus (mouse)	L Q N L G A V V G R L E A E L A E I R S E T E
Ggallus (rooster)	L Q N L G S V I S K L E A E M A E I K A E T E
Trubripes (fish)	
Drerio(zebrafish)	L Q N L G S V I G K L E A E L T D
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	L Q N L G S V I T K L E A E L E E I K M E T E

BFSP2, p.(Glu371Lys), CSA146



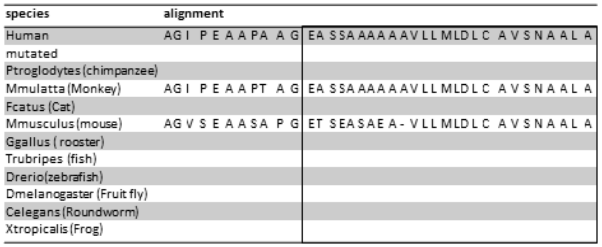
species	alignment
Human	N H W V Y W V G P I I G G G L S L L Y D F L L
mutated	N H W V Y W V G P I I E G G L S L L Y D F L
Ptroglodytes (chimpanzee)	N H W V Y W V G P I I G G G L S L L Y D F L
Mmulatta (Monkey)	N H W V Y W V G P I I G G G L S L L Y D F L
Fcatus (Cat)	
Mmusculus (mouse)	N H W V Y W V G P I I G G G L S L L Y D F L
Ggallus (rooster)	
Trubripes (fish)	N H W V Y W V G P M I G G A M A L L Y D F M
Drerio(zebrafish)	N H W V Y W V G P M I G A A M A L L Y D F M
Dmelanogaster (Fruit fly)	Y H W V Y W V G P I A G G L L G I I Y
Celegans (Roundworm)	
Xtropicalis (Frog)	N H W V Y W V G P I I G G A V G L V Y D F I

MIP, p.(Gly211Glu), CSA146



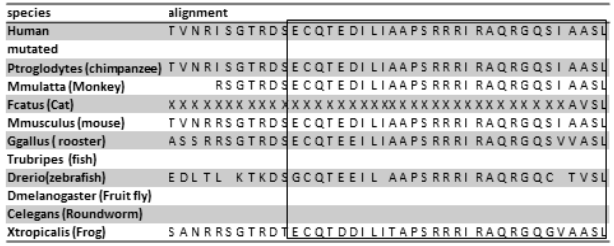
species	alignment
Human	Q E E M L Q R E E A E N T L Q S R Q D V D N A
mutated	Q E E M L Q R E E A E Y T L Q S R Q D V D N
Ptroglodytes (chimpanzee)	Q E E M L Q R E E A E N T L Q S R Q D V D N
Mmulatta (Monkey)	Q E E M L Q R E E A E N T L Q S R Q D V D N
Fcatus (Cat)	A V E A A N Y Q D T I G R L Q D I Q N M K E
Mmusculus (mouse)	Q E E M L Q R E E A E S T L Q S R Q D V D N
Ggallus (rooster)	Q E E M L Q R E E A E S T L Q S R Q D V D N
Trubripes (fish)	Q D E I A Q R E E A E N N M Q S R Q D V D N
Drerio(zebrafish)	Q E E M L Q R E D A E N S L R S R Q D V D N
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	Q D E M I Q K E E A E G N L Q S R Q D V D N

VIM, p.(Asn201Tyrp), JM



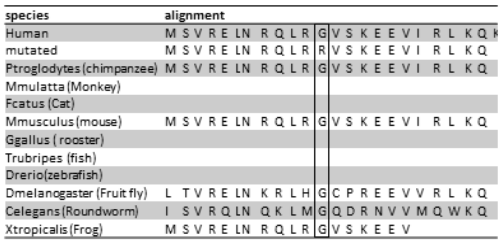
species	alignment
Human	A G I P E A A P A A G E A S S A A A A A A V L L M L D L C A V S N A A L A
mutated	
Ptroglodytes (chimpanzee)	
Mmulatta (Monkey)	A G I P E A A P T A G E A S S A A A A A A V L L M L D L C A V S N A A L A
Fcatus (Cat)	
Mmusculus (mouse)	A G V S E A A S A P G E T S E A S A E A - V L L M L D L C A V S N A A L A
Ggallus (rooster)	
Trubripes (fish)	
Drerio(zebrafish)	
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	

NHS, p.(Glu108*) , CSA147



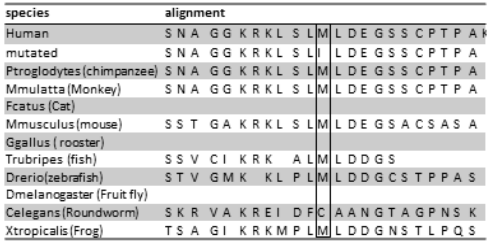
species	alignment
Human	T V N R I S G T R D S E C Q T E D I L I A A P S R R R I R A Q R G Q S I A A S L
mutated	
Ptroglodytes (chimpanzee)	T V N R I S G T R D S E C Q T E D I L I A A P S R R R I R A Q R G Q S I A A S L
Mmulatta (Monkey)	R S G T R D S E C Q T E D I L I A A P S R R R I R A Q R G Q S I A A S L
Fcatus (Cat)	X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X A V S L
Mmusculus (mouse)	T V N R R S G T R D S E C Q T E D I L I A A P S R R R I R A Q R G Q S I A A S L
Ggallus (rooster)	A S S R R S G T R D S E C Q T E E I L I A A P S R R R I R A Q R G Q S V V A S L
Trubripes (fish)	
Drerio(zebrafish)	E D L T L K T K D S G C Q T E E I L A A P S R R R I R A Q R G Q C T V S L
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	S A N R R S G T R D T E C Q T D D I L I T A P S R R R I R A Q R G Q G V A A S L

NHS, p.(Glu427Val*9), JM



species	alignment
Human	M S V R E L N R Q L R G V S K E E V I R L K Q K
mutated	M S V R E L N R Q L R R V S K E E V I R L K Q
Ptroglodytes (chimpanzee)	M S V R E L N R Q L R G V S K E E V I R L K Q
Mmulatta (Monkey)	
Fcatus (Cat)	
Mmusculus (mouse)	M S V R E L N R Q L R G V S K E E V I R L K Q
Ggallus (rooster)	
Trubripes (fish)	
Drerio(zebrafish)	
Dmelanogaster (Fruit fly)	L T V R E L N K R L H G C P R E E V V R L K Q
Celegans (Roundworm)	I S V R Q L N Q K L M G Q D R N V V M Q W K Q
Xtropicalis (Frog)	M S V R E L N R Q L R G V S K E E V

MAF, p.(Gly280Arg), CSA117



species	alignment
Human	S N A G G K R K L S L M L D E G S S C P T P A K
mutated	S N A G G K R K L S L I L D E G S S C P T P A
Ptroglodytes (chimpanzee)	S N A G G K R K L S L M L D E G S S C P T P A
Mmulatta (Monkey)	S N A G G K R K L S L M L D E G S S C P T P A
Fcatus (Cat)	
Mmusculus (mouse)	S S T G A K R K L S L M L D E G S A C S A S A
Ggallus (rooster)	
Trubripes (fish)	S S V C I K R K A L M L D D G S
Drerio(zebrafish)	S T V G M K K L P L M L D D G C S T P P A S
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	S K R V A K R E I D F C A A N G T A G P N S K
Xtropicalis (Frog)	T S A G I K R K M P L M L D D G N S T L P Q S

HSF4, p.Met212Ile, CSA163

Appendix 16

Protein sequence alignments demonstrating the conservation of the altered amino acid (boxed) in South East Asian paediatric cataract cohort. Families with protein alterations in A: structural proteins (GJA8, MIP, VIM and COL4A1); B: Crystallin (CRYGD and CRYBB1); C: transcription factor and signalling molecules (HSF4, PAX6 and TDRD7) and D in protein associate with enzymatic and syndromic paediatric cataracts (GALE, GALK1, AGK, GCNT2, NHS and NSDHL). The alignments are generated using Mutation Taster (http://www.mutationtaster.org/) and show alignments of 11 protein sequences to the human protein sequence. Both mutated and wild type human protein sequences are given beside sequences from other species. Empty rows mean that there is no homologue for a gene in that particular species. Family references for each mutation are shown in brackets under the alignments.

A

species	alignment
Human	E E E K I V S H Y F P L T E V G M V E T S P L
mutated	E E E K I V S H Y F R L T E V G M V E T S P
Ptroglodytes (chimpanzee)	E E E K I V S H Y F P L T E V G M V E T S P
Mmulatta (Monkey)	E E E K I V S H Y F P L T E V G M V E T - -
Fcatus (Cat)	
Mmusculus (mouse)	E E E K I V S H Y F P L T E V G M V E T S P
Ggallus (rooster)	E E E K P V S H Y F P L T E V G - V E P S P
Trubripes (fish)	
Drerio(zebrafish)	E E D K S T S H F F P L T E V G G M E A G R
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	D E E K V V S R F Y P M T E V G L - E A S

GJA8, p.(Pro280Arg), P1

species	alignment
Human	M G D W S F L G N I L E E V N E H S T
mutated	M G D W S F P G N I L E E V N E H S T V I
Ptroglodytes (chimpanzee)	M G D W S F L G N I L E E V N E H S T V I
Mmulatta (Monkey)	M G D W S F L G N I L E E V N E H S T V I
Fcatus (Cat)	
Mmusculus (mouse)	M G D W S F L G N I L E E V N E H S T V I
Ggallus (rooster)	M G D W S F L G N I L E Q V N E Q S T V I
Trubripes (fish)	
Drerio(zebrafish)	M G D W S F L G N I L E E V N E H S T V I
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	M G D W S F L G N I L E E V N E H S T V I

GJA8, p.(Leu7Pro), BB 16 cat

species	alignment
Human	T L E Q A I K S A H E C Y D D E I Q L Y N E Q I
mutated	T L E Q A I K S A H E S Y D D E I Q L Y N E Q
Ptroglodytes (chimpanzee)	T L E Q A I K S A H E C Y D D E I Q L Y N E Q
Mmulatta (Monkey)	T L E Q A I K S A H E C Y D D E I Q L Y N E Q
Fcatus (Cat)	A L E Q A I K D A H E C Y D D E I Q L Y N E Q
Mmusculus (mouse)	A L E Q A I K H A H E C Y D E E L Q L Y N E Q
Ggallus (rooster)	M L E Q A I K N T Q E S Y D D E I Q L Y N E Q
Trubripes (fish)	
Drerio(zebrafish)	T L E Q T I K T T Q E S Y D D E I Q L Y N E Q
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	

BFSP1, p.Cys259Ser, PCC01-97A

species	alignment
Human	F Y V F F G L G S S L R W A P G P L H V L Q A
mutated	F Y V F F G L G S S L C W A P G P L H V L Q
Ptroglodytes (chimpanzee)	F Y V F F G L G S S L R W A P G P L H V L Q
Mmulatta (Monkey)	F Y V F F G L G S S L R W A P G P L H V L Q
Fcatus (Cat)	
Mmusculus (mouse)	F Y V F F G L G A S L R W A P G P L H V L Q
Ggallus (rooster)	
Trubripes (fish)	F F V F F G L G A A L R W T T G P L N V L H
Drerio(zebrafish)	F F V F F G L G A A L R W T T G P H N V L Q
Dmelanogaster (Fruit fly)	F L I F V G V G S T T S G S - - - - - V P Q
Celegans (Roundworm)	
Xtropicalis (Frog)	F Y V F F G L G A S L K W A A G P A N V L N

MIP, p.(Arg33Cys), pp-50 cat

species	alignment
Human	E E I A F L K K L H E E E I Q E L Q A Q I Q
mutated	E E I A F L K K L H E G E I Q E L Q A Q I Q
Ptroglodytes (chimpanzee)	E E I A F L K K L H E E E I Q E L Q A Q I Q
Mmulatta (Monkey)	E E I A F L K K L H E E E I Q E L Q A Q I Q
Fcatus (Cat)	E E I A F L K K L H D E E I Q E L Q A Q I Q
Mmusculus (mouse)	E E I A F L K K L H D E E I Q E L Q A Q I Q
Ggallus (rooster)	E E I V F L K K L H D E E I R E L Q A Q L Q
Trubripes (fish)	E E I N F L K K L H D E E M L E L Q N Q V H
Drerio(zebrafish)	D E E L A E M Q I Q M Q
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	H D E E I R E L Q L Q I Q

VIM, p.(Glu240Val), PCC 02-105

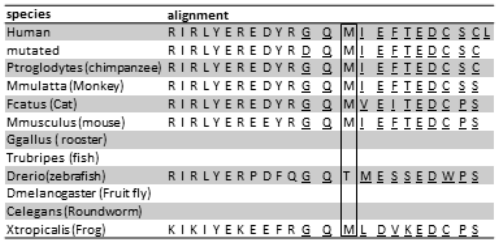
species	alignment
Human	D G D K G E K G S P G F P G E P G
mutated	D G D K G E K G R P G F P G E P G
Ptroglodytes (chimpanzee)	T G P P G P P G P P G P G E K G
Mmulatta (Monkey)	D G D K G E K G S P G F P G E P G
Fcatus (Cat)	D G E K G E K G S P G F P G D A G
Mmusculus (mouse)	D G E K G E R G S P G I P G D S G
Ggallus (rooster)	D G E P G P K G E P G L P G G F G
Trubripes (fish)	D G D V G S K G E K G F Q G A P G
Drerio(zebrafish)	D G Q S G E K G E I G F P G N P G
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	K G D R G L D G L G G I P G L P G
Xtropicalis (Frog)	D G D Q G E K G L Q G V P G I

COL4A1, p.Ser300Arg, SR 12 cat

species	alignment
Human	P G I P G Q D G P P G P P G I P G C N G T K G E
mutated	P G I P G Q D G P P G R P G I P G C N G T K G
Ptroglodytes (chimpanzee)	G V K G Q K G E R G L P G L Q G V I G F P G
Mmulatta (Monkey)	G I P G Q D G P P G P P G I P G C N G T K G
Fcatus (Cat)	P G I P G Q D G P P G P P G I P G C N G T K G
Mmusculus (mouse)	P G I P G Q D G P P G P P G I P G C N G T K G
Ggallus (rooster)	P G I P G Q D G P P G P P G I P G C N G T K G
Trubripes (fish)	P G I N G N D G P A G P P G I P G C N G T K G
Drerio(zebrafish)	P G L P G L E G A P G P Q G I P G C N G T K G
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	N G R P G E P G P P G A P G W D G C N G T D G
Xtropicalis (Frog)	P G I P G Q D G P P G P S G I P G C N G

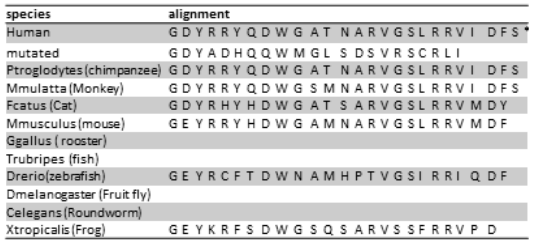
COL4A, p.(Pro119Arg), pp-50 cat-non segregating

B



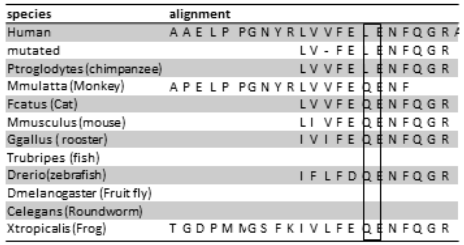
species	alignment
Human	R I R L Y E R E D Y R G Q M I E F T E D C S C L
mutated	R I R L Y E R E D Y R D Q M I E F T E D C S C
Ptroglodytes (chimpanzee)	R I R L Y E R E D Y R G Q M I E F T E D C S C
Mmulatta (Monkey)	R I R L Y E R E D Y R G Q M I E F T E D C S S
Fcatus (Cat)	R I R L Y E R E D Y R G Q M V E I T E D C P S
Mmusculus (mouse)	R I R L Y E R E E Y R G Q M I E F T E D C P S
Ggallus (rooster)	
Trubripes (fish)	
Drerio(zebrafish)	R I R L Y E R P D F Q G Q T M E S S E D W P S
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	K I K I Y E K E E F R G Q M L D V K E D C P S

CRYGD, p.(Gly100Asp), PCC 02-105



species	alignment
Human	G D Y R R Y Q D W G A T N A R V G S L R R V I D F S *
mutated	G D Y A D H Q Q W M G L S D S V R S C R L I
Ptroglodytes (chimpanzee)	G D Y R R Y Q D W G A T N A R V G S L R R V I D F S
Mmulatta (Monkey)	G D Y R R Y Q D W G S M N A R V G S L R R V I D F S
Fcatus (Cat)	G D Y R H Y H D W G A T S A R V G S L R R V M D Y
Mmusculus (mouse)	G E Y R R Y H D W G A M N A R V G S L R R V M D F
Ggallus (rooster)	
Trubripes (fish)	
Drerio(zebrafish)	G E Y R C F T D W N A M H P T V G S I R R I Q D F
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	G E Y K R F S D W G S Q S A R V S S F R R V P D

