Inherited cataracts: molecular genetics, clinical features, disease mechanisms and novel therapeutic approaches

Vanita Berry,1 Michalis Georgiou,1 Roy Quinlan,1,4 Anthony Moore,2,5 Michel Michaelides1,2

ABSTRACT
Cataract is the most common cause of blindness in the world; during infancy and early childhood, it frequently results in visual impairment. Congenital cataracts are phenotypically and genotypically heterogeneous and can occur in isolation or in association with other systemic disorders. Significant progress has been made in identifying the molecular genetic basis of cataract; 115 genes to date have been found to be associated with syndromic and non-syndromic cataract and 38 disease-causing genes have been identified to date to be associated with isolated cataract. In this review, we briefly discuss lens development and cataractogenesis, detail the variable cataract phenotypes and molecular mechanisms, including genotype–phenotype correlations, and explore future novel therapeutic avenues including cellular therapies and pharmacological treatments.

INTRODUCTION
Cataract is the most common, but treatable cause of blindness in the world. Recently, during the 70th World Health Assembly (October 2019), WHO estimated that 2.2 billion people are visually impaired around the world, out of which 65.2 million people are affected with cataract (https://www.who.int/publications-detail/world-report-on-vision). Congenital cataracts are detected at birth or during the first decade of life. WHO estimated that >14 million children are bilaterally blind from cataract, representing >50% of all causes of blindness globally.1 Congenital cataracts are present in 1/6–10 000 livebirths in developed countries and 5–15/10 000 births in developing countries. They are a prominent cause of vision loss in infants and children.2 3 Most vision loss is due to amblyopia, but some are due to post-operative complications such as glaucoma and retinal detachment.

approximately half of the congenital cataracts are characterised as inherited and are a clinical feature of nearly 200 syndromic genetic diseases,4 including for instance diabetes and cholesterol metabolism diseases. Congenital cataract was the first autosomal disease to identify lens development and cataractogenesis, detail the variable cataract phenotypes and molecular mechanisms, including genotype–phenotype correlations, and explore future novel therapeutic avenues including cellular therapies and pharmacological treatments.

LENS EMBRYOLOGY AND MORPHOLOGY
The ocular lens is a unique model to understand the important aspects of embryonic development, signalling, induction, cell differentiation, cell physiology, biochemistry, organelle degradation and cellular longevity. Several reviews can provide a detailed description.6 13 14

During gastrulation (day 22, Carnegie stage 9 in human development), a single eye field forms in the middle of the anterior neural plate, which separates into two optic vesicles and further induce the nearby surface ectoderm to form the lens placode (a precursor of the lens) by day 28 (Carnegie stages 12–13) (figure 1A,B). At this stage, a series of inductive interactions begin to shape the eye, orchestrated by signalling molecules such as bone morphogenetic proteins and fibroblastic growth factor 2, as well as TFs such as OTX2, PAX6 and PITX3.15 This programme can be mimicked in culture conditions to determine the differentiation pathway of pluripotent cells and their derivatives.16

The lens placode invaginates to form the lens pit (figure 1C), which makes a complete circle of cells and detaches from the surface ectoderm to develop into the lens vesicle (figure 1D). By the end of week 4 (Carnegie stages 10–13), the cells from the posterior vesicle start elongating towards the anterior epithelial cell layer to become the primary lens fibres that fill the lens vesicle and later become the embryonic nucleus of the mature lens (figure 1E). The portion of the optic vesicle that faced the lens placode gives rise to the lamellar, coralliform, blue dot (cerulean), cortical, pulvulrent and polymorphic.9–11 The majority of the inherited cataracts are autosomal dominant with complete penetrance, but variable expression, autosomal recessive and X-linked inheritance patterns are less frequent.

Over the last 10 years, enormous progress has been made in elucidating the molecular basis of congenital cataract, with causative mutations identified in genes encoding many different proteins including intracellular lens proteins (crystallins), membrane gap junction proteins (connexins), water channel proteins (aquaporins), cytoskeletal proteins (eg, BFSP1 (filensin), BFSP2 (phakinin) and vimentin) and transcription factors (TFs) (eg, FOXE3, PAX6, PITX3 and MAFA) (table 1). Recent advances in molecular genetics, particularly next-generation sequencing, has improved molecular diagnosis in the clinic.12
The retina, in turn, provides inductive signals that regulate the growth and apical-posterior axis of the lens. In the early optic cup stage, the lens vesicle releases signals that induce the overlying surface ectoderm to differentiate into the corneal epithelium. Around weeks 6–7 (Carnegie stages 16-19), lens fibres start to develop from the epithelial cells located at the equator where they begin to elongate and differentiate into the secondary lens fibres (fetal nucleus) of the developing lens (figure 1F). Around week 8 (Carnegie stage 20), the Y-shaped suture appears at the anterior and posterior poles of the embryonic nucleus of the lens as a result of the terminal ends of the secondary lens fibres abutting each other. The newly differentiated fibre cells continue to grow throughout life. During this process of terminal differentiation, fibre cells remove their nucleus and other cell organelles to minimise light scattering.

