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 Hejtmancik[1]

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

Correspondence: J. Fielding Hejtmancik, MOGS/OGVFB/NEI/NIH, Room 1120, 5635 Fisher's Lane, Rockville, MD 20852, USA; f3h@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

Citation: Chen J, Wang Q, Cabrera PE, et al. Molecular genetic analysis of Pakistani families with autosomal recessive congenital cataracts by homozygosity screening. *Invest Ophthalmol Vis Sci*. 2017;58:2207–2217. DOI:10.1167/iovs.17-21469

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.[1] Approximately one-third of CCs are familial; the cataract may be isolated or be associated with other systemic abnormalities.[2] Nonsyndromic CCs may account for approximately 70% of CC cases,[3] and have an estimated frequency of 1 to 6 per 10,000 live births.[4] 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.[5–8] 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).[9] 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

Investigative Ophthalmology & Visual Science

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 CCs 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 unlinked arCC families for homozygosity at 33 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 3 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 consanguineous 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-

TABLE 1. Summary of Gene-Specific Microsatellite Markers

No.	Chr.	Gene Name	Chr. Position, Mb	Locus, Genomic Region	Marker 1	Chr. Position, Mb	Marker 2	Chr. Position, Mb
1	1	*EPHA2*	16.45	1p36.1	D1S436	15.87	D1S2697	14.93
2	1	*FOXE3*	47.88	1p32	D1S2797	46.93	D1S2874	47.91
3	1	*GJA8*	147.37	1q21-q25	D1S442	145.63	D1S2612	147.78
4	2	*CRYGC*	208.99	2q33-q35	D2S2208	208.72		
5	2	*CRYGD*	208.98	2q33-q35	D2S2208	208.72		
6	3	*FYCO1*	45.95	3p21.31	D3S3582	45.38	D3S1767	46.95
7	3	*BFSP2*	133.11	3q21-q22	D3S3713	133.24		
8	3	*CRYGS*	186.25	3q26.3-qter	D3S1262	187.02	D3S3583	184.28
9	5	*SIL1*	138.28	5q31	D5S476	137.94	D5S1372	138.49
10	6	*GCNT2*	10.52	6p24-p23	D6S470	10.02	D6S1034	12.21
11	8	*EYA1*	72.11	8q13.3	D8S279	72.98		
12	9	*GALT*	34.64	9p13	D9S1805	34.18		
13	9	*TDRD7*	100.17	9q22.33	D9S180	100.64	D9S1851	99.57
14	10	*VIM*	17.27	10p13	D10S1476	16.72		
15	10	*SLC16A12*	91.19	10q23.13	D10S1143	91.24	D10S1753	92.41
16	10	*PITX3*	103.98	10q25	D10S1267	104.37	D10S1265	102.65
17	11	*CRYAB*	111.77	11q23.3-q24.2	D11S4078	112.25	D11S1987	112.68
18	12	*MIP*	56.84	12q12-q14.1	D12S1632	56.32	D12S1691	57.51
19	13	*GJA3*	20.71	13q11–13	D13S175	20.84	D13S1275	20.84
20	16	*TMEM114*	8.61	16p13.2	D16S406	8.45		
21	16	*HSF4*	67.19	16q22.1	D16S397	66.73	D16S3107	66.93
22	16	*MAF*	79.62	16q22-q23	D16S3119	79.66	D16S3073	79.95
23	17	*CRYBA1*	27.57	17q11-q12	D17S1873	27.45		
24	17	*GALK1*	73.75	17q24	D17S1301	72.68	D17S1603	74.06
25	19	*FTL*	49.46	19q13.33	D19S879	49.51	D19S866	50.75
26	19	*LIM2*	51.88	19q13.4	D19S206	52.55		
27	20	*BFSP1*	17.47	20p11.23-p12.1	D20S118	17.02	D20S112	17.37
28	20	*CHMP4B*	32.39	20q11.22	D20S890	32.13	D20S195	31.82
29	21	*CRYAA*	44.58	21q22.3	D21S1259	45.32	D21S1411	44.16
30	22	*CRYBB2*	25.61	22q11.2	D22S1167	27.03	D22S1144	27.68
31	22	*CRYBB3*	25.59	22q11.2	D22S1167	27.03	D22S1144	27.68
32	22	*CRYBB1*	26.99	22q11.2	D22S1167	27.03	D22S1144	27.68
33	22	*CRYBA4*	27.01	22q11.2	D22S1167	27.03	D22S1144	27.68

Chr., chromosome.

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, preferably in a conserved region (http://www.ebi.ac.uk/Tools/clustalw2/index.html, in the public domain); and it was judged significant in computational tests for pathogenicity. Missense variants were assessed for possible causality with the online programs Sorting Intolerant From Tolerant (SIFT; http://sift.jcvi.org/www/SIFT_enst_submit.html, in the public domain) and polymorphism phenotyping (Polyphen-2; http://genetics.bwh.harvard.edu/pph2/, in the public domain), as well as Condel (http://bg.upf.edu/fannsdb/, in the public domain), which uses input from multiple programs in its assessment.

