Molecular Genetic Analysis of Pakistani Families With Autosomal Recessive Congenital Cataracts by Homozygosity Screening

Jianjun Chen,1,2 Qiwei Wang,1,3 Patricia E. Cabrera,1 Zilin Zhong,1,2 Wenmin Sun,1,3 Xiaodong Jiao,1 Yabin Chen,1 Gowthaman Govindarajan,1 Muhammad Asif Naeem,4 Shaheen N. Khan,4 Muhammad Hassaan Ali,5 Muhammad Zaman Assir,5 Fawad Ur Rahman,6 Zaheeruddin A. Qazi,6 Sheikh Riazuddin,4,5,7 Javed Akram,5,7 S. Amer Riazuddin,8,9 and J. Fielding Hejtmancik1

1Ophthalmic Genetics and Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, Maryland, United States
2Department of Ophthalmology, Shanghai Tenth People’s Hospital, and Tongji Eye Institute, Tongji University School of Medicine, Shanghai, China
3State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, Guangdong, China
4National Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan
5Allama Iqbal Medical College, University of Health Sciences, Lahore, Pakistan
6Layton Rahmatulla Benevolent Trust Hospital, Lahore, Pakistan
7National Centre for Genetic Diseases, Shaheed Zulfiqar Ali Bhutto Medical University, Islamabad, Pakistan
8The Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States
9McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States

Correspondence: J. Fielding Hejtmancik, M0GS/OGVFBEHI/NIH, Room 1120, 5635 Fisher’s Lane, Rockville, MD 20852, USA; f5h@helix.nih.gov.
JC and QW are joint first authors.
SAR and JFH contributed equally to the work presented here and should therefore be regarded as equivalent authors.
Submitted: January 11, 2017
Accepted: February 24, 2017

PURPOSE: To identify the genetic origins of autosomal recessive congenital cataracts (arCC) in the Pakistani population.

METHODS. Based on the hypothesis that most arCC patients in consanguineous families in the Punjab areas of Pakistan should be homozygous for causative mutations, affected individuals were screened for homozygosity of nearby highly informative microsatellite markers and then screened for pathogenic mutations by DNA sequencing. A total of 83 unmapped consanguineous families were screened for mutations in 33 known candidate genes.

RESULTS. Patients in 32 arCC families were homozygous for markers near at least 1 of the 33 known CC genes. Sequencing the included genes revealed homozygous cosegregating sequence changes in 10 families, 2 of which had the same variation. These included five missense, one nonsense, two frame shift, and one splice site mutations, eight of which were novel, in EPHA2, FOXE3, FYCO1, TDRD7, MIP, GALK1, and CRYBA4.

CONCLUSIONS. The above results confirm the usefulness of homozygosity mapping for identifying genetic defects underlying autosomal recessive disorders in consanguineous families. In our ongoing study of arCC in Pakistan, including 83 arCC families that underwent homozygosity mapping, 3 mapped using genome-wide linkage analysis in unpublished data, and 30 previously reported families, mutations were detected in approximately 37.1% (43/116) of all families studied, suggesting that additional genes might be responsible in the remaining families. The most commonly mutated gene was FYCO1 (14%), followed by CRYBB3 (5.2%), GALK1 (3.5%), and EPHA2 (2.6%). This provides the first comprehensive description of the genetic architecture of arCC in the Pakistani population.

Keywords: homozygosity mapping, genetic analysis, autosomal recessive congenital cataracts, consanguineous

Congenital cataract (CC) is a significant cause of vision loss worldwide, causing approximately one-third of blindness in infants. Approximately one-third of CCs are familial; the cataract may be isolated or be associated with other systemic abnormalities. Nonsyndromic CCs may account for approximately 70% of CC cases, and have an estimated frequency of 1 to 6 per 10,000 live births. Congenital cataracts are clinically and genetically heterogeneous, with approximately 8.3% to 25.0% of nonsyndromic CCs being inherited, approximately 7% as autosomal recessive (ar), 76% to 89% as autosomal dominant (ad), or 2% to 10% as X-linked traits in European populations. Currently more than 48 CC loci have been identified, and more than 35 of them have been associated with causative mutations in specific genes, as delineated in the Cat-Map database (http://cat-map.wustl.edu/, in the public domain). Individuals in families having CCs as a result of the same mutation can show variable severity and morphology, probably reflecting effects of differences in their genetic backgrounds or environmental conditions.
factors. Conversely, cataracts with similar morphologies can result from mutations in genes involved in disparate biological pathways, suggesting that cataract is a final endpoint for a variety of different biological insults. That causative mutations have been identified in only a subset of patients with CAs suggests that additional cataract genes have yet to be identified. Although the fraction of families with cataracts caused by uncharacterized genes appears to vary in different populations, it is difficult to estimate from existing studies, most of which examine a subset of candidate genes and do not represent an exhaustive characterization of a random set of families.

