Original Article

A novel splice donor site mutation in EPHA2 caused congenital cataract in a Chinese family

Juan Bu[1], Sijie He[2,3]*, Lejin Wang[1], Jiankang Li[3], Jing Liu[1]*, Xiuqing Zhang[3,4,5]**

Background: Congenital cataract is a rare disorder characterized by crystallin denaturation, which becomes a major cause of childhood blindness. Although more than fifty pathogenic genes for congenital cataract have been reported, the genetic causes of many cataract patients remain unknown. In this study, the aim is to identify the genetic cause of a five-generation Chinese autosomal dominant congenital cataract family. **Methods:** Whole exome sequencing (WES) was performed on three affected and one unaffected member of the family, known causative genes were scanned first. Sanger sequencing was used to validate co-segregation of the candidate variant in the family. The impact on the transcript and amino acid sequences of the variant was further analyzed. **Results:** We identified a novel splice donor site mutation c. 2825+1G >A in EPHA2 that was absent in public and in-house databases and showed co-segregation in the family. This variant resulted in an altered splice that led to protein truncation. **Conclusions:** The mutation we identified was responsible for congenital cataract in our studied family. Our findings broaden the spectrum of causative mutations in EPHA2 gene for congenital cataract and suggest that WES is an efficient strategy to scan variants in known causative genes for genetically heterogeneous diseases.

Key words: Autosomal dominant congenital cataract, EPHA2, splice donor site mutation, whole exome sequencing

Access this article online

Website: www.ijo.in

DOI: 10.4103/0301-4738.185597

Quick Response Code:

Cataract is characterized by metabolic disturbance of crystalline lens that leads to crystallin denaturation, which is the primary cause of blindness worldwide. The estimated prevalence of nonsyndromic congenital cataracts is 1–6/10,000 live births[1] which becomes a major cause of childhood blindness.[2] About one-third of congenital cataracts are inherited, the majority pattern is autosomal dominant (AD) despite a few cases of autosomal recessive and X-linked inheritance is reported.[3,4] Several factors including the transparency and refractive index of lens, nutrition and intercellular communication in lens, cell motility and maintenance of cell volume and shape are related to the occurrence of cataract.

To date, more than fifty genes had been reported to be associated with congenital cataracts, among which more than twenty genes may cause AD pattern.[5] Several genes were reported to cause both dominant and recessive patterns, such as EPHA2 (OMIM 176946), GJA8 (OMIM 600897), SIL1 (OMIM 608005), CRYAB (OMIM 123590), HSF4 (OMIM 602438), CRYAA (OMIM 123580), CRYBB1 (OMIM 600929). The traditional strategy to scan known genes was time-consuming and expensive, high-throughput sequencing like whole exome sequencing (WES) was more efficient to do known genes scanning as well as novel genes discovery compared to traditional strategy.

In this study, we applied WES on a five-generation Chinese family with AD congenital cataract. We identified a novel splice donor site mutation (c. 2825+1G>A, p.D942E*) in the EPHA2 gene that resulted in protein truncation caused this disorder.

Methods

Participants and clinical diagnosis

A five-generation Han family including 13 affected and 25 unaffected individuals with congenital cataract from Anhui province in China participated in this study [Fig. 1]. All patients were performed dilated pupil examination, and their eyesight ranged from 0.02 to 0.3. Each patient showed the phenotype in a young age. The slit lamp photographs of two patients (II-1, IV-2) indicated a nuclear cataract in this family [Fig. 2]. Each patients' clinical information including diagnosis time and operation time was shown [Table 1]. Informed consents were signed by each participant or the guardians.

DNA extraction and whole exome sequencing

Genomic DNA was extracted from peripheral blood of each participant with QIAamp DNA Blood Mini Kit (Qiagen, Germany) according to the manufacturer's instructions.

*These authors contributed equally to this work.

[1]Department of Ophthalmology, Peking University Third Hospital, Key Laboratory of Vision Loss and Restoration, Ministry of Education, Beijing 100191, [2]BGI Education Center, University of Chinese Academy of Sciences, [3]BGI-Shenzhen, [4]The Guangdong Enterprise Key Laboratory of Human Disease Genomics, [5]Shenzhen Key Laboratory of Genomics, Shenzhen 518083, China

Correspondence to: Prof. Xiuqing Zhang, BGI-Shenzhen, Shenzhen 518083, China. The Guangdong Enterprise Key Laboratory of Human Disease Genomics, Shenzhen 518083, China. Shenzhen Key Laboratory of Genomics, Shenzhen 518083, China. E-mail: zhangxq@genomics.cn

Manuscript received: 23.11.15; Revision accepted: 28.04.16

Cite this article as: Bu J, He S, Wang L, Li J, Liu J, Zhang X. A novel splice donor site mutation in EPHA2 caused congenital cataract in a Chinese family. Indian J Ophthalmol 2016;64:364-8.

