HHS Public Access

Author manuscript

Hum Genet. Author manuscript; available in PMC 2018 February 01.

Published in final edited form as:

Hum Genet. 2017 February ; 136(2): 205–225. doi:10.1007/s00439-016-1747-6.

Novel phenotypes and loci identified through clinical genomics approaches to pediatric cataract

Nisha Patel[1], **Deepti Anand**[2], **Dorota Monies**[1,3], **Sateesh Maddirevula**[1], **Arif O. Khan**[1,4], **Talal Algoufi**[5], **Mohammed Alowain**[6], **Eissa Faqeih**[7], **Muneera Alshammari**[8], **Ahmed Qudair**[6], **Hadeel Alsharif**[1], **Fatimah Aljubran**[1], **Hessa S. Alsaif**[1], **Niema Ibrahim**[1], **Firdous M. Abdulwahab**[1], **Mais Hashem**[1], **Haifa Alsedairy**[1], **Mohammed A. Aldahmesh**[1], **Salil A. Lachke**[2,9], and **Fowzan S. Alkuraya**[1,10]

[1]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

[2]Department of Biological Sciences, University of Delaware, Newark, DE 19716, USA

[3]Saudi Human Genome Program, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia

[4]Eye Institute, Cleveland Clinic Abu Dhabi, Abu Dhabi, United Arab Emirates

[5]Department of Pediatrics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia

[6]Department of Medical Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia

[7]Department of Pediatrics, King Fahad Medical City, Riyadh, Saudi Arabia

[8]Department of Pediatrics, King Khalid University Hospital and College of Medicine, King Saud University, Riyadh, Saudi Arabia

[9]Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE 19716, USA

[10]Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

Abstract

Pediatric cataract is highly heterogeneous clinically and etiologically. While mostly isolated, cataract can be part of many multisystem disorders, further complicating the diagnostic process. In this study, we applied genomic tools in the form of a multi-gene panel as well as whole-exome sequencing on unselected cohort of pediatric cataract (166 patients from 74 families). Mutations in previously reported cataract genes were identified in 58% for a total of 43 mutations, including 15 that are novel. *GEMIN4* was independently mutated in families with a syndrome of cataract,

Corresponding author: Fowzan S. Alkuraya, falkuraya@kfshrc.edu.sa.

Compliance with ethical standards

Conflict of interest. On behalf of all authors, the corresponding author states that there is no conflict of interest.

global developmental delay with or without renal involvement. We also highlight a recognizable syndrome that resembles galactosemia (a fulminant infantile liver disease with cataract) caused by biallelic mutations in *CYP51A1*. A founder mutation in *RIC1 (KIAA1432)* was identified in patients with cataract, brain atrophy, microcephaly with or without cleft lip and palate. For nonsyndromic pediatric cataract, we map a novel locus in a multiplex consanguineous family on 4p15.32 where exome sequencing revealed a homozygous truncating mutation in *TAPT1*. We report two further candidates that are biallelically inactivated each in a single cataract family: *TAF1A* (cataract with global developmental delay) and WDR87 (non-syndromic cataract). In addition to positional mapping data, we use *iSyTE* developmental lens expression and gene-network analysis to corroborate the proposed link between the novel candidate genes and cataract. Our study expands the phenotypic, allelic and locus heterogeneity of pediatric cataract. The high diagnostic yield of clinical genomics supports the adoption of this approach in this patient group.

Introduction

Pediatric cataract is estimated to have a prevalence of 3–6 per 10,000 (Rahi and Dezateaux 2001; Foster et al. 1997; Stayte et al. 1993). Clinically, it is highly variable in its age of onset, severity and distribution (unilateral vs. bilateral and syndromic vs. isolated). Delayed intervention for this treatable disease can result in permanent blindness due to amblyopia. Indeed, many children in low-income countries are blind because of untreated cataract (Medsinge and Nischal 2015). The morbidity of pediatric cataract is also significant in higher income countries despite better access to surgical treatment, mostly driven by cases of delayed diagnosis (Zhang et al. 2012).

The etiology of pediatric cataract is heterogeneous but genetic factors account for 8–29% of cases (Shiels and Hejtmancik 2007, 2013; Hejtmancik 2008). All modes of inheritance have been reported, with autosomal dominant inheritance considered the most common form worldwide and autosomal recessive inheritance more common in the Middle East (Khan 2012, 2013; Khan et al. 2015). The online tool Cat-Map currently lists more than 38 genes that are mutated in isolated (non-syndromic) cataract (Shiels et al. 2010). Genes encoding the crystalline family of proteins account for a substantial proportion of mutation-positive pediatric cataract cases. Genes encoding transcription factors that control early lenticular development such as *EYA1* and *PITX3* are also an important source of cataract linked mutations. Interestingly, some genes are known to cause autosomal dominant as well as recessive forms of pediatric cataract depending on the nature of the mutation, e.g., *BFSP2*, *TDRD7* and *CRYAB* (Aldahmesh et al. 2011; Safieh et al. 2009; Lachke et al. 2011). Similarly, genes known to be mutated in syndromic forms of cataract have also been reported to cause apparently isolated cataract, e.g., *AGK* (Aldahmesh et al. 2012).

Identification of causal mutations in pediatric cataract can greatly improve our understanding of the mechanisms that control normal lenticular development. Practical benefits of mutation identification include improved diagnostic accuracy, refined recurrence risk estimates as well as the possibility of prevention. Unfortunately, the remarkable clinical and genetic heterogeneity described above makes it challenging to provide molecular diagnosis for pediatric cataract patients. Fortunately, the advent of genomics tools enables

the interrogation of a large number of genes simultaneously. The potential of this approach to improve the diagnostic yield in pediatric cataract has already been demonstrated in a number of studies (Gillespie et al. 2014, 2016; Ma et al. 2016; Musleh et al. 2016). The unbiased nature of this approach has unraveled the full phenotypic potential of known cataract genes and enabled the establishment of novel syndromic and isolated cataract genes (Aldahmesh et al. 2012). In this study, we show the power of implementing genomics tools in the diagnostic workup of pediatric cataract patients. In addition to broadening the allelic spectrum of known cataract genes, we describe novel candidate genes. Further, we use eye gene expression databases such as *iSyTE* (*i*ntegrated *Sy*stems *T*ool for *E*ye gene discovery) (Lachke et al. 2012) along with gene expression analysis in key mouse mutants that exhibit lens defects to indicate the potential regulatory pathways in which these newly identified cataract genes may function in the lens.

Materials and methods

Human subjects

All cataract patients seen in a pediatric ophthalmology clinic run by one of the authors (AOK) were eligible, regardless of family history. We have also enrolled a family referred from pediatric gastroenterology with unexplained lethal form of infantile liver disease and cataract. Informed consent was obtained from parents, and venous blood was collected from index and available family members as per an IRB-approved protocol (KFSHRC RAC# 2070023).

Multi–gene panel sequencing

A panel of 322 genes known to be mutated in various genetic eye conditions, including those involving cataract was designed as described before (Group SM 2015). All index cases were initially run on this panel as a first-tier test. Details of the bioinformatics analysis are published elsewhere (Group SM 2015). Variants were called according to the ACMG guidelines on variant interpretation.

Exome sequencing

All cases in which the multi-gene panel failed to identify a likely causal mutation were exome sequenced as described before (Group SM 2015). The surviving variants were analyzed based on zygosity (depending on family pedigree), predicted pathogenicity based on SIFT, Polyphen and combined annotation-dependent depletion (CADD) scores (for missense variants), prioritizing truncating variants, location within the autozygome (for AR cases) and frequency below 0.01 within in-house (2200 exomes), and ExAC databases. All variants reported here have been confirmed by Sanger sequencing and segregation analysis was completed in all available family members.

Positional mapping

Positional mapping was carried out using autozygome analysis as described before (Alkuraya 2010, 2012). Briefly, the Axiom SNP Chip platform was used for genome-wide genotyping followed by mapping regions of homozygosity (ROH) that are >2 Mb as

surrogates of autozygosity. Where applicable, exome variants were filtered by the coordinates of the candidate autozygome as described before (Alkuraya 2013, 2016).

