BMJ Open Ophthalmology

Pathogenic genetic variants identified in Australian families with paediatric cataract

Johanna L Jones ,[1] Bennet J McComish ,[1] Sandra E Staffieri ,[2,3] Emmanuelle Souzeau,[4] Lisa S Kearns,[2] James E Elder,[3,5] Jac C Charlesworth,[1] David A Mackey,[6] Jonathan B Ruddle,[2,3] Deepa Taranath,[4] John Pater,[7] Theresa Casey,[7] Jamie E Craig,[4] Kathryn P Burdon[1,4]

To cite: Jones JL, McComish BJ, Staffieri SE, *et al*. Pathogenic genetic variants identified in Australian families with paediatric cataract. *BMJ Open Ophthalmology* 2022;**7**:e001064. doi:10.1136/bmjophth-2022-001064

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjophth-2022-001064).

Received 11 May 2022
Accepted 5 August 2022

For numbered affiliations see end of article.

Correspondence to
Dr Johanna L Jones; johannaj@utas.edu.au

ABSTRACT

Objective Paediatric (childhood or congenital) cataract is an opacification of the normally clear lens of the eye and has a genetic basis in at least 18% of cases in Australia. This study aimed to replicate clinical gene screening to identify variants likely to be causative of disease in an Australian patient cohort.

Methods and analysis Sixty-three reported isolated cataract genes were screened for rare coding variants in 37 Australian families using genome sequencing.

Results Disease-causing variants were confirmed in eight families with variant classification as 'likely pathogenic'. This included novel variants *PITX3* p.(Ter303LeuextTer100), *BFSP1* p.(Glu375GlyfsTer2), and *GJA8* p.(Pro189Ser), as well as, previously described variants identified in genes *GJA3, GJA8, CRYAA, BFSP1, PITX3, COL4A1* and *HSF4*. Additionally, eight variants of uncertain significance with evidence towards pathogenicity were identified in genes: *GJA3, GJA8, LEMD2, PRX, CRYBB1, BFSP2,* and *MIP*.

Conclusion These findings expand the genotype–phenotype correlations of both pathogenic and benign variation in cataract-associated genes. They further emphasise the need to develop additional evidence such as functional assays and variant classification criteria specific to paediatric cataract genes to improve interpretation of variants and molecular diagnosis in patients.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Paediatric (congenital) cataract is a genetically and phenotypically heterogeneous disease. Genomic testing of children with cataract is increasingly being used to refine the diagnosis, determine prognosis and guide genetic counselling.

WHAT THIS STUDY ADDS

⇒ Three novel disease-causing variants, in genes *PITX3, BFSP1 and GJA8*, expand the genotype–phenotype spectrum of paediatric cataract and several previously described variants re-enforce their pathogenic classifications. Eight families had 'variants of uncertain significance' that could be reclassified as likely pathogenic with additional evidence.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Stringent classification of variants is required for clinical genetic testing and highlights that many variants do not have sufficient evidence for a clinically actionable classification, even in well-characterised disease-causing genes.

INTRODUCTION

Paediatric cataract (childhood or congenital cataract), a clouding of the crystalline lens of the eye during childhood, is a heritable condition in at least 18% of cases in Australia.[1] The total number of genes associated with a cataract phenotype is in excess of 200.[2] However, a core set of approximately 40 well-established genes are known to cause isolated paediatric cataract that include: crystallin genes (*CRYAA, CYRAB, CRYBA1, CRYBA2, CRYBA4, CRYBB1, CRYBB2, CRYBB3, CRYGA, CRYGB, CRYGC, CRYGD* and *CRYGS*), genes encoding membrane structural proteins (*GJA3, GJA8, MIP* and *LIM2*) and cytoskeletal proteins (*VIM, BFSP1* and *BFSP2*), transcription factor genes (*HSF4, PITX3, PAX6, FOXE3* and *MAF*) and genes for signalling molecules such as *EPHA2*. Other genes, such as *NHS, FTL, AGK, MIR184, GCNT2* and *GALK1*, are also routinely assessed and are associated with either isolated paediatric cataract or paediatric cataract as a characteristic phenotype as part of a syndrome. The success of genetic screens for familial paediatric cataract has varied greatly with reported solve rates between 25% and 77%.[3 4] Cohort, gene panel selection, sequencing methodology and the stringent use of variant classification criteria will have contributed to the varied successes. Despite this, routine gene screening and variant reporting in the research setting plays a very important role in expanding the known genetic and phenotypic spectrum of this condition, particularly regarding novel

and less-established cataract-associated genes. Variant reporting of both negative and positive findings improves the collective understanding of these genes, their products and the mechanisms that cause cataracts, and ultimately improves clinical diagnosis and outcomes for patients.

This study aimed to investigate a panel of 63 isolated cataract-associated genes in 37 Australian families. Disease-causing likely pathogenic variants in genes *GJA3, GJA8, CRYAA, BFSP1, PITX3, COL4A1* and *HSF4* were identified in eight families. Variants of uncertain significance (VUS) were identified in a range of well-established and other cataract genes that may be disease-causing but require additional evidence of pathogenicity following current scoring with the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines.

MATERIALS AND METHODS

All affected participants were diagnosed with inherited paediatric cataract, based on the observed cataract phenotype and a reported family history, by the examining ophthalmologist and genetic counsellors, respectively. Relatives were examined on recruitment to the study or following diagnosis during routine examination anytime thereafter. DNA was extracted from whole blood using a QiaAmp DNA blood Maxi Kit (Qiagen), buccal mucosa swabs using the PureGene DNA Isolation Kit (Gentra Systems) or saliva using Oragene DNA saliva collection kits (DNA Genotek, Ontario, Canada). Clinically actionable variants were returned to patients as research findings through our genetic counsellors, who facilitate subsequent nationally accredited genetic testing in Australia and appropriate counselling.

Genome sequencing was performed on the DNA of an affected individual (proband) from each of 37 families with European ancestry. Sequencing was either 150 bp paired-end sequencing on a HiSeq X Ten platform (30× coverage, Illumina) with an Illumina TrueSeq Nano Library Prep (V.2.5) at the Kinghorn Centre for Clinical Genomics (Sydney, Australia) or 250 bp paired-end sequencing on a NovaSeq 6000 platform (30× coverage, Illumina) with an Illumina Nextra DNA Flex library preparation at the Ramaciotti Centre for Genomics (Sydney, Australia). Variant calling was performed using the bcbio-next-gen pipeline (https://doi.org/10.5281/zenodo.3564938) with the BWA-MEM algorithm[5] for read alignment to human reference genome hg19 and variant calling with GATK[6] according to best practice guidelines. Variant annotation was performed using ANNOVAR.[7] MultiQC[8] reporting was used to assess sample quality and average target coverage, and read depth at variant sites exceeded 30 for most samples (online supplemental table S1). Five samples with lower target coverage and read depth were retained, although interpreted with caution.

Sixty-three genes were selected for assessment (online supplemental table S2) based on previous well-established congenital cataract genes, research-reported candidate genes or genes that are otherwise associated with syndromic conditions or other ocular phenotypes that have reports of isolated congenital cataracts or cataracts as an early presenting feature. The gene list is comparable to the panels used by accredited genetic testing laboratories for non-syndromic paediatric cataract. Variants within those genomic regions were filtered to functional 'exonic' or 'splicing' variants, with an MAF ≤0.00022 in gnomADv2.1.1[9] pop_max (highest population frequency) to match the reported Australian disease frequency of 2.2 per 10 000 live births.[1] Variants were prioritised for further analysis if they had a CADD PHRED score ≥10 or, for synonymous and non-coding RNA variants, a CADD PHRED score ≥15. Variant validation and cosegregation was performed with primers designed using NCBI primer blast[10] (online supplemental table S3). PCR was performed using MyTaq HS DNA polymerase (Bioline) prior to Sanger sequencing with either BrightDye Terminator (MCLAB) or BrilliantDye Terminator (Nimagen) Cycle sequencing kits and sequenced using an ABI 3500 Genetic Analyzer (Life Technologies), all according to manufacturer's instructions.

Predictive analysis of protein folding between wild-type and variant protein sequences was performed using the PredictProtein folding prediction tool (https://predictprotein.org).[11 12] HOPE protein structure analysis[13] was used to further assess the deleteriousness of missense variants on protein function. The mFold tool[14] (http://www.unafold.org/) was used for comparing microRNA folding between wild-type and variant sequences. Validated variants were interpreted using the ACMG-AMP guidelines[15] via InterVar[16] and manually adjusted as appropriate. Cosegregation considered based on informative meioses[17] and PP5 and BP6 criteria were excluded from use.[18] All variants reported in this study have been submitted to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/); ClinVar accession numbers SCV001573165-SCV001573189. Evidence for pathogenicity according to the ACMG-AMP guidelines is shown in online supplemental table S4).

RESULTS

Sixty-three paediatric cataract genes were screened in probands from 37 Australian families with inherited paediatric cataract. This cohort contained 13 probands/families that remained unsolved following a screen of 51 genes (families indicated in online supplemental table S1).[19] The other 24 probands/families were either unsolved following analysis of the *NHS,*[20] *EPHA2*[21] and crystallin genes[22] or have not previously been investigated.

Likely pathogenic variants

Eight probands/families had variants classified as likely pathogenic (8 of 37, 22%), including of three novel and five previously described variants (table 1).

Table 1 Rare coding variants detected in probands with paediatric cataract

Family ID	Phenotype	Gene	Genomic change	Coding change	Protein change	gnomAD popmax	CADD/Polyphen2/SIFT/PROVEAN	Seg	First publication	ACMG-AMP
Pathogenic/Likely pathogenic variants										
CRCH21	Central	*GJA3*	Chr13:20717252G>A	NM_021954.4:c.176C>T	p.(Pro59Leu)	0	27.7/P^{r}D/D/D	Yes	Bennett *et al*[37]	LP
CRCH90	Dense nuclear	*GJA3*	Chr13:20717372G>A	NM_021954.4:c.56C>T	p.(Thr19Met)	0	27.3/P^{r}D/D/D	Yes	Santhiya *et al*[38]	LP
CRCH29	–	*CRYAA*	Chr21:44589243C>T	NM_000394.4:c.34C>T	p.(Arg12Cys)	0	28/P^{r}D/D/D	Yes	Hansen *et al*[24]	LP
CRCH38	Central posterior cortical with dots	*COL4A1*	Chr13:110833673C>T	NM_001845.6:c.2159G>A	p.(Gly720Asp)	0	23.9/P^{r}D/D/D	Yes	Sibon *et al*[33]	LP
		MIR184	Chr15:79502181T>C	NR_029705.1:n.52T>C	n/a	0	16.87/-/-/-	Yes	–	VUS
CSA168	Lamellar/zonular	*HSF4*	Chr16:67199491A>G	NM_001040667.3:c.190A>G	p.(Lys64Glu)	0	29.4/P^{r}D/D/D	De novo?	Berry *et al*[23]	LP
CRCH28	Posterior subcapsular	*PITX3*	Chr10:103990272C>A	NM_005029.4:c.908G>T	p.(Ter303LeuextTer100)	0	18.39/-/-/-	Yes	–	LP
		EYA1	Chr8:72127864G>A	NM_000503.6:c.1460C>T	p.(Ser487Leu)	0.001	22.8/B/T/N	No	–	B
CSA182	Posterior sutural	*BFSP1*	Chr20:17475593del	NM_001195.5:c.1124del	p.(Glu375Glyfs Ter2)	0	32/-/-/-	Yes	–	LP
CRVEEH77	Nuclear or dense total	*GJA8*	Chr1:147380647C>T	NM_005267.5:c.565C>T	p.(Pro189Ser)	0	25.7/P^{r}D/D/D	Yes	*Hansen *et al*[24] Fernández-Alcalde *et al*[25]	LP
Variants of uncertain significance with evidence towards pathogenicity										
CRCH137	Lamellar	*GJA8*	Chr1:147380146G>A	NM_005267.5:c.64G>A	p.(Gly22Ser)	0	26.9/P^{r}D/D/D	Yes	*Wang *et al*[27] Fernández-Alcalde *et al*[25] Ye *et al*[28]	VUS
CTAS71	Central nuclear	*GJA8*	Chr1:147380470_147380472del	NM_005267.5:c.388_390del	p.(Lys131del)	0	14.76/-/-/N	n/a	–	VUS
CSA192	–	*GJA3*	Chr13:20717385G>T	NM_021954.4:c.43C>A	p.(Gln15Lys)	0	24.2/P^{s}D/D/N	n/a	–	VUS
CRVEEH79	Blue dot	*LEMD2*	Chr6:33756893T>C	NM_001348710.2:c.1A>G	p.(?)	0	23.2/B/D/N	Yes	–	VUS
CSA93	Dense post central with lenticonus	*PRX*	Chr19:40904522C>T	NM_020956.2:c.386G>A	p.(Arg129His)	0	18.86/P^{s}D/T/N	Yes	–	VUS
CTAS34	Anterior polar & nuclear	*CRYBB1*	Chr22:27008054A>T	NM_001887.4:c.281T>A	p.(Ile94Asn)	0	25.6/P^{r}D/D/D	Yes	–	VUS
		MIP	Chr12:56847410C>T	NM_012064.4:c.490G>A	p.(Val164Ile)	1.0×10^{-4}	15.54/B/T/N	RP?	–	VUS
CQLD130	–	*BFSP2*	Chr3:133119192C>T	NM_003571.4:c.265C>T	p.(Arg89Trp)	1.0×10^{-4}	23.2/P^{r}D/T/D	RP?	–	VUS
CRCH4	Nuclear	*MIP*	Chr12:56848060C>T	NM_012064.4:c.338G>A	p.(Arg113Gln)	0	29.1/P^{r}D/D/D	RP?	–	VUS
Variants of uncertain significance that are unlikely to be causing disease										
CSA158	Lamellar or cortical	*CYP51A1*	Chr7:91761099C>T	NM_000786.4:c.280G>A	p.(Ala94Thr)	0	28.8/P^{r}D/D/N	Yes	–	VUS
		WFS1	Chr4:6302631C>T	NM_001145853.1:c.1109C>T	p.(Ala370Val)	0	25.9/P^{r}D/T/N	No	–	VUS
CRCH5	Mid-periphery cortical dots, anterior polar	*LONP1*	Chr19:5694473G>A	NM_001276480.1:c.1657C>T	p.(Pro553Ser)	0	26.1/P^{r}D/D/D	No †1/1	–	VUS
CSA178	Lamellar	*IARS2*	Chr1:220279332G>C	NM_018060.4:c.1166G>C	p.(Gly389Ala)	0	26.8/P^{r}D/D/D	No	–	VUS

