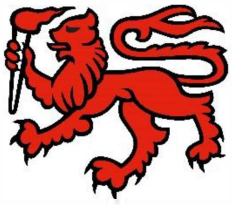


UNIVERSITY*of*
TASMANIA

Identifying genetic causes of paediatric cataract onset and severity: validation of candidate gene *HTR1F* and mapping of modifier genes

by

Duran Zhao

BMed, MMed

Menzies Institute for Medical Research | College of Health and Medicine

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (Medical Studies)

University of Tasmania, March 2022

Declaration of originality

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Duran Zhao

18/11/2021

Authority of access statement

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

Duran Zhao

18/11/2021

Statement of co-authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

Candidate — Duran Zhao, Menzies Institute for Medical Research

Author 1 — Kathryn Burdon, Menzies Institute for Medical Research

Author 2 — Guei-Sheung Liu, Menzies Institute for Medical Research

Author 3 — Johanna Jones, Menzies Institute for Medical Research

Author 4 — Robert Gasperini, School of Medicine, University of Tasmania

Author 5 — Jac Charlesworth, Menzies Institute for Medical Research

Contribution of work by co-authors for each paper:

PAPER 1: Located in Chapter 3

Zhao D, Jones JL, Gasperini RJ, Charlesworth JC, Liu GS, Burdon KP. Rapid and efficient cataract gene evaluation in F0 zebrafish using CRISPR-Cas9 ribonucleoprotein complexes. *Methods*. 2021 Oct;194:37-47.

Author contributions:

Conceived and designed experiment: Candidate, Author 1, Author 2

Performed the experiments: Candidate

Assisted the experiments: Author 3

Analysed the data: Candidate

Contributed reagents/materials/analysis tools: Author 4

Wrote the manuscript: Candidate

Critically review and final approval of manuscript: Author 1, Author 2, Author 3, Author 4, Author 5.

We, the undersigned, endorse the above stated contribution of work undertaken for each of the published (or submitted) peer-reviewed manuscripts contributing to this thesis:

Signed:			
	Duran Zhao	Professor Kathryn Burdon	Distinguished Professor Alison Venn
	Candidate	Primary supervisor	Director
	Menzies Institute for Medical Research	Menzies Institute for Medical Research	Menzies Institute for Medical Research
	University of Tasmania	University of Tasmania	University of Tasmania
Date:	18/11/21	19/11/21	22/11/2021

Statement of ethical conduct

The research associated with this thesis abides by the international and Australian codes on human experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committee of the University. Ethics Approval Numbers: H0014539 and A0017743.

Duran Zhao

18/11/2021

Acknowledgements

I would like to offer my sincere acknowledgements to numerous people who have helped, encouraged, and supported me throughout the almost four-year PhD journey.

To the University of Tasmania and Anhui Medical University for my full-time PhD scholarship, and the funding support from the National Health and Medical Research Council.

To my primary supervisor Prof Kathryn Burdon, it is a fortune to have the opportunity to work and learn from you. Your deep insights have always pushed me to sharpen my thinking and brought my work to a higher level. Your calm and optimistic attitude in the face of difficulties has profoundly inspired me. You are a responsible and respected supervisor.

To my co-supervisors, A/Prof Guei-Sheung Liu and Dr Jac Charlesworth, for your professional guidance and unwavering support whenever I need it. I could not have asked for a better supervision team.

To the research participants, collaborators, and clinicians involved in the paediatric cataract study, without whom this project would not exist.

To the members of the CompGen team and my lab-mates. A special thanks to Jo for all the lab skills and experience you have shared with me. The zebrafish experiments could not be completed without your help. To Elise, Rob, Hayden, Silvia, for the assistance in maintaining the operation of the

zebrafish lab. To Nizam and James, for sharing the experimental expertise with me without reservation. To Bennet and Nick, for your bioinformatic support. To Pat, Sionne, Alex, Kelsie, Raj, for your kind concern and help in my work and life.

To my dearest friend Ming, it is so lucky to have your accompany while studying for a PhD in a foreign country far away from home. We helped each other through all the ups and downs and I will cherish our friendship forever.

Finally, to my parents, thank you for your unconditional love and encouragement. You have always been my strongest backing and my warmest harbour. Without your endless care and support, I would not be able to get to where I am.

Abstract

Paediatric cataract is an opacity of the normally transparent lens in the eye, which occurs at birth (congenital) or develops during childhood. It is a leading cause of irreversible blindness in children. Up to 30% of paediatric cataracts are inherited. Although hundreds of cataract-related genes and loci have been found, genetic aetiology can not be identified in at least 40% of patients with inherited cataracts. Paediatric cataract is a highly heterogeneous disease with variable severity observed in affected individuals with the same causative mutation within a family. Several studies have suggested the involvement of modifier genes in paediatric cataract formation, but such genes have not been identified. This project consists of two studies and the overall aim is to identify causative and modifier genes for paediatric cataract.

The first study assessed the biological function of a novel candidate causative gene previously identified in a family with paediatric cataract. The *HTR1F* gene was knocked out in F0 zebrafish using CRISPR-Cas9 ribonucleoprotein complexes. There was no significant difference in cataract formation between the *htr1f*-knockout group and the control group. Therefore we concluded that knocking out *HTR1F* in the zebrafish models does not provide evidence for it being a cataract causative gene. Further investigations are warranted to obtain additional evidence.

The second study investigated the potential for modifier genes in two extended families (family CRCH13 and family CSA91/110) with autosomal

dominant paediatric cataract. The clinical phenotype information for each participant was updated and a high degree of variability for age of diagnosis and age of surgery within each family was noted. Genome-wide multipoint parametric linkage analyses of the affection status of cataract combined with whole-genome sequencing validated the previously reported causative genes. In addition, linkage analysis also indicated the existence of potential modifier genes through the presence of several suggestive linkage peaks.

Further study of genes that modify the disease severity was conducted in these two extended paediatric cataract families. Genome-wide multipoint variance components linkage analysis of disease severity classified by age of diagnosis and age of surgery revealed one significant linkage region and two suggestive linkage regions likely containing modifier variants in family CRCH13, but failed to detect any linkage regions in CSA91/110 likely due to limited statistical power.

In summary, we evaluated the function of a novel candidate cataract-causing gene in zebrafish models. In addition, we developed an approach for rapid and highly efficient evaluation of cataract candidate genes in F0 zebrafish using CRISPR-Cas9 ribonucleoprotein complexes. Zebrafish cataract assessment criteria were also generated, which to our knowledge, is the first systematic, definitive cataract assessment criteria for the zebrafish model. We confirmed the causative genes in two large paediatric cataract families and identified candidate linkage intervals for genetic modifiers in one of the families. The findings and novel strategies generated from this project can benefit the understanding of the pathogenesis of paediatric cataract and contribute to the identification of novel cataract-causing genes and modifier genes.

Table of contents

Declaration of originality ii
Authority of access statement iii
Statement of co-authorship iv
Statement of ethical conduct vi
Acknowledgements vii
Abstract ix
Table of contents xi
List of tables xvi
List of figures xviii
List of abbreviations xx
Chapter 1 Introduction 1
1.1 The human lens 1
1.1.1 Anatomy and function 1
1.1.2 Embryological development 4
1.2 Paediatric cataract 6
1.2.1 Phenotypes 6
1.2.2 Aetiology 9
1.2.3 Current treatment and its complications 9
1.3 Genes implicated in cataractogenesis 10
1.3.1 Crystallin genes 11
1.3.2 Membrane and cytoskeletal protein genes 12
1.3.3 Transcription factor genes 14
1.3.4 Metabolism-related genes 15
1.3.5 Novel cataract genes to be discovered 16

1.4 Phenotypic heterogeneity and genetic modifiers in paediatric cataract17
1.5 Overall hypothesis and aims....19
Chapter 2 General materials and methods....21
2.1 Materials....21
2.1.1 Reagents and equipment....21
2.1.2 Reagent setup....24
2.2 Ethics statement....26
2.3 Study participants....27
2.4 DNA preparation....28
2.4.1 Amplification....28
2.4.2 Quantification....28
2.5 Target sequence amplification....29
2.5.1 Primer design....29
2.5.2 Standard PCR and Reverse transcription-PCR....29
2.6 Agarose gel electrophoresis....30
2.7 Sanger sequencing....30
Chapter 3 Investigation of the role of *HTR1F* in cataract formation....32
3.1 Introduction....32
3.1.1 Family CSA92 background....32
3.1.2 Candidate variants identified in CSA92....35
3.1.3 *HTR1F*....37
3.1.4 Zebrafish model....38
3.1.5 CRISPR-Cas9....40
3.2 Hypothesis and Aim....42
3.3 Methods....43
3.3.1 Overall workflow....43
3.3.2 Zebrafish maintenance and breeding....43
3.3.3 Zebrafish genomic DNA extraction and amplification....44
3.3.4 Identification of *HTR1F* orthologs in zebrafish....45
3.3.5 Generation of the mutant zebrafish model....46

3.3.6 Quantification of pigment loss in the optimization of mutagenesis conditions 50
3.3.7 Evaluation of cataract formation by microscope imaging 52
3.3.8 Analysis of editing efficiency in F0 embryos 56
3.3.9 Statistical analysis 59
3.4 Results 59
3.4.1 Orthologs of *HTR1F* and their expression in zebrafish 59
3.4.2 Designed gRNAs 64
3.4.3 Optimizing dose and delivery route of CRISPR-Cas9 RNP 65
3.4.4 Determining the editing efficiency of *htr1f* gRNAs 69
3.4.5 Knockout of *htr1f* in F0 zebrafish 70
3.4.6 Cataract assessment 71
3.4.7 Editing efficiency analysis 73
3.5 Discussion 79
3.5.1 Is *HTR1F* the cataract-causing gene in family CSA92? 79
3.5.2 Functional analysis of cataract gene in F0 zebrafish using CRISPR-Cas9 RNP 83
3.6 Outcomes and significance of this study 87
Chapter 4 Validation of reported causative genes in two extended paediatric cataract families 89
4.1 Introduction 89
4.1.1 Family CRCH13 background 90
4.1.2 Background of the CSA families 90
4.2 Hypothesis and aims 91
4.3 Methods 92
4.3.1 Overall workflow 92
4.3.2 Whole-genome SNP genotyping 93
4.3.3 Genotyping data processing 94
4.3.4 Relationship testing 95
4.3.5 Inheritance error checking 95

4.3.6 Parametric linkage analysis 95
4.3.7 Whole-genome sequencing (WGS) 96
4.3.8 WGS data processing 97
4.3.9 Variant filtering 98
4.4 Results 101
4.4.1 Family phenotype and clinical information update 101
4.4.2 Whole-genome genotyping data 110
4.4.3 Family CRCH13 112
4.4.4 Family CSA91/110 117
4.5 Discussion 120
4.5.1 Causative genes in these two extended paediatric cataract families 120
4.5.2 Potential modifier genes? 122
4.5.3 The strategy of identifying cataract-causing genes in large families by combining linkage analysis and WGS 123
4.6 Outcomes and significance of this study 126
Chapter 5 Investigation of disease severity modifiers in two extended paediatric cataract families 127
5.1 Introduction 127
5.1.1 Genetic modifiers of cataractogenesis 128
5.2 Hypothesis and aim 129
5.3 Methods 130
5.3.1 Age of diagnosis/surgery 130
5.3.2 Disease severity assessment and classification 130
5.3.3 Variance components linkage analysis 131
5.4 Results 132
5.4.1 Family CSA91/110 132
5.4.2 Family CRCH13 135
5.5 Discussion 142
5.5.1 Mapping modifier loci linked to disease severity in extended pedigrees using variance components linkage analysis 142

5.5.2 Detection of disease severity modifiers in the two paediatric cataract families 145
5.6 Outcomes and significance of this study 149
Chapter 6 General discussion 150
6.1 Summary of the project 150
6.2 Knocking out the *HTR1F* gene in F0 zebrafish did not provide evidence for it being a cataract causative gene 151
6.3 Multiple genetic factors contribute to the phenotypic manifestation of paediatric cataract 156
6.4 Conclusion 160
Appendices 161
Appendix 1 List of primer sequences 161
Appendix 2 Publication of material from Chapter 3 163
Appendix 3 Script for Miseq paired-end data trimming 175
Appendix 4 Commands for analyzing genome editing outcomes in CRISPResso2 176
Appendix 5 Targeting sequence of *htr1f* gRNAs 178
Appendix 6 Indel position of the *htr1fa* knockout experiments 180
Appendix 7 Indel position of the *htr1fb* knockout experiments 182
Appendix 8 R script to convert Illumina final report to PLINK long-format files 184
Appendix 9 R script to categorise whole-genome sequencing data by sequencing quality 186
Appendix 10 R script to annotate the whole-genome sequencing data. 187
Appendix 11 The kinship of family CSA91 and family CSA110 188
References 191

List of tables

Table 2.1 List of reagents....21

Table 2.2 List of equipment....23

Table 2.3 Reaction components and cycling conditions of standard PCR and RT-PCR....29

Table 2.4 Reaction components and cycling conditions of Sanger sequencing....31

Table 3.1 Family CSA92 phenotype and clinical information....34

Table 3.2 Reaction components and cycling conditions of zebrafish genomic DNA amplification....45

Table 3.3 9μM gRNA working solution....48

Table 3.4 1.5μg/μl Cas9 protein working solution....48

Table 3.5 Reaction components and cycling conditions of barcoding PCR58

Table 3.6 gRNA sequences....65

Table 3.7 *htr1f* gRNA editing efficiency....70

Table 3.8 Cataract assessment of *htr1f* knockout experiments....72

Table 3.9 Information of samples for editing efficiency analysis....73

Table 3.10 Summary of characteristics of *htr1fa* editing efficiency analysis....75

Table 3.11 Summary of characteristics of *htr1fb* editing efficiency analysis....76

Table 4.1 Family CRCH13 clinical information....103

Table 4.2 Clinical information of family CSA91/110....108

Table 4.3 The number of finalised SNPs used for linkage analysis....111

Table 4.4 Affection status assigned to members involved in the parametric linkage analysis of family CRCH13....113

Table 4.5 Linkage regions identified in the parametric linkage analysis of affection status in family CRCH13 115

Table 4.6 Affection status assigned to members involved in the parametric linkage analysis of family CSA91/110 118

Table 4.7 Linkage regions identified in the parametric linkage analysis of affection status in family CSA91/110 118

Table 5.1 Disease severity assigned to members involved in the variance components linkage analysis of family CSA91/110 132

Table 5.2 Disease severity assigned to members involved in the variance components linkage analysis of family CRCH13 (four categories) 137

Table 5.3 Linkage regions identified in variance components linkage analysis of disease severity in family CRCH13 (four categories) 140

List of figures

Figure 1.1 Cross-section diagram of the human eye 1

Figure 1.2 Three-dimensional diagram showing the structure of the mature human lens 3

Figure 1.3 Diagrams showing embryonic development of the lens 5

Figure 1.4 Examples of some common paediatric cataract phenotypes 8

Figure 1.5 Genes associated with paediatric cataract 10

Figure 3.1 The pedigree of family CSA92 33

Figure 3.2 Lens photos of CSA92.07 showing nuclear lamellar cataract 35

Figure 3.3 RT-PCR showing *HTR1F* expression in human lens 38

Figure 3.4 The structure and functional mechanism of the CRISPR-Cas9 system 41

Figure 3.5 Overview of the key steps in Chapter 3 43

Figure 3.6 Zebrafish embryos at 2dpf 50

Figure 3.7 Quantification of the pigment loss 51

Figure 3.8 Zebrafish cataract assessment criteria (pitting) 54

Figure 3.9 Zebrafish cataract assessment criteria (central mass) 55

Figure 3.10 Alignment of zebrafish Htr1fa and human HTR1F 61

Figure 3.11 Alignment of zebrafish Htr1fb and human HTR1F 62

Figure 3.12 The transcripts and expressions of zebrafish *htr1fa* and *htr1fb* genes 63

Figure 3.13 CRISPR-Cas9 RNP dose optimization 67

Figure 3.14 Illustration of cell injection (A) and yolk injection (B) 68

Figure 3.15 CRISPR-Cas9 RNP delivery route optimization 69

Figure 3.16 Indel size distribution of the *htr1fa* knockout experiments 77

Figure 3.17 Indel size distribution of the *htr1fb* knockout experiments 78

Figure 3.18 The differences in the embryonic development of zebrafish (A-E) and mammalian (A'-E') lenses.....86

Figure 4.1 Workflow of the key steps in Chapter 4 92

Figure 4.2 Single nucleotide variants filtering pipeline 100

Figure 4.3 CRCH13 family pedigree.....102

Figure 4.4 Lens photos of family CRCH13 105

Figure 4.5 Pedigree of family CSA91/110.....107

Figure 4.6 Lens photos of family CSA91/110 109

Figure 4.7 Subfamilies consisting of individuals involved in the parametric linkage analysis of family CRCH13 114

Figure 4.8 Autosome-wide parametric linkage analysis of affection status in family CRCH13.....116

Figure 4.9 Autosome-wide parametric linkage analysis of affection status in family CSA91/110.....119

Figure 5.1 Classification and distribution of disease severity in family CSA91/110 134

Figure 5.2 Disease severity distribution in family CRCH13 (five categories) 136

Figure 5.3 Classification and distribution of disease severity in family CRCH13 (four categories) 139

Figure 5.4 Genome-wide variance components linkage analysis of cataract severity of family CRCH13 (four categories) 141

List of abbreviations

bp	base pair
cDNA	complementary DNA
chr	chromosome
cM	centiMorgan
CRISPR	clustered regularly interspaced short palindromic repeat
Cas9	CRISPR-associated system 9
crRNA	CRISPR RNA
dpf	day-post-fertilization
gRNA	guide RNA
GWAS	genome-wide association study
h^2	heritability
hg19	human reference genome version 19
HLOD	heterogeneity logarithm of the odds
indel	insertion and deletion
LOD	logarithm of the odds
MAF	minor allele frequency
mRNA	messenger RNA
NGS	next-generation sequencing
PCR	polymerase chain reaction
PTU	1-phenyl-2-thiourea
QC	quality control
RNP	ribonucleoprotein
RT-PCR	reverse transcription polymerase chain reaction
SEM	standard error of the mean
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SV	structure variant
tracrRNA	trans-activating RNA
WES	whole-exome sequencing
WGS	whole-genome sequencing

Chapter 1

Introduction

1.1 The human lens

1.1.1 Anatomy and function

The ocular lens is a transparent lamellar structure suspended behind the iris, which transmits and focuses light onto the retina to form clear images (Figure 1.1).

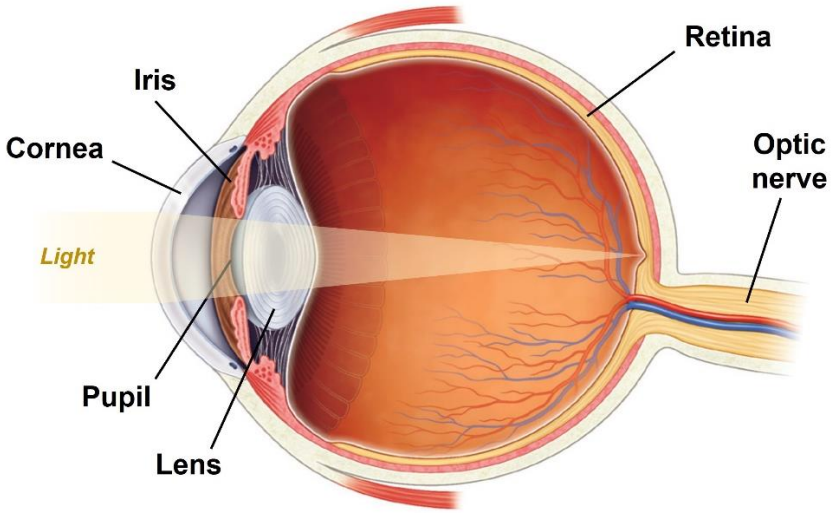


Figure 1.1 Cross-section diagram of the human eye

(Adapted from Levin *et al.*, 2011[1])

The human lens is a biconvex optical tissue with no blood vessels, nerves, or lymph. The three-dimensional structure of the lens is shown in Figure 1.2. The entire lens is enclosed within a transparent capsule. Beneath the anterior capsule is a single layer of lens epithelium, which exhibits lifelong mitotic activity. The epithelial cells located at the equator of the lens differentiate into secondary fibre cells. The developing secondary fibre cells elongate towards the anterior and posterior poles of the lens until they meet the elongating cells from the other side. The ends of the cells on both sides are joined at the anterior and posterior midlines to form sutures. During elongation and differentiation, secondary fibre cells gradually degrade organelles and nuclei, accumulate crystallin proteins in the cytoplasm, and eventually form mature secondary lens fibres. The newly generated secondary fibre cells cover the top of the older ones to form onion-like concentric layers, which is the cortex of the lens. Inside the cortex is the lens nucleus, which is composed of primary fibre cells derived from the posterior epithelial cells of the lens vesicle (see Section 1.1.2). The lens nucleus can be further divided into adult, infantile, fetal and embryonic nucleus from the outside to the inside. Overall, the fibre cells in the central lens are the oldest, and the ones in the outer layers are newer.

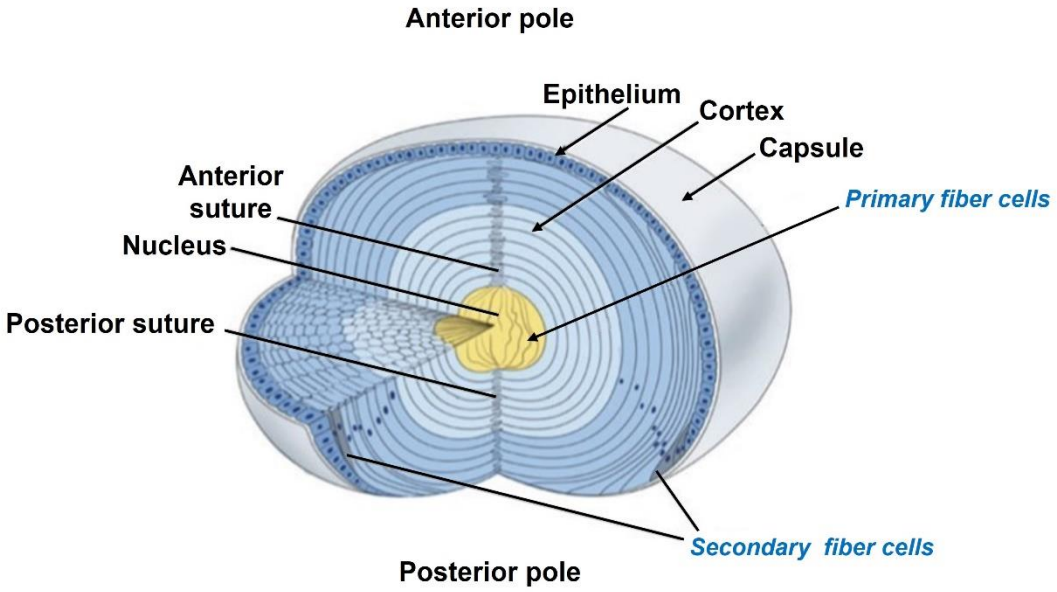


Figure 1.2 Three-dimensional diagram showing the structure of the mature human lens

(Adapted from Cvekl & Ashery-Padan, 2014[2])

The transparent structure and refractive properties of the lens allow it to smoothly transmit and focus the light onto the retina. Since all mature fibre cells in the lens are absent of nuclei and organelles, light rays can pass through them without any interference. It is reported that the concentration of crystallin proteins in lens fibre cells is 3 times the protein content of typical cells[3]. The refractive index of the lens is mainly provided by the high concentration of crystallin proteins, which makes the refractive index of the lens significantly higher than that of the surrounding fluid. The high content of the crystallin proteins is one of the important reasons for the transparency of the lens fibres as well. In addition, the hexagonal profile (cross-section) of the lens fibre cells enables them to be tightly arranged in a highly ordered manner, thus minimizing light scattering. The curvature of the refractive surfaces of the lens also contributes to the refractive power.

1.1.2 Embryological development

A sequence of inductive processes leads to lens formation[4]. Human lens development begins at 22 days' gestation (4mm embryonic stage), with the elongation of surface ectoderm cells to form the thickened lens placode[5] (Figure 1.3A and B). The lens placode then invaginates to form the lens pit which subsequently develops into the lens vesicle and separates from the surface ectoderm[6] (Figure 1.3C and D). Soon afterwards, the single layer of epithelial cells at the posterior end of the lens vesicle elongate rapidly to form primary fibre cells and fill in the cavity of the vesicle (Figure 1.3E and F). Thereafter secondary fibre cells are generated from the anterior epithelial cells in the equatorial zone[7]. The primary fibre cells gradually degrade the nuclei and organelles, and finally form the embryonic nucleus. The secondary fibre cells continue to be generated throughout life and encircle the primary cell to form the cortex.

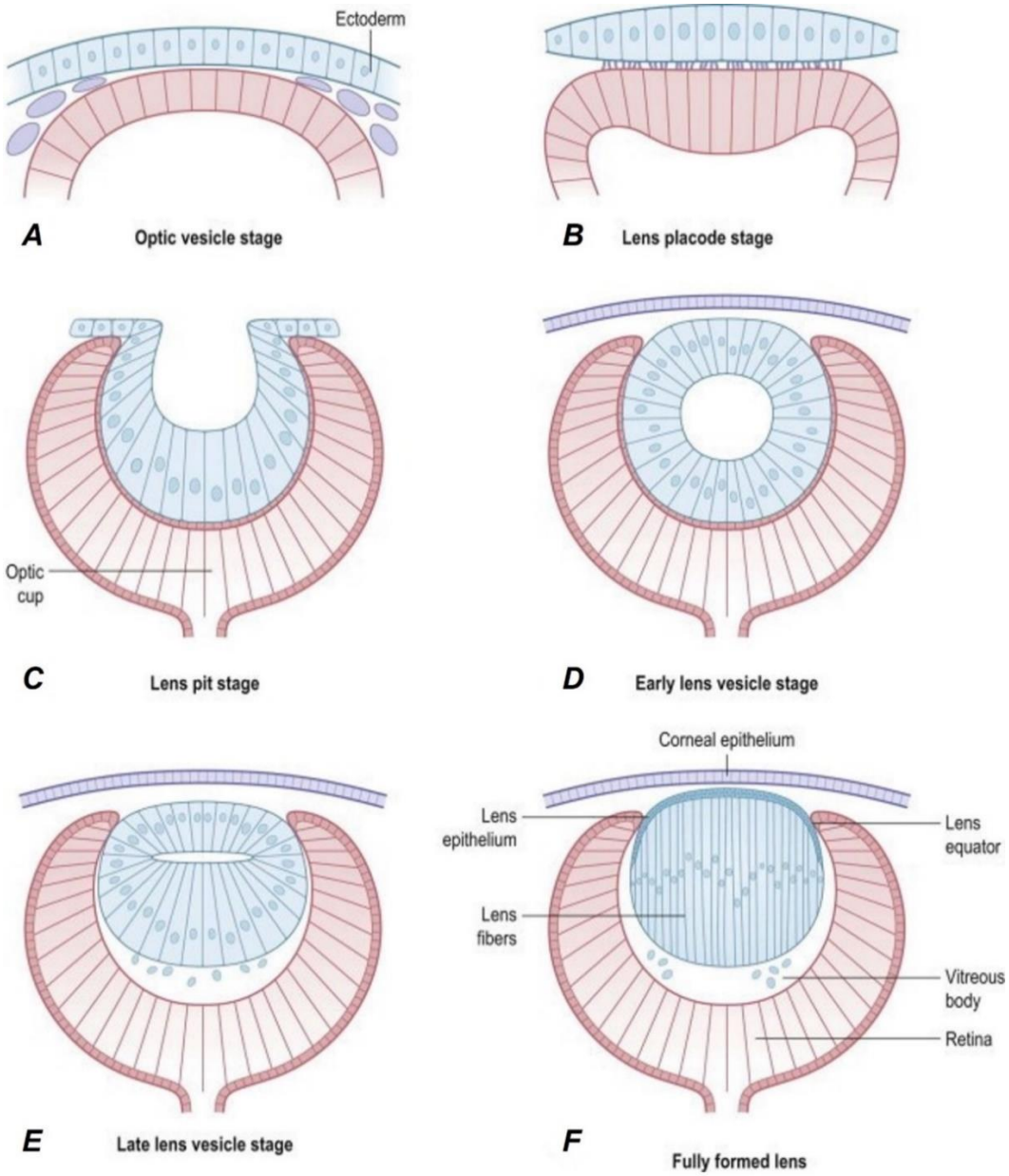


Figure 1.3 Diagrams showing embryonic development of the lens

A. The surface ectoderm and the optic vesicle adjacent to each other, lens start to form. **B.** The surface ectoderm cells in contact with the optic vesicle elongate to form the lens placode. **C.** The lens placode and the adjacent outer surface of the optic vesicle invaginate together to form the lens pit. **D.** Lens placode closes over and separates from the surface ectoderm to form the lens vesicle. **E.** The single layer of epithelial cells at the posterior pole of the lens vesicle elongates rapidly and differentiates into primary fibre cells. **F.** The primary fibre cells continue to elongate until they reach the anterior epithelial cells. At this stage, the bulk of the lens is filled with elongated primary fibre cells whose nuclei and organelles have not yet degenerated, and the secondary fibre cells have not yet developed. (Adapted from Levin *et al.*, 2011[1])

1.2 Paediatric cataract

Paediatric cataract is described as an opacity of the transparent ocular lens appearing at birth (congenital) or developing during childhood. The estimated prevalence of paediatric cataract ranges from 0.42-2.05 per 10,000 cases in developed countries to 0.63-13.6 per 10,000 cases in developing countries[8,9]. In Australia, the prevalence of paediatric cataract is 2.2 per 10,000 live births[10]. It is the principal cause of irreversible visual impairment or blindness in children. Congenital cataracts account for 5-20% of childhood blindness globally, but 22-30% of childhood blindness in developing countries[11]. Although paediatric cataracts are less common than age-related cataracts, they have severe and profound negative impacts on the visual development of young children. Without prompt intervention, opacity located on the central visual axis of the lens can result in permanent blindness due to stimulus deprivation amblyopia (lazy eye)[12]. In terms of considerable costs of diagnosis and treatment, long-term management and loss of productivity, paediatric cataract tends to be a heavy burden for not only the affected children but also their family and even the entire society[13,14].

1.2.1 Phenotypes

Paediatric cataracts may occur as unilateral or bilateral diseases. The majority of paediatric cataracts are isolated or non-syndromic (70%), while syndromic cases with other ocular disorders (15%) or systematic abnormalities (15%) have also been observed[15]. There is no universal method for the classification of paediatric cataracts. Based on the anatomic location in the lens (Figure 1.2), paediatric cataracts can be classified into polar (anterior or

posterior), subcapsular, cortical, nuclear, sutural, and total cataract. Alternatively, the morphology of the opacity is used to define paediatric cataracts, such as lamellar, pulverulent, membranous and cerulean (blue-dot) cataract. Examples of some clinically common types are given in Figure 1.4.

The location and morphology of the opacity provide a wealth of information for the onset, aetiology, and prognosis of paediatric cataract. For example, embryonic nuclear cataracts and fetal nuclear cataracts indicate abnormal gene expression or function arising during the first 2 to 3 months of gestation, respectively[16]. This is because the primary fibre cells that make up the embryonic and fetal lens nucleus are formed during these two time periods (Section 1.1.2). Since opacities are located in the centre of the lens, vision in patients with nuclear cataracts is commonly reduced. Lamellar cataract affects one or more of the concentric layers of secondary fibre cells, suggesting a short-term interruption of development (usually during the fetal period)[17]. Differing from lamellar cataract, cortical cataract is confined to a sector of the outer cortex that is adjacent to the lens capsule. Although the pathogenesis is unclear, the location of the opacity suggests that the disorder occurs in the later developmental stage[18]. Anterior polar cataracts are small white dots on the anterior surface of the lens, which usually have little impact on vision and rarely progress. In contrast, posterior polar cataracts often reduce vision as the opacities are close to the crucial point in the visual pathway.

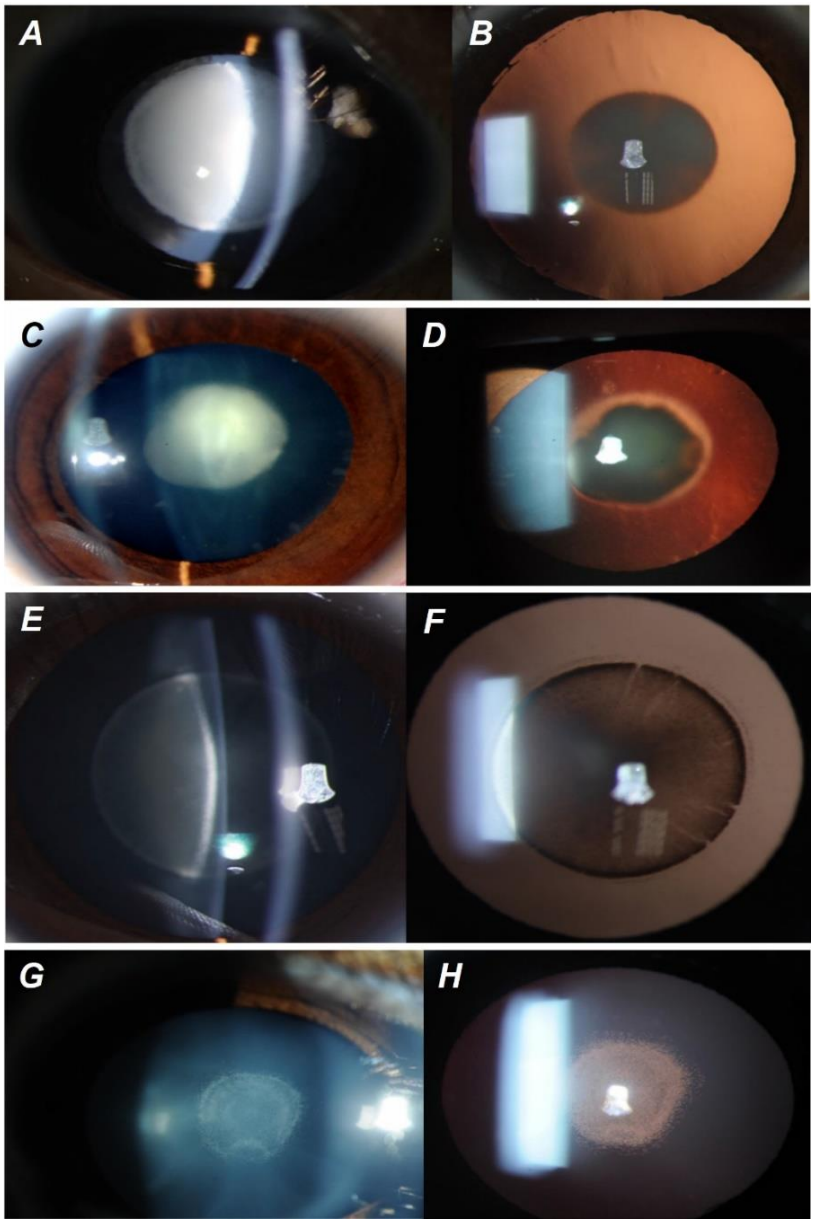


Figure 1.4 Examples of some common paediatric cataract phenotypes

Lens photos showing the common paediatric cataract phenotypes (**A, C, E, G**: slit-lamp view; **B, D, F, H**: retro-illumination view). **A** and **B.** Nuclear cataract. **C** and **D.** Posterior polar cataract. The slit-lamp photo shows the discoid shape opacity is located in the centre of the posterior capsule. **E** and **F.** Lamellar cataract. Disk-like opacities around the fetal nucleus are shown. **G** and **H.** Pulverulent cataract. Multiple punctate opacities can be observed in the lens nucleus. (**A, B, E, F** adapted from Chen, 2017[19]; **C, D** adapted from Kalantan, 2012[20]; **G, H** adapted from Khan *et al.* 2012[21])

1.2.2 Aetiology

Although paediatric cataract can be caused by a variety of factors, such as maternal infections, trauma, inflammation, metabolic disorders, and ocular malformations, a significant proportion of cases are attributed to genetic causes[16]. It is estimated that up to 30% of paediatric cataracts are inherited[22-24], but for bilateral cases, inherited genetic mutations are responsible for approximately 50% of the patients[25]. Unilateral paediatric cataract can be either genetic or acquired[22]. Hereditary paediatric cataract can be inherited in three patterns: autosomal dominant, autosomal recessive and X-linked, with autosomal dominant being the most frequent mode. *De novo* and somatic mutations are also reported to cause paediatric cataracts[26-28].

1.2.3 Current treatment and its complications

So far, the only available intervention for paediatric cataract is cataract-removal surgery combined with primary intraocular lens (IOL) implantation. In order to prevent deprivation amblyopia, ideally therapeutic intervention of paediatric cataracts that present from birth should be in the first 6 to 8 weeks of life[29]. However, the changeable axial length and corneal curvature as well as the intense inflammatory response in children make paediatric cataract surgery complicated and challenging. Despite great improvements that have been made on the surgical techniques, the outcomes are variable and the complication rate is high. Posterior capsule opacification (PCO) is the most frequent complication of paediatric cataract surgery. Children with PCO need a second intraocular surgery to remove the re-opacification[30]. Another well-recognised postoperative complication of paediatric cataract is secondary

glaucoma. It has been reported that infants who underwent cataract surgery before one month of age had the highest rate of secondary glaucoma[12]. Considering all these limitations of cataract surgery, understanding the genetics of paediatric cataract has become extremely important.

1.3 Genes implicated in cataractogenesis

By March 2021, more than 200 genes and loci have been identified in inherited paediatric cataracts according to the Cat-Map[31] (https://cat-map.wustl.edu/) database. Among them, over 52 genes have been discovered for isolated paediatric cataract[24]. These genes encode a broad range of functional proteins, including crystallins, membrane and cytoskeletal proteins, transcription factors, and metabolism-related proteins (Figure 1.5).

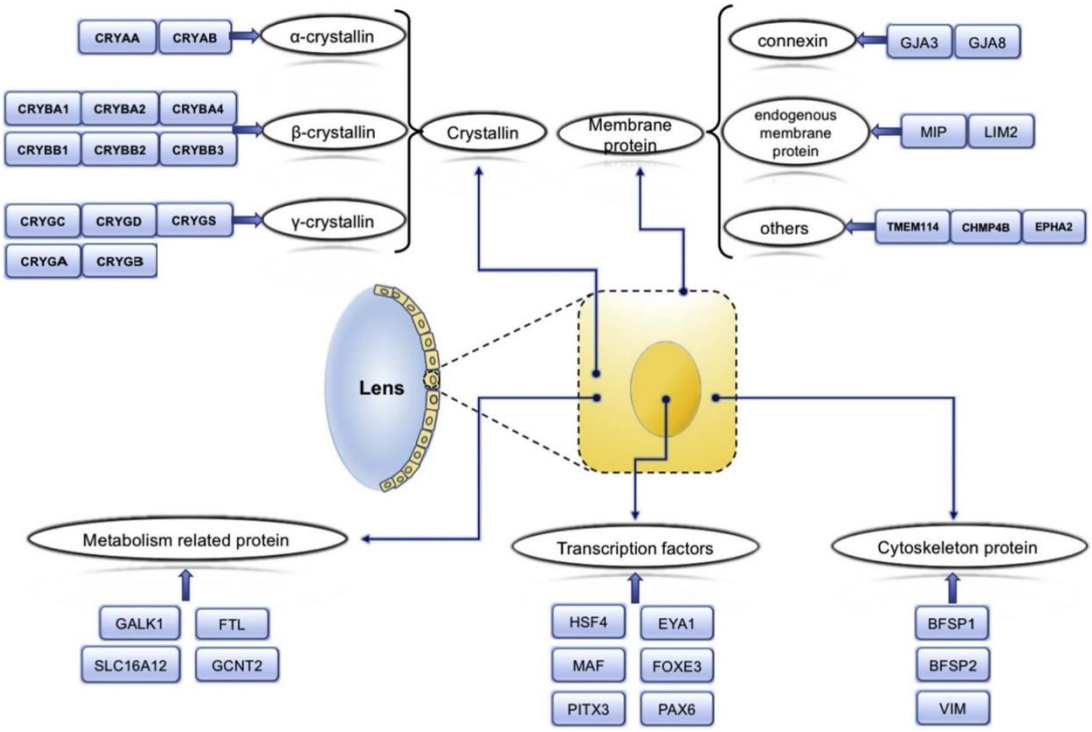


Figure 1.5 Genes associated with paediatric cataract

(Adapted from Zhu *et al.*, 2017[32])

1.3.1 Crystallin genes

Crystallins are the main protein component of the lens and provide both structure and refractive properties of the human lens[33]. They constitute about 90% of the water-soluble proteins in the lens[34]. There are 3 types of crystallins, the α, β, and γ crystallins, which are respectively encoded by the *CRYA*, *CRYB* and *CRYG* gene families.

α-crystallins account for about 40% of the total crystallins in the lens[35]. They play essential roles in maintaining the transparency and high refractive index of the lens. The large molecular weight (800-1000 kDa) and non-crystalline characteristics of the α-crystallin make it highly stable and soluble[35]. α-crystallins belong to the small heat shock protein family with chaperone-like activity and anti-apoptotic ability[36]. They can bind to denatured proteins to prevent the formation of light scattering aggregation precipitates[37]. Since the lens doesn't have any mechanism to dispose of the denatured proteins, this chaperone-like activity is essential for preserving lens transparency[35]. α-crystallins also play critical protective roles in the lens. Heat- and oxidative-induced stresses can lead to lens protein denaturation, hence resulting in opacities. It is reported that α-crystallin can inhibit the stress-induced apoptosis of lens epithelial cells[38,39].

β- and γ-crystallins are highly expressed in the lens as well, accounting for approximately 50% of the lens cytoplasmic proteins[40]. They exist in the lens fibre cells in an amorphous, dense liquid form, increasing the refractive index of the lens as well as keeping it transparent[41]. βγ-crystallins belong to one superfamily as they share a highly stable double-domain structure containing

a pair of tightly packed Greek key motifs[42]. Computer-based analysis suggests that Greek key motifs form intramolecular- and intermolecular-interdomain associations in γ-crystallins and β-crystallins, respectively[43]. The Greek key motifs pack the proteins densely in this way, which is essential for preventing light scattering. Mutations in βγ-crystallins show different pathogenic mechanisms from that of α-crystallins, mostly resulting in structural abnormalities by disrupting the protein folding. The unstable abnormal proteins aggregate and precipitate from the solution, consequently forming cataract[44].

To date, over 180 mutations in 13 crystallin genes have been reported to be implicated in the formation of paediatric cataract (Cat-Map[31] March 2021), including *CRYAA, CRYAB* (encoding α-crystallins); *CRYBA1, CRYBA2, CRYBA4, CRYBB1, CRYBB2*, and *CRYBB3* (encoding β-crystallins); and *CRYGA, CRYGB, CRYGC, CRYGD, CRYGS* (encoding γ-crystallins). Overall, crystallin mutations account for around 50% of non-syndromic familial cataracts with an identified genetic variant[45].

1.3.2 Membrane and cytoskeletal protein genes

There are two types of connexins (also known as gap-junction proteins) expressed in the lens: connexin46 (Cx46) and connexin50 (Cx50), with *GJA3* and *GJA8* encoding them respectively. Connexins are vital for nutrition transport and intercellular interaction in the avascular lens. They facilitate the transport of small molecules such as ions, nutrients and metabolic waste between neighbouring cells[46]. Due to these functions, connexins play indispensable roles in maintaining the metabolic homeostasis of the lens as

well as the transparency of the lens fibres[35]. More than 100 mutations in connexin genes have been reported to be implicated in the formation of paediatric cataracts (Cat-Map[31] March 2021).

The major intrinsic protein (MIP) is also known as Aquaporin 0 (AQP0), belonging to the aquaporin protein family (AQPs) of water channels. It is the most abundant membrane protein in the lens, representing 44.8% of the total membrane protein[32]. AQPs regulate the transmembrane permeation of water and other small molecules, forming microcirculation in the lens to nourish the central fibre cells as well as maintain the homeostasis of the lens[47,48]. Mutations in *MIP* lead to improper trafficking and decreased permeability of the AQP0, which leads to cataract formation[49-51]. So far, over 20 mutations in *MIP* have been discovered to be associated with autosomal dominant paediatric cataract (Cat-Map[31] March 2021).

Paediatric cataract related mutations are also detected in lens intrinsic membrane protein-2 (LIM2), which is a member of the claudin family of cell-junction proteins. The specific function of LIM2 is still unknown, but it is inferred to be involved in intercellular communication[32]. Four mutations in *LIM2* have been reported to cause paediatric cataract (Cat-Map[31] March 2021).

Mutations in a number of other genes that participate in membrane-associated signalling or transport processes are identified to cause paediatric cataract. These causative genes include *TMEM114* that encodes the transmembrane protein-114[52], *CHMP4B* that encodes the charge multivesicular protein-4B[53,54], and *EPHA2* that encodes the EPH receptor A2[55-57]. Mutations in these genes account for around 5% of non-syndromic familial cataract[31].

Cytoskeletal proteins are critical in providing structural support and maintaining lens transparency. Beaded filament structural proteins (BFSPs),

composed of BFSP1 (also called CP115) and BFSP2 (also called CP49), are unique lens-specific cytoskeletal proteins[58]. A previous study showed that beaded filament is involved in fibre cell differentiation and is necessary to maintain lens transparency[59]. However, the underlying mechanisms of *BFSPs* mutations leading to cataract formation remain to be elucidated. A total of 16 paediatric cataract-causing mutations were identified in *BSFP1* and *BFSP2* (Cat-Map[31] March 2021).

Another major cytoskeletal protein of the lens is vimentin, which is a type III intermediate filament protein encoded by the *VIM* gene. Three mutations in *VIM* are reported to be associated with paediatric cataract (Cat-Map[31] March 2021).

1.3.3 Transcription factor genes

Cataract-causing mutations have been found in several transcription factor-encoding genes, including *HSF4, MAF, PITX3, EYA1, FOXE3* and *PAX6* (Cat-Map[31] March 2021).

HSF4 is a heat shock protein transcription factor that regulates lens fibroblast differentiation and the expression of crystallins[32]. Two mechanisms may be involved in the formation of cataracts caused by the mutant HSF4. The expression of αB-crystallin is regulated by HSF4, hence cataracts induced by HSF4 disorder may be the result of the loss of chaperone-like activity of the abnormally expressed αB-crystallin[60]. In addition, mutations in HSF4 can prevent the expression of fibroblast growth factors and therefore delay the lens epithelial cell differentiation and lens fibre maturation[61]. This may also

contribute to cataract formation. Currently, more than 20 mutations in *HSF4* have been associated with paediatric cataract (Cat-Map[31] March 2021).

The *MAF* gene is implicated in the regulation of embryonic lens fibroblast development[62]. The *PITX3* gene regulates the expression of the paired-like homeodomain 3 protein, which plays a role in the formation of lens vesicles and the separation of ectoderm. *EYA1* is suggested to be involved in lens morphogenesis, as the protein is required for the survival of eye progenitor cells[63]. FOXE3 is a forkhead transcription factor that is specifically expressed in the lens. Mutations in the *FOXE3* gene have been reported to cause paediatric cataract[54,64]. The *PAX6* gene encodes the paired box 6 protein, which is vital for eye development as well as the expression of multiple lens crystallin genes[35]. A total of 37 paediatric cataract-related mutations have been reported in these genes (Cat-Map[31] March 2021).

1.3.4 Metabolism-related genes

Cataracts caused by metabolism-related genes are often accompanied by systemic metabolic diseases. *GALK1* and *SLC16A12* involved in glucose metabolism, and *FTL* and *GCNT2* involved in iron metabolism, are all implicated in cataractogenesis (Cat-Map[31] March 2021). The *GALK1* mutations lead to galactokinase deficiency, which results in paediatric cataract formation[65,66]. A mutation in *SLC16A12* was reported to cause juvenile cataract with microcornea and renal glucosuria[67]. Mutations in the iron-responsive element of the *FTL* gene lead to cataract-hyperferritinemia syndrome[68-70]. Paediatric cataracts induced by mutant GCNT2 are frequently associated with adult i blood-group phenotype[71-74].

1.3.5 Novel cataract genes remain to be discovered

Despite the considerable number of cataract genes identified with the application of high-throughput sequencing in the past decade, current gene panels still fail to provide molecular diagnoses in around 40% of children with inherited paediatric cataract[75-77]. This diagnostic gap may be caused by the inability of gene panels to detect structural or non-coding disease-causing mutations, but an alternative explanation is that current gene panels do not include all possible cataract causing genes. The latter indicates there are likely other unknown cataract-associated genes to be identified.

Traditional gene discovery approaches for inherited cataract utilise linkage analysis. However, prior to adding the novel identified genes to the genetic diagnostic panels, supplementary evidence is needed to confirm the role of these genes in paediatric cataract. The gold standard is to discover mutations in the same gene in multiple families, but this is difficult for paediatric cataract as it is very rare. An alternative strategy is to look for functional evidence of the novel genes in cataract formation using animal models. Functional evaluations of cataract genes have been successfully conducted in many animal models including zebrafish[78-81]. Jones *et al.* confirmed the pathogenicity of a novel mutation of *PGRMC1* in zebrafish models[82]. The mutation was identified by linkage analysis combined with whole-exome sequencing in an Australian family with paediatric cataract. Morpholino-induced *pgrmc1*-knockdown zebrafish developed significant cataracts, confirming the gene was causative. Similarly, a recent study

demonstrated the involvement of the *KPNA4* gene in cataractogenesis, as cataract phenotypes were observed in the *kpna4*-knockout zebrafish[83].

Once novel cataract genes are confirmed by functional analysis, including these genes in diagnostic panels could not only increase the success rate of molecular diagnosis, thereby ending the diagnostic odyssey and relieving parental guilt, but also improve the accuracy of genetic counselling regarding the risk of recurrence in future pregnancies. More importantly, identifying novel genetic causes of paediatric cataract will lead to a better understanding of the underlying mechanisms of cataractogenesis and promote the development of personalised medicine and new targeted therapies[84].

1.4 Phenotypic heterogeneity and genetic modifiers in paediatric cataract

Paediatric cataract is a highly heterogeneous disease. There are multiple reports of family members sharing identical rare mutations in the same cataract genes displaying considerable phenotypic variability, ranging from no cataract to severe paediatric cataract[85-90]. This intrafamilial phenotypic variation strongly suggests the presence of genetic modifiers in the development of cataract.

Genetic modifiers are defined as inherited allelic variations that do not determine whether a disease develops but influence the phenotypic expression of the primary disease-causing gene. Extensive evidence of genetic modifier effects has been reported in a wide range of diseases, such as retinitis pigmentosa[91,92], primary open-angle glaucoma (POAG)[93], Gaucher disease (GD)[94], long QT syndrome[95], cystic fibrosis[96,97], and Huntington's disease[98,99].

Genetic modifiers can affect any aspect of disease manifestation. such as the age of onset[93,98,100], severity[96,101], and penetrance[91,92,102]. The degree of modifier effects is also variable. Modifiers can slightly change the age of onset[98], or completely restore the normal condition[103]. The disease-modifying variants can be rare or common in the population, they can be located either in protein-coding regions or in regulatory regions, for example, untranslated regions (UTRs) or promoters[104].

The modification mechanisms are diverse as well. Modifiers can reduce or completely eliminate, enhance, or alter the function of disease-causing genes[105]. Modifier genes and primary causative genes can interact directly, or they can be molecularly independent but have similar biological effects and act on common or interactive pathways. The modifier effects can be generated by a single isolated gene or by a combination of multiple genes. Conversely, a single modifier gene can alter the phenotypic outcomes of one or more target genes or gene products.

Highly heterogeneous phenotypes, complex modification mechanisms, and the influence of environmental background make the identification of modifier genes extremely challenging, especially if multiple loci with small effects participate in the phenotypic modification. Despite ongoing research since the mid-20$^{\text{th}}$ century[106], our knowledge of modifier genes remains limited, particularly in rare Mendelian diseases. To date, two modifier genes of the congenital cataract have been identified in mouse models. The phosphodiesterase 6B, cGMP, rod receptor, beta polypeptide encoding gene (*Pde6b*) is reported to increase the severity of cataractogenesis of *Foxe*$^{\text{rct}}$ mice[100,107]. Matteson *et al.* discovered that the Forkhead box E3 gene (*Foxe3*) modulates the incidence of congenital cataracts caused by a mutation in the G protein-coupled receptor 161 gene (*Gpr161*) in the *vacuolated lens* (*vl*)

mice[108]. However, such cataract-modifying genes have not been identified in humans yet.

1.5 Overall hypothesis and aims

Although hundreds of genes and loci associated with paediatric cataract have been identified[31], genetic testing of these genes can only detect mutations in 30-70% of patients with inherited cataract[75-77]. The absence of molecular diagnoses in the remaining patients suggests there are likely more causative genes for paediatric cataract to be identified. In addition, as pointed out in Section 1.4, the modifier genes responsible for the phenotypic variation of paediatric cataract remain to be revealed. Therefore, the overall hypothesis of this thesis is:

There are more paediatric cataract causative genes to be identified in the unsolved families, and variants in undiscovered genes modify phenotype severity in families with mutations in well-characterised cataract genes

To test the hypothesis, the following specifical aims were set:

Aim 1: Determine the role of the candidate causative gene *HTR1F* in cataract formation

Our team previously identified a novel putative disease-causing gene, *HTR1F*, in an Australian family with paediatric cataract[109]. I performed the functional evaluation of this gene by knocking it out in zebrafish models using the CRISPR-Cas9 system. This study is described in Chapter 3.

Aim 2: Discover genetic modifiers that modulate the severity of paediatric cataract

Significant variation in disease severity was observed within two large paediatric cataract families with identified causative mutations[85,89,90]. To investigate the potential for genetic modifiers in these families, I first updated the clinical phenotype information of each family and validated the previously reported causative genes using the combination of parametric linkage analysis and whole-genome sequencing (WGS). I then applied variance components linkage analysis of disease severity to map potential modifier loci. The causative gene validation and modifier loci localisation are described in Chapter 4 and Chapter 5, respectively.

