Original Article

Deciphering the association of intronic single nucleotide polymorphisms of crystallin gene family with congenital cataract

Vidya Nair[1], Rajkumar Sankaranarayanan[1,2], Abhay Raghukant Vasavada[1]

Purpose: Introns play an important role in gene regulation and expression. Single nucleotide polymorphisms (SNPs) in introns have the potential to cause disease and alter the genotype–phenotype association. Hence, this study aimed to decipher the association of SNPs in the introns of the crystallin gene in congenital cataracts. **Methods:** SNPs in the introns of crystallin gene family – *CRYAA* (rs3788059), *CRYAB* (rs2070894), *CRYBA4* (rs2071861), and *CRYBB2* (rs5752083, rs5996863) – were genotyped in 248 participants consisting of 141 congenital cataracts and 107 healthy controls by allele-specific oligonucleotide polymerase chain reaction method. Around 10% of samples for each SNPs were sequenced to confirm the genotypes. The allele, genotype, and haplotype frequency were evaluated by the SHEsis online tool. **Results:** Using dominant model, the "A" allele of rs3788059 was found to have an increased risk toward congenital cataract development whereas the "G" allele was found to be protective (AA + AG vs. GG; odds ratio [*OR*] 95% confidence interval [CI] = 3.73 [1.71, 8.15], *P* = 0.0009). The "A" allele of both rs2070894 (AA + AG vs. GG; *OR* [95% CI] = 0.49 [0.29, 0.84], *P* = 0.012) and rs5752083 (AA + AC vs. CC; *OR* [95% CI] = 0.25 [0.08, 0.76], *P* = 0.016) were suggested to have a protective role by the dominant model. The A-C-T haplotype (rs2071861, rs5752083, and rs5996863) was found to be a significant risk factor for the development of congenital cataract. **Conclusion:** Intronic SNPs in crystallin genes may play a role in the predisposition toward congenital cataract. However, the present findings need to be replicated in a large cohort with more number of samples.

Access this article online

Website: www.ijo.in

DOI: 10.4103/ijo.IJO_3062_20

PMID: 34304179

Quick Response Code:

Key words: Crystallin, congenital cataract, intronic, SNP genotyping

Introns, the noncoding segments of DNA (deoxyribonucleic acid) are thought to play a vital role in genome evolution in eukaryotes.[1] Although once considered as junk DNA, introns are gaining importance as they perform a significant role in the regulation of gene expression, mRNA (messenger RNA [ribonucleic acid]) export, splicing, transcription coupling, and enhancing the protein diversity by exon shuffling and alternative splicing.[2-5] With the successful completion of the human genome project and the advent of next-generation sequencing platforms, a large number of intronic single nucleotide polymorphisms (SNPs) have been identified and associated with human diseases through several genome-wide association studies (GWAS).[6-9] Furthermore, introns may be the target for mutations at considerably higher proportion or mutational hotspots because they possess arrays of essential functional elements such as the intron splice enhancers and silencers, trans-splicing elements, and other controlling elements.[10-13] In addition to functional mutations, SNPs in introns may also cause increased susceptibility to disease and modulate the association between genotype and phenotype.[14]

Congenital cataract is characterized by the clouding of the lens, either completely or partially, that significantly affects normal vision either from the beginning or shortly after birth. It is one of the leading causes of treatable childhood blindness and has a prevalence rate of 1 to 6 per 10,000 live birth.[15] It may either be isolated or occur along with other ocular malformations and/or multisystemic disorder.[15] Although both genetic and environmental regulators are well-known causative factors, about 50% of congenital cataracts have been suggested owing to genetic factors.[16] It exhibits autosomal dominant, recessive, X-linked, and mitochondrial mode of inheritance pattern.[17]

More than 90% of the total water-soluble protein in the human eye lens is made up of crystallins that play a vital role in maintaining lens transparency.[18] They are characterized as α-, β-, and γ-crystallin families (encoded by *CRYAA*; *CRYAB*, *CRYBA1*, *CRYBA2*, *CRYBA4*, *CRYBB1*, *CRYBB2*, *CRYBB3*; and *CRYGA*, *CRYGB*, *CRYGC*, *CRYGD*, *CRYGS*, *CRYGN*, *CRYGEP*, *CRYGFP*, *CRYGGP* genes, respectively).[19] Mutation in more than 360 genes to be responsible for congenital cataract have been reported in several studies (Cat-Map; https://cat-map.wustl.edu/).[20] Although the majority of mutations that were identified till date in human congenital cataract is in crystallin