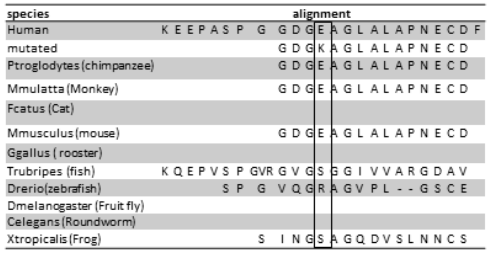
CRYGD, p.(Thr160Argfs*8), PCC 02-105



species	alignment
Human	A A E L P P G N Y R L V V F E L E N F Q G R A
mutated	L V - F E L E N F Q G R
Ptroglodytes (chimpanzee)	L V V F E L E N F Q G R
Mmulatta (Monkey)	A P E L P P G N Y R L V V F E Q E N F
Fcatus (Cat)	L V V F E Q E N F Q G R
Mmusculus (mouse)	L I V F E Q E N F Q G R
Ggallus (rooster)	I V I F E Q E N F Q G R
Trubripes (fish)	
Drerio(zebrafish)	I F L F D Q E N F Q G R
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	T G D P M M G S F K I V L F E Q E N F Q G R

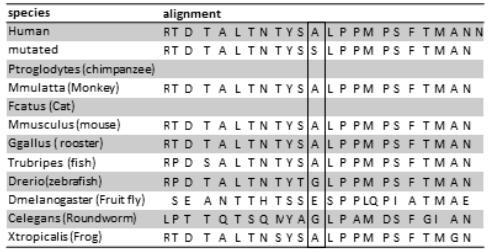
CRYBB1, p.Val62_Phe64del, PCC01-97A

C



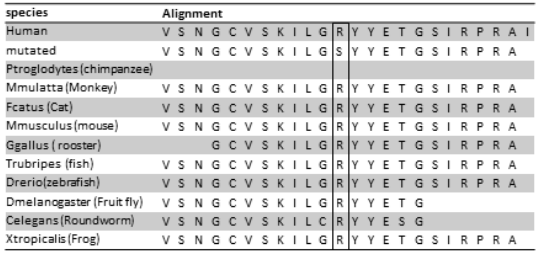
species	alignment
Human	K E E P A S P G G D G E A G L A L A P N E C D F
mutated	G D G K A G L A L A P N E C D
Ptroglodytes (chimpanzee)	G D G E A G L A L A P N E C D
Mmulatta (Monkey)	G D G E A G L A L A P N E C D
Fcatus (Cat)	
Mmusculus (mouse)	G D G E A G L A L A P N E C D
Ggallus (rooster)	
Trubripes (fish)	K Q E P V S P G V R G V G S G G I V V A R G D A V
Drerio(zebrafish)	S P G V Q G R A G V P L - - G S C E
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	S I N G S A G Q D V S L N N C S

HSF4, p.Glu304Lys, PCC 01-34



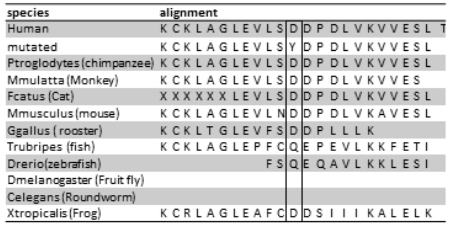
species	alignment
Human	R T D T A L T N T Y S A L P P M P S F T M A N N
mutated	R T D T A L T N T Y S S L P P M P S F T M A N
Ptroglodytes (chimpanzee)	
Mmulatta (Monkey)	R T D T A L T N T Y S A L P P M P S F T M A N
Fcatus (Cat)	
Mmusculus (mouse)	R T D T A L T N T Y S A L P P M P S F T M A N
Ggallus (rooster)	R T D T A L T N T Y S A L P P M P S F T M A N
Trubripes (fish)	R P D S A L T N T Y S A L P P M P S F T M A N
Drerio(zebrafish)	R P D T A L T N T Y T G L P P M P S F T M A N
Dmelanogaster (Fruit fly)	S E A N T T H T S S E S P P L Q P I A T M A E
Celegans (Roundworm)	L P T T Q T S Q N Y A G L P A M D S F G I A N
Xtropicalis (Frog)	R T D T A L T N S Y S A L P P M P S F T M G N

PAX6, p.Ala328Ser, PCC01-97A



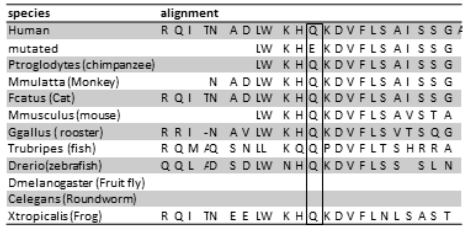
species	Alignment
Human	V S N G C V S K I L G R Y Y E T G S I R P R A I
mutated	V S N G C V S K I L G S Y Y E T G S I R P R A
Ptroglodytes (chimpanzee)	
Mmulatta (Monkey)	V S N G C V S K I L G R Y Y E T G S I R P R A
Fcatus (Cat)	V S N G C V S K I L G R Y Y E T G S I R P R A
Mmusculus (mouse)	V S N G C V S K I L G R Y Y E T G S I R P R A
Ggallus (rooster)	G C V S K I L G R Y Y E T G S I R P R A
Trubripes (fish)	V S N G C V S K I L G R Y Y E T G S I R P R A
Drerio(zebrafish)	V S N G C V S K I L G R Y Y E T G S I R P R A
Dmelanogaster (Fruit fly)	V S N G C V S K I L G R Y Y E T G
Celegans (Roundworm)	V S N G C V S K I L C R Y Y E S G
Xtropicalis (Frog)	V S N G C V S K I L G R Y Y E T G S I R P R A

PAX6, p.(Arg59Ser), SR 11 cat



species	alignment
Human	K C K L A G L E V L S D D P D L V K V V E S L T
mutated	K C K L A G L E V L S Y D P D L V K V V E S L
Ptroglodytes (chimpanzee)	K C K L A G L E V L S D D P D L V K V V E S L
Mmulatta (Monkey)	K C K L A G L E V L S D D P D L V K V V E S
Fcatus (Cat)	X X X X X X L E V L S D D P D L V K V V E S L
Mmusculus (mouse)	K C K L A G L E V L N D D P D L V K A V E S L
Ggallus (rooster)	K C K L T G L E V F S D D P L L L K
Trubripes (fish)	K C K L A G L E P F C Q E P E V L K K F E T I
Drerio(zebrafish)	F S Q E Q A V L K K L E S I
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	K C R L A G L E A F C D D S I I I K A L E L K

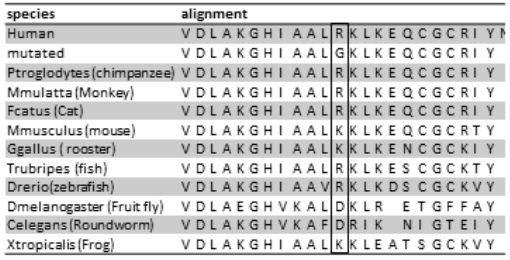
TDRD7, p.Asp587Tyr, PCC01-97A



species	alignment
Human	R Q I T N A D L W K H Q K D V F L S A I S S G A
mutated	L W K H E K D V F L S A I S S G
Ptroglodytes (chimpanzee)	L W K H Q K D V F L S A I S S G
Mmulatta (Monkey)	N A D L W K H Q K D V F L S A I S S G
Fcatus (Cat)	R Q I T N A D L W K H Q K D V F L S A I S S G
Mmusculus (mouse)	L W K H Q K D V F L S A V S T A
Ggallus (rooster)	R R I - N A V L W K H Q K D V F L S V T S Q G
Trubripes (fish)	R Q M A Q S N L L K Q Q P D V F L T S H R R A
Drerio(zebrafish)	Q Q L A D S D L W N H Q K D V F L S S S L N
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	R Q I T N E E L W K H Q K D V F L N L S A S T

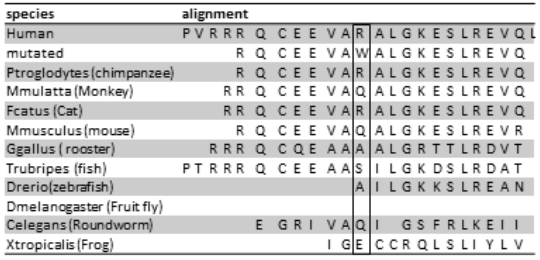
TDRD7, p.Glu304Lys, SR 12 cat

D



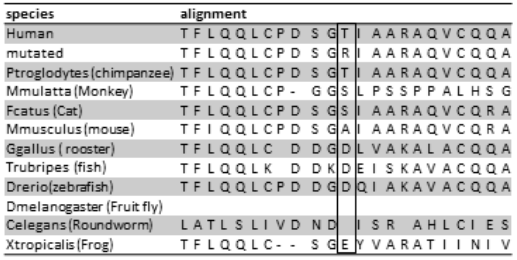
species	alignment
Human	V D L A K G H I A A L R K L K E Q C G C R I Y
mutated	V D L A K G H I A A L G K L K E Q C G C R I Y
Ptroglodytes (chimpanzee)	V D L A K G H I A A L R K L K E Q C G C R I Y
Mmulatta (Monkey)	V D L A K G H I A A L R K L K E Q C G C R I Y
Fcatus (Cat)	V D L A K G H I A A L R K L K E Q C G C R I Y
Mmusculus (mouse)	V D L A K G H I A A L K K L K E Q C G C R T Y
Ggallus (rooster)	V D L A K G H I A A L K K L K E N C G C K I Y
Trubripes (fish)	V D L A K G H I A A L R K L K E S C G C K T Y
Drerio(zebrafish)	V D L A K G H I A A V R K L K D S C G C K V Y
Dmelanogaster (Fruit fly)	V D L A E G H V K A L D K L R E T G F F A Y
Celegans (Roundworm)	V D L A K G H V K A F D R I K N I G T E I Y
Xtropicalis (Frog)	V D L A K G H I A A L K K L E A T S G C K V Y

GALE, p.(Arg256Gly), E1



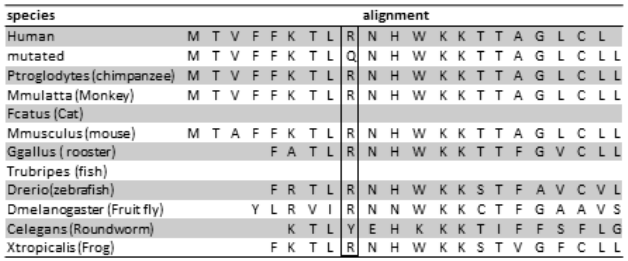
species	alignment
Human	P V R R R Q C E E V A R A L G K E S L R E V Q L
mutated	R Q C E E V A W A L G K E S L R E V Q
Ptroglodytes (chimpanzee)	R Q C E E V A R A L G K E S L R E V Q
Mmulatta (Monkey)	R R Q C E E V A Q A L G K E S L R E V Q
Fcatus (Cat)	R R Q C E E V A R A L G K E S L R E V Q
Mmusculus (mouse)	R Q C E E V A Q A L G K E S L R E V R
Ggallus (rooster)	R R R Q C Q E A A A A L G R T T L R D V T
Trubripes (fish)	P T R R R Q C E E A A S I L G K D S L R D A T
Drerio(zebrafish)	A I L G K K S L R E A N
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	E G R I V A Q I G S F R L K E I I
Xtropicalis (Frog)	I G E C C R Q L S L I Y L V

GALK1, p.Arg248Trp, W1



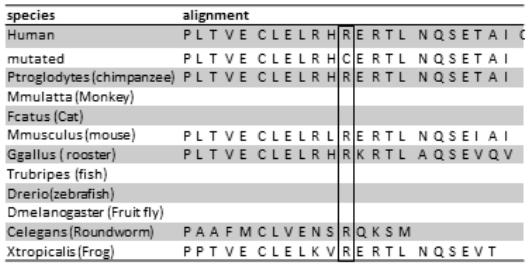
species	alignment
Human	T F L Q Q L C P D S G T I A A R A Q V C Q Q A
mutated	T F L Q Q L C P D S G R I A A R A Q V C Q Q A
Ptroglodytes (chimpanzee)	T F L Q Q L C P D S G T I A A R A Q V C Q Q A
Mmulatta (Monkey)	T F L Q Q L C P - G G S L P S S P P A L H S G
Fcatus (Cat)	T F L Q Q L C P D S G S I A A R A Q V C Q R A
Mmusculus (mouse)	T F I Q Q L C P D S G A I A A R A Q V C Q R A
Ggallus (rooster)	T F L Q Q L C D D G D L V A K A L A C Q Q A
Trubripes (fish)	T F L Q Q L K D D K D E I S K A V A C Q Q A
Drerio(zebrafish)	T F L Q Q L C P D D G D Q I A K A V A C Q Q A
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	L A T L S L I V D N D I S R A H L C I E S
Xtropicalis (Frog)	T F L Q Q L C - - S G E Y V A R A T I I N I V

GALK1, p.Thr162Arg, W1



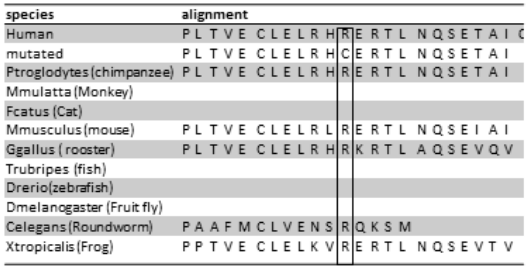
species	alignment
Human	M T V F F K T L R N H W K K T T A G L C L
mutated	M T V F F K T L Q N H W K K T T A G L C L L
Ptroglodytes (chimpanzee)	M T V F F K T L R N H W K K T T A G L C L L
Mmulatta (Monkey)	M T V F F K T L R N H W K K T T A G L C L L
Fcatus (Cat)	
Mmusculus (mouse)	M T A F F K T L R N H W K K T T A G L C L L
Ggallus (rooster)	F A T L R N H W K K T T F G V C L L
Trubripes (fish)	
Drerio(zebrafish)	F R T L R N H W K K S T F A V C V L
Dmelanogaster (Fruit fly)	Y L R V I R N N W K K C T F G A A V S
Celegans (Roundworm)	K T L Y E H K K K T I F F S F L G
Xtropicalis (Frog)	F K T L R N H W K K S T V G F C L L

AGK, p.Arg9Gln, PCC01-97A



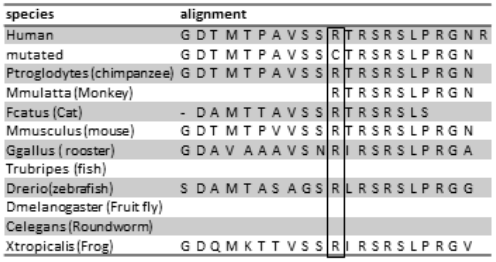
species	alignment
Human	P L T V E C L E L R H R E R T L N Q S E T A I C
mutated	P L T V E C L E L R H C E R T L N Q S E T A I
Ptroglodytes (chimpanzee)	P L T V E C L E L R H R E R T L N Q S E T A I
Mmulatta (Monkey)	
Fcatus (Cat)	
Mmusculus (mouse)	P L T V E C L E L R L R E R T L N Q S E I A I
Ggallus (rooster)	P L T V E C L E L R H R K R T L A Q S E V Q V
Trubripes (fish)	
Drerio(zebrafish)	
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	P A A F M C L V E N S R Q K S M
Xtropicalis (Frog)	P P T V E C L E L K V R E R T L N Q S E V T

GCNT2, p.Arg385Cys, PCC 10-188



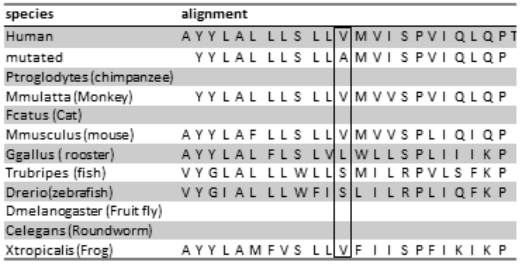
species	alignment
Human	P L T V E C L E L R H R E R T L N Q S E T A I C
mutated	P L T V E C L E L R H C E R T L N Q S E T A I
Ptroglodytes (chimpanzee)	P L T V E C L E L R H R E R T L N Q S E T A I
Mmulatta (Monkey)	
Fcatus (Cat)	
Mmusculus (mouse)	P L T V E C L E L R L R E R T L N Q S E I A I
Ggallus (rooster)	P L T V E C L E L R H R K R T L A Q S E V Q V
Trubripes (fish)	
Drerio(zebrafish)	
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	P A A F M C L V E N S R Q K S M
Xtropicalis (Frog)	P P T V E C L E L K V R E R T L N Q S E V T V

GCNT2, p.(Arg385Cys), PCC 10-183



species	alignment
Human	G D T M T P A V S S R T R S R S L P R G N R
mutated	G D T M T P A V S S C T R S R S L P R G N
Ptroglodytes (chimpanzee)	G D T M T P A V S S R T R S R S L P R G N
Mmulatta (Monkey)	R T R S R S L P R G N
Fcatus (Cat)	- D A M T T A V S S R T R S R S L S
Mmusculus (mouse)	G D T M T P V V S S R T R S R S L P R G N
Ggallus (rooster)	G D A V A A A V S N R I R S R S L P R G A
Trubripes (fish)	
Drerio(zebrafish)	S D A M T A S A G S R L R S R S L P R G G
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	G D Q M K T T V S S R I R S R S L P R G V

NHS, p.Arg480Cys, PCC 10-188



species	alignment
Human	A Y Y L A L L L S L L V M V I S P V I Q L Q P T
mutated	Y Y L A L L L S L L A M V I S P V I Q L Q P
Ptroglodytes (chimpanzee)	
Mmulatta (Monkey)	Y Y L A L L L S L L V M V V S P V I Q L Q P
Fcatus (Cat)	
Mmusculus (mouse)	A Y Y L A F L L S L L V M V V S P L I Q I Q P
Ggallus (rooster)	A Y Y L A L F L S L V L W L L S P L I I I K P
Trubripes (fish)	V Y G L A L L L W L L S M I L R P V L S F K P
Drerio(zebrafish)	V Y G I A L L L W F I S L I L R P L I Q F K P
Dmelanogaster (Fruit fly)	
Celegans (Roundworm)	
Xtropicalis (Frog)	A Y Y L A M F V S L L V F I I S P F I K I K P

NSDHL, p.Val311Ala, pp-50 cat-non segregating

Appendix 17

The coverage/read data statistics on CSA106, CRCH26, CRCH11 and CSA92.