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 unlinked arCC families. In the second stage, a second affected individual in each of the 81 remaining families was screened in regions that were homozygous in the first affected individual, and all markers were heterozygous in 49 families. In the remaining 32 families, two-point linkage analyses and haplotype 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 $\theta = 0$ (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, as are the sequence conservation in nine

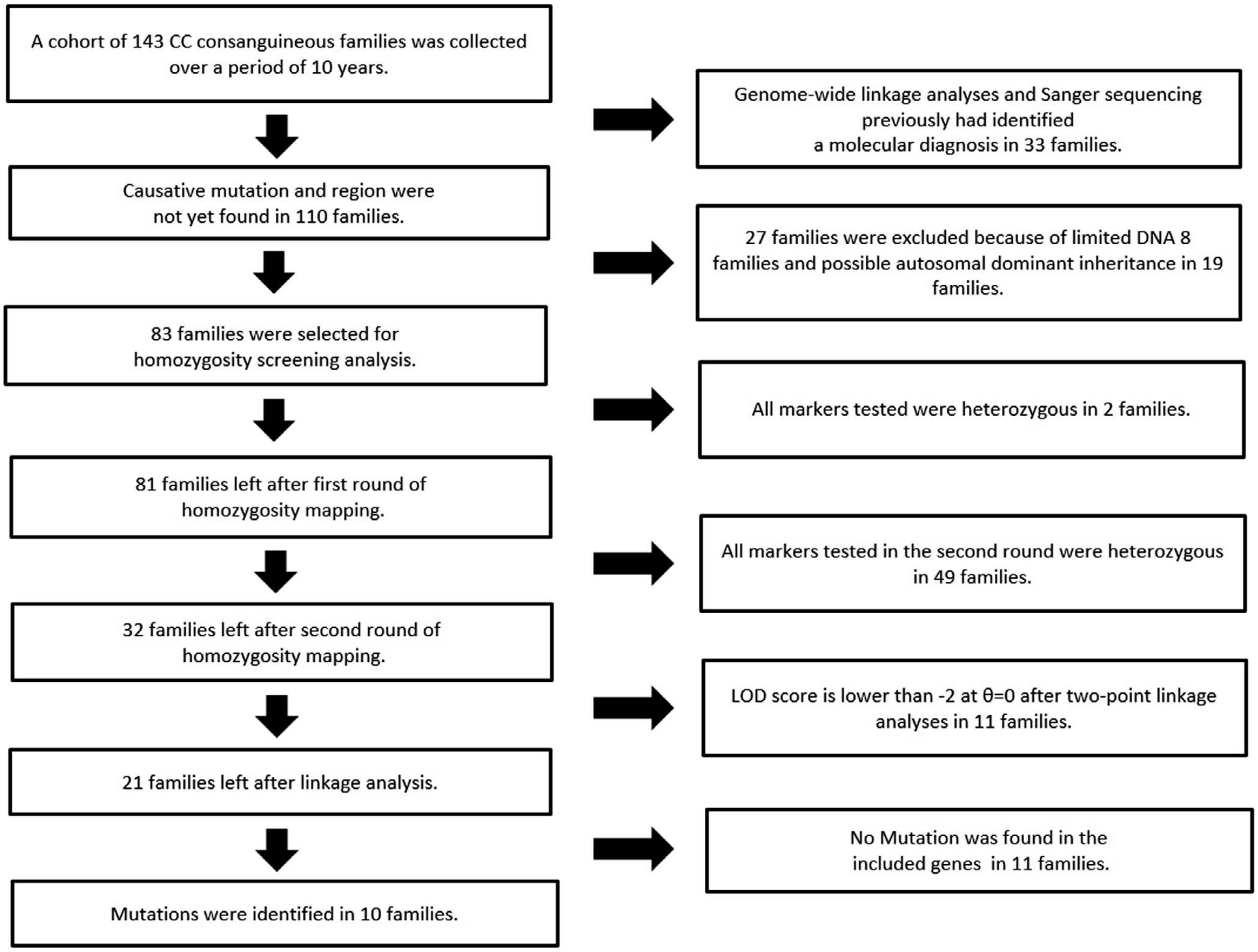

Figure 1. Work flow of the present study.

species ranging from humans to zebrafish for missense mutations.

Mutations identified included a novel homozygous substitution in *EPHA2* exon 10 (c.1814C>T, p.[Thr605Ile]) in families 60061 and 60157; a novel homozygous substitution in *FOXE3* (c.307G>A, p.[Glu103Lys]) in family 60039; a known homozygous substitution in *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 *CRYBA4* exon 5 in family 60038.

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 *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