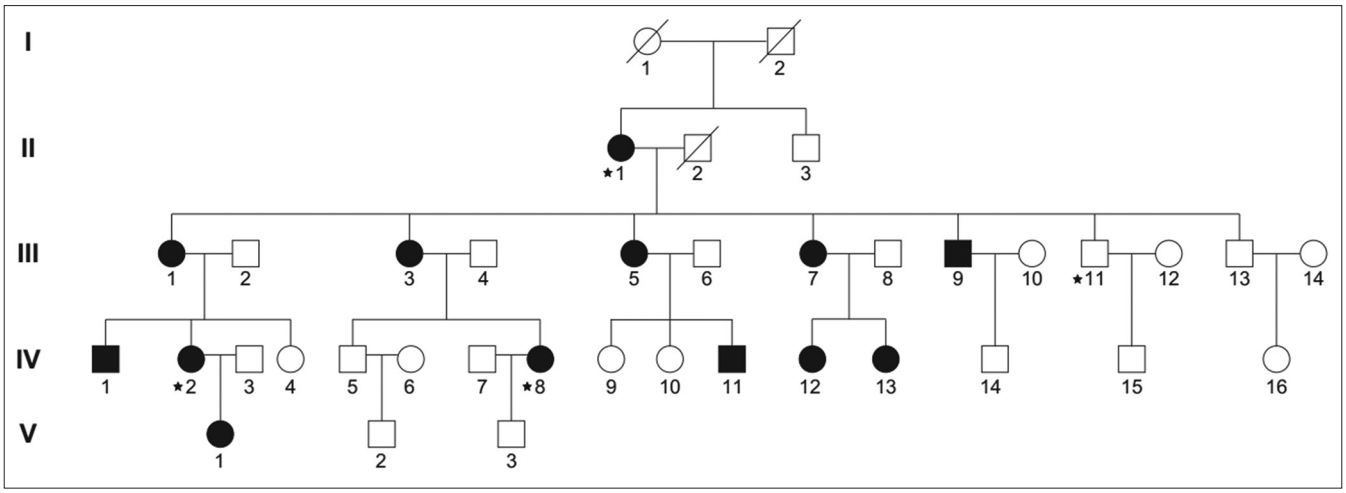

Figure 1: Pedigree of the Chinese cataract family. Affected individuals are indicated by filled symbols, the individuals marked star are applied whole exome sequencing

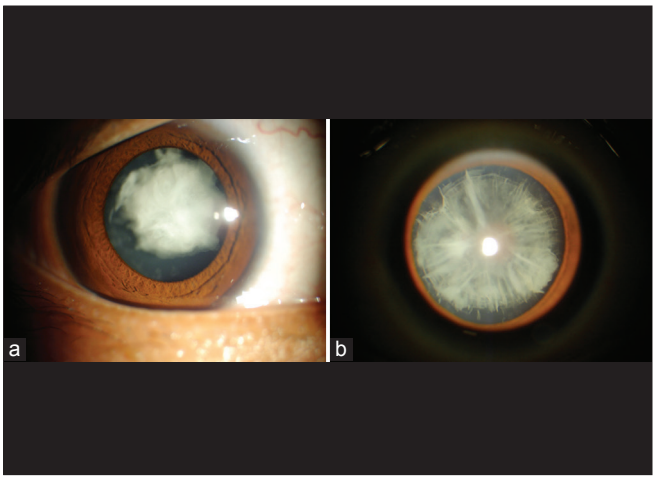

Figure 2: Slit lamp photographs of two patients. The slit lamp photographs indicated a nuclear cataract in this family. (a) Slit lamp photograph for patient II-1. (b) Slit lamp photograph for patient IV-2