Sample	Total reads	Average throughput depth of target regions (X)	% Coverage of target regions
CSA106.01	74,309,034	146.6	97.7%
CSA106.02	72,275,718	142.6	97.9%
CSA10.03	80,702,012	159.2	97.9%
CSA106.04	75,097,814	148.2	97.9%
CSA106.05	80,767,028	159.4	97.8%
CSA106.08	64,935,270	128.1	97.4%
CSA106.09	57,781,902	114.0	96.7%
CSA106.14	53,952,118	106.5	96.9%
CSA106.15	49,007,814	96.7	96.6%
CSA106.17	64,648,410	127.6	97.0%
CSA.106.19	63,748,790	125.8	96.9%
CRCH26.01	69,874,992	137.9	96.7%
CRCH26.02	70,997,214	140.1	96.1%
CRCH26.03	62,484,852	123.3	97.1%
CRCH26.04	72,735,780	143.5	97.2%
CRCH26.05	70,491,116	139.1	97.1%
CRCH26.06	64,385,216	127.0	96.2%
CRCH11.01	74,341,022	146.7	96.7%
CRCH11.02	55,396,286	109.3	97.3%
CRCH11.03	70,972,528	140.0	97.2%
CRCH11.04	63,052,478	124.4	97.0%
CRCH11.05	62,982,382	124.3	93.7%
CRCH11.06	71,392,008	140.9	97.5%

Appendix 17-continued

Sample	Total reads	Average throughput depth of target regions (x)	% Coverage of target regions
CSA92.01	122979012	64.7	93%
CSA92.02	98675304	84.9	95%
CSA92.03	94643944	106	95%
CSA92.04	204951664	78.4	94%
CSA92.05	95739178	79.2	94%

Summary statistics of exome sequenced family members showing total reads per sample, average read depth of target region and coverage percentage of target region covered more than 10 fold for CRCH11, CRCH26 and CSA106 (provided by Macrogen) and more than 8 fold for CSA92 (provided by our collaborator in State Key Laboratory of Medical Genetics, Central South University, Changsha, China).

Appendix 18

List of primers used to validate the variations detected in families CSA92 and CRCH11.

Family	Position	Gene	Forward primer	Reverse primer	Product size	±Q / Mg^{+2} (mM)	PCR cycles annealing temperature
CSA92	Chr11:83674028	*DLG2*	AGAAAAACCTTGGCATGCAG	GCATCCAAAACAACACCTGA	348	1.5	30/57°C
CSA92	Chr3: 88040764	*HTR1F*	TCCTTTTGGAGAGTGGTGAGA	GCTTTTGGAATGCTTTCTTGA	396	1.5	30/57°C
CSA92	Chr19: 55358681	*KIR2DS4*	AAGCAGGGGAAAGCTAGGTC	AGAAAGTCCTGCCTCTGTGG	386	1.5	30/ 57°C
CSA92	Chr11: 116649768	*ZNF259*	GTTCTTGGGAAGGTGGGAAT	CACTTGCATCACAAGCTGTTT	398	1.5	30/ 57°C
CSA92	Chr11: 88330507	*GRM5*	TCATGGAAAATTCCAGTGCTA	TGGGATTTCACAAGGAATGAG	446	1.5	30/57°C
CSA92	chr5:76373168	*ZBED3*	TGTGGAGGCACCTGAGGA	ATTGGCATTGGACGGAGA	492	1.5	30/7°C
CSA92	Chr3: 75786933	*ZNF717*	GAGAACTCATGCTGGCAAAA	TGAGGAATGACTTGCGATGA	242	1.5	30/57°C
CSA92	Chr16:67201032	*HSF4*	TAGCAACAGGCCTCAGCTCT	GGAGAAGGTTTGAGGGAGGA	391	+Q/ 1.5	30/63 °C
CSA92	Chr20:17479685	*BFSP1*	ACCGAGTCCAGGTTCCTTTT	GGCACACAATAGGCACTCAA	380	1.5	30/57°C
CRCH11	Chr1:6585916	*NOL9*	AGCACTGTAAACCCCTCTGG	GCTCATGCAGCTTTAAAAGAGA	398	1.5	35/58°C
CRCH11	Chr1:23885474	*ID3*	CTACAGCGCGTCATCGACTA	TGCACATCACATGGAAACTG	458	1.5	35/58°C
CRCH11	Chr2:234231596	*SAG*	GCAGGAAATTTTGGGAAGGT	CGCCTCCACACAAAATACAA	484	1.5	35/58°C

The *BFSP1* and *HSF4* variations in CSA92 detected via Ion torrent PGM sequencing. Other variants detected via exome sequencing

Appendix 19

List of primers used for sequencing the 5'UTR and translated region of *HTR1F* gene in 43 unsolved congenital cataract cases.

Family	Gene	Region	Forward primer	Reverse primer	±Q / Mg^{+2} (mM)	PCR cycles/ annealing temperature
CSA92	*HTR1F*	5 ' UTR	CCGATTCATAGGGAGACACAA	GAGTCAAGATGCACTGCAGC	1.5	57/30°C
CSA92	*HTR1F*	Exon1	AAAGGAAGAGAAAAGTTCTTGAAGC	CTATTCTGGAGGCACCAAGG	1.5	57/30°C
CSA92	*HTR1F*	Exon1	GGTCTGTGACATTTGGCTGA	AGAACGGAAAGCAGCCACTA	1.5	57/30°C
CSA92	*HTR1F*	Exon 1	AGGATTGCAAAGGAGGAGG	GGTTTTTGAGGGGAGGAATAA	1.5	57/30°C

Appendix 20

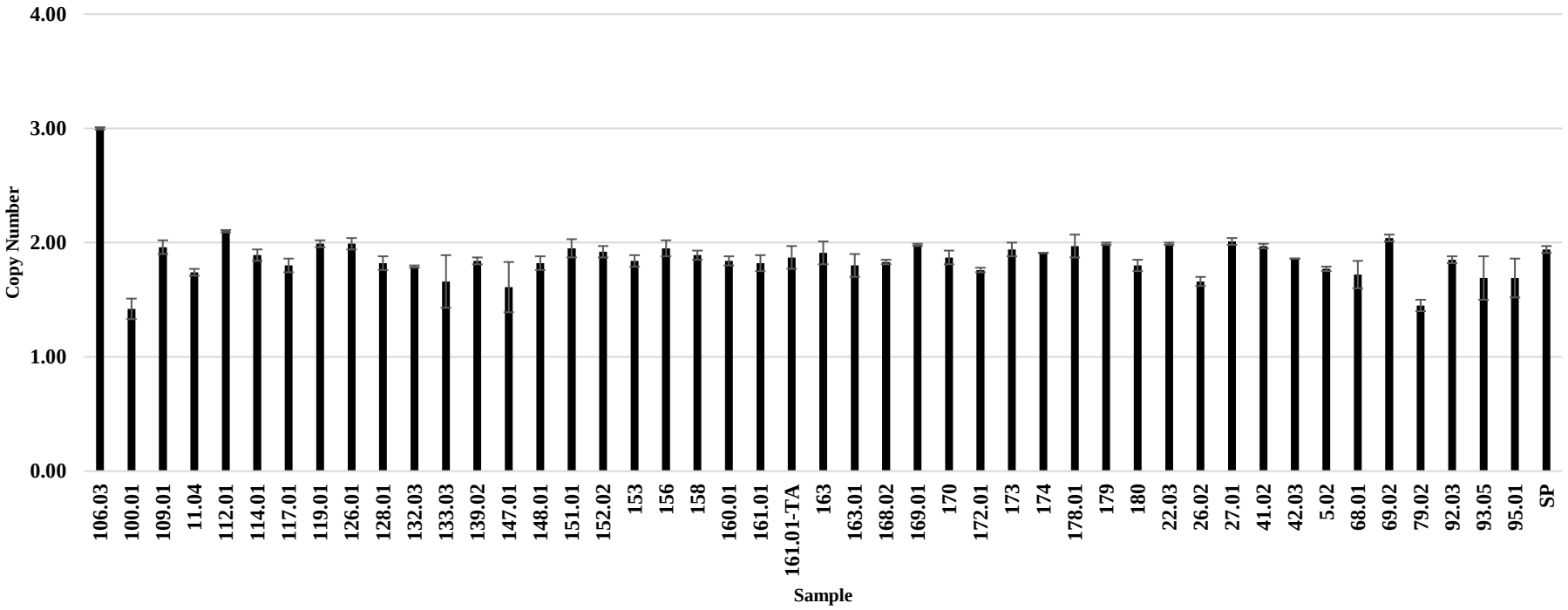
List of primers used for sequencing the 5'UTR and translated regions of *NOL9* gene in 47 unsolved congenital cataract cases.

Family	Gene	Region	Forward primer	Reverse primer
CRCH11	*NOL9*	5"UTR +Exon 1	GCTGGAACCTCCTAGTGTGC	GACCACGGTCCTCTCCTACC
CRCH11	*NOL9*	Exon 2	TCTTAGCCAGGATGGTCTCG	GGCAACAAGAGCAAAACTCC
CRCH11	*NOL9*	Exon 3	GTGTGCTAAGGACGTTGTGG	GGAGGATGAACATTTTTCACG
CRCH11	*NOL9*	Exon 4	GTGAAGGCTCGAGAGAAAGC	TGGGAAACAAACATTGATCG
CRCH11	*NOL9*	Exon 5	TCCAACTCCCCAGATTAAACC	GCAGGTAAGGCAAAGTCAGC
CRCH11	*NOL9*	Exon 6	CAGCGATTGTCATCTTGAGG	AAGTGTGAGCCACCACAGG
CRCH11	*NOL9*	Exon 7	GCTGAAACAAAACCCTCACC	AAATTAGCTGGGCAGAGTGG
CRCH11	*NOL9*	Exon 8	CTTCTGGCCTCAAGTGATCC	AGGTCTTTTTCGACGGAAGG
CRCH11	*NOL9*	Exon 9	GACCTGCCATTCTCAATTCC	TCCCTCAGCCTCTCTACACC
CRCH11	*NOL9*	Exon 10	AGTAGGCACTTTGGCAGACC	CCAGCAGTGACTGAGACAGC
CRCH11	*NOL9*	Exon 11	CCGGGTTCAAGCTATTCTCC	ACCCCTGACCCTACTTCAGC
CRCH11	*NOL9*	Exon 12	GACCATCACCATGACACTGC	GCCTTCTTCCTTCCTTCTCG

Primer design and sequencing performed via AGRF.

Appendix 21

Assessment of CRYBB1 gene dosage in duplicated region by quantitative polymerase chain reaction in congenital cataract probands without mutation in known genes.



References

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30(1):97-101.

Abou Tayoun AN, Tunkey CD, Pugh TJ, Ross T, Shah M, Lee CC, Harkins TT, Wells WA, Tafe LJ, Amos CI and others. 2013. A comprehensive assay for CFTR mutational analysis using next-generation sequencing. Clin Chem 59(10):1481-8.

Addison PK, Berry V, Holden KR, Espinal D, Rivera B, Su H, Srivastava AK, Bhattacharya SS. 2006. A novel mutation in the connexin 46 gene (GJA3) causes autosomal dominant zonular pulverulent cataract in a Hispanic family. Mol Vis 12:791-5.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7(4):248-9.

Akarsu AN, Turacli ME, Aktan SG, Barsoum-Homsy M, Chevrette L, Sayli BS, Sarfarazi M. 1996. A second locus (GLC3B) for primary congenital glaucoma (Buphthalmos) maps to the 1p36 region. Hum Mol Genet 5(8):1199-203.

Al-Haddad C, Abdulaal M, Badra R, Barikian A, Noureddine B, Farra C. 2015. Genotype/Phenotype Correlation in Primary Congenital Glaucoma Patients in the Lebanese Population: A Pilot Study. Ophthalmic Genet:1-6.

Aldahmesh MA, Khan AO, Mohamed J, Alkuraya FS. 2011. Novel recessive BFSP2 and PITX3 mutations: insights into mutational mechanisms from consanguineous populations. Genet Med 13(11):978-81.

Aldahmesh MA, Khan AO, Mohamed JY, Alghamdi MH, Alkuraya FS. 2012a. Identification of a truncation mutation of acylglycerol kinase (AGK) gene in a novel autosomal recessive cataract locus. Hum Mutat 33(6):960-2.

Aldahmesh MA, Khan AO, Mohamed JY, Hijazi H, Al-Owain M, Alswaid A, Alkuraya FS. 2012b. Genomic analysis of pediatric cataract in Saudi Arabia reveals novel candidate disease genes. Genet Med 14(12):955-62.

AlFadhli S, Abdelmoaty S, Al-Hajeri A, Behbehani A, Alkuraya F. 2012. Novel crystallin gamma B mutations in a Kuwaiti family with autosomal dominant congenital cataracts reveal genetic and clinical heterogeneity. Mol Vis 18:2931-6.

Alfadhli S, Behbehani A, Elshafey A, Abdelmoaty S, Al-Awadi S. 2006. Molecular and clinical evaluation of primary congenital glaucoma in Kuwait. Am J Ophthalmol 141(3):512-6.

Ali M, McKibbin M, Booth A, Parry DA, Jain P, Riazuddin SA, Hejtmancik JF, Khan SN, Firasat S, Shires M and others. 2009. Null mutations in LTBP2 cause primary congenital glaucoma. Am J Hum Genet 84(5):664-71.

Amaya L, Taylor D, Russell-Eggitt I, Nischal KK, Lengyel D. 2003. The Morphology and Natural History of Childhood Cataracts. Survey of Ophthalmology 48(2):125-144.

Ambros V. 2004. The functions of animal microRNAs. Nature 431(7006):350-5.

Andley UP. 2007. Crystallins in the eye: Function and pathology. Prog Retin Eye Res 26(1):78-98.