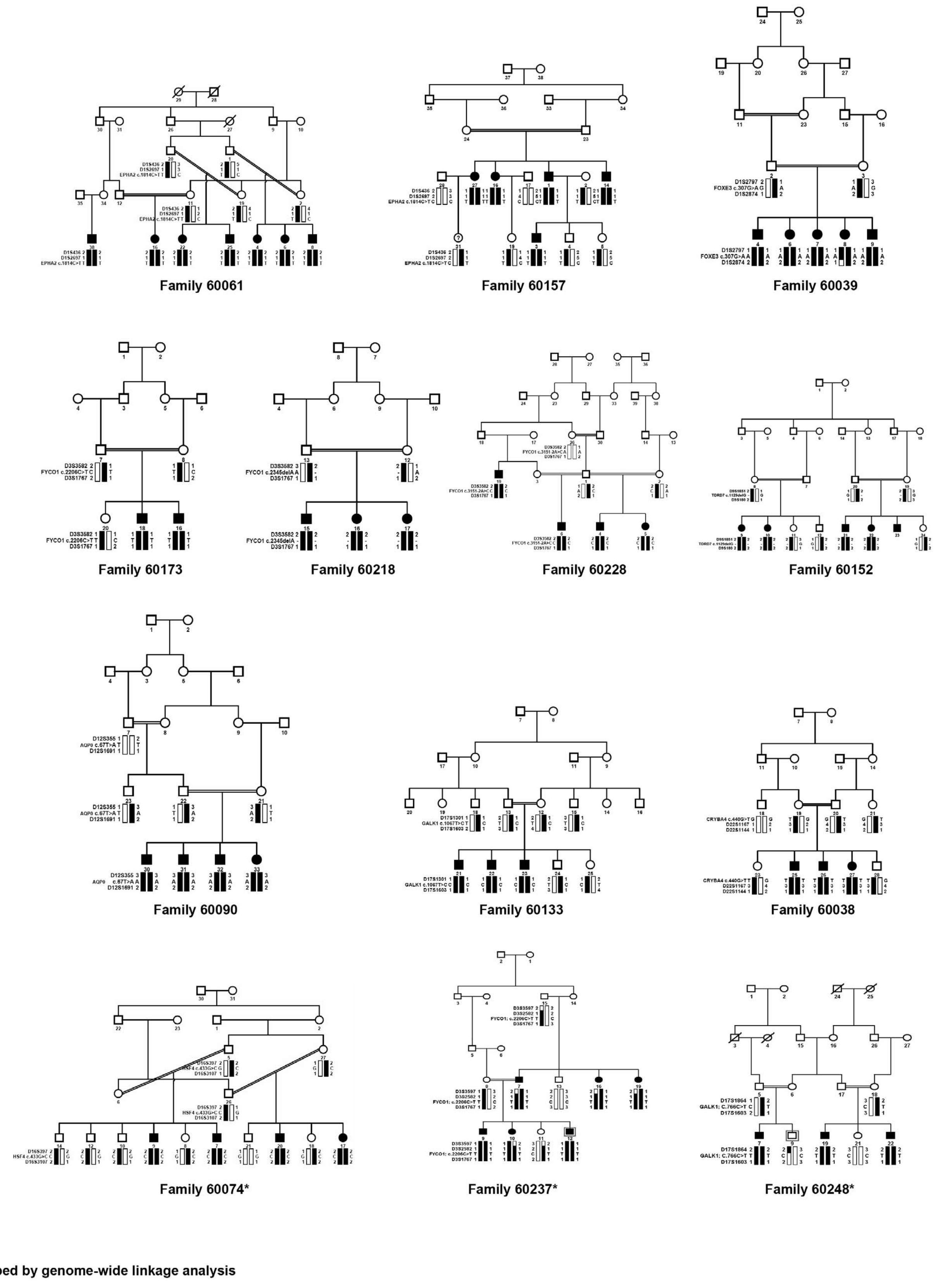

FIGURE 2. The 13 arCC pedigrees collected from Pakistan including 10 families that were mapped through homozygosity mapping and 3 families that were mapped by genome-wide linkage analysis (denoted by *asterisks*). *Filled symbols* denote affected individuals. Pedigrees include haplotypes for two microsatellite and gene mutations. The *blackened bars* correspond to affected haplotypes with alleles that cosegregate with the disease and that are homozygous in affected individuals.

Table 2. Two-Point LOD Scores of Known Cataract Gene Markers in the 13 arCC Families