Mouse lens expression analysis by *iSyTE* tool

To gain insights into the significance of each of the cataract-linked candidate genes in this study (*TAPT1*, *RIC1*, *CYP51A1*, *GEMIN4*, *TAF1A* and *WDR87*) we applied our published approach of using lens expression analysis (Lachke et al. 2012; Anand and Lachke 2016). Mouse orthologs of these genes were investigated for their expression and enrichment in mouse lens expression microarrays datasets using *iSyTE* database (Lachke et al. 2012) and publicly available mouse lens microarray data. Expression intensities scores were computed at different stages of lens development stages, namely, E10.5, E16.5, P0, P28 and P56. In addition, lens-enrichment was estimated based on whole embryonic body (WB)-based in silico subtraction approach. The "R" statistical environment (http://www.rproject.org) was used to import raw microarray files, which were pre-processed and background corrected using Affy package available at Bioconductor (http://www.bioconductor.org) (Gautier et al. 2004). Detailed analysis of microarrays is described elsewhere (Anand et al. 2015). Using RNA-seq data from mouse stage P0 (SRP040480) isolated lens epithelium (P0_epi) and fiber cells (P0_FC) (Hoang et al. 2014), expression values in counts per million (CPM) were obtained and plotted to test differential expression of candidate genes in these cell types.

Gene expression analysis in targeted gene knockout mouse mutant lens datasets

The expression of candidate genes (*Tapt1*, *Ric1*, *Cyp51a1*, *Gemin4*, *Taf1a* and *Wdr87*) was investigated in various targeted gene knockout mouse mutants that exhibit lens defects. Mouse lens tissue gene expression microarray datasets from mutant animals for *Pax6* conditional lens knockout (cKO) at E9.5 (GSE49227) and E10.5 (GSE49216); *Brg1* (dominant negative dnBrg1 mutant) at E15.5 (GSE22322), *Notch2* conditional lens knockout mutant at E19.5 (GSE31643), *E2f1:E2f2:E2f3* (triple null conditional lens knockout mutant) at P0 (GSE16533), *Hsf4* null at P0 (GSE22362), *Sparc* null at P28 (isolated lens epithelium) (GSE13402), *Tdrd7* null at P30 (GSE25776), *Klf4* null at P56 (GSE47694), and *Mafg−/−:Mafk+/−* compound mutants at P60 (GSE65500) were analyzed for differential expression of candidate genes. Further, transgenic mice over-expressing *Foxe3* in fiber compartment at P2 (GSE9711) were also analyzed. Mutant lens tissues that exhibited significant differential expression of candidate genes (*Tapt1*, *Ric1*, *Cyp51a1*, *Gemin4*, and *Taf1a*) at p value ≤ 0.05 were plotted.

Candidate gene–network analysis using protein–protein interaction (PPI) data and *iSyTE*

To derive molecular insights for the identified candidate genes (*Tapt1*, *Ric1*, *Cyp51a1*, and *Gemin4*), we used an inhouse Python script to fetch out statistically significant PPI with proteins that function in the lens as well as potential new lens-expressed candidates from the String database (http://string-db.org). The obtained interactions were then subjected to lens expression and enrichment analysis at E10.5 lens dataset in *iSyTE* dataset described above, and visualized using Cytoscape.

Results

Clinical phenotypes

A total of 166 cataract patients comprising 74 families were enrolled in this study. The demographics of the study cohort are detailed in Table 1. A positive family history was observed in 67%, and non-syndromic cataract was the most common presentation (72%). Both known and apparently novel forms of syndromic cataract were encountered (Table 1). A few syndromic forms of cataract are worth highlighting. The first is related to what we initially reported in 2015 in several families who all shared the same founder mutation in *GEMIN4*.

All patients shared global developmental delay and infantile cataract with or without renal involvement. Patient 14DG2265 provided independent confirmation of this association where his novel *GEMIN4* mutation (NM_015721.2:c.314C>T;p. (Pro105Leu)) was associated with an identical phenotype (Table 1, Table S1). The mutation segregated within the family, and both parents are carriers, is absent in our database and predicted to be pathogenic by Polyphen, SIFT and CADD. Another recognizable syndrome was observed in 16DG0226 who was found at 1 week of age to have cholestatic jaundice and cataract, and was referred to our center for further evaluation. His physical examination showed growth parameters on the 5th percentile, icterus and bilateral cataract. His laboratory investigations revealed elevated liver enzymes (ALT 143, AST 518, alkaline phosphatase 729, GGT 167), AFP (>50,000), and ferritin (7994). Urine was negative for succinylacetone and reducing substances, and blood had normal isoelectric focusing of transferring. A liver biopsy revealed cholestasis with diffuse giant cell transformation and pseudorosettes. Parents are consanguineous and there is history of one sister who died at age of 2 months with liver failure.

There was also positive family history on the paternal side of neonatal deaths in twins due to progressive cholestatic jaundice (see pedigree in Figure S1, Table S1). By combining the index and his affected cousin, we were able to map this phenotype to a locus on Chr7: 80,350,364-105,103,372 where exome sequencing revealed a mutation in *CYP51A1* (NM_000786.3:c.695T>C;p.(Leu232Pro)). Finally, in two families with a syndromic form of cataract consisting of global developmental delay, microcephaly, brain atrophy with or without cleft lip and palate we were able to identify a candidate locus on Chr9:5629029-5778014 where exome sequencing revealed a shared founder mutation in *RIC1* (NM_020829.3:c.3794G>C;p.(Arg1265Pro). Surprisingly, we also observed isolated cataract without aniridia in a patient with a novel de novo dominant *PAX6* mutation (10DG1895).

Expanding the allelic spectrum of pediatric cataract

The multi-gene panel and exome sequencing identified a likely causal mutation in 58% of our cohort (not including candidate genes). The most commonly mutated group of genes was the crystalline genes, and one founder mutation in *CRYBB1* was identified in 11 families (Table 1; Fig. 1). Table 1 lists all the mutations identified in known cataract genes, including 15 that are novel (20%). Because the design of the multi-gene panel was in August

2013, cataract genes published after that date were not included but mutations therein were identified by exome sequencing, which we performed on all cases with a negative panel result. Of particular interest is *LONP1*, which we found to be mutated in five families, thus representing the second most commonly mutated gene in our cohort after the crystalline genes. Furthermore, we note that not all *LONP1*–related cataract cases were syndromic, which suggests that *LONP1* is yet another example of genes that can be mutated in both syndromic and non-syndromic forms of cataract.

Expanding the genetic heterogeneity of pediatric cataract

In addition to revealing mutations in known cataract genes that postdate the design of the multi-gene panel, exome sequencing of negative panel cases revealed, as expected, mutations in candidate genes. Specifically, we confirmed *GEMIN4* as a disease gene for the syndrome of cataract and global developmental delay (Alazami et al. 2015). The same founder mutation in *GEMIN4* was identified in 10DG0703 who was previously reported to have a missense variant in *MFSD6L*, thus disproving the link proposed between cataract and *MFSD6L*, at least in that patient (Aldahmesh et al. 2012). Similarly, we have previously published CYP51A1 as a novel candidate gene for nonsyndromic cataract based on a family (10DG1249) with a pseudodominant inheritance of a novel missense variant in this gene (Khan et al. 2015; Aldahmesh et al. 2012). Subsequently, another group reported a mutation in this gene in a patient with cataract and liver disease (Gillespie et al. 2014). Thus, our finding of an independent mutation in 16DG0226 (Figure S1) confirm *CYP51A1* as a disease gene for the syndrome of cataract and cholestatic liver disease, although it can also be mutated in patients with isolated cataract.

In family 12DG2657, we were also able to map isolated cataract phenotype to a single locus (Chr4:13944470-16401420), in which exome sequencing revealed a splicing variant in the novel candidate TAPT1 (NM_153365.2:c.846 + 2insT). RTPCR confirmed the partially truncating nature of this variant (NM_153365.2:r.712_846del), (Figure S2). Furthermore we were able to identify the same mutation (NM_020829.3:c.3794G>C;p.(Arg1265Pro)) in the novel candidate *RIC1* in two apparently unrelated patients (07DG-0035/10DG1320 and 15DG2427) who nonetheless shared one autozygous interval thus confirming the founder nature of this mutation (Figure S3).