Table 1: Clinical evaluation of affected family members

Case	Age	Sex	Phenotype	Diagnosis time	Operation time
II-1	79	Female	Nuclear cataract	2005.3	2005.6
III-1	58	Female	Nuclear cataract	2012.9	2012.9
III-3	57	Female	Nuclear cataract	2012.7	2012.9
III-5	55	Female	Nuclear cataract	2012.7	2013.5
III-7	51	Female	Nuclear cataract	2012.7	-
III-9	48	Male	Nuclear cataract	2012.7	2012.9
IV-1	33	Male	Nuclear cataract	2014.1	-
IV-2	32	Female	Nuclear cataract	2012.7	2012.9
IV-8	28	Female	Nuclear cataract	2012.7	-
IV-11	21	Male	Nuclear cataract	2012.7	-
IV-12	26	Female	Nuclear cataract	2012.7	-
IV-13	23	Female	Nuclear cataract	2012.7	2012.9
V-1	5	Female	Nuclear cataract	2014.1	-

-: The individual didn't obtain an operation

WES was applied on 1 unaffected individual (III-11) and three affected individuals (II-1, IV-2 and IV-8). Exome capture was performed with NimbleGen SeqCap EZ Human Exome Library v2.0 (NimbleGen, Madison, WI, USA) covering 44MB of coding region and then sequenced on HiSeq2000 platform (Illumina, San Diego, USA). Briefly, genomic DNA samples were randomly fragmented into 250–300 bp and then adapters were ligated to both ends of the fragments. After exome region enrichment, the libraries were sequenced using the Hiseq2000 platform, and paired-end reads of 90-bp were generated.

Reads mapping and variants detection

Raw reads with low quality or containing adapters were filtered before mapping. For single nucleotide polymorphism (SNP) calling, filtered reads were aligned to the human genome reference (UCSC hg19) with Short Oligonucleotide Analysis Package (SOAP, version 2.21),[6] and then SOAPsnp software (version 1.05)[7] was used to detect SNPs. We eliminated low-quality SNPs if the genotype quality <20 or <4 reads covering this site. For indel calling, Burrows–Wheeler Aligner[8] was used to do the alignment and Genome Analysis Tool Kit[9] was used to call small indels. Indels were called heterozygous if the indel-supporting reads/total reads ranged from 0.3 to 0.7, whereas indels were called homozygous if the indel-supporting reads/total reads >0.7.

Variants analysis

The called variants were annotated and categorized by ANNOVAR.[10] Variants located in intron, intergenic region, and untranslated region as well as synonymous substitutions were excluded. Then, we filtered the variants observed in public databases including dbSNP137, 1000 genomes project, exome sequencing project and our in-house databases with a frequency > 0.005. Considering dominant inheritance model, we chose the heterozygous variants shared by three affected individuals and absent in the unaffected individual as candidates. We scanned 37 previous reported cataract-related genes[5] and 15 additional cataract-related genes obtained from OMIM (http://omim.org/) to find whether there were variants located in these genes. Sorting Intolerant From Tolerant (SIFT) (http://sift.jcvi.org/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph 2/index.

shtml) and MutationTaster (http://www.mutationtaster.org/) were used to predict whether the variants are harmful[11] and Genomic Evolutionary Rate Profiling (GERP) was used to predict the conservation of the variants. We further analyzed the impact on the transcript and amino acid sequences of the variant in the cataract-related gene.

Sanger sequencing validation

Sanger sequencing was used to validate the variant identified by WES in the four individuals. Polymerase chain reaction (PCR) primers were designed by Primer Z (http://genepipe.ncgm.sinica.edu.tw/primerz/begindesign.do), the sequences were as follows: 5′-CGGCACATAGCCCTCAGTAA-3′ and 5′-GAGGGGCAGCAGTAGTTACA-3′. After PCR amplification, the purified production was sequenced on ABI 3730XL DNA analyzer. Other family members available were Sanger sequenced to confirm co-segregation by the same method.

Results

Whole exome sequencing and bioinformatics identified a novel splice donor site mutation in EPHA2

WES was applied on four individuals in the cataract family. An average mean depth of target region (44 M) was 68.22 with coverage of 99.44%, and the coverage of target region that sequenced at least 10 times (depth ≥ 10×) was 96.84% [Table 2].

We identified 97156 SNPs and 9586 indels of each individual on average, among which 13881 SNPs and 1404 indels were protein-disrupting variants (PDV). Briefly, PDV were included missense, nonsense, splice and read through variants for SNPs and frameshift, cds-indel and splice variants for indels. After filtering against public databases and in-house databases and considering a dominant model, 23 SNPs and 1 indel were left [Table 2]. For the indel, one case was homozygosis while other two cases were heterozygosis, so we actually excluded the indel in the latter analysis (the indel was not in cataract-related genes). Then, we scanned 52 previously reported cataract-related genes [Supplementary Table 1] and found a splice donor site mutation in EPHA2, c. 2825+1G>A (chr1:16455928 C > T, hg19). We searched related papers and found this mutation had not been reported before. Other 23 variants were not in the 52 genes.