Chapter 2

General materials and methods

2.1 Materials

2.1.1 Reagents and equipment

The reagents and equipment utilised in this thesis are listed in Table 2.1 and Table 2.2.

Table 2.1 List of reagents

Item	**Supplier**	**Catalogue No.**
TAE buffer (Tris-acetate-EDTA) (50X)	Thermo Fisher Scientific	B49
Methylene Blue	AQUASONIC	PL084
NaCl, KCl, $MgSO_4$, KH_2PO_4, Na_2HPO_4, $CaCl_2$, $NaHCO_3$, KCl, NaOH	Sigma-Aldrich	-
PTU (1-phenyl-2-thiourea)	Sigma-Aldrich	P7629
Alt-R® CRISPR-Cas9 crRNA	Integrated DNA Technologies (IDT)	-
Alt-R® CRISPR-Cas9 tracrRNA	IDT	1072533
Alt-R® S.p.HiFi Cas9 Nuclease V3	IDT	1081060
IDTE pH 7.5 (1X TE Solution)	IDT	11-01-02-02
1M HEPES buffer	Sigma-Aldrich	H3375
QIAamp DNA blood maxi kit	Qiagen	51194

Nucleon BACC3 kit	Amersham Pharmacia Biotech	GERPN8512
PureGene DNA isolation kit	Gentra Systems	158445
Illustra GenomiPhi V2 DNA Amplification Kit	GE Healthcare	25-6600-31
Qubit™ dsDNA BR Assay kit	Thermo Fisher Scientific	Q32853
Qubit™ dsDNA HS Assay kit	Thermo Fisher Scientific	Q32854
Agarose powder	Bioline	BIO-41025
SYBR™ Safe DNA Gel Stain in 1x TAE	Thermo Fisher Scientific	S33112
DNA Loading Buffer Blue (5x)	Bioline	BIO-37045
HyperLadder™ 100bp	Bioline	BIO-33030
AMPure XP beads	Beckman Coulter	A63880
BigDye™ Terminator v3.1 Cycle Sequencing Kit	Thermo Fisher Scientific	43-374-55
BrightDye® Terminator Cycle Sequencing Kit	MCLAB	BDT3-100
CleanSEQ Dye-Terminator Removal kit	Beckman Coulter	A29151
POP-7™ Polymer	Applied Biosystems	4393714
Phire Tissue Direct PCR Master Mix	Thermo Fisher Scientific	F170s
Nuclease-Free Duplex buffer	IDT	11-01-03-01
RNeasy® Plus Mini Kit	Qiagen	74104
SuperScript® III First Strand synthesis system	Thermo Fisher Scientific	12574026
Tricaine methanesulfonate	Glentham Life Sciences	GE5936
RNA*later*™ RNA Stabilisation Reagent (50ml)	Qiagen	NC9786752
Ethanol	Sigma-Aldrich	E7023
Nextera XT Index kit v2 Set A (96 Indices, 384 Samples)	Illumina	FC-131-2001
Q5® Hot Start High-Fidelity 2X Master Mix	New England BioLabs	M0494S
Buffer EB (250ml)	Qiagen	19086
High Sensitivity D1000 ScreenTapes	Integrated Sciences	5067-5582

High Sensitivity D1000 Reagents (Sample Buffer & Ladder)	Integrated Sciences	5067-5583
Miseq Reagent Kit v3 (600-cycle)	Illumina	MS-102-3003
Infinium® Global Screening Array-24 v1.0 (GSA) kit	Illumina	20005132
TruSeq Nano Library Prep v2.5	Illumina	20015964

Table 2.2 List of equipment

Item	**Supplier**	**Catalogue No.**
Standard microcentrifuge tubes, 1.5ml	Labcon	#3039-560-000
RNase-free microfuge tubes, 1.5ml	Thermo Fisher Scientific	AM12400
Falcon™ 15ml conical centrifuge tubes	Thermo Fisher Scientific	14-959-53A
Filtered sterile pipette tips	Labcon	-
MicroAmp™ Optical 96-well reaction plate	Thermo Fisher Scientific	N8010560
Desktop microcentrifuges	Bioline Global	SF7000
Vortex mixer	Ratek	IC-VM1
Block heater	Ratek	DBH200
Qubit® 2.0 Fluorometer	Thermo Fisher Scientific	Q32866
Veriti™ 96-well Thermal cycler	Applied Biosystems	4375786
3500 Genetic Analyser	Applied Biosystems	4405673
8-Capillary Array	Applied Biosystems	4404685
Amersham Imager 600	GE Healthcare	450N120051
Capillary glass (1.0mm x 0.58 mm x 10 cm)	Harvard Apparatus	GC100F-10
Micropipette puller	Sutter Instrument	P-87
Pico-liter injector	Warner Instruments	PLI-10
Petri dish	NEST Scientific	752001

Nunc™ glass base dish (17mm)	Thermo Fisher Scientific	150682
Incubator	Nüve	EN025
Stereomicroscope	Olympus	SZX16
Inverted microscope	Nikon	Eclipse Ti
S-Series compact USB 2.0 camera	Mightex	SCE-B013-U
sCMOS camera	Andor	Zyla 4.2 PLUS
PowerPac™ basic power supply	Bio-Rad	1645050
Magnetic plate	ALAPAQUA	A001322
3500 Genetic Analyzer	Thermo Fisher Scientific	A27772
4200 TapeStation Loading Tips	Integrated Sciences	5067-5599
4200 TapeStation system	Integrated Sciences	G2991BA
MiSeq sequencer	Illumina	SY-410-1003
HiScan System	Illumina	SY-103-1001
Novaseq 6000 system	Illumina	20012850

2.1.2 Reagent setup

2.1.2.1 1x TAE electrophoresis working solution

TAE buffer (Tris-acetate-EDTA) (50X) was diluted with sterile Milli-Q water to 1x TAE electrophoresis working solution, and stored at room temperature (18-22°C) for up to 6 months.

Component	**Amount**
TAE buffer (Tris-acetate-EDTA) (50X)	20ml
Milli-Q H_2O	980ml
Final volume	1L

2.1.2.2 E2 embryo media with Methylene blue

The 0.5x E2 with 0.5mg/l Methylene Blue working solution was prepared following the fish nursey protocol[110] provided by the Zebrafish International Resource Center. The final concentrations of E2 media components are: 7.5mM NaCl, 0.25mM KCl, 0.5mM $MgSO_4$, 75µM KH_2PO_4, 25µM Na_2HPO_4, 0.5mM $CaCl_2$, and 0.35mM $NaHCO_3$.

2.1.2.3 1x PTU (200µM) egg water

0.076g PTU (1-phenyl-2-thiourea) were added to 50ml autoclaved Milli-Q water to prepare a 50x PTU (10,000µM) stock solution. Fresh 1x PTU (200µM) egg water was prepared each time as needed by diluting 50x PTU stock solution 50 times with E2 embryo media with Methylene Blue.

2.1.2.4 Alt-R CRISPR-Cas9 crRNA, tracrRNA and Cas9 nuclease aliquots

The Alt-R CRISPR-Cas9 crRNA and tracrRNA were resuspended in IDTE buffer to final concentrations of 100µM each, and stored according to supplier instructions. At the time of these experiments, the crRNA, tracrRNA and S.p.HiFi-Cas9 Nuclease were aliquoted into 2µl small stocks and stored at -20°C to reduce unnecessary freeze-thaw cycles.

2.1.2.5 Cas9 working buffer (20mM HEPES; 150mM KCl, pH 7.5)

The Cas9 working buffer was prepared by mixing the following components, and stored at room temperature (18-22°C) for up to 6 months.

Component	Amount
1M HEPES buffer	300µl
KCl powder	0.168g
RNase-free H_2O	14.7ml
Final volume	15ml

2.1.2.6 0.2N NaOH

Add 0.2g NaOH into a 15ml falcon tube, fill the tube up to 5ml with Milli-Q water to make 1N NaOH. Add 200µl of the 1N NaOH to 800µl Milli-Q water to dilute to 0.2N NaOH.

2.2 Ethics statement

Studies in this project were conducted according to the guidelines of the Declaration of Helsinki and followed the National Health and Medical Research Council statement of ethical conduct in research involving humans (2007, updated 2018). The project was approved by the Tasmania Health and Medical Human Research Ethics Committee (H0014539), the Southern Adelaide Clinical Human Research Ethics Committee (3-07) and the Royal Victorian Eye and the Ear Hospital Human Research Ethics Committee (01/432H/11). Informed consent was obtained from all participants, or a

parent/guardian for minors under 18 years of age. The animal research was approved by the University of Tasmania Animal Ethics Committee (project number: A0017743) in accordance with the Australian National Health and Medical Research Council Code of Practice for the Care and Use of Animals for Scientific Purpose.

2.3 Study participants

Family CRCH13 was identified through a database maintained by the Royal Children's Hospital, Melbourne, Australia and the Royal Victorian Eye and Ear Hospital, Melbourne, comprising paediatric cataract patients from south-eastern Australia with any type of lens opacity[10]. Probands in family CSA91 and CSA110 were identified from patients presenting to the eye clinic at Flinders Medical Centre (FMC), Adelaide, South Australia, or the Women's and Children's Hospital, Adelaide, South Australia. Detailed family histories were obtained, and additional affected and unaffected family members were invited to participate in the project. All participating family members were examined by multiple ophthalmologists. All affected individuals in the project were definitively diagnosed with paediatric cataracts. All affected individuals have bilateral cataracts unless otherwise noted. Patients over 40 years old were diagnosed with paediatric cataract because they had paediatric cataract phenotypes (e.g., fetal lamellar cataract, pulverulent cataract) or carried rare paediatric cataract variants that were shared within the families. The status of individuals with age-related cataracts was set to unaffected. DNA of the consenting participants was extracted from the whole blood with QIAamp DNA blood maxi kit (Table 2.1) or Nucleon BACC3 kit (Table 2.1), or buccal mucosal swabs with PureGene DNA isolation kit (Table 2.1)[85,89,90].

2.4 DNA preparation

2.4.1 Amplification

Genomic DNA samples with limited stocks (volume <10µl and concentration <50ng/µl) were whole-genome amplified using Illustra GenomiPhi V2 DNA Amplification Kit (Table 2.1). Approximately 20ng of template DNA was mixed with the sample buffer to a final volume of 10ul. The template sample was denatured by heating to 95°C for 3 minutes then cooling to 4°C on ice. A master mix containing 9µl of reaction buffer and 1µl of enzyme mix was added to the cooled sample, and the mixture was incubated at 30°C for 1.5 hours and heated to 65°C for 10 minutes then cooled to 4°C. The amplification reactions were stored at 4°C.

2.4.2 Quantification

Qubit™ dsDNA BR Assay kit (Table 2.1) was utilised to quantify the concentration of patient DNA and whole-genome amplified DNA. Barcoded PCR products generated from next-generation sequencing (NGS) (Section 3.3.8.2) were quantified using Qubit™ dsDNA HS Assay kit (Table 2.1). For both kits, Qubit reagents were 1:200 diluted in the Qubit buffer to prepare the working solution. 1µl of DNA sample was added to 199µl working solution and incubated at room temperature (18-22°C) for 2-3 minutes prior to being loaded to Qubit 2.0 fluorometer (Table 2.2). The sample concentration was automatically calculated by the Qubit 2.0 fluorometer.

2.5 Target sequence amplification

2.5.1 Primer design

Primers used in this thesis were designed using Primer-BLAST[111] on NCBI (National Center for Biotechnology Information) website. Primers were synthesised by Sigma-Aldrich or Integrated DNA Technologies (IDT) and diluted to 10µM or 3.3µM stocks as needed. The primer stocks were kept at 4°C. All the primers utilised in this project are listed in Appendix 1.

2.5.2 Standard PCR and Reverse transcription-PCR

Standard PCR (Polymerase chain reaction) and reverse transcription (RT)-PCR were performed using MyTaq™ HS Mix (Table 2.1). The reactions were run on Veriti™ 96-well Thermal cyclers (Table 2.2). Reaction components and cycling conditions are listed in Table 2.3.

Table 2.3 Reaction components and cycling conditions of standard PCR and RT-PCR

Component	Volume (µl)
MyTaq™ HS Mix	10
Milli-Q water	6
Forward primers (10µM)	1
Reverse primers (10µM)	1
DNA/cDNA (20µg/µl)	2
Final volume	20

Cycle step	Temperature (°C)	Time	Cycles
Initial denaturation	95	1min	1
Denaturation	95	10s	
Annealing	57[1]	15s	30[2]
Extension	72	15s	

The annealing temperature[1] and cycle number[2] were optimized when required.

2.6 Agarose gel electrophoresis

2% w/v agarose gels were used to visualise the PCR products. 0.6g agarose powder (Table 2.1) was mixed with 30ml SYBR™ Safe DNA Gel Stain (Table 2.1) in 1x TAE (Section 2.1.2.1) and microwaved for 1 minute to dissolve the agarose before pouring into the mould. 4µl of PCR products were loaded onto the gel with 1µl DNA Loading Buffer Blue (5x) (Table 2.1) and 2µl of 100bp HyperLadder™ (Table 2.1) was loaded alongside the samples as an indicator to estimate the size of the PCR products. Gels were electrophoresed in 1x TAE buffer (Section 2.1.2.1) at 80V for 30-80 minutes. The electrophoresis time was adjusted according to the size of the PCR products. Gels were imaged on the Amersham Imager 600 (Table 2.2) after the electrophoresis.

2.7 Sanger sequencing

Prior to being sequenced, the amplified PCR products were purified using AMPure XP beads (Table 2.1) according to the instructions provided by the manufacturer. The volume ratio of PCR products to be purified and the magnetic beads was 1:1.8. Sequencing reactions were conducted on Veriti™ 96-well Thermal cyclers using BigDye™ Terminator v3.1 Cycle Sequencing

Kit (Table 2.1) or BrightDye® Terminator Cycle Sequencing Kit (Table 2.1). The reaction components and cycling conditions are listed in Table 2.4. Sequencing reactions were purified using the CleanSEQ Dye-Terminator Removal kit (Table 2.1) following the manufacturer's instructions. 10µl purified sequencing reaction was denatured at 95°C for 5 minutes on Veriti™ 96-well Thermal cycler, then sequenced on a 3500 Genetic Analyser (Table 2.2) with POP-7™ Polymer (Table 2.1) and 8-Capillary Array (Table 2.2).

Table 2.4 Reaction components and cycling conditions of Sanger sequencing

Component	**Volume (µl)**
BigDye™ Terminator/ BrightDye® Terminator	0.25
Milli-Q water	6
Sequencing buffer	1.75
Forward/reverse primers (3.3µM)	1
Purified PCR product	1
Final volume	10

Cycle step	**Temperature (°C)**	**Time**	**Cycles**
Initial denaturation	96	1min	1
Denaturation	96	10s	
Annealing	50	5s	25
Extension	60	1min15s/4min[1]	

[1]1min15s was applied when purified PCR product is less than 500bp, and 4min was applied when the product is 500-1000bp.

Chapter 3

Investigation of the role of *HTR1F* in cataract formation

The methods development sections within this chapter have been published in *Methods* in 2021. The article, entitled 'Zhao D, Jones JL, Gasperini RJ, Charlesworth JC, Liu GS, Burdon KP. Rapid and efficient cataract gene evaluation in F0 zebrafish using CRISPR-Cas9 ribonucleoprotein complexes'[112], is included in this thesis as Appendix 2.

3.1 Introduction

A promising candidate gene *HTR1F* was previously identified in an Australian cataract family[109]. The pathogenicity of this novel potential cataract-causing gene needs to be verified. In this chapter, we investigated the biological effect of *HTR1F* and introduced a rapid and efficient method to evaluate cataract candidate genes in F0 zebrafish.

3.1.1 Family CSA92 background

Family CSA92 (Figure 3.1) from South Australia presented with autosomal dominant paediatric cataracts. The cataract phenotype in this

family varies significantly, ranging from very faint fetal lamellar cataracts that were undiagnosed until the family was examined for this research, to severe cataracts that were removed with intraocular lens (IOL) implantation surgery after diagnosis (Table 3.1). All family members were examined by multiple ophthalmologists to confirm the variable phenotype. The lens photos of CSA92.07 are shown in Figure 3.2 as an example of the cataract phenotypes.

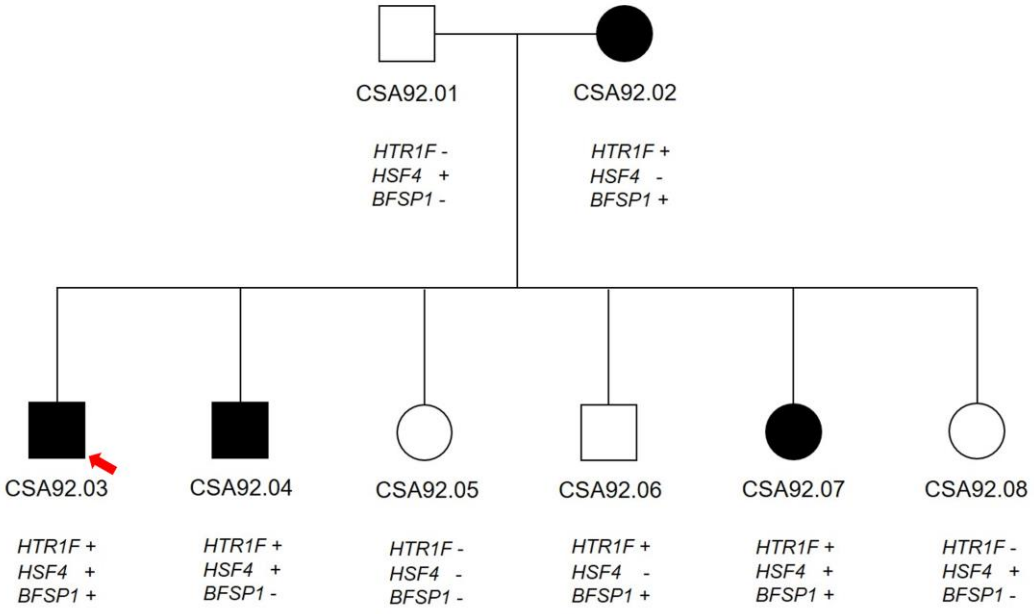


Figure 3.1 The pedigree of family CSA92

Males are indicated by squares and females by circles. Solid black symbols indicate affected individuals while open symbols indicate unaffected individuals. "+" indicates gene with mutations and "-" indicates gene with wild-type alleles. The red arrow indicates the proband.

Table 3.1 Family CSA92 phenotype and clinical information

Affected member	Cataract phenotype	Age of diagnosis	Age of surgery (right/left eye)	Age of recruitment
CSA92.02	Very faint fetal lamellar	42 years	No surgery at recruitment	42 years
CSA92.03	Bilateral IOL	4 years	Both eyes: 4 years	22 years
CSA92.04	Fetal nuclear lamellar, some central opacity	-	No surgery at recruitment	20 years
CSA92.07	Nuclear lamellar cataracts	4 years	22 years/4 years	14 years

IOL: intraocular lens implant. - : data unavailable. (Amended from Javadiyan, 2016[109])

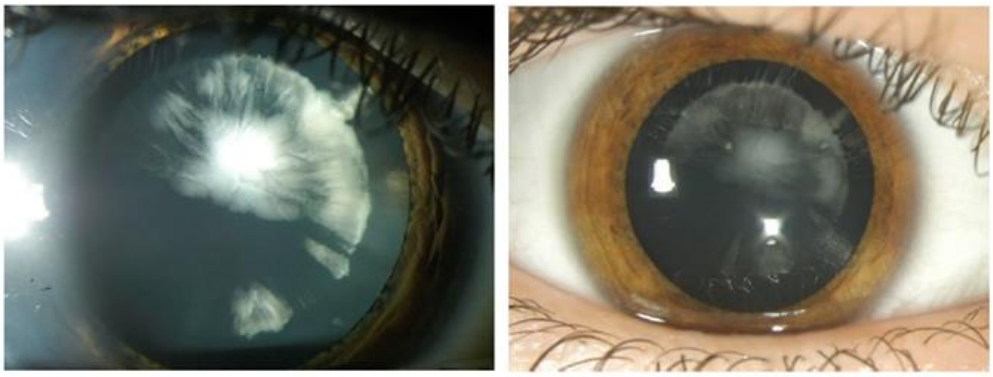


Figure 3.2 Lens photos of CSA92.07 showing nuclear lamellar cataract

(Photos from Javadiyan, 2016[109])

3.1.2 Candidate variants identified in CSA92

This work was undertaken as part of the PhD thesis of Dr Sharhbanou Javadiyan[109]. Through direct sequencing, none of 51 known paediatric cataract genes was identified as the causative gene in family CSA92[113]. Although rare potentially pathogenic variants in both *HSF4* (c.636G>T (p.M212I), rs199742128) and *BFSP1* (c.736A>G (p.T246A), rs143865632) were detected in this family, neither of them completely segregated with disease (Figure 3.1). The affected member CSA92.02 does not carry the variant in *HSF4* and another affected individual CSA92.04 does not carry the variant in *BFSP1.* Whole-exome sequencing (WES) of selected individuals (CSA92.01, CSA92.02, CSA92.03, CSA92.04 and CSA92.05) was performed to search for a potential novel disease-causing gene.

A novel c.865C>T (p.R289W) variant (rs767724180) in the *HTR1F* gene (NM_000866.5), which was the only rare coding variant segregating in the exome sequences of CSA92, was identified by Javadiyan[109]. The variant was

detected in all affected individuals but also presented in an unaffected individual CSA92.06, who had no clinically significant cataract at the age of 17. This variant is extremely rare, with a MAF (minor allele frequency) of $2x10^{-5}$ (5/250586 alleles) in the gnomAD database[114] (v2.1.1). In addition, given the phenotype in family CSA92 can be very mild and variable, and those gnomAD cohorts would not have been screened for this phenotype, it is completely feasible that gnomAD would contain participants with a cataract caused by this variant.

There is no report of the *HTR1F* c.865C>T (p.R289W) variant in public databases of clinical variants, but multiple pieces of evidence suggest it may have an impact on the protein function. The variant has a CADD[115] score of 27.1, indicating that it is in the top 1% of most deleterious substitutions in the human genome. It alters the amino acid at position 289 from Arginine to Tryptophan. Compared to the wild-type Arginine, the mutant Tryptophan is bigger and more hydrophobic, and the charge changes from positive to neutral. The increase in hydrophobicity can result in loss of hydrogen bonds and/or disturb correct folding. Loss of charge can cause loss of interactions with other molecules or residues. Moreover, the amino acid substitution is predicted to be deleterious and probably damaging by predictive tools SIFT[116] and PolyPhen-2[117], respectively. Conserved Domain Database (CDD)[118] indicates the variant locates in the 7tm_1 (pfam00001) conserved domain, which is important for the main activity of the protein. Collectively, this variant may affect the structure and function of the HTR1F protein as it changes the properties of the amino acid located in the conserved domain.

3.1.3 *HTR1F*

The *HTR1F* gene encodes the 5-hydroxytryptamine receptor 1F (HTR1F) protein. GTEx Portal[119] indicates that *HTR1F* is ubiquitously expressed in most mammalian tissues, and we have detected *HTR1F* expression in the human lens (Figure 3.3). The HTR1F protein is a G-protein coupled serotonin receptor. Although the exact mechanism regarding how serotonin is involved in cataractogenesis remains obscure, it has been shown to cause cataracts in both high and suppressed levels[120-123].

High concentrations of serotonin have been found in the aqueous humour of cataract patients[124], and have been shown to lead to dense cataracts in rats[120]. Selective serotonin reuptake inhibitors (SSRIs) increase serotonin levels by blocking its reuptake. Several case-control studies have reported that SSRI use increases the risk of cataract formation[125-127]. In addition, a teenager using SSRIs was reported to develop bilateral cataract[121]. It is hypothesised that high levels of serotonin reduce the blood flow to the ciliary body and anterior chamber, resulting in metabolic disorders of the lens, which leads to cataracts[120]. Conversely, serotonin antagonists also have been shown to cause cataracts in rats[122,123]. Melatonin is the downstream product of serotonin, synthesized in both the eyes and the pineal gland[128]. It is an antioxidant and free radical scavenger in the ocular lens and plays a protective role against cataract formation[129,130]. Hence the mechanism by which serotonin antagonists induce cataract formation may be the reduction of the sequentially synthesized melatonin, thereby triggering oxidative stress-induced cataractogenesis.

Taken together, serotonin is demonstrated to be crucial in lens transparency. Therefore, although there is no existing evidence in the literature that the *HTR1F* gene is associated with cataract or other eye phenotypes, its

protein product HTR1F, as the receptor of serotonin, could be involved in cataract formation.

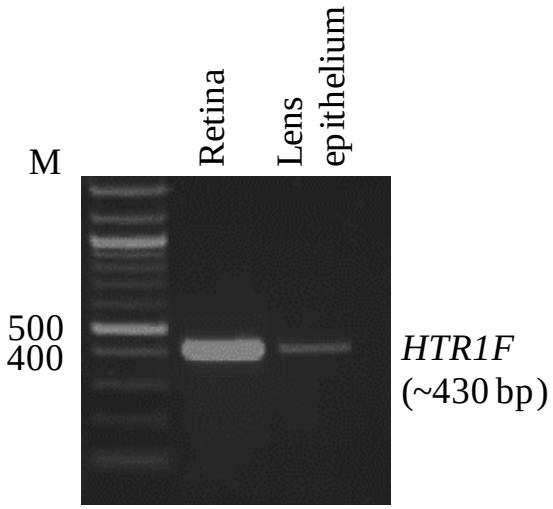


Figure 3.3 RT-PCR showing *HTR1F* expression in human lens

(The RT-PCR work was undertaken by Dr Sharhbanou Javadiyan (Department of Ophthalmology, School of Medicine, Flinders University))

3.1.4 Zebrafish model

To confirm the identified novel gene is the cause of a disease, the gold standard is to identify similar variants in the same gene in multiple affected families. However, since paediatric cataract is such a rare and heterogeneous disease, it can take many years to identify multiple families harbouring variants in the same gene to meet the burden of evidence. In this situation, seeking additional biological evidence of the candidate gene that involves in cataract formation becomes particularly important.

Genetically modified models that display cataract formation provide strong evidence that a candidate gene is important in lens biology and cataractogenesis. Several animal models have been used in cataract research,

including mice, rats, rabbits, dogs, and zebrafish[78-81,131]. However, the relatively small numbers of offspring and long generation period limit the application of mammalian models. In contrast, zebrafish (*Danio rerio*) have been increasingly used in eye research over the past few years[132].

Many factors make zebrafish a remarkable model for ocular research. A large number of zebrafish can be easily maintained at a low cost. The relatively short production time (2-4 months) and a large clutch of offspring (200-300 eggs from a single mating pair on a weekly basis) make them ideal models for genetic studies compared to mouse models. Zebrafish embryos develop outside of the female and remain transparent for the first few days of development, providing technical advantages for embryo microinjection. In particular, the eyes of the zebrafish are relatively large compared to the overall body size[133] and become functional as early as 3 day-post-fertilization (dpf)[134], giving them distinct advantages over other mammalian model organisms in eye research. The zebrafish ocular system is highly similar to that of humans both anatomically and functionally[135-137]. Additionally, the zebrafish genome has been fully sequenced and annotated, and around 70% of human genes have functional orthologs in zebrafish[138]. This high level of homology to humans means that most human genes can be interrogated in zebrafish and the functions of those genes are largely similar between the two species.

With these outstanding traits, zebrafish has become a popular model organism for ocular studies, especially lens research. In recent years, several cataract genes, such as crystallin genes[131,139], *HSF4*[140], *Apq0*[141-143], *FOXE3*[79] and *UNC45B*[144], have been successfully investigated in the zebrafish model, providing the feasibility of zebrafish models in cataract research.

3.1.5 CRISPR-Cas9

Clustered regularly interspaced short palindromic repeat (CRISPR)-CRISPR-associated system (Cas)9 has become a widely used targeted gene-editing technology in zebrafish models. The system is comprised of guide RNA (gRNA) and Cas9 nuclease (Figure 3.4A), which are combined to form a ribonucleoprotein (RNP) complex. The gRNA consists of a target-specific CRISPR RNA (crRNA) and a universal trans-activating RNA (tracrRNA). The crRNA is designed to be complementary to the target DNA site, a 20 base pair sequence next to a protospacer adjacent motif (PAM)[145]. The gRNA forms a complex with the specific target site, directing the Cas9 nuclease to that locus. Once at the appropriate site, the Cas9 nuclease specifically recognizes the PAM sequence and produces double-strand breaks (DSBs). There are two mechanisms to repair DSB in vertebrate cells: non-homologous end joining (NHEJ) and homologous recombination (HR) (Figure 3.4B). The error-prone NHEJ can produce insertions or deletions between the end of the broken strands, resulting in frame-shifts or loss of function mutations (knockout). The frequency of HR is much lower than that of NHEJ. This mechanism uses template sequences to repair the DSBs and the templates can be programmable sequences provided by geneticists. Therefore, through HR, sequences of interest can be inserted into the target genomic sites (knockin).

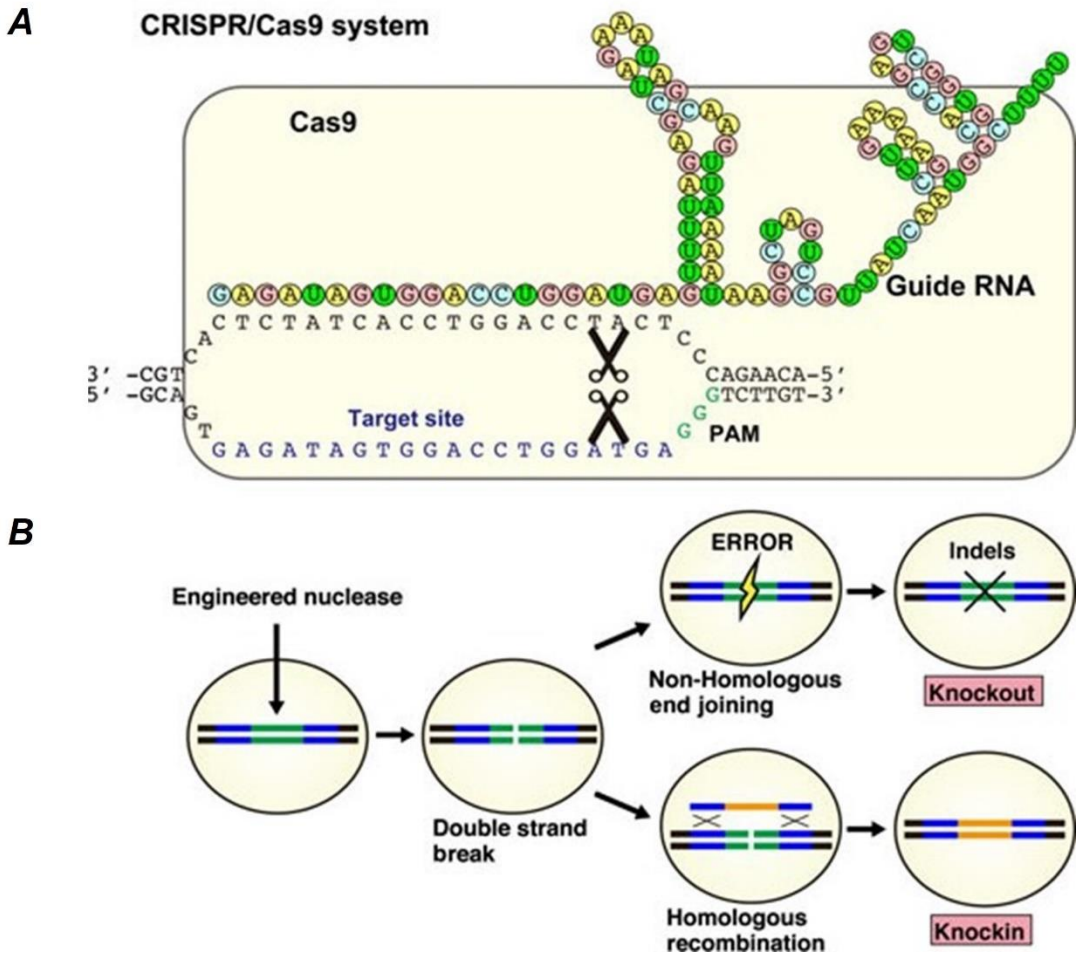


Figure 3.4 The structure and functional mechanism of the CRISPR-Cas9 system

A. The CRISPR/Cas9 system. CRISPR/Cas9 consists of a guide RNA and the Cas9 nuclease. The guide RNA recognizes sequences (20 bases) to the genomic target adjacent to the protospacer-adjacent motif (PAM) site of NGG (N: any nucleotide). CRISPR/Cas9 induces DSBs in the targeted genomic locus. **B.** DSBs can be repaired by nonhomologous end-joining (NHEJ) or homologous recombination (HR). In NHEJ, both broken ends are rejoined with a high frequency of insertions and/or deletions (indels), resulting in frameshift-mediated gene disruption (knockout). HR is an error-free repair pathway that uses homologous templates. Sequences of interest can be set as templates and inserted into target sites (knockin). (Image adapted from Hisano *et al.*, 2014[146])

Multiple advantages make the CRISPR-Cas9 system an outstanding genetic manipulation technique. Its advantages over other methods such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) include easy design processes and rapid synthesis of reagents.

Morpholinos (MOs) have been used extensively for the temporary knockdown of genes in zebrafish, but CRISPR-Cas9 provides a permanent genetic modification with relatively low off-target effects. Moreover, the CRISPR-Cas9 system can perform both knockout and knockin, and the gene modification can be transmitted through generations. Due to these excellent features, CRISPR-Cas9 has been widely and successfully used in disease investigation[145,147-150]. The application of using CRISPR technology to investigate paediatric cataract in zebrafish is under development. Recently, four studies have successfully used CRISPR-Cas9-edited zebrafish to demonstrate the critical roles of the disease genes in cataract formation. Vorontsova *et al.* used CRISPR-Cas9 to generate loss-of-function deletions in zebrafish *aqp0a* and/or *aqp0b* and illuminated that only *aqp0a* is essential for lens transparency[143]. Krall and colleagues demonstrated *foxe3* plays an important role in cataractogenesis using CRISPR-Cas9-mediated mutant fish line[79]. Mishra *et al.* disrupted the αB-crystallin genes utilizing the CRISPR-Cas9 system and showed that αB-crystallin is vital to maintaining lens transparency[131]. In another study, Ping's team injected CRISPR-Cas9 into zebrafish to knock out *kpna4* and identified this gene is involved in cataract formation[83]. These studies demonstrated the feasibility of employing CRISPR technology to investigate paediatric cataract in the zebrafish model.

3.2 Hypothesis and Aim

Hypothesis:

HTR1F is the causative gene for cataracts in family CSA92.

Aim:

Determine if *HTR1F* causes cataracts by knocking out this gene in zebrafish models.

3.3 Methods

3.3.1 Overall workflow

Figure 3.5 illustrates the overall experimental workflow.

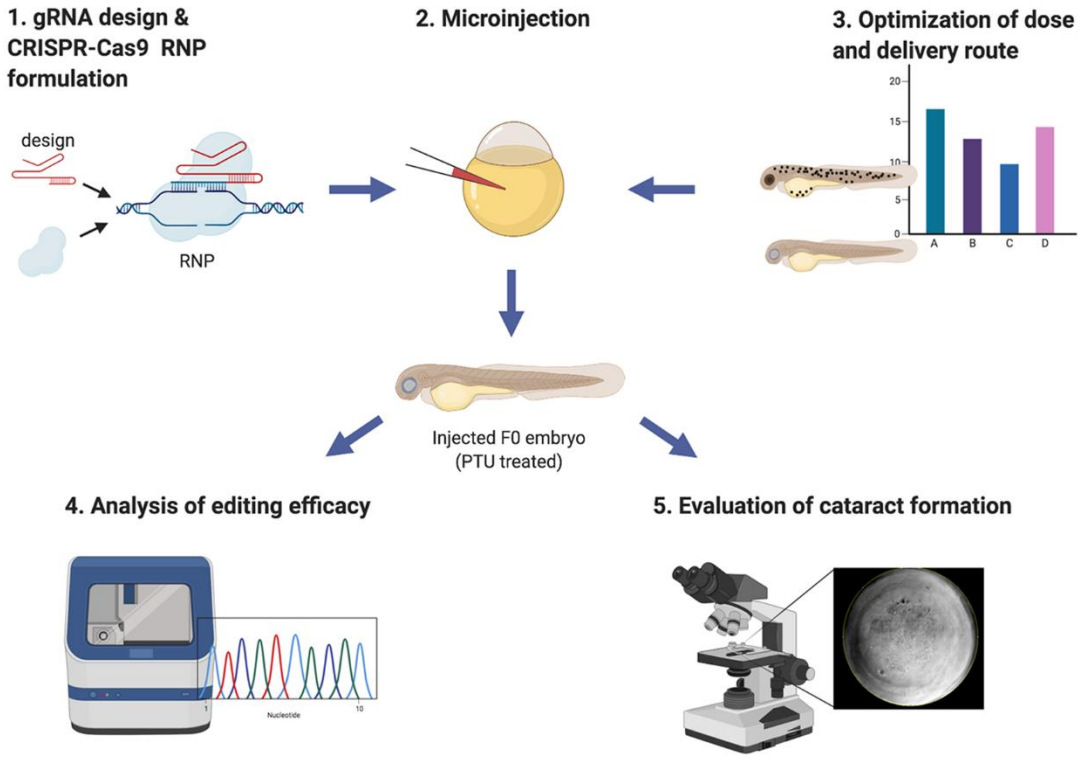


Figure 3.5 Overview of the key steps in Chapter 3

3.3.2 Zebrafish maintenance and breeding

AB strain zebrafish[151] were housed in a circulating water system under standard conditions[152], with a daily cycle of 14 hours light/10 hours dark.

Embryos were obtained from the natural pairwise spawns of the wild-type AB zebrafish and were collected and raised in an incubator (Table 2.2) at 28.5°C in E2 media with Methylene Blue (Section 2.1.2.2) until 5dpf.

3.3.3 Zebrafish genomic DNA extraction and amplification

The genomic DNA was extracted and amplified using the Phire Tissue Direct PCR Master Mix kit (Table 2.1) according to the Dilution & Storage protocol provided by the manufacturer. Briefly, individual larvae were placed into 20µl Dilution Buffer and 0.5µl DNARelease Additive to release the genomic DNA. The reaction was incubated at room temperature for 2 minutes and heated at 98°C for 2 minutes. The remaining tissue in the reaction solution was spun down and the supernatant was used for amplification or stored at -20°C if not used immediately. 1ul of 1:10 diluted supernatant was amplified using the components and cycling conditions listed in Table 3.2. The Tm values of primers and the annealing temperature were determined using the Tm calculator[153] on the Thermo Fisher Scientific website. The 2-step protocol was used when primer Tm values are higher than 69°C, otherwise the 3-step protocol was applied.

Table 3.2 Reaction components and cycling conditions of zebrafish genomic DNA amplification

Component	Volume (µl)
2x Phire Tissue Direct PCR Master Mix	10
Milli-Q water	7
Forward primer (10µM)	1
Reverse primer (10µM)	1
Sample	1
Final volume	20

The product information is listed in Table 2.1.

Cycle step	2-step		3-step		Cycles
	Temp. (°C)	Time	Temp. (°C)	Time	
Initial denaturation	98	5min	98	5min	1
Denaturation	98	5s	98	5s	35-40[2]
Annealing	-	-	x[1]	5s	
Extension	72	20s	72	20s	

Temp.: temperature.
[1]Annealing temperature is determined by the Tm calculator[153] on the Thermo Fisher Scientific website.
[2]Number of cycles can be optimized as required.

3.3.4 Identification of *HTR1F* orthologs in zebrafish

Information regarding the zebrafish *HTR1F* orthologs was collected from the online zebrafish databases (GRCz11) of the National Center for Biotechnology Information (NCBI: https://www.ncbi.nlm.nih.gov) and Ensembl (https://www.ensembl.org). The genomic and protein sequences of

both zebrafish *HTR1F* orthologs, *htr1fa* and *htr1fb*, were BLAST (https://blast.ncbi.nlm.nih.gov) against the corresponding human sequences to investigate their homology, and their expression was validated in AB wild-type fish. Zebrafish larvae produced by natural spawning were collected at 3dpf and used for gene expression validation. Total RNA was isolated from the larvae using an RNeasy® Plus Mini Kit (Table 2.1) and reverse transcribed to cDNA using Invitrogen SuperScript® III First Strand synthesis system (Table 2.1) following the manufacturer's instructions. Primers were designed to target the exons of the desired transcripts of *htr1fa* and *htr1fb*, and the cDNA was amplified using standard PCR (Section 2.5.2). The amplified PCR products were then visualized on 2% w/v agarose gel and confirmed by Sanger sequencing as previously described (Section 2.7).

3.3.5 Generation of the mutant zebrafish model

3.3.5.1 gRNA design

The gRNAs of *htr1fa* and *htr1fb* were designed using the online CRISPR design tool CHOPCHOP (http://chopchop.cbu.uib.no). The Ensembl reference genomic DNA sequence of *htr1fa* (ENSDARG00000109862) and *htr1fb* (ENSDARG00000053580) were used for gRNA target site search. Optimal gRNAs were selected based on the following criteria:

- ***Position.*** For each *htr1f* gene, the target sites of the gRNAs are located within the most upstream exons that are conserved across all target transcripts. This is most likely to generate frameshift mutations and early termination of the protein sequence.

- ***Editing efficiency.*** All the gRNAs were chosen based on CHOPCHOP's prediction to have optimal on-target efficiencies and no potential off-target sites with less than 3bp mismatches.
- ***Number.*** It has been shown that targeting a single locus in a coding region only generates biallelic loss-of-function alleles 44% of the time, however, targeting a coding region at two, three, or four independent loci generates biallelic loss of function variants 79%, 93%, and 98% of the time respectively[154]. Four gRNAs were designed for each target gene to increase the success rate of introducing mutations and producing loss of protein function.

To confirm the target sequences of the designed gRNAs in the AB wild-type fish match the zebrafish reference genome, primer pairs (Appendix 1) that surround the gRNA target sites were designed and the genomic DNA of AB wild-type fish was amplified at 4dpf as described in Section 3.3.3. The PCR products were assessed by agarose gel electrophoresis and Sanger sequencing.

3.3.5.2 CRISPR-Cas9 RNP formulation

The methods for CRISPR-Cas9 ribonucleoprotein (RNP) complex formulation were adapted from IDT (Integrated DNA Technologies) online instructions[155]. The crRNA, tracrRNA and HiFi (high fidelity)-Cas9 nuclease aliquots (Section 2.1.2.4) were thawed at room temperature to prepare gRNA and Cas9 protein working solution. The 9µM gRNA working solution was prepared by mixing the components shown in Table 3.3. In the case of multiple crRNAs to be used in a single injection, the crRNAs were mixed in equal amounts while keeping the total volume of crRNAs constant. The gRNA

working solution was heated at 95°C for 5 minutes and left on the benchtop to cool to room temperature.

Table 3.3 9µM gRNA working solution

Component	Volume (µl)
100µM Alt-R® CRISPR-Cas9 crRNA	0.9
100µM Alt-R® CRISPR-Cas9 tracrRNA	0.9
Nuclease-Free Duplex buffer	8.2
Final volume	10

The product information is listed in Table 2.1. The volumes provided here are the optimized results generated in Section 3.4.3.

The 1.5µg/µl Cas9 protein working solution was prepared by gently mixing the components in Table 3.4.

Table 3.4 1.5µg/µl Cas9 protein working solution

Component	Volume (µl)
10ug/ul Alt-R® S.p.HiFi Cas9 Nuclease V3	1.5
Cas9 working buffer	8.5
Final volume	10

The product information is listed in Table 2.1. The volumes provided here are the optimized results generated in Section 3.4.3.

For each experimental group, 3µl of 9µM gRNA working solution and 3µl of 1.5µg/µl Cas9 protein working solution was gently mixed to form the CRISPR-Cas9 RNP complexes which were incubated at 37°C for 10 minutes then left on the benchtop to cool to room temperature.

3.3.5.3 Microinjection

The embryos that were produced by natural spawns were collected on the morning of injection, and 4nl of the CRISPR-Cas9 RNP was injected into the yolk sac of each one-cell stage embryo. Preliminary data generated by Dr Johanna Jones (Menzies Institute for Medical Research, University of Tasmania) shows that 16% of embryos treated with a morpholino to knockdown expression of the aquaporin (AQP0) gene develop cataracts by 5dpf[156]. Based on that we calculated that at least 50 embryos were required in each group to achieve 85% power. While AQP0 knockdown is well known to cause cataracts in zebrafish[157], it should be noted that the morpholino technology used in this pilot is different to CRISPR-Cas9 gene editing and the calculations were used as a guide only.

In each round of injection, in addition to the experimental group that was injected with *htr1f* (*htr1fa*, *htr1fb*, or both) gRNA, the negative control group and the uninjected group were also included to rule out the interference caused by CRISPR-Cas9 RNP and the injection operation. Fish in the negative control group were injected with gRNA targeting the *lacZ* gene, which is a gene in *E.coli* that has been widely used as a negative control.

Both the injected and uninjected embryos were incubated at 28.5°C in E2 media with Methylene Blue in petri dishes (Table 2.1). From 24 hour-post-fertilization (hpf), the embryos were raised in 1x PTU (1-phenyl-2-thiourea) egg water (Section 2.1.2.3) to suppress pigment formation. The injected larvae were then assessed for cataracts (Section 3.3.7) and collected for determining the editing efficiency (Section 3.3.8) at 4dpf.

3.3.6 Quantification of pigment loss in the optimization of mutagenesis conditions

To achieve the optimal editing efficiency with minimum disruptions to the development of the zebrafish embryos, gRNA targeting *tyrosinase* (*tyr*) gene was applied as a positive control to optimize the dose and delivery route of CRISPR-Cas9 RNPs. *tyr* gene is one of the commonly used positive controls in zebrafish model experiments. Knocking out of this gene results in pigmentation defects, which can be easily observed and quantitatively analyzed as early as 2dpf (Figure 3.6).

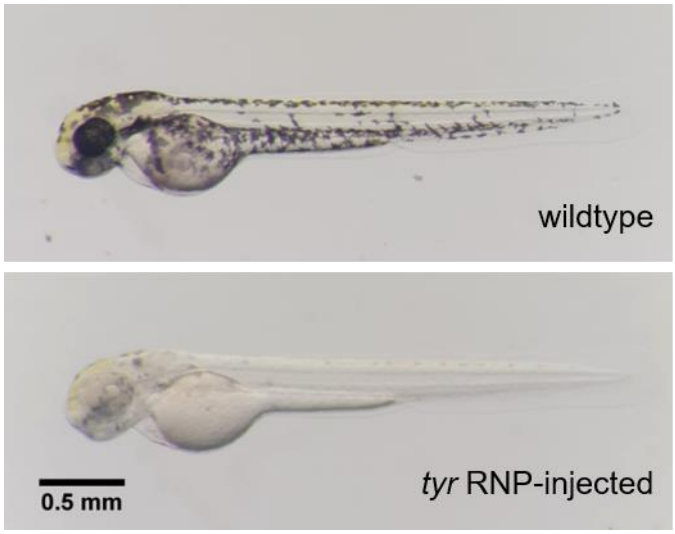


Figure 3.6 Zebrafish embryos at 2dpf

Embryos injected with CRISPR-Cas9 RNP against *tyr* gene (bottom) had observable pigment loss compared to uninjected wild-type zebrafish (top).

The pigment loss of *tyr*-RNP injected fish was quantitated by calculating the percentage of the unpigmented areas on the uveal tract as well as counting the pigment spots on the body (Figure 3.7). The injected larvae were anesthetized in tricaine (300mg/l) (Table 2.1) for 1-2 minutes at 2dpf, and a single anesthetized larva was transferred to a glass base dish (17mm diameter)

(Table 2.2) for brightfield microscope imaging. Whole-embryo images were taken using an SZX16 zoom stereomicroscope (Table 2.2) and an iPhone XR, then processed and analyzed using ImageJ[158] software. The percentage of the unpigmented areas on the uveal tract was manually calculated by dividing the pixels of the unpigmented areas by the ones of the total area. The pixels of the total area were measured after selecting the uveal tract of one eye and clearing the outside of the selected area, and on this basis, the unpigmented areas were selected and measured by adjusting the colour threshold. To count the body pigment spots, the uveal tract of both eyes was selected and removed. By adjusting the noise tolerance, the pigment spots were captured and automatically counted.

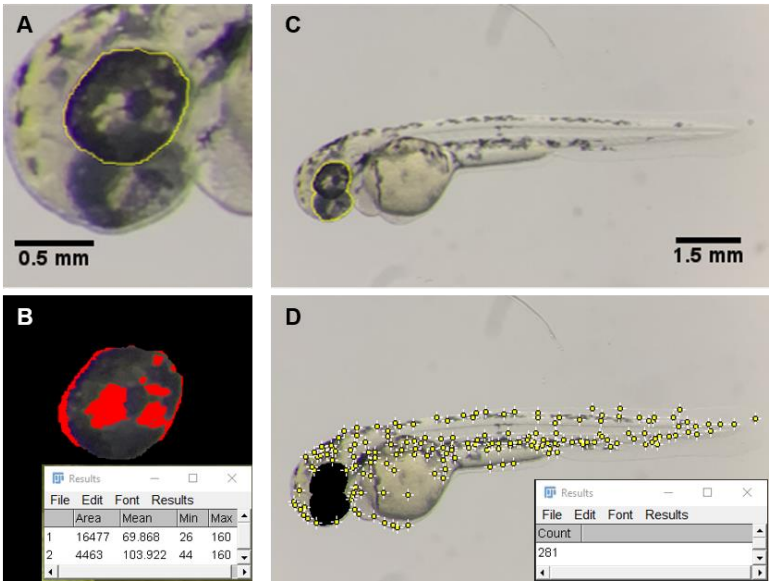


Figure 3.7 Quantification of the pigment loss

Calculating the percentage of the unpigmented area on the zebrafish uveal tract (**A, B**) and counting the pigment spots on the body (**C, D**). In ImageJ[158], select the uveal tract on one side of the eye (**A**). Remove the outside of the selected part and measure the pixels of the uveal tract area (16477 pixels for the example). Adjust the colour threshold to select the unpigmented area (red area in **B**) and measure the pixels of this area (4463 pixels for the example). Manually calculate the unpigmented area percentage. To count the body pigment spots, select and remove the uveal tract area of both eyes (**C**). Adjust the noise tolerance so that the yellow spots can perfectly match the corresponding pigment spots on the body. The spot number will be automatically counted out (281 for the example).

3.3.7 Evaluation of cataract formation by microscope imaging

3.3.7.1 Imaging

On the fourth day after fertilization, the larvae were anesthetized in tricaine (300mg/l) (Table 2.1) and transferred to a glass base dish (Table 2.1) for individual imaging. Whole-embryo images were taken using an SZX16 zoom stereomicroscope (Table 2.2) with Mightex S-Series compact USB 2.0 camera (Table 2.2) and SSClassicCamera software to record the gross morphology of the larvae. After that, the larva was mounted in 0.5% w/v agarose and quickly oriented to a position in which the lens can be viewed directly through the bottom of the dish before the agarose hardens. The lens of the mounted larva was then photographed using Differential Interference Contrast (DIC) microscopy settings with a Ti live cell microscope (Table 2.2), Zyla 4.2 PLUS sCMOS camera (Table 2.2) and NIS-Elements AR acquisition software.

3.3.7.2 Cataract assessment

The lens images were assessed for cataracts in ImageJ[158]. The images were cropped to display only the lens region of the eye to avoid calling bias based on the phenotype of surrounding ocular features. The processed images were pooled together for blinding and the blinded images were assessed and scored by two researchers independently based on the standard images in Figure 3.8 and Figure 3.9 as well as the cataract assessment criteria described below. Any

images with differing cataract classifications were assessed by a third researcher to break a tie, prior to unblinding.

• ***Pitting.*** Described as a dot-shaped depression on the surface of the lens. The pits can be diffused or aggregated in the central lens. Where pits exist, the uniform and smooth structure of the lens is destroyed.

Catagory	**Description**
Pitting=No/ Cataract=No	Can be: • Developmental pitting (Figure 3.8A and B). The pits are mostly fine and spread over the outer lens. They are undegraded organelles. The central lens is clear and uniform. • Artefacts (Figure 3.8C and D). High reflective pits with very round and regular shapes or pits with blurred edges are considered to be artefacts. The central lens is clear and uniform.
Pitting=?/ Cataract=No	Can be: • Unclear pits (Figure 3.8E and F), few in the centre of the lens. The rest of the lens is smooth and uniform. • Clear pits, but too few to be sure (Figure 3.8G and H). There are one or two clear pits in the centre of the lens. The rest of the lens is smooth and uniform.
Pitting=Yes/ Cataract=Yes (mild, score=1)	Can be: • A small number of clear pits, enough to be confident (Figure 3.8I and J). The pits are located in the central lens rather than the outer lens. • Pits may be blurred, but are numerous in the central lens (Figure 3.8K and L). This may be caused by the loss of focus during imaging.
Pitting=Yes/ Cataract=Yes (moderate, score=2)	Significant pits (Figure 3.8M-P). Can be seen all over the lens or cluster in the central lens.
Pitting=Yes/ Cataract=Yes (severe, score=3)	Significant pits, deep or high density (Figure 3.8Q-T). Seriously obstruct the visual pathway.