In addition to the above genes whose candidacy is supported by independent mutations (*GEMIN4* and *CYP51A1*) or linkage analysis (*TAPT1* and *RIC1*), exome sequencing also revealed homozygous truncating variants in two genes not previously linked to cataract. In patient 11DG2176, who presented with global developmental delay, unexplained hepatomegaly and cataract, we identified a homozygous frameshift deletion in *TAF1A* (NM_001201536.1:c.40_41del;p.(Asp14*)) (Figure S4). Patient 12DG2386, on the other hand, and his sibling presented with isolated congenital cataract and both were found to have a homozygous nonsense mutation in *WDR87* (NM_031951.4:c.856G>T;p.(Glu286*)) (Figure S5).

The candidate cataract genes are expressed in lens development

We next sought to investigate the relevance of the newly identified cataract-linked genes to lens biology. We first analyzed mouse lens microarray datasets at embryonic, early postnatal and late postnatal stages to examine the expression of *Tapt1*, *Ric1*, *Cyp51a1*, *Gemin4*, *Taf1a* and *Wdr87* during lens development. *Tapt1* is expressed in lens tissue at E10.5, E16.5, P0, and P56, and exhibits a trend toward high expression with developmental progression. Further, its expression was found to be high in P28 isolated lens epithelium as well (Fig. 2a). *Ric1* was expressed in the lens at all stages examined, albeit at low comparable levels, except in isolated lens epithelial cells where it exhibited higher expression (Fig. 2a). *Cyp51a1* was highly expressed in lens tissue at E10.5, E16.5, P0 and P56, and while it was also expressed in isolated lens epithelium, its levels are low in these cells compared to the whole lens tissue (Fig. 2a). *Gemin4* exhibited an expression trend that is high in early lens development at E10.5 and became progressively low in subsequent stages (Fig. 2a). Lens microarray indicates that *Taf1a* is expressed in various stages of mouse lens development (Fig. 2a). Finally, *WDR87* mouse ortholog, *4932431P20Rik*, is also expressed in the lens albeit at lower levels (Fig. 2a).

Next, we investigated if these candidate genes exhibit enriched expression in the lens as described (Lachke et al. 2012; Anand et al. 2015). *Tapt1* is significantly enriched in the lens from embryonic stage E16.5 through P56, with highest lens-enrichment in the P28 lens epithelium (Fig. 2b). Similarly, *Taf1a* exhibits enriched expression in several stages of lens development, namely, at E10.5, P0, P56 as well as in P28 lens epithelium (Fig. 2b). *Ric1* is enriched only in the P28 lens epithelium, while *Cyp51a1* and *Gemin4* exhibit lens-enrichment in embryonic stages E16.5 and E10.5, respectively (Fig. 2b).

We also examined RNA-seq data from newborn mouse lens epithelium and fiber cells to investigate if these genes are expressed within specific lenticular cell types. We find that while *Tapt1* is significantly expressed in both cell types in the lens, its expression in the epithelium is significantly higher compared to that in fiber cells (Fig. 2c). In contrast, *Cyp51a1* and *Ric1* that are also expressed in both lens cell types, exhibit significantly high expression in fiber cells compared to epithelial cells (Fig. 2c). *Gemin4* expression in both lens cell types was low in newborn mouse lenses (Fig. 2c), in agreement with the trend of low expression with lens development progression from embryonic to postnatal stages as observed in the microarray analysis. While *Taf1a* expression was found to be higher in fiber cells compared to epithelial cells, *4932431P20Rik* (*WDR87*) expression in both cell types was found to be low and not significantly different (Fig. 2c).

The candidate cataract genes are mis–expressed in key gene knockout mice with lens defects

We next sought to investigate whether *Tapt1*, *Ric1*, *Cyp51a1*, *Gemin4*, *Taf1a* and *4932431P20Rik* (*Wdr87*) were affected in different gene knockout mouse mutants that exhibited defects in lens development. *Tapt1* was significantly down-regulated in Pax6 conditional lens knockout (*Pax6* cKO) mouse lenses at E9.5, *Notch2* conditional lens knockout (*Notch2* cKO) mouse lenses at E19.5 and *E2f1−/−:E2f2−/−:E2f3−/−* triple conditional lens knockout (*E2f1/2/3* cKO) mouse lenses at P0 (Fig. 3). *Ric1* was down-

regulated in *Pax6* cKO lenses at E9.5 and E10.5, and up-regulated in *Sparc* null lens epithelium (Fig. 3). *Cyp51a1* exhibited mis-regulation in both directions in different gene knockout mouse lenses. In *Brg1* mutant (dnBrg1 mutant) lenses at E15.5 and *E2f1/2/3* cKO mouse lenses at P0, *Cyp51a1* exhibited significantly reduced expression (Fig. 3). Further, in transgenic mice that overexpress the lens epithelial transcription factor *Foxe3* in lens fiber cells, *Cyp51a1* expression was significantly reduced as well (Fig. 3). However, in stage P30 *Tdrd7* null mouse lenses, P56 *Klf4* conditional lens knockout (*Klf4* cKO) mouse lenses, as well as P60 *Mafg−/−:Mafk+/−* compound mouse mutant lenses *Cyp51a1* was significantly up-regulated (Fig. 3). *Gemin4* was found to be significantly reduced in Cpb:p300 conditional lens knockout (*Cpb:p300* cKO) mice at E10.5 (Fig. 3). Taf1a was found to be down-regulated in E15.5 *dnBrg1* and P0 *Hsf4* null mouse mutants, both of which exhibit lens defects (Fig. 3). Finally, *4932431P20Rik* (*WDR87*) was not identified to be mis-regulated in any of the mouse mutant datasets tested.

The candidates interact with proteins with known lens function or expression

To investigate if the new cataract associated candidates may potentially interact with proteins that are known to function in the lens or exhibit lens expression, we performed an integrated analysis with publically available protein-protein interaction (PPI) data and overlay of *iSyTE* lens gene expression data. Further, we investigated these networks for functional gene-ontology (GO) categories. Together, these analyses led to insights into their established connectivity with other candidates that are involved in lens defects or which may be expressed in the lens. This approach led to the outlining of 22 direct interacting partners of the nonsyndromic cataract candidate TAPT1 (Figure S6A, B). Further, from a total of 39 direct protein-protein level connections of the non-syndromic cataract candidate RIC1, 32 candidates were expressed in the lens, of which 14 were lens-enriched including GJA1, which had been shown to interact with RIC1 and mutations of which cause microphthalmia and cataract (Akiyama et al. 2005; Paznekas et al. 2003) (Figure S6C, D). This approach also revealed that the syndromic cataract candidate GEMIN4 is connected to 53 partners, of which 50 candidates exhibit lens expression and 37 exhibit lens-enrichment (Figure S7A, B). Similarly, CYP51A1, which is known to be involved in the synthesis of cholesterol, steroids and other lipids, is connected to 51 direct interactors, of which 35 candidates exhibit lens expression and 22 exhibit lens-enrichment (Figure S7C, D). As expected, GO analysis of the CYP51A1-PPI network reveals an enrichment for sterol biosynthetic process (GO:0016126) categories that includes 15 protein-protein interaction candidates, namely, TM7SF2, MVD, HMGCR, HSD3B7, HMGCS1, LSS, FDFT1, DHCR7, HSD17B7, NSDHL, DHCR24, FDPS, SIGMAR1, SQLE, MVK, that are expressed in the lens. Earlier studies on sterol profiling of the affected individuals with cataract and other eye disorders with causal mutation identified in CYP51A1, CYP27A1, SC5D, DHCR7 genes clearly suggest their role in sterol biosynthetic process/pathways (Gillespie et al. 2016). Further, CYP51A1 is directly connected to ALDH1A1 and MAFG, both of which are linked to cataracts (Agrawal et al. 2015; Lassen et al. 2007).