As part of an ongoing collaboration between the National Eye Institute (Bethesda, MD, USA) and the National Centre of Excellence in Molecular Biology and Allama Iqbal Medical College in Lahore, Pakistan, this study was designed to identify the genes underlying arCC in the Pakistani population. We screened 83 unrelated arCC families for homozygosity at 35 genes or loci commonly involved in CC and related disorders for possible involvement in disease. In 32 families showing homozygous regions encompassing known CC genes, the respective genes were sequenced, identifying nine disease-causing mutations in 10 families. Overall, including 5 unpublished and 30 previously reported families,10–20 mutations or loci were detected in 43 of arCC families tested, consistent with known cataract genes or loci being responsible for cataracts in 37.1% of the entire set of families.

Subjects and Methods
Ascertainment of Families and Clinical Analysis
This study was approved by institutional review boards (IRB) of the National Centre of Excellence in Molecular Biology and the Combined Neuroscience (CNS) IRB at the National Institutes of Health. Participating subjects gave informed consent consistent with the tenets of the Declaration of Helsinki. Ophthalmological examinations were performed at the Layton Rehmatullah Benevolent Trust Hospital in Lahore, Pakistan. Detailed family and medical histories were obtained from family members. Presence and types of cataract in both affected and unaffected individuals of the families were confirmed by slit lamp biomicroscopy. A cohort of 143 CC consanguineous families was collected over a period of 10 years. Genome-wide linkage analyses using 384 highly polymorphic microsatellite markers and Sanger sequencing had identified a molecular diagnosis in 3 unpublished and 30 previously reported families.10–21 From the remaining 110 unlinked families, 83 were selected for homozygosity screening analysis based on the availability of DNA samples and an ar inheritance pattern that in addition included consanguinous matings. In addition, control DNA samples were available from 96 unrelated, ethnically matched Pakistani individuals. Blood samples were obtained from study participants, and DNA was extracted using standard methods, as previously described.22

Homozygosity Mapping
Thirty-three candidate genes and loci involved in CC and related disorders or based on expression and function were selected for screening (Table 1). Each was screened for homozygosity by genotyping 1 or 2 microsatellite markers (total of 51 markers; Table 1). The screening algorithm and a summary of the results are shown in Figure 1. The microsatellite markers were selected based on reported high heterozygosity (0.75 or more) and were located within 1 to 2 megabases (Mb) of the candidate gene. If a single marker with 75% or greater heterozygosity was not available, two markers were genotyped. Information on the PCR primer reaction conditions, heterozygosity, and location was obtained from the UniSTS Human Genome Database and National Center for Biotechnology Information (NCBI) Mapview databases. The detection of homozygosity at a given locus in an affected family member was followed by genotyping a second affected family individual at the locus. A variant of the multiplexing short tandem repeat with tailed primers approach described by Oetting et al.23 using fluorescently labeled tagged primers homologous to extensions on initial primers in a two-PCR approach was used to genotype these microsatellite markers. The PCR products were multiplex electrophoresed on an ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA), and fragment sizes were determined by GeneMapper version 4.0 (Applied Biosystems). Primer sequences and PCR conditions are shown in Supplementary Table S1. If the second individual also was homozygous, linkage was carried out with all available family members to confirm cosegregation of markers with disease. Polymerase chain reaction products were separated on an ABI 3130 DNA Analyzer (Applied Biosystems), and alleles were assigned with GeneMapper Software version 4.0 (Applied Biosystems).