Continued

Table 1 Continued

Family ID	Phenotype	Gene	Genomic change	Coding change	Protein change	gnomAD popmax	CADD/Polyphen2/SIFT/ PROVEAN	Seg	First publication	ACMG-AMP
CQLD88	–	*FBN1*	Chr15:48779550G>A	NM_000138.5:c.3422C>T	p.(Pro1141Leu)	6.4×10^{-5}	28/P'D/D/D	RP?	–	VUS
CSA100	Nuclear	*LSS*	Chr21:47647553G>C	NM_001001438.3:c.232C>G	p.(Leu78Val)	6.5×10^{-5}	16.77/B/T/N	No	–	VUS

Rare coding variants identified in the probands and assessment of cosegregation in the family. Section 1: pathogenic variants, section 2: variants of uncertain significance and section 3: variants unlikely to be disease-causing. Genomic change reported using human reference genome hg19. '–' indicates data not available, n/a indicates not applicable. gnomAD popmax minor allele frequency represents highest frequency observed in V2.1.1 data release. Variant predicted functional effects using CADD version 1.6[39 40] with PHRED score given; Polyphen-2[41] 'PD' probably damaging, 'PD' possibly damaging, 'B' benign, SIFT[42] 'D' damaging, 'T' tolerated, PROVEAN[43] 'D' damaging, 'N' neutral, '–' scoring unavailable. Seg: cosegregation in the family either Yes, No, de novo?, RP? if possible reduced penetrance, or 'n/a' singleton and not applicable. Variant interpretation based on the ACMG-AMP guidelines.[15]
*Reported alternate amino acid change at same location.
†1/1 homozygous observation of variant in proband.
ACMG, American College of Medical Genetics and Genomics; B, benign; LP, likely pathogenic; n/a, not available; P, pathogenic; VUS, variant of uncertain significance.

Two previously described variants were identified in the *GJA3* gene (p.Pro59Leu and p.Thr19Met) in families CRCH21 and CRCH90 (figure 1A–B). A commonly reported pathogenic *CRYAA* (p.Arg12Cys) change was identified in family CRCH29 (figure 1C). A known *COL4A1* p.(Gly720Asp) change was determined to be disease-causing in family CRCH38 (figure 1D). Additionally, in family CRCH38, a *MIR184* n.52T>C change was also observed but deemed benign with no harmful predicted change to secondary structure of the microRNA molecule (online supplemental figure S1). A p.(Lys64Glu) *HSF4* change was identified in family CSA168 (figure 1E) and is located in the highly conserved DNA binding domain of the protein. This *HSF4* variant has previously been reported as pathogenic[23] in a family with lamellar paediatric cataract, with a comparable phenotype observed here in CSA168-01 (figure 2A).

A novel stop loss *PITX3* variant, in family CRCH28 (figure 1F), was predicted to extend the normal 302-residue long protein by an additional 100 amino acids, when assessed using the NCBI ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/). Comparative protein folding prediction indicated small structural changes within the DNA binding domain and the formation of beta-strands with potential DNA-binding and protein-binding affinity in the additional 100 amino acid protein extension (online supplemental figure S2). A benign *EYA1* variant was also observed but failed to cosegregate with the disease in the family. Affected individuals were diagnosed between 9 and 21 years of age with posterior subcapsular cataracts (table 1) and had no other reportable ocular features. This cataract phenotype is consistent with the posterior polar or posterior subcapsular opacifications reported in *PITX3* variants to date, with or without additional anterior segment mesenchymal dysgenesis features.[2]

A novel *BFSP1* c.1124delA p.(Glu375GlyfsTer2) frameshift variant segregated in an autosomal dominant manner in family CSA182 (figure 1G). A posterior sutural cataract phenotype with a pulverulent appearance (figure 2B) was consistently observed across all affected individuals in the family. Sutural, often pulverulent-like, cataracts are the most frequently reported phenotype with *BFSP2* variants while the range of phenotypes observed for *BFSP1* variants is wider and includes nuclear, lamellar and cortical cataracts.[2]

In family CRVEEH77, a novel p.(Pro189Ser) *GJA8* variant was identified. Previous reports of variants c.565C>G p.(Pro189Ala) and c.566C>T p.(Pro189Leu) have been made at this location in patients with isolated paediatric cataracts, all with nuclear or nuclear inclusive phenotypes.[24 25] Additionally, the same residue change to the equivalent conserved amino acid in the GJA3 protein, p.(Pro187Ser), has been reported as disease-causing.[26]

VUS with evidence towards pathogenicity

Eight probands/families had VUS with evidence towards pathogenicity (22%, table 1). Six of these reside in

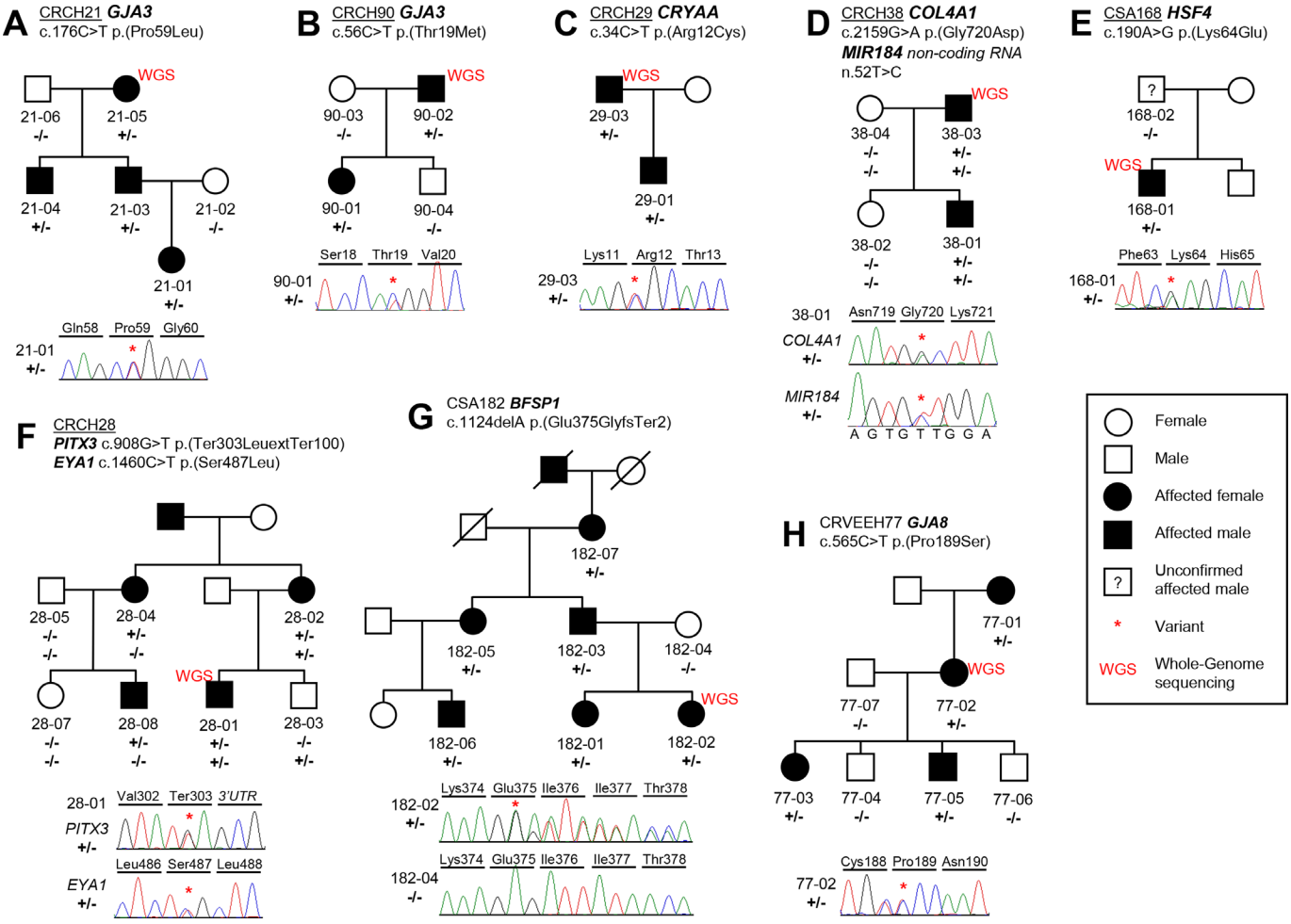

Figure 1 Families with likely pathogenic variants in isolated paediatric cataract causing genes. (A) Family CRCH21 with a segregating previously described *GJA3* p.(Pro59Leu) variant. (B) Family CRCH90 with a previously described *GJA3* p.(Thr19Met) variant. (C) Family CRCH29 with a previously described *CRYAA* p.(Arg12Cys) variant. (D) Family CRCH38 with previously described *COL4A1* p.(Gly720Asp) variant and *MIR184* n.52T>C variant of uncertain significance. (E) Family CSA168 with previously described *HSF4* p.(Lys64Glu) variant in the proband. (F) Family CRCH28 with a novel segregating *PITX3* p.(Ter303LeuextTer100) variant and non-segregating *EYA1* p.(Ser487Leu) variant. (G) Family CSA182 with a novel *BFSP1* c.1124delA frameshift variant. (H) Family CRVEEH77 with a novel segregating *GJA8* p.(Pro189Ser) variant altering an amino acid that has been previously associated with paediatric cataract.

well-established isolated cataract genes *GJA3, GJA8, CRYBB1, BFSP2* and *MIP*. With additional evidence of pathogenicity, such as additional meiosis demonstrating segregation or robust functional evaluation, these variants would likely be upgraded in classification to pathogenic variants.

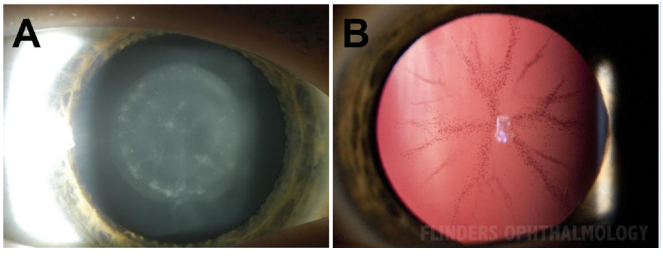

Figure 2 Cataract phenotypes. (A) Lamellar cataract observed in CSA168-01 via slit-lamp photography using direct illumination. (B) Transillumination of CSA182-06, displaying the posterior sutural cataract phenotype with a pulverulent appearance was consistently observed in all affected individuals in the family.

Three variants were identified in connexin genes *GJA8* and *GJA3*. A GJA8 p.(Gly22Ser) change, observed in family CRCH137, has been previously reported[25 27 28] and is one meiosis short of reclassification as likely pathogenic (table 1, figure 3A). In family CTAS71, the novel p.(Lys131del) change (table 1, figure 3B) is located centrally in the GJA8 protein's cytoplasmic loop which is a less conserved protein region in general. A *GJA3* c.43C>A p.(Gln15Lys) variant was observed in singleton CSA192 (table 1, figure 3C). This glutamine residue is in a very highly conserved region of the GJA3 protein, with only asparagine and arginine alternatively observed at this position in the softshell turtle and tetraodon fish, respectively. The CTAS71 and CSA192 variants are likely de novo but require parental analysis.