Armitage MM, Kivlin JD, Ferrell RE. 1995. A progressive early onset cataract gene maps to human chromosome 17q24. Nat Genet 9(1):37-40.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22(2):195-201.
Augusteyn RC. 2010. On the growth and internal structure of the human lens. Experimental Eye Research 90(6):643-654.
Awadalla MS, Fingert JH, Roos BE, Chen S, Holmes R, Graham SL, Chehade M, Galanopolous A, Ridge B, Souzeau E and others. 2015. Copy number variations of TBK1 in Australian patients with primary open-angle glaucoma. Am J Ophthalmol 159(1):124-30 e1.
Ayme S, Philip N. 1996. Fine-Lubinsky syndrome: a fourth patient with brachycephaly, deafness, cataract, microstomia and mental retardation. Clin Dysmorphol 5(1):55-60.
Azmanov DN, Dimitrova S, Florez L, Cherninkova S, Draganov D, Morar B, Saat R, Juan M, Arostegui JI, Ganguly S and others. 2011. LTBP2 and CYP1B1 mutations and associated ocular phenotypes in the Roma/Gypsy founder population. Eur J Hum Genet 19(3):326-33.
Azuma N, Hirakiyama A, Inoue T, Asaka A, Yamada M. 2000. Mutations of a human homologue of the Drosophila eyes absent gene (EYA1) detected in patients with congenital cataracts and ocular anterior segment anomalies. Hum Mol Genet 9(3):363-6.
Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12(11):745-55.
Bayle JH, Randazzo F, Johnen G, Kaufman S, Nagy A, Rossant J, Crabtree GR. 2002. Hyperphenylalaninemia and impaired glucose tolerance in mice lacking the bifunctional DCoH gene. J Biol Chem 277(32):28884-91.
Behnam M, Imagawa E, Chaleshtori AR, Ronasian F, Salehi M, Miyake N, Matsumoto N. 2016. A novel homozygous mutation in HSF4 causing autosomal recessive congenital cataract. J Hum Genet 61(2):177-9.
Bejjani BA, Lewis RA, Tomey KF, Anderson KL, Dueker DK, Jabak M, Astle WF, Otterud B, Leppert M, Lupski JR. 1998. Mutations in CYP1B1, the gene for cytochrome P4501B1, are the predominant cause of primary congenital glaucoma in Saudi Arabia. Am J Hum Genet 62(2):325-33.
Bejjani BA, Stockton DW, Lewis RA, Tomey KF, Dueker DK, Jabak M, Astle WF, Lupski JR. 2000. Multiple CYP1B1 mutations and incomplete penetrance in an inbred population segregating primary congenital glaucoma suggest frequent de novo events and a dominant modifier locus. Hum Mol Genet 9(3):367-74.
Belmouden A, Melki R, Hamdani M, Zaghloul K, Amraoui A, Nadifi S, Akhayat O, Garchon HJ. 2002. A novel frameshift founder mutation in the cytochrome P450 1B1 (CYP1B1) gene is associated with primary congenital glaucoma in Morocco. Clin Genet 62(4):334-9.
Bennett TM, Mackay DS, Knopf HL, Shiels A. 2004. A novel missense mutation in the gene for gap-junction protein alpha3 (GJA3) associated with autosomal dominant "nuclear punctate" cataracts linked to chromosome 13q. Mol Vis 10:376-82.
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(1):15-7.
Berry V, Francis P, Reddy MA, Collyer D, Vithana E, MacKay I, Dawson G, Carey AH, Moore A, Bhattacharya SS and others. 2001. Alpha-B crystallin gene

(CRYAB) mutation causes dominant congenital posterior polar cataract in humans. Am J Hum Genet 69(5):1141-5.
Berry V, Francis PJ, Prescott Q, Waseem NH, Moore AT, Bhattacharya SS. 2011. A novel 1-bp deletion in PITX3 causing congenital posterior polar cataract. Mol Vis 17:1249-53.
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(3):415-9.
Berry V, Mackay D, Khaliq S, Francis PJ, Hameed A, Anwar K, Mehdi SQ, Newbold RJ, Ionides A, Shiels A and others. 1999. Connexin 50 mutation in a family with congenital "zonular nuclear" pulverulent cataract of Pakistani origin. Hum Genet 105(1-2):168-70.
Beyer EC, Berthoud VM. 2014. Connexin hemichannels in the lens. Front Physiol 5:20.
Beyer EC, Ebihara L, Berthoud VM. 2013. Connexin mutants and cataracts. Front Pharmacol 4:43.
Bidinost C, Matsumoto M, Chung D, Salem N, Zhang K, Stockton DW, Khoury A, Megarbane A, Bejjani BA, Traboulsi EI. 2006. Heterozygous and homozygous mutations in PITX3 in a large Lebanese family with posterior polar cataracts and neurodevelopmental abnormalities. Invest Ophthalmol Vis Sci 47(4):1274-80.
Bielczyk-Maczynska E, Lam Hung L, Ferreira L, Fleischmann T, Weis F, Fernandez-Pevida A, Harvey SA, Wali N, Warren AJ, Barroso I and others. 2015. The Ribosome Biogenesis Protein Nol9 Is Essential for Definitive Hematopoiesis and Pancreas Morphogenesis in Zebrafish. PLoS Genet 11(12):e1005677.
Billingsley G, Santhiya ST, Paterson AD, Ogata K, Wodak S, Hosseini SM, Manisastry SM, Vijayalakshmi P, Gopinath PM, Graw J and others. 2006. CRYBA4, a novel human cataract gene, is also involved in microphthalmia. Am J Hum Genet 79(4):702-9.
Boerrigter RM, Siertsema JV, Kema IP. 1992. Serotonin (5-HT) and the rat's eye. Some pilot studies. Doc Ophthalmol 82(1-2):141-50.
Borck G, Kakar N, Hoch J, Friedrich K, Freudenberg J, Nurnberg G, Yilmaz R, Daud S, Baloch DM, Nurnberg P and others. 2012. An Alu repeat-mediated genomic GCNT2 deletion underlies congenital cataracts and adult i blood group. Hum Genet 131(2):209-16.
Bosio S, Campanella A, Gramaglia E, Porporato P, Longo F, Cremonesi L, Levi S, Camaschella C. 2004. C29G in the iron-responsive element of L-ferritin: a new mutation associated with hyperferritinemia-cataract. Blood Cells Mol Dis 33(1):31-4.
Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. 2013. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet 14(10):681-91.
Boyle DL, Takemoto L. 2000. A possible role for alpha-crystallins in lens epithelial cell differentiation. Mol Vis 6:63-71.
Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW. 2013. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. PLoS Comput Biol 9(4):e1003031.
Bremond-Gignac D, Bitoun P, Reis LM, Copin H, Murray JC, Semina EV. 2010. Identification of dominant FOXE3 and PAX6 mutations in patients with congenital cataract and aniridia. Mol Vis 16:1705-11.
Brian G, Taylor H. 2001. Cataract blindness--challenges for the 21st century. Bull World Health Organ 79(3):249-56.

Brooks DG, Manova-Todorova K, Farmer J, Lobmayr L, Wilson RB, Eagle RC, Jr., St Pierre TG, Stambolian D. 2002. Ferritin crystal cataracts in hereditary hyperferritinemia cataract syndrome. Invest Ophthalmol Vis Sci 43(4):1121-6.

Bu L, Jin Y, Shi Y, Chu R, Ban A, Eiberg H, Andres L, Jiang H, Zheng G, Qian M and others. 2002. Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. Nat Genet 31(3):276-8.

Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT. 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res 41(Web Server issue):W349-57.

Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, Rodnina MV, Komar AA. 2016. Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. Mol Cell 61(3):341-51.

Burdon KP, McKay JD, Sale MM, Russell-Eggitt IM, Mackey DA, Wirth MG, Elder JE, Nicoll A, Clarke MP, FitzGerald LM and others. 2003. Mutations in a novel gene, NHS, cause the pleiotropic effects of Nance-Horan syndrome, including severe congenital cataract, dental anomalies, and mental retardation. Am J Hum Genet 73(5):1120-30.

Burdon KP, Sharma S, Chen CS, Dimasi DP, Mackey DA, Craig JE. 2007. A novel deletion in the FTL gene causes hereditary hyperferritinemia cataract syndrome (HHCS) by alteration of the transcription start site. Hum Mutat 28(7):742.

Burdon KP, Wirth MG, Mackey DA, Russell-Eggitt IM, Craig JE, Elder JE, Dickinson JL, Sale MM. 2004a. Investigation of crystallin genes in familial cataract, and report of two disease associated mutations. Br J Ophthalmol 88(1):79-83.

Burdon KP, Wirth MG, Mackey DA, Russell-Eggitt IM, Craig JE, Elder JE, Dickinson JL, Sale MM. 2004b. A novel mutation in the Connexin 46 gene causes autosomal dominant congenital cataract with incomplete penetrance. J Med Genet 41(8):e106.

Cascella R, Strafella C, Germani C, Novelli G, Ricci F, Zampatti S, Giardina E. 2015. The Genetics and the Genomics of Primary Congenital Glaucoma. Biomed Res Int 2015:321291.

Cazzola M, Bergamaschi G, Tonon L, Arbustini E, Grasso M, Vercesi E, Barosi G, Bianchi PE, Cairo G, Arosio P. 1997. Hereditary hyperferritinemia-cataract syndrome: relationship between phenotypes and specific mutations in the iron-responsive element of ferritin light-chain mRNA. Blood 90(2):814-21.

Chacon-Camacho OF, Buentello-Volante B, Velazquez-Montoya R, Ayala-Ramirez R, Zenteno JC. 2014. Homozygosity mapping identifies a GALK1 mutation as the cause of autosomal recessive congenital cataracts in 4 adult siblings. Gene 534(2):218-21.

Chakrabarti S, Kaur K, Rao KN, Mandal AK, Kaur I, Parikh RS, Thomas R. 2009. The transcription factor gene FOXC1 exhibits a limited role in primary congenital glaucoma. Invest Ophthalmol Vis Sci 50(1):75-83.

Chalam KV, Ambati Bk, Beaver HA, Grover S, M.Levine L, Wells T, K.Isbey E. 2009-2010. Fundamental and Principles of ophthalmology American Academy of Ophthalmology

Chan WH, Biswas S, Ashworth JL, Lloyd IC. 2012. Congenital and infantile cataract: aetiology and management. Eur J Pediatr 171(4):625-30.

Chavarria-Soley G, Michels-Rautenstrauss K, Pasutto F, Flikier D, Flikier P, Cirak S, Bejjani B, Winters DL, Lewis RA, Mardin C and others. 2006. Primary congenital glaucoma and Rieger's anomaly: extended haplotypes reveal founder effects for eight distinct CYP1B1 mutations. Mol Vis 12:523-31.

Chen C, Sun Q, Gu M, Liu K, Sun Y, Xu X. 2015. A novel Cx50 (GJA8) p.H277Y mutation associated with autosomal dominant congenital cataract identified with targeted next-generation sequencing. Graefes Arch Clin Exp Ophthalmol 253(6):915-24.
Chen J, Ma Z, Jiao X, Fariss R, Kantorow WL, Kantorow M, Pras E, Frydman M, Pras E, Riazuddin S and others. 2011. Mutations in FYCO1 cause autosomal-recessive congenital cataracts. Am J Hum Genet 88(6):827-38.
Chen JH, Qiu J, Chen H, Pang CP, Zhang M. 2014. Rapid and cost-effective molecular diagnosis using exome sequencing of one proband with autosomal dominant congenital cataract. Eye (Lond) 28(12):1511-6.
Chen Q, Ma J, Yan M, Mothobi ME, Liu Y, Zheng F. 2009. A novel mutation in CRYAB associated with autosomal dominant congenital nuclear cataract in a Chinese family. Mol Vis 15:1359-65.
Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T and others. 2015. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am J Hum Genet 97(2):199-215.
Churchill A, Graw J. 2011. Clinical and experimental advances in congenital and paediatric cataracts. Philos Trans R Soc Lond B Biol Sci 366(1568):1234-49.
Coccia M, Brooks SP, Webb TR, Christodoulou K, Wozniak IO, Murday V, Balicki M, Yee HA, Wangensteen T, Riise R and others. 2009. X-linked cataract and Nance-Horan syndrome are allelic disorders. Hum Mol Genet 18(14):2643-55.
Cohen D, Bar-Yosef U, Levy J, Gradstein L, Belfair N, Ofir R, Joshua S, Lifshitz T, Carmi R, Birk OS. 2007. Homozygous CRYBB1 deletion mutation underlies autosomal recessive congenital cataract. Invest Ophthalmol Vis Sci 48(5):2208-13.
Colomb E, Kaplan J, Garchon HJ. 2003. Novel cytochrome P450 1B1 (CYP1B1) mutations in patients with primary congenital glaucoma in France. Hum Mutat 22(6):496.
Conley YP, Erturk D, Keverline A, Mah TS, Keravala A, Barnes LR, Bruchis A, Hess JF, FitzGerald PG, Weeks DE and others. 2000. A juvenile-onset, progressive cataract locus on chromosome 3q21-q22 is associated with a missense mutation in the beaded filament structural protein-2. Am J Hum Genet 66(4):1426-31.
Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. 2013. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet 132(10):1077-130.
Cooper MA, Son AI, Komlos D, Sun Y, Kleiman NJ, Zhou R. 2008. Loss of ephrin-A5 function disrupts lens fiber cell packing and leads to cataract. Proc Natl Acad Sci U S A 105(43):16620-5.
Costagliola C, Parmeggiani F, Sebastiani A. 2004. SSRIs and intraocular pressure modifications: evidence, therapeutic implications and possible mechanisms. CNS Drugs 18(8):475-84.
Craig JE, Clark JB, McLeod JL, Kirkland MA, Grant G, Elder JE, Toohey MG, Kowal L, Savoia HF, Chen C and others. 2003. Hereditary hyperferritinemia-cataract syndrome: prevalence, lens morphology, spectrum of mutations, and clinical presentations. Arch Ophthalmol 121(12):1753-61.
Craig JE, Friend KL, Gecz J, Rattray KM, Troski M, Mackey DA, Burdon KP. 2008. A novel locus for X-linked congenital cataract on Xq24. Mol Vis 14:721-6.

Curry SM, Daou AG, Hermanns P, Molinari A, Lewis RA, Bejjani BA. 2004. Cytochrome P4501B1 mutations cause only part of primary congenital glaucoma in Ecuador. Ophthalmic Genet 25(1):3-9.
Danysh BP, Duncan MK. 2009. The lens capsule. Experimental Eye Research 88(2):151-164.
Dave A, Laurie K, Staffieri SE, Taranath D, Mackey DA, Mitchell P, Wang JJ, Craig JE, Burdon KP, Sharma S. 2013. Mutations in the EPHA2 gene are a major contributor to inherited cataracts in South-Eastern Australia. PLoS One 8(8):e72518.
Devi RR, Reena C, Vijayalakshmi P. 2005. Novel mutations in GJA3 associated with autosomal dominant congenital cataract in the Indian population. Mol Vis 11:846-52.
Devi RR, Vijayalakshmi P. 2006. Novel mutations in GJA8 associated with autosomal dominant congenital cataract and microcornea. Mol Vis 12:190-5.
Devi RR, Yao W, Vijayalakshmi P, Sergeev YV, Sundaresan P, Hejtmancik JF. 2008. Crystallin gene mutations in Indian families with inherited pediatric cataract. Mol Vis 14:1157-70.
Dimasi DP, Hewitt AW, Straga T, Pater J, MacKinnon JR, Elder JE, Casey T, Mackey DA, Craig JE. 2007. Prevalence of CYP1B1 mutations in Australian patients with primary congenital glaucoma. Clin Genet 72(3):255-60.
Ding X, Wang B, Luo Y, Hu S, Zhou G, Zhou Z, Wang J, Ma X, Qi Y. 2011. A novel mutation in the connexin 46 (GJA3) gene associated with congenital cataract in a Chinese pedigree. Mol Vis 17:1343-9.
Ding X, Zhou N, Lin H, Chen J, Zhao C, Zhou G, Hejtmancik JF, Qi Y. 2014. A novel MIP gene mutation analysis in a Chinese family affected with congenital progressive punctate cataract. PLoS One 9(7):e102733.
Donaldson PJ, Musil LS, Mathias RT. 2010. Point: A critical appraisal of the lens circulation model--an experimental paradigm for understanding the maintenance of lens transparency? Invest Ophthalmol Vis Sci 51(5):2303-6.
Dorfman R, Nalpathamkalam T, Taylor C, Gonska T, Keenan K, Yuan XW, Corey M, Tsui LC, Zielenski J, Durie P. 2010. Do common in silico tools predict the clinical consequences of amino-acid substitutions in the CFTR gene? Clin Genet 77(5):464-73.
Doucette L, Green J, Fernandez B, Johnson GJ, Parfrey P, Young TL. 2011. A novel, non-stop mutation in FOXE3 causes an autosomal dominant form of variable anterior segment dysgenesis including Peters anomaly. Eur J Hum Genet 19(3):293-9.
Eiberg H, Lund AM, Warburg M, Rosenberg T. 1995. Assignment of congenital cataract Volkmann type (CCV) to chromosome 1p36. Hum Genet 96(1):33-8.
Erie JC, Brue SM, Chamberlain AM, Hodge DO. 2014. Selective serotonin reuptake inhibitor use and increased risk of cataract surgery: a population-based, case-control study. Am J Ophthalmol 158(1):192-197 e1.
Evers C, Paramasivam N, Hinderhofer K, Fischer C, Granzow M, Schmidt-Bacher A, Eils R, Steinbeisser H, Schlesner M, Moog U. 2015. SIPA1L3 identified by linkage analysis and whole-exome sequencing as a novel gene for autosomal recessive congenital cataract. Eur J Hum Genet 23(12):1627-33.
Faletra F, d'Adamo AP, Pensiero S, Athanasakis E, Catalano D, Bruno I, Gasparini P. 2013. A novel CRYBB2 missense mutation causing congenital autosomal dominant cataract in an Italian family. Ophthalmic Genet 34(1-2):115-7.

Farmer LD, Ng SK, Rudkin A, Craig J, Wangmo D, Tsang H, Southisombath K, Griffiths A, Muecke J. 2015. Causes of Severe Visual Impairment and Blindness: Comparative Data From Bhutanese and Laotian Schools for the Blind. Asia Pac J Ophthalmol (Phila) 4(6):350-6.