We used harmful prediction tools mentioned in method to predict the impact of these variants [Supplementary Table 2]. If the GERP score is >3, then we consider the site to be conservative. The mutation in EPHA2 had a GERP score 5.77 and we further checked the sequence of this site among different species from UCSC and we found nearly all species in this site have the same base which indicates that this site is highly conserved. Although SIFT and PolyPhen2_HVAR did not give a prediction for the EPHA2 splice site mutation, MutationTaster gave a disease-causing prediction.

This splice site mutation in EPHA2 was predicted to cause an alternative splicing and add 4 bases of intron 16 into mRNA, which brought in a *de novo* terminal codon and led to the loss of 34 amino acids encoded by exon 17 [Fig. 3].

Considering its functional impact and conservation as well as it located in cataract-related gene and fit a dominant inherited model, we regarded the rare EPHA2 splice site

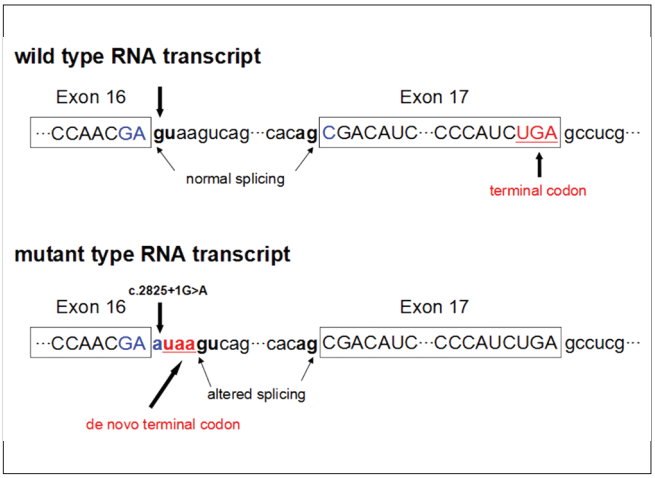

Figure 3: Schematic diagram of comparison between normal splicing and altered splicing of EPHA2. The upper shows the normal splicing of the wild type RNA transcript of EPHA2. The lower shows that the G>A variant at the splice donor site of intron 16 led to an altered splicing. The Asp encoded by GAC (marked blue in upper panel) was replaced by Glu encoded by GAA (marked blue in lower panel), and a *de novo* terminal codon (marked red in lower panel) resulted in termination of translation. The 34 amino acids encoded by exon 17 were lost

Table 2: Summary of exome sequencing data and variants filtration of the four individuals

Detail information for WES data and variants	III-11	II-1	IV-2	IV-8	Average
Mean depth of target region	69.85	63.14	67.78	72.10	68.22
Coverage of target region (%)	99.41	99.46	99.42	99.46	99.44
Fraction of target region covered ≥10× (%)	96.83	96.70	96.70	97.14	96.84
Total number of variants (SNP + indel)	96,791+9646	94,601+9181	98,515+9675	98,718+9840	97,156+9586
PDV	13829+850	13960+868	13823+832	13911+840	13881+848
PDV after filtered against dbSNP (db137)	1780+123	1779+123	1770+122	1774+124	1776+123
PDV after filtered against dbSNP + 1000 genomes	1071+77	1086+75	1092+76	1084+69	1083+74
PDV after filtered against dbSNP + 1000 genomes+ESP	1058+64	1078+64	1085+66	1076+58	1074+63
PDV after filtered against dbSNP + 1000 genomes + ESP + inhouse databases	670+64	698+64	700+66	679+58	687+63
Candidates fitting a dominant model	23+1				
Candidates in known genes	1+0				

ESP: Exome sequencing project, PDV: Protein-disrupting variants, WES: Whole exome sequencing, SNP: Single nucleotide polymorphism

mutation c. 2825+1G>A as a prior causative candidate then we did Sanger sequencing to validate this mutation.

Sanger validation

Sanger sequencing was used to exclude false positive (FP) and to do genetic validation for the EPHA2 splice site mutation c. 2825+1G>A. We used primers mentioned in method to do PCR followed by 3730 sequencing for the four samples to exclude the FP of high-throughput sequencing. Sanger results showed that control (III-11) was wild type while cases (II-1, IV-2, and IV-8) were C/T heterozygosis in this site [Fig. 4].