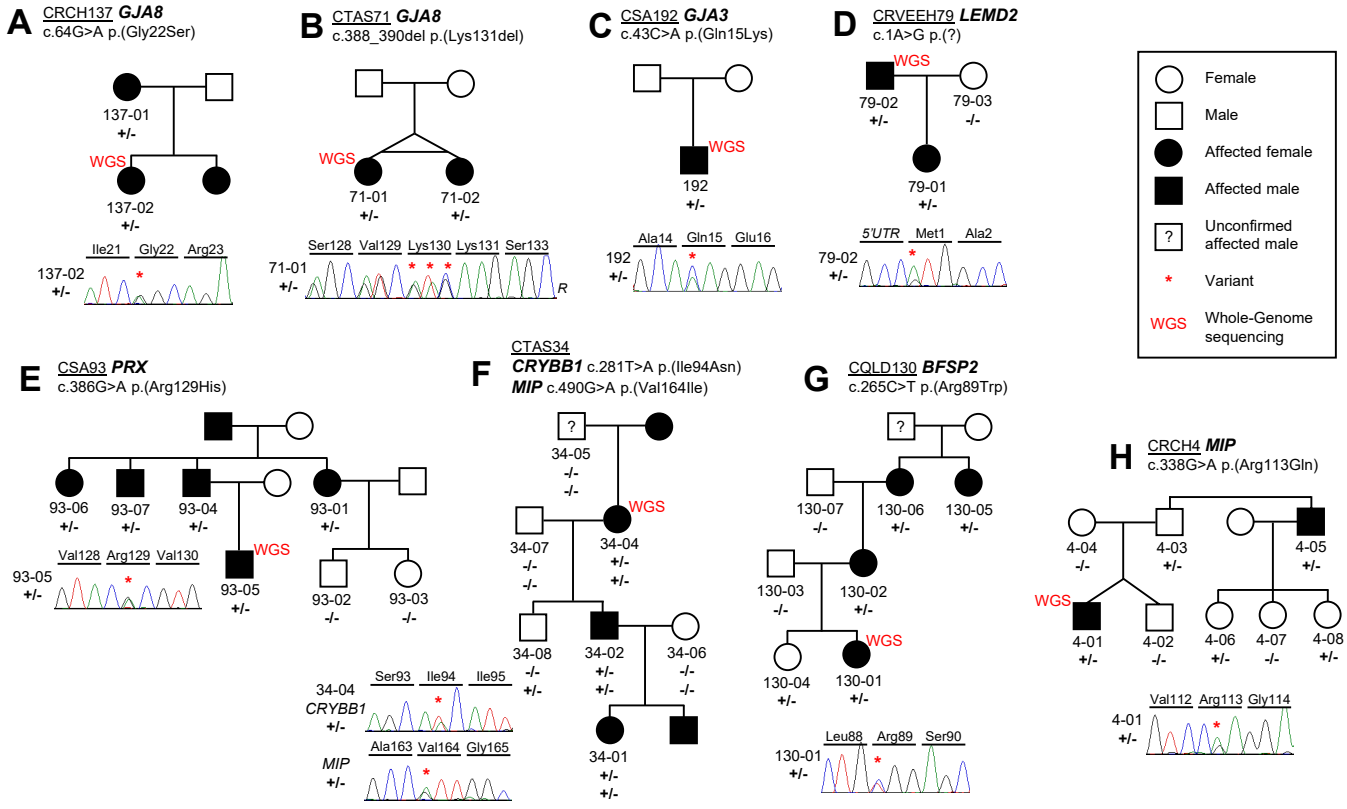

Figure 3 Families with variants of uncertain significance that have evidence towards pathogenicity. Families with gap junction variants include: (A) *GJA8* p.(Gly22Ser) in family CRCH137, (B) *GJA8* p.(Lys131del) in family CTAS71 shown in 'R' for sequencing with reverse primer, and (C) *GJA3* p.(Gln15Lys) in family CSA192. Five families were observed to have variants of uncertain significance in other cataract associated genes. (D) family CRVEEH79 with a start loss variant in the *LEMD2* gene. (E) Family CSA93 with a segregating *PRX* p.(Arg129His) change. (F) Family CTAS34 with a segregating *CRYBB1* p.(Ile94Asn) variant and *MIP* variant was also observed but also present in an unaffected individual. (G) family CQLD130 with a *BFSP2* p.(Arg89Trp) variant that is also observed in unaffected CQLD130-04 and is acting with possible reduced penetrance. (H) In family CRCH4 a *MIP* p.(Arg113Gln) variant was observed in the two affected individuals and obligate heterozygote CRCH4-03, as well as, unaffected siblings CRCH4-06 and CRCH4-08.

In family CRVEEH79, affected individuals were heterozygous for a start-loss variant in the *LEMD2* gene (table 1, figure 3D). Both individuals have mild blue dot cataracts that have not yet required surgery. The *LEMD2* c.1A>G variant was predicted to use an alternative methionine at position 233 in the native protein sequence, resulting in the loss of the conserved LEM domain, lamin A/C complex interacting region and one of two transmembrane domains. Alternatively, a methionine in a different reading frame closer to the 5'-untranslated region could be recruited and produce an 86 amino acid long protein or transcript likely to be subject to nonsense-mediated decay, which would result in a null allele.

The *PRX* c.386G>A variant observed to cosegregate in family CSA93 (table 1, figure 3E) causes a p.(Arg129His) change in the S-periaxin encoding NM_020956.2 transcript only, and is located in the intron of the transcript encoding the L-periaxin isoform c.381+5G>A (NM_181882.3). The variant causes the non-conserved arginine to be replaced with a smaller but still positively charged histidine residue at the C-terminal end of the translated protein.

A *CRYBB1* p.(Ile94Asn) change was identified in CTAS34 that likely accounts for their disease (table 1, figure 3F). Three residue types are observed across species at this site (Ile, Leu, Val), all of which are nonpolar compared with the polar asparagine reported in this family. Located in the first of four Greek Key motifs, the incorporation of a polar residue at this site was predicted to disrupt the hydrophobic interactions in the core of the protein.[13] Cataract-causing variants have been reported at adjacent residues p.Ser93[29] and p.Val96[30] indicating a region of functional importance. Additional segregation evidence from other known affected family members would be highly valuable and would assist in upgrading the classification of this variant. The *MIP* p.(Val164Ile) variant in the same family was also classified as a VUS, however, based on its higher population allele frequency, non-damaging in silico predictions and presence in an unaffected individual it is unlikely to be disease-causing.

In family CQLD130, the *BFSP2* p.(Arg89Trp) variant was observed in four affected individuals and a child who was unaffected as the 10 years old at last examination, but not yet old enough to be confirmed as unaffected

for a childhood onset disease (table 1, figure 3G). This residue change occurs within an evolutionary constrained block of the phakinin protein's N-terminus head region (amino acid 1–114) prior to the main α-helical rod that forms the majority of the 312 amino-acid-long structure. At this site arginine is observed in most species; however, tryptophan has been observed in some species. Changes to this *BFSP2* VUS classification will depend on future surveillance of CQLD130-04 for cataract development or additional reports in unrelated cataract patients.

In family CRCH4, the identified *MIP* p.(Arg113Gln) variant was assessed as being potentially disease-causing (table 1, figure 3H). The p.Arg113 residue is highly conserved across species, with positively charged residues in this extracellular domain known to be functionally important for AQP0 in cell–cell adhesion.[31] Reduced penetrance is observed in this family with obligate heterozygote CRCH4-03 showing no clinically significant opacities. Affected individual CRCH4-01 was diagnosed at birth and received surgery within a month, whereas CRCH4-05 was diagnosed at 2 years of age and did not require surgery until 38 years of age. The two other variant carriers, CRCH4-06 and CRCH4-08, were 7 and 4.5 years of age at last examination, respectively, at which time both still had clear lenses. This family is being screened regularly to assess for cataract development in unaffected individuals.

VUS that are unlikely to be disease-causing

Five families had identified variants classed as VUS unlikely to be causing disease (14%, table 1, online supplemental figure S3). In family CSA158, a *CYP51A1* p.(Ala94Thr) variant was the only variant observed to fully cosegregate with disease (online supplemental figure S3A) but is inconsistent with the recessive inheritance patterns that have been previously reported with variants in this gene and other cataract-associated genes involved in the cholesterol biosynthesis pathway. The primary evidence against pathogenicity in the remaining four families was poor cosegregation of the variant with disease in the context of reported inheritance patterns for the gene, population allele frequency and in silico predictions (online supplemental figure S3B-E).

DISCUSSION

A comprehensive selection of 63 isolated cataract-associated genes in 37 Australian families were investigated. Disease-causing likely pathogenic variants were identified in eight families in genes *GJA3, GJA8, CRYAA, BFSP1, PITX3, COL4A1* and *HSF4*. An additional eight families were identified to have VUS with evidence towards pathogenicity. The solved rate for this cohort resides between 22% and 43%, with many of the identified VUS likely to be disease-causing with the acquisition of additional evidence. With a subset of patients previously cleared of variants in cataract genes there was a reduced likelihood of achieving a solved rate comparable to screening a previously unstudied population. Despite this, these rates are not dissimilar to the 42% likely disease-causing in the previous study of our repository[19] and the recent 44.4% molecular diagnostic rate reported for clinical congenital cataract screening in the UK.[32]

Variants in connexin and crystallin genes again accounted for approximately half of disease-causing variants in Australian cohorts.[4 19] These gene products play critical and well-characterised roles in maintaining lens homeostasis and creating a high protein content that aids in achieving lens transparency.

Classification of the *COL4A1* p.(Gly720Asp) change as likely pathogenic, in family CRCH38, was greatly assisted by its previous observation in a family with a congenital cataract phenotype.[33 34] All individuals with the variant in that family presented with congenital cataract that were accompanied by a range of ophthalmological features but also leukoencephalopathy and stroke in some individuals,[33] which may have important health implications for our family.

The start-loss variant in the *LEMD2* gene presents an interesting finding in family CRVEEH79. *LEMD2* was only recently confirmed as a cataract gene, following the identification of a p.(Leu13Arg) change in the LEM domain in families with autosomal recessive juvenile cataracts in the Hutterite community of North America.[35] This previous report of the *LEMD2* gene in relation to paediatric cataract by Boone *et al*[35] found an additional relationship between variant carriers and sudden onset cardiac death, which may be pertinent to other individuals with *LEMD2* variants. However, autosomal dominant disease has not previously been reported with this gene and functional investigation of the potential for cataract development with this start-loss variant is needed. Increased screening of this gene in cataract patients will also assist in informing inheritance trends and genotype-phenotype correlations.

The findings in family CSA93, with the c.386G>A p.(Arg129His) change in *PRX* must be interpreted with caution. Impacting the coding region of only the S-periaxin isoform, this variant would greatly value independent confirmation of pathogenic changes to lens function and cataract formation. While the PRX protein has been shown as important for lens fibre cell structure in mice[36] the lack of cataract development and accompanying neurological features indicate this gene is still lacking the key evidence needed to confirm it causes congenital cataracts alone. This is further supported by *PRX* variants more often causing Charcot-Marie-Tooth disease (MIM:614895) and Dejerine-Sottas (MIM:145900), which are recessively inherited neurological conditions impacting the peripheral nervous system.

Variant interpretation continues to evolve with improved uptake in the reporting of cataract variants. Subsequently, our understanding of the disease-causing capacity of variants in cataract genes and their functional regions also continues to improve. This information critically underpins variant interpretation of newly identified variants using criteria such as the ACMG-AMP

guidelines.[15] Due to the rarity of the disease and the unique nature of the variants identified it would be expected that many families display variants classified as being of uncertain significance. This is compounded further by the breadth of genes associated with paediatric cataract that contribute to additional ocular and syndromic phenotypes. We have worked to stringently apply variant classification criteria and subclassify VUS based on their collective supporting evidence of pathogenicity. This has clearly identified those with the potential to reach likely pathogenic (or pathogenic) classification and are likely to account for the cataracts observed. Small family sizes frequently limited the use of cosegregation as a pathogenic evidence, but with variant reporting in databases such as ClinVar this may enable future reclassification with additional observations. The identification of these VUS in less-established cataract genes such as *LEMD2* and *PRX* highlights the work that remains, in the research setting, to better understand the role they play in cataractogenesis. For VUS identified in well-established cataract genes, such as the connexins, a move towards gene-specific variant classification criteria for isolated paediatric cataracts would be advantageous, as would establishing functional assays for routinely assessing the functional effects of novel variants.

Of the subset of families being reassessed following the previous screen of 51 cataract genes, all variants identified in those probands were in genes not previously assessed, with the exception of the likely pathogenic *HSF4* p.(Lys64Glu) variant in CSA168-01. All those variants received a VUS classification based on restrictions including inappropriate segregation or current limitations to our understanding of the gene, such as the capacity for autosomal dominant cataracts with variants in the *LEMD2* gene. These do, however, provide informative observations that may inform our future understanding of these genes. The genome sequencing data will allow for periodic reassessment of newly identified genes and further evaluation of copy number and non-coding variants. Currently, there are conflicting reports of increased molecular diagnostic rates when using genome sequencing in congenital cataract cases.[4 32] Our data currently indicate marginal difference with variants that would be equally identified with exome sequencing or a targeted gene panel. However, the full extent of the genomes sequencing benefits will be best measured in the coming years following the application of routine rescreening of genomic data in unsolved cases and an improved ability to identify and correctly interpret non-coding variants as pathogenic.