Family No.	Known Cataract Gene	Marker	0	0.01	0.05	0.1	0.2	0.3	0.4	Z_{max}	θ_{max}
60061	*EPHA2*	D1S436	5.21	5.11	4.72	4.22	3.17	2.07	0.98	5.22	0
		D1S2697	2.58	2.52	2.28	1.97	1.35	0.76	0.29	2.58	0
		c.1814C>T, p.(Thr605Ile)	6.32	6.22	5.81	5.26	4.10	2.79	1.35	6.32	0
60157	*EPHA2*	D1S436	2.62	2.56	2.32	2.01	1.39	0.79	0.29	2.62	0
		D1S2697	2.62	2.56	2.32	2.01	1.39	0.79	0.29	2.62	0
		c.1814C>T, p.(Thr605Ile)	3.61	3.55	3.28	2.93	2.21	1.46	0.71	3.61	0
60039	*FOXE3*	D1S2797	3.27	3.21	2.96	2.64	1.96	1.26	0.55	3.27	0
		c.307G>A, p.(Glu103Lys)	3.56	3.50	3.24	2.90	2.20	1.46	0.71	3.56	0
		D1S2874	$-\infty$	1.22	1.68	1.68	1.36	0.90	0.41	1.68	0.1
60173	*FYCO1*	D3S3582	0.82	0.80	0.72	0.62	0.42	0.24	0.10	0.82	0
		c.2206C>T, p.(Gln736*)	1.93	1.89	1.73	1.53	1.13	0.73	0.35	1.93	0
		D3S1767	1.43	1.39	1.24	1.05	0.68	0.34	0.09	1.43	0
60237	*FYCO1*	D3S3597	2.33	2.28	2.08	1.83	1.29	0.73	0.23	2.33	0
		D3S3582	3.09	3.03	2.79	2.47	1.80	1.09	0.41	3.09	0
		c.2206C>T, p.(Gln736*)	3.44	3.37	3.12	2.80	2.09	1.34	0.58	3.44	0
		D3S1767	3.09	3.03	2.79	2.47	1.80	1.09	0.41	3.09	0
60218	*FYCO1*	D3S3582	2.07	2.02	1.85	1.64	1.19	0.75	0.33	2.07	0
		c.2345delA, p.Gln782ArgfsX31	2.41	2.36	2.19	1.96	1.48	1.00	0.50	2.41	0
		D3S1767	1.30	1.27	1.16	1.01	0.72	0.44	0.19	1.30	0
60228	*FYCO1*	D3S3582	1.46	1.41	1.23	1.01	0.60	0.26	0.04	1.46	0
		c.3151-2A>C	2.00	1.94	1.69	1.38	0.76	0.23	-0.04	2.00	0
		D3S1767	1.46	1.41	1.23	1.01	0.60	0.26	0.04	1.46	0
60152	*TDRD7*	D9S1851	3.73	3.65	3.33	2.92	2.08	1.24	0.51	3.73	0
		c.1129delG, p.(Ala377Pro fs*2)	4.59	4.51	4.17	3.73	2.81	1.83	0.82	4.59	0
		D9S180	3.56	3.48	3.17	2.78	1.98	1.19	0.50	3.56	0
60090	*AQP0*	D12S355	2.71	2.66	2.45	2.19	1.63	1.06	0.49	2.71	0
	(MIP)	c.67T>A, p.(Tyr23Asn)	2.97	2.91	2.69	2.41	1.83	1.22	0.60	2.97	0
		D12S1691	2.67	2.61	2.40	2.14	1.58	1.01	0.45	2.67	0
60074	*HSF4*	D16S397	2.13	2.09	1.91	1.68	1.18	0.66	0.22	2.13	0
		c.433G>C, p.(Ala145Pro)	4.36	4.27	3.94	3.51	2.63	1.72	0.81	4.36	0
		D16S3107	3.59	3.51	3.20	2.79	1.98	1.16	0.41	3.59	0
60248	*GALK1*	D17S1864	2.15	2.10	1.89	1.63	1.10	0.59	0.17	2.15	0
		c.766C>T, p.(Arg256Trp)	2.41	2.36	2.16	1.89	1.33	0.76	0.25	2.41	0
		D17S1603	2.15	2.10	1.91	1.66	1.15	0.63	0.19	2.15	0
60133	*GALK1*	D17S1301	2.07	2.02	1.82	1.57	1.07	0.60	0.20	2.07	0
		c.1067T>C, p.(Leu356Pro)	2.66	2.61	2.39	2.13	1.58	1.04	0.52	2.66	0
		D17S1603	2.24	2.20	2.00	1.75	1.24	0.75	0.31	2.24	0
60038	*CRYBA4*	c.440G>T, p.(Gly147Val)	2.35	2.30	2.09	1.84	1.37	0.93	0.48	2.35	0
		D22S1167	2.07	2.02	1.83	1.59	1.15	0.73	0.33	2.07	0
		D22S1144	1.26	1.23	1.11	0.98	0.72	0.48	0.23	1.26	0

Table 3. Known Cataract Gene Mutations in Pakistani arCC Families

Chr.	Ped	Gene/Locus	Nucleotide	Amino Acid	MAF	PP2	SIFT	Condel	Ref.
1	60061	EPHA2	c.1814C>T*	p.(Thr605Ile)	5×10^{-5}	B	D	N	T
1	60157	EPHA2	c.1814C>T*	p.(Thr605Ile)	5×10^{-5}	B	D	N	T
1	60039	FOXE3	c.307G>A	p.(Glu103Lys)	NF	PD	D	D	T
3	60173	FYCO1	c.2206C>T	p.(Gln736*)†	3×10^{-5}	NA	NA	NA	[14]
3	60218	FYCO1	c.2345delA	p.(Gln782Argfs*32)†	1×10^{-5}	NA	NA	NA	T
3	60237	FYCO1	c.2206C>T	p.(Gln736*)†	3×10^{-5}	NA	NA	NA	U[14]
3	60228	FYCO1	c.3151-2A>C	p.(Ala1051Aspfs*27)†	NF	NA	NA	NA	T
9	60152	TDRD7	c.1129delG	p.(Ala377Profs*2)†	NF‡	NA	NA	NA	T
12	60090	AQPO§	c.67T>A	p.(Tyr23Asn)	NF	PD	D	D	T
16	60074	HSF4	c.433G>C	p.(Ala145Pro)	NF	PD	D	D	U,T
17	60248	GALK1	c.766C>T	p.(Arg256Trp)	2.1×10^{-4}	PD	D	D	U[48]
17	60133	GALK1	c.1067T>C	p.(Leu356Pro)	NF	PD	D	D	T
22	60038	CRYBA4	c.440G>T	p.(Gly147Val)	NF	PD	D	D	T

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.

* rs753345828.

† Predicted to result in nonsense-mediated decay.

‡ Frequency = 8×10^{-6} in dbSNP.

§ The first report of a mutation in this gene causing autosomal recessive CC.