### Cataract Phenotypes and Modes of Inheritance

The congenital cataract phenotype broadly reflects spatiotemporal insults experienced by the developing lens. Broadly these phenotypes can be divided into eight distinct clinical appearances: (1) nuclear (figure 2A), (2) anterior polar (figure 2B), (3) posterior polar (figure 2C), (4) lamellar, (5) coralliform, (6) blue dot (cerulean – figure 2D), (7) cortical (figure 2E) and (8) pulverulent (figure 2F).

There is considerable phenotypic variation in autosomal cataract, although there have been few systematic studies recording differences for specific mutations. To date (August 2019), 1314 novel and recurrent disease-causing sequence variants have been identified across 46 genes, including 45 autosomal-recessive (AR) genes, 21 autosomal-dominant (AD) genes, and 4 sex-linked genes.

### Table 1: Genes Implicated in Inherited Cataract

<table>
<thead>
<tr>
<th>Gene classification</th>
<th>Gene</th>
<th>Gene locus</th>
<th>Inheritance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystallins</td>
<td>CRYAA</td>
<td>21q22.3</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>CRYAB</td>
<td>11q22.1-q23.2</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>CRYBA1</td>
<td>17q11.2-q12</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>CRYBA2</td>
<td>2q34</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>CRYBA4</td>
<td>22q12.1</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>CRYBB1</td>
<td>22q12.1</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>CRYBB2</td>
<td>22q11.23</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>CRYBB3</td>
<td>22q11.23</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>CRYGA</td>
<td>2q33q35</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>CRYGB</td>
<td>2q33q35</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>CRYGC</td>
<td>2q33q35</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>CRYGD</td>
<td>2q33q35</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>CRYGS</td>
<td>3q25-pter</td>
<td>AD</td>
</tr>
<tr>
<td>Lens membrane protein</td>
<td>GJA1</td>
<td>13q11-q12</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>GJA8</td>
<td>1q21.1</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>MIP</td>
<td>12q13</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>TMEM114</td>
<td>16p13.2</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>EPHA2</td>
<td>1p36</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>SLC16A12</td>
<td>10q32.13.13</td>
<td>COMPLEX</td>
</tr>
<tr>
<td></td>
<td>LIM2</td>
<td>19q13.4</td>
<td>AR</td>
</tr>
<tr>
<td>Developmental factors</td>
<td>PAX6</td>
<td>11p13</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>PITX3</td>
<td>10q25</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>FOXE3</td>
<td>1p32</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>MAF</td>
<td>16q22-q23</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>EYA1</td>
<td>8q13.3</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>VSS2</td>
<td>14q24.3</td>
<td>AR</td>
</tr>
<tr>
<td></td>
<td>HSF4</td>
<td>16q21</td>
<td>AD/AR</td>
</tr>
<tr>
<td>Cytoskeletal proteins</td>
<td>Bfsp1</td>
<td>20p11.23-p12.1</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>Bfsp2</td>
<td>3q21-q22</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>VIM</td>
<td>10p13</td>
<td>AD/AR</td>
</tr>
<tr>
<td></td>
<td>NHS</td>
<td>Xp22.13</td>
<td>XL</td>
</tr>
<tr>
<td>Genes with special roles in the lens</td>
<td>TDRD7</td>
<td>9q22.33</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>CHMP4B</td>
<td>20q11.22</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>WFS1</td>
<td>4p16.1</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>FFOC1</td>
<td>3p21.31</td>
<td>AR</td>
</tr>
<tr>
<td></td>
<td>GCNT2</td>
<td>6p24.2</td>
<td>AR</td>
</tr>
<tr>
<td></td>
<td>FL1</td>
<td>19q13.33</td>
<td>AD</td>
</tr>
<tr>
<td></td>
<td>AGK</td>
<td>7q34</td>
<td>AR</td>
</tr>
</tbody>
</table>

As described in Cat-Map, an online chromosome map and reference database for inherited and age-related forms of cataract(s) in humans, mice and other vertebrates (https://cat-map.wustl.edu/).
identified (http://cat-map.wustl.edu/), with a well-defined distinct phenotype observed in 366 cases as shown in figure 3A. Interestingly, the most frequent phenotype was nuclear cataract (34%), followed by posterior polar (14%), total cataract (11%), lamellar (10%), cerulean and coralliform both (4%), and anterior polar (2%), and the remaining were age-related cataracts (21%) (figure 3B).