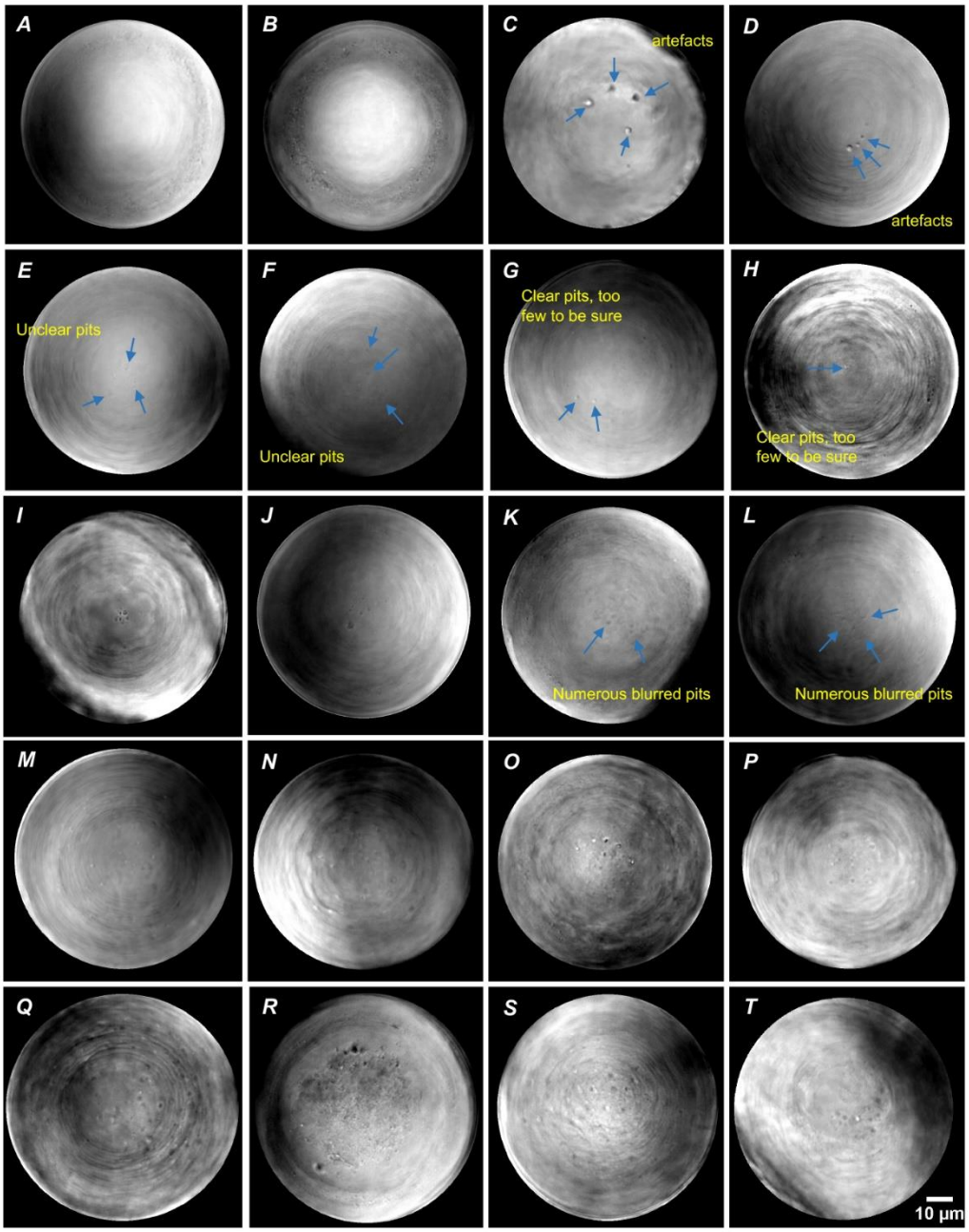


Figure 3.8 Zebrafish cataract assessment criteria (pitting)

• ***Central mass.*** The centre of the lens presents as an irregular shape protrusion, which is significantly different from the rest of the lens. The central lens has lost its smooth and high regular structure.

Catagory	Description
Central mass=No/ Cataract=No	The centre of the lens may seem to bulge out compared to the rest of the lens, but it is smooth, uniform and has a regular shape (Figure 3.9A-D). The outer lens is uniform and well developed. This could be an optical illusion caused by light or the different depths of the lens that had been captured during imaging.
Central mass=Yes/ Cataract=Yes (mild, score=1)	The central lens lost its round shape and become "flower-shaped" (Figure 3.9E). The outer lens is mostly uniform and well-developed. The "flower-shape" is presumed to be the outline of the original, undifferentiated cells in the lens nucleus.
Central mass=Yes/ Cataract=Yes (moderate, score=2)	The central lens is abnormal, lost its round shape and smooth, uniform texture (Figure 3.9F). In the lens nucleus, original, undifferentiated cells appear to be visible, meanwhile the outer lens is well developed.
Central mass=Yes/ Cataract=Yes (severe, score=3)	The central lens is significantly abnormal, completely lost its regular, smooth texture (Figure 3.9G and H).

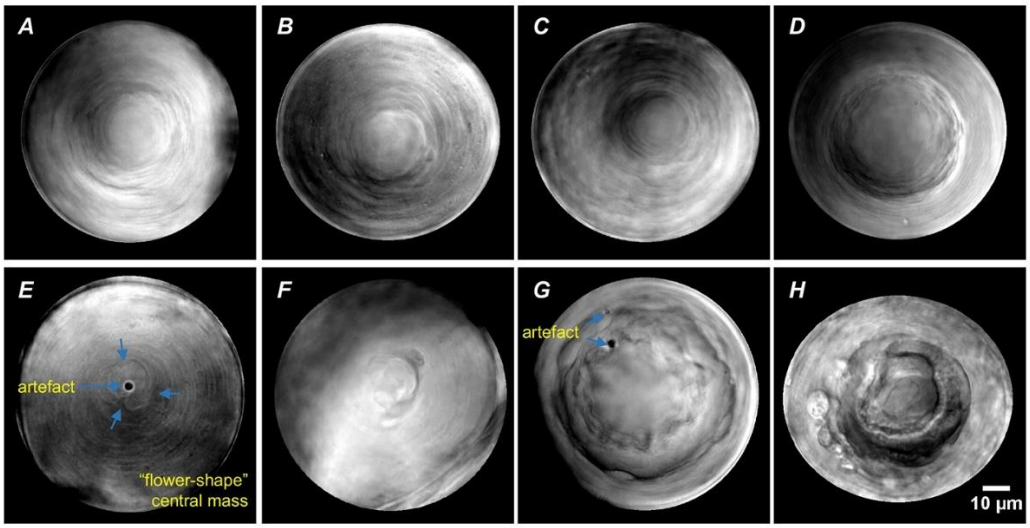


Figure 3.9 Zebrafish cataract assessment criteria (central mass)

3.3.8 Analysis of editing efficiency in F0 embryos

Due to the differences in editing complexity, two distinct analysis strategies were applied to single-guide CRISPR-Cas9 RNP and multi-guide CRISPR-Cas9 RNP induced edits.

3.3.8.1 Editing efficiency analysis of samples injected with single-guide CRISPR-Cas9 RNP

The injected larvae were collected at 4dpf, and the gRNA target region of the individual larva was amplified using the methods described in Section 3.3.3. The primers used for amplification are listed in Appendix 1. The PCR products were inspected on a 2% w/v agarose gel to confirm that the target region has been amplified correctly prior to being Sanger sequenced. The Sanger sequencing data (AB file) of both experimental and uninjected wild-type fish were uploaded to the free online CRISPR analysis tool – Sythego Inference of CRISPR Edits (ICE) (v2) (https://ice.synthego.com) and analysed for editing efficiency.

3.3.8.2 Editing efficiency analysis of samples injected with multi-guide Cas9 RNP

In the case of multi-guide induced editing, traditional Sanger sequencing and ICE are not suitable for analyzing complex editing products due to their relatively low sensitivity and accuracy. Instead, edits were analyzed by next-

generation sequencing (NGS) and a corresponding deep sequencing data analysis tool – CRISPResso2[159].

The targeted amplicon library was prepared following the Illumina 16S Metagenomic Sequencing Library Preparation protocol[160]. Primer pairs (Appendix 1) with target region covering all gRNA target sites were designed and assembled with the overhang adapter sequences (Forward adapter: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3', Reverse adapter: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'). Genomic DNA of embryos at 4dpf was extracted and amplified with the assembled primers as described in Section 3.3.3. The amplicons were purified by AMPure XP beads (Table 2.1) prior to being barcoded with the Nextera XT Index kit (Table 2.1). The barcoding PCR was performed using Q5® Hot Start High-Fidelity 2X Master Mix (Table 2.1) with the conditions listed in Table 3.5.

Table 3.5 Reaction components and cycling conditions of barcoding PCR

Component	Volume (µl)
Q5 Hot Start High-Fidelity 2X Mater Mix	12.5
Milli-Q water	5.5
Nextera XT Index 1 primers (N7xx)	2.5
Nextera XT Index 2 primers (S5xx)	2.5
Purified amplicons	2
Final volume	25

The product information is listed in Table 2.1.

Cycle step	Temperature (°C)	Time	Cycles
Initial denaturation	95	2min	1
Denaturation	95	15s	
Annealing	61	20s	15
Extension	72	20s	
Final extension	72	2min	1

The cycling conditions come from the well-developed protocol[161] provided by Gaudelli *et al.*, 2017[161].

The barcoded PCR products were then purified using AMPure XP beads and quantified using Qubit™ dsDNA HS Assay kit (Table 2.1). Subsequently, each sample was diluted to 4nM with Buffer EB (Table 2.1) and 5µl of the diluted samples were pooled into a final library. The concentration and size of the pooled library were checked using the High Sensitivity D1000 ScreenTape system (Table 2.1). The pooled amplicon library was denatured with fresh 0.2N NaOH (Section 2.1.2.6) and combined with Phix control (one of the High Sensitivity D1000 Reagents, (Table 2.1)) to increase the diversity of the library. The mixed library was loaded into the MiSeq sequencer (Table 2.2) and paired-end sequenced (2 x 301) using Miseq Reagent Kit v3 (Table 2.1).

The generated paired-end FASTQ files were trimmed with Timmomatic[162] (version 0.39) to remove the adapters using the script in Appendix 3. The trimmed paired-end reads were input into CRISPResso2[159], filtering out for low-quality reads, aligned to reference sequences and analyzed using the command list in Appendix 4.

3.3.9 Statistical analysis

All statistical analyses were performed using Prism 6 software. Between-group analyses were assessed using unpaired two-tailed t-test or one-way analysis of variance (ANOVA) and Tukey's multiple comparisons test. The Fisher's Exact test was used to compare the experimental group to the negative control group and uninjected group. The post-hoc power analysis (G*Power 3.1.9.4[163]) was applied to calculate the number of fish to be injected in each group. P values less than 0.05 were considered statistically significant. All the *HTR1F* knockout experiments were repeated to obtain robust results.

3.4 Results

3.4.1 Orthologs of *HTR1F* and their expression in zebrafish

The *HTR1F* gene has two orthologs in zebrafish, *htr1fa* and *htr1fb*. The predicted protein sequences of both orthologs were compared with the human HTR1F (Figure 3.10 and Figure 3.11), and 61% and 58% identical residues were identified, respectively. The *htr1fa* gene has 3 transcripts (*htr1fa-201* (ENSDART00000157259.2), *htr1fa-202* (ENSDART00000040449.7) and

htr1fa-203 (ENSDART00000185312.1)) in the Ensembl database and *htr1fb* has 1 transcript (ENSDART00000075626.5) (Figure 3.12A). The expression of both genes was detected in the AB strain fish at 3dpf (Figure 3.12B).

Score	Expect	Method	Identities	Positives	Gaps
408 bits(1048)	1e-146	Compositional matrix adjust.	215/351(61%)	277/351(78%)	11/351(3%)

```
Query  21   PSK-----ILVSLTLSGLALMTTTINSLVIAAIIVTRKLHHPANYLICSLAVTDFLVAVL  75
            PSK     +L+SLTLS LA++TT +N LVI AIIVTRKLHHP+NYLICSLAVTD LVA+L
Sbjct  15   PSKNNGDVVLLSLTLSFLAILTTAMNCLVITAIIVTRKLHHPSNYLICSLAVTDLLVAIL  74

Query  76   VMPFSIVYIVRESWIMGQVVCDIWLSVDITCCTCSILHLSAIALDRYRAITDAVEYARKR  135
            VMPFSI+YI +++W++G+ +C  WLSVDITCCTCSILHL+AIA+DRYRAITDAVEY+RKR
Sbjct  75   VMPFSIIYIAKDTWLIGEALCKFWLSVDITCCTCSILHLAAIAVDRYRAITDAVEYSRKR  134

Query  136  TPKHAGIMITIVWIISVFISMPPLFWRHQGTSRDDECIIKHDHIVSTIYSTFGAFYIPLA  195
            T   A IMI++VW++++ IS+PP+  R+     ++ECII H +I S +YSTFGAFYIPL
Sbjct  135  TSLRAAIMISVVWLLAIVISLPPILLRN---GEENECIIVHTNIASMLYSTFGAFYIPLI  191

Query  196  LILILYYKIYRAAKTLYHKRQASRIAKEEVNGQVLLESGEKSTKSVSTSYVLEKSLSDPS  255
            LILILYYKIYRAAKTLYH+R+ S + K E+NG +L +  ++   +  T    EKS+S+PS
Sbjct  192  LILILYYKIYRAAKTLYHRRRTSCLNKSEMNGHMLPKCSDREPTTPDTLSPPEKSMSEPS  251

Query  256  TDFDKIHSTVRSLRSEFKHEKSWRRQKISGTRERKAATTLGLILGAFVICWLPFFVKELV  315
            T+ D++  T RS +S  +      R +IS  RE+KAATTLGLI+GAFV+CWLPFF+ E++
Sbjct  252  TEGDRVRITARSPKSRRERSLR--RHRISNIREKKAATTLGLIIGAFVVCWLPFFIHEVI  309

Query  316  VNVCD-KCKISEEMSNFLAWLGYLNSLINPLIYTIFNEDFKKAFQKLVRCR  365
             N+C   C  S  ++NFL WLGYLNSLINPLIYTIFNEDFK+AFQK+++C+
Sbjct  310  YNICTVSCSRSVAVTNFLTWLGYLNSLINPLIYTIFNEDFKRAFQKIIKCK  360
```

Figure 3.10 Alignment of zebrafish Htr1fa and human HTR1F

The Htr1fa (NCBI RefSeq: XP_021334563.1) is indicated as “Sbjct” in the figure and human HTR1F (NCBI RefSeq: NP_000857.1) is indicated as “Query”. NCBI BLASTp shows that the two proteins are 61% identical.

Score	Expect	Method	Identities	Positives	Gaps
430 bits(1105)	3e-155	Compositional matrix adjust.	219/378(58%)	278/378(73%)	22/378(5%)

```
Query  1    MDFLNSSDQNLTSEELLNRM-PSKILVSLTLSGLALMTTTINSLVIAAIIVTRKLHHPAN  59
            MD +N +  +L S+ L ++M PSKIL+SLTLS LA+ TT INSLVI AI++TRKLH PAN
Sbjct  1    MDPINCTSVDL-SDVLASKMSPSKILLSLTLSFLAVATTAINSLVITAILITRKLHQPAN  59

Query  60   YLICSLAVTDFLVAVLVMPFSIVYIVRESWIMGQVVCDIWLSVDITCCTCSILHLSAIAL  119
            YLICSLAVTD LVAVLVMP SIVYI  E+W++G +VC +WL VD+TCCTCSILHL+AIAL
Sbjct  60   YLICSLAVTDLLVAVLVMPVSIVYIAEETWVLGPIVCHLWLGVDVTCCTCSILHLAAIAL  119

Query  120  DRYRAITDAVEYARKRTPKHAGIMITIVWIISVFISMPPLFWR-------HQGTSRDDEC  172
            DRYRAITDAV Y++KRT K   + I  +W +S+ +S+PPL WR         G     +C
Sbjct  120  DRYRAITDAVAYSQKRTYKRVIVTILSLWTLSILVSLPPLVWRKFPKVEFKDGKREPMDC  179

Query  173  IIKHDHIVSTIYSTFGAFYIPLALILILYYKIYRAAKTLYHKRQASRIAKEEVNGQVLLE  232
            +I+HDH+  T+YSTFGAFYIPLALIL+LYYKIY+AA+ L ++R +SR+ K+ V+  +L
Sbjct  180  LIEHDHVAFTVYSTFGAFYIPLALILVLYYKIYKAAEMLRNRRGSSRLVKQTVSSVML--  237

Query  233  SGEKSTKSVSTSY----VLEKSLSDPSTDFDKIHSTVRSLRSEFKHEKSWRRQKISGTRE  288
             G  S K ++ S      +EKS SDPSTD +++  T  S           R ++  G RE
Sbjct  238  PGMSSDKDIALSPDSFCPIEKSFSDPSTDGERVRITSSS-------GNLIRVRRNPGARE  290

Query  289  RKAATTLGLILGAFVICWLPFFVKELVVNVCDKCKISEEMSNFLAWLGYLNSLINPLIYT  348
            R+AA TLGLILGAFV+CWLPFF+KE++VN+C  C  S  +++FL WLGYLNSLINPLIYT
Sbjct  291  RRAALTLGLILGAFVVCWLPFFLKEVIVNICPTCTTSAVLADFLTWLGYLNSLINPLIYT  350

Query  349  IFNEDFKKAFQKLVRCRC  366
            IFNEDFKKAF+KL+  C
Sbjct  351  IFNEDFKKAFKKLLPLCC  368
```

Figure 3.11 Alignment of zebrafish Htr1fb and human HTR1F

The Htr1fb (NCBI RefSeq: XP_005159986.1) is indicated as “Sbjct” in the figure and the HTR1F (NCBI RefSeq: NP_000857.1) is indicated as “Query” in the figure. NCBI BLASTp shows that the two proteins are 58% identical.

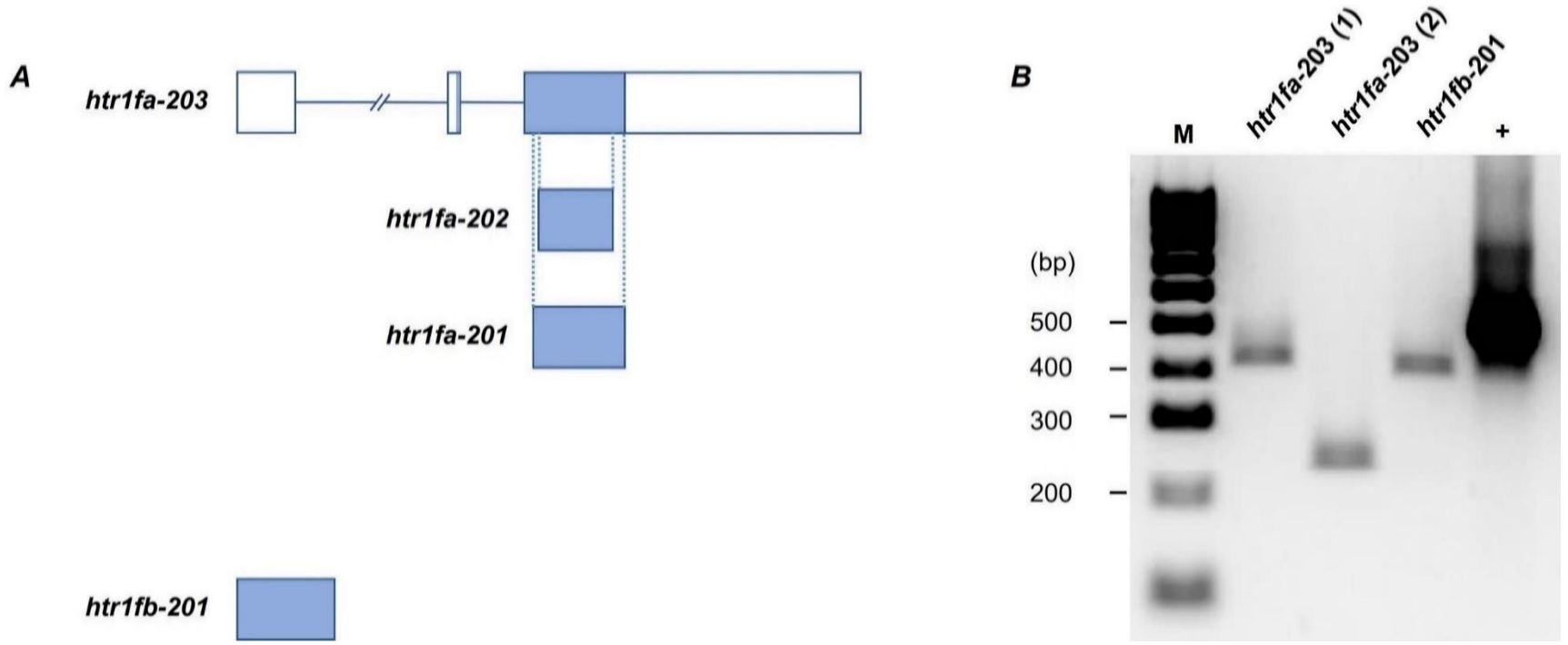


Figure 3.12 The transcripts and expressions of zebrafish *htr1fa* and *htr1fb* genes

A. Schematic of the transcripts of *htr1fa* and *htr1fb* in the Ensembl database. *htr1fa-202* and *htr1fa-203* are the transcripts of the primary assembled *htr1fa* gene and *htr1fa-201* is a product of the alternative sequence of *htr1fa* (ENSDARG00000053580.7). *htr1fb-201* is the only transcript of *htr1fb*. **B.** RT-PCR showing the expression of *htr1fa* and *htr1fb* in AB zebrafish embryos at 3dpf. As shown in panel **A**, the coding region of *htr1fa-201* and *htr1fa-202* are covered by *htr1fa-203*, so we utilized *htr1fa-203* in the *htr1fa* expression assessment. Due to the large size, two pairs of primers were designed for *htr1fa-203*. Expression for *htr1fa-203* (*htr1fa-203* (1): 416bp; *htr1fa-203* (2): 244bp) and *htr1fb-201* (397bp) was observed in AB zebrafish. M: 100bp ladder; +: β-actin positive control (491bp).

3.4.2 Designed gRNAs

The first and most critical step of the CRISPR-Cas9-induced knockout experiment is the design of gRNA. The editing efficiencies of each gRNA can vary greatly, thus a well-designed gRNA is essential for successful gene disruption. According to the criteria listed in Section 3.3.5.1, four independent gRNAs were designed for *htr1fa* and *htr1fb*, respectively. For each *htr1f* gene, the four gRNAs targeted different positions of the predicted coding region (Appendix 5). In particular, for *htr1fa*, the target region is shared by all three transcripts. The gRNA target sequences were verified in the AB wild-type fish.

The *tyr* gRNA used to optimize the dose and delivery route of CRISPR-Cas9 RNP (Section 3.4.3) is predesigned by the IDT company. The gRNA sequence of *lacZ* that was employed as the negative control in the *htr1f* knockout experiments (Section 3.4.5) was obtained from the Validated gRNA Sequence Data table of addgene[164] and was first published by Platt *et al.*[165]. The sequence was BLAST with the zebrafish genome to ensure that it has no target in zebrafish. Additionally, the potential off-target sites of the *lacZ* gRNA in the zebrafish genome were predicted in the IDT CRISPR-Cas9 guide RNA design checker[166], and the possibility of off-targeting to the *htr1fa* or *htr1fb* gene was ruled out.

The sequences of all the gRNAs are listed in Table 3.6.

Table 3.6 gRNA sequences

gRNA	Sequence (5' – 3')	Usage
htr1fa gRNA1	ATGTCTGAACCGTCCACCGA	*htr1fa* knockout
htr1fa gRNA2	GGCAACCAAGAGGTCGGTGA	
htr1fa gRNA3	CATCTGTTATAGCCCGGTAG	
htr1fa gRNA4	TCTACATCGCAAAAGACACG	
htr1fb gRNA1	GTATCGATCCAGCGCAATGG	*htr1fb* knockout
htr1fb gRNA2	TAACATCGACACCTAACCAC	
htr1fb gRNA3	GGCGAGGTGTAAAATTGAGC	
htr1fb gRNA4	AGTAGACGGTAAAAGCCACA	
tyr gRNA	CGTTGGGAAGGTCGGACACC	Positive control
lacZ gRNA	TGCGAATACGCCCACGCGAT	Negative control

3.4.3 Optimizing dose and delivery route of CRISPR-Cas9 RNP

In the optimization, *tyr*-targeting CRISPR-Cas9 RNP was injected into the F0 embryos and the pigment loss of the injected larvae was quantified using methods described in Section 3.3.6. In the initial dose optimization process, the CRISPR-Cas9 reagents were injected into the cell of one-cell stage embryos at a fixed volume of 1nl. A range of gRNA concentrations from 0.5µM to 13.5µM was tested while keeping Cas9 nuclease at 0.75µg/µl. A significant reduction of pigmentation in the uveal tract and body was observed in the 4.5µM and 13.5µM gRNA-injected fish (Figure 3.13A and B). However, no statistical difference in pigment loss was found between 4.5µM and 13.5µM gRNA-injected zebrafish (uveal tract: p=0.9427; body: p=0.9356). The 4.5µM and 13.5µM gRNAs were then applied to examine the effect of varying Cas9 concentration at 0.75µg/µl or 1µg/µl. In all experimental conditions that were tested in the Cas9 concentration optimization, we observed significant pigment

loss on both the uveal tract and body in the injected groups compared with the uninjected groups (Figure 3.13C and D). Although there were no significant differences in pigment loss between the injected groups in the optimization of Cas9 concentration (4.5µM gRNA/ 0.75µg/µl Cas9 vs 4.5µM gRNA/ 1µg/µl Cas9: uveal tract: p=0.6911, body: p=0.1863; 13.5µM gRNA /0.75µg/µl Cas9 vs 13.5µM gRNA/ 1µg/µl Cas9: uveal tract: p=0.8329, body: p=0.9795), the combination of 4.5µM gRNA and 0.75µg/µl Cas9 nuclease presented with the highest survival rates (93%, 57 of 61 injected embryos). We therefore set the concentration of gRNA and Cas9 nuclease to be 4.5µM and 0.75µg/µl.

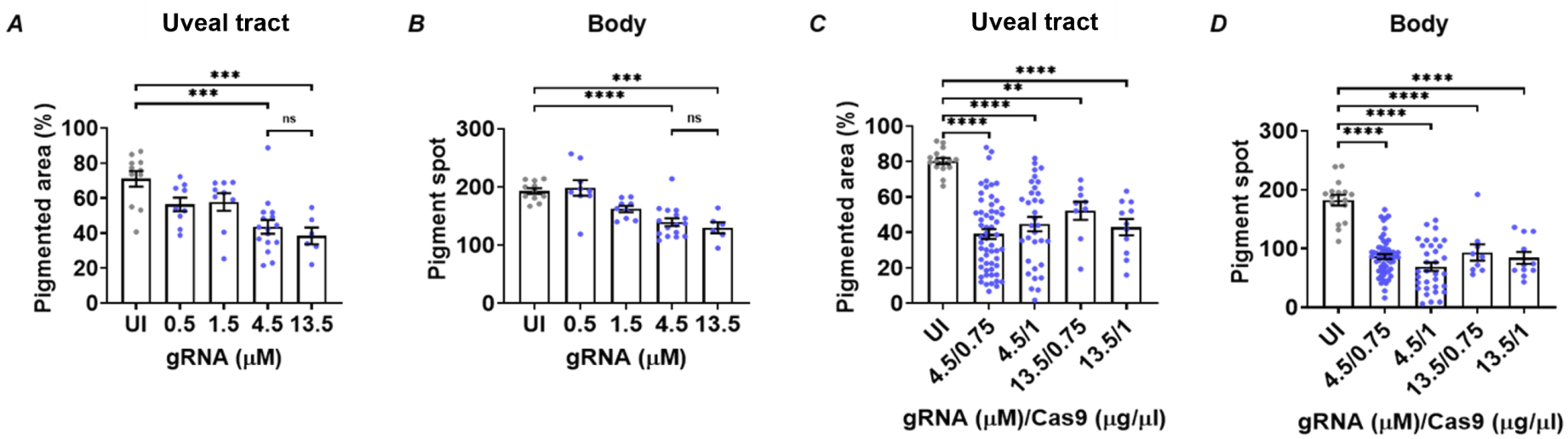


Figure 3.13 CRISPR-Cas9 RNP dose optimization

A, B. Optimisation of *tyr* gRNA concentration (UI: n=11, 0.5μM: n=9, 1.5μM: n=9, 4.5μM: n=16, 13.5μM: n=6). **C, D.** Optimisation of Cas9 concentration (UI: n=16, 4.5μM/0.75μg/μl: n=57, 4.5μM/1μg/μl: n=33, 13.5μM/0.75μg/μl: n=9, 13.5μM/1μg/μl: n=11). The pigment areas on the uveal tract (**A, C**) and body (**B, D**) of zebrafish injected with a series of different concentrations of *tyr* gRNA or Cas9 were measured and compared (UI: uninjected; **p < 0.01, ***p < 0.001, ****p < 0.0001; ns: not significant; one-way ANOVA and Tukey's multiple comparisons test; Mean ± SEM).

This condition was next used to determine the delivery route. There are two commonly used delivery routes in zebrafish embryo microinjection: inject into the single cell or the yolk sac (Figure 3.14). In theory, cell injections are more likely to induce gene editing and increase the chance of editing both alleles simultaneously. However, it is technically challenging and time-consuming, limiting the number of embryos that can be injected in a single experiment. In comparison, yolk injections are high throughput, enabling an increased number of injected embryos to be achieved in one round of injection. Additionally, yolk injections have been shown to have similar editing efficiencies to single-cell injections, with four times the reagent dose[167].

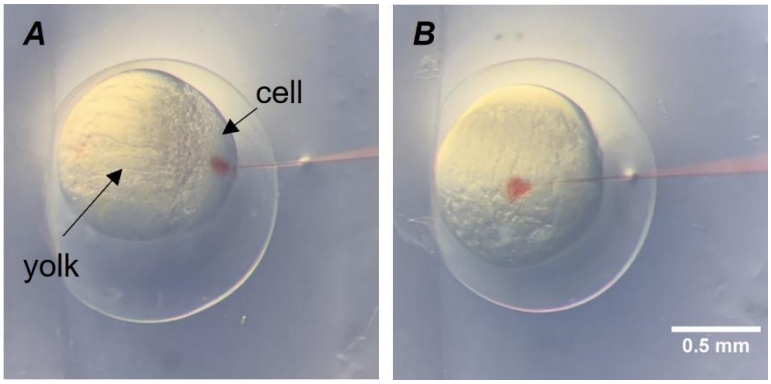


Figure 3.14 Illustration of cell injection (A) and yolk injection (B)

With the identical concentration of gRNA and Cas9 nuclease, the pigment loss of injecting 1nl of *tyr*-targeting CRISPR-Cas9 RNP into the cell and 4nl into the yolk of one-cell stage embryos was compared. It is shown that yolk injections had significantly higher body pigment loss (Figure 3.15, p=0.042) and a slightly higher survival rate (cell: 73%, 19 of 26 injected embryos; yolk: 88%, 30 of 34 injected embryos) than cell injections. In addition,

considering the simple operation and high throughput of yolk injection, we chose to perform yolk injections in all subsequent experiments.

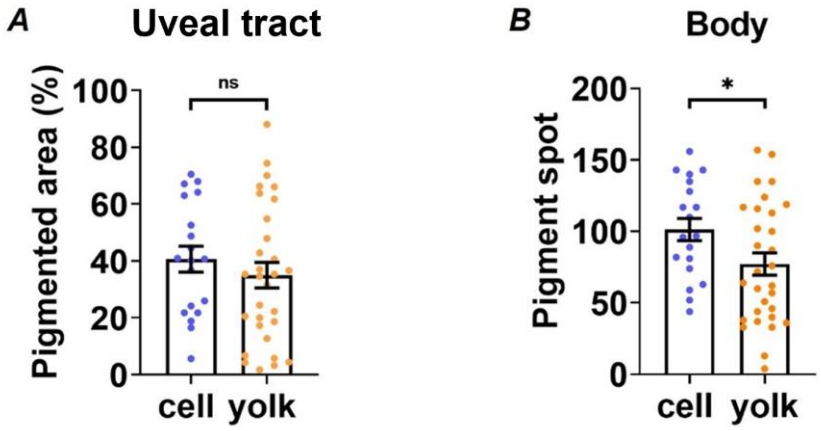


Figure 3.15 CRISPR-Cas9 RNP delivery route optimization

The zebrafish embryos were cell-injected (n=19) or yolk-injected (n=30) with *tyr*-targeting CRISPR-Cas9 RNP. The pigment areas on the uveal tract (**A**) and body (**B**) of zebrafish in the cell injection group and yolk injection group were measured and compared. (*p < 0.05, ns: not significant; unpaired two-tailed t-test; Mean ± SEM).

3.4.4 Determining the editing efficiency of *htr1f* gRNAs

Five larvae from each gRNA injected group were collected for the editing efficiency analysis (details described in Section 3.3.8.1). Samples with a low model fit ($R^2 < 0.8$) were excluded according to Synthego's recommendation[168]. The editing efficiency of each *htr1f* gRNA, which is indicated by the average insertions and/or deletions (indels) percentage of each individual edited fish, is listed in Table 3.7.

Table 3.7 *htr1f* gRNA editing efficiency

gRNA	Indel (%, per fish)
htr1fa gRNA1	82.6
htr1fa gRNA2	53.4
htr1fa gRNA3	89.8
htr1fa gRNA4	84.8
htr1fb gRNA1	94.6
htr1fb gRNA2	81.3
htr1fb gRNA3	72.8
htr1fb gRNA4	88.4

3.4.5 Knockout of *htr1f* in F0 zebrafish

To knockout both *htr1f* genes simultaneously, we first tried CRISPR-Cas9 RNP carrying all 8 *htr1f* gRNAs. The gRNAs were mixed in equal amounts while keeping the total concentration constant at 4.5µM prior to being assembled with 0.75µg/µl Cas9 protein. 1nl of the CRISPR-Cas9 RNP complexes were then yolk-injected to a single-cell stage F0 embryo. Subsequent multi-guide injections were performed in the same manner unless otherwise specified. Although a high survival rate was observed (87%, 68 of 78 injected embryos), common dysmorphic phenotypes related to toxicity, including pericardial edema, body curvature, hydrocephalus and short tail were observed at a rate of 54% (20 of 37 imaged embryos). The high dysmorphic rate suggests 8 gRNAs may be too toxic to the embryos. We therefore decreased the total number of gRNAs to 4 by injecting the mixture of the two most efficient gRNAs of both *htr1f* genes (*htr1fa*: gRNA3 and gRNA4; *htr1fb*: gRNA1 and gRNA4, Table 3.7). The four-gRNA injected embryos showed a similar survival rate (78%, 58 of 74 injected embryos) to the eight-gRNA embryos,

but the dysmorphic rate (12%, 7 of 58 imaged embryos) decreased dramatically to an acceptable level.

In order to investigate the effects of disrupting each *htr1f* gene, we further knocked out *htr1fa* and *htr1fb* separately by injecting CRISPR-Cas9 RNPs carrying the two most effective gRNAs corresponding to each gene. The survival rate and dysmorphic rate of *htr1fa*-RNP injected embryos were 84% (173 of 207 injected embryos) and 1% (2 of 207 imaged embryos), respectively. For the *htr1fb*-RNP injected embryos, the survival rate was 83% (191 of 229 injected embryos) and the dysmorphic rate was 3% (8 of 229 imaged embryos).

3.4.6 Cataract assessment

All the injected larvae (except for the eight-gRNA injection group), as well as 123 randomly selected uninjected wild-type larvae from the *HTR1F* knockout experiments, were imaged and assessed for cataracts using the strategies described in Section 3.3.7. Results are summarized in Table 3.8. No significant difference in cataract formation was observed between all the *htr1f* (*htr1fa, htr1fb* or both) knockout groups and their corresponding *lacZ* control groups. In addition, the results showed that the baseline prevalence of cataracts in the AB wild-type fish in our laboratory is 13.8%.

Table 3.8 Cataract assessment of *htr1f* knockout experiments

Target gene group		No.(%) fish		P-value, OR(95%CI)
		Cataract	**No cataract**	
htr1fa + htr1fb	Replicate 1 Injected	5(8.8%)	52	0.399, 3.329 (0.351,163.518)
	Replicate 1 Control	1(2.8%)	35	
	Replicate 2 Injected	7(14.3%)	42	1.000, 1.099 (0.271,4.813)
	Replicate 2 Control	5(13.2%)	33	
	Total Injected	**12(11.3%)**	**94**	**0.616, 1.444 (0.473,4.932)**
	Total Control	**6(8.1%)**	**68**	
htr1fa	Replicate 1 Injected	7(12.5%)	49	0.162, 3.674 (0.657,37.937)
	Replicate 1 Control	2(3.7%)	52	
	Replicate 2 Injected	10(18.2%)	45	0.801, 1.165 (0.373,3.750)
	Replicate 2 Control	8(16%)	42	
	Total Injected	**17(22%)**	**94**	**0.224, 1.696 (0.691,4.375)**
	Total Control	**10(9.6%)**	**94**	
htr1fb	Replicate 1 Injected	12(19.7%)	49	0.318, 1.845 (0.611,6.008)
	Replicate 1 Control	7(13.2%)	53	
	Replicate 2 Injected	9(16.4%)	46	0.399, 1.752 (0.482,7.191)
	Replicate 2 Control	5(10%)	45	
	Total Injected	**21(18.1%)**	**95**	**0.136, 1.801 (0.794,4.254)**
	Total Control	**12(10.9%)**	**98**	
Uninjected wild-type	Total	17(13.8%)	106	-

Statistical significance of experiment group in comparison to control (*lacZ*) calculated using Fisher's exact test.

3.4.7 Editing efficiency analysis

A total of 96 (Table 3.9) fish were selected and the editing efficiencies were analyzed following the methods described in Section 3.3.8.2. The same number of fish with and without cataracts were randomly selected from each injected group according to the results of the cataract assessment. Four randomly selected uninjected wild-type fish were also sequenced for reference.

Table 3.9 Information of samples for editing efficiency analysis

Group	**No. fish**	
	Cataract	**No cataract**
htr1fa + *htr1fb*	10	10
htr1fa	12	12
htr1fb	12	12
Control*	12	12
Uninjected wild-type	4	
Total	96	

*Control: zebrafish injected with CRISPR-Cas9 RNP targeting *lacZ*.

The characteristics of *htr1fa* and *htr1fb* editing analyses are summarised in Table 3.10 and Table 3.11, respectively. Two samples in the *htr1fa* and *htr1fb* double-knockout group were excluded from *htr1fa* mutagenesis analysis due to low aligned sequencing reads (<200 reads). Similarly, one double-knockout sample and one *htr1fb* single-knockout sample were excluded from the *htr1fb* mutagenesis analysis.

For *htr1fa* editing efficiency analysis, all the samples in both the double-knockout group and the *htr1fa* single-knockout group have been edited and the editing efficiencies of both groups were over 99%. For both groups, the main type of indel was deletion, followed by the coexistence of insertion and deletion, and the least was insertion. The size of most deletions is 0-20bp and 100-120bp (Figure 3.16A and C). The 100-120bp size deletions were generated by the entire excisions between the two gRNAs, as the distance between the two *htr1fa*-gRNA target sites is 98bp. The majority of insertions are between 0 and 40bp (Figure 3.16B and D). The specific distribution and nucleic acid composition of indels are shown in Appendix 6.

Similar outcomes were observed in *htr1fb* mutagenesis analysis. The samples were all edited with editing efficiencies above 97% in both the double-knockout group and the *htr1fb* single-knockout group. Most of the induced indels were deletions and the proportion of both insertions and deletions was slightly higher than that of only insertions. Most of the deletions are 0-40bp (Figure 3.17A and C) and the insertion size is mainly 0-20bp (Figure 3.17B and D). Details of indel distribution and nucleic acid composition are shown in Appendix 7.

In contrast, the editing efficiency of either *htr1fa* (0.005%±0.003%) or *htr1fb* (0.022%±0.006%) in the control group was close to 0%. The tiny percentage of editing is considered background noise in the assay.

Table 3.10 Summary of characteristics of *htr1fa* editing efficiency analysis

	Group	
	***htr1fa* + *htr1fb* gRNAs**	***htr1fa* gRNAs**
Total analyzed fish	18	24
Number of edited fish	18	24
Editing efficiency* (%, mean±SEM)	99.54±0.21	99.65±0.12
Indel type (%, mean±SEM)		
Only insertion	4.56±0.95	5.07±0.87
Only deletion	73.9±3.82	73.47±2.36
Insertion & deletion	21.08±3.58	21.1±1.9
Indel size	Figure 3.16	
Indel position	Appendix 6	

*Editing efficiency is indicated as the percentage of sequence reads showing modified alleles in a single fish calculated by CRISPResso2[159].

Table 3.11 Summary of characteristics of *htr1fb* editing efficiency analysis

	Group	
	***htr1fa* + *htr1fb* gRNAs**	***htr1fb* gRNAs**
Total analyzed fish	19	23
Edited fish	19	23
Editing efficiency* (%, mean±SEM)	99.29±0.44	98.6±0.64
Indel type (%, mean±SEM)		
Only insertion	7.21±2.43	6.02±2.02
Only deletion	67.05±4.8	59.38±5.97
Insertion & deletion	25.02±3.66	33.2±5.44
Indel size	Figure 3.17	
Indel position	Appendix 7	

*Editing efficiency is indicated as the percentage of sequence reads showing modified alleles in a single fish calculated by CRISPResso2[159].

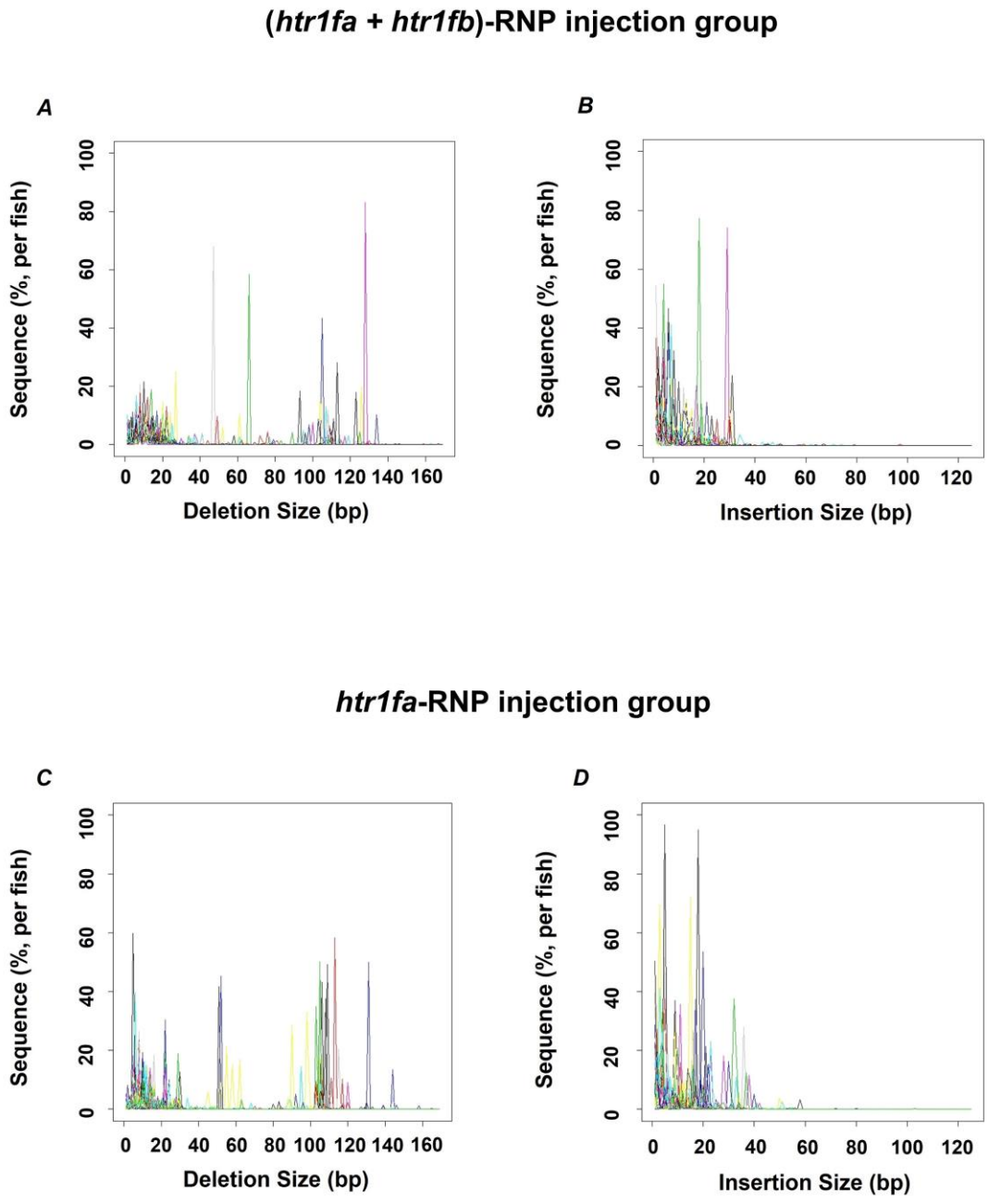


Figure 3.16 Indel size distribution of the *htr1fa* knockout experiments

Frequency distribution of indel size. Each coloured line represents a sample.

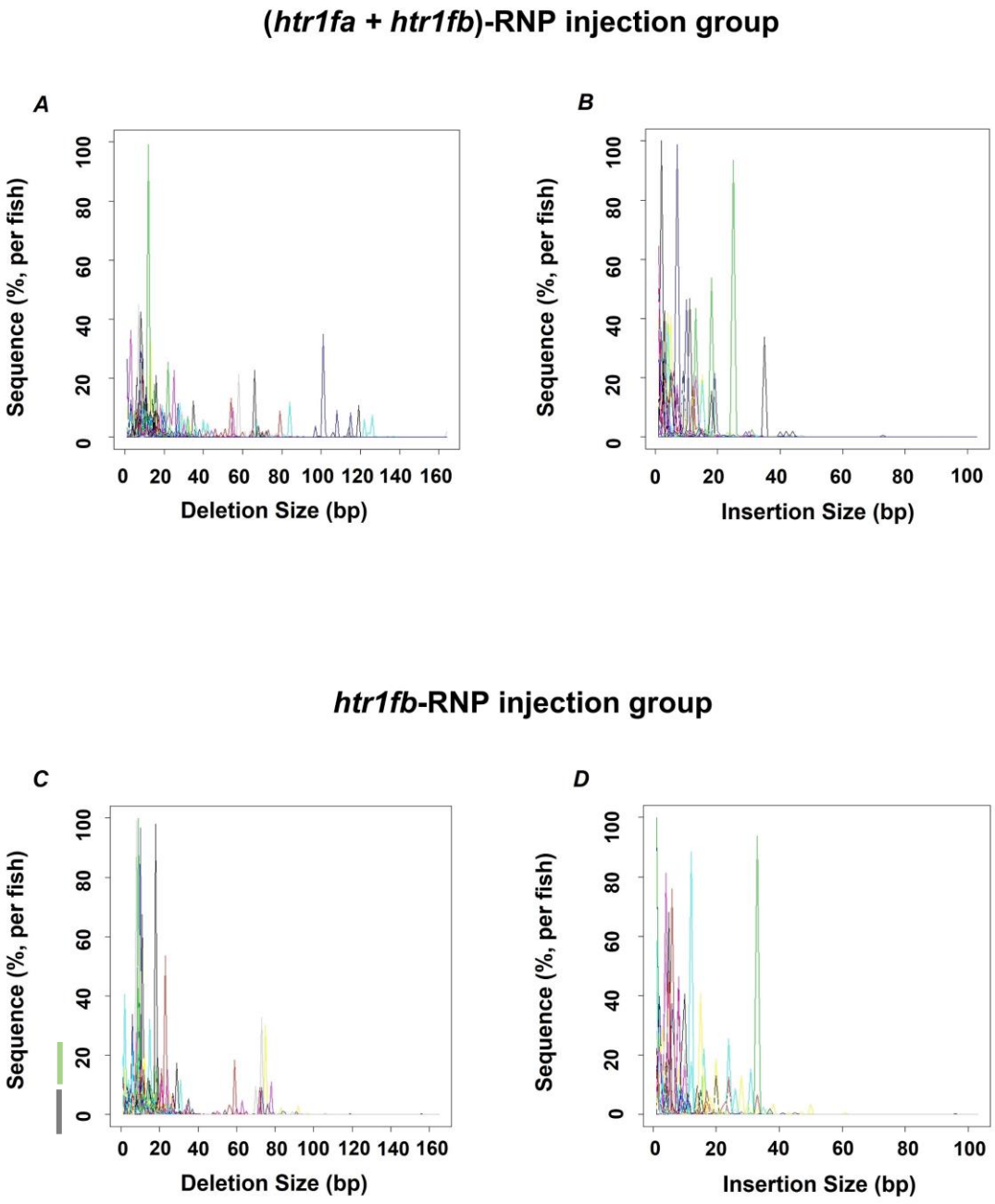


Figure 3.17 Indel size distribution of the *htr1fb* knockout experiments

Frequency distribution of indel size. Each coloured line represents a sample.

3.5 Discussion

3.5.1 Is *HTR1F* the cataract-causing gene in family CSA92?

HTR1F encodes G-protein coupled serotonin (5-hydroxytryptamine, 5-HT) receptor. Serotonin is a monoamine neurotransmitter that is mainly distributed in the digestive system and the central nervous system. It is involved in the regulation of a variety of complex biological processes, including sleep, mood and appetite. The regulatory function of serotonin is achieved by binding to specific receptors. There are 7 known serotonin receptor families (5-HT_1 to 5-HT_7), each of which mediates a different function of 5-HT. Some of the serotonin receptor families have subtypes, for example, HTR1F belongs to the 5-HT_1 family.

Multiple previous studies have shown the association between serotonin and cataract formation[120,125,126,169,170]. It is reported that both injections of serotonin and serotonin eyedrops lead to dense cataracts in rats[120]. Multiple pieces of evidence suggest selective serotonin reuptake inhibitors (SSRIs), which increase the concentration of serotonin by binding to the serotonin transporter and blocking the reuptake of serotonin, are involved in cataractogenesis. A 19-year-old teenager was reported to develop bilateral cataracts as a result of SSRI treatment[121]. Three population-based studies conducted in Taiwan, the United States and Canada consistently found SSRIs use is associated with an increased risk of cataract development[125,126,169]. This conclusion is supported by a meta-analysis of studies evaluating cataract formation in patients using antidepressants, which contains 447,672 cases and 1,510,391 controls[170]. These studies indicate excess serotonin plays an

important role in cataractogenesis, however, decreasing the concentration of serotonin also has been demonstrated to cause cataract[122,123]. RG 12915[122] and ICT 322[123], both are selective 5-HT_3 antagonists, are proven to result in lens opacities in rats.

Melatonin (N-acetyl-5-methoxytryptamine), a derivative of serotonin, has also been shown to be involved in cataract formation, but has a protective effect. It is a hormone that regulates circadian and seasonal rhythms[171,172]. Melatonin is produced in the pineal gland[173] and the eye. Specifically, the synthesis of melatonin has been detected in the crystalline lens[174,175]. Several studies demonstrated melatonin reduced the generation of cataracts in rats due to its antioxidant properties[129,130,176-180].

Collectively, the evidence indicates the homeostasis of serotonin and melatonin is essential for maintaining lens transparency. Therefore, as a receptor for serotonin, variants in HTR1F that alter its function are likely to affect the concentration of serotonin as well as its downstream product melatonin, leading to the formation of cataracts.

The cataract rates in all *htr1f* (*htr1fa*, *htr1fb*, or both) knockout fish groups were slightly higher than that in the control groups, however no statistically significant difference was reached (Table 3.8). We therefore concluded that knocking out of *HTR1F* orthologs in the zebrafish model did not lead to significant cataract formation.

There are multiple possible reasons for the absence of cataracts in the zebrafish *HTR1F* knockout models. Our observation period for the phenotype was limited to 4 dpf, as the pigment inhibitory effect of PTU gradually disappears on the fifth day after fertilization. It is possible that cataracts occur at a later time point beyond 4 dpf. The pathogenicity of the loss of function

of *HTR1F* may be relatively mild. Inspection of gnomAD[181] (Z score=1.03) and Missense Tolerance Ratio (MTR) viewer[182] (MTR score=0.917) indicates that the *HTR1F* gene is generally tolerant to missense variation, and that the p.Arg289 residue is located in one of the regions of the protein most tolerant to missense variation. Knockout of *HTR1F* orthologs in zebrafish might contribute to the development of cataracts, as there is a clear trend that all the *htr1f* (*htr1fa*, *htr1fb* or both) knockout groups had higher cataract rates than the controls, although the differences didn't reach statistical significance (Table 3.8). It may require more fish for better power to discriminate the small differences. In family CSA92, variants in two other known cataract genes, *HSF4* and *BSFP1*, were also identified. These variants were not directly responsible for cataracts in this family, as they did not segregate with the disease, but they may contribute to the significant cataract phenotypes in the presence of the *HTR1F* mutation. Serotonin receptors consist of 7 subfamilies (5-HT_1 to 5-HT_7), if more than one is expressed in the lens, the effects of knocking out *HTR1F* may be compensated by other redundant serotonin receptors. Notably, it is observed in our lab that the baseline cataract rates of the wild-type zebrafish vary with fish strains. The AB strain fish used in this study has a higher cataract rate compared to TU or TL fish, which are the other two commonly used wild-type fish strains (the baseline cataract rates of the fish strains other than AB were subsequently characterised by Dr Johanna Jones (Menzies Institute for Medical Research, University of Tasmania), unpublished data). The high baseline cataract rate of the AB wild-type fish could also affect the comparison of the cataract percentage between the knockout groups and the controls, obscuring the results. To minimise the potential confounding effect of the baseline cataract rates on the results, knockout experiments in other strains with a lower baseline cataract rate

should be done. In addition, due to the nature of the CRISPR-Cas9 system, it is unlikely to be 100% knockout of the target gene. Editing efficiency analysis shows that the editing in most fish is high, but not complete (Table 3.10 and Table 3.11). It is possible that there are functional wild-type proteins remaining in some cells. Furthermore, the various indels with different sizes (Figure 3.16 and Figure 3.17) and positions (Appendix 6 and Appendix 7) indicate a high level of mosaicism in the edited fish. It is likely that the editing rates were varied in different tissues, which might affect lens phenotypes. In our study, we knocked out *HTR1F* in zebrafish models to mimic the loss of function of this gene. However, the variant identified in CSA92 is possibly a gain-of-function mutation that increases the function of *HTR1F*, rather than reduces it. If so, there wouldn't be cataracts in the zebrafish knockout models. Finally, as shown in Section 3.4.1, approximately 40% of the human *HTR1F* gene is not homologous with its orthologs in zebrafish. This could also partly explain the lack of phenotypes in the zebrafish models, if the proteins play slightly different roles in lens biology between the two species.