Ferda Percin E, Ploder LA, Yu JJ, Arici K, Horsford DJ, Rutherford A, Bapat B, Cox DW, Duncan AM, Kalnins VI and others. 2000. Human microphthalmia associated with mutations in the retinal homeobox gene CHX10. Nat Genet 25(4):397-401.

Fine BA, Lubinsky M. 1983. Craniofacial and CNS anomalies with body asymmetry, severe retardation, and other malformations. J Clin Dysmorphol 1(4):6-9.

Fingert JH, Robin AL, Stone JL, Roos BR, Davis LK, Scheetz TE, Bennett SR, Wassink TH, Kwon YH, Alward WL and others. 2011. Copy number variations on chromosome 12q14 in patients with normal tension glaucoma. Hum Mol Genet 20(12):2482-94.

Firasat S, Riazuddin SA, Hejtmancik JF, Riazuddin S. 2008. Primary congenital glaucoma localizes to chromosome 14q24.2-24.3 in two consanguineous Pakistani families. Mol Vis 14:1659-65.

Flanagan SE, Patch AM, Ellard S. 2010. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet Test Mol Biomarkers 14(4):533-7.

Forshew T, Johnson CA, Khaliq S, Pasha S, Willis C, Abbasi R, Tee L, Smith U, Trembath RC, Mehdi SQ and others. 2005. Locus heterogeneity in autosomal recessive congenital cataracts: linkage to 9q and germline HSF4 mutations. Hum Genet 117(5):452-9.

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(2):105-9.

Francis PJ, Berry V, Moore AT, Bhattacharya S. 1999. Lens biology: development and human cataractogenesis. Trends Genet 15(5):191-6.

Fuentes Fajardo KV, Adams D, Mason CE, Sincan M, Tifft C, Toro C, Boerkoel CF, Gahl W, Markello T. 2012. Detecting false-positive signals in exome sequencing. Hum Mutat 33(4):609-13.

Gao X, Cheng J, Lu C, Li X, Li F, Liu C, Zhang M, Zhu S, Ma X. 2010. A novel mutation in the connexin 50 gene (GJA8) associated with autosomal dominant congenital nuclear cataract in a Chinese family. Curr Eye Res 35(7):597-604.

Gao Z, Muecke J, Edussuriya K, Dayawansa R, Hammerton M, Kong A, Sennanayake S, Senaratne T, Marasinghe N, Selva D. 2011. A survey of severe visual impairment and blindness in children attending thirteen schools for the blind in sri lanka. Ophthalmic Epidemiol 18(1):36-43.

Garber I, Pudek M. 2014. A novel deletion in the iron-response element of the L-ferritin gene, causing hyperferritinaemia cataract syndrome. Ann Clin Biochem 51(Pt 6):710-3.

Gencik A. 1989. Epidemiology and genetics of primary congenital glaucoma in Slovakia. Description of a form of primary congenital glaucoma in gypsies with autosomal-recessive inheritance and complete penetrance. Dev Ophthalmol 16:76-115.

Geyer DD, Spence MA, Johannes M, Flodman P, Clancy KP, Berry R, Sparkes RS, Jonsen MD, Isenberg SJ, Bateman JB. 2006. Novel single-base deletional mutation in major intrinsic protein (MIP) in autosomal dominant cataract. Am J Ophthalmol 141(4):761-3.

Gilbert C, Foster A. 2001. Blindness in children: control priorities and research opportunities. Br J Ophthalmol 85(9):1025-7.
Gill D, Klose R, Munier FL, McFadden M, Priston M, Billingsley G, Ducrey N, Schorderet DF, Heon E. 2000. Genetic heterogeneity of the Coppock-like cataract: a mutation in CRYBB2 on chromosome 22q11.2. Invest Ophthalmol Vis Sci 41(1):159-65.
Gillespie RL, O'Sullivan J, Ashworth J, Bhaskar S, Williams S, Biswas S, Kehdi E, Ramsden SC, Clayton-Smith J, Black GC and others. 2014. Personalized diagnosis and management of congenital cataract by next-generation sequencing. Ophthalmology 121(11):2124-37 e1-2.
Girelli D, Corrocher R, Bisceglia L, Olivieri O, Zelante L, Panozzo G, Gasparini P. 1997. Hereditary hyperferritinemia-cataract syndrome caused by a 29-base pair deletion in the iron responsive element of ferritin L-subunit gene. Blood 90(5):2084-8.
Glaser T, Jepeal L, Edwards JG, Young SR, Favor J, Maas RL. 1994. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. Nat Genet 7(4):463-71.
Gonzalez-Huerta LM, Messina-Baas OM, Cuevas-Covarrubias SA. 2007. A family with autosomal dominant primary congenital cataract associated with a CRYGC mutation: evidence of clinical heterogeneity. Mol Vis 13:1333-8.
Graw J. 2009a. Genetics of crystallins: cataract and beyond. Exp Eye Res 88(2):173-89.
Graw J. 2009b. Mouse models of cataract. J Genet 88(4):469-86.
Graw J, Klopp N, Illig T, Preising MN, Lorenz B. 2006. Congenital cataract and macular hypoplasia in humans associated with a de novo mutation in CRYAA and compound heterozygous mutations in P. Graefes Arch Clin Exp Ophthalmol 244(8):912-9.
Graw J, Schmidt W, Minogue PJ, Rodriguez J, Tong JJ, Klopp N, Illig T, Ebihara L, Berthoud VM, Beyer EC. 2009. The GJA8 allele encoding CX50I247M is a rare polymorphism, not a cataract-causing mutation. Mol Vis 15:1881-5.
Gripp KW, Nicholson L, Scott CI, Jr. 1996. Apparently new syndrome of congenital cataracts, sensorineural deafness, Down syndrome-like facial appearance, short stature, and mental retardation. Am J Med Genet 61(4):382-6.
Gu F, Li R, Ma XX, Shi LS, Huang SZ, Ma X. 2006. A missense mutation in the gammaD-crystallin gene CRYGD associated with autosomal dominant congenital cataract in a Chinese family. Mol Vis 12:26-31.
Gu F, Zhai H, Li D, Zhao L, Li C, Huang S, Ma X. 2007. A novel mutation in major intrinsic protein of the lens gene (MIP) underlies autosomal dominant cataract in a Chinese family. Mol Vis 13:1651-6.
Guleria K, Sperling K, Singh D, Varon R, Singh JR, Vanita V. 2007a. A novel mutation in the connexin 46 (GJA3) gene associated with autosomal dominant congenital cataract in an Indian family. Mol Vis 13:1657-65.
Guleria K, Vanita V, Singh D, Singh JR. 2007b. A novel "pearl box" cataract associated with a mutation in the connexin 46 (GJA3) gene. Mol Vis 13:797-803.
Guo Y, Su D, Li Q, Yang Z, Ma Z, Ma X, Zhu S. 2012. A nonsense mutation of CRYGC associated with autosomal dominant congenital nuclear cataracts and microcornea in a Chinese pedigree. Mol Vis 18:1874-80.
Gupta N, Yucel YH. 2007. Glaucoma as a neurodegenerative disease. Curr Opin Ophthalmol 18(2):110-4.
Hansen L, Eiberg H, Rosenberg T. 2007a. Novel MAF mutation in a family with congenital cataract-microcornea syndrome. Mol Vis 13:2019-22.

Hansen L, Mikkelsen A, Nurnberg P, Nurnberg G, Anjum I, Eiberg H, Rosenberg T. 2009. Comprehensive mutational screening in a cohort of Danish families with hereditary congenital cataract. Invest Ophthalmol Vis Sci 50(7):3291-303.
Hansen L, Yao W, Eiberg H, Funding M, Riise R, Kjaer KW, Hejtmancik JF, Rosenberg T. 2006. The congenital "ant-egg" cataract phenotype is caused by a missense mutation in connexin46. Mol Vis 12:1033-9.
Hansen L, Yao W, Eiberg H, Kjaer KW, Baggesen K, Hejtmancik JF, Rosenberg T. 2007b. Genetic heterogeneity in microcornea-cataract: five novel mutations in CRYAA, CRYGD, and GJA8. Invest Ophthalmol Vis Sci 48(9):3937-44.
Harding J. 1991. Cataract Biochemistry, Epidemiology and Pharmacology Chapman & Hall.
He H, Liyanarachchi S, Akagi K, Nagy R, Li J, Dietrich RC, Li W, Sebastian N, Wen B, Xin B and others. 2011a. Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. Science 332(6026):238-40.
He W, Li X, Chen J, Xu L, Zhang F, Dai Q, Cui H, Wang DM, Yu J, Hu S and others. 2011b. Genetic linkage analyses and Cx50 mutation detection in a large multiplex Chinese family with hereditary nuclear cataract. Ophthalmic Genet 32(1):48-53.
Hejtmancik JF. 2008. Congenital cataracts and their molecular genetics. Semin Cell Dev Biol 19(2):134-49.
Hejtmancik JF. 2015. Ophthalmology: Cataracts dissolved. Nature 523(7562):540-1.
Heon 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(5):1261-7.
Hershberg R, Petrov DA. 2008. Selection on codon bias. Annu Rev Genet 42:287-99.
Hollander DA, Sarfarazi M, Stoilov I, Wood IS, Fredrick DR, Alvarado JA. 2006. Genotype and phenotype correlations in congenital glaucoma: CYP1B1 mutations, goniodysgenesis, and clinical characteristics. Am J Ophthalmol 142(6):993-1004.
Hong N, Chen YH, Xie C, Xu BS, Huang H, Li X, Yang YQ, Huang YP, Deng JL, Qi M and others. 2014. Identification of a novel mutation in a Chinese family with Nance-Horan syndrome by whole exome sequencing. J Zhejiang Univ Sci B 15(8):727-34.
Hu S, Wang B, Zhou Z, Zhou G, Wang J, Ma X, Qi Y. 2010. A novel mutation in GJA8 causing congenital cataract-microcornea syndrome in a Chinese pedigree. Mol Vis 16:1585-92.
Huang B, He W. 2010. Molecular characteristics of inherited congenital cataracts. Eur J Med Genet 53(6):347-57.
Huang KM, Wu J, Brooks SP, Hardcastle AJ, Lewis RA, Stambolian D. 2007. Identification of three novel NHS mutations in families with Nance-Horan syndrome. Mol Vis 13:470-4.
Hughes AE, Bradley DT, Campbell M, Lechner J, Dash DP, Simpson DA, Willoughby CE. 2011. Mutation altering the miR-184 seed region causes familial keratoconus with cataract. Am J Hum Genet 89(5):628-33.
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 (Lond) 12 (Pt 2):224-6.
Iseri SU, Osborne RJ, Farrall M, Wyatt AW, Mirza G, Nurnberg G, Kluck C, Herbert H, Martin A, Hussain MS and others. 2009. Seeing clearly: the dominant and

recessive nature of FOXE3 in eye developmental anomalies. Hum Mutat 30(10):1378-86.
Itoh MT, Takahashi N, Abe M, Shimizu K. 2007. Expression and cellular localization of melatonin-synthesizing enzymes in the rat lens. J Pineal Res 42(1):92-6.
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(4):1432-6.
Jamieson RV, Farrar N, Stewart K, Perveen R, Mihelec M, Carette M, Grigg JR, McAvoy JW, Lovicu FJ, Tam PP and others. 2007. Characterization of a familial t(16;22) balanced translocation associated with congenital cataract leads to identification of a novel gene, TMEM114, expressed in the lens and disrupted by the translocation. Hum Mutat 28(10):968-77.
Jamieson RV, Perveen R, Kerr B, Carette M, Yardley J, Heon E, Wirth MG, van Heyningen V, Donnai D, Munier F and others. 2002. Domain disruption and mutation of the bZIP transcription factor, MAF, associated with cataract, ocular anterior segment dysgenesis and coloboma. Hum Mol Genet 11(1):33-42.
Jamuar SS, Tan EC. 2015. Clinical application of next-generation sequencing for Mendelian diseases. Hum Genomics 9:10.
Javadiyan S, Craig JE, Souzeau E, Sharma S, Lower KM, Pater J, Casey T, Hodson T, Burdon KP. 2016. Recurrent mutation in the crystallin alpha A gene associated with inherited paediatric cataract. BMC Res Notes 9:83.
Jiang H, Jin Y, Bu L, Zhang W, Liu J, Cui B, Kong X, Hu L. 2003. A novel mutation in GJA3 (connexin46) for autosomal dominant congenital nuclear pulverulent cataract. Mol Vis 9:579-83.
Jiang J, Jin C, Wang W, Tang X, Shentu X, Wu R, Wang Y, Xia K, Yao K. 2009. Identification of a novel splice-site mutation in MIP in a Chinese congenital cataract family. Mol Vis 15:38-44.
Kajiwara K, Berson EL, Dryja TP. 1994. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. Science 264(5165):1604-8.
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.
Karali A, Russell P, Stefani FH, Tamm ER. 2000. Localization of myocilin/trabecular meshwork--inducible glucocorticoid response protein in the human eye. Invest Ophthalmol Vis Sci 41(3):729-40.
Kataoka K. 2007. Multiple mechanisms and functions of maf transcription factors in the regulation of tissue-specific genes. J Biochem 141(6):775-81.
Katsnelson A. 2011. Breaking the silence. Nat Med 17(12):1536-8.
Kaul H, Riazuddin SA, Shahid M, Kousar S, Butt NH, Zafar AU, Khan SN, Husnain T, Akram J, Hejtmancik JF and others. 2010a. Autosomal recessive congenital cataract linked to EPHA2 in a consanguineous Pakistani family. Mol Vis 16:511-7.
Kaul H, Riazuddin SA, Yasmeen A, Mohsin S, Khan M, Nasir IA, Khan SN, Husnain T, Akram J, Hejtmancik JF and others. 2010b. A new locus for autosomal recessive congenital cataract identified in a Pakistani family. Mol Vis 16:240-5.
Kaur J, Kukreja S, Kaur A, Malhotra N, Kaur R. 2012. The oxidative stress in cataract patients. J Clin Diagn Res 6(10):1629-32.

Kaur K, Reddy AB, Mukhopadhyay A, Mandal AK, Hasnain SE, Ray K, Thomas R, Balasubramanian D, Chakrabarti S. 2005. Myocilin gene implicated in primary congenital glaucoma. Clin Genet 67(4):335-40.
Ke T, Wang QK, Ji B, Wang X, Liu P, Zhang X, Tang Z, Ren X, Liu M. 2006. Novel HSF4 mutation causes congenital total white cataract in a Chinese family. Am J Ophthalmol 142(2):298-303.
Khaliq S, Hameed A, Ismail M, Anwar K, Mehdi SQ. 2002. A novel locus for autosomal dominant nuclear cataract mapped to chromosome 2p12 in a Pakistani family. Invest Ophthalmol Vis Sci 43(7):2083-7.
Khan AO. 2011. Genetics of primary glaucoma. Curr Opin Ophthalmol 22(5):347-55.
Khan AO, Aldahmesh MA, Meyer B. 2007. Recessive congenital total cataract with microcornea and heterozygote carrier signs caused by a novel missense CRYAA mutation (R54C). Am J Ophthalmol 144(6):949-952.
Khan K, Rudkin A, Parry DA, Burdon KP, McKibbin M, Logan CV, Abdelhamed ZI, Muecke JS, Fernandez-Fuentes N, Laurie KJ and others. 2011. Homozygous mutations in PXDN cause congenital cataract, corneal opacity, and developmental glaucoma. Am J Hum Genet 89(3):464-73.
Kizhatil K, Ryan M, Marchant JK, Henrich S, John SW. 2014. Schlemm's canal is a unique vessel with a combination of blood vascular and lymphatic phenotypes that forms by a novel developmental process. PLoS Biol 12(7):e1001912.
Kobayashi K, Kobayashi H, Ueda M, Honda Y. 1998. Estrogen receptor expression in bovine and rat retinas. Invest Ophthalmol Vis Sci 39(11):2105-10.
Kondo Y, Saitsu H, Miyamoto T, Lee BJ, Nishiyama K, Nakashima M, Tsurusaki Y, Doi H, Miyake N, Kim JH and others. 2013. Pathogenic mutations in two families with congenital cataract identified with whole-exome sequencing. Mol Vis 19:384-9.
Kong XD, Liu N, Shi HR, Dong JM, Zhao ZH, Liu J, Li-Ling J, Yang YX. 2015. A novel 3-base pair deletion of the CRYAA gene identified in a large Chinese pedigree featuring autosomal dominant congenital perinuclear cataract. Genet Mol Res 14(1):426-32.
Kronengold J, Trexler EB, Bukauskas FF, Bargiello TA, Verselis VK. 2003. Single-channel SCAM identifies pore-lining residues in the first extracellular loop and first transmembrane domains of Cx46 hemichannels. J Gen Physiol 122(4):389-405.
Kumar M, Agarwal T, Khokhar S, Kumar M, Kaur P, Roy TS, Dada R. 2011. Mutation screening and genotype phenotype correlation of alpha-crystallin, gamma-crystallin and GJA8 gene in congenital cataract. Mol Vis 17:693-707.
Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4(7):1073-81.
Kumar R, Ha T, Pham D, Shaw M, Mangelsdorf M, Friend KL, Hobson L, Turner G, Boyle J, Field M and others. 2016. A non-coding variant in the 5' UTR of DLG3 attenuates protein translation to cause non-syndromic intellectual disability. Eur J Hum Genet.
Kwok JB, Raskin S, Morgan G, Antoniuk SA, Bruk I, Schofield PR. 2001. Mutations in the glycine receptor alpha1 subunit (GLRA1) gene in hereditary hyperekplexia pedigrees: evidence for non-penetrance of mutation Y279C. J Med Genet 38(6):E17.
Kwon YH, Fingert JH, Kuehn MH, Alward WL. 2009. Primary open-angle glaucoma. N Engl J Med 360(11):1113-24.