We then detected 18 other family members available (8 cases and 10 controls) to confirm co-segregation. The results showed that this mutation was detected in all affected individuals and absent in all healthy individuals [Supplementary Fig. 1]. Despite we had excluded variants with a frequency more than 0.005 in various databases, we further checked this mutation in more than 1000 additional in-house exome samples captured by the same library and did not find this mutation.

Discussion

Cataract is a severe disease with high clinical and inherited heterogeneity. There are several subtypes of cataract in clinical and more than fifty genes have been reported to cause cataract or syndromes with cataract. Scanning the known genes one after another by traditional Sanger sequencing is time-consuming and quite expensive. An efficient strategy is WES followed by known genes scanning which allows us to identify mutations at a global level. If no variants in known genes are identified, we may discover new genes with the whole exome data. In this study, we found a variant in a cataract-related gene EPHA2.

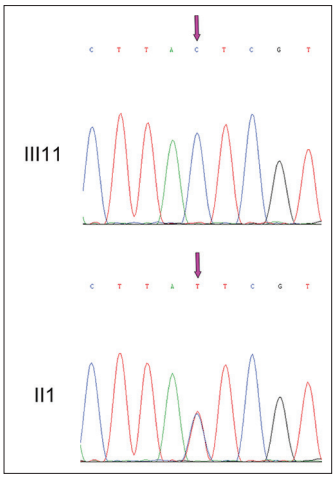

Figure 4: Results of reverse Sanger sequencing for the unaffected and one representative affected individual applied whole exome sequencing. Reverse sequencing showed that the healthy control (III-11) is wild type while the representative affected individual (II-1) is heterozygous C/T at this position. The other two affected individuals applied whole exome sequencing (IV-2 and IV-8) are heterozygous C/T and the Sanger results are shown in supplementary figure

EPHA2 is one of the causative genes and was estimated to explain 4.7% of inherited cataract cases in South-Eastern Australia.[12] The EPHA2 gene belongs to the ephrin receptor subfamily of the protein-tyrosine kinase family. EPH and EPH-related receptors have been implicated in mediating developmental events, particularly in the nervous system. This gene encodes a protein that binds ephrin-A ligands. Mutations in this gene are the cause of certain genetically-related cataract disorders (http://www.ncbi.nlm.nih.gov/gene/1969). The intracellular region of the ephrin receptor comprised a regulatory juxtamembrane domain, a tyrosine kinase domain, a sterile alpha motif (SAM) domain, and a PDZ-binding motif domain.[13,14] Several mutations in EPHA2 were reported to cause cataract,[12,15-18] most of which were located in SAM domain [Fig. 5]. This may suggest a crucial role of SAM domain in EPHA2 gene.

In this study, a novel splice donor site mutation c. 2825+1G>A located in SAM domain was identified [Fig. 5]. This mutation may change the splicing and lead to protein truncation that influence the structure of the protein and thus cause cataract. We predicted the variant may cause a p.D942E* change on the protein level, but further experiments were needed to confirm our presumption. Although our results showed evidence that the novel mutation is the cause of this cataract family, how this mutation would change protein structure, metabolic process and finally induce cataract was not clearly known.

Except for the mutation in EPHA2 gene, we had two rare variants in our candidate list that predicted to be damaging by three tools and with a high GERP score. The two variants were in two genes SLC12A8 and ABCD1. These two genes were related to psoriasis and spinocerebellar degeneration separately. We applied Sanger sequencing to confirm the two variants. Sanger results showed that variant in ABCD1 gene was FP while variant in SLC12A8 gene was not co-segregated with the disease in the family. So these two variants were not pathogenic variants. This situation implied that rare and probably harmful variants may not be pathogenic variants, so how to identify pathogenic variants from detrimental variants

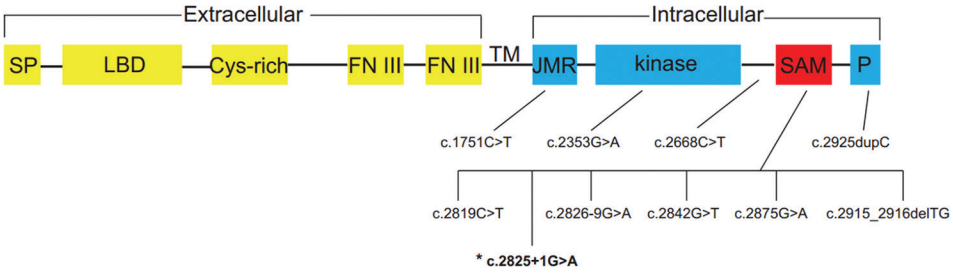

Figure 5: Schematic drawing of the EPHA2 protein and variants causing cataract in EPHA2. SP: Signal peptide, LBD: Ligand binding domain, Cys-rich: Cysteine rich region, FN III: Fironectin III type repeats, TM: Transmembrane domain, JMR: Juxtamembrane region, SAM: Sterile alpha motif, P: PDZ-binding motif. The variant marked star is the novel variant discovered in this study