Linkage Analysis
Haplotype comparisons used the Cyrillic 2.1 program (Cyrillic Software, Wallingford, Oxfordshire, UK) for inspection to identify homozygous regions common to affected individuals in each family. Two-point linkage analyses were performed with the FASTLINK version of MLINK from the LINKAGE Program Package.24,25 Maximum logarithm of the odds (LOD) scores were calculated with ILINK from the LINKAGE Program Package. Autosomal recessive cataracts were analyzed as a fully penetrant trait with a disease allele frequency of 0.0001, and mutation frequency of 0. The marker order and physical distances between the markers were obtained from the Marshfield database and the NCBI chromosome sequence maps.

DNA Sequencing
Mutation screening of candidate gene coding regions used PCR amplification of exons and adjacent intronic regions. Primer pairs for individual exons in the critical interval were designed online with the Primer3 program (http://primer3.sourceforge.net/, in the public domain). Polymerase chain reaction primers for each exon were used for bidirectional sequencing with Big Dye Terminator Ready reaction mix per instructions of the manufacturer (Applied Biosystems). Sequencing was performed using ABI PRISM 3130 automated sequencers (Applied Biosystems) and analyzed using Mutation Surveyor (Soft Genetics, Inc., State College, PA, USA) and the Seqman program of DNASTAR Software (DNASTAR, Inc., Madison, WI, USA). Sequence changes observed were checked for cosegregation in the family and for presence or absence in at least 96 healthy control individuals as well as the 1000 Genomes (http://www.internationalgenome.org/home, in the public domain) and ExAC (http://exac.broadinstitute.org/, in the public domain) databases, although low frequencies of heterozygous changes were not considered to exclude pathogenicity.

Pathogenicity Assessment of Identified Variants
A mutation was considered novel if it was not present in the Human Mutation Database (http://www.hgmd.cf.ac.uk/ac; in the public domain) or the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/index.html, in the public do-
main) and not in Cat-Map (http://cat-map.wustl.edu, in the public domain). A sequence variation was considered pathogenic when it cosegregated with the disease in the family; was not present in 96 randomly selected controls from the Pakistani population; altered a well-conserved amino acid, not present in 96 randomly selected controls from the Pakistani population; and was observed in nine mutations (eight of which were novel) cosegregating with and likely to be causative for CC in 10 of the respective families (Supplementary Fig. S1). In addition, mutations were identified in three encoded genes and the location of the mutations in them are shown in Figure 3, as are the sequence conservation in nine

**Results**

In the 83 unlinked arCC families, fluorescently labeled microsatellite markers flanking each of the 33 genes or loci were genotyped to test for homozygosity. In the first stage, a single affected individual from each family was screened, and all markers tested for the 33 loci were heterozygous in two of the 33 known genes with LOD scores higher than 2 at 0 = 0 (Table 2). Sequence analysis of these genes in the corresponding families revealed nine mutations (eight of which were novel) cosegregating with and likely to be causative for CC in 10 of the respective families (Supplementary Fig. S1). In addition, mutations were identified in three families that had undergone unpublished genome-wide linkage type analyses were performed with closely spaced microsatellite markers in regions that were homozygous in both the affected individuals tested in stages 1 and 2. An LOD less than −2 was obtained in 11 of the families, leaving 21 families requiring sequence analysis of candidate genes, of which mutations were identified in 10. The work flow is summarized in Figure 1 and the results of the linkage analysis are shown in Figure 2 and Table 2.

**Mutation Analysis of Known Cataract Genes in the Homozygous Regions**

Twenty-one families had homozygous regions containing 1 or more of the 33 known genes with LOD scores higher than −2 at 0 = 0 (Table 2). Sequence analysis of these genes in the corresponding families revealed nine mutations (eight of which were novel) cosegregating with and likely to be causative for CC in 10 of the respective families (Supplementary Fig. S1). In addition, mutations were identified in three families that had undergone unpublished genome-wide linkage analysis, including previously described mutations in FYCO1 and GALK1 and a novel mutation in HSF4, marked as U in Table 3. Pathogenicity of mutations was evaluated using a detailed in silico analysis (Table 3). Domain structures of the encoded genes and the location of the mutations in them are shown in Figure 3.
species ranging from humans to zebrafish for missense mutations.