Altogether, knocking out the *HTR1F* orthologs in AB zebrafish does not provide clear evidence for *HTR1F* being the cataract causative gene in family CSA92. Further investigations are warranted to obtain additional evidence. In the future, establishing a stable *HTR1F* knockout line in wild-type fish strains with low cataract rates is highly recommended. Moreover, extending the observation period to the larval stage or later may help the detection of cataracts. Experimental investigations in other animal models that are more homologous to the humans at the *HTR1F* gene could also provide more accurate results. Furthermore, it is worth knocking in the specific *HTR1F* variant identified in family CSA92, which will benefit the understanding of the exact mechanism underlying the variant and cataract formation. Re-analysis

of the WES data is also recommended. The WES analysis of family CSA92 was performed by Dr Sharhbanou Javadiyan in 2016[109]. Over the last five years, numerous new variants have been identified which could now be interrogated, it is possible that the *HTR1F* gene may no longer be the top candidate.

3.5.2 Functional analysis of cataract gene in F0 zebrafish using CRISPR-Cas9 RNP

Gene identification for inherited paediatric cataract continues and has accelerated in recent years with the advent of whole exome and genome sequencing to complement family-based linkage approaches. However, functional validation of these genes remains a bottleneck in the discovery pipeline. The rapid development of genome engineering techniques, particularly CRISPR-Cas9, has greatly promoted reverse genetic screening, making routine functional assessment possible. With gene discovery in human patients and their families routinely employing genome-wide approaches, it is not uncommon to identify several plausible candidate genes. Ideally, similar variants in those genes would be found in other similarly affected families, however, that process can take many years until just the right family makes it into research or diagnostic program. In this chapter, we present a rapid zebrafish-based functional validation approach using CRISPR-Cas9 RNP, which can provide valuable evidence for the role of candidate genes in cataract formation.

A major concern of the F0-based gene functional evaluation is the genetic mosaicism in the injected fish. The variety of unpredictable indels induced by CRISPR-Cas9 complicates phenotypic analysis. However, some recent studies

have demonstrated that well-designed CRISPR-Cas9 with editing efficiency approaching saturation can induce almost fully penetrant loss-of-function phenotypes with low off-targets, faithfully simulating the null mutants in F0 embryos[154,167,183,184]. Several critical strategies were applied in our study to maximize editing efficiency. Firstly, the gRNAs were designed and prioritised based on comprehensive criteria, and injection dose and delivery route were optimized using *tyr*-targeting CRISPR-Cas9 RNP. These laid the foundation for high editing efficiency. The final editing efficiencies of either *htr1fa/htr1fb* single-knockouts or double-knockouts were all above 97% (Table 3.10 and Table 3.11). Secondly, by designing gRNAs targeting multiple sites of a single gene and injecting these gRNAs simultaneously, the mutagenicity is further improved with the increase of the probability of including efficient guides and loss-of-function deletions. As shown in Appendix 6, several samples had large deletions, as the sequences between the two gRNAs target sites were completely excised. Finally, we applied CRISPR-Cas9 RNP, which is more efficient compared to the CRISPR-Cas9 mRNA complex[183,185], to achieve saturating mutagenesis.

Another vital factor for successful cataract candidate gene evaluation is the accurate identification of cataract phenotypes in the fish lens. Because of its simplicity and directness, lens images of live embryos are widely used for the phenotypic assessment of zebrafish cataract models. However, there is a lack of standardized reference material available to assess zebrafish cataract development. Multiple factors can interfere with the cataract assessment. Developmental delays are frequently observed in CRISPR-Cas9 RNP injected embryos, and the presence of developing lens fibre cells makes it difficult to distinguish real cataracts. A similar study reported that cataracts detected at 3dpf were absent at 4dpf[143]. It cannot be ruled out that the phenotypes

observed are the manifestation of immature lenses and not developmental cataracts. Therefore, it is suggested to analyse the phenotype at 4dpf or even 5dpf, to ensure the lens is of sufficient maturity to obtain reliable phenotypes. Artefact is another confounding factor. Based on our experience, lesions with a very round and regular shape, highly reflective dots or pitting with blurred edges are considered artefacts. In addition, CRISPR-Cas9 RNP injection itself is possible to lead to non-specific lens lesions. In this study, the false-positive phenotypes were effectively eliminated by applying the *lacZ* negative control. In our assessment criteria, we defined cataracts in the early zebrafish larvae by characterizing defects that lead to gross changes in lens morphology observable by brightfield microscopy with further classification by severity. Examples of false positives are also given.

It is worth noting that despite the ocular structure of zebrafish being quite similar to humans, there are still some differences in the lens formation process between the two. Unlike other vertebrate species, the zebrafish lens does not form a hollow lens vesicle made up of a single layer of ectodermal cells. Instead, the lens placode delaminates from the surface ectoderm as a solid cluster (Figure 3.18A-C, A'-C'). Besides, the primary fibre cells in the zebrafish lens don't form from the epithelial cells in the posterior of the lens vesicle that elongates to fill in the cavity of the vesicle. They are formed by cells in the centre of the cluster, and the cells elongate in a circular way (Figure 3.18D, D')[186]. These distinctions may have an impact on the phenotypic similarity between the zebrafish modelling mutants and the corresponding human diseases.

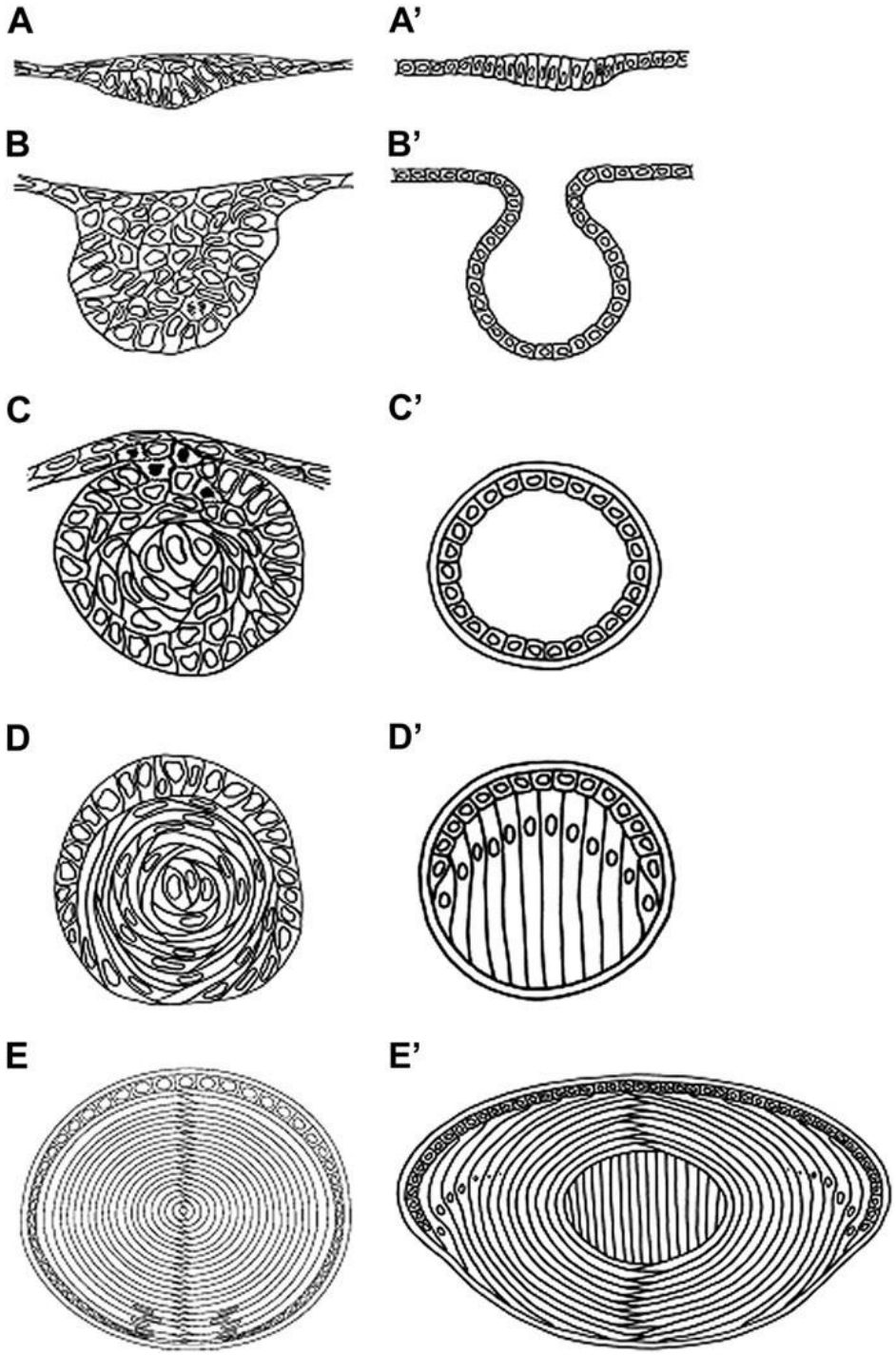


Figure 3.18 The differences in the embryonic development of zebrafish (A-E) and mammalian (A'-E') lenses

(Adapted from Dahm *et al.*, 2007[186])

In summary, a simple CRISPR-Cas9 RNP-based method for rapid and efficient evaluation of cataract candidate genes in F0 zebrafish was developed

in this study. Compared with traditional functional analysis strategies that need to establish stable mutant lines, our method creates significant labour- and time- savings. In addition, detailed criteria to allow standardized assessment of the lens phenotype using brightfield imaging were generated, and reference images for accurate and rigorous phenotypic analysis were provided. The use of single fish strain (AB) and a mild phenotype observed with the *htr1f* gene are the major limitations of the cataract assessment criteria to date. These should continue to be developed to apply to additional strains which may have different developmental rates and to account for more severe cataract that may be observed with different genes.

3.6 Outcomes and significance of this study

The role of *HTR1F* in cataract formation was assessed by knocking out its orthologs *htr1fa* and *htr1fb* in AB zebrafish. Compared with the controls, there was no significant difference in cataract formation in any *htr1f* (*htr1fa, htr1fb* or both) knockout groups. Therefore we determined that knocking out the *HTR1F* orthologs in AB zebrafish does not provide evidence for *HTR1F* being a cataract causative gene. Additional experiments should be conducted to draw a definitive conclusion.

We also developed a strategy for rapid and highly efficient evaluation of cataract candidate genes in F0 zebrafish using a CRISPR-Cas9 RNP. Zebrafish cataract assessment criteria were also generated, which to our knowledge, are the first systematic, definitive cataract assessment criteria for the zebrafish model. The approach developed in this study can be used to generate zebrafish

models for in-depth study of specific genes, or to rapidly assess the cataract forming potential of multiple genes.

Chapter 4

Validation of reported causative genes in two extended paediatric cataract families

4.1 Introduction

Three Australian families, family CRCH13, CSA91 and CSA110, were diagnosed with autosomal dominant paediatric cataracts, with variable disease severity and age of onset/surgery. Putative cataract-causing genes with reduced penetrance have been detected in these families by direct sequencing of genes that have been reported to be associated with pediatric cataracts[85,89,90]. However, no systematic analysis with statistical evidence has been conducted to determine whether the detected genes are the only causes of cataracts in these families, or if any other genetic factors could better explain the observed cataract phenotypes. In addition, new family members have been recruited into the study since the initial investigations, and the clinical information has changed for several individuals since their initial enrolment in the study. Therefore, in this chapter, we updated the clinical information and performed parametric linkage analysis to more comprehensively evaluate the evidence implicating disease-causing genes in family CRCH13, CSA91, and CSA110, prior to investigating the potential modifier genes that contribute to the

diverse disease severity observed in these paediatric cataract families (Chapter 5).

4.1.1 Family CRCH13 background

A c.227G>A (p.R76H) variant (rs121917827) in the *GJA3* gene (NM_021954.4, also known as *CX46*) was previously detected to be segregating with the disease in family CRCH13 by Dr Kathryn Burdon and colleagues in 2004[85]. Briefly, all affected individuals were genotyped at microsatellite markers representing 13 genes and 5 loci that had been associated with paediatric cataract at the time the study was performed, and linkage analysis of the genotyping data indicated linkage to the *GJA3* gene. *GJA3* was sequenced in the affected family members (except for CRCH13.70, CRCH13.73 and CRCH13.75, which were enrolled after Dr Kathryn Burdon's investigation) and the variant 227G>A that resulted in an R76H substitution was detected in all the affected individuals as well as one unaffected relative CRCH13.46 (Figure 4.3).

4.1.2 Background of the CSA families

The CSA families are two related families, CSA91 and CSA110. A c.62G>A (p.R21Q) variant (rs397515626) in the *CRYAA* gene (NM_000394.4) was independently detected by Dr Kate Laurie and colleagues in CSA91[89] and Dr Shari Javadiyan and colleagues in CSA110[90]. This rare coding variant was revealed by direct sequencing of known cataract genes in both families. Ten crystallin genes implicated in cataract formation were screened in family CSA91 and 51 paediatric cataract-associated genes were sequenced in family

CSA110. All the affected individuals in both families and an unaffected member CSA110.2 (Figure 4.5) carried the variant. CSA110.02 had no sign of cataract at the last examination in 2005 when he was 20 years old. No linkage screen has been conducted for these two families.

Given both family CSA91 and family CSA110 are from Adelaide, South Australia, carry the same extremely rare variant (minor allele frequency of $1.2x10^{-5}$ in the gnomAD database[181] (v2.1.1)), and exhibit similar cataract phenotypes we searched for shared ancestry. Haplotype analysis demonstrated all of the mutation carriers in both families shared an identical haplotype carrying the mutant allele, indicating these two families have a common ancestor and belong to a single extended family.

4.2 Hypothesis and aims

Hypothesis:

GJA3 and *CRYAA* are the causative genes for cataracts in family CRCH13 and the CSA families, respectively.

Aims:

1. Update the clinical information of family CRCH13 and the CSA families, and determine the relationship between family CSA91 and CSA110.
2. Performing parametric linkage analysis and whole-genome sequence (WGS) to confirm that *GJA3* and *CRYAA* are the cataract-causing genes in family CRCH13 and the CSA families, respectively.

4.3 Methods

4.3.1 Overall workflow

The overall workflow of this chapter is summarised in Figure 4.1. Briefly, all the family members with available DNA were whole-genome genotyped using SNP arrays. The generated genotyping data was utilised for parametric linkage analysis with the affection status of cataract as the trait. WGS was performed on selected family members and the data within the linkage regions was filtered for rare, pathogenic, segregating variants.

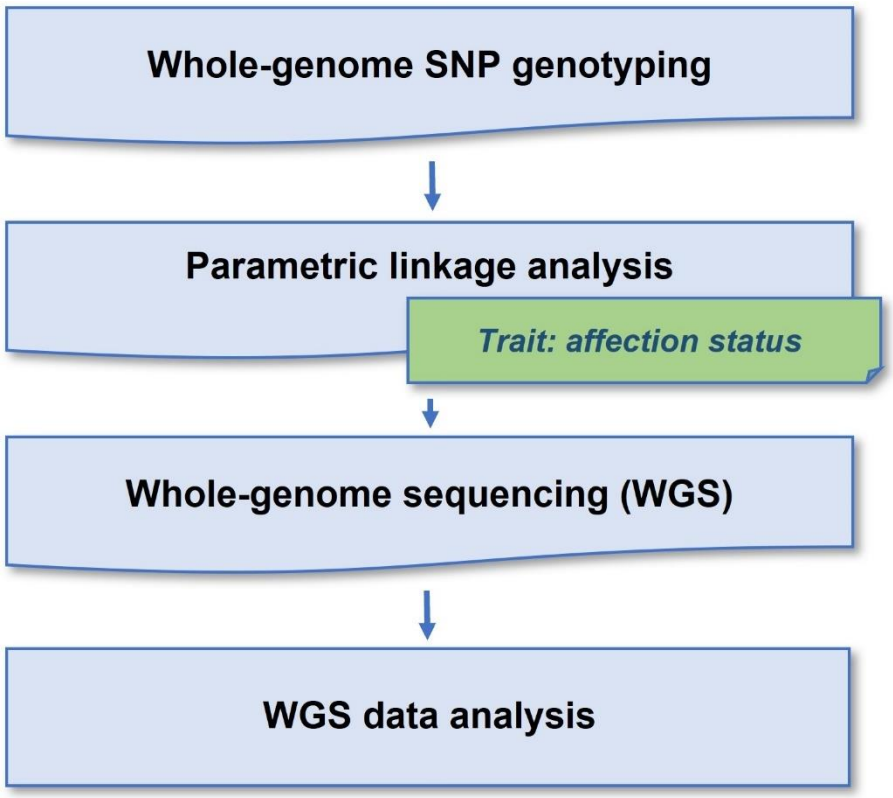


Figure 4.1 Workflow of the key steps in Chapter 4

4.3.2 Whole-genome SNP genotyping

All individuals with available DNA samples were genotyped in-house using the Infinium® Global Screening Array-24 v1.0 (GSA) kit (Table 2.1) according to the Infinium® HTS Assay Protocol Guide[187]. For each sample, 4µl of 50ng/µl genomic DNA was used for the single nucleotide polymorphism (SNP) array genotyping. Samples with limited DNA stock were whole-genome amplified using the Genomiphi method described in Section 2.4.1. The BeadChips were imaged in-house using the Illumina HiScan system (Table 2.2) following the instructions provided by the manufacturer.

The raw fluorescence intensity data (*.idat) generated by the Illumina HiScan system, as well as its version matching bead pool manifest file (GSA-24v1-0_C1.bpm) and cluster file (GSA-24v1-0_C1_ClusterFile.egt) were analysed in Illumina GenomeStudio® 2.0 Software. The manifest file and cluster file were downloaded from the Illumina website[188]. Quality control (QC) was performed following the Infinium® Genotyping Data Analysis guide provided by Illumina[189]. All samples with call rates less than 0.99 were excluded prior to recalling genotypes. Poorly performing SNPs were removed using the following criteria: (1) call frequency < 1; (2) AB R Mean <0.5; (3) AB T Mean ranging from 0-0.2 and 1-0.8; (4) have Mendelian inheritance errors; (5) have reproducibility errors. X-linked SNPs in the male samples were manually inspected and the heterozygous ones were removed. SNPs on chromosome (chr) Y, chr0 and chrXY (pseudo-autosomal region of chrX), as well as mitochondrial SNPs were also omitted as they were not used in downstream linkage analysis. After QC, the final genotyping data was exported to the GenomeStudio Final Report.

4.3.3 Genotyping data processing

The Final Report genotype data generated by GenomeStudio was converted to PLINK format files. The final report was first converted into PLINK long-format files using the R script (written by Ryan Schubert for the Wheeler Lab) presented in Appendix 8 in conjunction with a manually made SampleID_Gender.csv file containing sample identify and sex information. The PLINK long-format files were further converted to .map and .ped files in PLINK v1.9[190,191]. Triallelic SNPs were removed from .map and .ped files in PLINK v1.9 as PLINK is specialised to the analysis of biallelic SNPs and the information from the third allele can not be used.

It was found that the physical location (centiMorgan (cM)) of the SNPs provided by the Infinium® Global Screening Array-24 v1.0 (GSA) kit was incorrect. We therefore corrected the cM information in the .map file according to the correct genetic distance provided in GSA-24v1-0_C1.csv_Physical-and-Genetic-Coordinates.txt (downloaded from the Illumina website[188]).

Linkage disequilibrium (LD) based SNP pruning was performed in PLINK v1.9 to remove SNP pairs in high LD. An LD-pruned SNP set was obtained from the pruning of an unrelated cohort consisting of 129 individuals also genotyped with Illumina Infinium® Global Screening Array-24 v1.0. The LD pruning was performed using the `--indep-pairwise` function, with the window size in SNPs, the number of SNPs to shift the window at each step and the r^2 threshold set to 50, 5 and 0.5, respectively. The subset of LD-pruned SNPs was extracted from the family genotyping data using the PLINK `--extract` function and applied in subsequent linkage analysis.

4.3.4 Relationship testing

The familial relationships were assessed in PLINK v1.9 using `--genome` function to verify the collected pedigree information as well as to detect any possible sample contamination, swap or duplication.

In order to better understand the relationship between the two distantly related families — CSA91 and CSA110, the kinship between CSA91 and CSA110 was estimated by Dr Nicholas Blackburn (Menzies Institute for Medical Research, University of Tasmania) using the Identical by Descent via Identical by State (IBIS) v1.20.9[192] program.

4.3.5 Inheritance error checking

Mendelian inheritance errors were detected in Mega2 (Version 6.0.0)[193,194] software and any inconsistent genotypes were removed. All autosomal genotyping data required for downstream linkage analysis were exported by Mega2 into MERLIN[195] format files. Since cataracts in the target families were inherited in an autosomal dominant manner, X-chromosome data was excluded from subsequent analysis. In the MERLIN v1.1.2[195] software, genotyping errors were detected again using the `--error` function and removed by the PEDWIPE program.

4.3.6 Parametric linkage analysis

Genome-wide multipoint parametric linkage analysis of autosomal genotyping data was performed on family CRCH13 and the CSA families in the MERLIN v1.1.2[195] program. Prior to analysis, the data quality and format

consistency of input files were verified using the PEDSTATS program in MERLIN. A rare, dominant disease model was applied in the linkage analysis. Specifically, the disease allele frequency was set to 0.0001, and the penetrances were set to 0.0001, 0.90 and 1.0, which represents the probability of affected family members carrying 0, 1 and 2 copies of the disease allele, respectively.

Linkage plots were drawn to visualize the analysis results exported from MERLIN. Heterogeneity logarithm of the odds (HLOD) scores of 3.3 and 2 were respectively employed as thresholds for genome-wide significant and suggestive linkages, according to guidelines that have been widely applied for interpreting linkage results[196]. The boundary on both sides of the linkage region was defined at the site with an HLOD score equal to the maximum HLOD score minus one.

4.3.7 Whole-genome sequencing (WGS)

Individuals from different branches of the extended families, presenting a range of disease severity and with available DNA were selected for WGS. Specifically, five individuals from CRCH13 (CRCH13.07, CRCH13.09, CRCH13.73, CRCH13.74, and CRCH13.75) and four from the CSA families (CSA91.03, CSA91.07, CSA110.05, and CSA110.07) were whole-genome sequenced. These individuals are marked with yellow outlines in Figure 4.3 and Figure 4.5. WGS was performed using the TruSeq Nano Library Prep v2.5 (Table 2.1) with 150 bp paired-end sequencing on the Illumina Novaseq 6000 system (100Gb) (Table 2.2) at the Australian Genome Research Facility.

4.3.8 WGS data processing

4.3.8.1 Raw data alignment and variant calling

The generated raw data was manipulated using the bcbio-nextgen toolkit (v1.1.9)[197]. Briefly, paired-end reads were aligned to the human reference genome (hg19) using BWA-MEM (v0.7.17)[198]. Duplicate alignments were marked using Biobambam's bamsormadup (v2.0.87)[199]. Single nucleotide variants (SNVs) and small insertions and deletions (indels) were called for each sample using GATK (v4.1.4.1)[200] HaplotypeCaller, and then all samples were jointly genotyped using GATK GenotypeGVCFs. Structural variants (SVs) were called in LUMPY[201] software.

4.3.8.2 Extracting data from linkage regions of interest

Variants under the genome-wide significant linkage peaks and suggestive linkage peaks of each extended family were extracted from the WGS data using BCFtools v1.12[202]. Multi-allelic variants were separated into individual entries, and indels were left-aligned and normalized to the GRCh37/hg19 reference genome.

4.3.8.3 Categorising WGS data by sequencing quality

The extracted and normalized WGS data were categorised by read depth (DP) and genotype quality (GQ) with the VariantFiltration program in GATK v4.1.3.0[203,204] and Java[205] program. This procedure was performed using the R script in Appendix 9. The specific criteria for classification are as follows:

- ***High confidence:*** DP ≥ 10 and GQ ≥ 20

- ***Low confidence:*** DP $<$ 10 and GQ $<$ 20
- ***Low coverage but high quality:*** DP $<$ 10 and GQ $\geq$ 20
- ***High coverage but low quality:*** DP $\geq$ 10 and GQ $<$ 20

4.3.8.4 Variant annotation

Variants were annotated in ANNOVAR (2019Oct24)[206] using the GRCh37/hg19 database. The detailed commands are listed in Appendix 10. More specifically, variant basic information, including position, reference allele, alternative allele, function and gene was annotated using the UCSC RefGene database, a subset of NCBI's Reference Sequence (RefSeq) project[207]. Variant allele frequency was annotated using population minor allele frequency (MAF) data in the Genome Aggregation Database (gnomAD) v2.1.1[181], the Exome Aggregation Consortium (ExAC) v0.3[208] database, the 1000 Genomes (European, version 2015Aug)[209] database and the Kaviar Genomic Variant (version 20150923)[210] database. Functional consequences of variants were annotated with Combined Annotation Dependent Depletion (CADD) (v1.3)[115,211] database and Functional Analysis through Hidden Markov Models (FATHMM)[212] database.

4.3.9 Variant filtering

4.3.9.1 Single nucleotide variants

Cataract-related genes published in Cat-Map (https://cat-map.wustl.edu/)[31] (an online reference database for inherited cataract) within the linkage regions were first prioritised for single nucleotide variants (SNVs)

filtering, otherwise the filtering was performed in the entire linkage region. Exonic, ncRNA exonic, intronic, ncRNA intronic, intergenic variants and variants located in 5' or 3' untranslated region (UTR) were filtered. The filtering pipeline is summarized in Figure 4.2 and the criteria are as follows:

- ***High confidence.*** The sequencing data was first filtered based on sequencing quality. The data quality was categorised as described in Section 4.3.8.3. Only "High confidence" (DP $\geq$ 10 and GQ $\geq$ 20) variants were saved for subsequent filtering.
- ***Rare or novel.*** The variants were next filtered based on MAF. The MAF annotation was described in Section 4.3.8.4. Variants with MAF $\leq$ 0.01 in gnomAD (rare) or absent in all the available databases used for annotation (novel) were saved for the next step of filtering.
- ***Predicted pathogenic.*** The deleteriousness of variants was annotated as described in Section 4.3.8.4. For exonic and ncRNA exonic variants, a CADD score $\geq$ 10 was set as the threshold, that is, the variants are predicted to be the 10% most deleterious possible substitutions in the human genome. For intronic, ncRNA intronic, intergenic variants and variants located in 5' or 3' untranslated region (UTR), a CADD score $\geq$ 20 was set as the threshold, indicating the variants in the top 1% most deleterious possible substitutions in the human genome.
- ***Segregating with the disease.*** Variants were filtered to the ones that were fully segregated with disease in the selected WGS individuals. Variants that were shared by all the affected family members, but absent in the unaffected family members were saved.

Variants that passed all the above filtering criteria are considered candidates for the cause of cataracts.

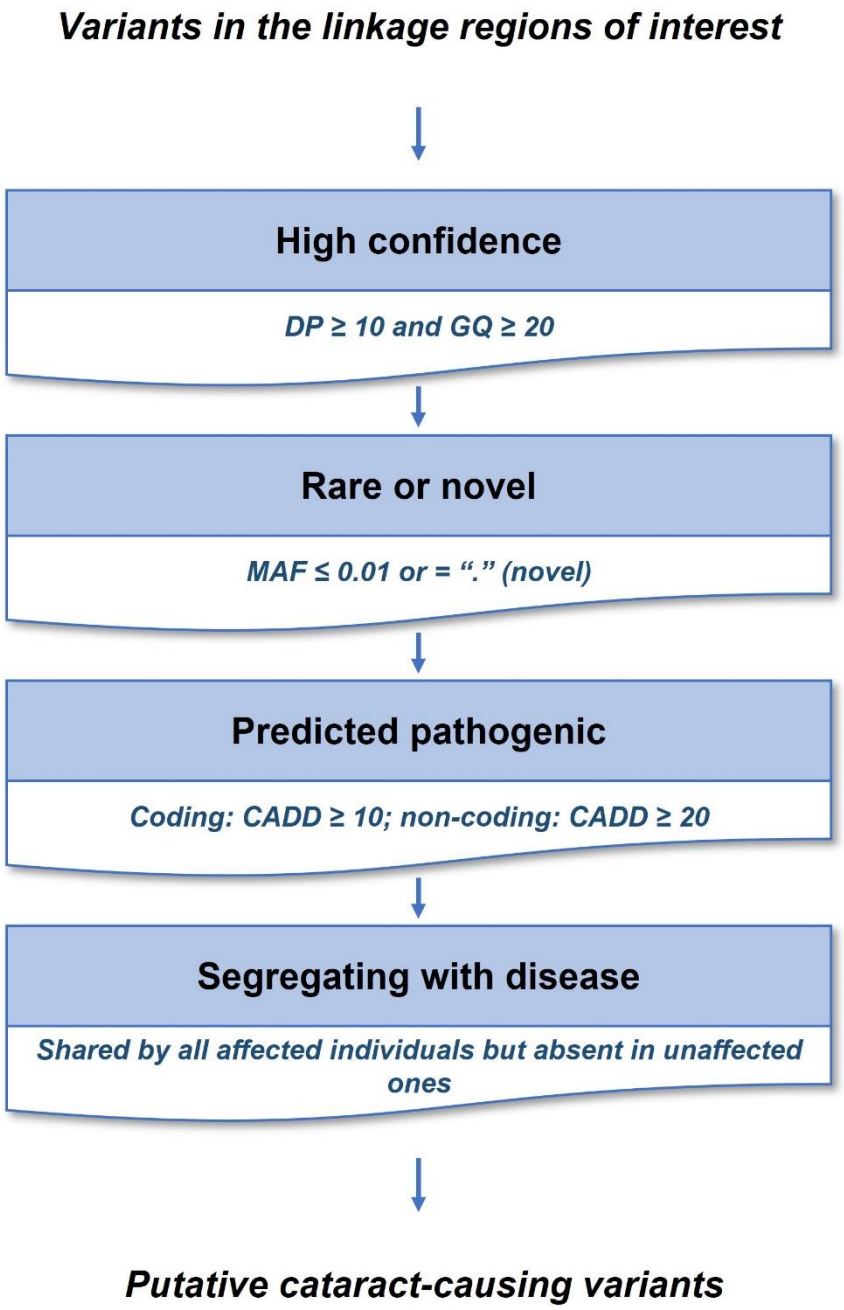


Figure 4.2 Single nucleotide variants filtering pipeline

Single nucleotide variants extracted from linkage regions of interest were filtered to identify putative causative variants for cataract. The contents with the blue background are the filtering criteria, and the corresponding parameters used for filtering are listed below them.

4.3.9.2 Structural variants

Structural variants (SVs) were extracted from the linkage regions of interest (Section 4.3.8.2) and manually filtered to the ones that were fully segregated with disease in the whole-genome sequenced individuals.

4.4 Results

4.4.1 Family phenotype and clinical information update

4.4.1.1 Family CRCH13

The multi-generational family CRCH13 presented with autosomal dominant paediatric cataract. Affected individuals showed diverse phenotypes, ranging from mild cortical pulverulent cataract to dense fetal nuclear cataract surrounding with cortical lamellar opacity. Great variability was also observed at the age of diagnosis/surgery, with the age of diagnosis ranging from birth to 19 years old, and the age of surgery ranging from 10 to 67 years. The updated pedigree of CRCH13 is shown in Figure 4.3. The clinical information and lens photos of affected individuals are displayed in Table 4.1 and Figure 4.4. Compared to the clinical data published by Burdon *et al.*[85], the affection statuses of 5 individuals were changed (CRCH13.38, CRCH13.48 and CRCH13.69 were updated from unaffected to affected; CRCH13.21 and CRCH13.51 were updated from unaffected to unknown), and 5 new family members (CRCH13.70, CRCH13.71, CRCH13.73, CRCH13.74 and CRCH13.75) were recruited in the study. The information of the age of diagnosis/surgery was also checked and supplemented.

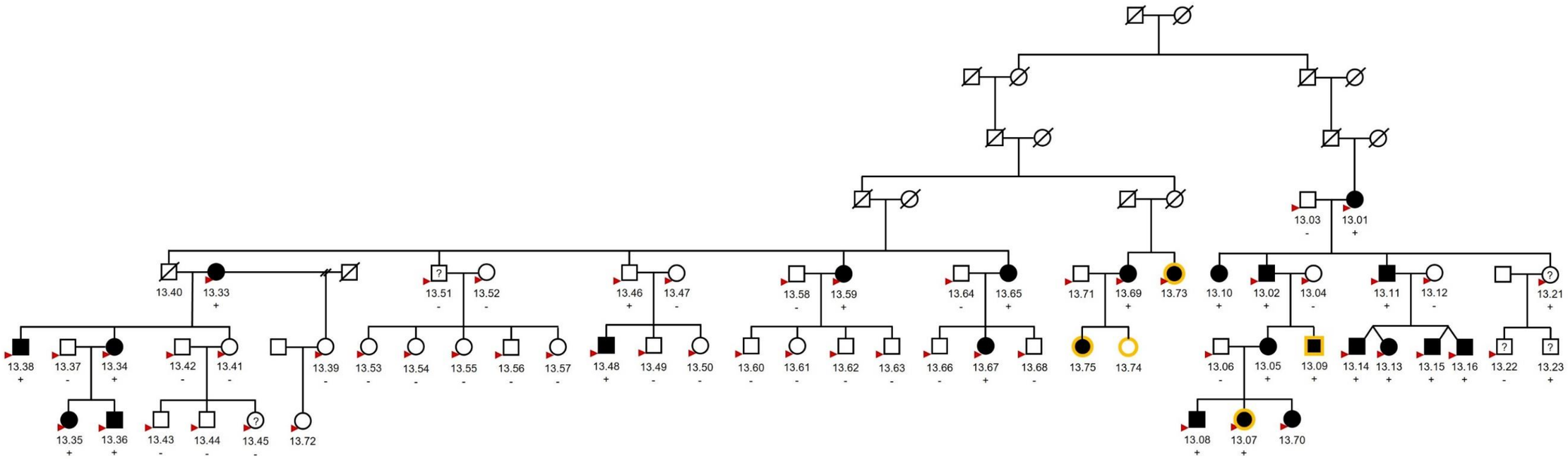


Figure 4.3 CRCH13 family pedigree

Males are indicated by squares and females by circles. Solid black symbols indicate affected individuals while open symbols indicate unaffected individuals and symbols with "?" correspond to individuals whose affection status is unknown. "+" indicates the *GJA3* R76H mutation carriers, "-" indicates non-carriers. Symbols with red triangles represent individuals with genotyping data, and symbols with yellow outlines represent individuals with whole-genome sequencing (WGS) data.

Table 4.1 Family CRCH13 clinical information

Affected member	Cataract phenotype	Age of diagnosis	Age of surgery (right/left eye)	Age of recruitment
CRCH13.01	-	5.5 years	67 years	21 years
CRCH13.02	-	7 years	45 years/42 years	54 years
CRCH13.05	Nuclear fine gold dots	2 years	26 years/no surgery at recruitment	30 years
CRCH13.07	Multiple irregular lamella opacities; fetal nuclear cataract with brownish-white dots	7 months	No surgery at 7 years	2 years
CRCH13.08	Fetal nuclear cataracts, dense inferiorly and superiorly, surrounded with cortical lamellar cataract	11 months	No surgery at 13 years	4 years
CRCH13.09	-	2 years	13 years/14 years	28 years
CRCH13.10	-	19 years	No surgery at recruitment	59 years
CRCH13.11	-	7 years	35 years/37 years	50 years
CRCH13.13	-	6 months	14 years	21 years
CRCH13.14	-	6 months	14 years	21 years
CRCH13.15	-	6 months	11 years/17 years	18 years
CRCH13.16	-	6 months	21 years/ 28 years	18 years
CRCH13.33	-	8 years	14 years/15 years	78 years

CRCH13.34	-	5 years	36 years/37 years	46 years
CRCH13.35	-	12 months	10 years/12 years	19 years
CRCH13.36	-	At birth	10 years	11 years
CRCH13.38	Bilateral pulverulent cataracts with good visual acuity	-	No surgery at recruitment	48 years
CRCH13.48	Right eye: cortical and pulverulent cataract with blue dots; left eye: pulverulent cataract	-	No surgery at recruitment	50 years
CRCH13.59	Bilateral pulverulent nuclear cataracts	-	No surgery at 77 years	73 years
CRCH13.65	Bilateral cortical (few) pulverulent cataracts	-	No surgery at 74 years	70 years
CRCH13.67	Nuclear sutural cataracts	7 years	35 years/36 years	44 years
CRCH13.69	Right eye: congenital nuclear cataract; left eye: pulverulent cataract with good visual acuity	6 years	No surgery at recruitment	61 years
CRCH13.70	Fetal nuclear lamellae with fine white flecks throughout the nucleus; minimal cortical changes	12 months	No surgery at 13 years	4 years
CRCH13.73	Bilateral pulverulent cataracts	-	No surgery at recruitment	63 years
CRCH13.75	Bilateral very small nuclear opacities	-	No surgery at recruitment	41 years

- : data unavailable. CRCH13.38, CRCH13.48, CRCH13.59, CRCH13.65, CRCH13.73, and CRCH13.75 with no available age of onset data were diagnosed as paediatric cataracts based on the observed cataract phenotype.

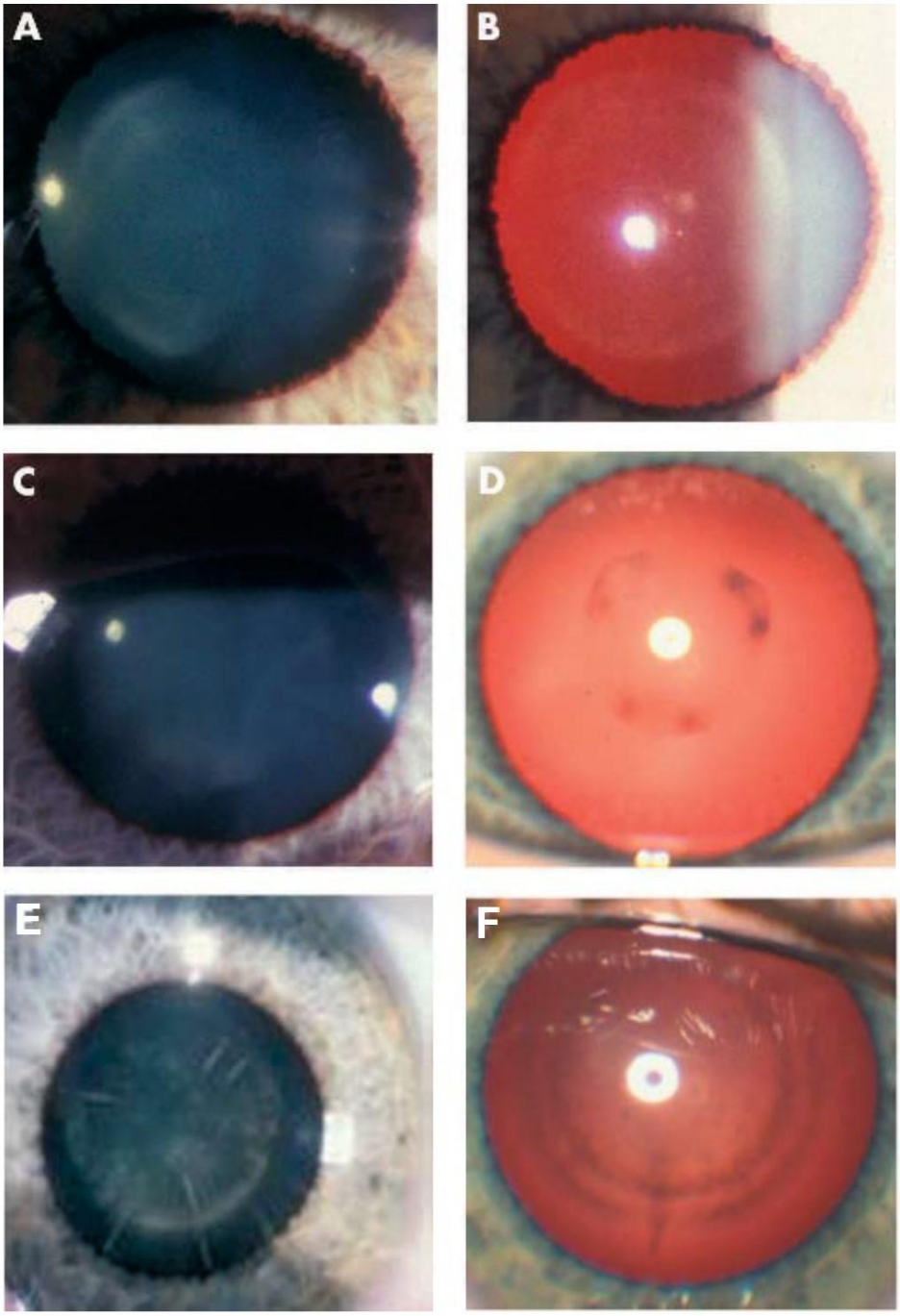


Figure 4.4 Lens photos of family CRCH13

The lens photos show a range of pulverulent cataract phenotypes of pedigree CRCH13 (left: direct illumination view; right: retro-illumination view). **A** and **B**. Mild pulverulent phenotype of the nucleus of CRCH13.05. A mild pulverulent phenotype of the nucleus with cortical lamella opacities was observed in CRCH13.08 (**C** and **D**) and CRCH13.07 (F). **E**. Pulverulent phenotype in the nucleus with cortical riders of CRCH13.36. (Adapted from Burdon *et al.*, 2004[85]).

4.4.1.2 The CSA families

The relationship between family CSA91 and CSA110 was updated and the pedigree is shown in Figure 4.5. The results of the IBIS kinship analysis (Appendix 11) suggested that there are 6^{th} and 7^{th}-degree relationships between family members in CSA91 and CSA110, confirming these two families belong to a single large family. The most related pairs between the two families are CSA91.02 with both CSA110.01 and CSA110.05. The kinship coefficients between CSA91.02/CSA110.01 and CSA91.02/CSA110.05 are 0.009 and 0.008 respectively, indicating that these people belong to the 6^{th}-degree relationship. From here on, these two families will be regarded as subfamilies of one big pedigree, family CSA91/110.

Both subfamilies presented with lamellar cataract, in an autosomal dominant pattern. A variety of disease severity was observed in CSA91 and CSA110, with a similar trend towards earlier onset in the younger generation. In CSA91, the age of diagnosis ranged from 18 months to 61 years, and the age of surgery ranged from 5 years to no surgery at 61 years (the time of recruitment). In CSA110, most of the data of age of diagnosis was missing, and the age of surgery ranged from 4 to 75 years. Details of the clinical information are summarized in Table 4.2 and the example lens photos refer to Figure 4.6. Compared to the clinical information published by Laurie *et al.*[89] and Javadiyan *et al.*[90], the data of the age of diagnosis/surgery of affected individuals were supplemented and verified.

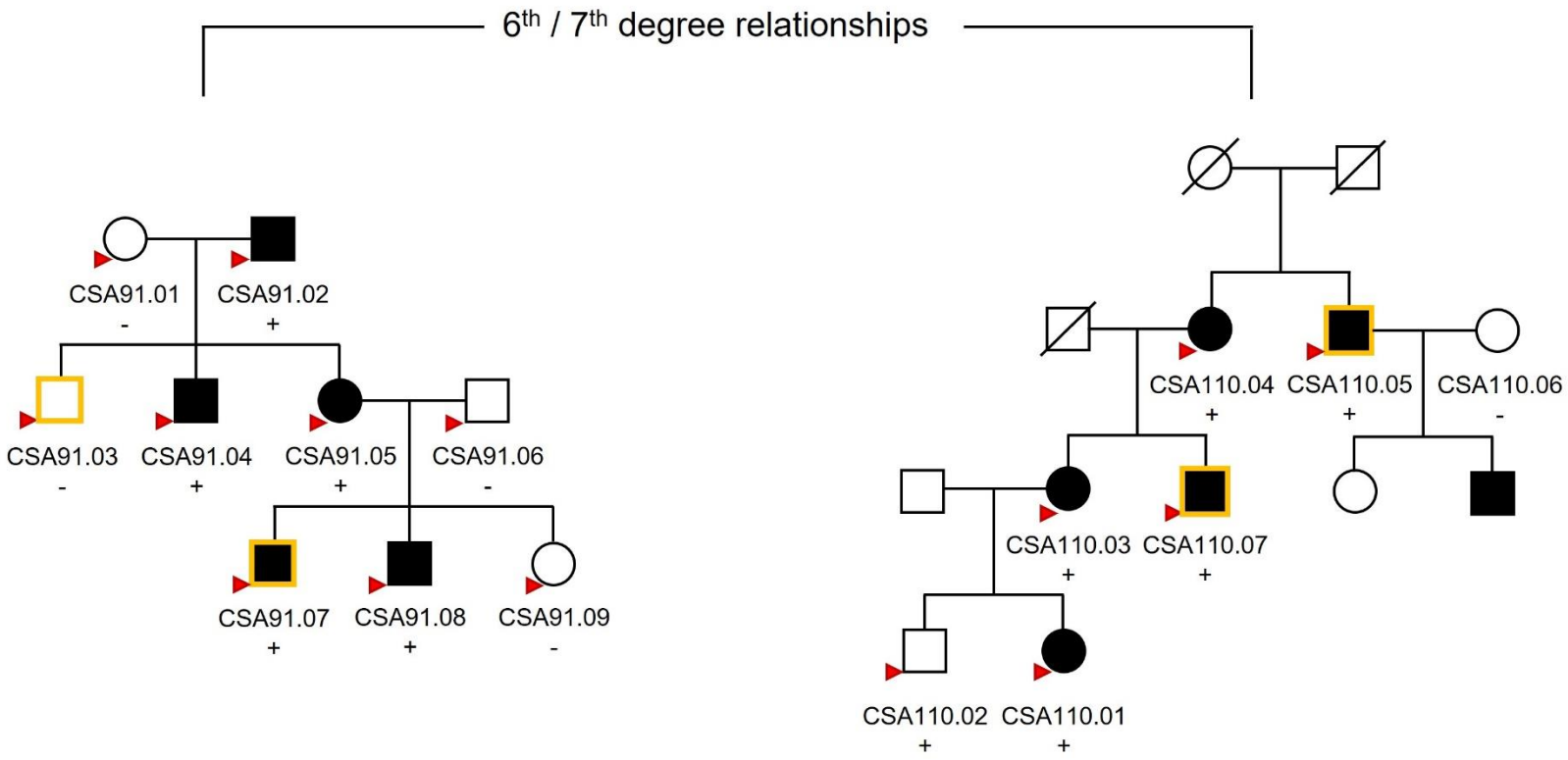


Figure 4.5 Pedigree of family CSA91/110

Males are indicated by squares and females by circles. Solid black symbols indicate affected individuals while open symbols indicate unaffected individuals. "+" indicates the *CRYAA* R21Q mutation carriers, "-" indicates non-carriers. Symbols with red triangles represent individuals with genotyping data, and symbols with yellow outlines represent individuals with whole-genome sequencing (WGS) data.

Table 4.2 Clinical information of family CSA91/110

Affected member	Cataract phenotype	Age of diagnosis	Age of surgery (right/left eye)	Age of recruitment
CSA91.02	Minimal lamellar white dots	61 years	No surgery at recruitment	61 years
CSA91.04	Mild lamellar white dots	30 years	No surgery at recruitment	34 years
CSA91.05	Mild lamellar opacity consisting of fine white dots in a single lamellar shell	-	No surgery at recruitment	32 years
CSA91.07	Nuclear lamellar cataracts	3 years	5 years	9 years
CSA91.08	Dense nuclear lamellar cataracts	18 months	7 years/6 years	6 years
CSA110.01	Lamellar cataracts	7 years	16 years	16 years
CSA110.03	-	12 years	46 years	51 years
CSA110.04	-	-	75 years	76 years
CSA110.05	Bilateral mild lamellar cataracts with anterior cortical spokes	-	-/61 years	74 years
CSA110.07	Severe bilateral cataracts	-	4 years	44 years

- : data unavailable. CSA91.02, CSA91.05, CSA110.4, CSA110.5 and CSA110.7 were diagnosed with paediatric cataracts by the cataract phenotype and/or the presence of the paediatric cataract-associated mutation, *CRYAA* c.62G>A (p.R21Q).

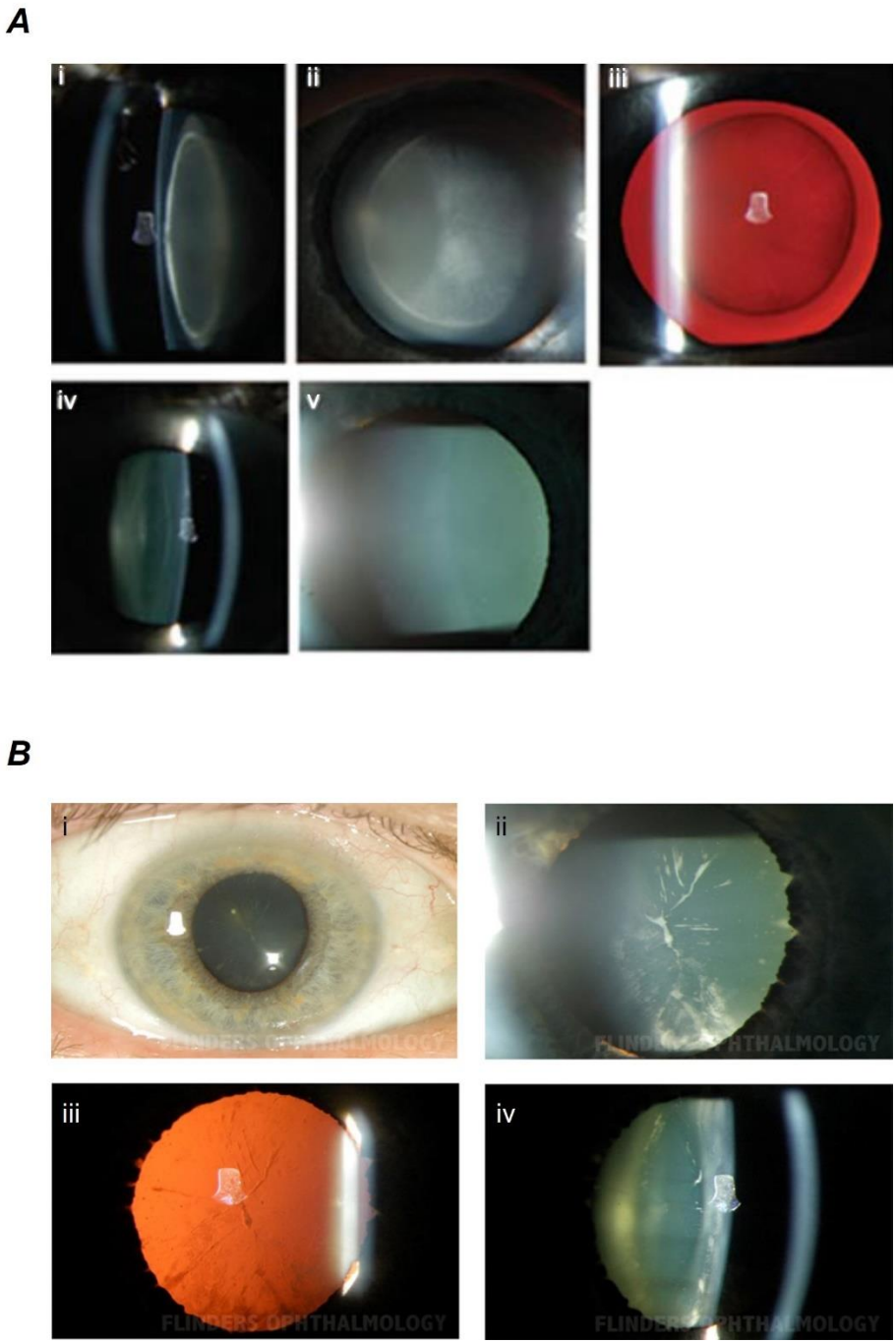


Figure 4.6 Lens photos of family CSA91/110

A. Example lens photos of family CSA91. Panel **i**, **ii**, and **iii** in section **A** show a moderate lamellar opacity with involvement of fetal nucleus in CSA91.08. (**i** and **ii**: direct illumination view; **iii**: retro-illumination view). Panel **iv** and **v** in section **A** show a milder phenotype of cataract in CSA91.04 which is lamellar opacity (**iv**) with fine white dots (**v**) (**iv** and **v**: direct illumination view). **B**. Example lens photos of family CSA110. Panel **i-iv** in section **B** shows the lamellar opacity with anterior cortical spokes observed in CSA110.05. (**i**, **ii** and **iv**: direct illumination view; **iii**: retro-illumination view). (Images in section **A** from Laurie *et al.*, 2013[89]; images in section **B** from Javadiyan *et al.*, 2016[90])

4.4.2 Whole-genome genotyping data

A total of 72 individuals (CRCH13: 56 individuals; CSA91:9 individuals, CSA110:7 individuals) were whole-genome genotyped and raw genotyping data of 618,540 SNPs were generated for downstream analysis. Two individuals (CRCH13.10 and CRCH13.09) presented with genotyping call rates < 0.99 in the GenomeStudio were discarded from subsequent analysis. All the 1st-degree relationships were verified, after correcting the two samples (CRCH13.64 and CRCH13.65) that were detected to be swapped by relationship testing. 931 and 58 Mendelian inconsistent genotypes were detected and removed from family CRCH13 and CSA91/110, respectively.

After GenomeStudio QC (Section 4.3.2), data processing (Section 4.3.3), relationship testing (Section 4.3.4) and inheritance error checking (Section 4.3.5), 54 individuals and 141,213 SNPs were included in the linkage analysis of family CRCH13, and 16 individuals (CSA91: 9 individuals, CSA110: 7 individuals) and 189,161 SNPs were included in the linkage analysis of family CSA91/110. Individuals with genotyping data are indicated with red triangles in the pedigrees shown in Figure 4.3 and Figure 4.5. The details of the SNPs used for linkage analysis are listed in Table 4.3.