Author affiliations
[1]Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia
[2]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia
[3]Department of Ophthalmology, Royal Children's Hospital, Melbourne, Victoria, Australia
[4]Department of Ophthalmology, Flinders University, Flinders Medical Centre, Bedford Park, South Australia, Australia
[5]Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia
[6]Centre for Ophthalmology and Visual Science, University of Western Australia, Lions Eye Institute, Perth, Western Australia, Australia
[7]Ophthalmology Department, Women's and Children's Hospital, Adelaide, South Australia, Australia

Acknowledgements We would like to thank the research participants involved in this study. We would like to thank Illumina and the Ramaciotti Centre for Genomics for the genome sequencing data obtained following receipt of the 2019 Genetic Society of Australasia student research grant schemes 'Illumina NovaSeq Mini Grant'.

Contributors Conceptualisation, KPB, JEC, DAM and JLJ; Methodology, JLJ, BJM, KPB and JCC; Software, BJM and JCC; Validation, JLJ; Formal analysis, JLJ and BJM; Investigation, JLJ, SES, ES and LSK; resources, JEC, DAM, JBR, JEE, DT, JP and TC; Data curation, BJM; Writing—original draft preparation, JLJ, KPB; Writing—review and editing, all; Visualisation, JLJ; Supervision, KPB and DAM; Project administration, ES, SES and LSK; Funding acquisition, KPB, JBR and DAM. Guarantor, KPB. All authors have read and agreed to the published version of the manuscript.

Funding This work was supported by a National Health and Medical Research Council (NHMRC) of Australia Ideas grant (GNT1185477) and Centre for Research Excellence grant (GNT1116360). JLJ was supported by the Australian Government Research Training Program and KPB was supported by a NHMRC Senior Research Fellowship. The Centre for Eye Research Australia receives Operational Infrastructure Support from the Victorian Government.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The study was 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). This study was approved by the Tasmania Health and Medical Human Research Ethics Committee (H0014539 14 Nov 2014 and 23875 12 Dec 2020), the Southern Adelaide Clinical Human Research Ethics Committee (3-07 19 Jul 2013 and 29 Nov 2019) and the Royal Victorian Eye and the Ear Hospital Human Research Ethics Committee (01/432H/11 6 Mar 2014). Study participants were recruited after they or their guardian gave written informed consent. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. All variants reported in this study have been submitted to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/); ClinVar accession numbers SCV001573165-SCV001573189. Additional data can be made available on request if within the ethical bounds of this study.

ORCID iDs
Johanna L Jones http://orcid.org/0000-0002-6475-9134
Bennet J McComish http://orcid.org/0000-0003-1940-0007
Sandra E Staffieri http://orcid.org/0000-0003-3131-9359

REFERENCES

1 Wirth MG, Russell-Eggitt IM, Craig JE, *et al*. Aetiology of congenital and paediatric cataract in an Australian population. *Br J Ophthalmol* 2002;86:782–6.
2 Shiels A, Bennett TM, Hejtmancik JF. Cat-map: putting cataract on the map. *Mol Vis* 2010;16:2007–15.
3 Li D, Wang S, Ye H, *et al*. Distribution of gene mutations in sporadic congenital cataract in a Han Chinese population. *Mol Vis* 2016;22:589–98.
4 Ma A, Grigg JR, Flaherty M, *et al*. Genome sequencing in congenital cataracts improves diagnostic yield. *Hum Mutat* 2021;42:1173–83.
5 Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 2009;25:1754–60.
6 DePristo MA, Banks E, Poplin R, *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
7 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
8 Ewels P, Magnusson M, Lundin S, *et al*. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;32:3047–8.
9 Karczewski KJ, Francioli LC, Tiao G, *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
10 Ye J, Coulouris G, Zaretskaya I, *et al*. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 2012;13:134.
11 Altschul SF, Madden TL, Schäffer AA, *et al*. Gapped blast and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402.
12 Bernhofer M, Dallago C, Karl T, *et al*. PredictProtein - Predicting Protein Structure and Function for 29 Years. *Nucleic Acids Res* 2021;49:W535–40.
13 Venselaar H, Te Beek TAH, Kuipers RKP, *et al*. Protein structure analysis of mutations causing inheritable diseases. an e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 2010;11:548.
14 Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;31:3406–15.
15 Richards S, Aziz N, Bale S, *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. *Genet Med* 2015;17:405–24.
16 Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am J Hum Genet* 2017;100:267–80.
17 Jarvik GP, Browning BL. Consideration of cosegregation in the pathogenicity classification of genomic variants. *Am J Hum Genet* 2016;98:1077–81.
18 Biesecker LG, Harrison SM, ClinGen Sequence Variant Interpretation Working Group. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med* 2018;20:1687–8.
19 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* 2017;7:3257–68.
20 Burdon KP, McKay JD, Sale MM, *et al*. Mutations in a novel gene, NHS, cause the pleiotropic effects of Nance-Horan syndrome, including severe congenital cataract, dental anomalies, and mental retardation. *Am J Hum Genet* 2003;73:1120–30.
21 Dave A, Laurie K, Staffieri SE, *et al*. Mutations in the EphA2 gene are a major contributor to inherited cataracts in south-eastern Australia. *PLoS One* 2013;8:e72518.
22 Burdon KP, Wirth MG, Mackey DA, *et al*. Investigation of crystallin genes in familial cataract, and report of two disease associated mutations. *Br J Ophthalmol* 2004;88:79–83.
23 Berry V, Pontikos N, Moore A, *et al*. A novel missense mutation in HSF4 causes autosomal-dominant congenital lamellar cataract in a British family. *Eye* 2018;32:806–12.
24 Hansen L, Yao W, Eiberg H, *et al*. Genetic heterogeneity in microcornea-cataract: five novel mutations in CRYAA, CRYGD, and GJA8. *Invest Ophthalmol Vis Sci* 2007;48:3937–44.
25 Fernández-Alcalde C, Nieves-Moreno M, Noval S, *et al*. Molecular and genetic mechanism of non-syndromic congenital cataracts. mutation screening in Spanish families. *Genes* 2021;12. doi:10.3390/genes12040580. [Epub ahead of print: 16 04 2021].
26 Ding X, Wang B, Luo Y, *et al*. A novel mutation in the connexin 46 (GJA3) gene associated with congenital cataract in a Chinese pedigree. *Mol Vis* 2011;17:1343–9.
27 Wang X, Wang D, Wang Q, *et al*. Broadening the mutation spectrum in GJA8 and CHMP4b: novel missense variants and the associated phenotypes in six Chinese Han congenital cataracts families. *Front Med* 2021;8:713284.
28 Ye Y, Wu M, Qiao Y, *et al*. Identification and preliminary functional analysis of two novel congenital cataract associated mutations of cx46 and Cx50. *Ophthalmic Genet* 2019;40:428–35.
29 Jin A, Zhang Y, Xiao D, *et al*. A novel mutation p.S93R in CRYBB1 associated with dominant congenital cataract and microphthalmia. *Curr Eye Res* 2020;45:483–9.
30 Reis LM, Tyler RC, Muheisen S, *et al*. Whole exome sequencing in dominant cataract identifies a new causative factor, CRYBA2, and a variety of novel alleles in known genes. *Hum Genet* 2013;132:761–70.
31 Varadaraj K, Kumari SS. Molecular mechanism of aquaporin 0-induced fiber cell to fiber cell adhesion in the eye lens. *Biochem Biophys Res Commun* 2018;506:284–9.
32 Jackson D, Malka S, Harding P, *et al*. Molecular diagnostic challenges for non-retinal developmental eye disorders in the United Kingdom. *Am J Med Genet C Semin Med Genet* 2020;184:578–89.
33 Sibon I, Coupry I, Menegon P, *et al*. Col4A1 mutation in axenfeld-rieger anomaly with leukoencephalopathy and stroke. *Ann Neurol* 2007;62:177–84.
34 Coupry I, Sibon I, Mortemousque B, *et al*. Ophthalmological features associated with COL4A1 mutations. *Arch Ophthalmol* 2010;128:483–9.
35 Boone PM, Yuan B, Gu S, *et al*. Hutterite-type cataract maps to chromosome 6p21.32-p21.31, cosegregates with a homozygous mutation in LEMD2, and is associated with sudden cardiac death. *Mol Genet Genomic Med* 2016;4:77–94.
36 Maddala R, Skiba NP, Lalane R, *et al*. Periaxin is required for hexagonal geometry and membrane organization of mature lens fibers. *Dev Biol* 2011;357:179–90.
37 Bennett TM, Mackay DS, Knopf HLS, *et al*. A novel missense mutation in the gene for gap-junction protein alpha3 (GJA3) associated with autosomal dominant "nuclear punctate" cataracts linked to chromosome 13q. *Mol Vis* 2004;10:376–82.
38 Santhiya ST, Kumar GS, Sudhakar P, *et al*. Molecular analysis of cataract families in India: new mutations in the CRYBB2 and GJA3 genes and rare polymorphisms. *Mol Vis* 2010;16:1837–47.
39 Kircher M, Witten DM, Jain P, *et al*. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5.
40 Rentzsch P, Witten D, Cooper GM, *et al*. Cadd: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 2019;47:D886–94.
41 Adzhubei IA, Schmidt S, Peshkin L, *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
42 Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001;11:863–74.
43 Choi Y, Sims GE, Murphy S, *et al*. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012;7:e46688.

Supplementary data

Table S1 Genome Sequencing QC data

Sample	Total read count	% reads mapped	Average target read coverage	Average sequencing depth ‡	Average read length
CRCH4-01	976.6 M	49.3%	22	18.3 X	148 bp
CRCH5-01 †	933.8 M	98.4%	43	38.0 X	148 bp
CRCH12-01	964.5 M	79.1%	36	29.6 X	148 bp
CRCH14-04	996.1 M	79.4%	37	32.2 X	148 bp
CRCH21-05	518.2 M	81.6%	33	18.3 X	236 bp
CRCH27-01 †	981.5 M	81.5%	37	32.9 X	148 bp
CRCH28-01	965.9 M	95.4%	43	38.8 X	148 bp
CRCH29-03	2072.8 M	50.4%	49	42.7 X	147 bp
CTAS34-04	937.5 M	99.1%	43	38.4 X	148 bp
CRCH38-01	900.9 M	88.4%	37	31.7 X	146 bp
CRCH41-03 †	937.1 M	81.7%	37	32.3 X	147 bp
CRCH65-01	1024.7 M	85.2%	41	38.0 X	147 bp
CTAS71-01	816.4 M	98.1%	37	33.0 X	147 bp
CTAS72-04	992.4 M	99.3%	46	42.4 X	148 bp
CRVEEH77-02	880.7 M	98.6%	40	35.8 X	147 bp
CRVEEH78-01	986.9 M	98.3%	45	41.2 X	148 bp
CRVEEH79-02 †	1003.5 M	98.7%	46	41.3 X	147 bp
CQLD88-04	1016.0 M	90.5%	43	38.2 X	148 bp
CRCH90-02	967.5 M	51.8%	23	19.3 X	147 bp
CSA93-05	973.5 M	98.4%	45	40.7 X	147 bp
CSA100-01 †	964.5 M	98.6%	44	39.5 X	147 bp
CRVEEH113-01	961.8 M	97.4%	44	39.1 X	146 bp
CSA119-05 †	1019.0 M	98.3%	47	40.9 X	147 bp
CSA126-01	568.6 M	99.9%	45	33.7 X	241 bp
CSA128-02 †	942.1 M	97.7%	43	34.3 X	147 bp
CQLD130-01	1964.9 M	69.7%	65	47.5 X	147 bp
CRCH137-02	847.2 M	84.8%	33	29.8 X	148 bp
CSA152-01 †	985.2 M	99.4%	46	40.9 X	148 bp
CSA158-02 †	869.5 M	99.3%	40	37.2 X	148 bp
CSA167	1011.2 M	98.8%	47	42.3 X	148 bp
CSA168-01 †	874.0M	98.7%	40	35.9 X	147 bp
CSA169-02 †	933.9 M	99.7%	43	37.9 X	148 bp
CSA178-02 †	995.0 M	99.3%	46	42.0 X	148 bp
CSA179	569.1 M	99.8%	45	32.9 X	240 bp
CSA181-01 †	930.7 M	98.8%	43	38.6 X	148 bp
CSA182-02	946.6 M	98.7%	44	37.7 X	147 bp
CSA192	906.5 M	99.2%	42	39.2 X	148 bp

MultiQC (v1.9) metrics. M; million. ‡ Average sequencing depth at sites in VCF file. Boxed data points if <30. †family with a proband screened in (Javadiyan et al. 2017).