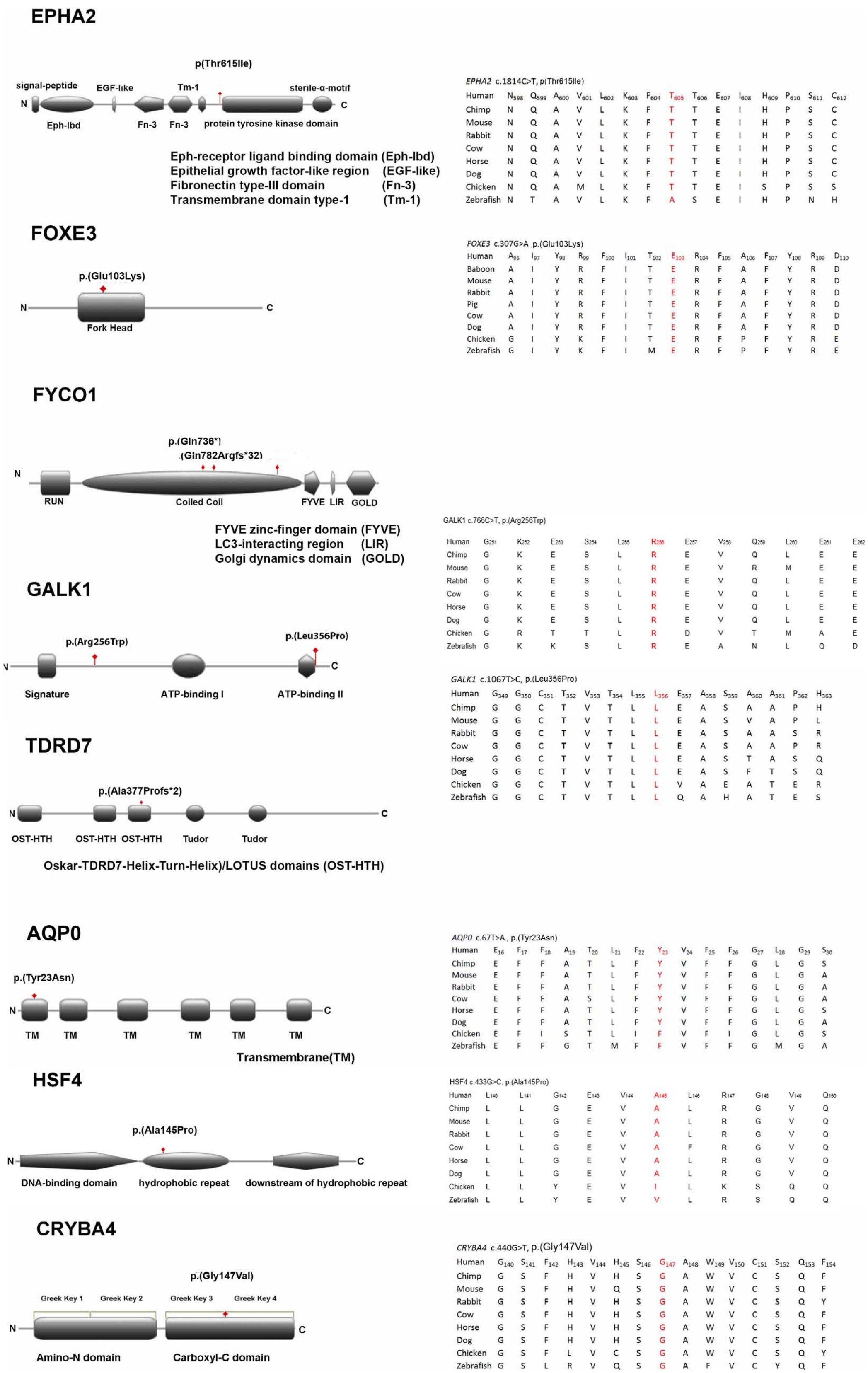

FIGURE 3. Domain structure and evolutionary conservation of proteins with missense mutations. Graphic overview of the proteins encoded by genes in which mutations were identified in *EPHA2*, *FOXE3*, *FYCO1*, *TDRD7*, *AQP0*, *HSF4*, *GALK1*, and *CRYBA4*. Structural or functional domains are depicted, 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 arCC, whereas only 13% had adCC or an ambiguous pedigree (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 arCC. 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.[26] First described as a cause of ad cataracts in a Caucasian family,[27,28] homozygous recessive mutations were subsequently implicated in arCC in a Pakistani family.[10] 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.[27,29–32] Here, we report a novel homozygous *EPHA2* 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 1p33, is a member of the forkhead box gene family,[33] 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.[34] 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 in the fork head domain (Fig. 3) and might change the ability of FOXE3 to bind DNA.

FYCO1 contains 18 exons and encodes for a coiled coil protein comprising 1478 amino acids (~167 kDa).[35] Expressed widely including the eye (UniGene, https://www.ncbi.nlm.nih.gov/unigene, in the public domain), it comprises an α-helical RUN domain following by a long coiled-coil region, an FYVE zinc-finger domain, an LC3-interacting region, and a Golgi dynamics domain.[14] FYCO1 interacts directly with LC3, affecting the maturation of p40phox$^+$ phagosomes,[36] to participate in autophagosomal trafficking.[37] The mutation seen in family 60173, c.2206C>T, p.(Gln736*), Table 3, was identified in three previously published families (60003, 60012, and 60069)[13] and one unpublished family (60237). The novel p.(Gln782Argfsx32) mutation was detected in family 60218 (Fig. 3). The c.3151–2A>C, p.(Ala1051Aspfs*27) mutation identified in family 60228 results in the inactivation of a splice acceptor site. All are predicted to cause nonsense-medicated decay of the *FYCO1* mRNA and a loss of FYCO1 function in the face of the requirement for turning over large amounts of protein and organelles as part of fiber cell differentiation.