Table 4.3 The number of finalised SNPs used for linkage analysis

	CRCH13	**CSA91/110**
Chr1	10,931	15,154
Chr2	11,762	15,552
Chr3	10,537	13,409
Chr4	9,400	12,504
Chr5	8,782	11,924
Chr6	9,006	12,280
Chr7	8,409	10,712
Chr8	7,888	10,340
Chr9	6,599	8,575
Chr10	7,187	9,803
Chr11	7,076	9,282
Chr12	7,434	9,742
Chr13	5,595	7,355
Chr14	4,975	6,512
Chr15	4,352	5,840
Chr16	4,273	6,087
Chr17	3,619	5,104
Chr18	4,378	6,012
Chr19	2,353	3,456
Chr20	3,273	4,675
Chr21	1,901	2,617
Chr22	1,483	2,226
Total	**141,213**	**189,161**

4.4.3 Family CRCH13

4.4.3.1 Parametric linkage analysis

Individuals that were not informative to the linkage analysis were removed and the remaining family members were divided into 5 subfamilies (Figure 4.7) to fit into the MERLIN software[195]. MERLIN has a bit-size limit of 24 bits for a single family. The bit-size is defined as two times the number of individuals with parents presented in the pedigree minus the number of pedigree founders. The multiple parametric linkage analysis of family CRCH13 was performed under a rare dominant inheritance model (Section 4.3.6) with cataract affection status as the trait. The affection status of each family member participating in the parametric linkage analysis was summarised in Table 4.4. All subfamilies were analysed together and the overall HLOD score was calculated.

Table 4.4 Affection status assigned to members involved in the parametric linkage analysis of family CRCH13

Affected	CRCH13.01, CRCH13.02, CRCH13.05, CRCH13.07, CRCH13.08, CRCH13.09, CRCH13.11, CRCH13.13, CRCH13.14, CRCH13.15, CRCH13.16, CRCH13.33, CRCH13.34, CRCH13.35, CRCH13.36, CRCH13.38, CRCH13.48, CRCH13.59, CRCH13.65, CRCH13.67, CRCH13.69, CRCH13.70, CRCH13.73, CRCH13.75
Unaffected	CRCH13.03, CRCH13.04, CRCH13.06, CRCH13.12, CRCH13.37, CRCH13.40, CRCH13.42, CRCH13.46, CRCH13.47, CRCH13.64, CRCH13.66, CRCH13.71, CRCH13.74
Unknown	CRCH13.21, CRCH13.51

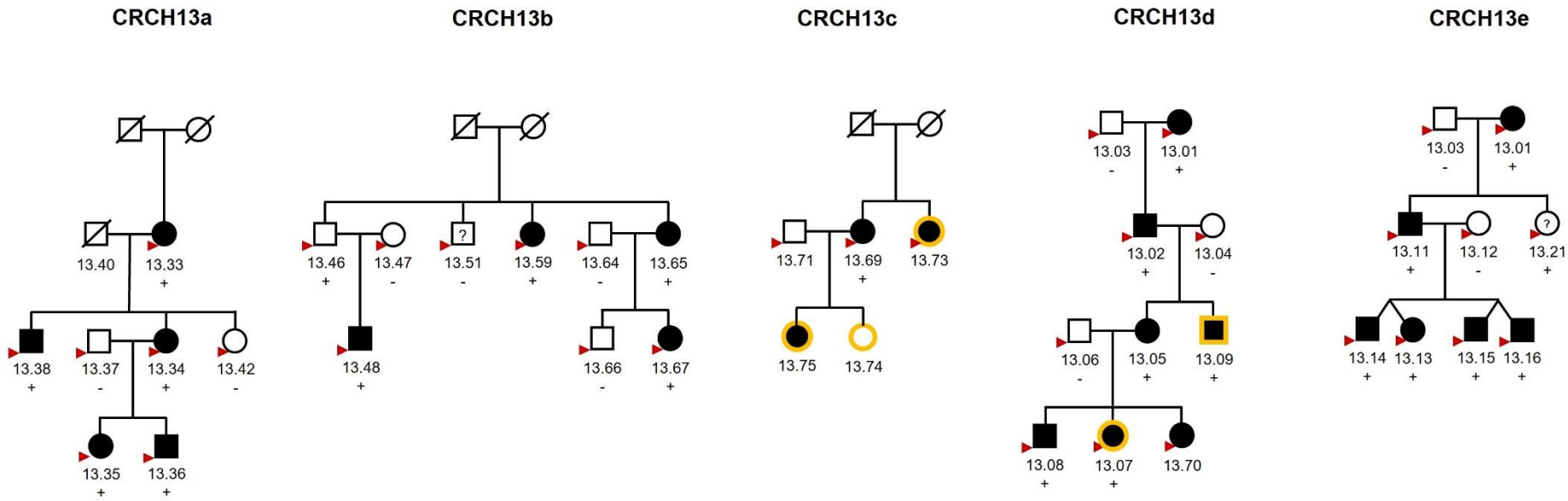


Figure 4.7 Subfamilies consisting of **individuals involved in the parametric linkage analysis of family CRCH13**

Individuals that were not informative to the linkage analysis were removed and the remaining family members were divided into 5 subfamilies (CRCH13a, CRCH13b, CRCH13c, CRCH13d, and CRCH13e) to fit into the MERLIN software[195]. Males are indicated by squares and females by circles. Solid black symbols indicate affected individuals while open symbols indicate unaffected individuals and symbols with "?" correspond to individuals whose affection status is unknown. "+" indicates the *GJA3* R76H mutation carriers, "-" indicates non-carriers. Symbols with red triangles represent individuals with genotyping data, and symbols with yellow outlines represent individuals with whole-genome sequencing (WGS) data.

Two linkage peaks were identified across all the autosomes (Figure 4.8). A significant linkage peak with an HLOD score of 3.586 was detected on chr13 and a suggestive linkage peak with an HLOD score of 2.62 was detected on chr10 (Table 4.5).

Table 4.5 Linkage regions identified in the parametric linkage analysis of affection status in family CRCH13

Chr	Position (hg19)	Size (bp)	Maximum HLOD	Significance
13	chr13:19,121,858-20,793,969	1,672,112	3.586	Significant
10	chr10:113,052,939-119,170,304	6,117,366	2.62	Suggestive

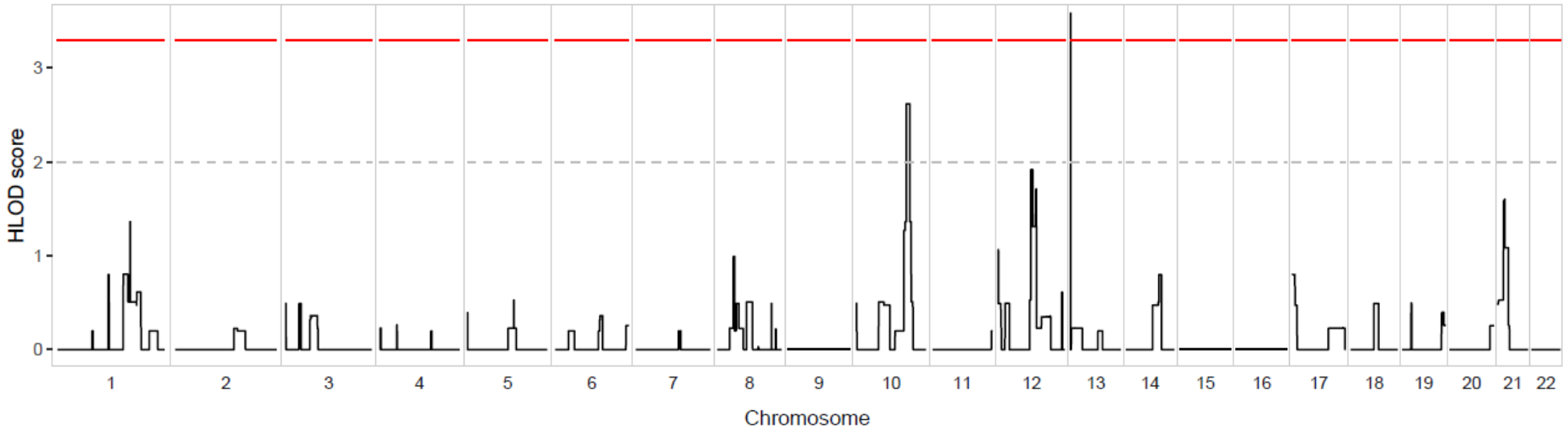


Figure 4.8 Autosome-wide parametric linkage analysis of affection status in family CRCH13

The X-axis represents the position in the genome ordered by chromosome and the y-axis represents the heterogeneity logarithm of the odds (HLOD) score. Significant (HLOD = 3.3) and suggestive (HLOD = 2) linkage thresholds are indicated with red solid line and grey dashed line, respectively.

4.4.3.2 Cataract causing gene identification

Under the significant linkage peak on chr13 (Table 4.5), two known cataract genes, *GJA3* and *GJB6*, were detected according to the online cataract gene database — Cat-Map[31]. The maximum HLOD score was detected on the *GJA3* gene, which was previously reported as the cause of cataract in this family. WGS data analysis showed that the *GJA3* R76H variant detected in all affected individuals included in the previous study was absent in the subsequently recruited affected individuals CRCH13.73 and CRCH13.75. No other rare, pathogenic, segregating variants were identified in *GJB6* and the rest of the significant linkage region.

Under the suggestive linkage peak on chr10 (Table 4.5), no genes known to be related to cataract were detected. We therefore filtered SNVs in the entire linkage region using the WGS data, but failed to detect any rare, pathogenic SNVs that fully segregated with the disease. Furthermore, only 5 SVs were called in this region and were manually excluded as none of them fully segregated with disease.

4.4.4 Family CSA91/110

4.4.4.1 Parametric linkage analysis

CSA91 and CSA110 were analysed together as a single extended family in the multiple parametric linkage analysis. CSA110.06 was excluded from the linkage analysis as she was not related to any other genotyped individuals by blood. A rare dominant inheritance model (Section 4.3.6) was employed and the affection status of cataract was set as the trait for the analysis. The

affection status of each family member participating in the parametric linkage analysis was summarised in Table 4.6.

Table 4.6 Affection status assigned to members involved in the parametric linkage analysis of family CSA91/110

Affected	CSA91.02, CSA91.04, CSA91.05, CSA91.07, CSA91.08, CSA110.01, CSA110.03, CSA110.04, CSA110.05, CSA110.07
Unaffected	CSA91.01, CSA91.03, CSA91.09, CSA110.02

Autosome-wide linkage analysis (Figure 4.9) detected a suggestive linkage peak on chr11, and a peak with HLOD>1 was identified on chr21. The detailed information of these two peaks was summarized in Table 4.7.

Table 4.7 Linkage regions identified in the parametric linkage analysis of affection status in family CSA91/110

Chr	Position (hg19)	Size (bp)	Maximum HLOD	Significance
11	chr11:120,783,824-124,740,659	3,956,836	2.585	Suggestive
21	chr21:43,413,453-48,099,710	4,686,258	1.581	N/A

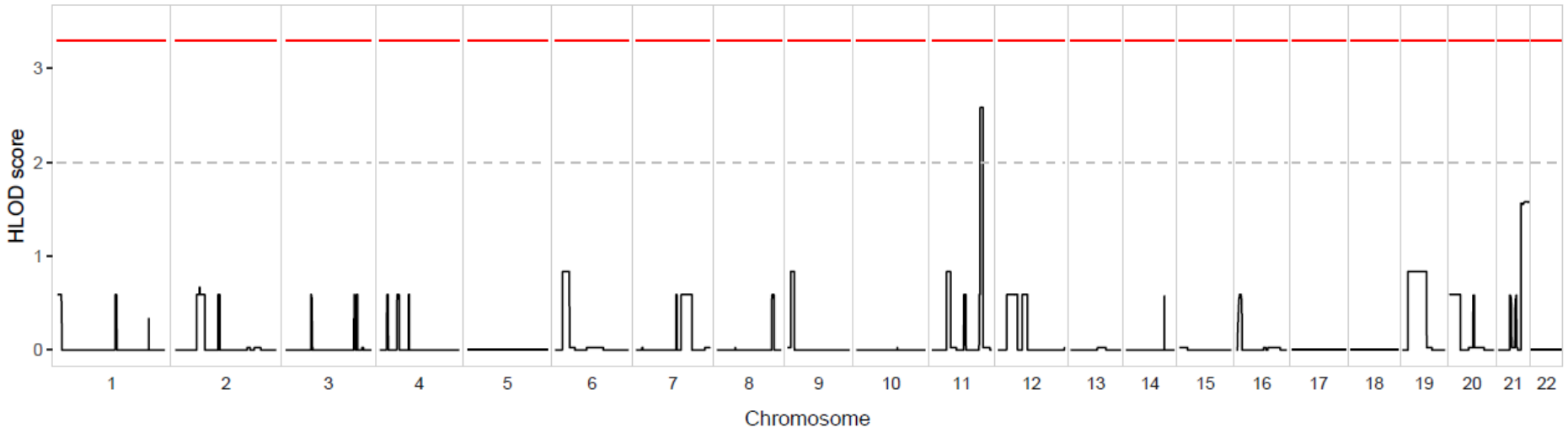


Figure 4.9 Autosome-wide parametric linkage analysis of affection status in family CSA91/110

The X-axis represents the position in the genome ordered by chromosome and the y-axis represents the heterogeneity logarithm of the odds (HLOD) score. Significant (HLOD = 3.3) and suggestive (HLOD = 2) linkage thresholds are indicated with red solid line and grey dashed line, respectively.

4.4.4.2 Cataract causing gene identification

There were no known cataract genes reported in Cat-Map[31] under the suggestive linkage peak on chr11 (Table 4.7). Filtering of the WGS data of this linkage region failed to detect any potential causative SNVs, and none of the 6 SVs in this linkage region was fully segregating with the disease.

Under the linkage peak on chr21 (Table 4.7), three known cataract genes *CRYAA*, *COL18A1* and *LSS* were identified. The previously reported cataract-causing mutation *CRYAA* c.62G>A (p.R21Q) was validated in the WGS data, while no rare, pathogenic, segregating variants were identified on *COL18A1* and *LSS*.

4.5 Discussion

4.5.1 Causative genes in these two extended paediatric cataract families

Previous investigation of family CRCH13 by direct sequencing of known cataract genes and loci detected that *GJA3* R76H was the cause of cataracts in this family[85]. However, the variant was also carried by an unaffected individual, CRCH13.46. This raised the question of whether other genetic factors in the family can better explain the observed cataract phenotype. We therefore conducted parametric linkage analysis of cataract affection status to more comprehensively identify the causative gene, with statistical evidence. Parametric linkage analysis verified the *GJA3* gene — it was located in the only significant linkage region autosome-wide and had the highest HLOD score (3.586). Interestingly, WGS analysis showed the newly recruited patients

CRCH13.73 and CRCH13.75 did not carry the *GJA3* R76H variant that was reported to be segregating with disease in the previous investigation.

Although CRCH13.73 and CRCH13.75 were diagnosed with cataract when they were recruited at the age of 63 and 41, respectively, their phenotypes were pulverulent cataracts and nuclear cataracts, which are the common types of paediatric cataract. Therefore, these two individuals were considered to have paediatric cataracts rather than age-related cataracts. CRCH13.73 and CRCH13.75 belong to a relatively independent small branch of family CRCH13 (Figure 4.3). Cataracts in this small family may be dominated by a gene other than *GJA3*, or maybe the result of a combination of multiple genetic factors. It is also possible that some environmental factors specific to this family lead to the formation of cataracts. Screening of known cataract genes in the WGS data of CRCH13.73 and CRCH13.75 is recommended as a future step to determine the etiology of cataract in these two family members.

However, although CRCH13.73 and CRCH13.75 didn't carry the *GJA3* R76H variant, the linkage analysis revealed that there was no evidence that an alternative gene or location was the primary cause of cataract in the extended family CRCH13, and the peak linkage region (HLOD=3.586) was located at *GJA3*, confirming that this gene is the most likely cause of cataract.

Similar to family CRCH13, the previously reported causative mutation in *CRYAA* in the CSA families also displayed reduced penetrance, as it was detected in an unaffected member CSA110.2. The *CRYAA* locus was identified by parametric linkage analysis located in a linkage region with an HLOD score of 1.581. The reported c.62G>A (p.R21Q) variant on the *CRYAA* gene was further confirmed in the WGS data. Although a suggestive linkage peak

(HLOD=2.585) was identified on chr11, no potential rare, pathogenic, segregating SNVs or SVs was detected. Taken together, we concluded the *CRYAA* gene is the cause of cataracts in the CSA families.

4.5.2 Potential modifier genes?

Great variability in disease severity was observed in both the CRCH13 family and the CSA91/110. Affected individuals displayed varying degrees of lens opacification, ranging from mild cataracts that were unrecognised until the families were examined, to severe cataracts that were diagnosed in the first year of age and required surgical removal during childhood. In particular, the age of diagnosis/surgery was significantly diverse within both families. In family CRCH13, the age range for diagnosis was 0-19 years old, and the age range for surgery was 10-67 years. Similarly, in family CSA91/CSA110, the age of diagnosis ranged from 18 months to 61 years, and the age of surgery ranged from 5 years to no surgery until 61 years (the time of recruitment). Additionally, it was noted that the severity of disease in both extended families increased with generations. The majority of individuals in the youngest generation were diagnosed with severe cataracts before 1 year old, and some required surgical intervention in childhood. Their parents were generally diagnosed between 1 to 20 years old. Moderate cataracts in this generation were surgically removed roughly in middle age. The disease severity in the grandparents' generation was mild. Most of the grandparents were not diagnosed until they were included in the study in old age. The age of surgery of the grandparents was mostly over 50, some of the patients hadn't even undergone surgery at the time of recruitment. The diverse severity of affected members carrying the same causative mutation within a family as well as the

clear trend of increasing severity over generations indicates the existence of disease severity modifier genes.

Furthermore, the results of parametric linkage analysis also provide supporting evidence for the hypothesis of modifier genes. The presence of suggestive linkage peaks in both CRCH13 (Table 4.5) and CSA91/110 (Table 4.7) indicates that potential modifier genes may play roles in the cataractogenesis in these two families.

Collectively, it is suggested that the development and severity of cataracts in the CRCH13 and CSA91/110 families may be modified by genes other than the disease-causing genes we have detected. We hypothesised that there are potential modifiers to be identified in these two extended families.

4.5.3 The strategy of identifying cataract-causing genes in large families by combining linkage analysis and WGS

Linkage analysis is a classic genetic method to map the chromosomal location of disease genes. It is particularly powerful for tracking causal loci for rare diseases such as paediatric cataract in extended families with multiple generations and enrichment of affected members. Several causative genes for paediatric cataract have been successfully identified using linkage analysis[213-216]. Compared with direct sequencing of candidate genes, linkage analysis can provide statistical evidence of the involvement of a genetic region in cataractogenesis. In addition, genome-wide linkage analysis can systematically and comprehensively identify the regions that link to the disease phenotype

across the whole genome. For example, since only candidate cataract genes were studied in family CRCH13 and the CSA families before, it is impossible to determine whether other genetic factors contribute to the formation of cataracts in these families. Genome-wide parametric linkage analysis of these two families not only verified the causative genes detected by direct sequencing of candidate genes, but also identified multiple suggestive linkage peaks indicating an investigation of potential modifier genes may be fruitful. Furthermore, thorough linkage analysis of the entire genome is also beneficial to the identification of any novel genes underlying the disease, which cannot be achieved by screening candidate genes known to be associated with cataracts.

Whole-genome sequencing technology is becoming widely used due to the significant decrease in cost. This allows the exploration of pathogenic variants in almost the complete genome (~98-99%) of an individual which whole-exome sequencing can not achieve[217]. According to reports, there remain 30%-40% of paediatric cataract cases that fail to identify the genetic causes using microarray, exome sequencing and panel-based methods[75,218,219]. WGS is helpful to fill the gap, as a proportion of such cases are likely caused by mutations in non-coding regions. In our study, we analysed WGS data of selected individuals to search for pathogenic variants. In addition to filtering the exonic variants, all non-coding variants, such as intronic variants, intergenic variants, variants in the 5' and 3' UTRs were also taken into account. Since no other appropriate candidates were identified in both coding and non-coding regions, we were able to conclude that the previously reported mutations are the causes of cataracts in CRCH13 and CSA91/110. WGS is also superior for SV detection compared to whole-exome sequencing (WES). SVs that are not in exons cannot be detected in WES data, and the detection of re-arrangements

is also difficult. In the CRCH13 and CSA91/110 families, the investigation of SVs using WGS data further ruled out the possibility of other causative variants, consolidating our conclusion that the previously reported variants were the causes of cataracts in these two families.

Although the combination of linkage analysis and WGS successfully identified the cataract-causing variants in both CRCH13 and CSA91/110, there were some limitations. Due to the bit size limitation of MERLIN[195] software for a single family, the large family CRCH13 was divided into 5 subfamilies in the linkage analysis. This might decrease the statistical power to detect linkage as the more affected members in a pedigree, the more informative the family is[220]. In the future, the combined use of multiple linkage analysis software is recommended, especially the ones suitable for analysing large pedigrees, such as SOLAR (Sequential Oligogenic Linkage Analysis Routines)[221] or MORGAN (Monte Carlo Genetic Analysis)[222], which will help to improve the accuracy of linkage detection.

One major limitation of WGS is the varied sequencing quality throughout the genome. In order to ensure the reliability of the sequencing data, we only filtered for "high-confidence" variants with read depth greater than 10 and genotype quality greater than 20. It is possible that variants with low sequencing quality were missed under the relatively strict threshold. However, there is no fixed standard for the quality control of whole-genome sequencing data, it needs to be adjusted according to the specific situation and requirements.

4.6 Outcomes and significance of this study

In this chapter, the clinical information available for two extended cataract families, family CRCH13 and CSA91/110, have been updated, and the previously reported putative causative genes of both families were verified by parametric linkage analysis coupled with WGS. Additionally, indications of potential disease-modifying genes also have been detected, which leads to the further investigation described in Chapter 5.

Chapter 5

Investigation of disease severity modifiers in two extended paediatric cataract families

5.1 Introduction

Family CRCH13 and CSA91/110 presented with autosomal dominant paediatric cataracts. The causative mutations have been identified for both families. Earlier work of our team identified the R76H mutation in the *GJA3* gene encoding connexin46 in family CRCH13[85], as well as the R21Q mutation in *CRYAA* in family CSA91/110[89,90]. In Chapter 4, the clinical phenotype information of each participant was updated and the reported causative genes for both families were confirmed. Notably, a high degree of variability of the age of diagnosis as well as the age of surgery was observed in affected individuals with the same causative mutation within each family, suggesting that the severity of the disease may be modified by additional genetic factors other than the causative genes. Furthermore, parametric linkage analysis also indicated the existence of potential modifiers through the presence of several suggestive linkage peaks (Section 4.4.3.1 and Section 4.4.4.1). In this chapter, we investigated potential disease severity modifier loci in these two extended paediatric cataract families.

5.1.1 Genetic modifiers of cataractogenesis

Genetic modifiers refer to DNA sequence variations that alter the phenotypic outcomes of primary causative genes but do not necessarily cause disease by themselves[217]. By directly interacting with the product of the pathogenic gene, or acting on the same or complementary biological pathway of the primary gene, genetic modifiers can subtly or profoundly modulate any aspect of disease expression, for example, age of onset[93,98,100], severity[96,101], and penetrance[91,92,102]. Genetic modifiers can be rare or common variants in protein-coding regions or regulatory regions, such as untranslated regions (UTRs) or promoters[104].

Several modifier loci and genes of cataractogenesis have been identified in mouse and rat models. A modifier locus that suppressed the cataract phenotype induced by *Gja8* mutation was located on chromosome 5 of the BN/Sea rat[223]. Narita *et al.* mapped two modifier loci on chromosome 3 and chromosome 10 that were associated with the distinct phenotypes observed in the Nakano cataract (NCT) mouse model[224]. The *vacuolated lens (vl)* mutation, which results in a frameshift deletion in the orphan G protein-coupled receptor 161 gene (*Gpr161*), causes full penetrance of paediatric cataract in C3H/HeSnj mouse model[108]. Millonig's team reported that the cataract incidence was decreased by 85% when crossing the $Gpr161^{vl/vl}$ C3H and MOLF/Ei mice, suggesting that cataracts resulting from *vl* mutation were rescued by MOLF modifiers[108]. The team mapped 4 modifier loci from MOLF background—*Modvl1*, *2*, *3*, and *4*, and they further determined that the Forkhead box E3 (*Foxe3*) gene is a gene responsible for the phenotypic modification of the *vl* cataract[108,225]. Similarly, by mating SJL/J-$Foxe3^{rct/rct}$ and MSM/Ms mice, a

strain-specific variant in the beta polypeptide (*Pde6b*) encoding gene was identified to modulate the onset of cataracts in *Foxe3*^{rct} mice[100,107].

Phenotypic modification has been observed in human paediatric cataract families as well. It has been frequently reported that affected family members with identical pathogenic mutation displayed a spectrum of disease severity, from undetectable cataracts in late adulthood to severe cataracts that required surgical removal in infancy[85,86,89,90]. The significant intrafamilial phenotypic heterogeneity indicates additional genetic factors modifying the severity of paediatric cataract. However, despite decades of research efforts, genetic modifiers of paediatric cataract are still largely undiscovered due to the complexity of modifier effects including the heterogeneity of the genetic background, the interference of environmental factors, the possibility of involvement of multiple modifier loci, and the challenge of physical mapping of the modifier genes.

5.2 Hypothesis and aim

Hypothesis:

As yet unknown genes modify the cataract severity in family CRCH13 and CSA91/110.

Aim:

Identify modifier loci linked to disease severity in family CRCH13 and CSA91/110 using variance components linkage analysis.

5.3 Methods

5.3.1 Age of diagnosis/surgery

The age of diagnosis is defined as the time when a definite cataract phenotype was first observed in the lens of a patient recorded by the ophthalmologists. The age of surgery is the time when the affected participant first underwent a cataract removal operation. For patients with bilateral cataracts, if the diagnosis/surgery age of the two eyes is different, the earlier one was utilized.

5.3.2 Disease severity assessment and classification

The disease severity of affected individuals was evaluated and classified based on the age of diagnosis as well as the age of surgery. Due to the mild cataract phenotype, some affected individuals were undiagnosed until they were examined for the study, therefore the accurate age of diagnosis (onset) cannot be obtained. Some patients had not had surgery as mild cataracts did not affect vision, hence there was no data on the age of surgery for these patients. Due to the missing data issue, age of diagnosis or age of surgery cannot be directly utilised as quantitative traits in the variance components linkage analysis. To overcome this shortage, we combined the two to assess the severity of the disease, then the disease severity was classified into ordinal grades and applied as a categorical trait in the linkage analysis (see Section 5.3.3). The disease severity category thresholds were specifically set according to the distribution of the age of diagnosis/surgery in each family.

5.3.3 Variance components linkage analysis

Genome-wide multipoint variance components linkage analysis of disease severity was performed in the MERLIN v1.1.2[195] software using SNP array genotyping data (Section 4.4.2) and the equivalent SNPs extracted from whole-genome sequencing data of newly recruited individuals CRCH13.73 and CRCH13.75. The quality of the input genotyping data was verified using the PEDSTATS program and problematic SNPs were removed with the PEDWIPE program in MERLIN. The logarithm of the odds (LOD) score of 3.3 and 2 were respectively employed as thresholds for genome-wide significant and suggestive linkages[196]. The boundary on both sides of the linkage region was defined at the site with a LOD score equal to the maximum LOD score minus one (LOD-1 region). The covariate employed in the linkage analysis was set in the Mega2 (Version 6.0.0)[193,194] software. An alternative phenotype file (*.phe) containing the data of covariate was made following the tutorial on the Mega2 website[194]. The genotype of the SNP array marker rs57547952 was set as the covariate in the variance component linkage analysis of family CRCH13 (described in Section 5.4.2). Since Mega2 enforces the covariate to be quantitative, the genotyping data was transformed into quantitative data with wild-type set to 0, mutant set to 10, and unknown set to -9. Due to the bit-size limitation (24 bits) of MERLIN on the individual family, the extended family CRCH13 was divided into 5 subfamilies, as described in Section 4.4.3.1. All subfamilies were analysed together in the variance components linkage analysis and the overall LOD score was calculated.

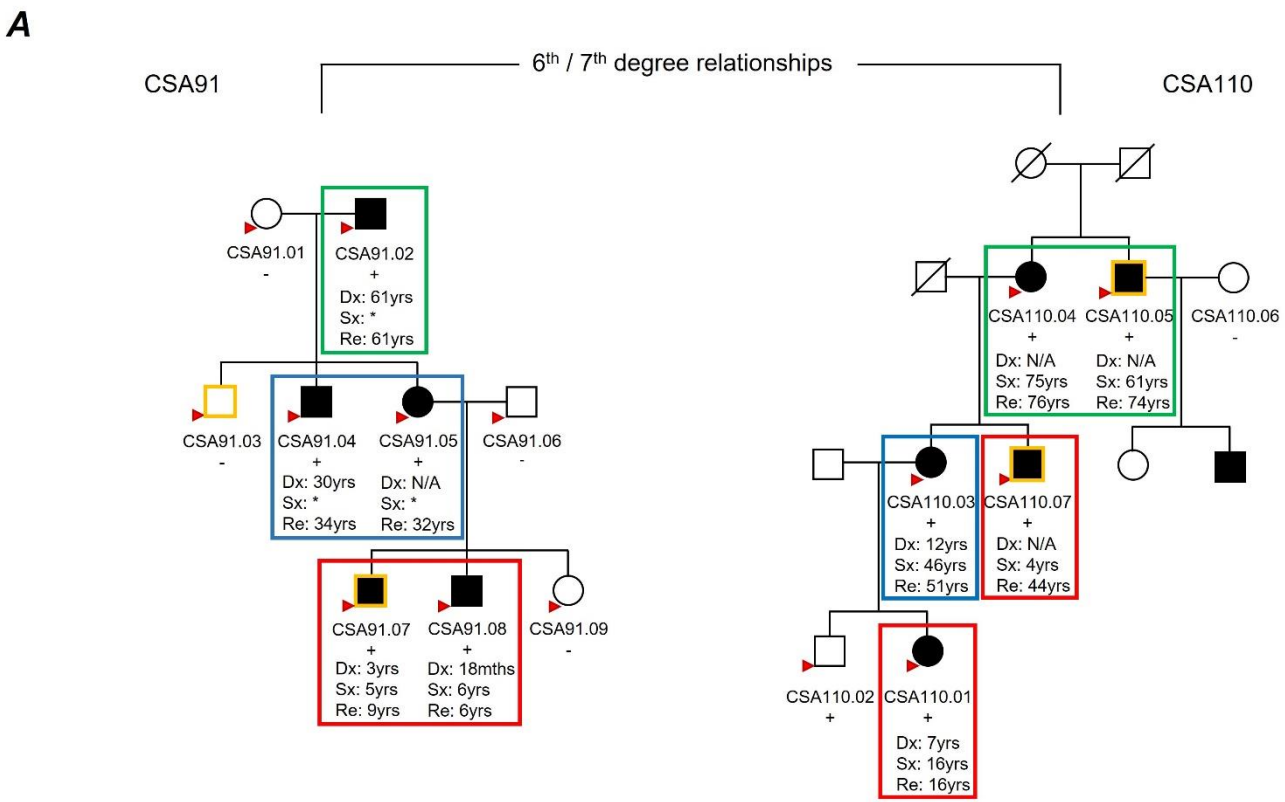
5.4 Results

5.4.1 Family CSA91/110

According to the spectrum of the age of diagnosis/surgery, we first classified family members in CSA91/110 into 4 categories: unaffected, mild, moderate, and severe. Patients who were diagnosed or had cataract-removal surgery during childhood (age of diagnosis: ≤10 years and/or age of surgery: ≤20 years) were considered to have severe cataracts, those were diagnosed or had surgery at middle age (age of diagnosis: 10-50 years and/or age of surgery: 20-50 years) were binned into the moderate category, and the ones were diagnosed or had surgery at old age (age of diagnosis/surgery: ≥50 years) fell into the mild category. Although the exact diagnosis age of CSA91.05 was not available, she already had cataracts when she was recruited at the age of 32, so she was classified into the moderate group. The severity classification of members in family CSA91/110 is shown in Table 5.1 and Figure 5.1A. The phenotype distribution is shown in Figure 5.1B.

Table 5.1 Disease severity assigned to members involved in the variance components linkage analysis of family CSA91/110

Unaffected	CSA91.01, CSA91.03, CSA91.06, CSA91.09, CSA110.02
Mild	CSA91.02, CSA110.04, CSA110.05
Moderate	CSA91.04, CSA91.05, CSA110.03
Severe	CSA91.07, CSA91.08, CSA110.01, CSA110.07


A
CSA91
6th / 7th degree relationships
CSA110
CSA91.01
-
CSA91.02
+
Dx: 61yrs
Sx: *
Re: 61yrs
CSA91.03
-
CSA91.04
+
Dx: 30yrs
Sx: *
Re: 34yrs
CSA91.05
+
Dx: N/A
Sx: *
Re: 32yrs
CSA91.06
-
CSA91.07
+
Dx: 3yrs
Sx: 5yrs
Re: 9yrs
CSA91.08
+
Dx: 18mths
Sx: 6yrs
Re: 6yrs
CSA91.09
-
CSA110.04
+
Dx: N/A
Sx: 75yrs
Re: 76yrs
CSA110.05
+
Dx: N/A
Sx: 61yrs
Re: 74yrs
CSA110.06
-
CSA110.03
+
Dx: 12yrs
Sx: 46yrs
Re: 51yrs
CSA110.07
+
Dx: N/A
Sx: 4yrs
Re: 44yrs
CSA110.02
+
CSA110.01
+
Dx: 7yrs
Sx: 16yrs
Re: 16yrs

B

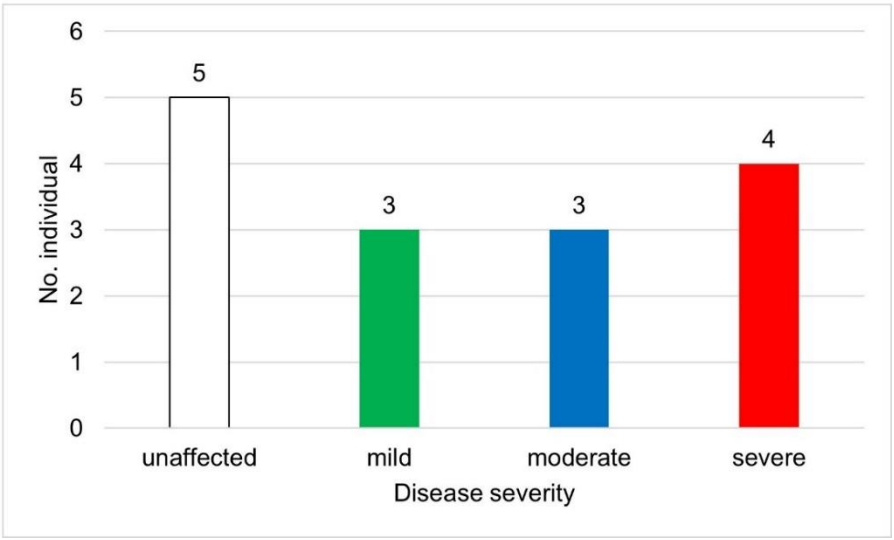


Figure 5.1 Classification and distribution of disease severity in family CSA91/110

A. Disease severity classification of family CSA91/110. The disease severity categories are indicated with coloured boxes. Green boxes: mild, blue boxes: moderate, red boxes: severe. Males are indicated by squares and females by circles. Solid black symbols indicate affected individuals while open symbols indicate unaffected individuals. Age of diagnosis (Dx)/surgery (Sx) is listed under each affected individual. *: no surgery at recruitment; N/A: data unavailable; Re: recruitment; yrs: years; mths: months. "+" indicates the *CRYAA* R21Q mutation carriers, "-" indicates non-carriers. Symbols with red triangles represent individuals with genotyping data, and symbols with yellow outlines represent individuals with whole-genome sequencing (WGS) data. **B.** Disease severity distribution in family CSA91/110.

Family members in the unaffected, mild, moderate, and severe groups were assigned values of 0, 1, 2, and 3 in the variance components linkage analysis. However, due to low sample numbers and the non-normal trait distribution, there was no detectable heritability (h^2=0.00) using MERLIN[195]. The phenotype was also recoded to 0, 10, 20, 30 and 0, 100, 200, 300, to see if the larger standard deviation impacted the estimate, but no heritability was detectable for this trait. We then collapsed the mild and moderate groups to further normalise the trait distribution, but there was still no detectable heritability due to the limited sample size.

5.4.2 Family CRCH13

Family CRCH13 was initially classified into 5 groups based on the distribution of the age of diagnosis/surgery. Patients who were diagnosed within the first year (age of diagnosis: ≤1 year) were considered to have severe cataracts. According to the age of surgery, patients with a diagnosis age >1 year were classified as moderate (age of surgery: ≤50 years) or mild (age of surgery: >50 years). In order to distinguish patients with different severity as much as possible, individuals with moderate cataracts were further divided into two subgroups: less severe (age of surgery: ≤30 years) and more severe (age of surgery: 30-50 years). In addition, affected individuals who were not diagnosed until recruitment and hadn't had surgery were classified into the mild category because the cataract phenotypes were too mild to be aware of or to affect vision. All together, family members of CRCH13 were classified into unaffected, mild, moderate (less severe), moderate (more severe), and severe categories. The heritability of the trait was 24.32%, however, no evidence for linkage analysis was observed across the genome (p-value=0.5). This is likely due to the trait distribution (Figure 5.2) deviating too much from normal to yield a reliable result.

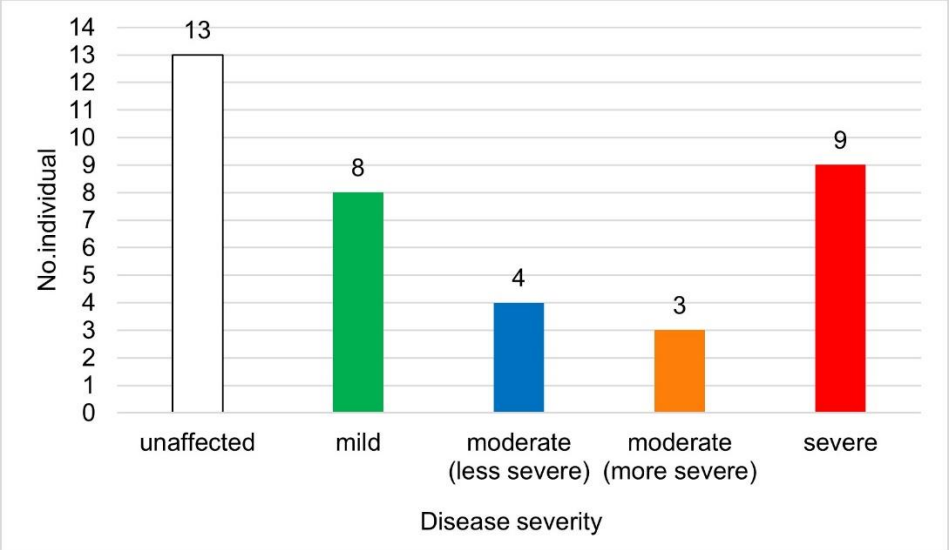


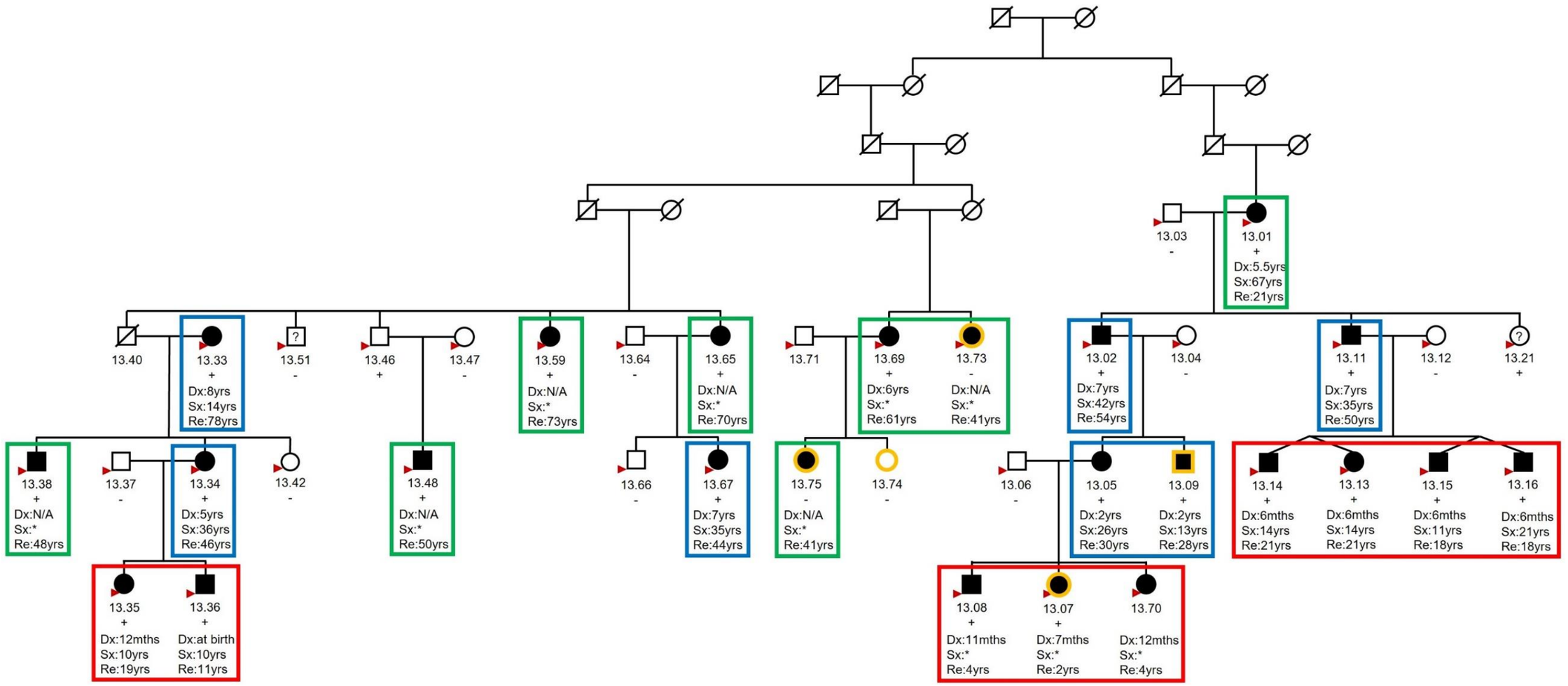
Figure 5.2 Disease severity distribution in family CRCH13 (five categories)

We therefore collapsed the two subgroups of the moderate category, resulting in four roughly balanced categories in the CRCH13 family: unaffected (n=13), mild (n=8), moderate (n=7), and severe (n=9) (Table 5.2 and Figure 5.3).

Table 5.2 Disease severity assigned to members involved in the variance components linkage analysis of family CRCH13 (four categories)

Unaffected	CRCH13.03, CRCH13.04, CRCH13.06, CRCH13.12, CRCH13.37, CRCH13.40, CRCH13.42, CRCH13.46, CRCH13.47, CRCH13.64, CRCH13.66, CRCH13.71, CRCH13.74
Mild	CRCH13.01, CRCH13.38, CRCH13.48, CRCH13.59, CRCH13.65, CRCH13.69, CRCH13.73, CRCH13.75
Moderate	CRCH13.02, CRCH13.05, CRCH13.09, CRCH13.11, CRCH13.33, CRCH13.34, CRCH13.67
Severe	CRCH13.07, CRCH13.08, CRCH13.13, CRCH13.14, CRCH13.15, CRCH13.16, CRCH13.35, CRCH13.36, CRCH13.70
Unknown	CRCH13.21, CRCH13.51

A


13.40
13.33
+
Dx:8yrs
Sx:14yrs
Re:78yrs
13.51
-
13.46
+
13.47
-
13.59
+
Dx:N/A
Sx:*
Re:73yrs
13.64
-
13.65
+
Dx:N/A
Sx:*
Re:70yrs
13.71
13.69
+
Dx:6yrs
Sx:*
Re:61yrs
13.73
-
Dx:N/A
Sx:*
Re:41yrs
13.03
-
13.01
+
Dx:5.5yrs
Sx:67yrs
Re:21yrs
13.02
+
Dx:7yrs
Sx:42yrs
Re:54yrs
13.04
-
13.11
+
Dx:7yrs
Sx:35yrs
Re:50yrs
13.12
-
13.21
+
13.38
+
Dx:N/A
Sx:*
Re:48yrs
13.37
-
13.34
+
Dx:5yrs
Sx:36yrs
Re:46yrs
13.42
-
13.48
+
Dx:N/A
Sx:*
Re:50yrs
13.66
-
13.67
+
Dx:7yrs
Sx:35yrs
Re:44yrs
13.75
-
Dx:N/A
Sx:*
Re:41yrs
13.74
-
13.06
-
13.05
+
Dx:2yrs
Sx:26yrs
Re:30yrs
13.09
+
Dx:2yrs
Sx:13yrs
Re:28yrs
13.14
+
Dx:6mths
Sx:14yrs
Re:21yrs
13.13
+
Dx:6mths
Sx:14yrs
Re:21yrs
13.15
+
Dx:6mths
Sx:11yrs
Re:18yrs
13.16
+
Dx:6mths
Sx:21yrs
Re:18yrs
13.35
+
Dx:12mths
Sx:10yrs
Re:19yrs
13.36
+
Dx:at birth
Sx:10yrs
Re:11yrs
13.08
+
Dx:11mths
Sx:*
Re:4yrs
13.07
+
Dx:7mths
Sx:*
Re:2yrs
13.70
Dx:12mths
Sx:*
Re:4yrs

B

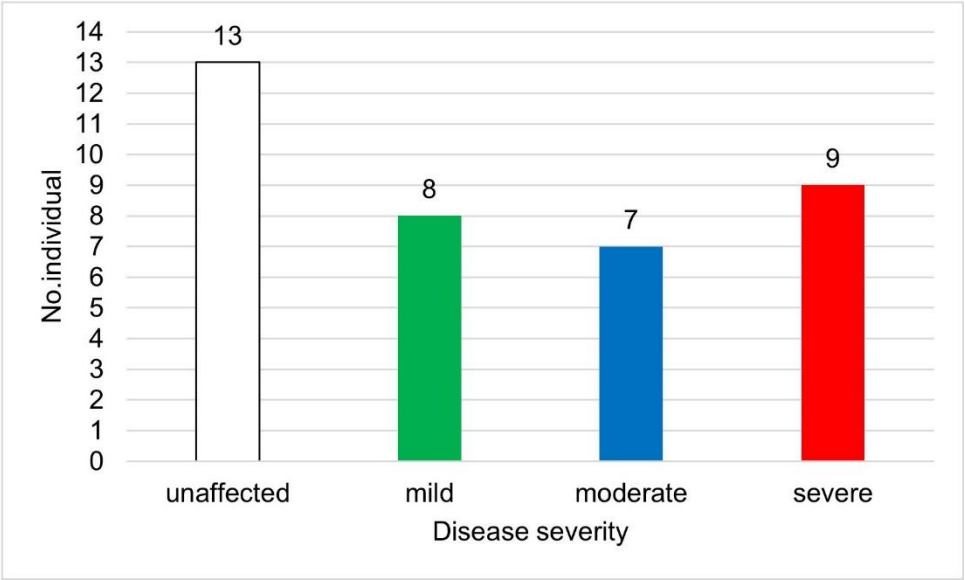


Figure 5.3 Classification and distribution of disease severity in family CRCH13 (four categories)

A. Disease severity classification of family CRCH13. The disease severity categories are indicated with coloured boxes. Green boxes: mild, blue boxes: moderate, red boxes: severe. Males are indicated by squares and females by circles. Solid black symbols indicate affected individuals while open symbols indicate unaffected individuals. Age of diagnosis (Dx)/surgery (Sx) is listed under each affected individual. *: no surgery at recruitment; N/A: data unavailable; Re: recruitment; yrs: years; mths: months. "+" indicates the *GJA3* R76H mutation carriers, "-" indicates non-carriers. Symbols with red triangles represent individuals with genotyping data, and symbols with yellow outlines represent individuals with whole-genome sequencing (WGS) data. **B.** Disease severity distribution in family CRCH13.

Under the four-category system, the estimated heritability was 19.45%. Multipoint variance components linkage analysis was performed with individuals in the unaffected, mild, moderate, and severe groups coded to 0, 1, 2, and 3. Three linkage regions were identified across the genome (Figure 5.4): one significant linkage region on chromosome (chr) 21 (LOD = 4.19, p=0.00001) and two suggestive linkage regions on chr1 (LOD = 2.53, p=0.0003) and chr13 (LOD = 2.07, p=0.001), respectively (Table 5.3).

Table 5.3 Linkage regions identified in variance components linkage analysis of disease severity in family CRCH13 (four categories)

Chr	Position (hg19)	Size (bp)	Maximum LOD	Linkage
21	chr21:19881273-22381705	2500433	4.19	Significant
1	chr1:110584887-111748425	1163539	2.53	Suggestive
13	chr13:19121858-21518461	2396604	2.07	Suggestive

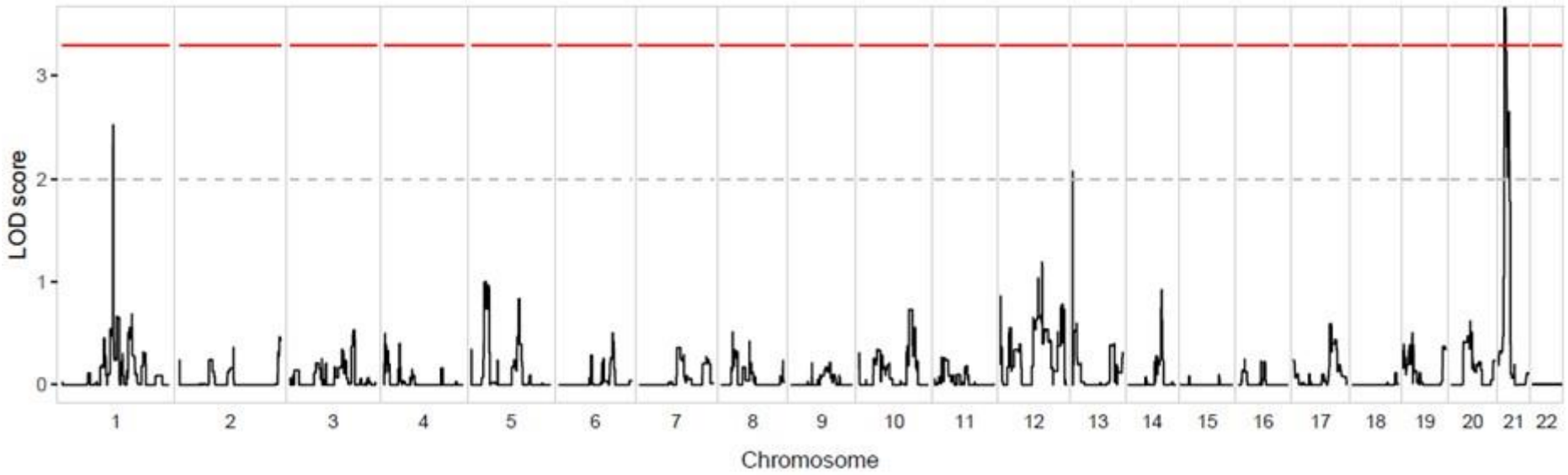


Figure 5.4 Genome-wide variance components linkage analysis of cataract severity of family CRCH13 (four categories)

The X-axis represents the position in the genome ordered by chromosome and the y-axis represents the logarithm of the odds (LOD) score. Significant (LOD = 3.3) and suggestive (LOD = 2) linkage thresholds are indicated with red solid line and grey dashed line, respectively.

It is worth noting that the linkage region detected on chr13 (hg19: chr13:19121858-21518461) was where the causative mutation *GJA3* R76H (chr13: 20717201) was located. In order to eliminate the possible interference of the causative mutation on the identification of modifier gene, the SNP array marker (rs57547952) of the mutation site was set as a covariate in the variance components linkage analysis. However, the discrete covariate reduced the overall variance to the point where the heritability dropped to 1.46%, which was too low to obtain reliable linkage results.

We also conducted the investigation by further collapsing the mild and moderate groups, to normalise the trait distribution more. However, the heritability was remarkably reduced to 2.47%.

5.5 Discussion

5.5.1 Mapping modifier loci linked to disease severity in extended pedigrees using variance components linkage analysis

Depending on the type of data available, two major strategies are applied in human modifier studies: association studies and linkage studies. Association studies involve family-based association studies and genome-wide association studies (GWASs). They both require large data sets, as family-based association studies utilise a number of consanguineous families with the same genetic variant but present variable phenotypes, while GWAS is based on population data. In contrast, modifier studies that utilise linkage-based approaches are typically performed on pedigrees. Initially, linkage analysis for

the quantitative trait that was used to map modifier loci was only applicable to pairs of relatives such as sibpairs[226,227]. While with continuous development in linkage methods, variance components linkage analyses were also successfully performed in large extended pedigrees[221,228,229], which provides greater linkage information than sibpairs, as all possible biological relationships are used at the same time. Variance components linkage analysis is designed to search for quantitative traits in pedigrees[221,230,231], it is powerful and flexible as it is penetrance model-free and can be applied to pedigrees of arbitrary size and complexity[232]. Since our study cohorts are extended pedigrees, and the clinical phenotype (age of diagnosis/surgery) utilised to locate modifiers is quantitative, variance components linkage analysis is the most suitable method for our research.