The most common mode of inheritance is autosomal dominant (44%), followed by autosomal recessive (18%), X-linked recessive (6%) with the remaining 32% being sporadic with no clear family history (https://cat-map.wustl.edu/) (figure 3A). Improvements in molecular diagnosis using next-generation sequencing will allow a better understanding of the mode of inheritance in sporadic cases, to establish whether the variant is homozygous/recessive or heterozygous/dominant.

MOLECULAR GENETICS
Disease-causing sequence variants in 115 genes have been identified in isolated and syndromic inherited cataract. Thirty-eight genes are responsible for distinct isolated cataract and can be divided into five main groups based on the encoded proteins: (1) crystallins, (2) membrane proteins, (3) transcription factors, (4) cytoskeletal proteins and (5) gene products with special roles in the lens (figure 4).

Crystallins
Crystallins (α, β and γ) account for nearly 90% of all lens proteins; they are crucial for the lens optical properties and its function, for the remarkable resilience to post-translational modifications due to ageing and the environment to maintain lens transparency.

Alpha-crystallins are members of the small heat shock protein family, which are molecular chaperones protecting lens proteins and enzymes from aggregation, which could otherwise lead to lens opacification. Alpha-crystallin comprises two related subunits (αA polypeptide and αB polypeptide) encoded by CRYAA on chromosome 21q22.3 and CRYAB on chromosome 11q22-q22.3, respectively. Variants in CRYAA cause both autosomal-dominant and autosomal-recessive cataracts and account for 25.8% of all crystallin mutations. CRYAA is primarily expressed in the lens; there is also extralenticular expression detected in mouse retina and cornea.

CRYAB is expressed in the lens epithelial cells and in the retina, skeletal muscle, heart, kidney and brain. Mutations in CRYAB cause not only cataract but also myopathies there being heart and lens-specific enhancers to regulate its expression in these tissues. In 2001, Berry and colleagues found the first dominant heterozygous mutation in a large

Figure 3 Frequency pie charts. (A) Frequency of various mode of inheritance. (B) Frequency of cataract phenotypes in families seen to date. AD, autosomal dominant; AR, autosomal recessive; XL, X-linked.

Figure 4 Cataract-causing variants by gene. Total number of 1314 disease-causing variants (novel and recurrent) are shown in various highly expressed genes to date (http://cat-map.wustl.edu/), including crystallins, gap junction proteins, membrane proteins, developmental and cytoskeletal proteins in the lens; remaining 476 variants are found in various other genes important for the normal function of the crystalline lens.
family with posterior polar cataract.\textsuperscript{27} To date, 22 mutations have been reported in the \textit{CRYAB} gene in both autosomal-dominant and autosomal-recessive cataract. Heterozygous missense variants have also been described in patients with desmin-related myopathies and cardiomyopathy.\textsuperscript{28}

The \(\beta\gamma\)-crystallins are derived by gene duplication and belong to a large superfamily that includes such proteins as AIM1.\textsuperscript{24} They consist of four homologous Greek key motifs organised into two domains. The \(\beta\gamma\)-crystallin family comprises four acidic (A) and three basic (B) forms encoded by genes—\textit{CRYBA2}, \textit{CRYBA1} and \textit{CRYBB3}, \textit{CRYBB2}, \textit{CRYBB1} and \textit{CRYBA4}, respectively. \(\gamma\)-Crystallin is encoded by the \(\gamma\)-gene cluster on chromosome 2q33–35 encompassing genes \(\gamma\)A to \(\gamma\)D. Fewer mutations are found in the \(\gamma\)A (\textit{CRYGA}) and \(\gamma\)B (\textit{CRYAB}) compared with \(\gamma\)C (\textit{CRYGC}) and \(\gamma\)D(\textit{CRYGD}).\textsuperscript{29–42}

Interestingly, most of the variants in \textit{CRYGC} and \textit{CRYGD} genes cause nuclear cataract and coralliform phenotypes inherited dominantly. A single \(\gamma\)S-crystallin gene resides on the long arm of chromosome 3q25-qter; mutations in this gene are mainly causing autosomal-dominant disease with various phenotypes. Crystallins is the most prevalent proteins in the lens; nearly 294 variants have been found in crystallins, which account for nearly 22.4% of all the inherited cataract variants found to date (figure 4).