Lachke SA, Alkuraya FS, Kneeland SC, Ohn T, Aboukhalil A, Howell GR, Saadi I, Cavallesco R, Yue Y, Tsai AC and others. 2011. Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. Science 331(6024):1571-6.
Lachke SA, Higgins AW, Inagaki M, Saadi I, Xi Q, Long M, Quade BJ, Talkowski ME, Gusella JF, Fujimoto A and others. 2012. The cell adhesion gene PVRL3 is associated with congenital ocular defects. Hum Genet 131(2):235-50.
Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227(5259):680-5.
Langle UW, Wolf A, Kammuller ME. 1993. Cataractogenic effects in rats following chronic administration of SDZ ICT 322, a selective 5-HT3 antagonist. Fundam Appl Toxicol 21(3):393-401.
Laurie KJ, Dave A, Straga T, Souzeau E, Chataway T, Sykes MJ, Casey T, Teo T, Pater J, Craig JE and others. 2013. Identification of a novel oligomerization disrupting mutation in CRYAlphaA associated with congenital cataract in a South Australian family. Hum Mutat 34(3):435-8.
Lemmers RJ, Tawil R, Petek LM, Balog J, Block GJ, Santen GW, Amell AM, van der Vliet PJ, Almomani R, Straasheijm KR and others. 2012. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet 44(12):1370-4.
Li A, Li B, Wu L, Yang L, Chen N, Ma Z. 2015. Identification of a novel NHS mutation in a Chinese family with Nance-Horan syndrome. Curr Eye Res 40(4):434-8.
Li B, Liu Y, Liu Y, Guo H, Hu Z, Xia K, Jin X. 2016. Identification of a GJA3 Mutation in a Large Family with Bilateral Congenital Cataract. DNA Cell Biol 35(3):135-9.
Li F, Wang S, Gao C, Liu S, Zhao B, Zhang M, Huang S, Zhu S, Ma X. 2008. Mutation G61C in the CRYGD gene causing autosomal dominant congenital coralliform cataracts. Mol Vis 14:378-86.
Li FF, Yang M, Ma X, Zhang Q, Zhang M, Wang SZ, Zhu SQ. 2010. Autosomal dominant congenital nuclear cataracts caused by a CRYAA gene mutation. Curr Eye Res 35(6):492-8.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078-9.
Li N, Yang Y, Bu J, Zhao C, Lu S, Zhao J, Yan L, Cui L, Zheng R, Li J and others. 2006. An autosomal dominant progressive congenital zonular nuclear cataract linked to chromosome 20p12.2-p11.23. Mol Vis 12:1506-10.
Li N, Zhou Y, Du L, Wei M, Chen X. 2011. Overview of Cytochrome P450 1B1 gene mutations in patients with primary congenital glaucoma. Exp Eye Res 93(5):572-9.
Li X, Buckton AJ, Wilkinson SL, John S, Walsh R, Novotny T, Valaskova I, Gupta M, Game L, Barton PJ and others. 2013. Towards clinical molecular diagnosis of inherited cardiac conditions: a comparison of bench-top genome DNA sequencers. PLoS One 8(7):e67744.
Li Y, Wang J, Dong B, Man H. 2004. A novel connexin46 (GJA3) mutation in autosomal dominant congenital nuclear pulverulent cataract. Mol Vis 10:668-71.
Liao HM, Niu DM, Chen YJ, Fang JS, Chen SJ, Chen CH. 2011. Identification of a microdeletion at Xp22.13 in a Taiwanese family presenting with Nance-Horan syndrome. J Hum Genet 56(1):8-11.

Lim SH, Tran-Viet KN, Yanovitch TL, Freedman SF, Klemm T, Call W, Powell C, Ravichandran A, Metlapally R, Nading EB and others. 2013. CYP1B1, MYOC, and LTBP2 mutations in primary congenital glaucoma patients in the United States. Am J Ophthalmol 155(3):508-517 e5.
Lin H, Hejtmancik JF, Qi Y. 2007. A substitution of arginine to lysine at the COOH-terminus of MIP caused a different binocular phenotype in a congenital cataract family. Mol Vis 13:1822-7.
Lin Q, Zhou N, Zhang N, Qi Y. 2014. Mutational screening of EFNA5 in Chinese age-related cataract patients. Ophthalmic Res 52(3):124-9.
Lin Y, Liu NN, Lei CT, Fan YC, Liu XQ, Yang Y, Wang JF, Liu B, Yang ZL. 2008. [A novel GJA8 mutation in a Chinese family with autosomal dominant congenital cataract]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 25(1):59-62.
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(5):665-8.
Litt M, Kramer P, LaMorticella DM, Murphey W, Lovrien EW, Weleber RG. 1998. Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. Hum Mol Genet 7(3):471-4.
Liu BF, Liang JJ. 2005. Interaction and biophysical properties of human lens Q155* betaB2-crystallin mutant. Mol Vis 11:321-7.
Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012:251364.
Liu M, Ke T, Wang Z, Yang Q, Chang W, Jiang F, Tang Z, Li H, Ren X, Wang X and others. 2006a. Identification of a CRYAB mutation associated with autosomal dominant posterior polar cataract in a Chinese family. Invest Ophthalmol Vis Sci 47(8):3461-6.
Liu Q, Wang KJ, Zhu SQ. 2014. A novel p.G112E mutation in BFSP2 associated with autosomal dominant pulverulent cataract with sutural opacities. Curr Eye Res 39(10):1013-9.
Liu Y, Allingham RR. 2011. Molecular genetics in glaucoma. Exp Eye Res 93(4):331-9.
Liu Y, Zhang X, Luo L, Wu M, Zeng R, Cheng G, Hu B, Liu B, Liang JJ, Shang F. 2006b. A novel alphaB-crystallin mutation associated with autosomal dominant congenital lamellar cataract. Invest Ophthalmol Vis Sci 47(3):1069-75.
Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell. J. 2000. Molecular Cell Biology. 4th edition ed: W. H. Freeman.
Loewen CJ, Moritz OL, Molday RS. 2001. Molecular characterization of peripherin-2 and rom-1 mutants responsible for digenic retinitis pigmentosa. J Biol Chem 276(25):22388-96.
Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30(5):434-9.
Lorenz B. 2007. Genetische Untersuchungen bei kongenitaler Katarakt. Der Ophthalmologe 104(7):559-565.
Lou D, Tong JP, Zhang LY, Chiang SW, Lam DS, Pang CP. 2009. A novel mutation in CRYBB2 responsible for inherited coronary cataract. Eye (Lond) 23(5):1213-20.

Lv H, Huang C, Zhang J, Liu Z, Zhang Z, Xu H, You Y, Hu J, Li X, Wang W. 2014. A novel HSF4 gene mutation causes autosomal-dominant cataracts in a Chinese family. G3 (Bethesda) 4(5):823-8.
Ma AS, Grigg JR, Ho G, Prokudin I, Farnsworth E, Holman K, Cheng A, Billson FA, Martin F, Fraser C and others. 2016. Sporadic and Familial Congenital Cataracts: Mutational Spectrum and New Diagnoses Using Next-Generation Sequencing. Hum Mutat 37(4):371-84.
Ma X, Li FF, Wang SZ, Gao C, Zhang M, Zhu SQ. 2008. A new mutation in BFSP2 (G1091A) causes autosomal dominant congenital lamellar cataracts. Mol Vis 14:1906-11.
Ma ZW, Zheng JQ, Li J, Li XR, Tang X, Yuan XY, Zhang XM, Sun HM. 2005. Two novel mutations of connexin genes in Chinese families with autosomal dominant congenital nuclear cataract. Br J Ophthalmol 89(11):1535-7.
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(5):1357-64.
Mackay DS, Andley UP, Shiels A. 2003. Cell death triggered by a novel mutation in the alphaA-crystallin gene underlies autosomal dominant cataract linked to chromosome 21q. Eur J Hum Genet 11(10):784-93.
MacKinnon JR, Giubilato A, Elder JE, Craig JE, Mackey DA. 2004. Primary infantile glaucoma in an Australian population. Clin Experiment Ophthalmol 32(1):14-8.
Maeda YY, Funata N, Takahama S, Sugata Y, Yonekawa H. 2001. Two interactive genes responsible for a new inherited cataract (RCT) in the mouse. Mamm Genome 12(4):278-83.
Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. 2011. What can exome sequencing do for you? J Med Genet 48(9):580-9.
Makley LN, McMenimen KA, DeVree BT, Goldman JW, McGlasson BN, Rajagopal P, Dunyak BM, McQuade TJ, Thompson AD, Sunahara R and others. 2015. Pharmacological chaperone for alpha-crystallin partially restores transparency in cataract models. Science 350(6261):674-7.
Makrythanasis P, Antonarakis SE. 2013. Pathogenic variants in non-protein-coding sequences. Clin Genet 84(5):422-8.
Mashima Y, Suzuki Y, Sergeev Y, Ohtake Y, Tanino T, Kimura I, Miyata H, Aihara M, Tanihara H, Inatani M and others. 2001. Novel cytochrome P4501B1 (CYP1B1) gene mutations in Japanese patients with primary congenital glaucoma. Invest Ophthalmol Vis Sci 42(10):2211-6.
McAvoy JW, Chamberlain CG, de Longh RU, Hales AM, Lovicu FJ. 1999. Lens development. Eye 13(3b):425-437.
McGinnity FG, Page AB, Bryars JH. 1987. Primary congenital glaucoma: twenty years experience. Ir J Med Sci 156(12):364-5.
McKay JD, Patterson B, Craig JE, Russell-Eggitt IM, Wirth MG, Burdon KP, Hewitt AW, Cohn AC, Kerdraon Y, Mackey DA. 2005. The telomere of human chromosome 1p contains at least two independent autosomal dominant congenital cataract genes. Br J Ophthalmol 89(7):831-4.
McLeod JL, Craig J, Gumley S, Roberts S, Kirkland MA. 2002. Mutation spectrum in Australian pedigrees with hereditary hyperferritinaemia-cataract syndrome reveals novel and de novo mutations. Br J Haematol 118(4):1179-82.
Medina-Trillo C, Aroca-Aguilar JD, Mendez-Hernandez CD, Morales L, Garcia-Anton M, Garcia-Feijoo J, Escribano J. 2015. Rare FOXC1 variants in congenital glaucoma: identification of translation regulatory sequences. Eur J Hum Genet.

Mehra S, Kapur S, Vasavada AR. 2011. Polymorphisms of the gamma crystallin A and B genes among Indian patients with pediatric cataract. J Postgrad Med 57(3):201-5.
Melki R, Colomb E, Lefort N, Brezin AP, Garchon HJ. 2004. CYP1B1 mutations in French patients with early-onset primary open-angle glaucoma. J Med Genet 41(9):647-51.
Merriman B, Rothberg JM. 2012. Progress in ion torrent semiconductor chip based sequencing. Electrophoresis 33(23):3397-417.
Messina-Baas OM, Gonzalez-Huerta LM, Chima-Galan C, Kofman-Alfaro SH, Rivera-Vega MR, Babayan-Mena I, Cuevas-Covarrubias SA. 2007. Molecular analysis of the CYP1B1 gene: identification of novel truncating mutations in patients with primary congenital glaucoma. Ophthalmic Res 39(1):17-23.
Messina-Baas OM, Gonzalez-Huerta LM, Cuevas-Covarrubias SA. 2006. Two affected siblings with nuclear cataract associated with a novel missense mutation in the CRYGD gene. Mol Vis 12:995-1000.
Meyer E, Rahman F, Owens J, Pasha S, Morgan NV, Trembath RC, Stone EM, Moore AT, Maher ER. 2009. Initiation codon mutation in betaB1-crystallin (CRYBB1) associated with autosomal recessive nuclear pulverulent cataract. Mol Vis 15:1014-9.
Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, Prathalingham R, Dennis J, Bolla MK, Wang Q and others. 2013. Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. Am J Hum Genet 93(6):1046-60.
Michels-Rautenstrauss KG, Mardin CY, Zenker M, Jordan N, Gusek-Schneider GC, Rautenstrauss BW. 2001. Primary congenital glaucoma: three case reports on novel mutations and combinations of mutations in the GLC3A (CYP1B1) gene. J Glaucoma 10(4):354-7.
Millat G, Chanavat V, Rousson R. 2014a. Evaluation of a new high-throughput next-generation sequencing method based on a custom AmpliSeq library and ion torrent PGM sequencing for the rapid detection of genetic variations in long QT syndrome. Mol Diagn Ther 18(5):533-9.
Millat G, Chanavat V, Rousson R. 2014b. Evaluation of a new NGS method based on a custom AmpliSeq library and Ion Torrent PGM sequencing for the fast detection of genetic variations in cardiomyopathies. Clin Chim Acta 433:266-71.
Minogue PJ, Tong JJ, Arora A, Russell-Eggitt I, Hunt DM, Moore AT, Ebihara L, Beyer EC, Berthoud VM. 2009. A mutant connexin50 with enhanced hemichannel function leads to cell death. Invest Ophthalmol Vis Sci 50(12):5837-45.
Mohammed H, D'Santos C, Serandour AA, Ali HR, Brown GD, Atkins A, Rueda OM, Holmes KA, Theodorou V, Robinson JL and others. 2013. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. Cell Rep 3(2):342-9.
More MI, Kirsch FP, Rathjen FG. 2001. Targeted ablation of NrCAM or ankyrin-B results in disorganized lens fibers leading to cataract formation. J Cell Biol 154(1):187-96.
Mothobi ME, Guo S, Liu Y, Chen Q, Yussuf AS, Zhu X, Fang Z. 2009. Mutation analysis of congenital cataract in a Basotho family identified a new missense allele in CRYBB2. Mol Vis 15:1470-5.

Muller M, Bhattacharya SS, Moore T, Prescott Q, Wedig T, Herrmann H, Magin TM. 2009. Dominant cataract formation in association with a vimentin assembly disrupting mutation. Hum Mol Genet 18(6):1052-7.
Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, Makarov SS, Maixner W, Diatchenko L. 2006. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science 314(5807):1930-3.
Narooie-Nejad M, Paylakhi SH, Shojaee S, Fazlali Z, Rezaei Kanavi M, Nilforushan N, Yazdani S, Babrzadeh F, Suri F, Ronaghi M and others. 2009. Loss of function mutations in the gene encoding latent transforming growth factor beta binding protein 2, LTBP2, cause primary congenital glaucoma. Hum Mol Genet 18(20):3969-77.
Narumi Y, Nishina S, Tokimitsu M, Aoki Y, Kosaki R, Wakui K, Azuma N, Murata T, Takada F, Fukushima Y and others. 2014. Identification of a novel missense mutation of MAF in a Japanese family with congenital cataract by whole exome sequencing: a clinical report and review of literature. Am J Med Genet A 164A(5):1272-6.
Niceta M, Stellacci E, Gripp KW, Zampino G, Kousi M, Anselmi M, Traversa A, Ciolfi A, Stabley D, Bruselles A and others. 2015. Mutations Impairing GSK3-Mediated MAF Phosphorylation Cause Cataract, Deafness, Intellectual Disability, Seizures, and a Down Syndrome-like Facies. Am J Hum Genet 96(5):816-25.
Nishiguchi S, Wood H, Kondoh H, Lovell-Badge R, Episkopou V. 1998. Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. Genes Dev 12(6):776-81.
Nishio SY, Usami SI, Watanabe M, Hayashi Y. 2015. Clinical Application of a Custom AmpliSeq Library and Ion Torrent PGM Sequencing to Comprehensive Mutation Screening for Deafness Genes. Genet Test Mol Biomarkers.
Nonnenmacher L, Langer T, Blessing H, Gabriel H, Buchwald HJ, Meneksedag C, Kohne E, Gencik M, Debatin KM, Cario H. 2011. Hereditary hyperferritinemia cataract syndrome: clinical, genetic, and laboratory findings in 5 families. Klin Padiatr 223(6):346-51.
Panicker SG, Reddy AB, Mandal AK, Ahmed N, Nagarajaram HA, Hasnain SE, Balasubramanian D. 2002. Identification of novel mutations causing familial primary congenital glaucoma in Indian pedigrees. Invest Ophthalmol Vis Sci 43(5):1358-66.
Parmley JL, Chamary JV, Hurst LD. 2006. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol Biol Evol 23(2):301-9.
Pasta SY, Raman B, Ramakrishna T, Rao Ch M. 2003. Role of the conserved SRLFDQFFG region of alpha-crystallin, a small heat shock protein. Effect on oligomeric size, subunit exchange, and chaperone-like activity. J Biol Chem 278(51):51159-66.
Pasutto F, Chavarria-Soley G, Mardin CY, Michels-Rautenstrauss K, Ingelman-Sundberg M, Fernandez-Martinez L, Weber BH, Rautenstrauss B, Reis A. 2010. Heterozygous loss-of-function variants in CYP1B1 predispose to primary open-angle glaucoma. Invest Ophthalmol Vis Sci 51(1):249-54.
Pauli S, Soker T, Klopp N, Illig T, Engel W, Graw J. 2007. Mutation analysis in a German family identified a new cataract-causing allele in the CRYBB2 gene. Mol Vis 13:962-7.