Properly defining the trait for variance components linkage analysis is the key to successfully identifying modifiers. The clinical phenotypic information that can accurately reflect the severity of the disease is the basis of this study. In our study, the age of diagnosis/surgery was used as the measure of disease severity, however, the progress of cataract treatment and management over time would affect the timing of surgery for patients of different generations, causing inevitable bias to the investigation. Additionally, as subtle opacities were hard to be aware of, many patients with mild cataracts were undiagnosed until they were examined for this study, hence the exact age of onset is unknown. Similarly, some affected individuals hadn't had surgery as the mild phenotype did not affect the vision, so there was no data on the age of surgery for these people. Therefore, neither the age of diagnosis nor the age of surgery could be employed as quantitative traits in the linkage analysis, because incomplete data could not provide sufficient statistical power. In order to overcome this limitation, we combined the two in the assessment of disease

severity to complement each other. However, in this way, we were forced to use categorical data that classified family members by combining the age of diagnosis and the age of surgery, rather than using the original quantitative data as required by variance components linkage analysis. An appropriate description of the phenotypic variability in the studied families is essential for the detection of modifiers. Disease severity classification strategies that are too detailed or too rough are not conducive to finding modifiers. In family CRCH13, the classification of participants into 5 categories captured the spectrum of severity, but the number of individuals in each category was quite diverse and the non-normal trait distribution made the linkage results unreliable (p=0.5). Although dividing family members into 3 categories made the trait more normally distributed, the classification was too rough to adequately describe the differences among patients with different severity, so there was not enough heritability to detect the variance of the trait (h^2=2.47%). Only by classifying the family members into 4 categories with roughly equal numbers, the variance components linkage of disease severity successfully mapped three chromosomal regions for candidate modifier loci, even though the accuracy was still limited due to the deviation from the normal distribution.

The strategy of mapping modifier loci using variance components linkage analysis is more valuable in large pedigrees with multiple generations and dozens to hundreds of affected individuals. Generally, the effects of modifiers are not as strong as that of disease-causing genes, hence a large pedigree size is required to provide sufficient statistical power to detect even subtle genetic linkage. While modifier loci were successfully located in the extended family CRCH13, we were unable to identify potential modifiers in the smaller family CSA91/110 using variance components linkage analysis. The limited pedigree size and further classification of participants led to insufficient statistical

power, which was the decisive reason for the failure of the modifier detection in CSA91/110.

The complexity upper bound of the MERLIN[195] software (24 bits, described in Section 4.4.3.1) is also a limitation of this study. We had to split the original pedigree of CRCH13 into 5 subfamilies to fit into MERLIN. In theory, SOLAR (Sequential Oligogenic Linkage Analysis Routines)[221] software would be better as it can handle large pedigrees so that family CRCH13 could be analysed as a whole, but it does not apply to the categorical trait that we generated to solve the problem of missing data.

5.5.2 Detection of disease severity modifiers in the two paediatric cataract families

5.5.2.1 Family CSA91/110

In Chapter 4, the suggestive linkage peak identified in the parametric linkage analysis of CSA91/110 suggested the presence of potential modifiers. However, variance components linkage analysis was unable to map the loci that modify the disease severity in this family as no heritability for the phenotype was detectable. This is most likely due to the finite family size. Although we tried different disease severity classification strategies, limited family members could not provide sufficient phenotypic variance, and thereby sufficient statistical power, to detect the linkage. We determined that variance components linkage analysis may not be a suitable method to identify modifiers in CSA91/110, and alternative approaches should be explored.

5.5.2.2 Family CRCH13

By classifying the severity into 4 categories, linkage regions on chr1, chr13, and chr21 were successfully mapped in family CRCH13. Noteworthily, the primary causative mutation *GJA3* R76H was under the linkage peak on chr13, indicating this linkage signal is likely driven by the causative mutation. We tried to adjust the effect of the causative mutation by conducting linkage analysis with the SNP array marker of the mutation site as the covariate, but the trait heritability dropped from 19.45% to 1.64%. It is inferred that the affection status accounted for most of the variance among the family members, therefore the discrete covariate representing the disease-causing mutation significantly reduced the overall trait variance to an undetectable level. The remaining heritability was insufficient to identify the relatively subtle modification effect that explained the severity variability in affected individuals.

We were unable to identify candidate disease-modifying variants within the linkage regions due to the lack of sufficient WGS data. As described in Section 4.3.7, We did WGS on CRCH13.07, CRCH13.09, CRCH13.73, CRCH13.74, and CRCH13.75 to confirm the primary causative gene and rule out other highly penetrant variants in the suggestive linkage peaks identified in Chapter 4. These individuals were selected as they belonged to different branches of the pedigree, their phenotypes were distributed across the entire severity spectrum from unaffected to severely affected (Figure 5.3A). However, the WGS data of CRCH13.73 and CRCH13.75 were not appropriate for modifier variants screening as we later determined these two affected members did not carry the *GJA3* R76H causative mutation (described in Section 4.4.3.2). CRCH13.73 and CRCH13.75 were newly recruited members and were not

included in the previously published report of this family. Due to the timing of their recruitment, SNP array data were not available and they were included in the linkage analysis using SNPs extracted from the WGS data. Since they both had paediatric cataract and CRCH13.69 (mother of CRCH13.75, sibling of CRCH13.73) did carry the rare *GJA3* R76H causative mutation, it was not until the WGS data were obtained that the two newly enrolled patients were found not to carry the causative mutation shared by all other affected individuals in this extended family. After eliminating these two individuals, the WGS data of the remaining members were insufficient for the modifier variants filtering. This is one of the limitations of this study. In the future, WGS should be performed on as many family members as possible in order to identify the variants possibly acting as modifiers.

The nuclear family of CRCH13.73 and CRCH13.75, as part of the CRCH13 family, displayed similar disease severity variability with the rest of the extended family. It is possible that the paediatric cataract of CRCH13.73 and CRCH13.75 may be caused by genetic factors other than the *GJA3* mutation or even environmental factors, but the severity of cataracts in the entire family was modulated by shared modifiers. In this case, it is suggested to whole-genome sequence CRCH13.33, CRCH13.34, CRCH13.38 and CRCH13.36, as they locate in a different branch of the pedigree other than CRCH13.07 and CRCH13.09 (WGS data available for these two individuals), and displayed variable disease severity (Figure 5.3A). If the modifiers contribute to the disease, WGS data should be filtered for variants with 0, 1, and 2 copies of the alternate allele in individuals with mild cataract (CRCH13.38), moderate cataract (CRCH13.33, CRCH13.34, CRCH13.09), and severe cataract (CRCH13.36, CRCH13.07), respectively. Conversely, if the modifiers have a protective effect against the disease, variants with 2, 1, and

0 copies of the alternate alleles in patients with mild, moderate, and severe cataracts, respectively, should be considered. Additionally, although paediatric cataract is a rare disease, its modifiers are not necessarily rare. Since modifiers modulate rather than determine disease phenotypes, they can be either rare or common variants. Therefore, there is no need to filter on minor allele frequency (MAF) as in screening for causative variants.

It is also possible that CRCH13.73 and CRCH13.75 independently carried a causative mutation that can lead to mild/late-onset cataract, which can modulate the severity of the remaining members of the CRCH13 family that shared the *GJA3* mutation. In the strict sense, the mutation carried by CRCH13.73 and CRCH13.75 is no longer a genetic modifier, as it can independently lead to cataract formation. This condition is closer to a digenic disorder. When the mutant genotypes at two loci better explain the phenotype of some patients and their mildly affected relatives than the genotype at a single locus, inheritance is digenic[233]. Digenic diseases include situations with two loci of roughly equal importance that jointly determine the onset of disease, as well as situations where one of the loci has a primary effect but presents with variable age of onset or disease severity[234]. As yet there is no report of digenic cataract families, but digenic inheritance has been identified in several other diseases, such as keratoconus[235], retinitis pigmentosa (RP)[236], deafness[237], and Hirschprung's disease[233]. The next work is to screen known cataract-causing genes in the WGS data of CRCH13.73 and CRCH13.75, to determine whether they carry another disease-causing mutation other than the *GJA3* R76H.

Certainly, it cannot be ruled out that the variable disease severity observed in family CRCH13 is a comprehensive result of multiple genetic and environmental factors, however, this situation is beyond the scope of this study.

5.6 Outcomes and significance of this study

In this chapter, linkage regions likely containing variants associated with cataract severity were identified in one of the extended paediatric cataract families. To the best of our knowledge, this is the first study to investigate genetic modifiers of paediatric cataract. Our strategy can serve as a starting point to study genetic modifiers in extended pedigrees with paediatric cataract or other rare monogenic diseases.

Chapter 6

General discussion

6.1 Summary of the project

So far, more than 200 genes and loci have been identified to be related to inherited paediatric cataract[31]. However, these genes only account for around 60% of the patients with hereditary paediatric cataracts[75-77]. The molecular aetiology in the remaining affected individuals remains to be determined. Furthermore, diverse phenotypes have been observed in patients with the same causative mutations within paediatric cataract families. Multiple studies support the idea that genetic modifiers contribute to the intrafamilial phenotypic heterogeneity[85,86,89,90], but no studies to date have discovered such modifiers in paediatric cataract. The overall aim of this project was to identify novel causative genes and modifiers in paediatric cataract families. Three key findings were generated in this thesis: 1) knocking out of the novel candidate gene *HTR1F* in F0 zebrafish did not provide evidence for it being a cataract causing gene; 2) the primary causative gene for two extended families with variable phenotype was confirmed; 3) three putative linkage regions for modifier loci were identified in an extended paediatric cataract family with a range of severity.

In the first study, we knocked out a novel identified putative causative gene *HTR1F* in F0 zebrafish models using clustered regularly interspaced short

palindromic repeat (CRISPR)-Cas9 (CRISPR-associated system 9) ribonucleoprotein (RNP) to explore its role in cataractogenesis, as described in Chapter 3. We found the cataract rate in the *htr1f* knockout fish were slightly higher than that in the controls, but the difference did not reach statistical significance. In the second study, we first updated the clinical information of two extended paediatric cataract families and validated the primary causative genes previously reported in both families by performing parametric linkage analysis and whole-genome sequencing (WGS), as described in Chapter 4. We then conducted variance components linkage analysis of disease severity to investigate the potential severity modifiers, as described in Chapter 5. We detected one significant linkage region on chr21 (LOD = 4.19, p=0.00001) as well as two suggestive linkage regions on chr1 (LOD = 2.53, p=0.0003) and chr13 (LOD = 2.07, p=0.001) in one of the families, CRCH13, but failed to detect any linkage regions in the other family, CSA91/110, likely due to limited statistical power.

6.2 Knocking out the *HTR1F* gene in F0 zebrafish did not provide evidence for it being a cataract causative gene

With the application of high-throughput sequencing technologies such as whole-exome sequencing (WES) and WGS in the screening of paediatric cataract, a significant number of new genes have been identified in the past decade[31]. However, these genes that are discovered through genetic data need supplementary evidence to confirm their pathogenicity[238]. While the gold standard is to find multiple families carrying mutations in the same gene, it

is difficult to achieve for rare diseases such as paediatric cataract. Functional analyses therefore become widely applied to provide biological evidence for the novel genes being causative.

We evaluated the role of a novel gene, *HTR1F*, in the development of cataracts. Our team previously identified this gene by WES of a paediatric cataract family, CSA92[109]. In this thesis, our study reveals that knocking out of the *HTR1F* gene in F0 zebrafish did not provide evidence for it being a cataract causative gene.

One possible explanation is that the pathogenicity of the *HTR1F* variant is relatively weak, which results in mild phenotypes that require a larger sample size for better power to detect the pathogenic effect. The p.Arg289 residue is located in one of the regions of HTR1F most tolerant to missense variation, therefore the p.R289W *HTR1F* variant may not have a strong impact on the function of the resultant protein. Additionally, as described in Chapter 3, affected individuals in family CSA92 displayed a spectrum of severity. It is worth noting that except for the primary mutation in *HTR1F*, non-segregating variants in two well-characterized cataract genes, *HSF4* and *BFSP1* were also identified in this family. Interestingly, the most severe patients in this family harboured all three variants, while patients with two variants were relatively milder. It is likely that the *HSF4* and *BFSP1* variants modified the disease severity in family CSA92. The mild cataract phenotypes induced by the primary *HTR1F* mutation may become significant with the contribution of the two modifier variants in *HSF4* and *BFSP1*. However, it is difficult to determine whether the *HSF4* and *BFSP1* variants are modifiers or just benign variants in family CSA92 when the primary gene is not confirmed. Besides, the small family size of CSA92 may limit the statistical power to detect the modifiers as well.

By targeted disruption of the *HTR1F* gene in the zebrafish model, we assessed the biological effects of HTR1F protein dysfunction. However, it is possible that the mutation identified in CSA92 induced disease by creating a new protein function, or the variant may be completely unrelated to the disease. The former, although not frequent, has been reported. A misfolding gain-of-function (GOF) pathogenic mechanism is believed to be involved in myocilin-associated glaucoma[239]. Signal transducer and activator of transcription 1 (*STAT1*) GOF mutations have been identified as the most common cause of inherited syndromic chronic mucocutaneous candidiasis (CMC)[240]. A recent study identified a GOF mutation in the *CLCN2* chloride channel gene leading to primary aldosteronism[241]. Further investigation by introducing the exact mutation identified in CSA92 to the zebrafish model using CRISPR-Cas9-based knockin technique[242] will help to test whether the *HTR1F* mutation is a GOF.

From our study, we summarized some factors that may affect the results of cataract gene evaluation studies using the zebrafish model. The cataract rate of the uninjected wild-type fish utilized as negative controls may confound the results. It was observed in our laboratory that the baseline cataract rate of the AB strain wild-type zebrafish used in this study is higher than that of the TU or TL zebrafish, which are two other broadly used wild-type strains. In fact, the cataract rate of the AB wild-type fish (13.8%) in our laboratory is also higher than that of the fish of the same strain in other studies (Vorontsova *et al.*'s study[143]: ~1%; Goishi *et al.*'s study[243]: 0%; Mishra *et al.*'s study[131]: ~10%), this may be due to high levels of inbreeding or a specific variant that has arisen in our facility. The high cataract rate of the AB wild-type control fish used in this study may obscure the differences between the cataract rates of the *htr1f* knockout groups and the control groups. While some similar cataract gene

evaluation studies using zebrafish models only applied dummy-injected fish as controls[78,244,245], our results highlight the necessity of including uninjected wild-type fish as negative controls and we suggest selecting a wild-type fish strain with a low baseline cataract rate. The observation period also can affect the detection of the disease phenotype. The action duration of 1-phenyl-2-thiourea (PTU) (to suppress pigment formation) limited our observation period to 4 day-post-fertilization (dpf). Potential cataracts that occur later than this period will be missed. In a study that verified the function of DNase1l1l in cataract formation in zebrafish models, lens opacities in the DNase1l1l knockout fish was observed from the 4th month and worsened at the 8th months[80]. The onset of cataract phenotypes in this study far exceeded our observation period. The differences in human and zebrafish lenses and genomes should be kept in mind as well when interpreting the results. Despite the fact that the morphology and function of human and zebrafish lenses are similar after maturation, they are slightly different in the early stage of embryonic development. The way the lens originates from the surface ectoderm and the formation of primary fibre cells are different in humans and zebrafish (discussed in Chapter 3). In addition, the human *HTR1F* gene and its orthologs in zebrafish are only about 60% homologous. These differences in lens structure development and genome may affect the modelling effect of zebrafish on human diseases. Further investigations utilizing other animal models with high homology to humans (e.g. rats, mice), or patient induced pluripotent stem (iPS) cells if available will provide complementary evidence to draw definitive conclusions.

To our knowledge, this is the first study to evaluate a cataract gene using a dual-guide CRISPR-Cas9 RNP in the zebrafish model. Previous cataract gene studies mostly employed morpholinos (MOs) technology to generate

zebrafish mutants[141,244-246]. However, MOs do have some unavoidable limitations such as significant off-target toxicity and short inhibition time window (2-4 days)[247]. Moreover, Kok *et al.* showed about 80% of the published morpholino-mediated phenotypes they tested were not recapitulated in mutant zebrafish embryos[248]. With the rapid development of genome engineering technology, the CRISPR-Cas9 system, which can produce permanent gene modifications with relatively low off-target effects, has begun to replace MOs in functional studies of cataract genes in recent years[79,83,131,143]. These studies used a single guide RNA (gRNA) with Cas9 mRNA to establish mutant zebrafish. In order to maximize the editing efficiency of F0 mutant fish, we modified the current methods, using dual-gRNA targeting two sites of the gene of interest, and the editing efficiencies in our study reached more than 97%. Furthermore, we used Cas9 protein instead of Cas9 mRNA to reduce the potential mosaicism caused by the translation delay of Cas9 mRNA into active enzymatic form[249]. However, due to the nature of the CRISPR-Cas9 system, the genetic mosaicism in F0 mutant zebrafish and potential off-target effects are still systematic limitations of our study. Stable zebrafish knockout lines should be established to validate the findings in F0 mutant fish.

In this study, we developed a rapid and highly efficient CRISPR-Cas9 RNP-mediated mutagenesis strategy to evaluate cataract genes in F0 zebrafish models. The protocol provides a valuable platform that either can facilitate simple and cost-effective phenotypic screening of multiple candidate cataract genes, or can be applied as a primary experiment for in-depth study of specific genes to quickly assess the function of the promising gene in cataract formation prior to spending time and effort to build up stable mutant fish lines. More importantly, for the first time, we established systematic and standardized

cataract assessment criteria for the zebrafish model, which is expected to promote the consistency and comparability of research in the community.

6.3 Multiple genetic factors may contribute to the phenotypic manifestation of paediatric cataract

It is increasingly realized that the classic concept of "one gene, one phenotype" is an oversimplification of the genetic causality, with a more deep and comprehensive understanding of the genetic architecture of diseases enabled by next-generation sequencing technologies. In addition, it has been realized since the last century that no gene acts alone[250-253], the ultimate phenotypic expression of most variants is the result of interaction with additional genetic factors[254-256]. Thus, to be precise, 'monogenic' diseases are defined in terms of the number of genes that trigger the pathogenesis, there might be one or more modifiers that modulate the phenotypes. Paediatric cataract is typically regarded as a monogenic disease. However, remarkable phenotypic variation among family members carrying the same cataract-causing mutation is observed in our study as well as the literature[85,86,89,90], strongly indicating there are additional genetic factors that modify the phenotypic manifestation.

Since inbred laboratory mouse strains are genetically homogeneous and their living environment can be closely controlled, they become a powerful model for modifier gene studies. So far, two modifier genes have been discovered for paediatric cataract using mouse models. Matteson *et al.* identified that the Forkhead box E3 (*Foxe3*) gene, a key regulator of lens

development, modified the cataract phenotype induced by a deletion mutation in the G protein-coupled receptor 161 (*Gpr161*) gene in the *vacuolated lens* (*vl*) mouse[108]. It is reported by Wada *et al.* that the severity and onset time of the cataract phenotypes in *Foxe3* mutant mice were modified by a variant in the phosphodiesterase 6B, cGMP, rod receptor, beta polypeptide (*Pde6b*) encoding gene[107].

While successes have been made in identifying genetic modifiers of cataractogenesis in mouse models, such cataract modifiers have not been identified in humans. Currently, the discovery of human modifiers has mainly relied on large-scale research. A genome-wide association study (GWAS) of 6,365 cystic fibrosis (CF) patients with variants in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene revealed five modifier loci of lung disease severity[96]. An HD MAPS Study performed variance components linkage analysis and genome-wide scans on 295 pedigrees containing 629 sibling pairs and mapped the modifiers of age at neurological onset of Huntington disease (HD) to three linkage regions[257]. In a similar study, a genome-wide linkage scan was conducted in 45 independent Venezuelan HD kindreds containing 1332 individuals, and five regions harboured potential HD modifier genes were detected[258]. However, it is difficult to collect thousands of patients for paediatric cataract—a rare disease with a prevalence of 4.24 per 10,000 cases worldwide[8]. Instead, we investigated modifiers by performing variance components linkage analysis in extended families. In a multi-generation paediatric cataract family, CRCH13, we successfully identified three candidate linkage regions that harboured one or more genetic modifiers for disease severity. This finding provides supporting evidence for multilocus involvement in the phenotype formation of paediatric cataract. The strategy of using the linkage approach to map modifiers in extended pedigrees can minimize

environmental interference, but there remain some challenges. A primary obstacle is the recruitment of large families with sufficient statistical power to detect modifiers of relatively small effect. As discussed in Chapter 5, it is most likely that the pedigree size of family CSA91/110 limited the power to detect potential modifiers. Another major challenge is to identify the specific genetic modifiers within the mapped linkage regions. Although it may be feasible to map an approximate chromosomal location of the modifier loci, the way to determine the exact variants whithout knowing the number, population frequency as well as the magnitude of modification effect of the variants remains to be explored. Nevertheless, our study is the first to investigate modifiers in paediatric cataract and one of the first for any rare monogenic diseases. This research has achieved initial success in locating the regions for potential genetic factors that modulate the severity of paediatric cataract. More broadly, it serves as a starting point for the investigation of the molecular mechanisms underlying multifactorial traits. In the future, more families with paediatric cataract should be studied, and smaller families could be pooled to increase the power to detect modifiers.

It has been observed that some common single nucleotide polymorphisms (SNPs), usually benign on their own, can also modify disease phenotypes when co-segregating with the pathogenic mutation. The common SNPs in the nitric oxide 1 adaptor protein (*NOS1AP*) gene were reported to modulate the QT interval duration and risk of arrhythmias in the long QT syndrome[259]. A recent study demonstrated that a common *SCN5A* polymorphism H558R modifies the clinical phenotype of Brugada syndrome[260]. The variable disease severity observed in family CRCH13 and CSA91/110 may also be driven by the background polygenic risk. It is possible that affected family members with severe early-onset cataracts carry more genetic risk factors. Currently, there

are only two GWAS for cataract, a multiethnic genome-wide association meta-analysis combining results from the GERA (Genetic Epidemiology Research in Adult Health and Aging) and UK Biobank cohorts[261], and a smaller one of the International Cataract Genetics Consortium[262]. None of the polygenic risk loci reported in these two cataract GWAS was found within the parametric or variance components linkage peaks detected in family CRCH13. More GWAS are needed to describe the full spectrum of cataract genetics. Once a substantial proportion of the genetic risk loci for age-related cataract are identified, we would be able to calculate the genetic risk score for each family member to determine if polygenic risk loci contribute to the severity variability.

With the advance in understanding the molecular basis of cataractogenesis, we suggest a more complete and global perspective on the genetic architecture of cataract genomics. It is increasingly realised that cataract is a continuous spectrum of disease, with paediatric cataract with high penetrance at one end, and age-related cataract with major environmental contributors at the other end. At least eight genes such as *EPHA2*, *GJA8*, *GALT*, *SLC16A12*, *HSF4*, *GALK*, *FTL*, and *CRYAA* have been reported to be involved in both the paediatric cataract and age-related cataract[263-271], demonstrating the molecular link between these two forms of cataract. Modifier genes and background polygenic risk factors may act as a bridge between the two ends of cataract, modulate the phenotypic manifestation and contribute to the disease spectrum. Therefore, investigating modifier genes may not only provide opportunities to discover novel therapies for paediatric cataract, but also the more common age-related cataract.

6.4 Conclusion

In this thesis, we investigated the role of a novel gene *HTR1F* in cataractogenesis, and there was a lack of evidence for it being cataract causing in zebrafish models. In addition, three candidate regions for genetic modifiers of disease severity were identified in an extended paediatric cataract family. We also developed an approach for rapid and highly efficient evaluation of cataract candidate genes in F0 zebrafish using CRISPR-Cas9 ribonucleoprotein complexes, as well as a strategy to map potential modifiers in extended paediatric cataract pedigrees. These findings, along with the tools and methods generated in this thesis, will contribute to the identification of novel causative and modifier genes for paediatric cataract. This will lead to a better understanding of cataract pathogenesis and promote the development of non-surgical therapies to prevent or delay the progression of cataracts.

Appendices

Appendix 1 List of primer sequences

Primer	Sequence (5' – 3')	Usage
Zebrafish htr1f expression validation		
zfish_*htr1fa*_F1	GCGCACCTGTGAAACACATT	Validate the *htr1fa-203*
zfish_*htr1fa*_R1	CGGCTAGTGAGCAGATCAGG	(ENSDART00000185312.1) transcript*
zfish_*htr1fa*_F2	ACCCCTGCAAAGAAAATGGC	
zfish_*htr1fa*_R2	CACCCATTTCCCTATACTTCACA	
zfish_*htr1fb*_F	TACATGCTGCACCTGCTCAA	Validate the *htr1fb-201*
zfish_*htr1fb*_R	CATCACACTGCTGACGGTCT	(ENSDART00000075626.5) transcript
zfish_β*actin*_F	ACATGGAGAAGATCTGGC	Positive control
zfish_β*actin*_R	GCATACAGGTCCTTACGGA	
gRNA target sequences validation		

*htr1fa*_PCR_F	AGGCATGGATTCAAACCACACCG	Validate the gRNA target sequences of
*htr1fa*_PCR_R	ACAGCAACAGAACGACTGCAGG	*htr1fa*
*htr1fa*_seq_F	ACACCGGGCCCAGCAAAAACAA	
*htr1fa*_seq_R	AGCCAGCAGACGACAAACGCTC	
*htr1fb*_PCR_F	GCGACCACTGCCATCAACTCCC	Validate the gRNA target sequences of
*htr1fb*_PCR_R	AGAGGTGGTGCAGGTCGGACAA	*htr1fb*
*htr1fb*_seq_F	TCCTGATCACCCGCAAGCTCCA	
*htr1fb*_seq_R	ACCGGCTGCTTCCTCTACGGTT	
NGS library preparation		
*htr1fa*_NGS_F	AGGCATGGATTCAAACCACACCG	Amplicon PCR for the Miseq sequencing
*htr1fa*_NGS_R	ACAGCAACAGAACGACTGCAGG	
*htr1fb*_NGS_F	GCGACCACTGCCATCAACTCCC	
*htr1fb*_NGS_R	AGAGGTGGTGCAGGTCGGACAA	

*Two pairs of primers were designed to detect *htr1fa-203* due to their large size.

Appendix 2 Publication of material from Chapter 3

The following article was published in *Methods.*

Zhao D, Jones JL, Gasperini RJ, Charlesworth JC, Liu GS, Burdon KP. Rapid and efficient cataract gene evaluation in F0 zebrafish using CRISPR-Cas9 ribonucleoprotein complexes. *Methods*. 2021 Oct;194:37-47.

Methods 194 (2021) 37–47

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Rapid and efficient cataract gene evaluation in F0 zebrafish using CRISPR-Cas9 ribonucleoprotein complexes

Duran Zhao [a], Johanna L. Jones [a], Robert J. Gasperini [b], Jac C. Charlesworth [a], Guei-Sheung Liu [a,c,1], Kathryn P. Burdon [a,d,1,*]

[a] Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia
[b] School of Medicine, University of Tasmania, Hobart, Tasmania, Australia
[c] Ophthalmology, Department of Surgery, University of Melbourne, East Melbourne, Victoria, Australia
[d] Department of Ophthalmology, Flinders University, Bedford Park, South Australia, Australia

ARTICLE INFO

Keywords:
CRISPR
Cas9
Ribonucleoprotein (RNP) complex
Genome editing
Cataract
Zebrafish

ABSTRACT

Cataract is the leading cause of blindness worldwide. Congenital or paediatric cataract can result in permanent visual impairment or blindness even with best attempts at treatment. A significant proportion of paediatric cataract has a genetic cause. Therefore, identifying the genes that lead to cataract formation is essential for understanding the pathological process of inherited paediatric cataract as well as to the development of new therapies. Despite clear progress in genomics technologies, verification of the biological effects of newly identified candidate genes and variants is still challenging. Here, we provide a step-by-step pipeline to evaluate cataract candidate genes in F0 zebrafish using CRISPR-Cas9 ribonucleoprotein complexes (RNP). Detailed descriptions of CRISPR-Cas9 RNP design and formulation, microinjection, optimization of CRISPR-Cas9 RNP reagent dose and delivery route, editing efficacy analysis as well as cataract formation evaluation are included. Following this protocol, any cataract candidates can be readily and efficiently evaluated within 2 weeks using basic laboratory supplies.

1. Introduction

Cataract is an opacity of the transparent crystalline lens in the eye that interferes with the path and refraction of light to the retina. It is the principal cause of reversible visual impairment or blindness worldwide [1]. While cataract in adults is usually treatable with access to surgery, congenital or paediatric cataract can lead to life-long irreversible vision impairment or blindness, despite best attempts at treatment. A significant proportion of paediatric cataract is inherited and it often occurs alongside other syndromic features. Identifying the causative genes for inherited paediatric cataract can improve diagnostic accuracy through facilitating genetic testing. Gene identification can also improve understanding of the molecular mechanisms underlying cataractogenesis which may facilitate the development of medical therapy and individualized treatment [2]. So far, more than 40 genes have been linked to congenital or early-onset cataract [3–5] and they account for 60–70% of children with the disease [6–8]. However, significant challenges exist for identifying additional genes. Genome sequencing technologies are able to identify candidate variants, but validating these genes in multiple families is increasingly difficult for rare genetic causes of disease. Medium to high throughput techniques for evaluating the role of candidate genes in cataract formation are required.

These human cataract candidate genes can be verified in animal models. Gene knockout models that display cataract formation provide strong evidence that a candidate gene is important in lens biology and cataract development. Several animal models have been used in cataract research, including mice, rats, rabbits, dogs, and zebrafish. However, the relatively small number of offspring and long generation period limit the application of mammalian models. In contrast, zebrafish (*Danio rerio*) become an outstanding model system due to a number of advantages. Their rapid development and simple husbandry enable the generation of large numbers of fish for genetic research work from a single mating. Zebrafish embryos develop externally and remain transparent for the first few days of development, providing technical advantages for embryo microinjection. The eyes of the embryos are relatively large [9] and become functional by 3 days-post-fertilization (dpf) [10], allowing for

* Corresponding author at: 17 Liverpool Street (Private Bag 23), Hobart, Tas 7000, Australia.
E-mail address: kathryn.burdon@utas.edu.au (K.P. Burdon).
[1] These senior authors contributed equally to this work.

https://doi.org/10.1016/j.ymeth.2020.12.004
Received 5 November 2020; Received in revised form 22 December 2020; Accepted 29 December 2020
Available online 6 January 2021
1046-2023/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

the observation of lens morphology at an early stage. The genome of zebrafish has been sequenced and annotated. Around 70% of human genes have functional orthologs in zebrafish [11]. This high level of homology to humans means that most human genes can be interrogated in zebrafish and the functions of those genes are largely similar between the two species.

Clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 nuclease system has become a widely used targeted gene editing technology in zebrafish models. Its advantages over other methods such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) include easy design processes and rapid synthesis of reagents. Morpholinos (MOs) have been used extensively for temporary knockdown of genes in zebrafish, but CRISPR-Cas9 provides a permanent genetic modification with relatively low off-target effects. The system is comprised of guide RNA (gRNA) and Cas9 nuclease, which are combined to form a ribonucleoprotein (RNP) complex. The gRNA consists of a target-specific CRISPR RNA (crRNA) and a universal trans-activating RNA (tracrRNA). The crRNA is designed to be complementary to the target DNA site, a 20 base pair sequence next to a protospacer adjacent motif (PAM) [12]. The gRNA forms a complex with the specific target site, directing the Cas9 nuclease to that locus. Once at the appropriate site, Cas9 nuclease can induce double-strand breaks (DSB), triggering non-homologous end joining (NHEJ) to repair the break. The error-prone NHEJ can produce small insertions or deletions between the end of the broken DNA strands, resulting in efficient mutagenesis [12–14].

Although many articles have described disease modelling in zebrafish, specific protocols for evaluation of cataracts in CRISPR-Cas9 edited zebrafish have not been reported. In this article, we provide a rapid and highly efficient method for evaluating cataract candidates in F0 zebrafish using CRISPR-Cas9 RNP. Detailed pipelines of CRISPR-Cas9 RNP design and formulation, microinjection, dose and delivery route optimization of CRISPR-Cas9 RNP, editing efficacy analysis as well as cataract formation evaluation are included.

2. Overview

To better describe the pipeline of using CRISPR-Cas9 RNP to evaluate cataract genes in F0 zebrafish, we use *HTR1F* (5-hydroxytryptamine receptor 1F) as an example cataract candidate gene to illustrate the methods applied in this article. The *HTR1F* protein is a G-protein coupled serotonin receptor. Serotonin is crucial in lens transparency [15] and increased serotonin level has been shown to lead to lens opacities in rats [16]. The data presented in Section 4.4 and lens images shown in Section 4.5.2 were generated from *HTR1F* knockout experiments. Fig. 1 illustrates the overall experimental workflow.

3. Materials

3.1. Reagents and equipment

The reagents and equipment utilized in this protocol are summarized in Table 1 and Table 2.

3.1.1. Reagent setup

3.1.1.1. E2 embryo media with methylene blue. The 0.5x E2 with 0.5 mg/l Methylene Blue working solution was prepared following the Zebrafish International Resource Center (ZIRC) protocol: https://zebrafish.org/wiki/protocols/nursery. The final concentrations of E2 media components are: 7.5 mM NaCl, 0.25 mM KCl, 0.5 mM MgSO4, 75 μM KH2PO4, 25 μM Na2HPO4, 0.5 mM CaCl2, and 0.35 mM NaHCO3.

3.1.1.2. 1x PTU (200 μM) egg water. Add 0.076 g PTU (1-phenyl-2-thiourea) to 50 ml autoclaved Milli-Q water to prepare a 50x PTU (10,000 μM) stock solution. Store the 50x PTU stock solution at room temperature. Prepare fresh 1x PTU (200 μM) egg water each time as needed by diluting 50x PTU stock solution 50 times with E2 embryo media with Methylene Blue.

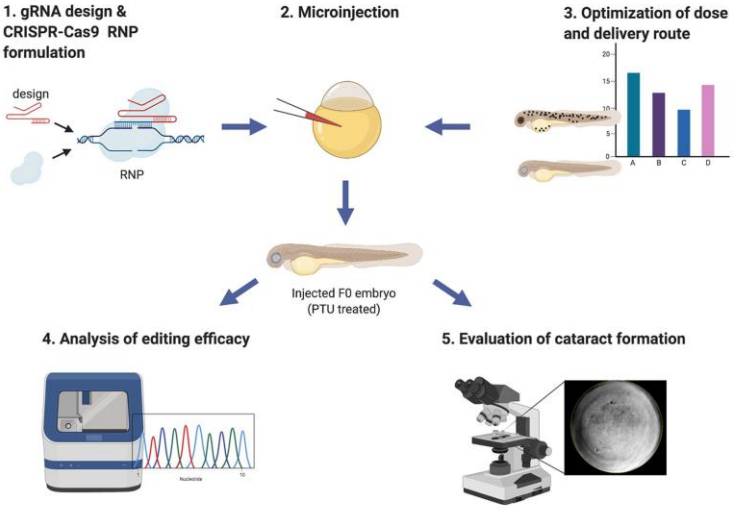


Fig. 1. Overview of the main steps of the cataract gene evaluation pipeline. Each step corresponds to a section of the protocol with the same title.

Table 1
Reagents used in this protocol.

Item	Supplier	Catalogue No.
NaCl, KCl, $MgSO_4$, KH_2PO_4, Na_2HPO_4, $CaCl_2$, $NaHCO_3$. KCl powder	Sigma-Aldrich	–
Methylene Blue	AQUASONIC	PL084
Alt-R® CRISPR-Cas9 crRNA	IDT	–
Alt-R® CRISPR-Cas9 tracrRNA	IDT	1072533
Alt-R® S.p.HiFi Cas9 Nuclease V3	IDT	1081060
IDTE pH 7.5 (1X TE Solution)	IDT	11-01-02-02
1 M HEPES buffer	Sigma-Aldrich	H3375
Nuclease-Free Duplex buffer	IDT	11-01-03-01
PTU (1-phenyl-2-thiourea)	Sigma-Aldrich	P7629
Tricaine methanesulfonate	Glentham Life Sciences	GE5936
Phire Tissue Direct PCR Master Mix	Thermo Scientific	F170s
Agarose powder	Bioline	BIO-41025
TAE buffer (Tris-acetate-EDTA) (50X)	Thermo Scientific	B49
HyperLadder™ 100 bp	Meridian Bioscience	BIO-33056
AMPure XP beads	Beckman Coulter	A63880
CleanSEQ Dye-Terminator Removal kit	Beckman Coulter	A29151
BrilliantDye™ Terminator (v3.1) Cycle Sequencing kit	NimaGen	BRD3-1000
BigDye™ Terminator v1.1 & v3.1 5X Sequencing Buffer	Thermo Scientific	4336699
Ethanol	Sigma-Aldrich	E7023

NOTE:

- PTU is highly toxic by ingestion and is a skin sensitizer. Weigh and prepare the PTU in a fume hood and always have gloves on during use.

3.1.1.3. Alt-R CRISPR-Cas9 crRNA, tracrRNA and Cas9 nuclease aliquots. The CRISPR oligos are usually received dehydrated. Resuspend the Alt-R CRISPR-Cas9 crRNA and tracrRNA in IDTE buffer to final concentrations of 100 μM each. Store the rehydrated CRISPR oligos and S.p.HiFi-Cas9 nucleases according to supplier instructions. At the time of these experiments, we aliquoted crRNA, tracrRNA and S.p.HiFi-Cas9 Nuclease into 2 μl small stocks and stored at −20 °C to reduce unnecessary freeze–thaw cycles.

Table 2
Equipment required for this protocol.

Item	Supplier	Catalogue No.
Standard microcentrifuge tubes, 1.5 ml	Labcon	#3039-560-000
RNase-free microfuge tubes, 1.5 ml	Thermo Scientific	AM12400
Filtered sterile pipette tips	Labcon	–
Desktop microcentrifuges	Bioline Global	SF7000
Vortex mixer	Ratek	IC-VM1
Block heater	Ratek	DBH200
Capillary glass (1.0 mm × 0.58 mm × 10 cm)	Harvard Apparatus	GC100F-10
Micropipette puller	Sutter Instrument	P-87
Pico-liter injector	Warner Instruments	PLI-10
Petri dish	NEST Scientific	752001
Nunc™ glass base dish (27 mm)	Thermo Scientific	150682
Incubator	Nüve	EN025
Stereomicroscope	Olympus	SZX16
Inverted microscope	Nikon	Eclipse Ti
S-Series compact USB 2.0 camera	Mightex	SCE-B013-U
sCMOS camera	Andor	Zyla 4.2 PLUS
Veriti™ 96-well Thermal cycler	Applied Biosystems	4375786
PowerPac™ basic power supply	Bio-Rad	1645050
Magnetic plate	ALAPAQUA	A001322
3500 Genetic Analyzer	Thermo Scientific	A27772

3.1.1.4. Cas9 working buffer (20 mM HEPES; 150 mM KCl, pH 7.5). Mix the following:

Component	Amount
1 M HEPES buffer	300 μl
KCl powder	0.168 g
RNase-free H_2O	14.7 ml
Final volume	15 ml

After KCl powder completely dissolves in the solution, aliquot the Cas9 working buffer into 1 ml stocks and store at 4 °C.

3.2. Zebrafish maintenance and breeding

AB strain zebrafish (source: Zebrafish International Resource Center (ZIRC)) were housed in a circulating water system under standard conditions [17], with a daily cycle of 14 h light/10 h dark. Embryos were obtained from the natural pairwise spawns of the wildtype AB zebrafish and were collected and raised in an incubator at 28.5 °C in E2 media with Methylene Blue until 5 dpf. The animal research in this study was approved by the University of Tasmania Animal Ethics Committee (project number: A0017743) in accordance with the Australian National Health and Medical Research Council Code of Practice for the Care and Use of Animals for Scientific Purpose.

4. Methods

4.1. gRNA design and CRISPR-Cas9 RNP formulation

4.1.1. gRNA design

The first and most critical step is the design of gRNA. The editing efficiencies of each gRNA can vary greatly, thus a well-designed gRNA is essential for successful CRISPR-Cas9-based gene disruptions. It is worth noting that zebrafish frequently have more than one ortholog of human genes [11]. For example, the *HTR1F* gene has two orthologs in the zebrafish genome, *HTR1Fa* and *HTR1Fb*. Collect the genomic DNA sequence, transcripts and protein information of all the orthologs of your target gene before starting gRNA design. There are multiple zebrafish genomic resources including Zebrafish Information Network (ZFIN) (http://zfin.org/), Ensembl (http://www.ensembl.org/Danio_rerio/Info/Index), UCSC genome browser (https://genome.ucsc.edu/) and NCBI zebrafish genome (https://www.ncbi.nlm.nih.gov/). Several CRISPR design tools are available online, such as CHOPCHOP (https://chopchop.cbu.uib.no/), CRISPOR (http://crispor.tefor.net/) and E-CRISP (http://www.e-crisp.org/E-CRISP/). For beginners we recommend CHOPCHOP, as it is simple and intuitive. Here we use CHOPCHOP as an example to explain the steps of gRNA design.

1. Input the gene identifiers or genomic coordinates or paste the genomic sequence of your target gene into CHOPCHOP, specify the organism and CRISPR mode to be "Danio rerio (danRer11/GRCz11)" and "CRISPR-Cas9" "knock-out". Under the "Options", by default, the target region, PAM sequence and number of mismatches of predicted off-targets have been set to "coding region", "NGG" and "up to 3" respectively. These parameters can be adjusted according to individual needs. After setting, click "Find Target Sites!" to perform guide search.
2. A list of candidate gRNAs with relevant information such as target sequence, genomic location, number and type of predicted off-targets, GC content and efficiency score should be displayed. Detailed information of predicted off-targets including location, number of mismatches, and sequence can be found by clicking on an individual guide. Based on this information, select suitable gRNAs according to the principles below:

- **Position.** It is recommended that the target sites are located within the most upstream exon that is conserved across all target transcripts

of the gene of interest. This is most likely to generate frameshift mutations and early termination of the protein sequence. Alternatively, gRNAs can be designed to target exons encoding known functional protein domains. In this way, the protein function could be disrupted even with non-frameshift alleles.

- **Editing efficiency.** Select gRNAs with the highest on-target efficiency and the fewest predicted off-targets. Generally, guides with predicted off-targets containing three or fewer mismatches to the target sequence should be avoided [13]. When multiple gRNAs have similar on and off-target efficiencies, priority is given to gRNAs with off-target loci on chromosomes other than the one that the target gene resides, or in non-coding regions (e.g. intergenic) that are unlikely to have impacts on gene expression.
- **Number.** It has been shown that targeting a single locus in a coding region only generates biallelic loss-of-function alleles 44% of the time, however, targeting a coding region at two, three, or four independent loci generates biallelic loss of function variants 79%, 93%, and 98% of the time respectively [14]. To increase the success rate of introducing mutations and producing loss of protein function, we recommend designing 2 to 4 gRNAs for each target gene. Nonetheless, it is suggested the total number of gRNAs injected per embryo does not exceed 8, otherwise it may affect the survival and development of zebrafish embryos as well as increase the off-target effects [14]. The exact number of gRNAs to be injected into the zebrafish embryo depends on which gene is being targeted, the number of orthologs of the target gene, and how many genes are to be targeted simultaneously.
- **GC% of the guide sequence.** A previous study has identified that gRNAs with over 50% GC content exhibit high mutagenesis activity [18].

3. Order CRISPR oligos. The fully functional gRNA consists of the target-specific crRNA (designed above) and the tracrRNA with a universal sequence. The two components can be ordered pre-synthesized into the gRNA which can provide greater stability for challenging experimental conditions. Alternatively, for greater flexibility, they can be ordered individually and assembled in house.

NOTE:

- The gRNA target sequences should be verified by Sanger sequencing in the strain of fish to be used in your experiment. It is possible, depending on the strain and source of the fish, that polymorphisms are present compared to the reference used to design the gRNAs. Even a single base mismatch can significantly reduce the targeting efficiency of the CRISPR system [19].
- To eliminate the interference caused by CRISPR protein as well as injection-induced defects, it is necessary to involve a negative control in the experiments. gRNA against a non-existent gene in the zebrafish genome such as *LacZ*, a gene of *E. coli*, can be designed as a negative control. Align the target sequence of your negative control to the zebrafish genome to make sure there is no target in the zebrafish, and check the predicted off-targets of the negative control to avoid targeting your gene of interest.

4.1.2. CRISPR-Cas9 RNP formulation

Two approaches have been commonly used to construct the CRISPR-Cas9 system for *in vivo* microinjection. Early methods co-inject gRNA and Cas9 mRNA into zebrafish embryos. With the proliferation of CRISPR-Cas9 reagents and suppliers, the more handy and cost-effective way is to directly deliver gRNA and Cas9 protein as RNP complex. Compared to mRNA-mediated delivery, RNP has the advantages of high stability, fast onset, and more importantly, by skipping the transcription process, it is likely to reduce genetic mosaicism due to Cas9 expression delay [20,21]. The protocol below used the Alt-R CRISPR-Cas9 system to synthesize CRISPR-Cas9 RNP and has been adapted from instructions provided by IDT (online instructions: Zebrafish embryo microinjection: Ribonucleoprotein delivery using the Alt-R CRISPR-Cas9 System).

1. On the day of injection, thaw crRNA, tracrRNA and HiFi-Cas9 nuclease stocks at room temperature.
2. Preheat the heat block to 95 °C.
3. Prepare 9 μM gRNA working solution by mixing the following components:

Component	Volume (μl)
100 μM Alt-R® CRISPR-Cas9 crRNA	0.9
100 μM Alt-R® CRISPR-Cas9 tracrRNA	0.9
Nuclease-Free Duplex buffer	8.2
Final volume	10

In the case of multiple crRNAs, mix the crRNAs equally while keeping the total amount of crRNAs constant. The volumes provided here are the optimized results generated in our experiments (see Section 4.3). The user can adjust the volumes as required.

4. Heat gRNA working solution at 95 °C for 5 min.
5. Remove the working solution from the heat block and leave on the benchtop to cool to room temperature.
6. Reset the heat block to 37 °C.
7. Prepare 1.5 μg/μl Cas9 protein working solution. Mix the following reagents by flicking the tube and spin down briefly. Do not vortex.

Component	Volume (μl)
10ug/ul Alt-R® S.p.HiFi Cas9 Nuclease V3	1.5
Cas9 working buffer	8.5
Final volume	10

The volumes provided here are the optimized results generated in our experiments (see Section 4.3). The user can adjust the volumes as required.

8. Assemble the CRISPR-Cas9 RNP complexes. For each experimental group, combine 3 μl of 9 μM gRNA working solution with 3 μl of 1.5 μg/μl Cas9 protein working solution. Mix the reagents by flicking the tube and spin down briefly. Do not vortex.
9. Incubate the CRISPR-Cas9 RNP complexes at 37 °C for 10 min.
10. Remove the CRISPR-Cas9 RNP complexes from the heat block and leave on the benchtop to cool to room temperature.

NOTE:

- Keep the whole process RNase-free. Clean work surfaces with RNase-inactivating solution, handle all materials with gloves and use RNase-free tips and tubes.
- In this protocol we used HiFi Cas9 nuclease, which has significantly reduced off-target effects compare to wildtype S.p.Cas9 nuclease [22].

4.2. Microinjection

Two possible delivery strategies can be applied in embryo microinjection: inject into the single cell or the yolk sac. In theory, cell injections are more likely to induce gene editing and increase the chance of editing both alleles simultaneously. However, it is technically challenging and time-consuming, limiting the number of embryos that can be injected in a single experiment. In comparison, yolk injections are high throughput, enabling an increased number of injected embryos to be achieved in one sitting. Additionally, yolk injections have been shown to have similar mutagenesis efficiencies to single-cell injections, with four times the

reagent dose [23].

1. The afternoon before scheduled injection, set up zebrafish in breeding tanks and separate the males and females with dividers. We use two males to three females to increase the number of fertilised eggs produced for injection.
2. On the morning of injection, pull out the dividers and allow the fish to spawn naturally. After approximately 10–15 min, collect the embryos and inject 1 nl of the CRISPR-Cas9 RNP into the cell of each one-cell stage embryo or 4 nl into the yolk sac instead. Use forceps to gently remove the embryos that have divided past the single-cell stage or fail to be injected. Save uninjected embryos from each clutch as controls.
3. Incubate both injected and uninjected embryos at 28.5 °C in E2 media with Methylene Blue in petri dishes. After 4–6 h, remove unfertilized or dead eggs, refresh the E2 media and record the number of surviving embryos. From 24hpf, raise the embryos in 1x PTU egg water to suppress pigment formation. Change the 1x PTU egg water twice a day as well as remove any dead embryos. Check and record the survival and dysmorphic rates. The injected larvae can be assessed for cataract and collected for editing efficiency analysis at 4dpf.

NOTE:

- Inject embryos as soon as possible and not beyond the one-cell stage. Early injection maximizes editing and reduces the genetic mosaicism, making the subsequent evaluation of editing efficiency easier.
- It is recommended to line up and orient the embryos before injection. This can speed up the injection process when performing cell injections, and avoids accidentally injecting the relatively large amount of CRISPR-Cas9 RNP (4 nl) into the cell when performing yolk injections.
- The injection should be a single smooth motion. Do not adjust the orientation of the needle in the embryo. After injecting the RNP complexes, gently withdraw the needle and avoid tear out of the cytoplasm or yolk sac.
- Power calculations should be performed to determine the minimum number of injected embryos required for statistical analysis. Inject more embryos than required to allow for natural attrition of early embryos. The number can be based on the survival rate calculated from optimization experiments.
- In general, more than 50% of the embryos in the injected group are expected to show normal development similar to the uninjected siblings.
- High levels of embryo death may be caused by excessive reagent concentration or may indicate that the target gene is essential for embryo survival. Repeat injections with a decreased reagent concentration.
- Cloudy yolk sacs in surviving embryos likely indicate infection. Rinse the embryos carefully with E2 media prior to injection to remove any debris. Timely removal of dead embryos also helps to maintain the health of remaining embryos.

4.3. Optimization of dose and delivery route of CRISPR-Cas9 RNP

To achieve the optimal editing efficiency with minimum disruption to the development of the zebrafish embryo, we used gRNA targeting *tyrosinase (tyr)* gene to optimize the dose and delivery route of CRISPR-Cas9 RNPs. Knockout of the *tyr* gene results in pigmentation defects, which can be easily observed as early as 2 dpf (Fig. 2A) and quantitatively analyzed by measuring the pigment areas on the retina and the body of the injected zebrafish (Figure S1). *tyr*-gRNA is commercially available from IDT, or it can be designed by the user through the web-based CRISPR design tools. In the initial dose optimization process, the reagents were injected into the cell of one-cell-stage embryos at a fixed volume of 1 nl. We first tested a range of gRNA concentrations from 0.5 μM to 13.5 μM, while keeping Cas9 nuclease at 0.75 μg/μl. A significant reduction of pigmentation in the retina and body were observed in 4.5 μM and 13.5 μM gRNA-injected zebrafish (Fig. 2B and C). However, no statistical difference in pigment loss was found between 4.5 μM and 13.5 μM gRNA-injected zebrafish (retina: $p = 0.9427$; body: $p = 0.9356$). We then applied 4.5 μM and 13.5 μM gRNA to examine the effect of varying Cas9 concentration at 0.75 μg/μl or 1 μg/μl. In all experimental conditions that tested in the Cas9 concentration optimization, we observed significant pigment loss on both retina and body in the injected groups compared with the uninjected group (Fig. 2D and E), indicating high mutagenesis load. Although there were no significant differences in pigment loss between the injected groups in the optimization of Cas9 concentration (data not shown), the combination of 4.5 μM gRNA and 0.75 μg/μl Cas9 nuclease presented with the highest survival rates (93%, 57 of 61 injected embryos). We therefore set the concentration of gRNA and Cas9 nuclease to be 4.5 μM and 0.75 μg/μl.

We next used this condition to determine the delivery route. With the identical concentration of gRNA and Cas9 nuclease, we compared the pigment loss of injecting 1 nl of CRISPR-Cas9 RNP into the cell (Fig. 3A) and 4 nl into the yolk (Fig. 3B) of one-cell stage embryos. Our data showed that yolk injections had significantly higher body pigment loss (Fig. 3D, $p = 0.042$) and slightly higher survival rate (cell: 73%, 19 of 26 injected embryos; yolk: 88%, 30 of 34 injected embryos) than cell injections. In addition, considering the simple operation and high throughput of yolk injection, we chose to perform yolk injections in all subsequent experiments.

4.4. Analysis of editing efficacy in F0 embryos:

According to the principles described in Section 4.1.1, we designed 4 gRNAs targeting different locations of *HTR1Fa* exon 2 and *HTR1Fb* exon 1 respectively (Figure S2 and S3, Table S1). We first injected embryos with individual gRNAs and analyzed the editing efficiency. A detailed protocol of the single gRNA editing efficacy analysis is presented below. To improve consistency and convenience, we employed Phire Tissue Direct PCR Master Mix kit in the target region amplification step.

1. Collect 5 larvae from each gRNA injected group at 4pdf, amplify the gRNA target region of individual larvae using the Phire Tissue Direct PCR Master Mix kit according to the manufacturer's instructions.
2. After PCR, inspect the PCR products on a 2% (w/v) gel to confirm that the target region has been amplified correctly, then Sanger sequence the amplicons.
3. Upload the Sanger sequencing data (AB1 file) of both experimental and uninjected wildtype samples to the free online CRISPR analysis tool — Inference of CRISPR Edits (ICE) (https://www.synthego.com/products/bioinformatics/crispr-analysis) and specify the guide sequence. Once all the samples have been uploaded, click "Analyze" to perform the CRISPR editing efficiency analysis.

The 8 designed gRNAs displayed various editing efficiencies, with the average indel percentage of individual embryos ranging from 53.4% to 88.4% (Fig. 4). We next pooled all 8 gRNAs at equal concentrations and co-injected the gRNA mixture with Cas9 nuclease into the embryos. Although a high survival rate was observed (87%, 68 of 78 injected embryos), a high dysmorphic rate (54%, 20 of 37 imaged embryos) was also observed in the injected embryos, suggesting 8 gRNAs may be too toxic to the development of embryos. We then injected CRISPR-Cas9 RNPs carrying a mixture of the two most efficient gRNAs of both *HTR1F* orthologs into the embryos. The four-gRNA injected embryos showed a similar survival rate (78%, 58 of 74 injected embryos) to the eight-gRNA ones, but the dysmorphic rate (12%, 7 of 58 imaged embryos) decreased dramatically to an acceptable level.