[1]Department of Pediatric Ophthalmology, Iladevi Cataract and IOL Research Centre, Ahmedabad, Gujarat, [2]Department of Ophthalmic Genetics, Aditya Jyot Foundation for Twinkling Little Eyes, Mumbai, Maharashtra, India

Correspondence to: Dr. Rajkumar Sankaranarayanan, Department of Ophthalmic Genetics, Aditya Jyot Foundation for Twinkling Little Eyes, Mumbai - 400 031, Maharashtra, India. E-mail: srajkumar31@gmail.com

Received: 25-Sep-2020 Revision: 21-Feb-2021
Accepted: 21-Mar-2021 Published: 23-Jul-2021

Cite this article as: Nair V, Sankaranarayanan R, Vasavada AR. Deciphering the association of intronic single nucleotide polymorphisms of crystallin gene family with congenital cataract. Indian J Ophthalmol 2021;69:2064-70.

genes,[17] only a few studies have reported the association of intronic SNPs of crystallin genes with congenital cataract.[15,21-23] Furthermore, all these studies have been performed in different ethnic groups and different cohorts of the Indian population. In all these studies, the population from the western region of India was always kept isolated. Considering the importance of introns in human genomes and the dearth of genetic association studies in the western Indian population, the present study was designed to elucidate the association of intronic SNPs of crystallin genes (*CRYAA*, *CRYAB*, *CRYBA4*, and *CRYBB2*) with congenital cataract in a cohort of western India. This study may assist in identifying the disease-associated loci and further help in the implementation of tools for prenatal diagnosis and risk prediction of congenital cataract.

Methods

Recruitment and ocular examination

The study participants were recruited essentially from the western region of India. All the study procedures adhered to the tenets of the Declaration of Helsinki and were approved by the Institutional Ethical Committee. Written informed consent was obtained from the parents and/or the guardians.

A thorough ophthalmic examination was performed. Different visual acuity assessments were performed for children of different age-groups, using Cardiff's acuity test for 1 to 2 years, LEA symbols for 2 to 3 years, Lippman's HOTV test for 3 to 5 years, and Snellen's chart for older than 5 years. The distant direct examination was done to look for anterior segment abnormalities such as corneal opacity, shallow anterior chamber, peripheral anterior synechiae (Peter's anomaly), microcornea, posterior synechiae, the keyhole pupil (iridofundal coloboma), and enlarged ciliary processes with vessels on the lens (PFV). Red reflex screening (Bruckner's test) with direct ophthalmoscope focusing on each dilated pupil (with homatropine 2%) separately from 30 cm distance was performed to identify lenticular opacity,[24,25] and both eyes were visualized simultaneously from 3 ft to identify anisometropia, strabismus, and asymmetric cataract and fixation pattern based on different glows.[26] The cataract was classified based on the zone and morphology of lens opacification observed through either slit-lamp biomicroscopy or under an operating microscope.

Participants with no symptoms of cataract and other ocular disorders were considered as controls and those with cataract were identified as cases. Participants with a history of traumatic cataract, viral infection, neurodevelopmental disorder, chromosomal abnormality, systemic diseases, and in-born errors of metabolism were excluded from this study.

SNP genotyping

About 2 mL of peripheral venous blood was collected from the cases and healthy controls. Genomic DNA was extracted by the salting out method.[27] A total of 5 SNPs from four different genes – *CRYAA* (rs3788059), *CRYAB* (rs2070894), *CRYBA4* (rs2071861), and *CRYBB2* (rs5752083, rs5996863) – were selected from the 1000 Genomes project (http://www.1000genomes.org/). All the SNPs were genotyped by allele-specific oligonucleotide–polymerase chain reaction (ASO-PCR) method. The ASO primers were designed using the WASP online tool (http://bioinfo.biotec.or.th/wasp)[28] and are listed in Table 1. The PCR reaction for the wild and the mutant allele was carried out in two separate tubes each containing 1X Emerald GT PCR Master Mix (TaKaRa Bio Inc., Japan), 50 ng genomic DNA, 20 pM each of allele-specific primers. The thermal cycling steps consisted of one cycle of initial denaturation at 94°C/1 minute, and 40 cycles of the second denaturation at 94°C/30 seconds, annealing at 53–57°C/30 seconds, extension at 72°C/30 seconds, and a final extension at 72°C/3 minutes. All the amplicons were resolved using 4% agarose gel and visualized by UV (ultraviolet) transilluminator on ethidium bromide staining. The allele and genotype frequencies were scored by direct counting. About 10% of both case and control samples were sequenced to confirm the genotypes.