**Membrane proteins**

**Connexins**

A sophisticated cell-to-cell communication network is important in maintaining lens cell homeostasis. In the developing lens, this communication is maintained via gap junction channels which allow the flow of ions, second messengers and metabolites between lens fibre cells. In the lens, these channels are made up of three connexin isoforms: \textit{GJA1} (Cx43), \textit{GJA3} (Cx46) and \textit{GJA8} (Cx50). Six connexin molecules assemble to form a hemichannel or connexon. Each connexon is either made from a single type of connexin, or from more than one type, which leads to either homomeric or heteromeric hemichannels, respectively. These hemichannels dock with a counterpart in an adjacent cell to make a gap junction channel linked by their extracellular loops.\textsuperscript{13} 34 \textit{GJA1} is expressed only in the lens epithelial cells during early stages of lens development, but not associated with lens pathology.\textsuperscript{15} 36 Fifty-five heterozygous variants and 1 homozygous variant have been found in \textit{GJA3}, with various associated lens phenotypes, including pupilvulner, nuclear, lamellar, coralliform and total (http://cat-map.wustl.edu/). To date, 90 heterozygous variants have been described in families with autosomal-dominant cataract, and a single homozygous variant in autosomal-recessive cataract has been found in \textit{GJA8}, associated not only with inherited cataract but also age-related cataract and other eye anomalies including microcornea, microphthalmia and corneal opacification.\textsuperscript{37–39}

**MIP**

\textit{MIP26} is the major intrinsic protein of the lens, encoded by \textit{AQP0}, a member of the ubiquitous family of water channel proteins called aquaporins that allow rapid movements of water across cell membranes. \textit{MIP} is highly expressed in terminally differentiated lens fibre, comprising nearly half of the total lens fibre cell membrane proteins.\textsuperscript{38} Berry and colleagues\textsuperscript{39} 40 identified two autosomal-dominant variants (G134E and T138R) leading to polymorphic and lamellar cataract. So far, 37 heterozygous variants have been found in \textit{MIP} causing autosomal-dominant cataract.\textsuperscript{7}

**LIM2**

\textit{LIM2} is a lens-specific integral membrane protein, also referred to as MP19. This gene encodes an eye lens-specific protein found at the junctions of lens fibre cells, where it may contribute to cell junctional organisation. It acts as a receptor for calmodulin and may play an important role in both lens development and cataractogenesis. Mutations in \textit{LIM2} have been associated recessive cataracts and age-related cataracts.\textsuperscript{31–43}

**Transcription factors**

TFs including \textit{PAX6}, \textit{FOXE3}, \textit{HSF4}, \textit{MAF} and \textit{PITX3} among many others play an important role in lens development.\textsuperscript{13} 14 \textit{PAX6}, the paired-box gene, is one of the key players in vertebrate eye development. In the lens, \textit{PAX6} plays a key role in the regulation of lens-specific crystallins.\textsuperscript{44} 45 and thus is an important TF in lens development. Heterozygous null mutations in \textit{PAX6} typically give rise to aniridia often with associated lens opacities but missense changes may result in milder phenotypes including foveal hypoplasia, Peters anomaly and isolated cataract.\textsuperscript{46} 47

\textit{FOXE3} is a forkhead-box TF, required for morphogenesis and differentiation of the anterior segment of the eye.\textsuperscript{48} Semina \textit{et al} reported the first human variants in this gene causing anterior segment mesenchymal dysgenesis and congenital cataracts.\textsuperscript{49} More than 20 homozygous and heterozygous variants have been reported, displaying severe developmental eye anomalies including cataract.\textsuperscript{50}

\textit{MAF}, is a basic region leucine zipper (bZIP), an oncogene, which is expressed in early lens development.\textsuperscript{51} \textit{MAF} regulates the expression of the lens crystallins.\textsuperscript{52} Mutations in \textit{MAF} are not only responsible for cataract and ocular abnormalities but also cause Aymé-Gripp syndrome.\textsuperscript{53} 54 Recently, a missense mutation in \textit{MAFA} gene has been reported to cause autosomal-dominant insulinomatosis and diabetes mellitus along with cataract and glaucoma.\textsuperscript{55}

\textit{PITX3} gene is a member of the \textit{REIG/PITX} family of homeobox TFs.\textsuperscript{56} To date, 26 variants in \textit{PITX3} have been identified (including a hotspot in exon 4, c.640_656dup17bp) to cause mainly posterior cataracts and anterior segment dysgenesis in different ethnicities.\textsuperscript{57}