NOTE:

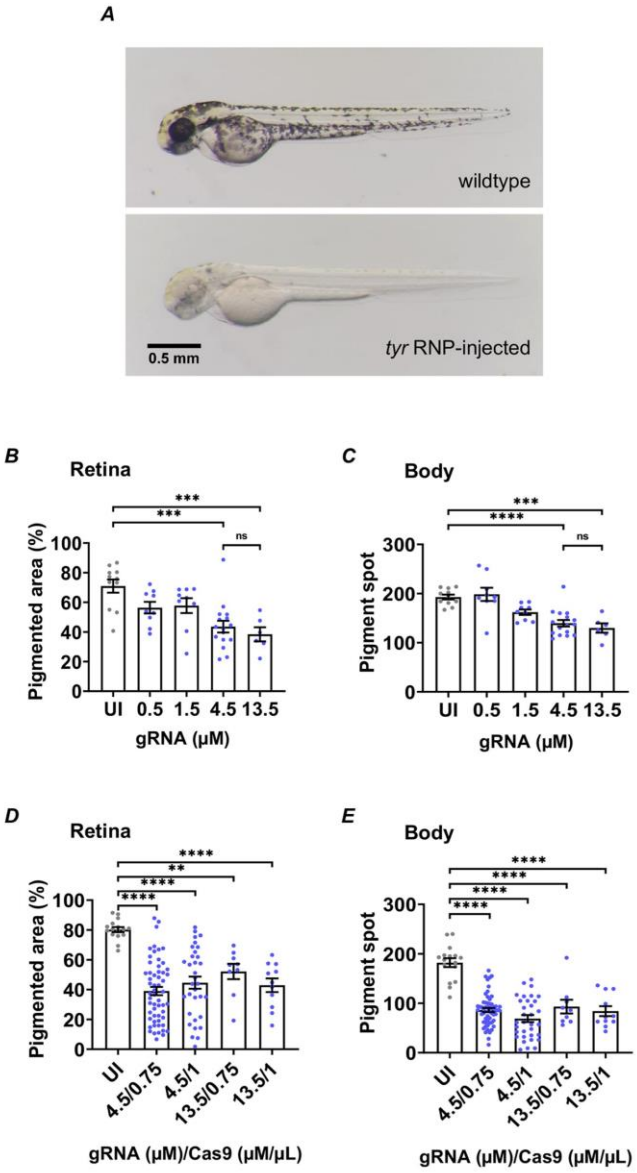


Fig. 2. CRISPR-Cas9 RNP dose optimization. A. Zebrafish embryos at 2dpf. Embryos injected with CRISPR-Cas9 RNP against *tyr* gene (bottom) had observable pigment loss compare to uninjected wildtype zebrafish (top). B and C. Optimization of *tyr* gRNA concentration (UI: n = 11, 0.5 μM: n = 9, 1.5 μM: n = 9, 4.5 μM: n = 16, 13.5 μM: n = 6). D and E. Optimization of Cas9 concentration (UI: n = 16, 4.5 μM/0.75 μg/μl: n = 57, 4.5 μM/1μg/μl: n = 33, 13.5 μM/0.75 μg/μl: n = 9, 13.5 μM/1μg/μl: n = 11). The pigment areas on the retina (B, D) and body (C, E) of zebrafish injected with a series of different concentrations of *tyr* gRNA or Cas9 were measured and compared (UI: uninjected; $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$; ns: not significant; one-way ANOVA and Tukey's multiple comparisons test; Mean ± SEM).

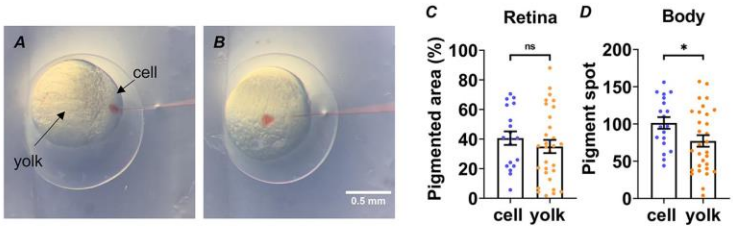


Fig. 3. CRISPR-Cas9 RNP delivery route optimization. A and B. Illustrations of cell injection (A) and yolk injection (B). The zebrafish embryos were cell-injected (n = 19) or yolk-injected (n = 30) with *tyr* RNP. The pigment areas on the retina (C) and body (D) of zebrafish in the cell injection group and yolk injection group were measured and compared. (*$p < 0.05$, ns: not significant; unpaired two-tailed *t*-test; Mean ± SEM).

- Alternative methods for assessing genome editing include T7E1 assay [24], high-resolution melting analysis (HRMA) [25], heteroduplex mobility assay (HMA) [26], CRISPR-STAT assay [27], and next-generation sequencing (NGS) [28].
- In the case of multi-guide-induced complex editing, we recommend applying NGS for editing efficacy analysis. ICE cannot clearly distinguish between noise and sample signal less than 5%, cannot detect deletions >150 bp for a multi-guide experiment, and can only analyze up to 3 gRNAs for the multi-guide knockout.
- When amplifying the target region to detect mutagenesis activities that are induced by multiple gRNAs, properly design primer pairs to capture all gRNA target sites in a single amplicon, otherwise deletions generated between targets will remove primer binding sites and prevent PCR based amplification.

4.5. Evaluation of cataract formation by microscope imaging

4.5.1. Imaging

1. On the fourth day after injection, anesthetize the larvae in tricaine (300 mg/l) for 1–2 min.
2. Transfer a single larva to a glass base dish (27 mm diameter).
3. Take whole-embryo images to record the gross morphology of the larvae. Our lab uses an SZX16 zoom stereomicroscope with Mightex S-Series compact USB 2.0 camera and SSClassicCamera software.
4. Mount the larva in 0.5% (w/v) agarose and quickly orient it to a position in which the lens can be viewed directly through the bottom of the dish before the agarose hardens.
5. Photograph the lens of the mounted larva using Differential Interference Contrast (DIC) microscopy settings with a Ti live cell microscope, Zyla 4.2 PLUS sCMOS camera and NIS-Elements AR acquisition software (or equivalent). Label the images with the fish number, experiment group and the time of imaging.

NOTE:

- Keep the melted agarose at approximately 40 °C, avoiding too high or too low temperatures. High temperatures will damage the larva, while low temperatures will cause the agarose to set too quickly, without allowing enough time to orient the larva.
- Remove as much liquid as possible from the larva before mounting to prevent the larva from floating in the liquid and not being fixed.
- Carefully wipe the glass bottom of the dish every time before mounting the larva, as any dirt or droplets on the dish bottom will be photographed into the lens images and interfere with the subsequent assessment of cataract.
- Position the larva with one eye facing and touching the bottom of the dish. Avoid bubbles or agarose between the larva eye and the dish bottom, which will create artefacts or blur the lens image, making cataract assessment difficult.

4.5.2. Cataract assessment

The lens images were used to determine the cataract status of each fish. There are currently no standard criteria for the characterization of cataracts in zebrafish models. The difficulties in cataract assessment of zebrafish are mainly focused on the identification of normal developing

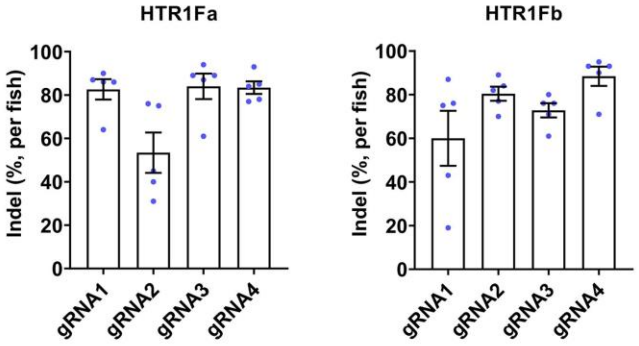


Fig. 4. Editing efficiency analysis for individual gRNA. 5 injected larvae were randomly selected at 4dpf from each gRNA injected group, and the target region of individual embryos was amplified and analyzed for indel percentage.

lenses and artefacts from real cataracts. We reviewed the literature on the developing zebrafish lens morphology [29–33] as well as studies of cataract in zebrafish models [32,33], and reviewed the lens images produced from wildtype (uninjected) embryos from our experiments to understand the morphology of normal developing zebrafish lens. We then generated a detailed cataract assessment and scoring criteria for zebrafish models based on more than 700 lens images obtained from our experiments (Fig. 5 and Fig. 6).

1. Crop lens images in ImageJ [34] to display only the lens region of the eye to avoid calling bias based on the phenotype of surrounding ocular features.

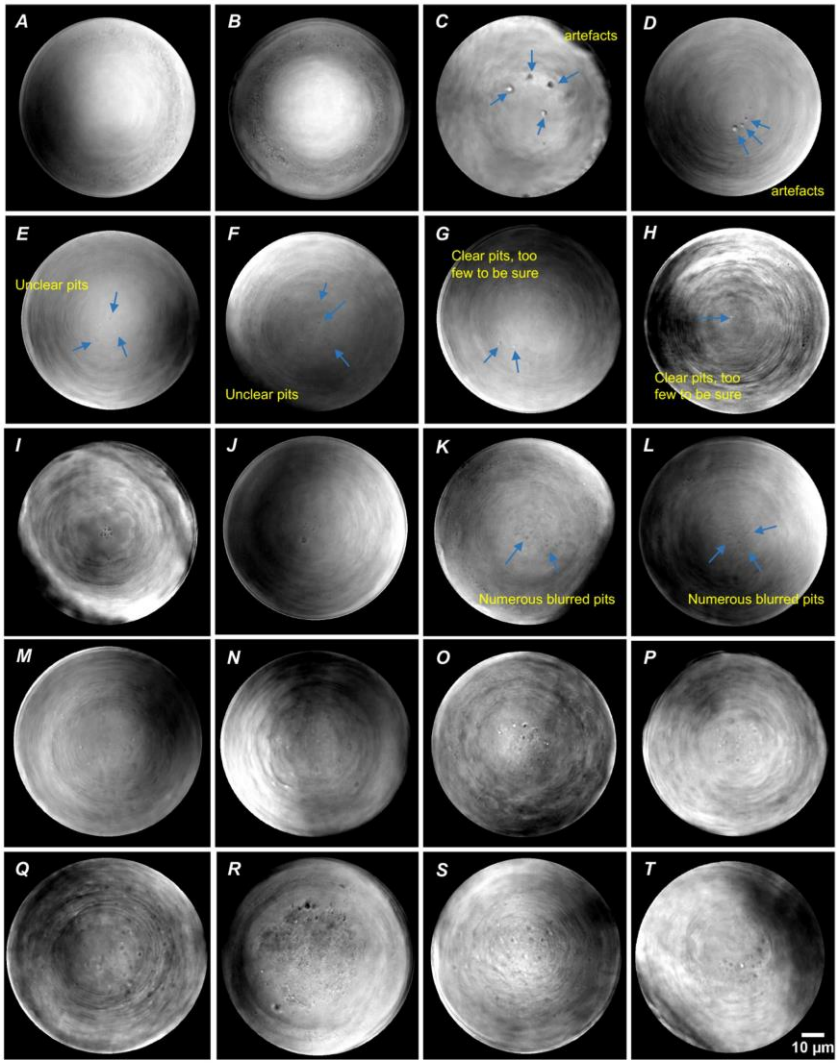


Fig. 5. Zebrafish cataract assessment criteria (pitting).

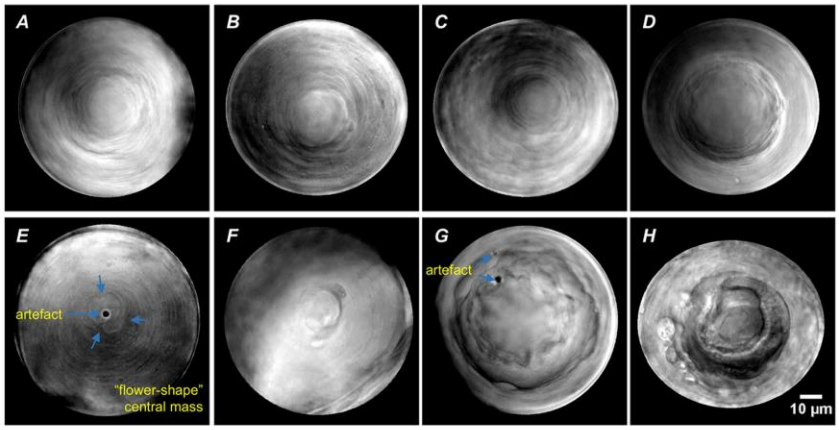


Fig. 6. Zebrafish cataract assessment criteria (central mass).

2. Collate all images into a single group-set and perform blinding using a random number generator, to randomise and reassign image identification numbers.
3. Provide the blinded data to two researchers who should independently assess and score the images based on the standard images in Fig. 5 and Fig. 6 and descriptions below.

- **Pitting.** Described as a dot-shaped depression on the surface of the lens. The pits can be diffused or aggregated in the central lens. Where pits exist, the uniform and smooth structure of the lens is destroyed.

Catagory	Description
Pitting = No/ cataract = No	Can be: • Developmental pitting (Fig. 5A and B). The pits are mostly fine and spread over the outer lens. They are undegraded organelles. The central lens is clear and uniform. • Artefacts (Fig. 5C and D). High reflective pits with very round and regular shape or pits with blurred edges are considered to be artefacts. The central lens is clear and uniform.
Pitting=?/ cataract = No	Can be: • Unclear pits (Fig. 5E and F), few in number in the centre of the lens. The rest of the lens is smooth and uniform. • Clear pits, but too few to be sure (Fig. 5G and H). There are one or two clear pits in the centre of the lens. The rest of the lens is smooth and uniform.
Pitting = Yes/ cataract = Yes (mild, score = 1)	Can be: • A small number of clear pits, enough to be confident (Fig. 5I and J). The pits are located in the central lens rather than the outer lens. • Pits may be blurred, but are numerous in the central lens (Fig. 5K and L). This may be caused by loss of focus during imaging.
Pitting = Yes/ cataract = Yes (moderate, score = 2)	Significant pits (Fig. 5M-P). Can be seen all over the lens or cluster in the central lens.
Pitting = Yes/ cataract = Yes (severe, score = 3)	Significant pits, deep or high density (Fig. 5Q-T). Seriously obstruct the visual pathway.

- **Central mass.** The centre of the lens presents as an irregular shape protrusion, which is significantly different from the rest of the lens. The central lens has lost its smooth and high regular structure.

Catagory	Description
Central mass = No/ cataract = No	The centre of the lens may seem to bulge out compared to the rest of the lens, but it is smooth, uniform and has a regular shape (Fig. 6A-D). The outer lens is uniform and well developed. This could be an optical illusion caused by light or the different depths of the lens that had been captured during imaging.
Central mass = Yes/ cataract = Yes (mild, score = 1)	The central lens lost its round shape and become "flower-shape" (Fig. 6E). The outer lens is mostly uniform and well developed. The "flower-shape" is presumed to be the outline of the original, undifferentiated cells in the lens nucleus.
Central mass = Yes/ cataract = Yes (moderate, score = 2)	The central lens is abnormal, lost its round shape and smooth, uniform texture (Fig. 6F). In the lens nucleus, original, undifferentiated cells appear to be visible, meanwhile the outer lens is well developed.
Central mass = Yes/ cataract = Yes (severe, score = 3)	The central lens is significantly abnormal, completely lost its regular, smooth texture (Fig. 6G and H).

4. Any images with differing cataract classifications should then be assessed by a third researcher to break the tie, prior to unblinding.

NOTE:

- Frequently, regular concentric circles (Fig. 7A and B) or irregular "ring-like" (Fig. 7C-E) or "wave-like" structures (Fig. 7F-H) can be observed in the lenses with the central area uniform and smooth. They are believed to be maturing lens fiber cells rather than developmental cataracts, as we observed similar structures in the lens images of developing wildtype zebrafish that showed in the literature [29–33]. No cataract is called unless there is obvious disruption to make the "flower-shape" that is highlighted in the central mass category.

5. Discussion

Gene identification for inherited paediatric cataract continues and has accelerated in recent years with the advent of whole exome and

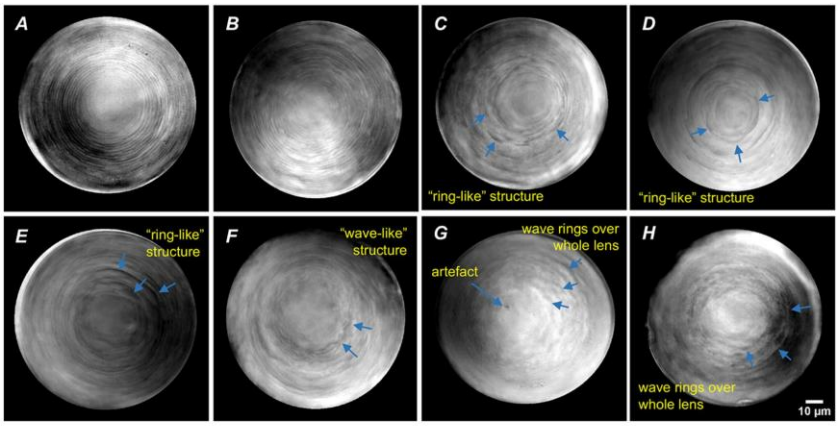


Fig. 7. Normal developing lens images with maturing lens fiber cells.

genome sequencing to complement family-based linkage approaches. However, functional validation of these genes remains a bottleneck in the discovery pipeline. The rapid development of genome engineering techniques, particularly CRISPR-Cas9, has greatly promoted reverse genetic screening, making routine functional assessment possible. Here we provide a step-by-step protocol as well as hands-on tips for rapid and highly efficient evaluation of cataract candidate genes in F0 zebrafish using a CRISPR-Cas9 RNP strategy. We generated zebrafish cataract assessment criteria, which to our knowledge, are the first systematic, definitive cataract assessment criteria for the zebrafish model. This protocol can be used to develop zebrafish models for in-depth study of specific genes, or to rapidly assess the cataract forming potential of multiple genes. With gene discovery in human patients and their families routinely employing genome-wide approaches, it is not uncommon to identify several plausible candidate genes. Ideally, similar variants in those genes would be found in additional affected families, however, that process can take many years until just the right family makes it into a research or diagnostic program. The use of a rapid zebrafish-based functional validation protocol can provide valuable evidence for the role of candidate genes in cataract formation.

A major concern of the F0-based gene functional evaluation is the genetic mosaicism in the injected fish. The variety of unpredictable indels induced by CRISPR-Cas9 complicate phenotypic analysis. However, some recent studies have demonstrated that well-designed CRISPR-Cas9 with mutagenesis efficiency approaching saturation can induce almost fully penetrant loss-of-function phenotypes with low off-targets, faithfully simulating the null mutants in F0 embryos [21,23,35,36]. Several critical strategies were applied in our protocol to maximize mutagenesis efficiency. Firstly, we provide comprehensive gRNA design instructions and detailed injection condition optimization pipeline to obtain appropriately designed guides and optimal injections, thus lay the foundation for high mutagenesis efficiency. In our hands, five of the eight tested gRNAs had an editing efficiency of over 80%, with the maximum editing efficiency reached 88.4%. Secondly, by designing gRNAs targeting multiple sites of a single gene and injecting these gRNAs simultaneously, the mutagenicity is further improved with the increase of the probability of including efficient guides and loss-of-function deletions. Finally, we applied CRISPR-Cas9 RNP, which is more mutagenic compare to the CRISPR-Cas9 mRNA complex [18,21], to achieve saturating mutagenesis. After successful evaluation following this protocol, any promising cataract candidates can be further confirmed in stable germline-transmitted mutant lines.

Another vital factor for successful cataract candidate gene evaluation is the accurate identification of cataract phenotypes in the fish lens. Because of its simplicity and directness, lens images of live embryos are widely used for phenotypic assessment of zebrafish cataract models. However, there is a lack of standardized reference material available to assess zebrafish cataract development. Multiple factors can interfere with the cataract assessment. Developmental delays are frequently observed in CRISPR-Cas9 injected embryos, and the presence of developing lens fiber cells makes it difficult to distinguish real cataracts. A similar study reported that cataracts detected at 3dpf were absent at 4dpf [37]. It cannot be ruled out that the phenotypes observed are the manifestation of immature lenses and not developmental cataracts. Therefore, we suggest analysing the phenotype at 4dpf or even 5dpf, to ensure the lens is of sufficient maturity to obtain reliable phenotypes. Artefact is another confounding factor. Based on our experience, lesions with a very round and regular shape, highly reflective dots or pitting with blurred edges are considered artefacts. In addition, CRISPR injection itself is possible to lead to non-specific lens lesions. In this situation, the false positive phenotypes can be effectively eliminated by the use of a properly designed negative control such as *LacZ*. In our assessment criteria, we defined cataract in the early zebrafish larvae by characterizing defects that lead to gross changes in lens morphology observable by brightfield microscopy with further classification by severity. Examples of false positives are also given. Our use of single fish strain (AB) and a mild phenotype observed with our chosen candidate gene are the major limitations of our cataract assessment criteria to date. These should continue to be developed to apply to additional strains which may have different developmental rates and to account for more severe cataract that may be observed with different genes.

In summary, we report a simple CRISPR-Cas9 RNP based method for rapid and efficient evaluation of cataract candidate genes in F0 zebrafish. Compared with traditional functional analysis strategies that need to establish stable mutant lines, our protocol creates significant labour- and time-savings. In addition, we also generated detailed criteria to allow standardized assessment of the lens phenotype using brightfield imaging and provide reference images for accurate and rigorous phenotypic analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Elise Yeaman for technical support and maintenance of zebrafish lines.

Funding

This work was supported by the National Health and Medical Research Council (NHMRC) of Australia [grants GNT1185477, GNT1059954 and GNT1116360]. Duran Zhao was supported by a Health Tasmania Graduate Research Scholarship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2020.12.004.

References

[1] S.R. Flaxman, R.R. Bourne, S. Resnikoff, P. Ackland, T. Braithwaite, M.V. Cicinelli, et al., Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis, Lancet Global Health 5 (12) (2017) e1221–e1234.
[2] L. Zhao, X.-J. Chen, J. Zhu, Y.-B. Xi, X. Yang, L.-D. Hu, et al., Lanosterol reverses protein aggregation in cataracts, Nature 523 (7562) (2015) 607–611.
[3] J.F. Hejtmancik, Congenital cataracts and their molecular genetics, Semin. Cell Dev. Biol. 19 (2008) 134–149.
[4] A. Shiels, T.M. Bennett, J.F. Hejtmancik, Cat-Map: putting cataract on the map, Mol. Vision 16 (2010) 2007.
[5] A. Shiels, J.F. Hejtmancik, Mutations and mechanisms in congenital and age-related cataracts, Exp. Eye Res. 156 (2017) 95–102.
[6] R.L. Gillespie, J. O'Sullivan, J. Ashworth, S. Bhaskar, S. Williams, S. Biswas, et al., Personalized diagnosis and management of congenital cataract by next-generation sequencing, Ophthalmology 121 (11) (2014), 2124–2137.e2.
[7] S. Javadiyan, J.E. Craig, E. Souzeau, S. Sharma, K.M. Lower, D.A. Mackey, et al., High-throughput genetic screening of 51 pediatric cataract genes identifies causative mutations in inherited pediatric cataract in South Eastern Australia, G3: Genes, Genomes, Genetics 7 (10) (2017) 3257–3268.
[8] A.S. Ma, J.R. Grigg, G. Ho, I. Prokudin, E. Farnsworth, K. Holman, et al., Sporadic and familial congenital cataracts: Mutational spectrum and new diagnoses using next-generation sequencing, Hum. Mutat. 37 (4) (2016) 371–384.
[9] P. Goldsmith, Modelling eye diseases in zebrafish, NeuroReport 12 (13) (2001) A73–A77.
[10] G.J. Lieschke, P.D. Currie, Animal models of human disease: zebrafish swim into view, Nat. Rev. Genet. 8 (5) (2007) 353.
[11] K. Howe, M.D. Clark, C.F. Torroja, J. Torrance, C. Berthelot, M. Muffato, et al., The zebrafish reference genome sequence and its relationship to the human genome, Nature 496 (7446) (2013) 498–503.
[12] W.Y. Hwang, Y. Fu, D. Reyon, M.L. Maeder, S.Q. Tsai, J.D. Sander, et al., Efficient genome editing in zebrafish using a CRISPR-Cas system, Nat. Biotechnol. 31 (3) (2013) 227.
[13] N. Chang, C. Sun, L. Gao, D. Zhu, X. Xu, X. Zhu, et al., Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos, Cell Res. 23 (4) (2013) 465–472.
[14] P.D. Hsu, E.S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering, Cell 157 (6) (2014) 1262–1278.
[15] S. Vivekanandan, M.F. Lou, Evidence for the presence of phosphoinositide cycle and its involvement in cellular signal transduction in the rabbit lens, Curr. Eye Res. 8 (1) (1989) 101–111.
[16] R. Boerrigter, J. Siertsema, I. Kema, Serotonin (5-HT) and the rat's eye, Doc. Ophthalmol. 82 (1–2) (1992) 141–150.
[17] M. Westerfield, The zebrafish book: a guide for the laboratory use of zebrafish, 2000 http://zfin.org/zf_info/zfbook/zfbk.html.
[18] J.A. Gagnon, E. Valen, S.B. Thyme, P. Huang, L. Ahkmetova, A. Pauli, et al., Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs, PLoS ONE 9 (5) (2014), e98186.
[19] S. Lessard, L. Francioli, J. Alfoldi, J.C. Tardif, P.T. Ellinor, D.G. MacArthur, et al., Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci, Proc. Natl. Acad. Sci. U S A 114 (52) (2017) E11257–E11266.
[20] S. Chen, S. Sun, D. Moonen, C. Lee, A.-Y.-F. Lee, D.V. Schaffer, et al., CRISPR-READI: Efficient generation of knockin mice by CRISPR RNP electroporation and AAV donor infection, Cell Rep. 27 (13) (2019).
[21] A. Burger, H. Lindsay, A. Felker, C. Hess, C. Anders, E. Chiavacci, et al., Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes, Development 143 (11) (2016) 2025–2037.
[22] B.P. Kleinstiver, V. Pattanayak, M.S. Prew, S.Q. Tsai, N.T. Nguyen, Z. Zheng, et al., High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects, Nature 529 (7587) (2016) 490–495.
[23] R.S. Wu, I.I. Lam, H. Clay, D.N. Duong, R.C. Deo, S.R. Coughlin, A rapid method for directed gene knockout for screening in G0 Zebrafish, Dev. Cell 46 (1) (2018) 112–25 e4.
[24] M.C. Huang, W.C. Cheong, L.S. Lim, M.H. Li, A simple, high sensitivity mutation screening using Ampligase mediated T7 endonuclease I and Surveyor nuclease with microfluidic capillary electrophoresis, Electrophoresis 33 (5) (2012) 788–796.
[25] H.R. Thomas, S.M. Percival, B.K. Yoder, J.M. Parant, High-throughput genome editing and phenotyping facilitated by high resolution melting curve analysis, PLoS ONE 9 (12) (2014), e114632.
[26] S. Ota, Y. Hisano, M. Muraki, K. Hoshijima, T.J. Dahlem, D.J. Grunwald, et al., Efficient identification of TALEN-mediated genome modifications using heteroduplex mobility assays, Genes Cells 18 (6) (2013) 450–458.
[27] G.K. Varshney, B. Carrington, W. Pei, K. Bishop, Z. Chen, C. Fan, et al., A high-throughput functional genomics workflow based on CRISPR/Cas9-mediated targeted mutagenesis in zebrafish, Nat. Protoc. 11 (12) (2016) 2357–2375.
[28] A.N. Shah, C.F. Davey, A.C. Whitebirch, A.C. Miller, C.B. Moens, Rapid reverse genetic screening using CRISPR in zebrafish, Nat. Methods 12 (6) (2015) 535–540.
[29] T.M. Greiling, J.I. Clark, Early lens development in the zebrafish: a three-dimensional time-lapse analysis, Dev. Dyn.: An Official Publication of the Am. Assoc. Anatomists 238 (9) (2009) 2254–2265.
[30] R. Dahm, H.B. Schonthaler, A.S. Soehn, J. Van Marle, G.F. Vrensen, Development and adult morphology of the eye lens in the zebrafish, Exp. Eye Res. 85 (1) (2007) 74–89.
[31] T.M. Greiling, J.I. Clark, New insights into the mechanism of lens development using zebra fish. International review of cell and molecular biology, Elsevier (2012) 1–61.
[32] N. Gath, J.M. Gross, Zebrafish mab21l2 mutants possess severe defects in optic cup morphogenesis, lens and cornea development, Dev. Dyn. 248 (7) (2019) 514–529.
[33] K. Taler, O. Weiss, S. Rotem-Bamberger, A.M. Rubinstein, P. Seritrakul, J.M. Gross, et al., Lysyl hydroxylase 3 is required for normal lens capsule formation and maintenance of lens epithelium integrity and fate, Dev. Biol. 458 (2) (2020) 177–188.
[34] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, et al., Fiji: an open-source platform for biological-image analysis, Nat. Methods 9 (7) (2012) 676–682.
[35] C.J. Watson, A.T. Monstad-Rios, R.M. Bhimani, C. Gistelinck, A. Willaert, P. Coucke, et al., Phenomics-Based Quantification of CRISPR-Induced Mosaicism in Zebrafish, Cell Syst. (2020).
[36] K. Hoshijima, M.J. Jurynec, D.K. Shaw, A.M. Jacobi, M.A. Behlke, D.J. Grunwald, Highly efficient CRISPR-Cas9-based methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish, Dev. Cell 51 (5) (2019).
[37] I. Vorontsova, I. Gehring, J.E. Hall, T.F. Schilling, Aqp0a Regulates Suture Stability in the Zebrafish Lens, Invest. Ophthalmol. Vis. Sci. 59 (7) (2018) 2869–2879.

Appendix 3 Script for Miseq paired-end data trimming

```
#This script automates running trimmomatic for  multiple
fastq.gz files


for f1 in *_R1_001.fastq.gz

do


f2=${f1%%_R1_001.fastq.gz}"_R2_001.fastq.gz"


java -jar trimmomatic-0.39.jar PE  -phred33  -trimlog
trimlog.txt  $f1  $f2  paired_$f1  unpaired_$f1
paired_$f2  unpaired_$f2  ILLUMINACLIP:NexteraPE-
PE.fa:2:30:10  LEADING:3  TRAILING:3  SLIDINGWINDOW:4:15
MINLEN:36


done
```

Appendix 4 Commands for analyzing genome editing outcomes in CRISPResso2

***htr1fa*:**

```
CRISPRessoBatch --batch_settings CRISPResso2_htr1fa.bat
-a
TACTGACAACAGCCATGAACTGCCTGGTCATCACAGCCATCATTGTCACACGCAAA
CTCCACCATCCCTCAAACTACCTGATCTGCTCACTAGCCGTCACCGACCTCTTGGT
TGCCATTTTAGTCATGCCCTTCAGCATCATCTACATCGCAAAAGACACGTGGCTCA
TTGGTGAAGCATTGTGTAAATTCTGGCTAAGTGTGGATATAACCTGCTGCACGTGC
TCCATTTTACACTTGGCGGCCATCGCCGTGGACCGCTACCGGGCTATAACAGATGC
TGTGGAGTATTCCAGAAAAAGGACATCTTTGCGGGCTGCCATTATGATTAGCGTAG
TTTGGCTTCTCGCTATAGTCATCTCACTTCCACCAATACTTTTGAGAAATGGAGAG
GAAAATGAATGTATTATTGTACATACAAACATTGCATCCATGCTT -an htr1fa
-g TCTACATCGCAAAAGACACG,CATCTGTTATAGCCCGGTAG -gn
sgRNA4,sgRNA3 -p 96 -bo EditingEfficiencyAnalysis_htr1fa
--write_cleaned_report --default_min_aln_score 60 -q 0 -
s 0 --min_bp_quality_or_N 0 --exclude_bp_from_left 15 --
exclude_bp_from_right 15 --min_paired_end_reads_overlap
10 --max_paired_end_reads_overlap 100 -w 1 -wc -3 --
ignore_substitutions
```

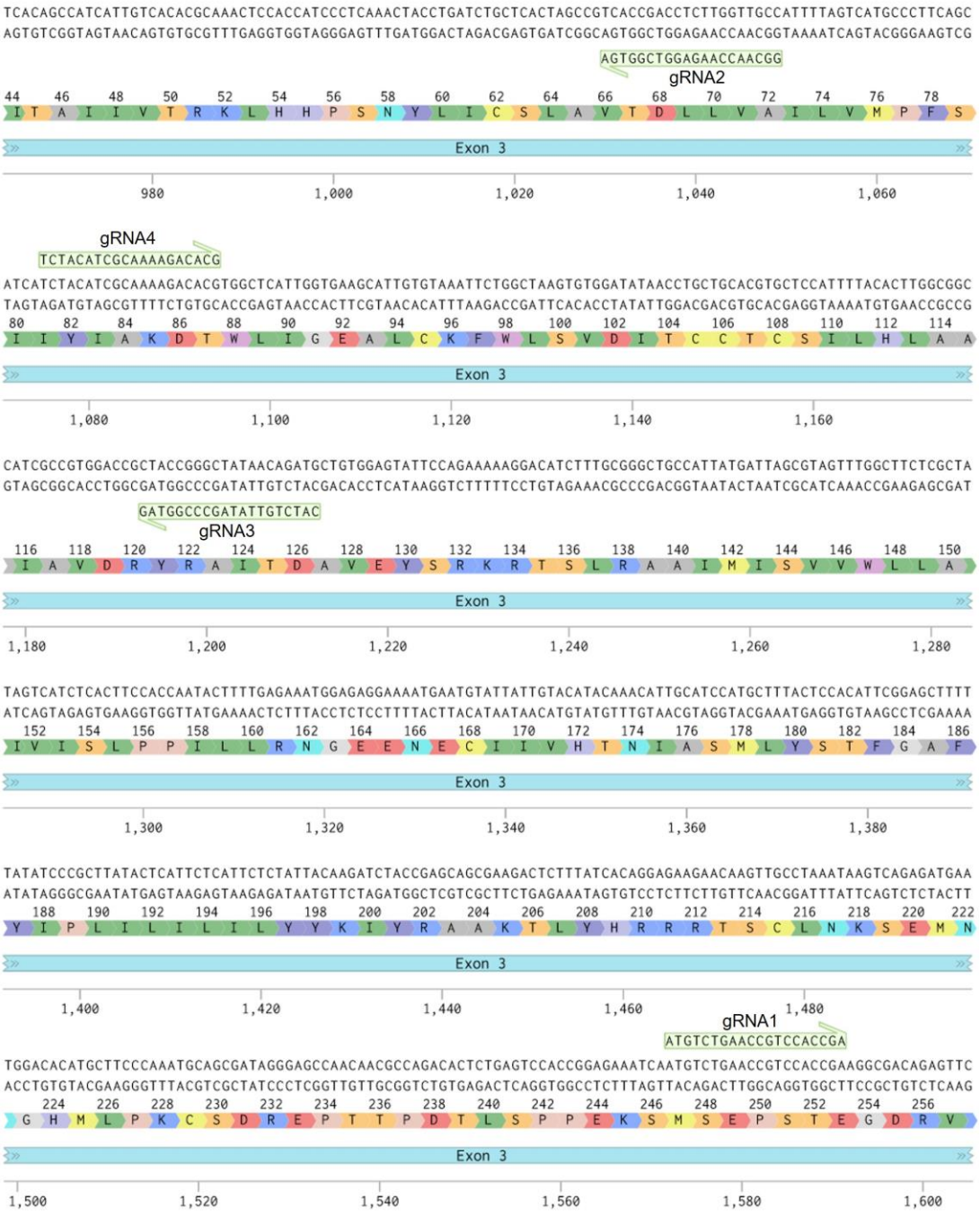
***htr1fb*:**

```
CRISPRessoBatch --batch_settings CRISPResso2_htr1fb.bat
-a
ATTGCAGAAGAGACTTGGGTCCTCGGCCCCATTGTTTGCCACCTGTGGTTAGGTGT
CGATGTTACATGCTGCACCTGCTCAATTTTACACCTCGCCGCCATTGCGCTGGATC
GATACCGTGCCATCACTGATGCTGTGGCGTATTCGCAAAAACGCACATATAAAAGA
GTGATTGTAACCATTTTGTCTCTGTGGACACTTTCCATTCTGGTGTCCCTTCCACC
TTTAGTATGGAGGAAATTCCCTAAAGTAGAATTTAAGGACGGGAAGAGGGAACCCA
TGGACTGTTTGATTGAACACGACCATGTGGCTTTTACCGTCTACTCAACTTTTGGA
GCATTTTACATTCCTTTGGCACTCATTTTGGTTCTTTACTACAAGATCTACAAGGC
```

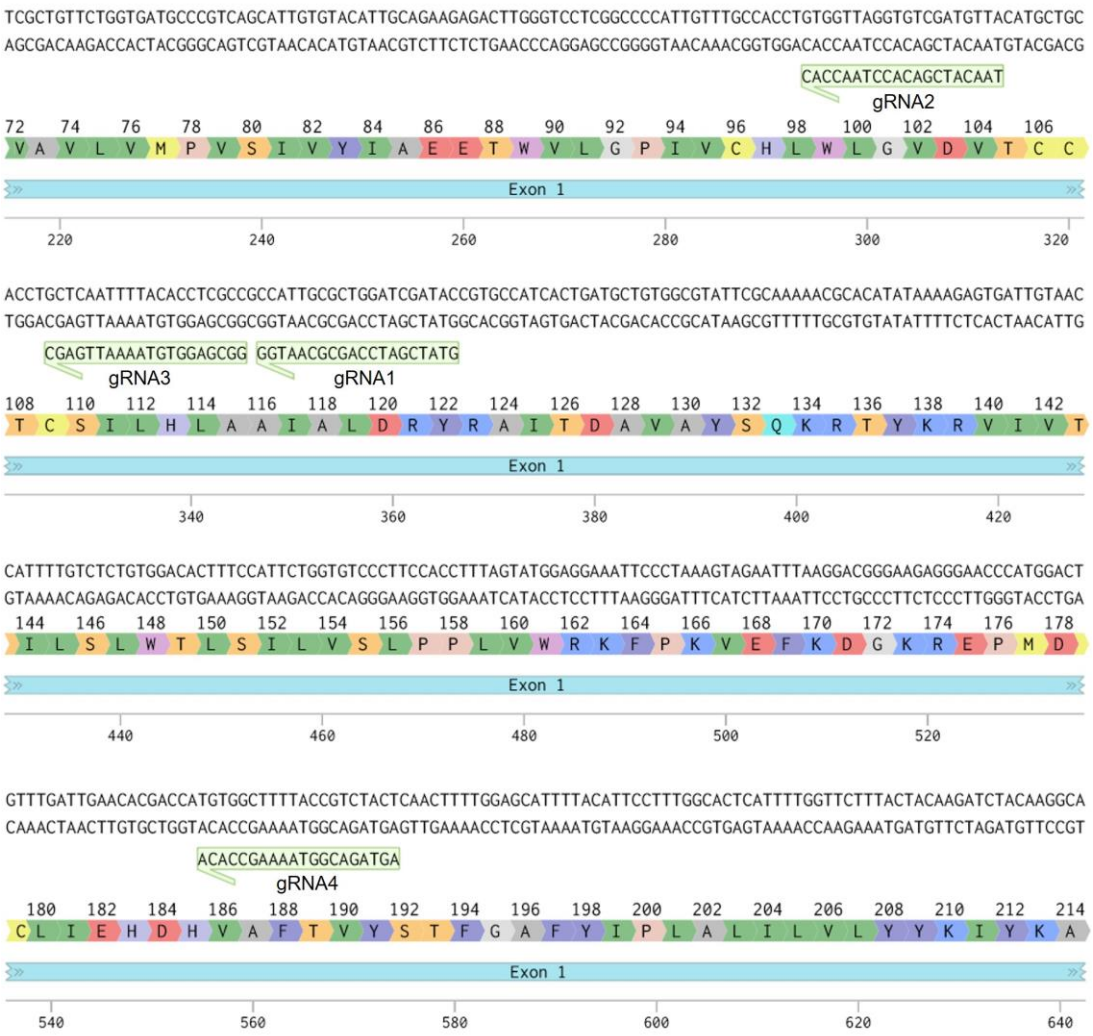
```
AGCGGAGATGCTTCGTAACCGTA -an htr1fb -g
GTATCGATCCAGCGCAATGG,AGTAGACGGTAAAAGCCACA -gn
sgRNA1,sgRNA4 -p 96 --batch_output_folder
EditingEfficiencyAnalysis_htr1fb --write_cleaned_report
--default_min_aln_score 60 -q 0 -s 0 --
min_bp_quality_or_N 0 --exclude_bp_from_left 15 --
exclude_bp_from_right 15 --min_paired_end_reads_overlap
10 --max_paired_end_reads_overlap 100 -w 1 -wc -3 --
ignore_substitutions
```

Appendix 5 Targeting sequence of htr1f gRNAs

***htr1fa*:**



***htr1fb*:**

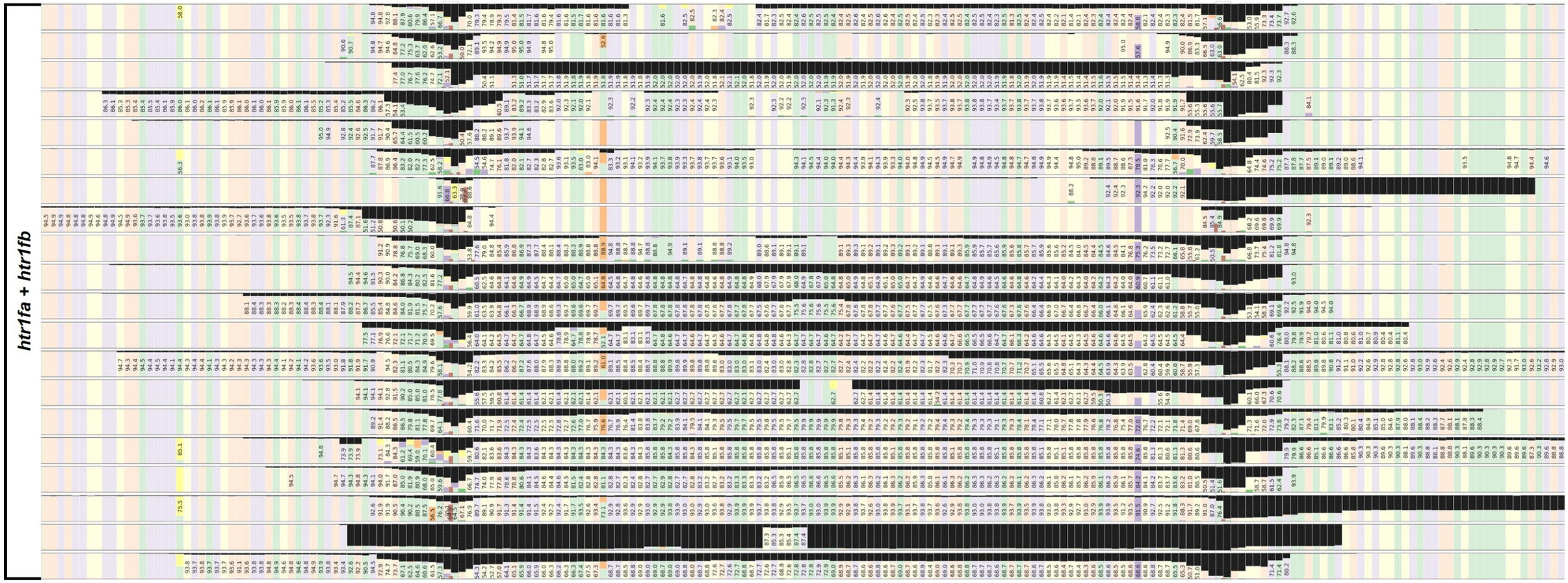


The figures are adapted from the export files of Benchling (https://benchling.com)

Appendix 6 Indel position of the *htr1fa* knockout experiments

(The figure is displayed in two parts due to the large size. Part A is on this page, Part B and figure legend are on the next page)

Part A



Part B

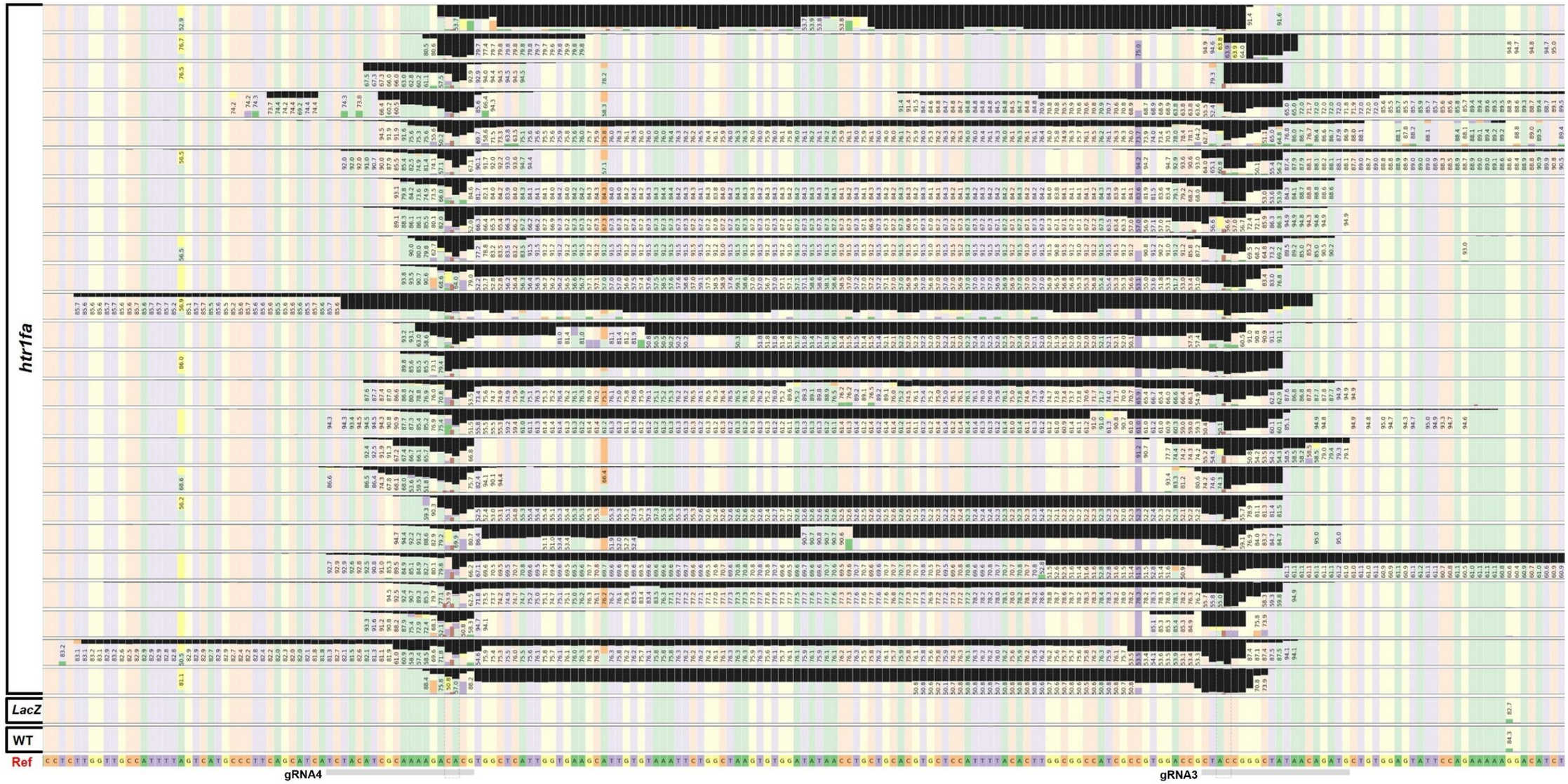
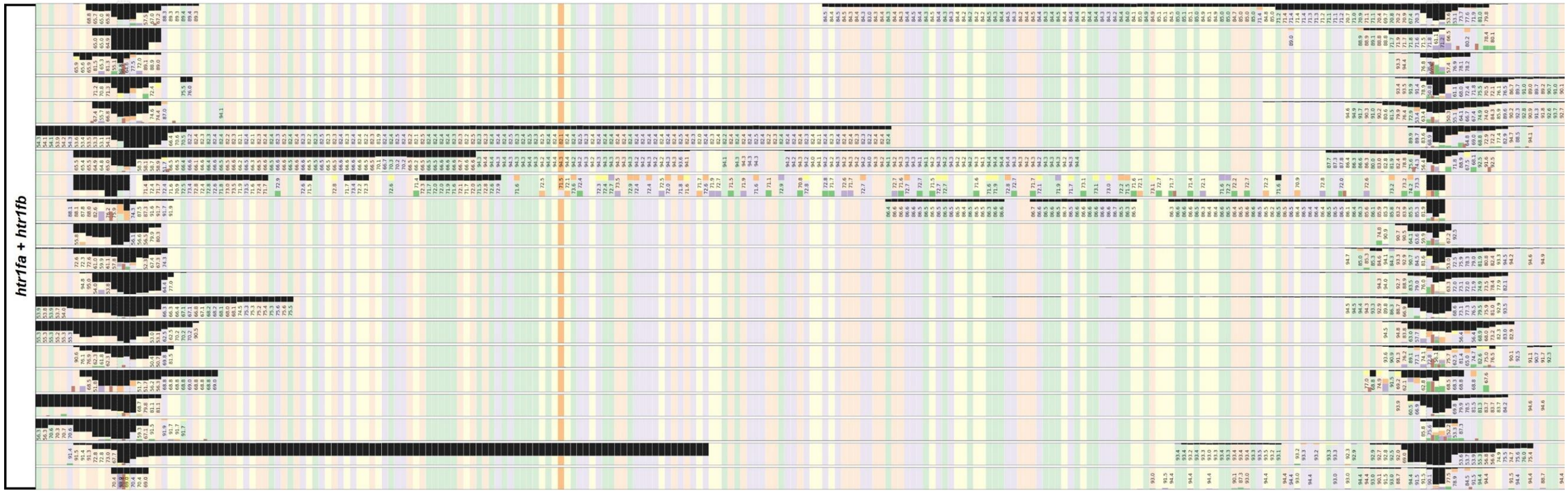


Figure illustrates the indel (insertion or deletion) distribution across amplicon. Reference sequence and gRNAs are shown at the bottom of the figure. At each base in the reference amplicon, the percentage of each base as observed in sequencing reads is shown. Black boxes on the left side of the sequences indicate the groups of the samples. WT: wild-type. Ref: reference. Figure is adapted from the export file of CRISPResso2[159].

Appendix 7 Indel position of the *htr1fb* knockout experiments

(The figure is displayed in two parts due to the large size. Part A is on this page, Part B and figure legend are on the next page)

Part A



Part B

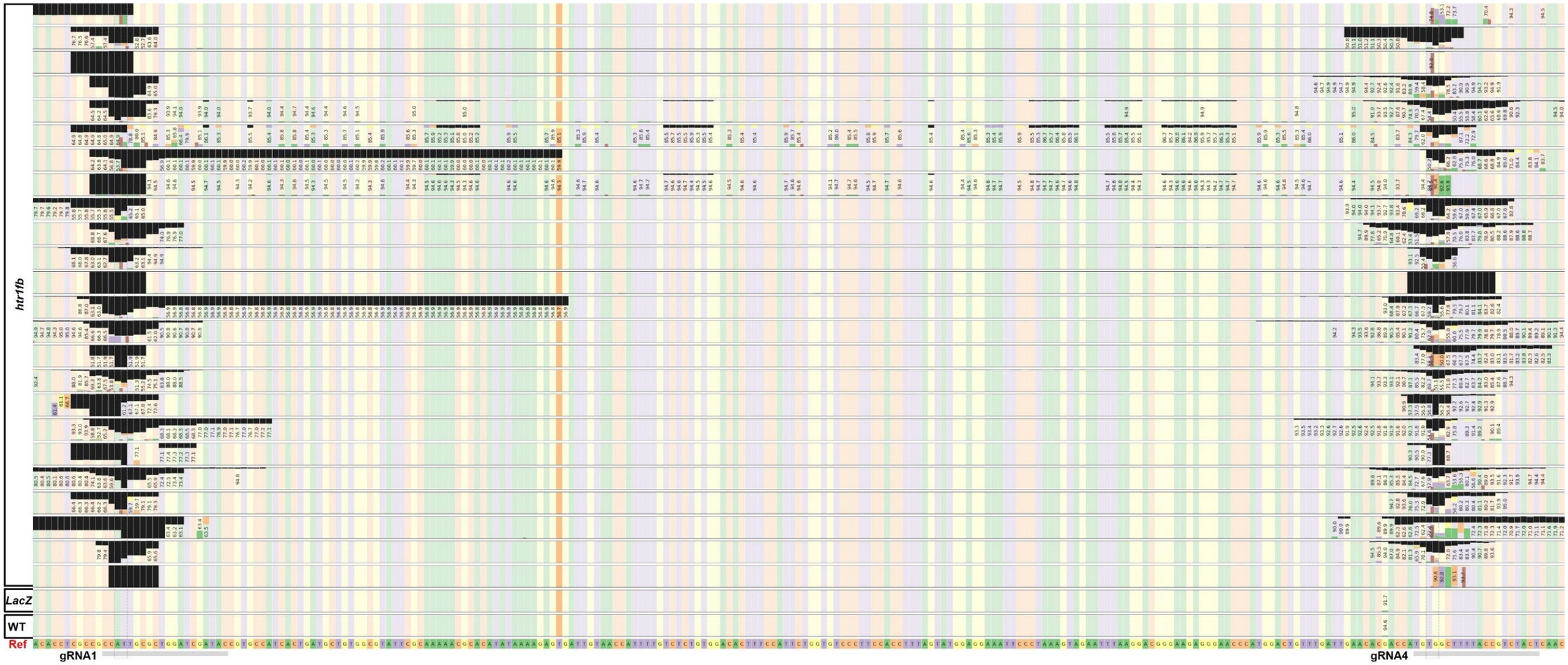


Figure illustrates the indel (insertion or deletion) distribution across amplicon. Reference sequence and gRNAs are shown at the bottom of the figure. At each base in the reference amplicon, the percentage of each base as observed in sequencing reads is shown. Black boxes on the left side of the sequences indicate the groups of the samples. WT: wild-type. Ref: reference. Figure is adapted from the export file of CRISPResso2[159].