Discussion

Molecular characterization of pediatric cataract has many practical applications. It provides accurate diagnosis, ends an otherwise expensive and protracted diagnostic odyssey and empowers families to make informed reproductive choices. Molecular diagnosis also has the potential to alter patient management. One good example is patient 13DG2254 whose cataract was found to be caused by a novel *GALT* mutation prompting urgent referral to the metabolic specialist for close dietary management of galactosemia. The marked clinical and genetic heterogeneity of pediatric cataract often complicates clinically-guided molecular testing, although this is quickly changing with the advent of clinical genomics. In this study, we show that a genomics approach can provide a likely molecular diagnosis in the majority of pediatric cataract patients. Our data also show that the genetic heterogeneity of pediatric cataract has not yet been fully captured, and we add to this genetic heterogeneity four loci defined by mutations in *GEMIN4, TAPT1, RIC1* and *CYP51A*, as well as biallelic loss of function mutations in *TAF1A* and *WDR87*.

GEMIN4 is an intron-less gene that encodes Gem (nuclear organelle)-associated protein 4, a ubiquitously expressed component of the Gemin protein complex that also includes SMN1 and the core components Gemin proteins 2, 3, 5, 6, 7 and 8 as well as Unrip (Charroux et al. 2000; Lorson et al. 2008). The complex is known to associate with the spliceosomal complex U snRNP (Fischer et al. 1997). The exact biological role of the complex is unknown so it is unclear how deficiency of *GEMIN4* can lead to the syndrome of global developmental delay and congenital cataract, and whether or not this mediated through perturbation of the complex. However, our finding of two independent homozygous mutations in *GEMIN*4 in patients with a similar phenotype strongly implicates *GEMIN4* in the etiology of this syndrome.

CYP51A1 encodes lanosterol 14α-demethylase, an enzyme that catalyzes a late step in cholesterol synthesis (Acimovic and Rozman 2013). Complete deficiency of the murine ortholog is embryonic lethal, which may explain why all the mutations, with the exception of one heterozygous stop-gain, observed thus far in this gene are all missense, rather than truncating (Keber et al. 2011). The hepatocyte-specific *Cyp51* partial KO mice display poor weight gain, increased liver/body size ratio as well as severe liver inflammation and fibrosis, findings reminiscent of the phenotype we observe in patient 16DG0226, as well as the family reported by Gillespie et al. (2014, 2016) (Lorbek et al. 2015). Further, other genes such as *CYP27A1*, *SC5D*, *DHCR7*, which encode enzymes involved in cholesterol and sterol biosynthesis are also linked with syndromic cataract. Thus, accumulation of precursor metabolites such as lanosterol in the lens and liver may be causative of tissuespecific defects observed in the patient in the present study. The link between *TAPT1* and cataract was unexpected.

TAPT1 encodes transmembrane anterior posterior transformation 1 protein that was found by Symoens et al. to be mutated in two families with osteogenesis imperfect alike skeletal dysplasia (Symoens et al. 2015). On the other hand, the family we describe in which cataract maps to a single locus in which a homozygous splicing *TAPT1* mutation was identified did not have any evidence of skeletal involvement. It is possible that the apparent discrepancy in

phenotype represents a genuine example of allelism especially since both our mutation and that identified by Symoen cause in-frame truncations mediated by entire exon skipping (exon 6 in this report and 10 in Symoen's).

Future cataract patients with different mutations in *TAPT1* will help clarify the true phenotypic spectrum. Similar to *TAPT1*, we have identified *RIC1* as a novel cataract candidate based on strong positional mapping data that point to a single locus. Significantly, the connection in the PPI based network between *RIC1* and the cataract-linked gene *GJA1* was due to an established direct interaction between these proteins as shown by a previous study (Akiyama et al. 2005). Further, that study also showed that knockdown of *RIC1* resulted in defective localization of GJA1 to gap junctions, affecting gap junction conductivity, which may offer a potential explanation for the cataract associated with *RIC1* mutations.

We have previously shown that the mutation spectrum of genetically heterogeneous diseases is dominated by autosomal recessive mutations in our highly inbred populations (Patel et al. 2015; Anazi et al. 2016; Alazami et al. 2016). We show in this study that cataract displays a similar trend with recessive mutations accounting for 87% of all identified mutations. Interestingly, we show that some cataract genes that had only been reported to cause the disease in a dominant fashion, can also cause autosomal recessive cataract, e.g., *EPHA2* in patient 10DG0428. These examples are very helpful in shedding light on the molecular pathogenesis of these genes since they can challenge the notion of haploinsufficiency of dominant mutations when carriers of loss of function recessive mutations (parents) appear normal.

It has been shown that enriched expression in developing lens tissue can be used as a criterion to evaluate potential function in lens development (Lachke et al. 2011, 2012a, b; Anand and Lachke 2016; Anand et al. 2015; Agrawal et al. 2015; Kasaikina et al. 2011; Wolf et al. 2013; Manthey et al. 2014; Dash et al. 2015; Audette et al. 2016). Consistent with those data, we find that all six candidates are significantly expressed in mouse lens development, and five exhibit lens-enrichment. We also examined microarray data from several targeted gene mouse mutants lens/presumptive lens tissue for their expression of these candidate genes, and performed PPI analysis. Several interesting observations emerged from these analyses. For example, the downregulation of *Tapt1* and *Ric1* in *Pax6* cKO presumptive lens ectoderm suggests that these genes are expressed early in lens development. Similarly, *Tapt1* is down-regulated in *Notch2* cKO lens indicating that *Tapt1* is under Notch signaling pathway, which is essential for proper lens development. We note that *Cyp51a1* is abnormally expressed in *Mafg−/−:Mafk+/−*, which exhibit mis-regulation of genes involved in the sterol synthesis pathway (Agrawal et al. 2015). PPI network analysis for Cyp51a1 independently shows an enrichment for sterol biosynthetic process, which is particularly significant because lanosterol synthase mutations can cause cataracts in humans and rat (PMC1350995) and the sterol pathway is important for maintenance of lens transparency by prevention of protein aggregation in the lens (Makley et al. 2015; Zhao et al. 2015). Thus, analysis of specific gene perturbation mouse mutants that exhibit lens defects demonstrated mis-regulation of these newly identified cataract genes, and PPI analysis revealed novel connections that is suggestive of function in lens development.

In conclusion, we show the value of applying research and clinical genomics in the analysis of pediatric cataract, which in turn will lead to improved diagnostic accuracy in the near future. Our study confirms the candidacy of some previously reported novel genes as well as adds a number of novel candidates whose potential role in lenticular development and cataract should be verified by future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by King Salman Center for Disability Research (FSA) and the National Eye Institute of the National Institutes of Health under Award Number R01EY021505 (SAL). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. SAL is a Pew Scholar in Biomedical Sciences.

References

Acimovic J, Rozman D. Steroidal triterpenes of cholesterol synthesis. Molecules. 2013; 18(4):4002–4017. [PubMed: 23558541]

Agrawal SA, Anand D, Siddam AD, et al. Compound mouse mutants of bZIP transcription factors Mafg and Mafk reveal a regulatory network of non-crystalline genes associated with cataract. Hum Genet. 2015; 134(7):717–735. [PubMed: 25896808]

Akiyama M, Ishida N, Ogawa T, et al. Molecular cloning and functional analysis of a novel Cx43 partner protein CIP150. Biochem Biophys Res Commun. 2005; 335(4):1264–1271. [PubMed: 16112082]

Alazami AM, Patel N, Shamseldin HE, et al. Accelerating novel candidate gene discovery in neurogenetic disorders via whole exome sequencing of prescreened multiplex consanguineous families. Cell Rep. 2015; 10(2):148–161. [PubMed: 25558065]

Alazami AM, Al-Qattan SM, Faqeih E, et al. Expanding the clinical and genetic heterogeneity of hereditary disorders of connective tissue. Hum Genet. 2016; 135(5):525–540. [PubMed: 27023906]

Aldahmesh MA, Khan AO, Mohamed J, Alkuraya FS. Novel recessive BFSP2 and PITX3 mutations: insights into mutational mechanisms from consanguineous populations. Genet Med. 2011; 13(11): 978–981. [PubMed: 21836522]

Aldahmesh MA, Khan AO, Mohamed JY, et al. Identification of a truncation mutation of acylglycerol kinase (AGK) gene in a novel autosomal recessive cataract locus. Hum Mutat. 2012a; 33(6):960–962. [PubMed: 22415731]

Aldahmesh MA, Khan AO, Mohamed JY, et al. Genomic analysis of pediatric cataract in Saudi Arabia reveals novel candidate disease genes. Genet Med. 2012b; 14(12):955–962. [PubMed: 22935719]

Alkuraya FS. Autozygome decoded. Genet Med. 2010; 12(12):765–771. [PubMed: 21189493]

Alkuraya FS. Discovery of rare homozygous mutations from studies of consanguineous pedigrees. Curr Protoc Hum Genet. 2012; 75:6.12.1–6.12.13.