TDRD7 (OMIM 611258) belongs to a large family of Tudor domain–containing proteins, and as an RNA granule component interacts with methylated arginine residues and RNA to control the levels of mRNAs posttranscriptionally.[38] *TDRD7* is highly expressed in differentiating fiber cells of the lens. To date, only two mutations in *TDRD7* have been reported: a balanced chromosomal rearrangement disrupting the *TDRD7* gene causing juvenile cataracts in an isolated patient, and as a cause of arCCs in a family.[38] *TDRD7* has also been implicated in age-related cataract.[39] Here, we report a novel homozygous c.1129delG, p.(Ala377Profs*2) *TDRD7* frameshift mutation cosegregating with cataracts in a consanguineous Pakistani family (60152; Table 3), possibly resulting in nonsense-medicated decay of the *TDRD7* mRNA.

The *AQP0* (major intrinsic protein [MIP], OMIM 154050) gene on chromosome 12q13 is a member of the aquaporin family, a ubiquitous family of membrane water transport proteins that confers rapid movements of water across cell membranes. This 263–amino acid intrinsic membrane protein is expressed only in terminally differentiated fiber cells, constituting more than 50% of the total membrane protein in the lens.[40,41] Members of the aquaporin family are predicted to share a unique structure with six transmembrane bilayer-spanning domains (TM1–TM6, Fig. 3).[41] Fourteen different mutations in *MIP* have been identified in 15 families with different types of ad cataract, as listed in Cat-Map.[9] However, this novel homozygous missense mutation is the first to be associated with ar cataracts. The c.67T>A, p.(Tyr23Asn) 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.[42]

HSF4 (OMIM 116800) mutations were originally identified in ad cataract,[43] and later in ar cataract families.[44–46] A novel missense mutation, c.433G>C, p.(Ala145Pro), 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,[47] and two mutations, c.410delG, p.(Gly137Valfs*27), and c.416T>C, p.(Leu139Pro), were reported in two Pakistani families.[13] Family 60248 in this study showed ac.766C>T, p.(Arg256Trp), mutation previously reported by Asada et al.,[48] 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 βγ-crystallin ancestor. All have a highly conserved two-domain, four Greek key motif structure. *CRYBA4* (βA4-crystallin, OMIM 123631) encodes a 196–amino acid protein. The c.440G>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

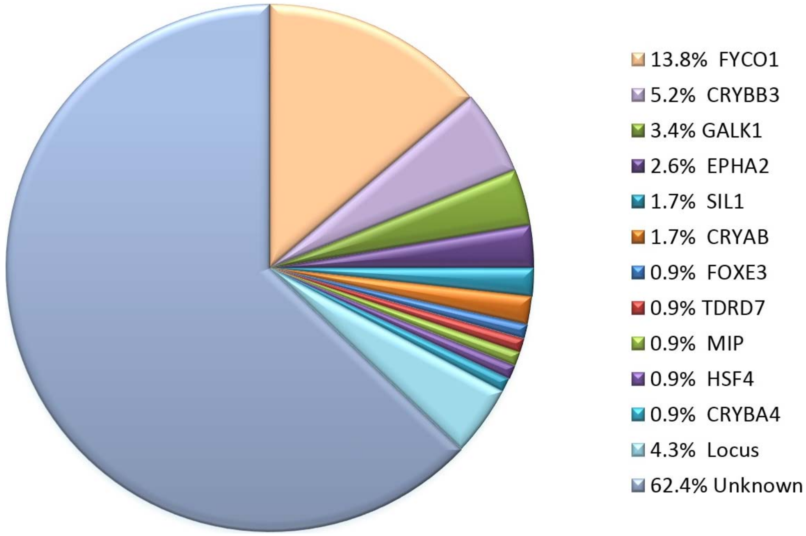

Figure 4. Frequency of cataract gene mutations in the Pakistani population. *Pie chart* showing the frequencies of cataract gene mutations in the Pakistani population as seen in this study and our previous studies.

lens beyond that of a structural crystallin. The mutation is in the fourth Greek key motif (Fig. 3), changes a highly conserved amino acid, and thus probably damages the protein structure. This glycine residue is a critical part of a tryptophan corner motif, occurring at the junction of a Greek key and the following β-strand of the barrel, and thought to aid in folding of the Greek key.[49] The glycine residue is two amino acids to the N-terminal of a tryptophan residue (W-2), which forms a hydrogen bond with the hydrophilic W-3 residue (Supplementary Fig. S2).

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 genomic regions.[50,51] 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 *GALK1* (4/116), approximately 2.6% for *EPHA2* (3/116), approximately 1.7% each for *CRYAB* (2/116) and *SIL1* (2/116), and approximately 0.9% each for *FOXE3* (1/116), *TDRD7* (1/116), *MIP* (1/115), *HSF4* (1/116), and *CRYBA4* (1/116). Figure 4 summarizes the genetic causes of arCC in the Pakistani families studied. In all, the 11 genes are responsible for approximately 32.8% of inherited cataracts in these 116 Pakistani families, with cataracts in the remaining families excluded from these loci.