Supplementary data

Table S2 Paediatric cataract genes selected for screening

Gene	Locus	Inheritance	MIM	References	Additional ocular phenotypes	Other
AGK	7q34	AR	610345	(Aldahmesh et al. 2012a) †		Sengers syndrome (MIM: 212350)
BFSP1	20p12.1	AD/AR	603307	(Ramachandran et al. 2007)		
BFSP2	3q22.1	AD/AR	603212	(Conley et al. 2000) and (Jakobs et al. 2000)		
CHMP4B	20q11.22	AD	610897	(Shiels et al. 2007)		
COL4A1	13q34	AD	120130	(Xia et al. 2014) †		Small-vessel brain disease 1 with or without ocular anomalies (MIM: 175780)
COL4A2	13q34	AD	120090	(Ha et al. 2016)		Brain small vessel disease 2 (BSVD2, [MIM: 614483])
CRYAA	21q22.3	AD/AR	123580	(Litt et al. 1998)	Microcornea, iris coloboma, nystagmus, microphthalmia	
CRYAB	11q23.1	AD/AR	123590	(Berry et al. 2001) †		Myofibrillar myopathy (MFM2, [MIM: 608810])
CRYBA1	17q11.2	AD/AR	123610	(Kannabiran et al. 1998)	Nystagmus	
CRYBA2	2q35	AD	600836	(Reis et al. 2013)		
CRYBA4	22q12.1	AD/AR	123631	(Billingsley et al. 2006)	Microphthalmia, microcornea	
CRYBB1	22q12.1	AD/AR	600929	(Willoughby et al. 2005)	Microphthalmia, microcornea, nystagmus	
CRYBB2	22q11.23	AD	123620	(Litt et al. 1997)	Microphthalmia, microcornea, strabismus	
CRYBB3	22q11.23	AD/AR	123630	(Riazuddin et al. 2005)	Microcornea	
CRYGA	2q33.3	AD	123660	(Li et al. 2016)		
CRYGB	2q33.3	AD	123670	(AlFadhli et al. 2012)		

Supplementary data

CRYGC	2q33.3	AD	123680	(Heon et al. 1999)	Microcornea, glaucoma, microphthalmia, nystagmus	
CRYGD	2q33.3	AD	123690	(Stephan et al. 1999) and (Heon et al. 1999)	Nystagmus, microcornea	
CRYGS	3q27.3	AD	123730	(Sun et al. 2005)	Lens subluxation	
CTDP1	18q23	AR	604927	(Tzifi et al. 2011)		Congenital cataract, facial dysmorphism and neuropathy (CCFDN, [MIM: 604168])
CYP27A1	2q35	AR	606530	(Khan et al. 2015)		Cerebrotendinous xanthomatosis (MIM: 21370)
CYP51A1	7q21.2	AR	601637	(Aldahmesh et al. 2012b)		
DNMBP	10q24.2	AR	611282	(Ansar et al. 2018)	Nystagmus, amblyopia, exotropia	
EPHA2	1p36.13	AD/AR	176946	(Shiels et al. 2008)	Posterior lenticonus	
EYA1	8q13.3	AD	601653	(Azuma et al. 2000) †	Persistence of pupillary membrane, corneal opacity	Branchio-oto-renal syndrome with or without cataract (BOR1, [MIM:113650])
FBN1	15q21.1	Sporadic	134797	(Li et al. 2016) †		Marfan syndrome (MIM: 154700), Weill-Marchesani syndrome (WMS2 [MIM: 608328])
FOXE3	1p33	AD/AR	601094	(Semina et al. 2001)	Anterior segment dysgenesis 2 (ASGD2, [MIM: 610256])	
FTL	19q13.33	AD	134790	(Girelli et al. 1995) and (Beaumont et al. 1995)		Hyperferritinemia with or without cataract (HHCS, [MIM: 600886])
FYCO1	3p21.31	AR	607182	(Chen et al. 2011)		

Supplementary data

GALK1	17q25.1	AR	604313	(Stambolian et al. 1995)		Galactokinase deficiency with cataract (MIM: 230200)
GCNT2	6p24.3-p24.2	AR	600429	(Yu et al. 2001)	Nystagmus	Adult i blood group with cataract (MIM: 110800)
GJA3	13q12.11	AD	121015	(Mackay et al. 1999)		
GJA8	1q21.2	AD/AR	600897	(Shiels et al. 1998)	Microphthalmia, nystagmus, secondary glaucoma microcornea, corneal opacity, sclerocornea, coloboma	
HSF4	16q22.1	AD/AR	602438	(Bu et al. 2002)	Nystagmus	
IARS2	1q41	AR	612801	(Li et al. 2018) †		Cataracts, growth hormone deficiency, sensory neuropathy, sensorineural hearing loss and skeletal dysplasia (CAGSSS, [MIM: 616007])
LEMD2	6p21.31	AR	616312	(Boone et al. 2016)		
LIM2	19q13.41	AR	154045	(Pras et al. 2002)	Nystagmus, amblyopia	
LONP1	19p13.3	AR	605490	(Khan et al. 2015)		CODAS syndrome (MIM: 600373)
LSS	21q22.3	AR	600909	(Zhao et al. 2015)		
MAF	16q23.2	AD	177075	(Jamieson et al. 2002) †	Microcornea, iris coloboma, amblyopia	Ayme-Gripp syndrome (MIM: 601088)
MIP	12q13.3	AD	154050	(Berry et al. 2000)	Nystagmus, strabismus	
MIR184	15q25.1	AD	613146	(Iliff et al. 2012)	EDICT syndrome (MIM: 614303)	
NHS	Xp22.2-p22.1	XL	300457	(Coccia et al. 2009) †	Microcornea, nystagmus, secondary glaucoma, strabismus	Nance-Horan syndrome (MIM: 302350)
PANK4	1p36.32	AD	606162	(Sun et al. 2019)	Nystagmus	

Supplementary data

PAX6	11p13	AD	607108	(Glaser et al. 1994)	Aniridia (MIM: 106210), nystagmus, corneal abnormalities, glaucoma	
PEX11B	1q21.1	AR	603867	(Taylor et al. 2017)		Peroxisome biogenesis disorder 14B (PBD14B, [MIM: 614920])
PITX3	10q24.32	AD/AR	602669	(Semina et al. 1998)	Anterior segment dysgenesis 1 (ASGD1, [MIM: 107250])	
PRX	19q13.2	AD	605725	(Yuan et al. 2016)	Amblyopia	
LOC105378949	1p36.33	AD	-	(Eiberg et al. 2019)		
RRAGA	9p22.1	AD	612194	(Chen et al. 2016)		
SLC16A12	10q23.31	AD	611910	(Kloeckener-Gruissem et al. 2008)	Microcornea	Glucosuria
SLC40A1	2q32.2	AD	604653	(Yamakawa et al. 2016)		Hemochromatosis (MIM: 606069)
SLC7A8	14q11.2	AR	604235	(Knopfel et al. 2019)		
SIPA1L3	19q13.1-q13.2	AR	616655	(Greenlees et al. 2015) and (Evers et al. 2015)	Corneal clouding, microphthalmia, iridocorneal and lenticular adhesions	
TDRD7	9q22.33	AR	611258	(Lachke et al. 2011)		
TMEM114	16p13.2	AD	611579	(Jamieson et al. 2007)		
TRPM3	9q21.12-q21.13	AD	608961	(Bennett et al. 2014)	Open-angle glaucoma	
UNC45B	17q12	AD	611220	(Hansen et al. 2014)		
VIM	10p13	AD	193060	(Muller et al. 2009)		
VSX2	14q24.3	AR	142993	(Percin et al. 2000)	Microphthalmia, iris coloboma, anophthalmia	
WDR36	5q22.1	Sporadic	609669	(Li et al. 2016)		
WDR87	19q13.13	AR	-	(Khan et al. 2015)		

Supplementary data

WFS1	4p16.1	AD	606201	(Berry et al. 2013) †	Iris coloboma	Wolfram syndrome 1 (WFS1, [MIM: 222300])

Inheritance; 'AD' autosomal dominant, 'AR' autosomal recessive, 'XL' X-linked, and 'sporadic' if only sporadic non-syndromic case has been reported in Cat-Map prior to publication (Shiels et al. 2010). MIM; gene reference. '-' data unavailable. Reference; initial gene reference. † reference for report of non-syndromic occurrence when gene is primarily known to cause syndromic congenital cataracts. Additional ocular phenotypes; as reported in Cat-Map (Shiels et al. 2010). Other; report of syndrome or condition associated with this gene.

Supplementary data

Table S3 Primers use for reported variant validation and segregation analysis

Family ID	Gene/variant	Primer pair (5' to 3')	Anneal temperature (°C)
CRCH4	*MIP* p.(Arg113Gln)	CCACCTGTCAATCCTCACCA TGTTCTGCAGGTGGCTATGG	57
CRCH5	*LONP1* p.(Pro553Ser)	AATGGGAATGGCTTTGGGGT TACAAGATTGTCAGCGGCGA	59
CRCH21	*GJA3* p.(Pro59Leu)	CTCTTCCATGCGCACGATGT GGAATCTGAAGCAATGGGCG	60
CRCH28	*EYA1* p.(Ser487Leu)	AATGCTGGGATGAGCTGAGTAG TAAATCCTCAGGTCTGCTTGG	57
	PITX3 p.(Ter303LeuextTer100)	TGAAAACGAGGGAGGGGAAG ACCCGTGTAACTCGAGCCT	61
CRCH29	*CRYAA* p.(Arg12Cys)	CTCCATTCTGCTGGTGGCA CAAGACCAGAGTCCATCGCT	59
CTAS34	*MIP* p.(Val164Ile)	CTACCTTGGGGTCAAGAAGGA CTTGAGGAGGTAACACTGTGGC	57
	CRYBB1 p.(Ile94Asn)	ATTTCTCCAGAGCCCAGAACCA GGATGGGAGGACAGGATCATT	57
CRCH38	*COL4A1* p.(Gly720Asp)	ATGTCCTGGGACGTTCACAA AAGTGGGGAACGGCATTGTA	57
	MIR184 n.52T>C	CCGGGAAATCAAACGTCCAT AACGCCAGGTTTTCCCCATC	57
CTAS71	*GJA8* p.(Lys131del)	CATGGAGGAGAAGCGCAAAAG GAAGTAGTGGCCCACGATGA	57
CRVEEH77	*GJA8* p.(Pro189Ser)	ACCCTGCTGAGGACCTACAT GACACAGAGGCCACAGACAA	57
CRVEEH79	*LEMD2* p.(?)	TTGTTGCGGTAGACATCCCG AAAGGCCAAGTGCAGACCTT	57
CQLD88	*FBN1* p.(Pro1141Leu)	GCTTCCAACTTTGGCAATGA GAGGCCCCCACCTTTAACAT	57
CRCH90	*GJA3* p.(Thr19Met)	TGTCGTAGCAGACGTTCTCG CCCGGTGTTCATGAGCATTT	57
CSA93	*PRX* p.(Arg129His)	AGGGGCAGAGGGTGGAATTA ATGCGCCGAGCCTTACAAAG	57
CSA100	*LSS* p.(Leu78Val)	AGTGGGCCACCATAATCACC TTGGGCTGTATGTGAAGAGGG	57
CQLD130	*BFSP2* p.(Arg89Trp)	CTCCAGGACCAATGCCATGAG TGTTTCCAGCTCCTGACTGAC	57
CRCH137	*GJA8* p.(Gly22Ser)	TCGGGGCCTTCTTTGTTCTC GCGAATGTGGGAGATGGGAA	57
CSA158	*WFS1* p.(Ala370Val)	CCCACGCACCACATCAAC CATAGGGCTCCAGGTGGTTC	63
	CYP51A1 p.(Ala94Thr)	ACCCCAGGACATGGGAAAAG GGTCATGAAACGAAACTGGG	60
CSA168	*HSF4* p.(Lys64Glu)	TGGTAGAGCGGGACCAGTTT CACCTTCCGAAAACCGTCTG	63
CSA178	*IARS2* p.(Gly389Ala)	CCCCACAGGTGTAGATTTGGA CAGTACCATGGGCAGGTTGT	57
CSA182	*BFSP1* p.(Glu375GlyfsTer2)	ATCCTCTGGAGCCCCTTCTT GCCTATTTTCCAACCAGCGT	57
CSA192	*GJA3* p.(Gln15Lys)	TGTCGTAGCAGACGTTCTCG CCCGGTGTTCATGAGCATTT	57

+Q; Optimized with Qiagen Q solution

Supplementary data

Table S4 Evidence for pathogenicity according to ACMG-AMP guidelines.