Pébay A. 2014. Regenerative Biology of the Eye: New York, NY : Springer New York : Imprint: Springer.
Pfenniger A, Wohlwend A, Kwak BR. 2011. Mutations in connexin genes and disease. Eur J Clin Invest 41(1):103-16.
Plasilova M, Stoilov I, Sarfarazi M, Kadasi L, Ferakova E, Ferak V. 1999. Identification of a single ancestral CYP1B1 mutation in Slovak Gypsies (Roms) affected with primary congenital glaucoma. J Med Genet 36(4):290-4.
Ponnam SP, Ramesha K, Tejwani S, Matalia J, Kannabiran C. 2008. A missense mutation in LIM2 causes autosomal recessive congenital cataract. Mol Vis 14:1204-8.
Ponnam SP, Ramesha K, Tejwani S, Ramamurthy B, Kannabiran C. 2007. Mutation of the gap junction protein alpha 8 (GJA8) gene causes autosomal recessive cataract. J Med Genet 44(7):e85.
Ponnam SP, Ramesha K, Tejwani S, Ramamurthy B, Kannabiran C. 2009. Mutation of the gap junction protein alpha 8 (GJA8) gene causes autosomal recessive cataract. BMJ Case Rep 2009.
Pras E, Frydman M, Levy-Nissenbaum E, Bakhan T, Raz J, Assia EI, Goldman B, Pras E. 2000. A nonsense mutation (W9X) in CRYAA causes autosomal recessive cataract in an inbred Jewish Persian family. Invest Ophthalmol Vis Sci 41(11):3511-5.
Pras E, Levy-Nissenbaum E, Bakhan T, Lahat H, Assia E, Geffen-Carmi N, Frydman M, Goldman B, Pras E. 2002. A missense mutation in the LIM2 gene is associated with autosomal recessive presenile cataract in an inbred Iraqi Jewish family. Am J Hum Genet 70(5):1363-7.
Pras E, Raz J, Yahalom V, Frydman M, Garzozi HJ, Pras E, Hejtmancik JF. 2004. A nonsense mutation in the glucosaminyl (N-acetyl) transferase 2 gene (GCNT2): association with autosomal recessive congenital cataracts. Invest Ophthalmol Vis Sci 45(6):1940-5.
Preda L, Dinca O, Bucur A, Dragomir C, Severin E. 2010. Identical Mutation in SH3BP2 Gene Causes Clinical Phenotypes with Different Severity in Mother and Daughter - Case Report. Mol Syndromol 1(2):87-90.
Preus M, Cooper AR, O'Leary E. 1984. Sensorineural hearing loss, small facial features, submucous cleft palate, and myoclonic seizures. J Clin Dysmorphol 2(1):30-1.
Prokudin I, Simons C, Grigg JR, Storen R, Kumar V, Phua ZY, Smith J, Flaherty M, Davila S, Jamieson RV. 2014. Exome sequencing in developmental eye disease leads to identification of causal variants in GJA8, CRYGC, PAX6 and CYP1B1. Eur J Hum Genet 22(7):907-15.
Qi Y, Jia H, Huang S, Lin H, Gu J, Su H, Zhang T, Gao Y, Qu L, Li D and others. 2004. A deletion mutation in the betaA1/A3 crystallin gene (CRYBA1/A3) is associated with autosomal dominant congenital nuclear cataract in a Chinese family. Hum Genet 114(2):192-7.
Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341.
Quigley HA. 2011. Glaucoma. Lancet 377(9774):1367-77.
Quigley HA, Broman AT. 2006. The number of people with glaucoma worldwide in 2010 and 2020. Br J Ophthalmol 90(3):262-7.

Rahi JS, Dezateux 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(7):1444-8.

Ramachandran RD, Perumalsamy V, Hejtmancik JF. 2007. Autosomal recessive juvenile onset cataract associated with mutation in BFSP1. Hum Genet 121(3-4):475-82.

Reches A, Yaron Y, Burdon K, Crystal-Shalit O, Kidron D, Malcov M, Tepper R. 2007. Prenatal detection of congenital bilateral cataract leading to the diagnosis of Nance-Horan syndrome in the extended family. Prenat Diagn 27(7):662-4.

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(2):206-9.

Reis LM, Tyler RC, Muheisen S, Raggio V, Salviati L, Han DP, Costakos D, Yonath H, Hall S, Power P and others. 2013. Whole exome sequencing in dominant cataract identifies a new causative factor, CRYBA2, and a variety of novel alleles in known genes. Hum Genet 132(7):761-70.

Reis LM, Tyler RC, Semina EV. 2014. Identification of a novel C-terminal extension mutation in EPHA2 in a family affected with congenital cataract. Mol Vis 20:836-42.

Reis LM, Tyler RC, Volkmann Kloss BA, Schilter KF, Levin AV, Lowry RB, Zwijnenburg PJ, Stroh E, Broeckel U, Murray JC and others. 2012. PITX2 and FOXC1 spectrum of mutations in ocular syndromes. Eur J Hum Genet 20(12):1224-33.

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(5):531-7.

Rhode SJ, Ginsberg Sp. 1987. Ophthalmic Technology A guide for the eye care assistant New York: Raven Press.

Riazuddin SA, Yasmeen A, Yao W, Sergeev YV, Zhang Q, Zulfiqar F, Riaz A, Riazuddin S, Hejtmancik JF. 2005. Mutations in betaB3-crystallin associated with autosomal recessive cataract in two Pakistani families. Invest Ophthalmol Vis Sci 46(6):2100-6.

Roshan M, Vijaya PH, Lavanya GR, Shama PK, Santhiya ST, Graw J, Gopinath PM, Satyamoorthy K. 2010. A novel human CRYGD mutation in a juvenile autosomal dominant cataract. Mol Vis 16:887-96.

Safieh LA, Khan AO, Alkuraya FS. 2009. Identification of a novel CRYAB mutation associated with autosomal recessive juvenile cataract in a Saudi family. Mol Vis 15:980-4.

Sajjad N, Goebel I, Kakar N, Cheema AM, Kubisch C, Ahmad J. 2008. A novel HSF4 gene mutation (p.R405X) causing autosomal recessive congenital cataracts in a large consanguineous family from Pakistan. BMC Med Genet 9:99.

Santana A, Waiswo M. 2011. The genetic and molecular basis of congenital cataract. Arquivos Brasileiros de Oftalmologia 74:136-142.

Santana A, Waiswol M, Arcieri ES, Cabral de Vasconcellos JP, Barbosa de Melo M. 2009. Mutation analysis of CRYAA, CRYGC, and CRYGD associated with autosomal dominant congenital cataract in Brazilian families. Mol Vis 15:793-800.

Santhiya ST, Kumar GS, Sudhakar P, Gupta N, Klopp N, Illig T, Soker T, Groth M, Platzer M, Gopinath PM and others. 2010. Molecular analysis of cataract families in India: new mutations in the CRYBB2 and GJA3 genes and rare polymorphisms. Mol Vis 16:1837-47.
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(5):352-8.
Santhiya ST, Soker T, Klopp N, Illig T, Prakash MV, Selvaraj B, Gopinath PM, Graw J. 2006. Identification of a novel, putative cataract-causing allele in CRYAA (G98R) in an Indian family. Mol Vis 12:768-73.
Sarfarazi M, Akarsu AN, Hossain A, Turacli ME, Aktan SG, Barsoum-Homsy M, Chevrette L, Sayli BS. 1995. Assignment of a locus (GLC3A) for primary congenital glaucoma (Buphthalmos) to 2p21 and evidence for genetic heterogeneity. Genomics 30(2):171-7.
Sarfarazi M, Stoilov I. 2000. Molecular genetics of primary congenital glaucoma. Eye (Lond) 14 (Pt 3B):422-8.
Sarfarazi M, Stoilov I, Schenkman JB. 2003. Genetics and biochemistry of primary congenital glaucoma. Ophthalmol Clin North Am 16(4):543-54, vi.
Schulpis KH, Papaconstantinou ED, Koidou A, Michelakakis H, Tzamouranis J, Patsouras A, Shin Y. 1993. UDP galactose-4-epimerase deficiency in a 5.5-year-old girl with unilateral cataract. J Inherit Metab Dis 16(5):903-4.
Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11(4):361-2.
Sebastian E. 2010. The Complexity and Origins of the Human Eye: A Brief Study on the Anatomy, Physiology, and Origin of the Eye: Liberty university.
Seifi M, Footz T, Taylor SA, Elhady GM, Abdalla EM, Walter MA. 2016. Novel PITX2 gene mutations in patients with Axenfeld-Rieger syndrome. Acta Ophthalmol.
Semina EV, Brownell I, Mintz-Hittner HA, Murray JC, Jamrich M. 2001. Mutations in the human forkhead transcription factor FOXE3 associated with anterior segment ocular dysgenesis and cataracts. Hum Mol Genet 10(3):231-6.
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(2):167-70.
Senthil Kumar G, Kyle JW, Minogue PJ, Dinesh Kumar K, Vasantha K, Berthoud VM, Beyer EC, Santhiya ST. 2013. An MIP/AQP0 mutation with impaired trafficking and function underlies an autosomal dominant congenital lamellar cataract. Exp Eye Res 110:136-41.
Sharma S, Burdon KP, Dave A, Jamieson RV, Yaron Y, Billson F, Van Maldergem L, Lorenz B, Gecz J, Craig JE. 2008. Novel causative mutations in patients with Nance-Horan syndrome and altered localization of the mutant NHS-A protein isoform. Mol Vis 14:1856-64.
Sharma S, Koh KS, Collin C, Dave A, McMellon A, Sugiyama Y, McAvoy JW, Voss AK, Gecz J, Craig JE. 2009. NHS-A isoform of the NHS gene is a novel interactor of ZO-1. Exp Cell Res 315(14):2358-72.
Shentu X, Miao Q, Tang X, Yin H, Zhao Y. 2015. Identification and Functional Analysis of a Novel MIP Gene Mutation Associated with Congenital Cataract in a Chinese Family. PLoS One 10(5):e0126679.

Shentu XC, Zhao SJ, Zhang L, Miao Q. 2013. A novel p.R890C mutation in EPHA2 gene associated with progressive childhood posterior cataract in a Chinese family. Int J Ophthalmol 6(1):34-8.
Shiels A, Bennett TM, Hejtmancik JF. 2010. Cat-Map: putting cataract on the map. Mol Vis 16:2007-15.
Shiels A, Bennett TM, Knopf HL, Maraini G, Li A, Jiao X, Hejtmancik JF. 2008. The EPHA2 gene is associated with cataracts linked to chromosome 1p. Mol Vis 14:2042-55.
Shiels A, Bennett TM, Knopf HL, Yamada K, Yoshiura K, Niikawa N, Shim S, Hanson PI. 2007. CHMP4B, a novel gene for autosomal dominant cataracts linked to chromosome 20q. Am J Hum Genet 81(3):596-606.
Shiels A HJ. 2007. GEnetic origins of cataract. Archives of Ophthalmology 125(2):165-173.
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(3):526-32.
Shin YS, Rieth M, Endres W, Haas P. 1984. Sorbitol dehydrogenase deficiency in a family with congenital cataracts. J Inherit Metab Dis 7 Suppl 2:151-2.
Sia DI, Muecke J, Hammerton M, Ngy M, Kong A, Morse A, Holmes M, Piseth H, Hamilton C, Selva D. 2010. A survey of visual impairment and blindness in children attending four schools for the blind in Cambodia. Ophthalmic Epidemiol 17(4):225-33.
Sibon I, Coupry I, Menegon P, Bouchet JP, Gorry P, Burgelin I, Calvas P, Orignac I, Dousset V, Lacombe D and others. 2007. COL4A1 mutation in Axenfeld-Rieger anomaly with leukoencephalopathy and stroke. Ann Neurol 62(2):177-84.
Sitorus R, Ardjo SM, Lorenz B, Preising M. 2003. CYP1B1 gene analysis in primary congenital glaucoma in Indonesian and European patients. J Med Genet 40(1):e9.
Skuta GL, Cantor LB, Weiss JS. 2009-2010. Lens and Catract Singapour: American Academy of Ophthalmology
Smaoui N, Beltaief O, BenHamed S, M'Rad R, Maazoul F, Ouertani A, Chaabouni H, Hejtmancik JF. 2004. A homozygous splice mutation in the HSF4 gene is associated with an autosomal recessive congenital cataract. Invest Ophthalmol Vis Sci 45(8):2716-21.
Smith KR, Bromhead CJ, Hildebrand MS, Shearer AE, Lockhart PJ, Najmabadi H, Leventer RJ, McGillivray G, Amor DJ, Smith RJ and others. 2011. Reducing the exome search space for mendelian diseases using genetic linkage analysis of exome genotypes. Genome Biol 12(9):R85.
Solomon BD, Pineda-Alvarez DE, Balog JZ, Hadley D, Gropman AL, Nandagopal R, Han JC, Hahn JS, Blain D, Brooks B and others. 2009. Compound heterozygosity for mutations in PAX6 in a patient with complex brain anomaly, neonatal diabetes mellitus, and microophthalmia. Am J Med Genet A 149A(11):2543-6.
Song S, Landsbury A, Dahm R, Liu Y, Zhang Q, Quinlan RA. 2009. Functions of the intermediate filament cytoskeleton in the eye lens. J Clin Invest 119(7):1837-48.
Song Z, Wang L, Liu Y, Xiao W. 2015. A novel nonsense mutation in the MIP gene linked to congenital posterior polar cataracts in a Chinese family. PLoS One 10(3):e0119296.

Souma T, Tompson SW, Thomson BR, Siggs OM, Kizhatil K, Yamaguchi S, Feng L, Limviphuvadh V, Whisenhunt KN, Maurer-Stroh S and others. 2016. Angiopoietin receptor TEK mutations underlie primary congenital glaucoma with variable expressivity. J Clin Invest.

Souzeau E, Hayes M, Ruddle JB, Elder JE, Staffieri SE, Kearns LS, Mackey DA, Zhou T, Ridge B, Burdon KP and others. 2015. CYP1B1 copy number variation is not a major contributor to primary congenital glaucoma. Mol Vis 21:160-4.

Stambolian D, Ai Y, Sidjanin D, Nesburn K, Sathe G, Rosenberg M, Bergsma DJ. 1995. Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts. Nat Genet 10(3):307-12.

Stephan DA, Gillanders E, Vanderveen D, Freas-Lutz D, Wistow G, Baxevanis AD, Robbins CM, VanAuken A, Quesenberry MI, Bailey-Wilson J and others. 1999. Progressive juvenile-onset punctate cataracts caused by mutation of the gammaD-crystallin gene. Proc Natl Acad Sci U S A 96(3):1008-12.

Stitziel NO, Kiezun A, Sunyaev S. 2011. Computational and statistical approaches to analyzing variants identified by exome sequencing. Genome Biol 12(9):227.

Stoilov I, Akarsu AN, Alozie I, Child A, Barsoum-Homsy M, Turacli ME, Or M, Lewis RA, Ozdemir N, Brice G and others. 1998. Sequence analysis and homology modeling suggest that primary congenital glaucoma on 2p21 results from mutations disrupting either the hinge region or the conserved core structures of cytochrome P4501B1. Am J Hum Genet 62(3):573-84.

Stoilov I, Akarsu AN, Sarfarazi M. 1997. Identification of three different truncating mutations in cytochrome P4501B1 (CYP1B1) as the principal cause of primary congenital glaucoma (Buphthalmos) in families linked to the GLC3A locus on chromosome 2p21. Hum Mol Genet 6(4):641-7.

Stoilov IR, Costa VP, Vasconcellos JP, Melo MB, Betinjane AJ, Carani JC, Oltrogge EV, Sarfarazi M. 2002. Molecular genetics of primary congenital glaucoma in Brazil. Invest Ophthalmol Vis Sci 43(6):1820-7.

Su D, Guo Y, Li Q, Guan L, Zhu S, Ma X. 2012. A novel mutation in CRYAA is associated with autosomal dominant suture cataracts in a Chinese family. Mol Vis 18:3057-63.