Mutations identified included a novel homozygous substitution in \textit{EPHA2} exon 10 (c.1814C>T, p.[Thr605Ile]) in families 60061 and 60157; a novel homozygous substitution in \textit{FOXE3} (c.307G>A, p.[Glu103Lys]) in family 60039; a known homozygous substitution in \textit{FYCO1} exon 8 (c.2206C>T; p.[Gln736*]), in families 60218 and 60228; a c.1129delG frameshift mutation predicted to result in premature termination, p.(Ala377Profs*2), in family 60152; novel homozygous substitution in exon 1 (c.67T>A; p.[Tyr23Asn]) in family 60090; a c.1067T>C p.(Leu356Pro) missense mutation in family 60133; and a novel homozygous c.440G>T (p.[Gly147Val]) substitution in \textit{CRYBA4} exon 5 in family 60038.

**\textbf{EPH Receptor A2 (EPHA2) Variation of Uncertain Significance}**

This variant (rs753345828) has a reported minor allele frequency of 0.00005 in dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/) and was not seen in 192 ethnically matched control chromosomes (96 individuals). Families 60061 and 60157 share a common haplotype of 11 consecutive SNP markers across \textit{EPHA2}, suggesting that they derive the mutant allele from a common ancestor (Supplementary Table S2). The Thr605 residue is conserved among species from humans to chickens, but not in the zebrafish (Fig. 3), suggesting that it is essential for protein function. The SIFT score for this change was 0, predicting that it is deleterious to the protein. However, the PolyPhen-2 program predicts this mutation to be damaging using the HumDiv dataset but benign using the HumVar dataset for comparisons. It is also predicted to be neutral by program Condel (Table 3). Thus, although likely to be pathogenic, the significance of this sequence change is currently uncertain.

**DISCUSSION**

Here, we describe the results of screening 83 unlinked arCC families for homozygosity at 33 genes or loci known to be involved in arCC. Nine disease-causing mutations were identified in 10 families, and in 11 families no mutations were identified in the linked gene (Supplementary Table S3). We also describe the results of genome-wide linkage analysis in three families for which the mutation had been identified by using a standard linkage approach, but for which the results had not yet been published. Overall, including previously and newly identified mutations, causative genes or loci were identified in 37.1% of the entire set of families studied as part of this project.

The high degree of genetic heterogeneity in arCC makes genetic screening and gene identification expensive and time-consuming. Although this can be approached efficiently by...
FIGURE 2.  The 13 arCC pedigrees collected from Pakistan including 10 families that were mapped through homozygosity mapping and 3 families that were mapped by genome-wide linkage analysis (denoted by asterisks). Filled symbols denote affected individuals. Pedigrees include haplotypes for two microsatellite and gene mutations. The blackened bars correspond to affected haplotypes with alleles that cosegregate with the disease and that are homozygous in affected individuals.

*Mapped by genome-wide linkage analysis
### Table 2. Two-Point LOD Scores of Known Cataract Gene Markers in the 13 arCC Families

<table>
<thead>
<tr>
<th>Family No.</th>
<th>Known Cataract Gene</th>
<th>Marker</th>
<th>0</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>Zmax</th>
<th>( \theta_{\text{max}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>60061</td>
<td>EPHA2</td>
<td>D1S436</td>
<td>5.21</td>
<td>5.11</td>
<td>4.72</td>
<td>4.22</td>
<td>3.17</td>
<td>2.07</td>
<td>0.98</td>
<td>5.22</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>c.1814C&gt;T; p.(Thr605Ile)</td>
<td>D1S2097</td>
<td>2.58</td>
<td>2.52</td>
<td>2.28</td>
<td>1.97</td>
<td>1.55</td>
<td>0.76</td>
<td>0.29</td>
<td>2.58</td>
<td>0</td>
</tr>
<tr>
<td>60157</td>
<td>EPHA2</td>
<td>D1S436</td>
<td>2.62</td>
<td>2.56</td>
<td>2.32</td>
<td>2.01</td>
<td>1.39</td>
<td>0.79</td>
<td>0.29</td>
<td>2.62</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>c.1814C&gt;T; p.(Thr605Ile)</td>
<td>D1S2097</td>
<td>2.62</td>
<td>2.56</td>
<td>2.32</td>
<td>2.01</td>
<td>1.39</td>
<td>0.79</td>
<td>0.29</td>
<td>2.62</td>
<td>0</td>
</tr>
<tr>
<td>60039</td>
<td>FOXE3</td>
<td>D1S347</td>
<td>3.27</td>
<td>3.21</td>
<td>2.96</td>
<td>2.64</td>
<td>1.96</td>
<td>1.26</td>
<td>0.55</td>
<td>3.27</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>c.307G&gt;A; p.(Glu103Lys)</td>
<td>D1S2097</td>
<td>3.56</td>
<td>3.50</td>
<td>3.24</td>
<td>2.90</td>
<td>2.20</td>
<td>1.46</td>
<td>0.71</td>
<td>3.56</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 3. Known Cataract Gene Mutations in Pakistani arCC Families