Alkuraya FS. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. Hum Genet. 2013; 132(11):1197–1211. [PubMed: 23907654]

Alkuraya FS. Discovery of mutations for Mendelian disorders. Hum Genet. 2016; 135(6):615–623. [PubMed: 27068822]

Anand D, Lachke SA. Systems biology of lens development: a paradigm for disease gene discovery in the eye. Exp Eye Res. 2016; doi: 10.1016/j.exer.2016.03.010

Anand D, Agrawal S, Siddam A, et al. An integrative approach to analyze microarray datasets for prioritization of genes relevant to lens biology and disease. Genom Data. 2015; 5:223–227. [PubMed: 26185746]

Anazi S, Maddirevula S, Faqeih E, et al. Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Mol Psychiatry. 2016; doi: 10.1038/mp.2016.113

Audette DS, Anand D, So T, et al. Prox1 and fibroblast growth factor receptors form a novel regulatory loop controlling lens fiber differentiation and gene expression. Development. 2016; 143(2):318–328. [PubMed: 26657765]

Charroux B, Pellizzoni L, Perkinson RA, et al. Gemin4 a novel component of the SMN complex that is found in both gems and nucleoli. J Cell Biol. 2000; 148(6):1177–1186. [PubMed: 10725331]

Dash S, Dang CA, Beebe DC, Lachke SA. Deficiency of the RNA binding protein caprin2 causes lens defects and features of Peters anomaly. Dev Dyn. 2015; 244(10):1313–1327. [PubMed: 26177727]

Fischer U, Liu Q, Dreyfuss G. The SMN–SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell. 1997; 90(6):1023–1029. [PubMed: 9323130]

Foster A, Gilbert C, Rahi J. Epidemiology of cataract in childhood: a global perspective. J Cataract Refract Surg. 1997; 23:601–604. [PubMed: 9278811]

Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004; 20(3):307–315. [PubMed: 14960456]

Gillespie RL, O'Sullivan J, Ashworth J, et al. Personalized diagnosis and management of congenital cataract by next-generation sequencing. Ophthalmology. 2014; 121(11):2124–2137. e2. [PubMed: 25148791]

Gillespie RL, O'Sullivan J, Ashworth J, et al. Personalized diagnosis and management of congenital cataract by next-generation sequencing. Ophthalmology. 2014; 121(11):2124–37. e1–2. [PubMed: 25148791]

Gillespie RL, Urquhart J, Anderson B, et al. Next-generation sequencing in the diagnosis of metabolic disease marked by pediatric cataract. Ophthalmology. 2016; 123(1):217–220. [PubMed: 26233629]

Group SM. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. Genome Biol. 2015; 16:134. [PubMed: 26112015]

Hejtmancik JF. Congenital cataracts and their molecular genetics. Semin Cell Dev Biol. 2008; 19(2): 134–149. [PubMed: 18035564]

Hoang TV, Kumar PKR, Sutharzan S, et al. Comparative transcriptome analysis of epithelial and fiber cells in newborn mouse lenses with RNA sequencing. Mol Vis. 2014; 20:1491–1517. [PubMed: 25489224]

Kasaikina MV, Fomenko DE, Labunskyy VM, et al. Roles of the 15-kDa selenoprotein (Sep15) in redox homeostasis and cataract development revealed by the analysis of Sep 15 knockout mice. J Biol Chem. 2011; 286(38):33203–33212. [PubMed: 21768092]

Keber R, Motaln H, Wagner KD, et al. Mouse knockout of the cholesterogenic cytochrome P450 lanosterol 14alpha-demethylase (Cyp51) resembles Antley–Bixler syndrome. J Biol Chem. 2011; 286(33):29086–29097. [PubMed: 21705796]

Khan AO. Hereditary pediatric cataract on the Arabian Peninsula. Saudi J Ophthalmol. 2012; 26(1): 67–71. [PubMed: 23960971]

Khan AO. Ocular genetic disease in the Middle East. Curr Opin Ophthalmol. 2013; 24(5):369–378. [PubMed: 23846189]

Khan AO, Aldahmesh MA, Alkuraya FS. Phenotypes of recessive pediatric cataract in a cohort of children with identified homozygous gene mutations (an American ophthalmological society thesis). Trans Am Ophthalmol Soc. 2015; 113:T7–1. [PubMed: 26622071]

Lachke SA, Alkuraya FS, Kneeland SC, et al. Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. Science. 2011; 331(6024):1571–1576. [PubMed: 21436445]

Lachke SA, Ho JW, Kryukov GV, et al. iSyTE: integrated Systems Tool for Eye gene discovery. Invest Ophthalmol Vis Sci. 2012a; 53(3):1617–1627. [PubMed: 22323457]

Lachke SA, Higgins AW, Inagaki M, et al. The cell adhesion gene PVRL3 is associated with congenital ocular defects. Hum Genet. 2012b; 131(2):235–250. [PubMed: 21769484]

Lassen N, Bateman JB, Estey T, et al. Multiple and additive functions of ALDH3A1 and ALDH1A1: cataract phenotype and ocular oxidative damage in Aldh3a1(−/−)/Aldh1a1(−/−) knock-out mice. J Biol Chem. 2007; 282(35):25668–25676. [PubMed: 17567582]

Lorbek G, Perse M, Jeruc J, et al. Lessons from hepatocyte-specific Cyp51 knockout mice: impaired cholesterol synthesis leads to oval cell-driven liver injury. Sci Rep. 2015; 5:8777. [PubMed: 25739789]

Lorson MA, Dickson AM, Shaw DJ, et al. Identification and characterisation of a nuclear localisation signal in the SMN associated protein, Gemin4. Biochem Biophys Res Commun. 2008; 375(1):33–37. [PubMed: 18675250]

Ma AS, Grigg JR, Ho G, Prokudin I, Farnsworth E, Holman K, Cheng A, Billson FA, Martin F, Fraser C, Mowat D, Smith J, Christodoulou J, Flaherty M, Bennetts B, Jamieson RV. Sporadic and familial congenital cataracts: mutational spectrum and new diagnoses using next-generation sequencing. Hum Mutat. 2016; 37(4):371–384. [PubMed: 26694549]

Makley LN, McMenimen KA, DeVree BT, et al. Pharmacological chaperone for alpha-crystalline partially restores transparency in cataract models. Science. 2015; 350(6261):674–677. [PubMed: 26542570]

Manthey AL, Lachke SA, FitzGerald PG, et al. Loss of Sip1 leads to migration defects and retention of ectodermal markers during lens development. Mech Dev. 2014; 131:86–110. [PubMed: 24161570]

Medsinge A, Nischal KK. Pediatric cataract: challenges and future directions. Clin Ophthalmol (Auckland, NZ). 2015; 9:77.