Sun H, Ma Z, Li Y, Liu B, Li Z, Ding X, Gao Y, Ma W, Tang X, Li X and others. 2005. Gamma-S crystallin gene (CRYGS) mutation causes dominant progressive cortical cataract in humans. J Med Genet 42(9):706-10.

Sun W, Xiao X, Li S, Guo X, Zhang Q. 2011a. Mutation analysis of 12 genes in Chinese families with congenital cataracts. Mol Vis 17:2197-206.

Sun W, Xiao X, Li S, Guo X, Zhang Q. 2011b. Mutational screening of six genes in Chinese patients with congenital cataract and microcornea. Mol Vis 17:1508-13.

Sun W, Xiao X, Li S, Guo X, Zhang Q. 2014. Exome sequencing of 18 Chinese families with congenital cataracts: a new sight of the NHS gene. PLoS One 9(6):e100455.

Tanwar M, Kumar M, Dada T, Sihota R, Dada R. 2010. MYOC and FOXC1 gene analysis in primary congenital glaucoma. Mol Vis 16:1996-2006.

Thomson BR, Heinen S, Jeansson M, Ghosh AK, Fatima A, Sung HK, Onay T, Chen H, Yamaguchi S, Economides AN and others. 2014. A lymphatic defect causes ocular hypertension and glaucoma in mice. J Clin Invest 124(10):4320-4.

Traupe H, Has C. 2000. The Conradi-Hunermann-Happle syndrome is caused by mutations in the gene that encodes a 8- 7 sterol isomerase and is biochemically related to the CHILD syndrome. Eur J Dermatol 10(6):425-8.

Tsuchiya M, Misaka R, Nitta K, Tsuchiya K. 2015. Transcriptional factors, Mafs and their biological roles. World J Diabetes 6(1):175-83.
Tug E, Dilek NF, Javadiyan S, Burdon KP, Percin FE. 2013. A Turkish family with Nance-Horan Syndrome due to a novel mutation. Gene 525(1):141-5.
Tuteja G, Kaestner KH. 2007. SnapShot: forkhead transcription factors I. Cell 130(6):1160.
Vaca G, Ibarra B, Bracamontes M, Garcia-Cruz D, Sanchez-Corona J, Medina C, Wunsch C, Gonzalez-Quiroga G, Cantu JM. 1982. Red blood cell sorbitol dehydrogenase deficiency in a family with cataracts. Hum Genet 61(4):338-41.
van der Smagt JJ, Vink A, Kirkels JH, Nelen M, ter Heide H, Molenschot MM, Weger RA, Schellekens PA, Hoogendijk J, Dooijes D. 2014. Congenital posterior pole cataract and adult onset dilating cardiomyopathy: expanding the phenotype of alphaB-crystallinopathies. Clin Genet 85(4):381-5.
van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. Trends Genet 30(9):418-26.
Van Esch H, Jansen A, Bauters M, Froyen G, Fryns JP. 2007. Encephalopathy and bilateral cataract in a boy with an interstitial deletion of Xp22 comprising the CDKL5 and NHS genes. Am J Med Genet A 143(4):364-9.
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(6):392-6.
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(2):509-14.
Vanita V, Hennies HC, Singh D, Nurnberg P, Sperling K, Singh JR. 2006a. A novel mutation in GJA8 associated with autosomal dominant congenital cataract in a family of Indian origin. Mol Vis 12:1217-22.
Vanita V, Singh D, Robinson PN, Sperling K, Singh JR. 2006b. A novel mutation in the DNA-binding domain of MAF at 16q23.1 associated with autosomal dominant "cerulean cataract" in an Indian family. Am J Med Genet A 140(6):558-66.
Vanita V, Singh JR, Singh D, Varon R, Sperling K. 2008. A novel mutation in GJA8 associated with jellyfish-like cataract in a family of Indian origin. Mol Vis 14:323-6.
Varon R, Gooding R, Steglich C, Marns L, Tang H, Angelicheva D, Yong KK, Ambrugger P, Reinhold A, Morar B and others. 2003. Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome. Nat Genet 35(2):185-9.
Vasiliou V, Gonzalez FJ. 2008. Role of CYP1B1 in glaucoma. Annu Rev Pharmacol Toxicol 48:333-58.
Venturini G, Rose AM, Shah AZ, Bhattacharya SS, Rivolta C. 2012. CNOT3 is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance. PLoS Genet 8(11):e1003040.
Verdin H, Sorokina EA, Meire F, Casteels I, de Ravel T, Semina EV, De Baere E. 2014. Novel and recurrent PITX3 mutations in Belgian families with autosomal dominant congenital cataract and anterior segment dysgenesis have similar phenotypic and functional characteristics. Orphanet J Rare Dis 9:26.

Vincent A, Billingsley G, Priston M, Glaser T, Oliver E, Walter M, Ritch R, Levin A, Heon E. 2006. Further support of the role of CYP1B1 in patients with Peters anomaly. Mol Vis 12:506-10.
Vincent A, Billingsley G, Priston M, Williams-Lyn D, Sutherland J, Glaser T, Oliver E, Walter MA, Heathcote G, Levin A and others. 2001. Phenotypic heterogeneity of CYP1B1: mutations in a patient with Peters' anomaly. J Med Genet 38(5):324-6.
Vincent AL, Billingsley G, Buys Y, Levin AV, Priston M, Trope G, Williams-Lyn D, Heon E. 2002. Digenic inheritance of early-onset glaucoma: CYP1B1, a potential modifier gene. Am J Hum Genet 70(2):448-60.
Vulliamy T, Marrone A, Goldman F, Dearlove A, Bessler M, Mason PJ, Dokal I. 2001. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature 413(6854):432-5.
Wang H, Zhang T, Wu D, Zhang J. 2013. A novel beaded filament structural protein 1 (BFSP1) gene mutation associated with autosomal dominant congenital cataract in a Chinese family. Mol Vis 19:2590-5.
Wang J, Ma X, Gu F, Liu NP, Hao XL, Wang KJ, Wang NL, Zhu SQ. 2007. A missense mutation S228P in the CRYBB1 gene causes autosomal dominant congenital cataract. Chin Med J (Engl) 120(9):820-4.
Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, Jiang H, Zhang P, Shen L, Guo JF and others. 2010a. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain 133(Pt 12):3510-8.
Wang K, Wang B, Wang J, Zhou S, Yun B, Suo P, Cheng J, Ma X, Zhu S. 2009a. A novel GJA8 mutation (p.I31T) causing autosomal dominant congenital cataract in a Chinese family. Mol Vis 15:2813-20.
Wang KJ, Li SS, Yun B, Ma WX, Jiang TG, Zhu SQ. 2011a. A novel mutation in MIP associated with congenital nuclear cataract in a Chinese family. Mol Vis 17:70-7.
Wang KJ, Wang BB, Zhang F, Zhao Y, Ma X, Zhu SQ. 2011b. Novel beta-crystallin gene mutations in Chinese families with nuclear cataracts. Arch Ophthalmol 129(3):337-43.
Wang L, Chen X, Lu Y, Wu J, Yang B, Sun X. 2011c. A novel mutation in gammaD-crystallin associated with autosomal dominant congenital cataract in a Chinese family. Mol Vis 17:804-9.
Wang L, Lin H, Gu J, Su H, Huang S, Qi Y. 2009b. Autosomal-dominant cerulean cataract in a chinese family associated with gene conversion mutation in beta-B2-crystallin. Ophthalmic Res 41(3):148-53.
Wang W, Jiang J, Zhu Y, Li J, Jin C, Shentu X, Yao K. 2010b. A novel mutation in the major intrinsic protein (MIP) associated with autosomal dominant congenital cataracts in a Chinese family. Mol Vis 16:534-9.
Wang WH, McNatt LG, Shepard AR, Jacobson N, Nishimura DY, Stone EM, Sheffield VC, Clark AF. 2001. Optimal procedure for extracting RNA from human ocular tissues and expression profiling of the congenital glaucoma gene FOXC1 using quantitative RT-PCR. Mol Vis 7:89-94.
Wei X, Cai SP, Zhang X, Li X, Chen X, Liu X. 2012. Is low dose of estrogen beneficial for prevention of glaucoma? Med Hypotheses 79(3):377-80.
Weisschuh N, Aisenbrey S, Wissinger B, Riess A. 2012. Identification of a novel CRYBB2 missense mutation causing congenital autosomal dominant cataract. Mol Vis 18:174-80.

Weisschuh N, Wolf C, Wissinger B, Gramer E. 2008. A novel mutation in the FOXC1 gene in a family with Axenfeld-Rieger syndrome and Peters' anomaly. Clin Genet 74(5):476-80.
Willoughby CE, Arab S, Gandhi R, Zeinali S, Arab S, Luk D, Billingsley G, Munier FL, Heon E. 2003. A novel GJA8 mutation in an Iranian family with progressive autosomal dominant congenital nuclear cataract. J Med Genet 40(11):e124.
Willoughby CE, Ponzin D, Ferrari S, Lobo A, Landau K, Omidi Y. 2010. Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function – a review. clinical and experimental ophthalmology 38:2-11.
Willoughby CE, Shafiq A, Ferrini W, Chan LL, Billingsley G, Priston M, Mok C, Chandna A, Kaye S, Heon E. 2005. CRYBB1 mutation associated with congenital cataract and microcornea. Mol Vis 11:587-93.
Wistow G. 2012. The human crystallin gene families. Hum Genomics 6:26.
Wride MA. 2011. Lens fibre cell differentiation and organelle loss: many paths lead to clarity. Philos Trans R Soc Lond B Biol Sci 366(1568):1219-33.
Wussuki-Lior O, Abu-Horowitz A, Netzer I, Almer Z, Morad Y, Goldich Y, Yahalom V, Pras E, Pras E. 2011. Hematologic biomarkers in childhood cataracts. Mol Vis 17:1011-5.
Xia XY, Wu QY, An LM, Li WW, Li N, Li TF, Zhang C, Cui YX, Li XJ, Xue CY. 2014. A novel P20R mutation in the alpha-B crystallin gene causes autosomal dominant congenital posterior polar cataracts in a Chinese family. BMC Ophthalmol 14:108.
Xiao X, Li W, Wang P, Li L, Li S, Jia X, Sun W, Guo X, Zhang Q. 2011. Cerulean cataract mapped to 12q13 and associated with a novel initiation codon mutation in MIP. Mol Vis 17:2049-55.
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(7):535-9.
Yan M, Xiong C, Ye SQ, Chen Y, Ke M, Zheng F, Zhou X. 2008. A novel connexin 50 (GJA8) mutation in a Chinese family with a dominant congenital pulverulent nuclear cataract. Mol Vis 14:418-24.
Yang G, Xing B, Liu G, Lu X, Jia X, Lu X, Wang X, Yu H, Fu Y, Zhao J. 2011a. A novel mutation in the GJA3 (connexin46) gene is associated with autosomal dominant congenital nuclear cataract in a Chinese family. Mol Vis 17:1070-3.
Yang G, Zhang G, Wu Q, Zhao J. 2011b. A novel mutation in the MIP gene is associated with autosomal dominant congenital nuclear cataract in a Chinese family. Mol Vis 17:1320-3.
Yang J, Zhu Y, Gu F, He X, Cao Z, Li X, Tong Y, Ma X. 2008. A novel nonsense mutation in CRYBB1 associated with autosomal dominant congenital cataract. Mol Vis 14:727-31.
Yang Z, Li Q, Ma X, Zhu SQ. 2015a. Mutation Analysis in Chinese Families with Autosomal Dominant Hereditary Cataracts. Curr Eye Res 40(12):1225-31.
Yang Z, Li Q, Ma Z, Guo Y, Zhu S, Ma X. 2011c. A G-->T splice site mutation of CRYBA1/A3 associated with autosomal dominant suture cataracts in a Chinese family. Mol Vis 17:2065-71.
Yang Z, Li Q, Zhu S, Ma X. 2015b. A G57W Mutation of CRYGS Associated with Autosomal Dominant Pulverulent Cataracts in a Chinese Family. Ophthalmic Genet 36(3):281-3.

Yao K, Li J, Jin C, Wang W, Zhu Y, Shentu X, Wang Q. 2011a. Characterization of a novel mutation in the CRYBB2 gene associated with autosomal dominant congenital posterior subcapsular cataract in a Chinese family. Mol Vis 17:144-52.
Yao K, Wang W, Zhu Y, Jin C, Shentu X, Jiang J, Zhang Y, Ni S. 2011b. A novel GJA3 mutation associated with congenital nuclear pulverulent and posterior polar cataract in a Chinese family. Hum Mutat 32(12):1367-70.
Yasmeen A, Riazuddin SA, Kaul H, Mohsin S, Khan M, Qazi ZA, Nasir IA, Zafar AU, Khan SN, Husnain T and others. 2010. Autosomal recessive congenital cataract in consanguineous Pakistani families is associated with mutations in GALK1. Mol Vis 16:682-8.
Yi J, Yun J, Li ZK, Xu CT, Pan BR. 2011. Epidemiology and molecular genetics of congenital cataracts. Int J Ophthalmol 4(4):422-32.
Yu Chan JY, Choy BN, Ng AL, Shum JW. 2015. Review on the Management of Primary Congenital Glaucoma. J Curr Glaucoma Pract 9(3):92-9.
Yu LC, Twu YC, Chang CY, Lin M. 2001. Molecular basis of the adult i phenotype and the gene responsible for the expression of the human blood group I antigen. Blood 98(13):3840-5.
Yuan L, Guo Y, Yi J, Xiao J, Yuan J, Xiong W, Xu H, Yang Z, Zhang J, Deng H. 2015. Identification of a novel GJA3 mutation in congenital nuclear cataract. Optom Vis Sci 92(3):337-42.
Zeng L, Liu W, Feng W, Wang X, Dang H, Gao L, Yao J, Zhang X. 2013. A novel donor splice-site mutation of major intrinsic protein gene associated with congenital cataract in a Chinese family. Mol Vis 19:2244-9.
Zhang J, Zhang Y, Fang F, Mu W, Zhang N, Xu T, Cao Q. 2014. Congenital cataracts due to a novel 2bp deletion in CRYBA1/A3. Mol Med Rep 10(3):1614-8.
Zhang L, Qu X, Su S, Guan L, Liu P. 2012. A novel mutation in GJA3 associated with congenital Coppock-like cataract in a large Chinese family. Mol Vis 18:2114-8.
Zhang LY, Yam GH, Fan DS, Tam PO, Lam DS, Pang CP. 2007. A novel deletion variant of gammaD-crystallin responsible for congenital nuclear cataract. Mol Vis 13:2096-104.
Zhang T, Hua R, Xiao W, Burdon KP, Bhattacharya SS, Craig JE, Shang D, Zhao X, Mackey DA, Moore AT and others. 2009. Mutations of the EPHA2 receptor tyrosine kinase gene cause autosomal dominant congenital cataract. Hum Mutat 30(5):E603-11.
Zhang W, Hawse J, Huang Q, Sheets N, Miller KM, Horwitz J, Kantorow M. 2002. Decreased expression of ribosomal proteins in human age-related cataract. Invest Ophthalmol Vis Sci 43(1):198-204.
Zhao L, Chen XJ, Zhu J, Xi YB, Yang X, Hu LD, Ouyang H, Patel SH, Jin X, Lin D and others. 2015. Lanosterol reverses protein aggregation in cataracts. Nature 523(7562):607-11.
Zhao R, Yang Y, He X, Liu Z, Wang P, Zhou L, Tang J, Xu W, Li L, Zhu Y. 2011. An autosomal dominant cataract locus mapped to 19q13-qter in a Chinese family. Mol Vis 17:265-9.
Zheng JQ, Ma ZW, Sun HM. 2005. [A heterozygous transversion of connexin 50 in a family with congenital nuclear cataract in the northeast of China]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 22(1):76-8.
Zhou D, Ji H, Wei Z, Guo L, Li Y, Wang T, Zhu Y, Dong X, Wang Y, He L and others. 2013. A novel insertional mutation in the connexin 46 (gap junction alpha 3)

gene associated with autosomal dominant congenital cataract in a Chinese family. Mol Vis 19:789-95.
Zhou G, Zhou N, Hu S, Zhao L, Zhang C, Qi Y. 2010a. A missense mutation in CRYBA4 associated with congenital cataract and microcornea. Mol Vis 16:1019-24.
Zhou Z, Hu S, Wang B, Zhou N, Zhou S, Ma X, Qi Y. 2010b. Mutation analysis of congenital cataract in a Chinese family identified a novel missense mutation in the connexin 46 gene (GJA3). Mol Vis 16:713-9.
Zhu Y, Yu H, Wang W, Gong X, Yao K. 2014. A novel GJA8 mutation (p.V44A) causing autosomal dominant congenital cataract. PLoS One 9(12):e115406.
Zhuang X, Wang L, Song Z, Xiao W. 2015. A Novel Insertion Variant of CRYGD Is Associated with Congenital Nuclear Cataract in a Chinese Family. PLoS One 10(7):e0131471.