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.

Accession codes. GenBank: EPHA2 mRNA, NM_004431.3; EPHA2 protein, NP_004422.2. FOXE3 mRNA, NM_012186.2, protein, NP_036318.1. FYCO1 mRNA, NM_024513.3, FYCO1 protein, NP_078789.2. TDRD7 mRNA NM_014290.2; TDRD7 protein NP_055105.2. MIP mRNA, NM_012064.3; MIP protein, NP_036196.1. GALK1 mRNA, NM_000154.1; GALK1 protein, NP_000145.1. CRYBA4 mRNA, NM_001886.2; CRYBA4 protein, NP_001877.1.

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.

Disclosure: **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. Naeem**, 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

1. Robinson GC, Jan JE, Kinnis C. Congenital ocular blindness in children, 1945 to 1984. *Am J Dis Child*. 1987;141:1321–1324.
2. Foster A. Cataract—a global perspective: output, outcome and outlay. *Eye*. 1999;13:449–453.
3. Hejtmancik JF. Congenital cataracts and their molecular genetics. *Semin Cell Dev Biol*. 2008;19:134–149.

4. Lambert SR, Drack AV. Infantile cataracts. *Surv Ophthalmol*. 1996;40:427–458.
5. Francois J. Genetics of cataract. *Ophthalmologica*. 1982;184: 61–71.
6. Haargaard B, Wohlfahrt J, Fledelius HC, Rosenberg T, Melbye M. A nationwide Danish study of 1027 cases of congenital/infantile cataracts: etiological and clinical classifications. *Ophthalmology*. 2004;111:2292–2298.
7. Merin S. Inherited Cataracts. In: Merin S, ed. *Inherited Eye Diseases*. New York: Marcel Dekker, Inc.; 1991:86–120.
8. Wirth MG, Russell-Eggitt IM, Craig JE, Elder JE, Mackey DA. Aetiology of congenital and paediatric cataract in an Australian population. *Br J Ophthalmol*. 2002;86:782–786.
9. Shiels A, Bennett TM, Hejtmancik JF. Cat-Map: putting cataract on the map. *Mol Vis*. 2010;16:2007–2015.
10. Kaul H, Riazuddin SA, Shahid M, et al. Autosomal recessive congenital cataract linked to EPHA2 in a consanguineous Pakistani family. *Mol Vis*. 2010;16:511–517.
11. Riazuddin SA, Miri-Kordestani L, Kaul H, et al. Novel SIL1 mutations in consanguineous Pakistani families mapping to chromosomes 5q31. *Mol Vis*. 2009;15:1050–1056.
12. Riazuddin SA, Yasmeen A, Yao W, et al. Mutations in βB3-crystallin associated with autosomal recessive cataract in two Pakistani families. *Invest Ophthalmol Vis Sci*. 2005;46:2100–2106.
13. Yasmeen A, Riazuddin SA, Kaul H, et al. Autosomal recessive congenital cataract in consanguineous Pakistani families is associated with mutations in GALK1. *Mol Vis*. 2010;16:682–688.
14. Chen J, Ma Z, Jiao X, et al. Mutations in FYCO1 cause autosomal-recessive congenital cataracts. *Am J Hum Genet*. 2011;88:827–838.
15. Jiao X, Kabir F, Irum B, et al. A common ancestral mutation in CRYBB3 identified in multiple consanguineous families with congenital cataracts. *PLoS One*. 2016;11:e0157005.
16. Butt T, Yao W, Kaul H, et al. Localization of autosomal recessive congenital cataracts in consanguineous Pakistani families to a new locus on chromosome 1p. *Mol Vis*. 2007;13: 1635–1640.
17. Sabir N, Riazuddin SA, Butt T, et al. Mapping of a new locus associated with autosomal recessive congenital cataract to chromosome 3q. *Mol Vis*. 2010;16:2634–2638.
18. Kaul H, Riazuddin SA, Yasmeen A, et al. A new locus for autosomal recessive congenital cataract identified in a Pakistani family. *Mol Vis*. 2010;16:240–245.
19. Sabir N, Riazuddin SA, Kaul H, et al. Mapping of a novel locus associated with autosomal recessive congenital cataract to chromosome 8p. *Mol Vis*. 2010;16:2911–2915.
20. Riazuddin SA, Yasmeen A, Zhang Q, et al. A new locus for autosomal recessive nuclear cataract mapped to chromosome 19q13 in a Pakistani family. *Invest Ophthalmol Vis Sci*. 2005; 46:623–626.
21. Jiaox X, Khan SY, Irum B, et al. Missense mutations in CRYAB are liable for recessive congenital cataracts. *PLoS One* 2015; 10:e0137973.
22. Grimberg J, Nawoschik S, Belluscio L, McKee R, Turck A, Eisenberg A. A simple and efficient non-organic procedure for the isolation of genomic DNA from blood. *Nucleic Acids Res*. 1989;17:8390.
23. Oetting WS, Lee HK, Flanders DJ, Wiesner GL, Sellers TA, King RA. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics*. 1995;30:450–458.
24. Cottingham RW, Idury RM, Schaffer AA. Faster sequential genetic linkage computations. *Am J Hum Genet*. 1993;53: 252–263.