In silico analyses

The effect of intronic SNPs on splicing as well on transcription factor binding was checked using Human Splicing Finder (HSF)[29] and TRANSFAC[30] online tools, respectively.

Statistical analysis

All continuous variables were analyzed by Student's *t* test, and the values were presented as mean ± standard deviation (*SD*). The allele, genotype, haplotype, and Hardy–Weinberg equilibrium (HWE) were analyzed using the SHEsis online tool (http://analysis.bio-x.cn/SHEsisMain.htm).[31] The association of the alleles and the genetic models with the disease was calculated by taking odds ratio (*OR*) at 95% confidence interval (CI; http://www.hutchon.net/confidor.htm).[32] The strength of association of SNPs with the disease between the cases and the controls was tested by Chi-square test (www.socscistatistics.com/pvalues/chidistribution.aspx). Bonferroni's correction was applied for multiple SNPs testing by dividing the alpha error of 0.05 by the total number of SNPs tested. Hence, a *P* value <0.01 was considered statistically significant. Genetic models were considered significant if Yates corrected *P* value is <0.01.

Results

Demography of the participants

All the study participants were from western India, including Gujarat, Madhya Pradesh, Maharashtra, and Rajasthan. A total of 248 participants were recruited, consisting of 141 congenital cataract cases and 107 age-matched healthy controls. All the cases had isolated congenital cataracts of different phenotypes such as nuclear (18.44%), lamellar (22.70%), posterior subcapsular (31.20%), and total cataract (27.66%). There was no significant difference in age (range 0.1–10 years, $P = 0.1$) between the cases (5.87 ± 3.37 years) and the controls (6.45 ± 3.77 years). The demography of the recruited participants is shown in Table 2.

Association of allele, genotype, and haplotype frequencies with disease risk

A total of five intronic SNPs from four crystallin genes were genotyped in this study. The allele and genotype frequencies of all the polymorphisms in both cases and controls were scored by ASO-PCR followed by agarose gel electrophoresis [Fig. 1].

The allele and genotype frequency of SNP-rs3788059 was in HWE, whereas SNPs-rs2070894, rs2071861, rs5752083, and rs5996863 showed a deviation from HWE in both cases and controls [Table 3].

Table 1: List of ASO Primers for SNP Genotyping

Gene and SNP ID	Primer ID	Sequence (5'-3')	Amplicon size (bp)	MAF (1,000 G)
CRYAA-rs3788059 (c.190-370G > A)	SNP1-WRP	GTTGGTCCGTTAGGGTCAATAG	174	A: 0.0004
	SNP1-MRP	GTTGGTCCGTTAGGGTCAATAA		
	SNP1-CFP	GTGAGAAGGAGCATGTGGAAG		
CRYAB-rs2070894 (c.324 + 214G > A)	SNP3-WRP	ATCCCATCATCCCATCTAAGGAG	185	A: 0.26
	SNP3-MRP	ATCCCATCATCCCATCTAAGGAA		
	SNP3-CFP	ATAGTCCAGGTAGTGCTATCAGCTTT		
CRYBA4-rs2071861 (c.159-256A > G)	SNP5-WRP	TGATGTTTCGGGCTGGATAA	265	G: 0.28
	SNP5-MRP	TGATGTTTCGGGCTGGATAG		
	SNP5-CFP	AGGGTAGAGTGTGCAGGAGGTA		
CRYBB2-rs5752083 c.54 + 1112C > A	SNP7-WRP	ATGCTCTCATCAACCCTGGC	110	A: 0.30
	SNP7-MRP	ATGCTCTCATCAACCCTGGA		
	SNP7-CFP	GAGGTGGGAGGACTGTTTGAA		
CRYBB2-rs5996863 (g.17486C > T)	SNP8-WFP	CAATTCCCTTGCCTCTGACC	208	C: 0.38
	SNP8-MFP	CAATTCCCTTGCCTCTGACT		
	SNP8-CRP	TCAGGGTTCTTGGCTTCTCTT		