Musleh M, Hall G, Lloyd I, et al. Diagnosing the cause of bilateral paediatric cataracts: comparison of standard testing with a next-generation sequencing approach. Eye (Lond). 2016; 30(9):1175–1181. [PubMed: 27315345]

Patel N, Aldahmesh MA, Alkuraya H, et al. Expanding the clinical, allelic, and locus heterogeneity of retinal dystrophies. Genet Med. 2015; 18(6):554–562. [PubMed: 26355662]

Paznekas WA, Boyadjiev SA, Shapiro RE, et al. Connexin 43 (GJA1) mutations cause the pleiotropic phenotype of oculodentodigital dysplasia. Am J Hum Genet. 2003; 72(2):408–418. [PubMed: 12457340]

Rahi JS, Dezateaux C. Measuring and interpreting the incidence of congenital ocular anomalies: lessons from a national study of congenital cataract in the UK. Invest Ophthalmol Vis Sci. 2001; 42(7):1444–1448. [PubMed: 11381045]

Safieh LA, Khan A, Alkuraya F. Identification of a novel CRYAB mutation associated with autosomal recessive juvenile cataract in a Saudi family. Mol Vis. 2009; 15:980–984. [PubMed: 19461931]

Shiels A, Hejtmancik JF. Genetic origins of cataract. Arch Ophthalmol. 2007; 125(2):165–173. [PubMed: 17296892]

Shiels A, Hejtmancik J. Genetics of human cataract. Clin Genet. 2013; 84(2):120–127. [PubMed: 23647473]

Shiels A, Bennett TM, Hejtmancik JF. Cat-Map: putting cataract on the map. Mol Vis. 2010; 16:2007–2015. [PubMed: 21042563]

Stayte M, Reeves B, Wortham C. Ocular and vision defects in preschool children. Br J Ophthalmol. 1993; 77(4):228–232. [PubMed: 8494859]

Symoens S, Barnes AM, Gistelinck C, et al. Genetic defects in TAPT1 disrupt ciliogenesis and cause a complex lethal osteochondrodysplasia. Am J Hum Genet. 2015; 97(4):521–534. [PubMed: 26365339]

Wolf L, Harrison W, Huang J, et al. Histone posttranslational modifications and cell fate determination: lens induction requires the lysine acetyltransferases CBP and p300. Nucleic Acids Res. 2013; 41(22):10199–10214. [PubMed: 24038357]

Zhang H, Xie L, Wu X, Tian J. Long-term results of pediatric cataract surgery after delayed diagnosis. J Am Assoc Pediatr Ophthalmol Strabismus. 2012; 16(1):65–69.

Zhao L, Chen XJ, Zhu J, et al. Lanosterol reverses protein aggregation in cataracts. Nature. 2015; 523(7562):607–611. [PubMed: 26200341]

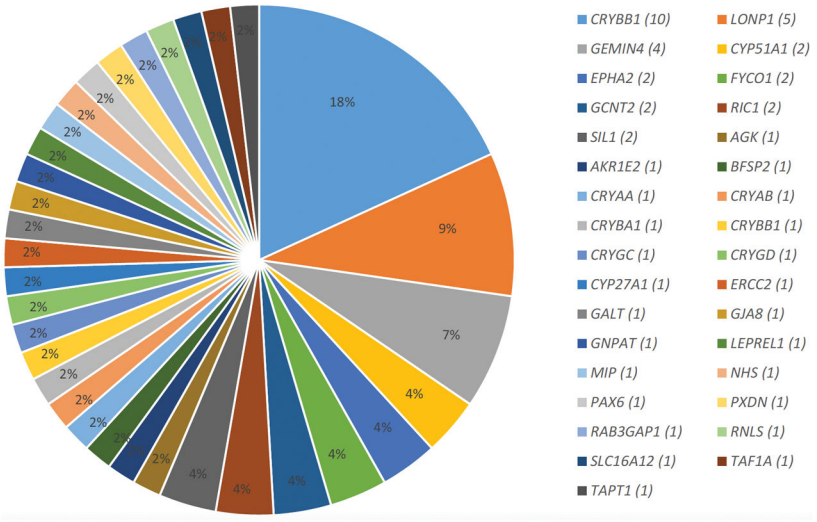

Fig. 1.

Expanding the allelic and locus heterogeneity of pediatric cataract. Distribution for mutations identified in known and novel candidate genes for cataract by NGS

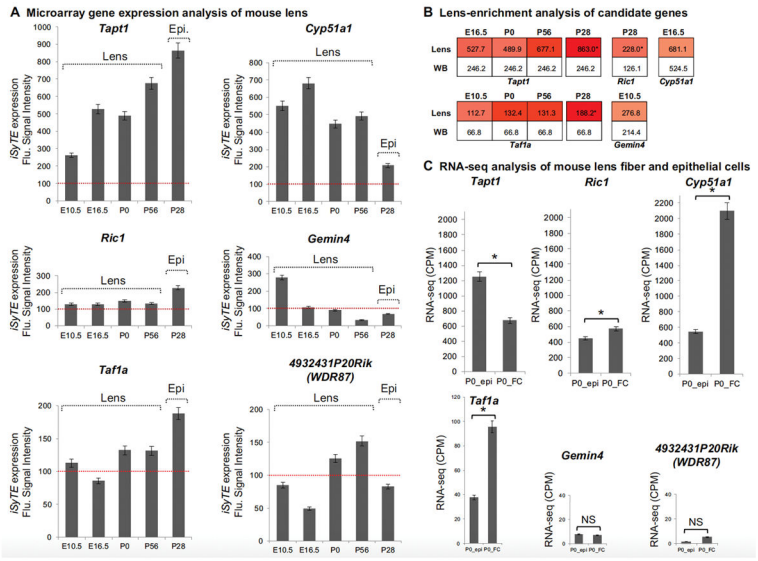

Fig. 2.

The mouse orthologs of the novel cataract candidate genes TAPT1, RIC1, CYP51A1, GEMIN4, TAF1A, WDR87 (4932431P20Rik) are expressed and enriched in lens development. (a) Lens expression of candidate genes Tapt1, Ric1, Cyp51a1, Gemin4, Taf1a, and 4932431P20Rik (WDR87) was analyzed in whole lens microarray datasets at mouse embryonic day (E) 10.5, E16.5 and postnatal day (P) 0, and P56, as well as isolated lens epithelium (Epi.) dataset at P28. The red dotted line in “a” indicates expression cut-off score of 100 fluorescence intensity units. (b) Lens-enrichment of candidate genes was evaluated by comparing their fluorescence expression intensity scores in the lens against that in the mouse whole embryonic body (WB) reference dataset. The color intensities in the heat map indicate the fold-change differences between lens expression over WB. (c) RNA-seq expression of newborn (P0) mouse isolated lens epithelium (epi) and fiber cells (FC). Error bars represent standard error of mean (SEM). Asterisk represents significant difference between comparisons in FC and Epi. expression ($p < 0.05$)

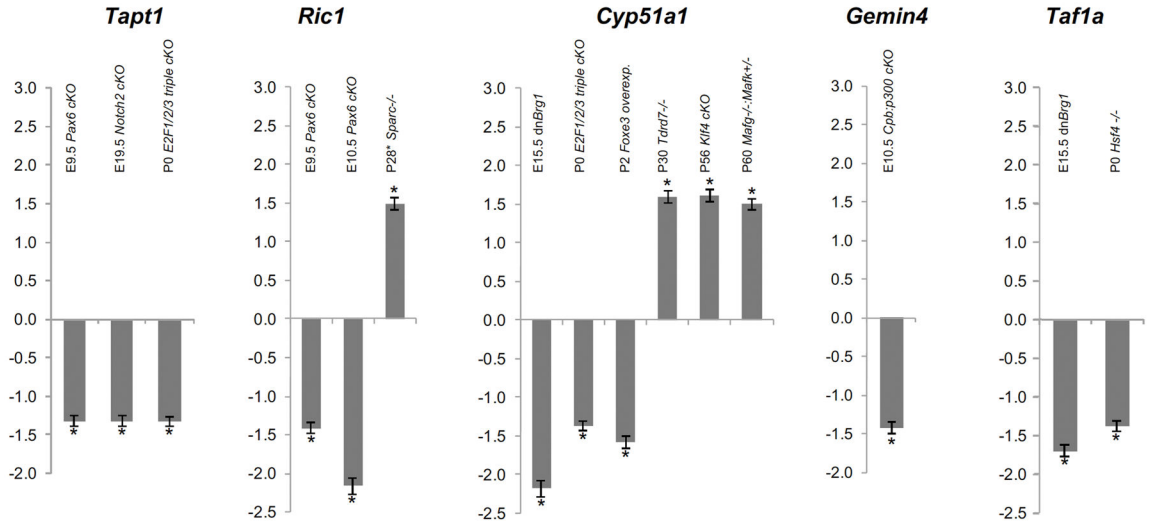

Fig. 3.