<table>
<thead>
<tr>
<th>Chr.</th>
<th>Ped</th>
<th>Gene/Locus</th>
<th>Nucleotide</th>
<th>Amino Acid</th>
<th>MAF</th>
<th>PP2</th>
<th>SIFT</th>
<th>Condel</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60061</td>
<td>EPHA2</td>
<td>c.1814C&gt;T*</td>
<td>p.(Thr605Ile)</td>
<td>5 × 10^{-5}</td>
<td>B</td>
<td>D</td>
<td>N</td>
<td>T</td>
</tr>
<tr>
<td>1</td>
<td>60157</td>
<td>EPHA2</td>
<td>c.1814C&gt;T*</td>
<td>p.(Thr605Ile)</td>
<td>5 × 10^{-5}</td>
<td>B</td>
<td>D</td>
<td>N</td>
<td>T</td>
</tr>
<tr>
<td>3</td>
<td>60039</td>
<td>FOXE3</td>
<td>c.307G&gt;A</td>
<td>p.(Glu103Lys)</td>
<td>ND</td>
<td>D</td>
<td>D</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>60133</td>
<td>GALK1</td>
<td>c.766C&gt;T</td>
<td>p.(Arg256Trp)</td>
<td>2.15</td>
<td>2.10</td>
<td>1.89</td>
<td>1.63</td>
<td>1.05</td>
</tr>
<tr>
<td>5</td>
<td>60152</td>
<td>TDRD7</td>
<td>c.1129delG</td>
<td>p.(Ala377Profs*2)</td>
<td>4.59</td>
<td>4.51</td>
<td>4.17</td>
<td>3.75</td>
<td>2.83</td>
</tr>
<tr>
<td>6</td>
<td>60090</td>
<td>AQP0</td>
<td>c.67T&gt;A</td>
<td>p.(Tyr23Asn)</td>
<td>2.97</td>
<td>2.91</td>
<td>2.69</td>
<td>2.41</td>
<td>1.83</td>
</tr>
<tr>
<td>6</td>
<td>60133</td>
<td>GALK1</td>
<td>c.1067T&gt;C</td>
<td>p.(Leu356Pro)</td>
<td>3.59</td>
<td>3.51</td>
<td>3.20</td>
<td>2.79</td>
<td>1.98</td>
</tr>
<tr>
<td>6</td>
<td>60218</td>
<td>GALK1</td>
<td>c.2345delA</td>
<td>p.(Gln782Argfs*32)</td>
<td>1.30</td>
<td>1.27</td>
<td>1.16</td>
<td>1.01</td>
<td>0.72</td>
</tr>
<tr>
<td>6</td>
<td>60237</td>
<td>GALK1</td>
<td>c.2206C&gt;T</td>
<td>p.(Gln736*)</td>
<td>3.73</td>
<td>3.65</td>
<td>3.33</td>
<td>2.92</td>
<td>2.08</td>
</tr>
<tr>
<td>6</td>
<td>60248</td>
<td>GALK1</td>
<td>c.766C&gt;T</td>
<td>p.(Arg256Trp)</td>
<td>2.15</td>
<td>2.10</td>
<td>1.89</td>
<td>1.63</td>
<td>1.10</td>
</tr>
<tr>
<td>6</td>
<td>60038</td>
<td>CRYBA4</td>
<td>c.440G&gt;T</td>
<td>p.(Gly147Val)</td>
<td>2.35</td>
<td>2.30</td>
<td>2.09</td>
<td>1.84</td>
<td>1.57</td>
</tr>
<tr>
<td>6</td>
<td>60038</td>
<td>CRYBA4</td>
<td>c.440G&gt;T</td>
<td>p.(Gly147Val)</td>
<td>2.35</td>
<td>2.30</td>
<td>2.09</td>
<td>1.84</td>
<td>1.57</td>
</tr>
</tbody>
</table>

B, benign; Chr., chromosome; D, damaging; MAF, minor allele frequency in ExAC Browser, no homozygotes unless otherwise noted; N, neutral; NF, not found in ExAC or 1000 Genomes databases; PD, probably damaging; Ped, pedigree; PP2, PolyPhen 2; Ref., reference; T, this study (not found in the Human Gene Mutation Database [HGMD]); U, unpublished.