ASO: Allele-specific oligonucleotide, SNP: Single nucleotide polymorphism, *CRYAA*: Crystallin alpha-A, *CRYAB*: Crystallin alpha-B, *CRYBA4*: Crystallin beta-B4, *CRYBB2*: Crystallin beta-B2, WRP: Wild-type reverse primer, MRP: Mutant reverse primer, WFP: Wild-type forward primer, MFP: Mutant forward primer, CFP: Common forward primer, CRP: Common reverse primer, MAF: Minor allele frequency

Table 2: Demography of the recruited participants

Demography	Congenital cataracts (*n*=141)	Controls (*n*=107)
Female, *n* (%)	60 (42.55)	48 (44.86)
Male, *n* (%)	81 (57.45)	59 (55.14)
Age in years (mean±SD)	5.87±3.37	6.45±3.77
P	0.1	
Cataract type (%)		
Nuclear	26 (18.44)	-
Lamellar	32 (22.70)	-
PSC	44 (31.20)	-
Total	39 (27.66)	-

PSC: Posterior subcapsular cataract, SD: Standard deviation

SNP1: rs3788059

The frequency of "A" allele (*OR* [95% CI] = 3.55 [1.68, 7.51], *P* = 0.0005), and "AG" genotype (*OR* [95% CI] = 3.53 [1.61, 7.73], *P* = 0.001) of SNP-rs3788059 were significantly higher in cases than in the controls. The dominant model for SNP-rs3788059 indicated that "A" allele is associated with increased risk (AA + AG vs. GG; *OR* [95% CI] = 3.73 [1.71, 8.15], *P* = 0.0009) of disease, whereas "G" allele showed protective effect.

SNP2: rs2070894

The frequency of "AG" genotype of SNP-rs2070894 was significantly different between the cases and the controls (*OR* [95% CI] = 0.48 [0.28, 0.82], *P* = 0.007). The dominant model suggested that "A" allele of SNP-rs2070894 is protective (AA + AG vs. GG; *OR* [95% CI] = 0.49 [0.29, 0.84], *P* = 0.012).

SNP3: rs5752083

The frequency of "AC" genotype of SNP-rs5752083 was significantly different between the cases and the controls (*OR* [95% CI] = 0.22 [0.07, 0.68], *P* = 0.005]. The dominant model for rs5752083 indicated a protective effect with "A" allele (AA + AC vs. CC; *OR* [95% CI] = 0.25 [0.08, 0.76], *P* = 0.016).

There was no significant difference in the allele and genotype frequencies of SNPs-rs2071861 and rs5996863 between the cases and the controls.

The allele, genotype, and haplotype frequencies of all the tested SNPs are shown in Tables 3 and 4. Haplotype distribution for SNPs-rs2071861, rs5752083, and rs5996863 was evaluated, as they present on the chromosome number 22. Although the haplotype analysis showed an increased frequency of A-C-T haplotype in cases than in controls (*OR* [95% CI] = 2.66 [1.09, 6.43], *P* = 0.025), these SNPs were not in linkage disequilibrium (*D'* = 0.04, r^2 = 0.00) [Table 5]. After correcting the *P* value, the rs3788059 "GG" genotype, rs2070894 "G" allele, and rs5752083 "C" allele were found to confer protection from congenital cataract (*P* = 0.01), whereas the A-C-T haplotype was found to be a risk factor for the causation of congenital cataract.

Prediction of the effect of intronic SNPs on splicing and transcription factor binding

Analyses of intronic variations using the HSF algorithm showed potential alteration of ESE (exon splicing enhancer) and ESS (exon splicing silencer) motifs for two of the SNPs (rs3788059 and rs5752083). The SNP rs3788059 of *CRYAA* gene located in chromosome 21 at position 43170137 was found to increase the strength of cryptic acceptor site from 39.44 to 67.31, an increase by ~71%. It is found to be responsible for the activation of a cryptic acceptor site [Supplementary Table 1]. Similarly, TRANSFAC analyses showed changes in binding sites for transcription factors [Supplementary Table 1 and Supplementary Figs. 1-3].