Tapt1, Ric1, Cyp51a1, Gemin4 and Taf1a are mis-regulated in targeted gene deletion mouse mutants with lens defects. Expression of candidate genes in various mouse mutants that exhibit lens defects including Pax6 lens-conditional null (Pax6 cKO) at E9.5 and E10.5, Notch2 cKO at E19.5, E2f1−/−:E2f2−/−:E2f3−/− triple cKO (E2f1/2/3 cKO) at P0, Sparc null at P28 (isolated lens epithelium only), Tdrd7 null at P30, Mafg−/−:Mafk+/− compound mutant at P60, Klf4 cKO at P56, Foxe3 lens overexpression mutant at P2, Cpb:p300 cKO mutant at E9.5, dnBrg1 mutant at E15.5 and Hsf4 null at P0. Differential expression in fold-change of candidate genes between mutant and control is plotted. Error bars represent standard error of mean (SEM). Asterisk represents significant expression differences between mutant vs. control lens datasets ($p < 0.05$)

Table 1

Overview of all the families included in the study and the outcome of NGS

Patient ID	Phenotype	Syndromic: yes/no	Mode of inheritance	Number of aff available	NGS outcome	Gene	Nomenclature	Zygosity	HGMD accession number	MAF: ExAC frequency	Ref
06-00549	Congenital cataract	No (intermittent 3-methylglutaconic aciduria but no clinical signs of Sanger's syndrome)	AR	3	Variant found in a known gene	AGK	NM_018238.3:c.424-3C>G	Homo	CS123669	0.00004947	PMID: 22415731
07DG0005	Congenital cataract (complete) with microcomea	No	AD	4	Unsolved						This study
07DG-0035/10DG1320	Pediatric posterior lenticonus cataract and global developmental delay	Yes	Simplex	1	Variant found in a novel candidate gene	RIC1	NM_020829.3:c.3794G>C;p.(Arg1265Pro)	Homo	Novel	Not reported	This study
08DG00152	Fetal (nuclear) cataract	No	AR	2	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	PMID: 22267527
09DG00751	Pulverulent cataract	No	AR	3	Variant found in a known gene	BFSP2	NM_003571.3:c.598_599dup; p.(Ala201Argfs*19)	Homo	CI118941	Not reported	PMID: 22935719
09DG01255	Pediatric cataract with posterior lenticonus	No	AR	2	Unsolved						This study
09DG01467	Congenital cataract	No	AR	2	Variant found in a known gene	GCNT2	NM_001491.2:c.1040A>G;p.(Tyr347Cys)	Homo	CM1212280	0.00004943	PMID: 22935719
09DG01472	Congenital cataract	No	AR	2	Variant found in a candidate gene (previously published)	RNLS	NM_001031709.2:c.215_216delinsT;p.(Lys72Ilefs*10)	Homo	CX1212286	Not reported	PMID: 22935719
10DG0339	Infantile cataract	No	AR	3	Variant found in a candidate gene (previously published)	AKR1E2	NM_001040177.2:c.582 + 1G>A	Homo	CS1212288	Not reported	PMID: 22935719
10DG0428	Pediatric cataract with posterior lenticonus	No	AR	1	Variant found in a known gene	EPHA2	NM_004431.3:c.1315C>T;p.(Pro439Ser)	Homo	Novel	Not reported	This study
10DG0498	Pediatric pulverulent cataract	No	AR	3	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	PMID: 22267527
10DG0703	Cataract and severe global developmental delay	Yes	AR	2	Variant found in a novel candidate gene	GEMIN4	NM_015721.2:c.2452T>C;p.(Trp818Arg)	Homo	CM150867	0.00001022	PMID: 25558065
10DG0870	Congenital cataract with microcomea	No	AD	2	Variant found in a known gene	CRYGD	NM_006891.3:c.134T>C;p.(Leu45Pro)	Homo	Novel	Not reported	This study
10DG1094	Pediatric pulverulent cataract	No	AR	2	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	PMID: 22267527
10DG1249	Congenital cataract	No	Pseudodominant (AR)	2	Variant found in a novel candidate gene	CYP51A1	NM_000786.3:c.829C>T;p.(Arg277Cys)	Homo	CM1212289	Not reported	PMID: 22935719
10DG1375	Congenital cataract	No	Pseudodominant (AR)	2	Variant found in a known gene	CRYAA	NM_000394.3: c.161G>C;p.(Arg54His)	Homo	CM1212054	0.0000165	This study
10DG1393	Congenital cataract	No	AR	2	Variant found in a known gene	FYCO1	NM_024513.3:c.2505delA;p.(Ala836Profs*80)	Homo	CD1212282	Not reported	PMID: 22935719
10DG1526	Congenital lamellar cataract	No	Simplex	1	Unsolved						This study
10DG1811	Congenital nuclear pulverulent cataract	No	AR	2	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	PMID: 22267527
10DG1895	Congenital cataract with microcomea	No	Simplex	1	Variant found in a known gene	PAX6	NM_000280.3:c.76C>T;p.(Arg26Trp)	Het	Novel	Not reported	This study
10DG1905	Congenital cataract with Peters anomaly	No	Simplex	1	Variant found in a known gene	PXDN	NM_012293.2:c.1018 + 1G>A	Homo	Novel	Not reported	This study
10DG1932	Pediatric cataract	No	AD	2	Unsolved						This study
10DG1948	Pediatric pulverulent cataract	No	Simplex	1	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	PMID: 22267527
10DG2001	Cataract, long face, bulbous nose, abnormal dentition	Yes	X-linked	6	Variant found in a known gene	NHS	NM_198270.2:c.2232del;p.(Lys744Asnfs*15)	Hemi	CD126120	Not reported	PMID: 22229851
11DG0108	Pediatric pulverulent cataract	No	Simplex	1	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	PMID: 22267527
11DG0190	Pediatric pulverulent cataract	No	Simplex	1	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	PMID: 22267527
11DG0228	Congenital cataract	No	AR	3	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	This study