* rs755345828.
† Predicted to result in nonsense-mediated decay.
‡ Frequency = 8 × 10^{-4} in dbSNP.
§ The first report of a mutation in this gene causing autosomal recessive CC.
FIGURE 3. Domain structure and evolutionary conservation of proteins with missense mutations. Graphic overview of the proteins encoded by genes in which mutations were identified in EPHA2, FOXE3, FYCO1, TDRD7, AQP0, HSF4, GALK1, and CRYBA4. Structural or functional domains are depicted, as well as the position of the mutation. Amino acid sequence conservation around residues affected by missense mutations are shown for the five known cataract genes identified in this study. The red bar indicates the position of the mutation. The sequences of proteins or predicted translation products from nine species from humans to zebrafish have been compared and aligned.
using high-throughput sequencing, this approach is generally more expensive than homozygosity mapping and requires full knowledge of the causative genes and their structure. In contrast to CC in European populations studied, 87% of the families in this project, collected in an unbiased fashion, had ar CC and 13% had ad CC in a Pakistani population (Fig. 1). Because of these considerations and the high levels of consanguinity in our families, we chose to use homozygosity testing disease gene loci for ar CC. This enables relatively rapid and economical screening of many loci and is particularly useful in analysis of consanguineous families in which regions of several centimorgans adjacent to the disease gene are expected to be identical by descent. Screening of 33 genes or loci in the present study identified putative pathogenic alterations in seven different genes in 10 (12%) of 83 families. Five missense, one nonsense, two frame shift, and one splice site mutations were detected, of which eight were novel.

It is unclear why 11 of the 21 families remaining after linkage analysis did not show a mutation in the included candidate gene. The most likely explanation is that these families were too small to yield a statistically significant LOD score (Table 2), so that the homozygosity is fortuitous and the true locus has yet to be mapped. Another possibility is that these families harbor mutations that might be missed by Sanger sequencing, either in introns or currently unidentified exons or control regions. Also, while studying offspring of consanguineous matings should decrease compound heterozygosity, it is possible that this is responsible for the dropout of some families during homozygosity mapping.

**EPHA2** (OMIM 176946) belongs to the A-subclass of receptor tyrosine kinase and interacts with its cognate membrane-anchored ligands to activate cell bidirectional signaling pathway. First described as a cause of ad cataracts in a Caucasian family, homoygous recessive mutations were subsequently implicated in ar CC in a Pakistani family. To date, nine different mutations in **EPHA2** have been reported (see Cat-Map) in 15 families, and **EPHA2** has also been implicated in age-related cataract. Here, we report a novel homozygous missense mutation in two consanguineous Pakistani families. DNA sequencing revealed the transition c.1814C>T, p.(Thr605Ile) in exon 10 located near the protein tyrosine kinase domain of the protein (Fig. 3), suggesting it might alter the tyrosine kinase activity of the EPHA2 protein.

The gene **FOXE3** (OMIM 601094), on chromosome 1p35, is a member of the forkhead box gene family, consisting of a single exon encoding a 319-amino acid DNA-binding transcription factor, consistent with a role in the development of the lens placode. The c.307G>A, p.(Glu103Lys), mutation reported in this study is a novel homozygous missense mutation associated with posterior subcapsular cataract. This mutation occurs in a highly conserved amino acid located within the first transmembrane region of the protein (Fig. 3), suggesting that it might alter the water pore channel function, possibly through affecting water-permeability properties or trafficking. Consistent with the ar cataracts resulting from a loss of function, a knockout mouse model also shows bilateral cataracts.