Family ID	Gene	Variant	Supporting evidence for pathogenicity	Supporting evidence for benign	ACMG-AMP	ClinVar
Pathogenic/Likely pathogenic						
CRCH21	*GJA3*	(NC_000013.10)Chr13:g.20717252G>A NM_021954.4:c.176C>T p.(Pro59Leu)	PM2, PP1_strong, PP3 Absent/near absent from population databases. Same variant previously established as pathogenic variant (report with cataract PMID: 15208569, 19182255, 21866213, 25148791, 26694549, 27609163), segregation (strong) with the disease in three families with 19 meioses (this study and PMID: 15208569, 27609163). Multiple predictive tools assessing variant as damaging/pathogenic.	-	LP	SCV001573165
CRCH90	*GJA3*	(NC_000013.10)Chr13:g.20717372G>A NM_021954.4:c.56C>T p.(Thr19Met)	PM2, PP1_strong, PP3 Absent/near absent from population databases. Same variant has been reported as pathogenic (report with cataract PMID: 28839118, 21031021, 29461512), segregation (strong) with the disease in four families with five meioses (this study and PMID: 28839118, 21031021, 29461512). Multiple predictive tools assessing variant as damaging/pathogenic.	-	LP	SCV001573166
CRCH29	*CRYAA*	(NC_000021.8)Chr21:g.44589243C>T NM_000394.4:c.34C>T p.(Arg12Cys)	PM2, PP1_strong, PP3 Absent/near absent from population databases. Same variant previously established pathogenic (report with cataract PMID: 23508780, 19503744, 17724170, 18587492, 19390652, 21686328, 30078984, 32010934), segregation (strong) with the disease in nine families with 20 meioses (this study and PMID:	-	LP	SCV001573167

Supplementary data

			23508780, 19503744, 17724170, 18587492, 19390652, 21686328, 30078984, 32010934). Multiple predictive tools assessing variant as damaging/pathogenic.			
CRCH38	*COL4A1*	(NC_000013.10)Chr13:g.110833673C>T NM_001845.6:c.2159G>A p.(Gly720Asp)	PM2, PP1_strong, PP3 Absent/near absent from population databases. Alternate source reports variant as pathogenic in family with additional congenital cataract phenotype (PMID: 17696175), segregation (strong) with the disease in two families with 6 meioses (this study and PMID: 17696175). Multiple predictive tools assessing variant as damaging/pathogenic.	-	LP	SCV001573168
	MIR184	(NC_000015.9)Chr15:g.79502181T>C NR_029705.1:n.52T>C	PM2, PP3 Absent/near absent from population databases. Multiple predictive tools assessing variant as damaging/pathogenic.	BP4 RNA folding predictive evidence of non-damaging/benign effect on product.	VUS	SCV001573169
CSA168	*HSF4*	(NC_000016.9)Chr16:g.67199491A>G NM_001040667.3:c.190A>G p.(Lys64Glu)	PM1, PM2, PP1_strong, PP3 Variant located in functional protein region, a highly conserved DNA binding domain with multiple pathogenic variants. Absent/near absent from population databases. Segregation (strong) with the disease in one family previously reported with eight meioses (PMID: 29243736). Multiple predictive tools assessing variant as damaging/pathogenic.	-	LP	SCV001573170
CRCH28	*PITX3*	(NC_000010.10)Chr10:g.103990272C>A NM_005029.4:c.908G>T p.(Ter303LeuextTer100)	PM2, PM4, PP1_supporting, PP3 Absent/near absent from population databases. Protein length altered due to loss of native stop codon. Supporting segregation evidence in the family with ≥3 meioses. Multiple predictive tools assessing variant as damaging/pathogenic.	-	LP	SCV001573171

Supplementary data

	EYA1	(NC_000008.10)Chr8:g.72127864G>A NM_000503.6:c.1460C>T p.(Ser487Leu)	-	BS4, BP4, BS1 Lack of segregation in affected family members, multiple predictive tools assessing the variant as non-damaging/benign. Present in population databases more than expected for the condition.	B	SCV0015 73172
CSA182	*BFSP1*	(NC_000020.10)Chr20:g.17475593del NM_001195.5:c.1124del p.(Glu375GlyfsTer2)	PM2, PM4, PP1_strong, PP3 Absent/near absent from population databases. Protein length changes due to introduction of premature stop codon. PP1_strong segregation evidence with ≥5 meioses in family. Multiple predictive tools assessing variant as damaging/pathogenic.	-	LP	SCV0015 73173
CRVEEH77	*GJA8*	(NC_000001.10)Chr1:g.147380647C>T NM_005267.5:c.565C>T p.(Pro189Ser)	PM2, PM5, PP1_supporting, PP3 Absent/near absent from population databases. Variant at an amino acid where a different missense variant was predicted pathogenic (PMID: 17724170). Segregation (supporting) with the disease in the family with ≥3 meioses. Multiple predictive tools assessing variant as damaging/pathogenic.	-	LP	SCV0015 73174
Variants of uncertain significance with evidence towards pathogenicity						
CRCH137	*GJA8*	(NC_000001.10)Chr1:g.147380146G>A NM_005267.5:c.64G>A p.(Gly22Ser)	PM2, PP3 Absent/near absent from population databases and multiple predictive tools assessing variant as damaging/pathogenic.	-	VUS	SCV0015 73175
CTAS71	*GJA8*	(NC_000001.10)Chr1:g.147380470_147380472del NM_005267.5:c.388_390del p.(Lys131del)	PM2, PM4_supporting, PP3 Absent/near absent from population databases. In-frame deletion changes protein length, but in cytoplasmic region	-	VUS	SCV0015 73176

Supplementary data

			with variable amino acid conservation between species. Multiple predictive tools assessing variant as damaging/pathogenic.			
CSA192	*GJA3*	(NC_000013.10)Chr13:g.20717385G>T NM_021954.4:c.43C>A p.(Gln15Lys)	PM2, PP3 Absent/near absent from population databases and multiple predictive tools assessing variant as damaging/pathogenic.	-	VUS	SCV001573177
CRVEEH79	*LEMD2*	(NC_000006.11)Chr6:g.33756893T>C NM_001348710.2:c.1A>G p.(?)	PM2, PP3 Absent/near absent from population databases. Multiple predictive tools assessing variant as damaging/pathogenic. Note: Variant causes loss of start codon with likely loss of function, but haploinsufficiency has not been shown to cause disease before.	-	VUS	SCV001573178
CSA93	*PRX*	(NC_000019.9)Chr19:g.40904522C>T NM_020956.2:c.386G>A p.(Arg129His)	PM2, PP1_supporting, PP3 Absent/near absent from population databases. Segregation (supporting) with the disease in the family with ≥3 meioses. Multiple predictive tools assessing variant as damaging/pathogenic.	-	VUS	SCV001573179
CTAS34	*CRYBB1*	(NC_000022.10)Chr22:g.27008054A>T NM_001887.4:c.281T>A p.(Ile94Asn)	PM2, PP3 Absent/near absent from population databases and multiple predictive tools assessing variant as damaging/pathogenic.	-	VUS	SCV001573180
	MIP	(NC_000012.11)Chr12:g.56847410C>T NM_012064.4:c.490G>A p.(Val164Ile)	PM2 Variant observed at a low rate in population data.	BP4 Multiple predictive tools assess amino acid change as non-damaging.	VUS	SCV001573181
CQLD130	*BFSP2*	(NC_000003.11)chr3:g.133119192C>T NM_003571.4:c.265C>T p.(Arg89Trp)	PM2, PP1_supporting, PP3 Supporting for absent/near absent from population databases. Segregation (supporting) with the disease in the family with ≥3 meioses. Multiple predictive tools assessing variant as damaging/pathogenic.	-	VUS	SCV001573182

Supplementary data

CRCH4	*MIP*	(NC_000012.11)Chr12:g.56848060C>T NM_012064.4:c.338G>A p.(Arg113Gln)	PM2, PP3 Absent/near absent from population databases. Multiple predictive tools assessing variant as damaging/pathogenic.	BS2 Observed in health adult obligate carrier (and 2 other unaffected children) with full penetrance expected at an early age.	VUS	SCV001573183
Variants of uncertain significance that are unlikely to be causing disease						
CSA158	*CYP51A1*	(NC_000007.13)Chr7:g.91761099C>T NM_000786.4:c.280G>A p.(Ala94Thr)	PM2, PP3 Absent/near absent from population databases and multiple predictive tools assessing variant as damaging/pathogenic.	-	VUS	SCV001573184
	WFS1	(NC_000004.11)Chr4:g.6302631C>T NM_001145853.1:c.1109C>T p.(Ala370Val)	PM2, PP3 Absent/near absent from population databases and multiple predictive tools assessing variant as damaging/pathogenic.	BS4 Lack of segregation in an affected member of the family.	VUS	SCV001573185
CRCH5	*LONP1*	(NC_000019.9)Chr19:g.5694473G>A NM_001276480.1:c.1657C>T p.(Pro553Ser)	PM2, PP3 Absent/near absent from population databases and multiple predictive tools assessing variant as damaging/pathogenic.	BS4 Variant does not segregate with the disease in the family, absent from affected individual.	VUS	SCV001573186
CSA178	*IARS2*	(NC_000001.10)Chr1:g.220279332G>C NM_018060.4:c.1166G>C p.(Gly389Ala)	PM2, PP3 Absent/near absent from population databases and multiple predictive tools assessing variant as damaging/pathogenic.	BS4 Lack of segregation in affected member of family.	VUS	SCV001573187
CQLD88	*FBN1*	(NC_000015.9)Chr15:g.48779550G>A NM_000138.5:c.3422C>T p.(Pro1141Leu)	PP3 Multiple predictive tools assessing variant as damaging/pathogenic.	BS1, BS2 Present in the population database more than expected for the disease. Variant is observed in unaffected individuals in the family.	VUS	SCV001573188

Supplementary data

CSA100	*LSS*	(NC_000021.8)Chr21:g.47647553G>C NM_001001438.3:c.232C>G p.(Leu78Val)	PM2 Absent/near absent from population databases.	BS4, BP4 Lack of segregation with affected members of the family. Multiple predictive tools assess the amino acid change as non-pathogenic.	VUS	SCV001573189

Variant interpretation based on the American College of Medical Genetics and Genomics (ACMG-AMP) guidelines (Richards et al. 2015), 'P' Pathogenic, 'LP' Likely Pathogenic, 'VUS' Variant of Uncertain Significance, and 'B' Benign. ClinVar; variant accession number.

Supplementary data

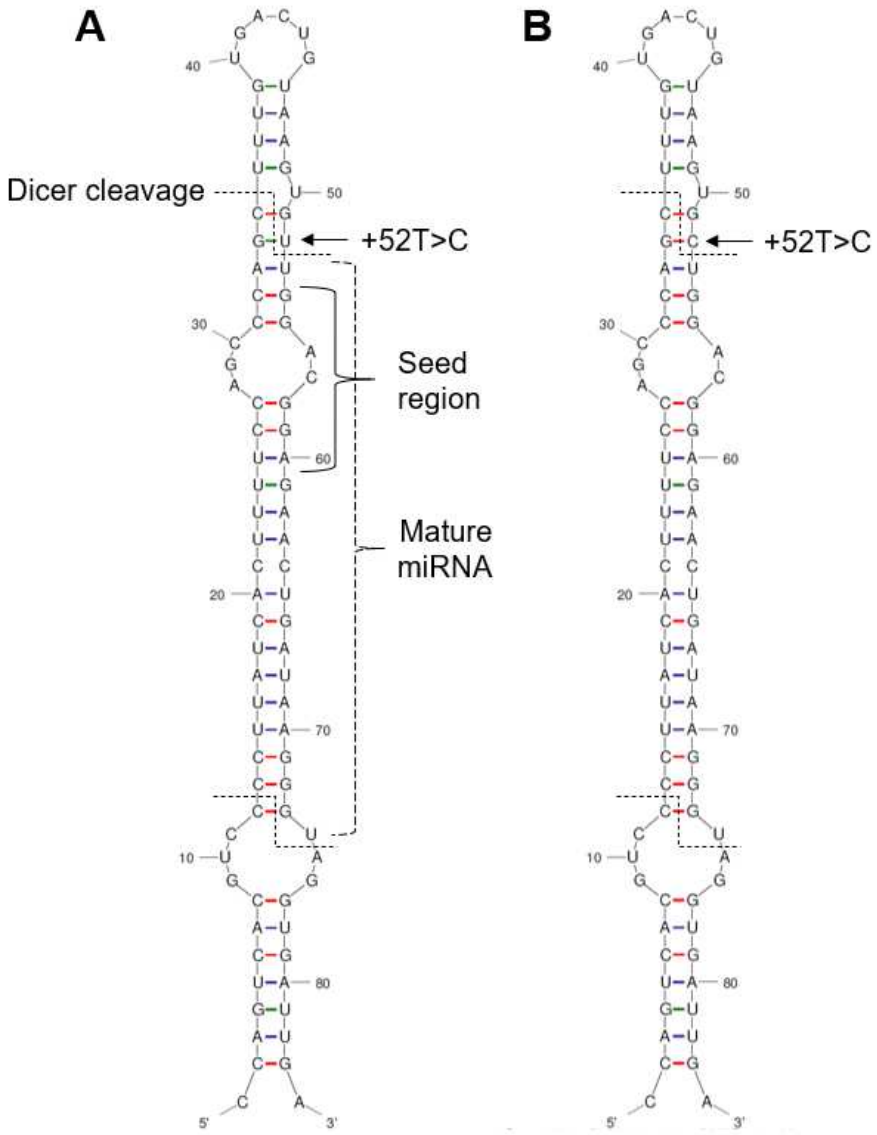

Figure S1 Predicted impact of the NR_029705.1:n.+52T>C variant on miR-184 structure and stability. The *MIR184* variant in family CRCH38 was assessed using mFold (Zuker 2003). No difference in miRNA folding was observed between wild-type miR-184 (**A**) and +52T>C variant carrier miR-184 (**B**). Minimum free energy prediction dropped from -35.50kcal/mol to -37.90kcal/mol with the T>C nucleotide change, increasing the stability of the molecule.