Patient ID	Phenotype	Syndromic: yes/no	Mode of inheritance	Number of aff available	NGS outcome	Gene	Nomenclature	Zygosity	HGMD accession number	MAF: ExAC frequency	Ref
11DG0243	Congenital cataract	No	AR	4	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	This study
11DG0436	Nuclear pulverulent cataract, severe myopia	No	Simplex	1	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	PMID: 22267527
11DG0440	Pediatric cataract as part of cerebrotendinous xanthomatosis	Yes	AR	3	Variant found in a known gene	CYP27A1	NM_000784.3:c.1263 + 1G>A	Homo	CS961547	0.00004944	PMID: 22935719
11DG0619	Pediatric cataract with high hyperopia	No	Simplex	1	Unsolved						This study
11DG0994	Pediatric cataract with microcornea	No	AD	7	Variant found in a known gene	CRYGC	NM_020989.3:c.403G>T;p.(Glu135*)	Het	Novel	0.00005768	This study
11DG1104	Pediatric nuclear cataract	No	Simplex	1	Variant found in a known gene	LONP1	NM_004793.3:c.1612C>T;p.(Arg538Cys)	Homo	Novel	0.0003781	This study
11DG1176	Pediatric posterior subcapsular cataract	No	AR	2	Unsolved						This study
11DG1504	Congenital cataract, dysmorphic facies, cleft palate, severe global developmental delay, multiple renal cysts	Yes	AR	3	Unsolved						This study
11DG1744	Pediatric cataract	No	AR	4	Variant found in a known gene	CRYAB	NM_001885.2: c.166C>T;p.(Arg56Trp)	Homo	CM092933	Not reported	This study
11DG1761	Pediatric cataract	No	AD	6	Variant found in a known gene	GJA8	NM_005267.4:c.460C>G;p.(His154Asp)	Het	Novel	Not reported	This study
11DG2176	Congenital cataract and global developmental delay	Yes	Simplex	1	Variant found in a novel candidate gene	TAF1A	NM_001201536.1:c.40_41del;p.(Asp14*)	Homo	Novel	Not reported	This study
11DG2480	Congenital cataract, global developmental delay, epilepsy, mild epiphyseal dysplasia, nephrocalcinosis, brain hypomyelination and callosal thinning	Yes	AR	2	Variant found in a novel candidate gene	GEMIN4	NM_015721.2:c.2452T>C;p.(Trp818Arg)	Homo	CM150867	0.00001022	PMID: 25558065
11DG2497	Pediatric pulverulent-like cataract	No	AR	3	Unsolved						This study
12DG0105	Pediatric cataract, microcephaly, intellectual disability and dystonia	Yes	AR	4	Unsolved						This study
12DG0449	Pediatric nuclear cataract	No	Simplex	1	Unsolved						This study
12DG0750	Congenital cataract	No	Simplex	1	Variant found in a known gene	SLC16A12	NM_213606.3:c.404C>T;p.(Ala135Val)	Het	Novel	Not reported	This study
12DG1540	Congenital cataract	No	Simplex	1	Unsolved						This study
12DG2168	Congenital cataract	No	AR	2	Variant found in a known gene	GCNT2	NM_001491.2:c.1019A>G;p.(Tyr340Cys)	Homo	Novel	Not reported	This study
12DG2185	Pediatric nuclear cataract with microcomea	No	Simplex	1	Variant found in a known gene	MIP	NM_012064.3:c.530A>G;p.(Tyr177Cys)	Het	CM113696	Not reported	This study
12DG2369	Congenital cataract	No	AR	2	Variant found in a known gene	FYCO1	NM_024513.3:c.2714_2715del;p.(Thr905Serfs*2) &NM_024513.3:c.2345del;p.(Gln782Argfs*32)	Compound het	Novel	Not reported	This study
12DG2386	Congenital cataract	No	AR	2	Variant found in a novel candidate gene	WDR87	NM_031951.4:c.856G>T;p.(Glu286*)	Homo	CM1513752	0.0003235	PMID: 26622071
12DG2657	Pediatric posterior lenticonus cataract	No	AR	3	Variant found in a novel candidate gene	TAPT1	NM_153365.2:c.846 + 2insT	Homo	Novel	Not reported	This study
13DG0017	Congenital cataract	No	AR	5	Variant found in a known gene	LONP1	NM_004793.3:c.2014C>T;p.(Arg672Cys)	Homo	CM156332	0.00004278	PMID: 26622071
13DG0019	Congenital cataract	No	AR	2	Unsolved						This study
13DG0140	Bilateral cataract, deafness, developmental delay and nystagmus	Yes	Simplex	1	Unsolved						This study
13DG0323	Pediatric cataract	No	AR	2	Variant found in a known gene	CRYBB1	NM_001887.3:c.171del;p.(Asn58Thrfs*107)	Homo	CD072392	Not reported	This study
13DG0326	Pediatric cataract and ectopia lentis	No	AR	4	Variant found in a known gene	LEPREL1	NM_018192.3:c.297del;p.(Gly100Alafs*104)	Homo	CD151406	Not reported	PMID: 25469533
13DG0345	Pediatric cataract	No	AR	2	Variant found in a known gene	CRYBA1	NM_005208.4:c.588_591del;p.(Arg196Serfs*21)	Homo	CD1513750	Not reported	PMID: 26622071
13DG1449	Pediatric cataract as part of Micro Warburg syndrome	Yes	Simplex	1	Variant found in a known gene	RAB3GAP1	NM_012233.2:c.1009C>T;p.(Arg337*)	Homo	CM1510147	0.000008256	This study

Patient ID	Phenotype	Syndromic: yes/no	Mode of inheritance	Number of aff available	NGS outcome	Gene	Nomenclature	Zygosity	HGMD accession number	MAF: ExAC frequency	Ref
13DG1542	Congenital cataract, global developmental delay, tubulopathy and severe osteopenia	Yes	AR	2	Variant found in a novel candidate gene	GEMIN4	NM_015721.2:c.2452T>C;p.(Trp818Arg)	Homo	CM150867	0.00001022	PMID: 25558065
13DG1939	Pediatric cataract with iris coloboma	No	AD	3	Unsolved						This study
13DG2254	Pediatric cataract	No (no clinical signs of galactosemia)	Simplex	1	Variant found in a known gene	GALT	NM_000155.3 :c.200G>A.NM_000155.3:p.(Arg67His)	Homo	CM012753	Not reported	This study
14DG0067	Pediatric cataract	No	AR	2	Variant found in a known gene	LONP1	NM_004793.3:c.2014C>T;p.(Arg672Cys)	Homo	CM156332	0.00004278	This study
14DG0179	Congenital cataract as part of CODAS syndrome	Yes	Simplex	1	Variant found in a known gene	LONP1	NM_004793.3:c.44G>C;p.(Arg15Pro)	Homo	Novel	Not reported	This study
14DG0182	Pediatric posterior cataract	No	AD	1	Unsolved						This study
14DG0246	Congenital cataract	No	Simplex	1	Variant found in a known gene	LONP1	NM_004793.3:c.2014C>T;p.(Arg672Cys)	Homo	CM156332	0.00004278	This study
14DG0727	Pediatric cataract as part of Marinesco Sjogren syndrome	Yes	AR	1	Variant found in a known gene	SIL1	NM_ 022464.4:c.1030-9G>A;p.(Phe345Alafs*9)	Homo	CS083273	0.00004248	This study
14DG1268	Pediatric cataract as part of cerebrooculofacioskeletal syndrome	Yes	AR	2	Variant found in a known gene	ERCC2	NM_000400.3:c.1997G>A;p.(Arg666Gln)	Homo	Novel	Not reported	This study
14DG1505	Pediatric cataract as part of Marinesco Sjogren syndrome	Yes	Simplex	1	Variant found in a known gene	SIL1	NM_ 022464.4:c.1030-9G>A;p.(Phe345Alafs*9)	Homo	CS083273	0.00004248	This study
14DG1506	Unilateral persistent fetal vasculature cataract	No	Simplex	1	Unsolved						This study
14DG1568	Infantile cataract	No	AD	2	Variant found in a known gene	EPHA2	NM_004431.3:c.2007G>T;p.(Gln669His)	Het	Novel	Not reported	This study
14DG1618	Cataract, global developmental delay and brain atrophy	Yes	AR	8	Unsolved						This study
14DG1686	Pediatric cataract	No	AR	2	Unsolved						This study
14DG2068	Cataract as part of rhizomelic chondrodysplasia punctata	Yes	Simplex	2	Variant found in a known gene	GNPAT	NM_014236.3:c.487C>G;p.(Arg163Gly)	Homo	Novel	Not reported	This study
14DG2265	Cataract, global developmental delay, ataxia	Yes	Simplex	2	Variant found in a novel candidate gene	GEMIN4	NM_015721.2:c.314C>T;p.(Pro105Leu)	Homo	Novel	Not reported	This study
15DG2427	Intellectual disability, cleft lip and palate, strabismus, brain atrophy	Yes	AR	3	Variant found in a novel candidate gene	RIC1	NM_020829.3:c.3794G>C;p.(Arg1265Pro)	Homo	Novel	Not reported	This study
16DG0226	Congenital cataract, neonatal fulminant hepatic failure and global developmental delay	Yes	AR	2	Variant found in a novel candidate gene	CYP51A1	NM_000786.3:c.695T>C;p.(Leu232Pro)	Homo	Novel	Not reported	This study

All mutations have been confirmed by Sanger sequencing and segregated with all affected and unaffected family members available