25. Lathrop GM, Lalouel JM. Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet*. 1984;36: 460–465.
26. Pasquale EB. Eph-ephrin bidirectional signaling in physiology and disease. *Cell*. 2008;133:38–52.
27. Shiels A, Bennett TM, Knopf HL, et al. The EPHA2 gene is associated with cataracts linked to chromosome 1p. *Mol Vis*. 2008;14:2042–2055.
28. Zhang T, Hua R, Xiao W, et al. Mutations of the EPHA2 receptor tyrosine kinase gene cause autosomal dominant congenital cataract. *Hum Mutat*. 2009;30:E603–E611.
29. Jun G, Guo H, Klein BE, et al. EPHA2 is associated with age-related cortical cataract in mice and humans. *PLoS Genet*. 2009;5:e1000584.
30. Yang J, Luo J, Zhou P, Fan Q, Luo Y, Lu Y. Association of the ephreceptor tyrosinekinase-type A2 (EPHA2) gene polymorphism rs3754334 with age-related cataract risk: a meta-analysis. *PLoS One*. 2013;8:e71003.
31. Tan W, Hou S, Jiang Z, Hu Z, Yang P, Ye J. Association of EPHA2 polymorphisms and age-related cortical cataract in a Han Chinese population. *Mol Vis*. 2011;17:1553–1558.
32. Sundaresan P, Ravindran RD, Vashist P, et al. EPHA2 polymorphisms and age-related cataract in India. *PLoS One*. 2012;7:e33001.
33. Katoh M. Human FOX gene family (Review). *Int J Oncol*. 2004;25:1495–1500.
34. Blixt A, Mahlapuu M, Aitola M, Pelto-Huikko M, Enerback S, Carlsson P. A forkhead gene, FoxE3, is essential for lens epithelial proliferation and closure of the lens vesicle. *Genes Dev*. 2000;14:245–254.
35. Kiss H, Darai E, Kiss C, et al. Comparative human/murine sequence analysis of the common eliminated region 1 from human 3p21.3. *Mamm Genome*. 2002;13:646–655.
36. Ma J, Becker C, Reyes C, Underhill DM. Cutting edge: FYCO1 recruitment to dectin-1 phagosomes is accelerated by light chain 3 protein and regulates phagosome maturation and reactive oxygen production. *J Immunol*. 2014;192:1356–1360.
37. Pankiv S, Johansen T. FYCO1: linking autophagosomes to microtubule plus end-directing molecular motors. *Autophagy*. 2010;6:550–552.
38. Lachke SA, Alkuraya FS, Kneeland SC, et al. Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. *Science*. 2011;331:1571–1576.
39. Zheng C, Wu M, He CY, et al. RNA granule component TDRD7 gene polymorphisms in a Han Chinese population with age-related cataract. *J Int Med Res*. 2014;42:153–163.
40. Berry V, Francis P, Kaushal S, Moore A, Bhattacharya S. Missense mutations in MIP underlie autosomal dominant 'polymorphic' and lamellar cataracts linked to 12q. *Nat Genet*. 2000;25:15–17.
41. Francis P, Chung JJ, Yasui M, et al. Functional impairment of lens aquaporin in two families with dominantly inherited cataracts. *Hum Mol Genet*. 2000;9:2329–2334.
42. Shiels A, Bassnett S. Mutations in the founder of the MIP gene family underlie cataract development in the mouse. *Nat Genet*. 1996;12:212–215.
43. Bu L, Jin YP, Shi YF, et al. Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. *Nat Genet*. 2002;31:276–278.
44. Smaoui N, Beltaief O, Benhamed S, et al. A homozygous splice mutation in the HSF4 gene is associated with an autosomal recessive congenital cataract. *Invest Ophthalmol Vis Sci*. 2004;45:2716–2721.
45. Forshew T, Johnson CA, Khaliq S, et al. Locus heterogeneity in autosomal recessive congenital cataracts: linkage to 9q and germline HSF4 mutations. *Hum Genet*. 2005;117:452–459.

Investigative Ophthalmology & Visual Science

46. Sajjad N, Goebel I, Kakar N, Cheema AM, Kubisch C, Ahmad J. A novel HSF4 gene mutation (p.R405X) causing autosomal recessive congenital cataracts in a large consanguineous family from Pakistan. *BMC Med Genet*. 2008;9:99.
47. Stambolian D, Ai Y, Sidjanin D, et al. Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts. *Nat Genet*. 1995;10:307–312.
48. Asada M, Okano Y, Imamura T, Suyama I, Hase Y, Isshiki G. Molecular characterization of galactokinase deficiency in Japanese patients. *J Hum Genet*. 1999;44:377–382.
49. Hemmingsen JM, Gernert KM, Richardson JS, Richardson DC. The tyrosine corner: a feature of most Greek key beta-barrel proteins. *Protein Sci*. 1994;3:1927–1937.
50. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010; 42:30–35.
51. Riazuddin SA, Parker DS, McGlumphy EJ, et al. Mutations in LOXHD1, a recessive-deafness locus, cause dominant late-onset Fuchs corneal dystrophy. *Am J Hum Genet*. 2012;90:533–539.

Investigative Ophthalmology & Visual Science