**Cytoskeletal proteins**

The cytoskeleton of a cell comprises three major filaments: microfilaments, microtubules and intermediate filaments. In the lens, two beaded filaments, a type of intermediate filament, \textit{BFSF1} (filensin) and \textit{BFSF2} (phakinin), are expressed.\textsuperscript{58} Several variants in \textit{BFSF2} lead to sutural opacities and nuclear cataract in association with \textit{BFSF1} variants. NHS (Nance-Horan Syndrome) is also associated with abnormalities in the lens cytoskeleton and epithelial cell junctions. A total of 55 sequence variants in the gene underlying the X-linked dominant NHS have been identified. Affected men have dense nuclear cataracts and frequently have microcornea, whereas heterozygous women show sutural cataracts with microcornea, craniofacial dysmorphism, nystagmus, strabismus and dental anomalies.\textsuperscript{59} 60 (table 1).

Studies in various animal models have identified several genes and their pathways linked to lens defects and cataract that are good candidates for examining human patients; for example, iSyTE (https://research.bioinformatics.udel.edu/iSyTE/ppi) has impacted many cataract gene discoveries both in human and...
animal models. Recently, Siddam et al have shown that deficiency of Celf1 (RNA binding protein) causes cataract in fish and mouse, and serves as a potential candidate to examine in patients with cataract.

CATARACT THERAPEUTICS AND FUTURE DIRECTION

Understanding cataractogenesis: latency and ageing

The timing of the appearance of cataract depends on whether it is due to a harmful mutation or primarily due to accumulated biomolecular damage. Both can be described as accumulated cataractogenic load, manifesting as cataract either during infancy or at other life stages. Genetic, environmental and stochastic factors all contribute to ageing of the lens and presbyopia at the end of the fourth decade of life is evidence of this. Presbyopia is both a consistent, universal, totally penetrant phenotypic change in the ageing human lens. In line with ageing hypotheses, treatment with antioxidants such as a lipoic acid derivative can reverse presbyopia in mouse models. Indeed, Nactey cysteine derivatives are also effective cataract-reversing molecules in other animal models, indicating that cataract, though a histopathological endpoint, is not necessarily irreversible if treatment and timing are optimised. For each individual, the manifestation of their cataract will depend on their accumulated cataractogenic load due to various genetic and environmental insults. This explains why the age of onset for age-related cataract is so variable and why for congenital cataract the age of onset is much more consistent. This can be interpreted as representing a threshold.

The lens relies on biomolecular longevity because of the lifelong retention of its components and particularly the proteins. The lens continues to grow throughout life and as such bimolecular lifelong retention of its components, and particularly the proteins, racemization/isomerisation and advanced glycation end-modified either arrest or even reverse cataract.

Small molecules

Cataracts are a major burden on public health, and effective therapeutic approaches are yet to be established for preventing or mitigating the process. Lens opacity triggers due to malfunction of alpha-crystallins, when they lose their chaperone properties either due to genetic basis or ageing. Due to their reduced functional capacity, crystallin tends to misfold the protein and aggregate into amyloids. Several studies have suggested few compounds that can reverse the light scattering lens aggregates in vitro and in animal studies.

Recently, Makley and colleagues identified a small molecule, ‘compound 29’, which binds to several members of the small heat shock protein family including the alpha-crystallins. This compound, an oxysterol, was shown to improve lens transparency in the mouse model and able to restore the protein solubility in the lenses of aged mice in vivo and in human lenses in vitro. These findings suggest an approach to treating cataracts by stabilising alpha-crystallins.

Another group identified variants in the LSS gene in two autosomal-recessive congenital cataract families. The LSS gene encodes lanosterol synthase, a key enzyme in the cholesterol biosynthesis pathway, is highly expressed in the lens and has water-lipid solubility properties. Lanosterol treatment has shown to reverse crystallin aggregation in vitro. It was also suggested to improve lens transparency in rabbits and dogs with naturally occurring cataract. Due to its solubility properties, lanosterol has been suggested as a possible topical eye-drop therapy for cataract.

CONCLUSIONS

The identification of genetic variants causing congenital cataract has not only improved our understanding of the pathogenesis of infantile cataract, the most frequently treatable cause of blindness in childhood, but also its more common counterpart, adult-onset cataract. For example, LIM2, EPHA2 and TDRD7 that have variants described in both congenital and age-related cataract. It is thus pivotal to identify genetic variations associated with a high risk of developing cataract. This may lead towards new strategies for the prevention of cataract or mitigating the progression of early lens opacity, thus reducing the global huge demand for surgery. There is immense potential in cataract research to elucidate the relationship between neurological and vascular diseases.
References