Appendix 8 R script to convert Illumina final report to PLINK long-format files

```
## Originally written by Ryan Schubert for the Wheeler Lab,
github@RyanSchu Lab github@wheelerlab


#load library
library(dplyr)

#GenomeStudio FinalReport.txt to lgen and map file:
Finalreport<-as.data.frame(read.table("FinalReport.txt",
sep='\t', skip = 9, header = T))
Finalreport<-filter(Finalreport, Allele1...Forward !=
'I')
Finalreport["empty"]="0"
Finalreport['fid']<-Finalreport$Sample.Group
map<-select(Finalreport, Chr, SNP.Name, empty, Position)
map<-map[!duplicated(map),]
map<-map[complete.cases(map),]
lgen<-select(Finalreport, fid, Sample.ID, SNP.Name,
Allele1...Forward, Allele2...Forward)
lgen<-lgen[!duplicated(lgen),]
lgen<-lgen[complete.cases(lgen),]
lgen<-filter(lgen, Allele1...Forward != "-" &
Allele2...Forward != "-")

write.table(map, " output_file.map",sep = "\t",
col.names = F, row.names = F, quote = F)
write.table(lgen, "output_file.lgen", sep = "\t",
col.names = F, row.names = F, quote = F)
```

```
#SampleID_Gender.csv to fam file:
sex<-
as.data.frame(read.table(file="SampleID_Gender.csv",
sep=',', header = T))
sex["empty"]<-"0"
sex["empty1"]<-"0"
sex["empty2"]<-"0"
sex['fid']<-sex$Sample.Group
sex<-select(sex, fid, Sample.ID, empty, empty1, Gender,
empty2)
sex$Gender<-gsub('Female', '2', sex$Gender, fixed = T)
sex$Gender<-gsub('Male', '1', sex$Gender, fixed = T)

write.table(sex, file = "output_file.fam", sep = "\t",
col.names = F, row.names = F, quote = F)
```

Appendix 9 R script to categorise whole-genome sequencing data by sequencing quality

```
## Originally written by Dr. Nicholas Blackburn (Menzies
Institute for Medical Research, University of Tasmania.
This script requires Java and GATK programs.

#index the input file
tabix -p vcf input_file.vcf

#categorise the variants
gatk VariantFiltration -R
/data/Genome_Reference/hg19/hg19.fa -O output_file.vcf -
variant input_file.vcf --genotype-filter-expression
"DP >= 10 && GQ >= 20" --genotype-filter-name
"High_Confidence" --genotype-filter-expression "DP < 10
&& GQ < 20" --genotype-filter-name "Low_Confidence" --
genotype-filter-expression "DP < 10 && GQ >= 20" --
genotype-filter-name "Low_coverage_High_quality" --
genotype-filter-expression "DP >= 10 && GQ < 20" --
genotype-filter-name "High_coverage_Low_quality
```

Appendix 10 R script to annotate the whole-genome sequencing data

```
table_annovar.pl input_file.vcf /data/annovar-db/ -
buildver hg19 -out output_file -remove -otherinfo -
protocol
refGene,ucscGenePfam,cytoBand,fathmm,cadd13,gnomad211_ge
nome,eigen,avsnp150,popfreq_max_20150413,1000g2015aug_eu
r,esp6500siv2_all,esp6500siv2_ea,esp6500siv2_aa,exac03,e
xac03nontcga,exac03nonpsych,kaviar_20150923,hrcr1,dbscsn
v11,UK10K_All_AF,clinvar_20200316,dbnsfp35a,gwava -
operation g,r,r,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f
-nastring . -vcfinput
```

Appendix 11 The kinship of family CSA91 and family CSA110

Individual1	Individual2	Kinship Coefficient	Degree
CSA91:CSA91.01	CSA110:CSA110.02	0.003139	-1
CSA91:CSA91.01	CSA110:CSA110.05	0.006885	7
CSA91:CSA91.01	CSA110:CSA110.07	0.007541	6
CSA91:CSA91.01	CSA110:CSA110.04	0.00698	6
CSA91:CSA91.01	CSA110:CSA110.03	0.0047	7
CSA91:CSA91.01	CSA110:CSA110.01	0.003816	-1
CSA91:CSA91.02	CSA110:CSA110.02	0.006829	7
CSA91:CSA91.02	CSA110:CSA110.05	0.008251	6
CSA91:CSA91.02	CSA110:CSA110.07	0.007269	6
CSA91:CSA91.02	CSA110:CSA110.04	0.006985	6
CSA91:CSA91.02	CSA110:CSA110.03	0.006854	7
CSA91:CSA91.02	CSA110:CSA110.01	0.008864	6
CSA91:CSA91.03	CSA110:CSA110.02	0.007537	6
CSA91:CSA91.03	CSA110:CSA110.05	0.003346	-1
CSA91:CSA91.03	CSA110:CSA110.07	0.003773	-1
CSA91:CSA91.03	CSA110:CSA110.04	0.004413	7
CSA91:CSA91.03	CSA110:CSA110.03	0.004426	7
CSA91:CSA91.03	CSA110:CSA110.01	0.006822	7
CSA91:CSA91.04	CSA110:CSA110.02	0.007529	6
CSA91:CSA91.04	CSA110:CSA110.05	0.007841	6
CSA91:CSA91.04	CSA110:CSA110.07	0.006221	7
CSA91:CSA91.04	CSA110:CSA110.04	0.007157	6
CSA91:CSA91.04	CSA110:CSA110.03	0.005247	7

Individual1	Individual2	Kinship Coefficient	Degree
CSA91:CSA91.04	CSA110:CSA110.01	0.006063	7
CSA91:CSA91.05	CSA110:CSA110.02	0.006771	7
CSA91:CSA91.05	CSA110:CSA110.05	0.005149	7
CSA91:CSA91.05	CSA110:CSA110.07	0.004595	7
CSA91:CSA91.05	CSA110:CSA110.04	0.004341	7
CSA91:CSA91.05	CSA110:CSA110.03	0.003945	-1
CSA91:CSA91.05	CSA110:CSA110.01	0.004392	7
CSA91:CSA91.06	CSA110:CSA110.02	0.004049	-1
CSA91:CSA91.06	CSA110:CSA110.05	0.007852	6
CSA91:CSA91.06	CSA110:CSA110.07	0.00457	7
CSA91:CSA91.06	CSA110:CSA110.04	0.001896	-1
CSA91:CSA91.06	CSA110:CSA110.03	0.004963	7
CSA91:CSA91.06	CSA110:CSA110.01	0.004443	7
CSA91:CSA91.07	CSA110:CSA110.02	0.002663	-1
CSA91:CSA91.07	CSA110:CSA110.05	0.007821	6
CSA91:CSA91.07	CSA110:CSA110.07	0.002748	-1
CSA91:CSA91.07	CSA110:CSA110.04	0.003116	-1
CSA91:CSA91.07	CSA110:CSA110.03	0.004952	7
CSA91:CSA91.07	CSA110:CSA110.01	0.001914	-1
CSA91:CSA91.08	CSA110:CSA110.02	0.003353	-1
CSA91:CSA91.08	CSA110:CSA110.05	0.007998	6
CSA91:CSA91.08	CSA110:CSA110.07	0.003003	-1
CSA91:CSA91.08	CSA110:CSA110.04	0.002494	-1
CSA91:CSA91.08	CSA110:CSA110.03	0.002876	-1
CSA91:CSA91.08	CSA110:CSA110.01	0.005498	7
CSA91:CSA91.09	CSA110:CSA110.02	0.003893	-1

Individual1	Individual2	Kinship Coefficient	Degree
CSA91:CSA91.09	CSA110:CSA110.05	0.006938	6
CSA91:CSA91.09	CSA110:CSA110.07	0.004688	7
CSA91:CSA91.09	CSA110:CSA110.04	0.00465	7
CSA91:CSA91.09	CSA110:CSA110.03	0.003228	-1
CSA91:CSA91.09	CSA110:CSA110.01	0.004541	7

The kinship was calculated using IBIS (Identical by Descent via Identical by State) v1.20.9[192] program. 1 = first degree, 2 = second degree and so on, -1 = unrelated

References

1. Levin LA, Kaufman PL, Alm A. Adler's Physiology of the Eye. 2011.

2. Cvekl A, Ashery-Padan R. The cellular and molecular mechanisms of vertebrate lens development. *Development.* 2014;141(23):4432-4447.

3. Fagerholm PP, Philipson BT, Lindström B. Normal human lens—the distribution of protein. *Experimental eye research.* 1981;33(6):615-620.

4. Grainger RM, Mannion JE, Cook Jr TL, Zygar CA. Defining intermediate stages in cell determination: acquisition of a lens - forming bias in head ectoderm during lens determination. *Developmental genetics.* 1997;20(3):246-257.

5. Snell RS, Lemp MA. *Clinical anatomy of the eye.* John Wiley & Sons; 2013.

6. Schook P. A review of data on cell actions and cell interaction during the morphogenesis of the embryonic eye. *Acta Morphologica Neerlando-scandinavica.* 1978;16(4):267-286.

7. McAvoy J. Developmental biology of the lens. *Mechanisms of cataract formation in the human lens.* 1981:7-46.

8. Wu X, Long E, Lin H, Liu Y. Prevalence and epidemiological characteristics of congenital cataract: a systematic review and meta-analysis. *Scientific reports.* 2016;6(1):1-10.

9. Khokhar SK, Pillay G, Dhull C, Agarwal E, Mahabir M, Aggarwal P. Pediatric cataract. *Indian journal of ophthalmology.* 2017;65(12):1340.

10. Wirth MG, Russell-Eggitt IM, Craig JE, Elder JE, Mackey DA. Aetiology of congenital and paediatric cataract in an Australian population. *Br J Ophthalmol.* 2002;86(7):782-786.

11. Liu Y-C, Wilkins M, Kim T, Malyugin B, Mehta JS. Cataracts. *The Lancet.* 2017;390(10094):600-612.

12. Churchill A, Graw J. Clinical and experimental advances in congenital and paediatric cataracts. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 2011;366(1568):1234-1249.

13. Stager DR, Felius J, Beauchamp GR. Congenital cataract cost. *Ophthalmology.* 2009;116(12):2484-2484. e2482.

14. Wilson M, Pandey S, Thakur J. Paediatric cataract blindness in the developing world: surgical techniques and intraocular lenses in the new millennium. *British Journal of Ophthalmology.* 2003;87(1):14-19.

15. Yi J, Yun J, Li Z-K, Xu C-T, Pan B-R. Epidemiology and molecular genetics of congenital cataracts. *International journal of ophthalmology.* 2011;4(4):422.

16. Srivastava RM, Agrawal S. Etiology of Pediatric Cataract. In: *Pediatric Cataract.* Springer; 2021:37-55.

17. He W, Li S. Congenital cataracts: gene mapping. *Human genetics.* 2000;106(1):1-13.

18. Messina-Baas OM, Gonzalez-Huerta LM, Cuevas-Covarrubias SA. Two affected siblings with nuclear cataract associated with a novel missense mutation in the CRYGD gene. *Mol Vis.* 2006;12(995):1000.

19. Chen W. Classification and Morphology of Pediatric Cataracts. In: *Pediatric Lens Diseases.* Springer; 2017:69-84.

20. Kalantan H. Posterior polar cataract: A review. *Saudi Journal of Ophthalmology.* 2012;26(1):41-49.

21. Khan AO, Aldahmesh MA, Mohamed JY, Alkuraya FS. Clinical and molecular analysis of children with central pulverulent cataract from the Arabian Peninsula. *British journal of ophthalmology.* 2012;96(5):650-655.

22. Haargaard B, Wohlfahrt J, Fledelius HC, Rosenberg T, Melbye M. A nationwide Danish study of 1027 cases of congenital/infantile cataracts: etiological and clinical classifications. *Ophthalmology.* 2004;111(12):2292-2298.

23. Shiels A, Hejtmancik JF. Genetics of human cataract. *Clin Genet.* 2013;84(2):120-127.

24. Reis LM, Semina EV. Genetic landscape of isolated pediatric cataracts: extreme heterogeneity and variable inheritance patterns within genes. *Human genetics.* 2019;138(8):847-863.

25. Francis PJ, Moore AT. Genetics of childhood cataract. *Current opinion in ophthalmology.* 2004;15(1):10-15.

26. Graw J, Klopp N, Illig T, Preising MN, Lorenz B. Congenital cataract and macular hypoplasia in humans associated with a de novo mutation in CRYAA and compound heterozygous mutations in P. *Graefe's Archive for Clinical and Experimental Ophthalmology.* 2006;244(8):912-919.

27. Mohebi M, Akbari A, Babaei N, Sadeghi A, Heidari M. Identification of a de novo 3bp deletion in CRYBA1/A3 gene in autosomal dominant congenital cataract. *Acta Medica Iranica.* 2016:778-783.

28. Redwood A, Douzgou S, Waller S, et al. Congenital cataracts in females caused by BCOR mutations; report of six further families demonstrating clinical variability and diverse genetic mechanisms. *European journal of medical genetics.* 2020;63(2):103658.

29. Solebo AL, Hammond CJ, Rahi JS. Improving outcomes in congenital cataract. *Nature.* 2018;556(7699):E1-E2.

30. Whitman MC, Vanderveen DK. Complications of pediatric cataract surgery. Paper presented at: Seminars in ophthalmology2014.

31. Shiels A, Bennett TM, Hejtmancik JF. Cat-Map: putting cataract on the map. *Molecular vision.* 2010;16:2007.

32. Zhu X, Zhang S, Chang R, Lu Y. New cataract markers: Mechanisms of disease. *Clin Chim Acta.* 2017;472:41-45.

33. Augusteyn RC. On the growth and internal structure of the human lens. *Experimental eye research.* 2010;90(6):643-654.

34. Hejtmancik JF, Shiels A. Overview of the Lens. *Progress in molecular biology and translational science.* 2015;134:119-127.

35. Pichi F, Lembo A, Serafino M, Nucci P. Genetics of Congenital Cataract. *Dev Ophthalmol.* 2016;57:1-14.

36. Tikhomirova T, Selivanova O, Galzitskaya O. α-Crystallins are small heat shock proteins: Functional and structural properties. *Biochemistry (Moscow).* 2017;82(2):106-121.

37. Basha E, O'Neill H, Vierling E. Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *Trends in biochemical sciences.* 2012;37(3):106-117.

38. Andley UP, Song Z, Wawrousek EF, Fleming TP, Bassnett S. Differential protective activity of αA-and αB-crystallin in lens epithelial cells. *Journal of Biological Chemistry.* 2000;275(47):36823-36831.

39. Andley UP. Crystallins in the eye: function and pathology. *Progress in retinal and eye research.* 2007;26(1):78-98.

40. Slingsby C, Wistow GJ. Functions of crystallins in and out of lens: roles in elongated and post-mitotic cells. *Progress in biophysics and molecular biology.* 2014;115(1):52-67.

41. Rocha MA, Sprague-Piercy MA, Kwok AO, Roskamp KW, Martin RW. Chemical Properties Determine Solubility and Stability in βγ-Crystallins of the Eye Lens. *ChemBioChem.* 2021;22(8):1329-1346.

42. Jaenicke R, Slingsby C. Lens crystallins and their microbial homologs: structure, stability, and function. *Critical Reviews in Biochemistry and Molecular Biology.* 2001;36(5):435-499.

43. Graw J. Genetics of crystallins: cataract and beyond. *Experimental eye research.* 2009;88(2):173-189.

44. Shiels A, Hejtmancik JF. Mutations and mechanisms in congenital and age-related cataracts. *Exp Eye Res.* 2017;156:95-102.

45. Li J, Xia Ch, Wang E, Yao K, Gong X. Screening, genetics, risk factors, and treatment of neonatal cataracts. *Birth defects research.* 2017;109(10):734-743.

46. Beyer EC, Ebihara L, Berthoud VM. Connexin mutants and cataracts. *Frontiers in pharmacology.* 2013;4:43.

47. Gao J, Wang H, Sun X, et al. The effects of age on lens transport. *Investigative ophthalmology & visual science.* 2013;54(12):7174-7187.

48. Agre P. Membrane water transport and aquaporins: looking back. In: Wiley Online Library; 2005.

49. Kumar GS, Kyle JW, Minogue PJ, et al. An MIP/AQP0 mutation with impaired trafficking and function underlies an autosomal dominant congenital lamellar cataract. *Experimental eye research.* 2013;110:136-141.

50. Zeng L, Liu W, Feng W, et al. A novel donor splice-site mutation of major intrinsic protein gene associated with congenital cataract in a Chinese family. *Molecular vision.* 2013;19:2244.

51. Shentu X, Miao Q, Tang X, Yin H, Zhao Y. Identification and functional analysis of a novel MIP gene mutation associated with congenital cataract in a Chinese family. *PloS one.* 2015;10(5):e0126679.

52. Jamieson RV, Farrar N, Stewart K, et al. 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. *Human mutation.* 2007;28(10):968-977.

53. Shiels A, Bennett TM, Knopf HL, et al. CHMP4B, a novel gene for autosomal dominant cataracts linked to chromosome 20q. *The American Journal of Human Genetics.* 2007;81(3):596-606.

54. Zhang XH, Da Wang J, Jia HY, et al. Mutation profiles of congenital cataract genes in 21 northern Chinese families. *Molecular vision.* 2018;24:471.

55. Kaul H, Riazuddin SA, Shahid M, et al. Autosomal recessive congenital cataract linked to EPHA2 in a consanguineous Pakistani family. *Molecular vision.* 2010;16:511.

56. Berry V, Pontikos N, Albarca-Aguilera M, et al. A recurrent splice-site mutation in EPHA2 causing congenital posterior nuclear cataract. *Ophthalmic genetics.* 2018;39(2):236-241.

57. Li D, Wang S, Ye H, et al. Distribution of gene mutations in sporadic congenital cataract in a Han Chinese population. *Molecular vision.* 2016;22:589.

58. Alizadeh A, Clark J, Seeberger T, Hess J, Blankenship T, FitzGerald PG. Targeted Deletion of the Lens Fiber Cell–Specific Intermediate Filament Protein Filensin. *Investigative ophthalmology & visual science.* 2003;44(12):5252-5258.

59. Alizadeh A, Clark JI, Seeberger T, et al. Targeted genomic deletion of the lens-specific intermediate filament protein CP49. *Investigative ophthalmology & visual science.* 2002;43(12):3722-3727.

60. Forshew T, Johnson CA, Khaliq S, et al. Locus heterogeneity in autosomal recessive congenital cataracts: linkage to 9q and germline HSF4 mutations. *Hum Genet.* 2005;117(5):452-459.

61. Fujimoto M, Izu H, Seki K, et al. HSF4 is required for normal cell growth and differentiation during mouse lens development. *The EMBO journal.* 2004;23(21):4297-4306.

62. Reza HM, Yasuda K. Roles of Maf family proteins in lens development. *Developmental dynamics: an official publication of the American Association of Anatomists.* 2004;229(3):440-448.

63. Bonini NM, Leiserson WM, Benzer S. The eyes absent gene: genetic control of cell survival and differentiation in the developing Drosophila eye. *Cell.* 1993;72(3):379-395.

64. Brémond-Gignac D, Bitoun P, Reis LM, Copin H, Murray JC, Semina EV. Identification of dominant FOXE3 and PAX6 mutations in patients with congenital cataract and aniridia. *Molecular vision.* 2010;16:1705.

65. Singh R, Ram J, Kaur G, Prasad R. Galactokinase deficiency induced cataracts in Indian infants: identification of 4 novel mutations in GALK gene. *Current eye research.* 2012;37(10):949-954.

66. Reich S, Hennermann J, Vetter B, et al. An unexpectedly high frequency of hypergalactosemia in an immigrant Bosnian population revealed by newborn screening. *Pediatric research.* 2002;51(5):598-601.

67. Kloeckener-Gruissem B, Vandekerckhove K, Nürnberg G, et al. Mutation of solute carrier SLC16A12 associates with a syndrome combining juvenile cataract with microcornea and renal glucosuria. *The American Journal of Human Genetics.* 2008;82(3):772-779.

68. GIRELLI D, OUVIERI O, Franceschi LD, Corrocher R, Bergamaschi G, CAZZOLA M. A linkage between hereditary hyperferritinaemia not related to iron overload and autosomal dominant congenital cataract. *British journal of haematology.* 1995;90(4):931-934.

69. Luscieti S, Tolle G, Aranda J, et al. Novel mutations in the ferritin-L iron-responsive element that only mildly impair IRP binding cause hereditary hyperferritinaemia cataract syndrome. *Orphanet journal of rare diseases.* 2013;8(1):1-10.

70. Cazzola M, Bergamaschi G, Tonon L, et al. Hereditary hyperferritinemia-cataract syndrome: relationship between phenotypes and specific mutations in the iron-responsive element of ferritin light-chain mRNA. *Blood, The Journal of the American Society of Hematology.* 1997;90(2):814-821.

71. Irum B, Khan SY, Ali M, et al. Deletion at the GCNT2 locus causes autosomal recessive congenital cataracts. *PloS one.* 2016;11(12):e0167562.

72. Inaba N, Hiruma T, Togayachi A, et al. A novel I-branching β-1, 6-N-acetylglucosaminyltransferase involved in human blood group I antigen expression. *Blood, The Journal of the American Society of Hematology.* 2003;101(7):2870-2876.

73. Pras E, Raz J, Yahalom V, et al. A nonsense mutation in the glucosaminyl (N-acetyl) transferase 2 gene (GCNT2): association with autosomal recessive congenital cataracts. *Investigative ophthalmology & visual science.* 2004;45(6):1940-1945.

74. Cheong S-S, Hull S, Jones B, et al. Pleiotropic effect of a novel mutation in GCNT2 causing congenital cataract and a rare adult i blood group phenotype. *Human genome variation.* 2017;4(1):1-5.

75. Ma AS, Grigg JR, Ho G, et al. Sporadic and familial congenital cataracts: Mutational spectrum and new diagnoses using next-generation sequencing. *Human mutation.* 2016;37(4):371-384.

76. Haer-Wigman L, van Zelst-Stams WA, Pfundt R, et al. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur J Hum Genet.* 2017;25(5):591-599.

77. Javadiyan S, Craig JE, Souzeau E, et al. High-throughput genetic screening of 51 pediatric cataract genes identifies causative mutations in inherited pediatric cataract in South Eastern Australia. *G3: Genes, Genomes, Genetics.* 2017;7(10):3257-3268.

78. Wu S-Y, Zou P, Mishra S, Mchaourab HS. Transgenic zebrafish models reveal distinct molecular mechanisms for cataract-linked αA-crystallin mutants. *Plos one.* 2018;13(11):e0207540.

79. Krall M, Htun S, Anand D, Hart D, Lachke S, Slavotinek A. A zebrafish model of foxe3 deficiency demonstrates lens and eye defects with dysregulation of key genes involved in cataract formation in humans. *Human genetics.* 2018:1-14.

80. Zhang J, Cui W-w, Du C, et al. Knockout of DNase1l1l abrogates lens denucleation process and causes cataract in zebrafish. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease.* 2020;1866(5):165724.

81. Gelatt KN, Das ND. Animal models for inherited cataracts: a review. *Current eye research.* 1984;3(5):765-778.

82. Jones JL, Corbett MA, Yeaman E, et al. A 127 kb truncating deletion of PGRMC1 is a novel cause of X-linked isolated paediatric cataract. *European Journal of Human Genetics.* 2021:1-10.

83. Ping X, Cheng Y, Bao J, Shi K, Zou J, Shentu X. KPNA4 is involved in cataract formation via the nuclear import of p53. *Gene.* 2021;786:145621.

84. Burdon KP. The utility of genomic testing in the ophthalmology clinic: A review. *Clinical & Experimental Ophthalmology.* 2021;49(6):615-625.

85. Burdon K, Wirth M, Mackey D, et al. A novel mutation in the Connexin 46 gene causes autosomal dominant congenital cataract with incomplete penetrance. *Journal of medical genetics.* 2004;41(8):e106-e106.

86. Devi RR, Reena C, Vijayalakshmi P. Novel mutations in GJA3 associated with autosomal dominant congenital cataract in the Indian population. *Mol Vis.* 2005;11(100-1):846-852.

87. Gonzalez-Huerta L, Ramirez-Sanchez V, Rivera-Vega M, Messina-Baas O, Cuevas-Covarrubias S. A family with hereditary hyperferritinaemia cataract syndrome: evidence of incomplete penetrance and clinical heterogeneity. *British journal of haematology.* 2008;143(4):596-598.

88. Sacconi S, Féasson L, Antoine JC, et al. A novel CRYAB mutation resulting in multisystemic disease. *Neuromuscular Disorders.* 2012;22(1):66-72.

89. Laurie KJ, Dave A, Straga T, et al. Identification of a novel oligomerization disrupting mutation in CRYAA associated with congenital cataract in a South Australian family. *Human mutation.* 2013;34(3):435-438.

90. Javadiyan S, Craig JE, Souzeau E, et al. Recurrent mutation in the crystallin alpha A gene associated with inherited paediatric cataract. *BMC Res Notes.* 2016;9:83.

91. McGee T, Devoto M, Ott J, Berson E, Dryja T. Evidence that the penetrance of mutations at the RP11 locus causing dominant retinitis pigmentosa is influenced by a gene linked to the homologous RP11 allele. *The American Journal of Human Genetics.* 1997;61(5):1059-1066.

92. Venturini G, Rose AM, Shah AZ, Bhattacharya SS, Rivolta C. CNOT3 is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance. *PLoS genetics.* 2012;8(11):e1003040.

93. Craig JE, Baird PN, Healey DL, et al. Evidence for genetic heterogeneity within eight glaucoma families, with the GLC1A Gln368STOP mutation being an important phenotypic modifier. *Ophthalmology.* 2001;108(9):1607-1620.

94. Davidson BA, Hassan S, Garcia EJ, Tayebi N, Sidransky E. Exploring genetic modifiers of Gaucher disease: The next horizon. *Human mutation.* 2018;39(12):1739-1751.

95. Crotti L, Monti MC, Insolia R, et al. NOS1AP is a genetic modifier of the long-QT syndrome. *Circulation.* 2009;120(17):1657-1663.

96. Corvol H, Blackman SM, Boëlle P-Y, et al. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nature communications.* 2015;6(1):1-8.

97. Emond MJ, Louie T, Emerson J, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nature genetics.* 2012;44(8):886-889.

98. Lee J-M, Wheeler VC, Chao MJ, et al. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell.* 2015;162(3):516-526.

99. Li JL, Hayden MR, Warby SC, et al. Genome-wide significance for a modifier of age at neurological onset in Huntington's disease at 6q23-24: the HD MAPS study. *BMC Med Genet.* 2006;7:71.

100. Maeda YY, Funata N, Takahama S, Sugata Y, Yonekawa H. Two interactive genes responsible for a new inherited cataract (RCT) in the mouse. *Mammalian genome.* 2001;12(4):278-283.

101. Mouton J, Van der Merwe L, Goosen A, et al. MYBPH acts as modifier of cardiac hypertrophy in hypertrophic cardiomyopathy (HCM) patients. *Human genetics.* 2016;135(5):477-483.

102. Timberlake AT, Choi J, Zaidi S, et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. *Elife.* 2016;5:e20125.

103. Oprea GE, Kröber S, McWhorter ML, et al. Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy. *Science.* 2008;320(5875):524-527.

104. Rodríguez JM, Wolfrum S, Robblee M, et al. Altered expression of raet1e, a major histocompatibility complex class 1–like molecule, underlies the atherosclerosis modifier locus ath11 10b. *Circulation research.* 2013;113(9):1054-1064.

105. Riordan JD, Nadeau JH. From peas to disease: modifier genes, network resilience, and the genetics of health. *The American Journal of Human Genetics.* 2017;101(2):177-191.

106. Haldane J. The relative importance of principal and modifying genes in determining some human diseases. *Journal of Genetics.* 1941;41(2-3):149-157.

107. Wada K, Saito J, Yamaguchi M, et al. Pde6brd1 mutation modifies cataractogenesis in Foxe3rct mice. *Biochemical and biophysical research communications.* 2018;496(1):231-237.

108. Matteson PG, Desai J, Korstanje R, et al. The orphan G protein-coupled receptor, Gpr161, encodes the vacuolated lens locus and controls neurulation and lens development. *Proceedings of the National Academy of Sciences.* 2008;105(6):2088-2093.

109. Javadiyan S. *Gene Identification and Characterisation for Blinding Inherited Eye Diseases*, Flinders University, School of Medicine.; 2016.

110. Fish Nursery Protocols. Zebrafish International Resource Center (ZIRC). https://zebrafish.org/wiki/protocols/nursery. Accessed March 5, 2017.

111. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics.* 2012;13(1):1-11.

112. Zhao D, Jones JL, Gasperini RJ, Charlesworth JC, Liu G-S, Burdon KP. Rapid and efficient cataract gene evaluation in F0 zebrafish using CRISPR-Cas9 ribonucleoprotein complexes. *Methods.* 2021.

113. Javadiyan S, Craig JE, Souzeau E, et al. High throughput genetic screening of 51 paediatric cataract genes identifies causative mutations in inherited paediatric cataract in South Eastern Australia. *G3: Genes, Genomes, Genetics.* 2017:g3. 300109.302017.

114. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.

115. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research.* 2019;47(D1):D886-D894.

116. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research.* 2003;31(13):3812-3814.

117. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods.* 2010;7(4):248-249.

118. Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. *Nucleic acids research.* 2004;32(suppl_2):W327-W331.

119. GTEx Portal. Accessed 2.7, 2022.

120. Boerrigter R, Siertsema J, Kema I. Serotonin (5-HT) and the rat's eye. *Documenta Ophthalmologica.* 1992;82(1-2):141-150.

121. Kisilevsky E, Margolin EA. Case of rapid bilateral cataract development in teenager using selective serotonin reuptake inhibitors. *Canadian Journal of Ophthalmology.* 2014;49(5):e114-e115.

122. Lerman S, Delongeas J, Plard J, et al. Cataractogenesis in rats induced by in utero exposure to RG 12915, a 5-HT3 antagonist. *Toxicological Sciences.* 1995;27(2):270-276.

123. Langle UW, Wolf A, Kammuller ME. Cataractogenic effects in rats following chronic administration of SDZ ICT 322, a selective 5-HT3 antagonist. *Fundam Appl Toxicol.* 1993;21(3):393-401.

124. Veglio F, De Sanctis U, Schiavone D, et al. Evaluation of serotonin levels in human aqueous humor. *Ophthalmologica.* 1998;212(3):160-163.

125. Chou P-H, Chu C-S, Chen Y-H, et al. Antidepressants and risk of cataract development: A population-based, nested case-control study. *Journal of affective disorders.* 2017;215:237-244.

126. Erie JC, Brue SM, Chamberlain AM, Hodge DO. Selective serotonin reuptake inhibitor use and increased risk of cataract surgery: a population-based, case-control study. *American journal of ophthalmology.* 2014;158(1):192-197. e191.

127. Etminan M, Mikelberg FS, Brophy JM. Selective serotonin reuptake inhibitors and the risk of cataracts: a nested case-control study. *Ophthalmology.* 2010;117(6):1251-1255.

128. Zawilska JB, Nowak JZ. Regulatory mechanisms in melatonin biosynthesis in retina. *Neurochemistry international.* 1992;20(1):23-36.

129. Li ZR, Reiter RJ, Fujimori O, Oh CS, Duan YP. Cataractogenesis and lipid peroxidation in newborn rats treated with buthionine sulfoximine: preventive actions of melatonin. *Journal of pineal research.* 1997;22(3):117-123.

130. Abe M, Reiter RJ, Orhii PB, Hara M, Poeggeler B. Inhibitory effect of melatonin on cataract formation in newborn rats: evidence for an antioxidative role for melatonin. *Journal of pineal research.* 1994;17(2):94-100.

131. Mishra S, Wu S-Y, Fuller AW, et al. Loss of αB-crystallin function in zebrafish reveals critical roles in the development of the lens and stress resistance of the heart. *Journal of Biological Chemistry.* 2018;293(2):740-753.

132. Zheng S-S, Han R-Y, Xiang L, Zhuang Y-Y, Jin Z-B. Versatile genome engineering techniques advance human ocular disease researches in zebrafish. *Frontiers in cell and developmental biology.* 2018;6:75.

133. Goldsmith P. Modelling eye diseases in zebrafish. *Neuroreport.* 2001;12(13):A73-A77.

134. Lieschke GJ, Currie PD. Animal models of human disease: zebrafish swim into view. *Nature Reviews Genetics.* 2007;8(5):353-367.

135. Schmitt EA, Dowling JE. Early - eye morphogenesis in the zebrafish, Brachydanio rerio. *Journal of Comparative Neurology.* 1994;344(4):532-542.

136. Chhetri J, Jacobson G, Gueven N. Zebrafish—on the move towards ophthalmological research. *Eye.* 2014;28(4):367-380.

137. Richardson R, Tracey-White D, Webster A, Moosajee M. The zebrafish eye—a paradigm for investigating human ocular genetics. *Eye.* 2017;31(1):68-86.

138. Howe K, Clark MD, Torroja CF, et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature.* 2013;496(7446):498-503.

139. Goishi K, Shimizu A, Najarro G, et al. αA-crystallin expression prevents γ-crystallin insolubility and cataract formation in the zebrafish cloche mutant lens. *Development.* 2006;133(13):2585-2593.

140. Gao M, Huang Y, Wang L, et al. HSF4 regulates lens fiber cell differentiation by activating p53 and its downstream regulators. *Cell death & disease.* 2017;8(10):e3082-e3082.

141. Froger A, Clemens D, Kalman K, Nemeth-Cahalan KL, Schilling TF, Hall JE. Two distinct aquaporin 0s required for development and transparency of the zebrafish lens. *Invest Ophthalmol Vis Sci.* 2010;51(12):6582-6592.

142. Clemens DM, Nemeth-Cahalan KL, Trinh L, Zhang T, Schilling TF, Hall JE. In vivo analysis of aquaporin 0 function in zebrafish: permeability regulation is required for lens transparency. *Invest Ophthalmol Vis Sci.* 2013;54(7):5136-5143.

143. Vorontsova I, Gehring I, Hall JE, Schilling TF. Aqp0a Regulates Suture Stability in the Zebrafish Lens. *Investigative Ophthalmology & Visual Science.* 2018;59(7):2869-2879.

144. Hansen L, Comyn S, Mang Y, et al. The myosin chaperone UNC45B is involved in lens development and autosomal dominant juvenile cataract. *European Journal of Human Genetics.* 2014;22(11):1290-1297.

145. Hwang WY, Fu Y, Reyon D, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology.* 2013;31(3):227.

146. Hisano Y, Ota S, Kawahara A. Genome editing using artificial site-specific nucleases in zebrafish. *Development, growth & differentiation.* 2014;56(1):26-33.

147. Jao L-E, Wente SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences.* 2013;110(34):13904-13909.

148. Auer TO, Duroure K, De Cian A, Concordet J-P, Del Bene F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome research.* 2014;24(1):142-153.

149. Ablain J, Durand EM, Yang S, Zhou Y, Zon LI. A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Developmental cell.* 2015;32(6):756-764.

150. Armstrong GAB, Liao M, You Z, Lissouba A, Chen BE, Drapeau P. Homology directed knockin of point mutations in the zebrafish tardbp and fus genes in ALS using the CRISPR/Cas9 system. *PloS one.* 2016;11(3):e0150188.

151. AB zebrafish Zebrafish International Resource Center (ZIRC) https://zfin.org/action/genotype/view/ZDB-GENO-960809-7. Accessed May 15, 2021.

152. Westerfield M. The zebrafish book: a guide for the laboratory use of zebrafish. *http://zfin org/zf_info/zfbook/zfbk html.* 2000.

153. Thermo Fisher Scientific Tm Calculator http://www.thermoscientific.com/pcrwebtools. Accessed March 21, 2021.

154. Hoshijima K, Jurynec MJ, Shaw DK, Jacobi AM, Behlke MA, Grunwald DJ. Highly efficient CRISPR-Cas9-based methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish. *Developmental cell.* 2019;51(5):645-657. e644.

155. Essner J. Zebrafish embryo microinjection: Ribonucleoprotein delivery using the Alt-R™ CRISPR-Cas9 System. https://sfvideo.blob.core.windows.net/sitefinity/docs/default-source/user-submitted-method/crispr-cas9-rnp-delivery-zebrafish-embryos-j-essnerc46b5a1532796e2eaa53ff00001c1b3c.pdf?sfvrsn=52123407_10. Accessed February 4, 2018.

156. Jones JL. *Identifying the genetic causes of paediatric cataract in Australian families*, University of Tasmania; 2020.

157. Kumari SS, Gandhi J, Mustehsan MH, Eren S, Varadaraj K. Functional characterization of an AQP0 missense mutation, R33C, that causes dominant congenital lens cataract, reveals impaired cell-to-cell adhesion. *Experimental eye research.* 2013;116:371-385.

158. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods.* 2012;9(7):676-682.

159. Clement K, Rees H, Canver MC, et al. Accurate and rapid analysis of genome editing data from nucleases and base editors with CRISPResso2. *Nature biotechnology.* 2019;37(3):224.

160. Amplicon P, Clean-Up P, Index P. 16s metagenomic sequencing library preparation. In:2013.

161. Gaudelli NM, Komor AC, Rees HA, et al. Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. *Nature.* 2017;551(7681):464-471.

162. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120.

163. Faul F, Erdfelder E, Lang A-G, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods.* 2007;39(2):175-191.

164. Validated gRNA Sequence Datatable (addgene). https://www.addgene.org/crispr/reference/grna-sequence/. Accessed April 9, 2019.

165. Platt RJ, Chen S, Zhou Y, et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell.* 2014;159(2):440-455.

166. IDT CRISPR-Cas9 guide RNA design checker. https://sg.idtdna.com/site/order/designtool/index/CRISPR_PREDESIGN. Accessed April 10, 2019.

167. Wu RS, Lam, II, Clay H, Duong DN, Deo RC, Coughlin SR. A Rapid Method for Directed Gene Knockout for Screening in G0 Zebrafish. *Dev Cell.* 2018;46(1):112-125 e114.

168. Low R2 Values with ICE. Synthego. https://www.synthego.com/help/low-r2. Accessed July 18, 2020.

169. Etminan M, Mikelberg FS, Brophy JM. Selective serotonin reuptake inhibitors and the risk of cataracts: a nested case-control study. *Ophthalmology.* 2010;117(6):1251-1255.

170. Fu Y, Dai Q, Zhu L, Wu S. Antidepressants use and risk of cataract development: a systematic review and meta-analysis. *BMC ophthalmology.* 2018;18(1):1-6.

171. Bartness T, Goldman B. Mammalian pineal melatonin: a clock for all seasons. *Experientia.* 1989;45(10):939-945.

172. Arendt J. Melatonin and the pineal gland: influence on mammalian seasonal and circadian physiology. *Reviews of reproduction.* 1998;3:13-22.

173. Lerner AB, Case JD, Takahashi Y. Isolation of melatonin and 5-methoxyindole-3-acetic acid from bovine pineal glands. *Journal of Biological Chemistry.* 1960;235(7):1992-1997.

174. Quay W. Increases in volume, fluid content, and lens weight of eyes following systemic administration of melatonin. *Journal of pineal research.* 1984;1(1):3-13.

175. Abe M, Itoh MT, Miyata M, Shimizu K, Sumi Y. Circadian rhythm of serotonin N-acetyltransferase activity in rat lens. *Experimental eye research.* 2000;70(6):805-808.

176. Bardak Y, Özertürk Y, Özgüner F, Durmus M, Delibas N. Effect of melatonin against oxidative stress in ultraviolet-B exposed rat lens. *Current eye research.* 2000;20(3):225-230.

177. Karslioğlu I, Ertekin MV, Taysi S, et al. Radioprotective effects of melatonin on radiation-induced cataract. *Journal of radiation research.* 2005;46(2):277-282.

178. Yağci R, Aydin B, Erdurmuş M, et al. Use of melatonin to prevent selenite-induced cataract formation in rat eyes. *Current eye research.* 2006;31(10):845-850.

179. Taysi S, Memisogullari R, Koc M, et al. Melatonin reduces oxidative stress in the rat lens due to radiation-induced oxidative injury. *International journal of radiation biology.* 2008;84(10):803-808.

180. Shirazi A. Radioprotective effect of melatonin in reducing oxidative stress in rat lenses. *Cell Journal (Yakhteh).* 2011;13(2):79.

181. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.

182. Silk M, Petrovski S, Ascher DB. MTR-Viewer: identifying regions within genes under purifying selection. *Nucleic acids research.* 2019;47(W1):W121-W126.

183. Burger A, Lindsay H, Felker A, et al. Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development.* 2016;143(11):2025-2037.

184. Watson CJ, Monstad-Rios AT, Bhimani RM, et al. Phenomics-Based Quantification of CRISPR-Induced Mosaicism in Zebrafish. *Cell Systems.* 2020.

185. Gagnon JA, Valen E, Thyme SB, et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PloS one.* 2014;9(5):e98186.

186. Dahm R, Schonthaler HB, Soehn AS, Van Marle J, Vrensen GF. Development and adult morphology of the eye lens in the zebrafish. *Experimental eye research.* 2007;85(1):74-89.

187. Illumina. Infinium® HTS Assay Protocol Guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/infinium_assays/infinium-hts/infinium-hts-assay-reference-guide-15045738-04.pdf. Accessed May 18, 2020.

188. Infinium Global Screening Array Product Files. https://sapac.support.illumina.com/array/array_kits/infinium-global-screening-array/downloads.html. Accessed June 11, 2021.

189. Technical Note: Infinium Genotyping Data Analysis. https://www.illumina.com/Documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf. Accessed June 11, 2021.

190. Purcell S. PLINK v1.9. http://pngu.mgh.harvard.edu/purcell/plink/. Accessed 10.2, 2019.

191. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics.* 2007;81(3):559-575.

192. Seidman DN, Shenoy SA, Kim M, et al. Rapid, phase-free detection of long identity-by-descent segments enables effective relationship classification. *The American Journal of Human Genetics.* 2020;106(4):453-466.

193. Baron RV, Kollar C, Mukhopadhyay N, Weeks DE. Mega2: validated data-reformatting for linkage and association analyses. *Source code for biology and medicine.* 2014;9(1):1-6.

194. Baron RV KC, Mukhopadhyay N, Almasy L, Schroeder M, Mulvihill WP, Weeks DE. Mega2 (Version 6.0.0). https://watson.hgen.pitt.edu/mega2.html. Published 2018. Accessed.

195. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics.* 2002;30(1):97-101.

196. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature genetics.* 1995;11(3):241-247.

197. bcbio-nextgen. https://github.com/bcbio/bcbio-nextgen. Accessed June 13, 2021.

198. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:13033997.* 2013.

199. Tischler G, Leonard S. biobambam: tools for read pair collation based algorithms on BAM files. *Source Code for Biology and Medicine.* 2014;9(1):1-18.

200. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research.* 2010;20(9):1297-1303.

201. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome biology.* 2014;15(6):1-19.

202. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *GigaScience.* 2021;10(2).

203. Van der Auwera G, O'Connor B. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. In: O'Reilly Media; 2020.

204. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics.* 2011;43(5):491.

205. Arnold K, Gosling J, Holmes D. *The Java programming language.* Addison Wesley Professional; 2005.

206. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research.* 2010;38(16):e164-e164.

207. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research.* 2016;44(D1):D733-D745.

208. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-291.

209. Consortium GP. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Abecasis GR. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.

210. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: an accessible system for testing SNV novelty. *Bioinformatics.* 2011;27(22):3216-3217.

211. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics.* 2014;46(3):310-315.

212. Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human mutation.* 2013;34(1):57-65.

213. Jiao X, Khan SY, Kaul H, et al. Autosomal recessive congenital cataracts linked to HSF4 in a consanguineous Pakistani family. *PloS one.* 2019;14(12):e0225010.

214. Sun M, Chen C, Hou S, et al. A novel mutation of PANK4 causes autosomal dominant congenital posterior cataract. *Human mutation.* 2019;40(4):380-391.

215. Jia X, Zhang F, Bai J, et al. Combinational analysis of linkage and exome sequencing identifies the causative mutation in a Chinese family with congenital cataract. *BMC medical genetics.* 2013;14(1):1-6.

216. Evers C, Paramasivam N, Hinderhofer K, et al. SIPA1L3 identified by linkage analysis and whole-exome sequencing as a novel gene for autosomal recessive congenital cataract. *European Journal of Human Genetics.* 2015;23(12):1627-1633.

217. Rahit K, Tarailo-Graovac M. Genetic modifiers and rare mendelian disease. *Genes.* 2020;11(3):239.

218. Javadiyan S, Craig JE, Sharma S, et al. Novel missense mutation in the bZIP transcription factor, MAF, associated with congenital cataract, developmental delay, seizures and hearing loss (Aymé-Gripp syndrome). *BMC medical genetics.* 2017;18(1):1-6.

219. Lenassi E, Clayton-Smith J, Douzgou S, et al. Clinical utility of genetic testing in 201 preschool children with inherited eye disorders. *Genetics in Medicine.* 2020;22(4):745-751.

220. Teare MD, Barrett JH. Genetic linkage studies. *The Lancet.* 2005;366(9490):1036-1044.

221. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *The American Journal of Human Genetics.* 1998;62(5):1198-1211.

222. Thompson E. Monte Carlo in genetic analysis. *University of Washington.* 1995.

223. Yamashita S, Furumoto K, Nobukiyo A, Kamohara M, Ushijima T, Furukawa T. Mapping of A gene responsible for cataract formation and its modifier in the UPL rat. *Investigative ophthalmology & visual science.* 2002;43(10):3153-3159.

224. Narita M, Wang Y, Kita A, Omi N, Yamada Y, Hiai H. Genetic analysis of Nakano Cataract and its modifier genes in mice. *Experimental eye research.* 2002;75(6):745-751.

225. Li BI, Ababon MR, Matteson PG, Lin Y, Nanda V, Millonig JH. Congenital cataract in Gpr161vl/vl mice is modified by proximal chromosome 15. *PloS one.* 2017;12(1):e0170724.

226. Haseman J, Elston R. The investigation of linkage between a quantitative trait and a marker locus. *Behavior genetics.* 1972;2(1):3-19.

227. Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *American journal of human genetics.* 1995;57(2):439.

228. Williams JT, Blangero J. Comparison of variance components and sibpair-based approaches to quantitative trait linkage analysis in unselected samples. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society.* 1999;16(2):113-134.

229. Williams JT, Blangero J. Power of variance component linkage analysis to detect quantitative trait loci. *Annals of human genetics.* 1999;63(6):545-563.

230. Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. *American journal of human genetics.* 1994;54(3):535.

231. Duggirala R, Williams JT, Williams-Blangero S, Blangero J. A variance component approach to dichotomous trait linkage analysis using a threshold model. *Genetic epidemiology.* 1997;14(6):987-992.

232. Blangero J, Williams JT, Almasy L. 12 Variance component methods for detecting complex trait loci. 2001.

233. Yan D, Liu X-Z. Modifiers of hearing impairment in humans and mice. *Current genomics.* 2010;11(4):269-278.

234. Schäffer AA. Digenic inheritance in medical genetics. *Journal of medical genetics.* 2013;50(10):641-652.

235. Burdon KP, Coster DJ, Charlesworth JC, et al. Apparent autosomal dominant keratoconus in a large Australian pedigree accounted for by digenic inheritance of two novel loci. *Human genetics.* 2008;124(4):379-386.

236. Kajiwara K, Berson EL, Dryja TP. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science.* 1994;264(5165):1604-1608.

237. Tam PK, Garcia-Barceló M. Genetic basis of Hirschsprung's disease. *Pediatric surgery international.* 2009;25(7):543-558.

238. MacArthur D, Manolio T, Dimmock D, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature.* 2014;508(7497):469-476.

239. Sharma R, Grover A. Myocilin-associated Glaucoma: A Historical Perspective and Recent Research Progress. *Molecular Vision.* 2021;27:480.

240. Toubiana J, Okada S, Hiller J, et al. Heterozygous STAT1 gain-of-function mutations underlie an unexpectedly broad clinical phenotype. *Blood, The Journal of the American Society of Hematology.* 2016;127(25):3154-3164.

241. Fernandes-Rosa FL, Daniil G, Orozco IJ, et al. A gain-of-function mutation in the CLCN2 chloride channel gene causes primary aldosteronism. *Nature genetics.* 2018;50(3):355-361.

242. Albadri S, Del Bene F, Revenu C. Genome editing using CRISPR/Cas9-based knock-in approaches in zebrafish. *Methods.* 2017;121:77-85.

243. Goishi K, Shimizu A, Najarro G, et al. αA-crystallin expression prevents γ-crystallin insolubility and cataract formation in the zebrafish cloche mutant lens. 2006.

244. Clemens DM, Németh-Cahalan KL, Trinh L, Zhang T, Schilling TF, Hall JE. In vivo analysis of aquaporin 0 function in zebrafish: permeability regulation is required for lens transparency. *Investigative ophthalmology & visual science.* 2013;54(7):5136-5143.

245. Zou P, Wu S-Y, Koteiche HA, et al. A conserved role of αA-crystallin in the development of the zebrafish embryonic lens. *Experimental eye research.* 2015;138:104-113.

246. Posner M, Skiba J, Brown M, Liang JO, Nussbaum J, Prior H. Loss of the small heat shock protein alphaA-crystallin does not lead to detectable defects in early zebrafish lens development. *Exp Eye Res.* 2013;116:227-233.

247. Varshney GK, Sood R, Burgess SM. Understanding and Editing the Zebrafish Genome. *Adv Genet.* 2015;92:1-52.

248. Kok FO, Shin M, Ni C-W, et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Developmental cell.* 2015;32(1):97-108.

249. Oh B, Hwang S, McLaughlin J, Solter D, Knowles BB. Timely translation during the mouse oocyte-to-embryo transition. *Development.* 2000;127(17):3795-3803.

250. East EM. Inheritance in crosses between Nicotiana langsdorffii and Nicotiana alata. *Genetics.* 1916;1(4):311.

251. Wright S. Evolution in Mendelian populations. *Genetics.* 1931;16(2):97.

252. Wright S. Surfaces of selective value revisited. *The American Naturalist.* 1988;131(1):115-123.

253. Fisher RA. *The genetical theory of natural selection.* Рипол Классик; 1958.

254. Mackay TF. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nature Reviews Genetics.* 2014;15(1):22-33.

255. Sackton TB, Hartl DL. Genotypic context and epistasis in individuals and populations. *Cell.* 2016;166(2):279-287.

256. Williams EG, Auwerx J. The convergence of systems and reductionist approaches in complex trait analysis. *Cell.* 2015;162(1):23-32.

257. Li J-L, Hayden MR, Almqvist EW, et al. A genome scan for modifiers of age at onset in Huntington disease: The HD MAPS study. *The American Journal of Human Genetics.* 2003;73(3):682-687.

258. Gayán J, Brocklebank D, Andresen JM, et al. Genomewide linkage scan reveals novel loci modifying age of onset of Huntington's disease in the Venezuelan HD kindreds. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society.* 2008;32(5):445-453.

259. Tomás M, Napolitano C, De Giuli L, et al. Polymorphisms in the NOS1AP gene modulate QT interval duration and risk of arrhythmias in the long QT syndrome. *Journal of the American College of Cardiology.* 2010;55(24):2745-2752.

260. Matsumura H, Nakano Y, Ochi H, et al. H558R, a common SCN5A polymorphism, modifies the clinical phenotype of Brugada syndrome by modulating DNA methylation of SCN5A promoters. *Journal of biomedical science.* 2017;24(1):1-10.

261. Choquet H, Melles RB, Anand D, et al. A large multiethnic GWAS meta-analysis of cataract identifies new risk loci and sex-specific effects. *Nature Communications.* 2021;12(1):1-12.

262. Yonova-Doing E, Zhao W, Igo RP, et al. Common variants in SOX-2 and congenital cataract genes contribute to age-related nuclear cataract. *Communications biology.* 2020;3(1):1-8.

263. Shiels A, Bennett TM, Knopf HL, et al. The EPHA2 gene is associated with cataracts linked to chromosome 1p. *Molecular vision.* 2008;14:2042.

264. Jun G, Guo H, Klein BE, et al. EPHA2 is associated with age-related cortical cataract in mice and humans. *PLoS Genet.* 2009;5(7):e1000584.

265. Liu Y, Ke M, Yan M, et al. Association between gap junction protein-alpha 8 polymorphisms and age-related cataract. *Molecular biology reports.* 2011;38(2):1301-1307.

266. Karas N, Gobec L, Pfeifer V, Mlinar B, Battelino T, Lukac-Bajalo J. Mutations in galactose-1-phosphate uridyltransferase gene in patients with idiopathic presenile cataract. *Journal of inherited metabolic disease.* 2003;26(7):699-704.

267. Zuercher J, Neidhardt J, Magyar I, et al. Alterations of the 5′ untranslated region of SLC16A12 lead to age-related cataract. *Investigative ophthalmology & visual science.* 2010;51(7):3354-3361.

268. Shi Y, Shi X, Jin Y, et al. Mutation screening of HSF4 in 150 age-related cataract patients. *Molecular vision.* 2008;14:1850.

269. Okano Y, Asada M, Fujimoto A, et al. A genetic factor for age-related cataract: identification and characterization of a novel galactokinase variant,"Osaka," in Asians. *The American Journal of Human Genetics.* 2001;68(4):1036-1042.

270. Faniello MC, Di Sanzo M, Quaresima B, et al. Bilateral cataract in a subject carrying a C to A transition in the L ferritin promoter region. *Clinical biochemistry.* 2009;42(9):911-914.

271. Bhagyalaxmi S, Srinivas P, Barton KA, et al. A novel mutation (F71L) in α A-crystallin with defective chaperone-like function associated with age-related cataract. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease.* 2009;1792(10):974-981.