Supplementary data

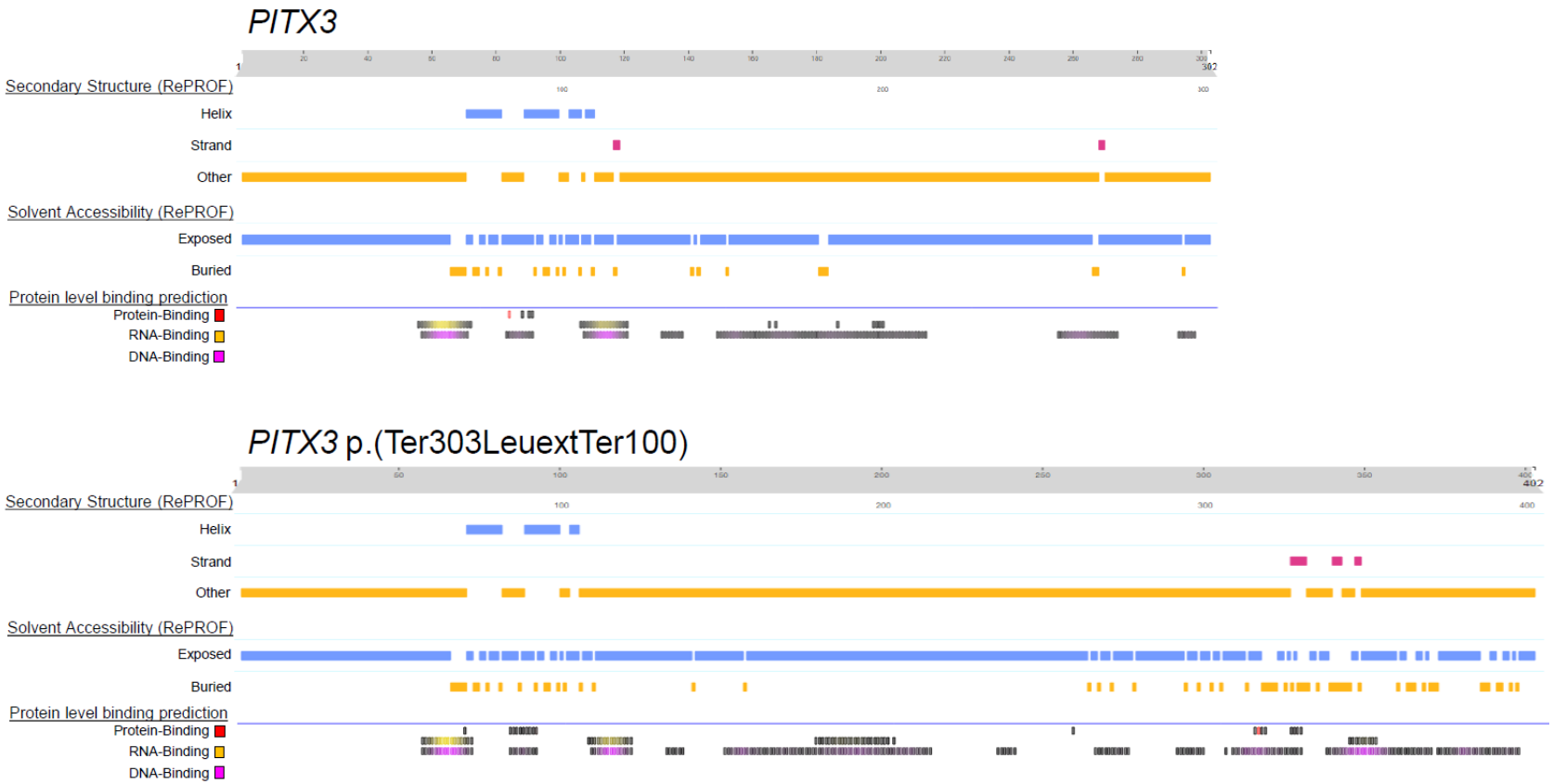

Figure S2 *PITX3* protein folding prediction. Comparison of wild-type *PITX3* and CRCH28 family stop loss p.(Ter303LeuextTer100) *PITX3* variant sequence using PredictProtein folding prediction tool (https://predictprotein.org). Each panel displays predicted secondary structure and solvent accessibility, with protein level binding predictions appended below. *PITX3*_variant sequence has been stretched to enable vertical comparison of the complementary initial 1-302 amino acid positions.

Supplementary data

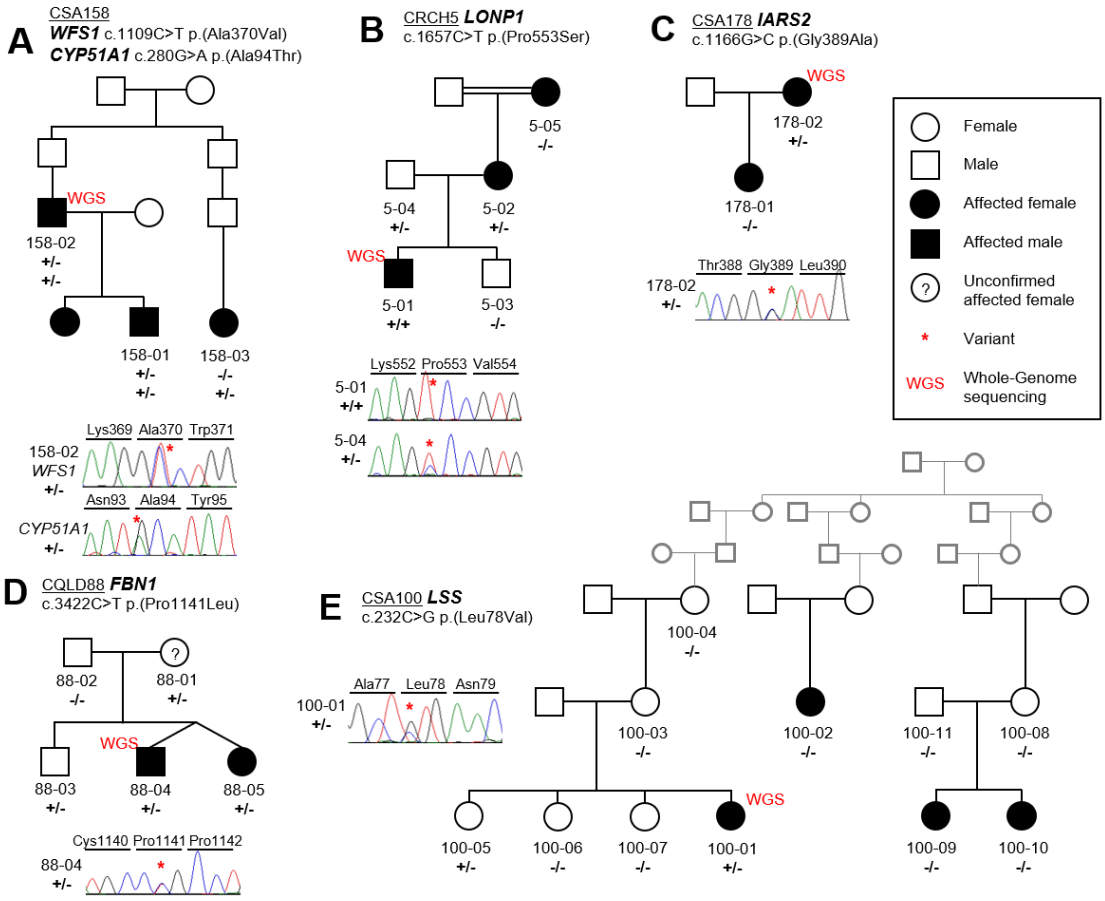

Figure S3 Variants of uncertain significance that are unlikely to be disease causing. Family pedigrees and accompanying sequencing chromatograms from variant analysis. A; Family CSA158 with a segregating *CYP51A1* p.(Ala94Thr) and non-segregating *WFS1* p.(Ala370Val) variants. B; *LONP1* variant enters pedigree twice to enable homozygous status in the proband in family CRCH5. C; In family CSA178, an *IARS2* variant was observed affected mother only. D; In family CQLD88, a *FBN1* variant was observed in an unaffected individual. E; Large pedigree CSA100 displays a *LSS* variant in an affected and unaffected sibling pair, that has entered in the pedigree in the previous generation.

Supplementary data

References:

Aldahmesh MA, Khan AO, Mohamed JY, Alghamdi MH, Alkuraya FS. 2012a. Identification of a truncation mutation of acylglycerol kinase (agk) gene in a novel autosomal recessive cataract locus. Human mutation. 33(6):960-962.

Aldahmesh MA, Khan AO, Mohamed JY, Hijazi H, Al-Owain M, Alswaid A, Alkuraya FS. 2012b. Genomic analysis of pediatric cataract in saudi arabia reveals novel candidate disease genes. Genetics in medicine : official journal of the American College of Medical Genetics. 14(12):955-962.

AlFadhli S, Abdelmoaty S, Al-Hajeri A, Behbehani A, Alkuraya F. 2012. Novel crystallin gamma b mutations in a kuwaiti family with autosomal dominant congenital cataracts reveal genetic and clinical heterogeneity. Molecular vision. 18:2931-2936.

Ansar M, Chung HL, Taylor RL, Nazir A, Imtiaz S, Sarwar MT, Manousopoulou A, Makrythanasis P, Saeed S, Falconnet E et al. 2018. Bi-allelic loss-of-function variants in dnmbp cause infantile cataracts. American journal of human genetics. 103(4):568-578.

Azuma N, Hirakiyama A, Inoue T, Asaka A, Yamada M. 2000. Mutations of a human homologue of the drosophila eyes absent gene (eya1) detected in patients with congenital cataracts and ocular anterior segment anomalies. Human molecular genetics. 9(3):363-366.

Beaumont C, Leneuve P, Devaux I, Scoazec JY, Berthier M, Loiseau MN, Grandchamp B, Bonneau D. 1995. Mutation in the iron responsive element of the l ferritin mrna in a family with dominant hyperferritinaemia and cataract. Nature genetics. 11(4):444-446.

Bennett TM, Mackay DS, Siegfried CJ, Shiels A. 2014. Mutation of the melastatin-related cation channel, trpm3, underlies inherited cataract and glaucoma. PloS one. 9(8):e104000.

Berry V, Francis P, Kaushal S, Moore A, Bhattacharya S. 2000. Missense mutations in mip underlie autosomal dominant 'polymorphic' and lamellar cataracts linked to 12q. Nature genetics. 25(1):15-17.

Berry V, Francis P, Reddy MA, Collyer D, Vithana E, MacKay I, Dawson G, Carey AH, Moore A, Bhattacharya SS et al. 2001. Alpha-b crystallin gene (cryab)

mutation causes dominant congenital posterior polar cataract in humans. American journal of human genetics. 69(5):1141-1145.

Berry V, Gregory-Evans C Fau - Emmett W, Emmett W Fau - Waseem N, Waseem N Fau - Raby J, Raby J Fau - Prescott D, Prescott D Fau - Moore AT, Moore At Fau - Bhattacharya SS, Bhattacharya SS. 2013. Wolfram gene (wfs1) mutation causes autosomal dominant congenital nuclear cataract in humans. (1476-5438 (Electronic)).

Billingsley G, Santhiya ST, Paterson AD, Ogata K, Wodak S, Hosseini SM, Manisastry SM, Vijayalakshmi P, Gopinath PM, Graw J et al. 2006. Cryba4, a novel human cataract gene, is also involved in microphthalmia. American journal of human genetics. 79(4):702-709.

Boone PM, Yuan B, Gu S, Ma Z, Gambin T, Gonzaga-Jauregui C, Jain M, Murdock TJ, White JJ, Jhangiani SN et al. 2016. Hutterite-type cataract maps to chromosome 6p21.32-p21.31, cosegregates with a homozygous mutation in lemd2, and is associated with sudden cardiac death. Molecular genetics & genomic medicine. 4(1):77-94.

Bu L, Jin Y, Shi Y, Chu R, Ban A, Eiberg H, Andres L, Jiang H, Zheng G, Qian M et al. 2002. Mutant DNA-binding domain of hsf4 is associated with autosomal dominant lamellar and marner cataract. Nature genetics. 31(3):276-278.

Chen J, Ma Z, Jiao X, Fariss R, Kantorow WL, Kantorow M, Pras E, Frydman M, Pras E, Riazuddin S et al. 2011. Mutations in fyco1 cause autosomal-recessive congenital cataracts. American journal of human genetics. 88(6):827-838.

Chen JH, Huang C, Zhang B, Yin S, Liang J, Xu C, Huang Y, Cen LP, Ng TK, Zheng C et al. 2016. Mutations of raga gtpase in mtorc1 pathway are associated with autosomal dominant cataracts. PLoS genetics. 12(6):e1006090.

Coccia M, Brooks SP, Webb TR, Christodoulou K, Wozniak IO, Murday V, Balicki M, Yee HA, Wangensteen T, Riise R et al. 2009. X-linked cataract and nance-horan syndrome are allelic disorders. Human molecular genetics. 18(14):2643-2655.

Conley YP, Erturk D, Keverline A, Mah TS, Keravala A, Barnes LR, Bruchis A, Hess JF, FitzGerald PG, Weeks DE et al. 2000. A juvenile-onset, progressive cataract locus on chromosome 3q21-q22 is associated with a missense

mutation in the beaded filament structural protein-2. American journal of human genetics. 66(4):1426-1431.

Eiberg H, Mikkelsen AF, Bak M, Tommerup N, Lund AM, Wenzel A, Sabarinathan R, Gorodkin J, Bang-Berthelsen CH, Hansen L. 2019. A splice-site variant in the lncrna gene rp1-140a9.1 cosegregates in the large volkmann cataract family. Molecular vision. 25:1-11.