was a major challenge, especially when we study a small family or sporadic samples.[19]

Conclusions

We identified a novel heterozygous splice donor site mutation c. 2825+1G>A in EPHA2 gene caused cataract in a Chinese family by WES. Our finding broadens the causative mutation spectrum of EPHA2 gene and indicates the efficiency of scanning variants in known genes for inherited heterogeneous diseases by WES.

Acknowledgments

The authors would like to thank the patients and healthy members in the cataract family and other healthy controls participate in this study. This study was supported by the National Natural Science Foundation (No. 81300789, No. 81470665, No. 31427801) and the Shenzhen Municipal Government of China (No. GJHZ20130417140916986).

Financial support and sponsorship

This study was supported by the National Natural Science Foundation (No. 81300789, No. 81470665, No. 31427801) and the Shenzhen Municipal Government of China (No. GJHZ20130417140916986).

Conflicts of interest

There are no conflicts of interest.

References

1. Apple DJ, Ram J, Foster A, Peng Q. Elimination of cataract blindness: A global perspective entering the new millenium. Surv Ophthalmol 2000;45 Suppl 1:S1-196.
2. Gilbert C, Foster A. Childhood blindness in the context of VISION 2020 – The right to sight. Bull World Health Organ 2001;79:227-32.
3. Hejtmancik JF. Congenital cataracts and their molecular genetics. Semin Cell Dev Biol 2008;19:134-49.
4. Huang B, He W. Molecular characteristics of inherited congenital cataracts. Eur J Med Genet 2010;53:347-57.
5. Reis LM, Tyler RC, Muheisen S, Raggio V, Salviati L, Han DP, *et al.* Whole exome sequencing in dominant cataract identifies a new causative factor, CRYBA2, and a variety of novel alleles in known genes. Hum Genet 2013;132:761-70.
6. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, *et al.* SOAP2: An improved ultrafast tool for short read alignment. Bioinformatics 2009;25:1966-7.
7. Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, *et al.* SNP detection for massively parallel whole-genome resequencing. Genome Res 2009;19:1124-32.
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
9. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491-8.
10. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
11. Li MX, Kwan JS, Bao SY, Yang W, Ho SL, Song YQ, *et al.* Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. PLoS Genet 2013;9:e1003143.
12. Dave A, Laurie K, Staffieri SE, Taranath D, Mackey DA, Mitchell P, *et al.* Mutations in the EPHA2 gene are a major contributor to inherited cataracts in South-Eastern Australia. PLoS One 2013;8:e72518.
13. Pasquale EB. Eph receptor signalling casts a wide net on cell behaviour. Nat Rev Mol Cell Biol 2005;6:462-75.
14. Pasquale EB. Eph-ephrin bidirectional signaling in physiology and disease. Cell 2008;133:38-52.
15. Shiels A, Bennett TM, Knopf HL, Maraini G, Li A, Jiao X, *et al.* The EPHA2 gene is associated with cataracts linked to chromosome 1p. Mol Vis 2008;14:2042-55.
16. Zhang T, Hua R, Xiao W, Burdon KP, Bhattacharya SS, Craig JE, *et al.* Mutations of the EPHA2 receptor tyrosine kinase gene cause autosomal dominant congenital cataract. Hum Mutat 2009;30:E603-11.
17. Kaul H, Riazuddin SA, Shahid M, Kousar S, Butt NH, Zafar AU, *et al.* Autosomal recessive congenital cataract linked to EPHA2 in a consanguineous Pakistani family. Mol Vis 2010;16:511-7.
18. Shentu XC, Zhao SJ, Zhang L, Miao Q. A novel p.R890C mutation in EPHA2 gene associated with progressive childhood posterior cataract in a Chinese family. Int J Ophthalmol 2013;6:34-8.
19. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, *et al.* Guidelines for investigating causality of sequence variants in human disease. Nature 2014;508:469-76.