**HSF4** (OMIM 116800) mutations were originally identified in ad cataract, and later in ar cataract families. A novel missense mutation, c.433G>C, p.(Ala144Pro), in the sixth exon of **HSF4** was found to cosegregate with the disease phenotype in this ar congenital nuclear cataract family (Fig. 1). The Ala145 residue is conserved among different species (Fig. 3), suggesting that it is essential for protein function.

**GALK1** (galactokinase, OMIM 604313) contains eight exons and is located on chromosome 17q25.1. It codes for a 392-amino acid protein containing two ATP binding sites (Fig. 3). Mutations in **GALK1** cause recessive cataracts, and two mutations, c.410delG, p.(Gly137Valfs*27), and c.416T>C, p.(Leu140Pro), were reported in two Pakistani families. Family 60248 in this study showed ac.766C>T, p.(Tyr252Asn) mutation previously reported by Asada et al., and family 61133 showed a novel c.1067T>C, p.(Leu356Pro), mutation at the junction of the second ATP binding site, which might be important for ATP binding.

The β-crystallin gene family includes three basic (CRYBB) and four acidic (CRYBA) crystallin proteins, believed to derive from a common β-crystallin ancestor. All have a highly conserved two-domain, four Greek key motif structure. **CRYBA4** (βA4-crystallin, OMIM 125631) encodes a 196-amino acid protein. The c.440C>T, p.(Gly147Val) mutation (Table 3) is the first **CRYBA4** mutation to be associated with ar cataracts, suggesting a lack of function in CRYBA4 causes the cataracts, and further that **CRYBA4** might have a functional role in the
In total, members in 21 of the 83 families had at least one homozygous region harboring a mutation in a known CC gene. In 10 probands, the causative mutation was identified in the included genes. In the remaining 11 arCC families, sequencing of the known genes in the mapped loci did not reveal a mutation that cosegregated with the disease phenotype (Supplementary Table S3). Clarifying the origin of cataracts in these families remains a challenge. Next-generation sequencing (NGS) has already proven valuable in identifying novel disease genes, both through whole exome or whole genome sequencing and targeted sequencing of linkage intervals or specific genes, both through whole exome or whole genome sequencing and targeted sequencing of linkage intervals or specific genomic regions. However, because of the expense of NGS, homozygosity mapping remains effective in terms of cost and time for localizing mutations in patients with arCC in populations with a high frequency of consanguineous matings, such as the Pakistani population.

Taken together with our previous work, mutations and loci were identified in 43 of 116 Pakistani arCC families. FYCO1 was implicated most commonly, with causative mutations identified in 13.8% (16/116) of arCC families, whereas CRYBB3 accounted for 5.2% (6/116) of arCC in the families studied. In addition, the percentage of arCC cases that can be attributed to the other genes in our study cohort is approximately 3.4% for FYCO1, 1.7% for SIL1, 1.7% for CRYAB, 0.9% for FOXE3, 0.9% for TDRD7, 0.9% for MIP, 0.9% for HSF4, 0.9% for CRYBA4, 4.3% for Locus, and 62.4% Unknown.

Figure 4 summarizes the genetic causes of arCC in the Pakistani population as seen in this study and our previous studies. Overall, this work demonstrates that homozygosity mapping is an efficacious and economical initial step in localizing genetic defects of consanguineous arCC families, allowing insight into the genetic architecture of arCC in the Pakistani population. In addition, these results lay the groundwork for screening larger groups of arCC families by using a similar approach followed by NGS to identify the causative genes in all families. The current advances in conventional and genetic therapies mean that knowledge of the genetic causes of disease in these patients is becoming increasingly valuable for their medical treatment.

Acknowledgments
The authors thank all the family members for their participation in this study. Supported in part by National Eye Institute Grant R01EY022714 (SAR), and by the Higher Education Commission, Islamabad, Pakistan.

Disclosures: J. Chen, None; Q. Wang, None; P.E. Cabrera, None; Z. Zhong, None; W. Sun, None; X. Jiao, None; Y. Chen, None; G. Govindarajan, None; M.A. Naem, None; S.N. Khan, None; M.H. Ali, None; M.Z. Assir, None; F.U. Rahman, None; Z.A. Qazi, None; S. Riazuddin, None; J. Akram, None; S.A. Riazuddin, None; J.F. Hejtmancik, None

References