Evers C, Paramasivam N, Hinderhofer K, Fischer C, Granzow M, Schmidt-Bacher A, Eils R, Steinbeisser H, Schlesner M, Moog U. 2015. Sipa1l3 identified by linkage analysis and whole-exome sequencing as a novel gene for autosomal recessive congenital cataract. European journal of human genetics : EJHG. 23(12):1627-1633.

Girelli D, Corrocher R, Bisceglia L, Olivieri O, De Franceschi L, Zelante L, Gasparini P. 1995. Molecular basis for the recently described hereditary hyperferritinemia-cataract syndrome: A mutation in the iron-responsive element of ferritin l-subunit gene (the "verona mutation"). Blood. 86(11):4050-4053.

Glaser T, Jepeal L, Edwards JG, Young SR, Favor J, Maas RL. 1994. Pax6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. Nature genetics. 7(4):463-471.

Greenlees R, Mihelec M, Yousoof S, Speidel D, Wu SK, Rinkwitz S, Prokudin I, Perveen R, Cheng A, Ma A et al. 2015. Mutations in sipa1l3 cause eye defects through disruption of cell polarity and cytoskeleton organization. Human molecular genetics. 24(20):5789-5804.

Ha TT, Sadleir LG, Mandelstam SA, Paterson SJ, Scheffer IE, Gecz J, Corbett MA. 2016. A mutation in col4a2 causes autosomal dominant porencephaly with cataracts. American journal of medical genetics Part A. 170a(4):1059-1063.

Hansen L, Comyn S, Mang Y, Lind-Thomsen A, Myhre L, Jean F, Eiberg H, Tommerup N, Rosenberg T, Pilgrim D. 2014. The myosin chaperone unc45b is involved in lens development and autosomal dominant juvenile cataract. European journal of human genetics : EJHG. 22(11):1290-1297.

Heon E, Priston M, Schorderet DF, Billingsley GD, Girard PO, Lubsen N, Munier FL. 1999. The gamma-crystallins and human cataracts: A puzzle made clearer. American journal of human genetics. 65(5):1261-1267.

Supplementary data

Iliff BW, Riazuddin SA, Gottsch JD. 2012. A single-base substitution in the seed region of mir-184 causes edict syndrome. Investigative ophthalmology & visual science. 53(1):348-353.

Jakobs PM, Hess JF, FitzGerald PG, Kramer P, Weleber RG, Litt M. 2000. Autosomal-dominant congenital cataract associated with a deletion mutation in the human beaded filament protein gene bfsp2. American journal of human genetics. 66(4):1432-1436.

Jamieson RV, Farrar N, Stewart K, Perveen R, Mihelec M, Carette M, Grigg JR, McAvoy JW, Lovicu FJ, Tam PP et al. 2007. Characterization of a familial t(16;22) balanced translocation associated with congenital cataract leads to identification of a novel gene, tmem114, expressed in the lens and disrupted by the translocation. Human mutation. 28(10):968-977.

Jamieson RV, Perveen R, Kerr B, Carette M, Yardley J, Heon E, Wirth MG, van Heyningen V, Donnai D, Munier F et al. 2002. Domain disruption and mutation of the bzip transcription factor, maf, associated with cataract, ocular anterior segment dysgenesis and coloboma. Human molecular genetics. 11(1):33-42.

Javadiyan S, Craig JE, Souzeau E, Sharma S, Lower KM, Mackey DA, Staffieri SE, Elder JE, Taranath D, Straga T et al. 2017. High-throughput genetic screening of 51 pediatric cataract genes identifies causative mutations in inherited pediatric cataract in south eastern australia. G3 (Bethesda, Md). 7(10):3257-3268.

Kannabiran C, Rogan PK, Olmos L, Basti S, Rao GN, Kaiser-Kupfer M, Hejtmancik JF. 1998. Autosomal dominant zonular cataract with sutural opacities is associated with a splice mutation in the betaa3/a1-crystallin gene. Molecular vision. 4:21.

Khan AO, Aldahmesh MA, Alkuraya FS. 2015. Phenotypes of recessive pediatric cataract in a cohort of children with identified homozygous gene mutations (an american ophthalmological society thesis). Transactions of the American Ophthalmological Society. 113:T7.

Kloeckener-Gruissem B, Vandekerckhove K, Nürnberg G, Neidhardt J, Zeitz C, Nürnberg P, Schipper I, Berger W. 2008. Mutation of solute carrier slc16a12 associates with a syndrome combining juvenile cataract with microcornea and renal glucosuria. American journal of human genetics. 82(3):772-779.

Supplementary data

Knopfel EB, Vilches C, Camargo SMR, Errasti-Murugarren E, Staubli A, Mayayo C, Munier FL, Miroshnikova N, Poncet N, Junza A et al. 2019. Dysfunctional lat2 amino acid transporter is associated with cataract in mouse and humans. Frontiers in physiology. 10:688.

Lachke SA, Alkuraya FS, Kneeland SC, Ohn T, Aboukhalil A, Howell GR, Saadi I, Cavallesco R, Yue Y, Tsai AC et al. 2011. Mutations in the rna granule component tdrd7 cause cataract and glaucoma. Science (New York, NY). 331(6024):1571-1576.

Li D, Wang S, Ye H, Tang Y, Qiu X, Fan Q, Rong X, Liu X, Chen Y, Yang J et al. 2016. Distribution of gene mutations in sporadic congenital cataract in a han chinese population. Molecular vision. 22:589-598.

Li J, Leng Y, Han S, Yan L, Lu C, Luo Y, Zhang X, Cao L. 2018. Clinical and genetic characteristics of chinese patients with familial or sporadic pediatric cataract. Orphanet journal of rare diseases. 13(1):94.

Litt M, Carrero-Valenzuela R, LaMorticella DM, Schultz DW, Mitchell TN, Kramer P, Maumenee IH. 1997. Autosomal dominant cerulean cataract is associated with a chain termination mutation in the human beta-crystallin gene crybb2. Human molecular genetics. 6(5):665-668.

Litt M, Kramer P, LaMorticella DM, Murphey W, Lovrien EW, Weleber RG. 1998. Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene cryaa. Human molecular genetics. 7(3):471-474.

Mackay D, Ionides A, Kibar Z, Rouleau G, Berry V, Moore A, Shiels A, Bhattacharya S. 1999. Connexin46 mutations in autosomal dominant congenital cataract. American journal of human genetics. 64(5):1357-1364.

Muller M, Bhattacharya SS, Moore T, Prescott Q, Wedig T, Herrmann H, Magin TM. 2009. Dominant cataract formation in association with a vimentin assembly disrupting mutation. Human molecular genetics. 18(6):1052-1057.

Percin EF, Ploder LA, Jessica JY, Arici K, Horsford DJ, Rutherford A, Bapat B, Cox DW, Duncan AM, Kalnins VI. 2000. Human microphthalmia associated with mutations in the retinal homeobox gene chx10. Nature genetics. 25(4):397-401.

Pras E, Levy-Nissenbaum E Fau - Bakhan T, Bakhan T Fau - Lahat H, Lahat H Fau - Assia E, Assia E Fau - Geffen-Carmi N, Geffen-Carmi N Fau - Frydman M,

Supplementary data

Frydman M Fau - Goldman B, Goldman B Fau - Pras E, Pras E. 2002. A missense mutation in the lim2 gene is associated with autosomal recessive presenile cataract in an inbred iraqi jewish family. (0002-9297 (Print)).

Ramachandran RD, Perumalsamy V, Hejtmancik JF. 2007. Autosomal recessive juvenile onset cataract associated with mutation in bfsp1. Human genetics. 121(3-4):475-482.

Reis LM, Tyler RC, Muheisen S, Raggio V, Salviati L, Han DP, Costakos D, Yonath H, Hall S, Power P et al. 2013. Whole exome sequencing in dominant cataract identifies a new causative factor, cryba2, and a variety of novel alleles in known genes. Human genetics. 132(7):761-770.

Riazuddin SA, Yasmeen A, Yao W, Sergeev YV, Zhang Q, Zulfiqar F, Riaz A, Riazuddin S, Hejtmancik JF. 2005. Mutations in betab3-crystallin associated with autosomal recessive cataract in two pakistani families. Investigative ophthalmology & visual science. 46(6):2100-2106.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E. 2015. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology. Genetics in medicine. 17(5):405-423.

Semina EV, Brownell I, Mintz-Hittner HA, Murray JC, Jamrich M. 2001. Mutations in the human forkhead transcription factor foxe3 associated with anterior segment ocular dysgenesis and cataracts. Human molecular genetics. 10(3):231-236.

Semina EV, Ferrell RE, Mintz-Hittner HA, Bitoun P, Alward WL, Reiter RS, Funkhauser C, Daack-Hirsch S, Murray JC. 1998. A novel homeobox gene pitx3 is mutated in families with autosomal-dominant cataracts and asmd. Nature genetics. 19(2):167-170.

Shiels A, Bennett TM, Hejtmancik JF. 2010. Cat-map: Putting cataract on the map. Molecular vision. 16:2007-2015.

Shiels A, Bennett TM, Knopf HL, Maraini G, Li A, Jiao X, Hejtmancik JF. 2008. The epha2 gene is associated with cataracts linked to chromosome 1p. Molecular vision. 14:2042-2055.

Shiels A, Bennett TM, Knopf HL, Yamada K, Yoshiura K, Niikawa N, Shim S, Hanson PI. 2007. Chmp4b, a novel gene for autosomal dominant cataracts

linked to chromosome 20q. American journal of human genetics. 81(3):596-606.

Shiels A, Mackay D, Ionides A, Berry V, Moore A, Bhattacharya S. 1998. A missense mutation in the human connexin50 gene (gja8) underlies autosomal dominant "zonular pulverulent" cataract, on chromosome 1q. American journal of human genetics. 62(3):526-532.

Stambolian D, Ai Y, Sidjanin D, Nesburn K, Sathe G, Rosenberg M, Bergsma DJ. 1995. Cloning of the galactokinase cdna and identification of mutations in two families with cataracts. Nature genetics. 10(3):307-312.

Stephan DA, Gillanders E, Vanderveen D, Freas-Lutz D, Wistow G, Baxevanis AD, Robbins CM, VanAuken A, Quesenberry MI, Bailey-Wilson J et al. 1999. Progressive juvenile-onset punctate cataracts caused by mutation of the gammad-crystallin gene. Proceedings of the National Academy of Sciences of the United States of America. 96(3):1008-1012.

Sun H, Ma Z, Li Y, Liu B, Li Z, Ding X, Gao Y, Ma W, Tang X, Li X et al. 2005. Gamma-s crystallin gene (crygs) mutation causes dominant progressive cortical cataract in humans. Journal of medical genetics. 42(9):706-710.

Sun M, Chen C, Hou S, Li X, Wang H, Zhou J, Chen X, Liu P, Kijlstra A, Lin S et al. 2019. A novel mutation of pank4 causes autosomal dominant congenital posterior cataract. Human mutation. 40(4):380-391.

Taylor RL, Handley MT, Waller S, Campbell C, Urquhart J, Meynert AM, Ellingford JM, Donnelly D, Wilcox G, Lloyd IC et al. 2017. Novel pex11b mutations extend the peroxisome biogenesis disorder 14b phenotypic spectrum and underscore congenital cataract as an early feature. Investigative ophthalmology & visual science. 58(1):594-603.

Tzifi F, Pons R, Athanassaki C, Poulou M, Kanavakis E. 2011. Congenital cataracts, facial dysmorphism, and neuropathy syndrome. Pediatric neurology. 45(3):206-208.

Willoughby CE, Shafiq A, Ferrini W, Chan LL, Billingsley G, Priston M, Mok C, Chandna A, Kaye S, Heon E. 2005. Crybb1 mutation associated with congenital cataract and microcornea. Molecular vision. 11:587-593.

Xia XY, Li N, Cao X, Wu QY, Li TF, Zhang C, Li WW, Cui YX, Li XJ, Xue CY. 2014. A novel col4a1 gene mutation results in autosomal dominant non-syndromic congenital cataract in a chinese family. BMC medical genetics. 15:97.

Supplementary data

Yamakawa N, Oe K, Yukawa N, Murakami K, Nakashima R, Imura Y, Yoshifuji H, Ohmura K, Miura Y, Tomosugi N et al. 2016. A novel phenotype of a hereditary hemochromatosis type 4 with ferroportin-1 mutation, presenting with juvenile cataracts. Internal medicine (Tokyo, Japan). 55(18):2697-2701.

Yu LC, Twu YC, Chang CY, Lin M. 2001. Molecular basis of the adult i phenotype and the gene responsible for the expression of the human blood group i antigen. Blood. 98(13):3840-3845.

Yuan L, Yi J, Lin Q, Xu H, Deng X, Xiong W, Xiao J, Jiang C, Yuan X, Chen Y et al. 2016. Identification of a prx variant in a chinese family with congenital cataract by exome sequencing. QJM : monthly journal of the Association of Physicians. 109(11):731-735.

Zhao L, Chen XJ, Zhu J, Xi YB, Yang X, Hu LD, Ouyang H, Patel SH, Jin X, Lin D et al. 2015. Lanosterol reverses protein aggregation in cataracts. Nature. 523(7562):607-611.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research. 31(13):3406-3415.