Discussion

Majority of congenital cataracts are manifested as a result of genetic variations in crystallin genes. Crystallin gene clusters

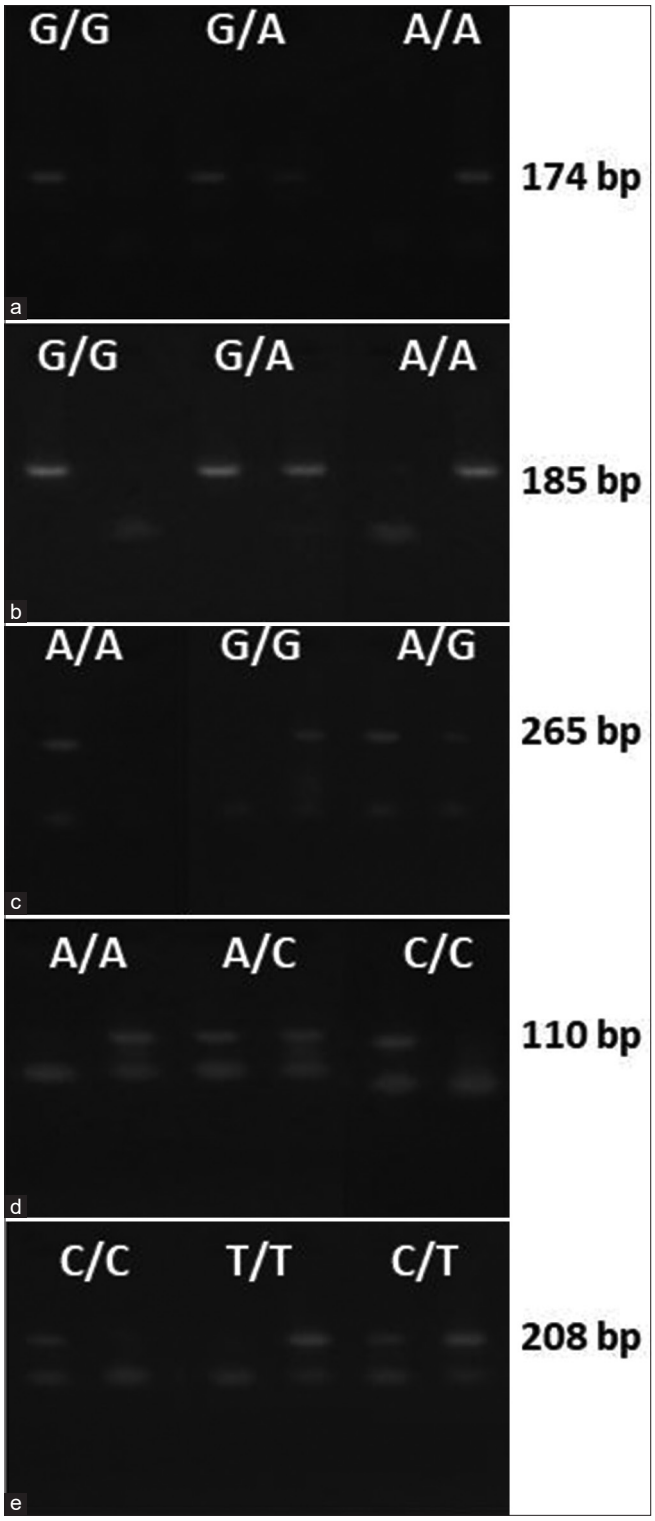


Figure 1: Four percent agarose gel shows the amplification of wild-type and rare alleles of the polymorphisms (a) rs3788059 (G > A), (b) rs2070894 (G > A), (c) rs2071861 (A > G), (d) rs5752083 (C > A), (e) rs5996863 (C > T) with their appropriate amplicon size

are responsible for the synthesis of two major crystallin protein families: α-crystallin and β/γ crystallins. The α-crystallin inhibits lens cell apoptosis and maintains protein stability.[33] Mutations in CRYAA is linked to the loss of α-crystallin protein, which ultimately leads to excessive light scattering and lens opacification.[34,35] On the other hand, β-crystallins aid in lens development and maintaining lens transparency.[36] Mutations in the β-crystallin genes are known to cause abnormality of the protein structure that makes the protein unstable, which precipitates from the solution. This in turn leads to additional protein denaturation and precipitation that subsequently leads to the formation of congenital cataract.[37]

Although 90% of the genome comprises introns, to date only very few reports are available on intronic variations or SNPs associated with congenital cataracts. Even if the intronic SNP does not have a functional consequence, it may exist in linkage disequilibrium with other functional SNPs and thereby help recognize the disease loci. Considering the potential association of SNPs with congenital cataract and the dearth of information on genetic association studies using intronic SNPs,[38,39] the present study was performed to understand the distribution of intronic SNPs rs3788059 (*CRYAA*), rs2070894 (*CRYAB*), rs2071861 (*CRYBA4*), rs5752083 (*CRYBB2*), and rs5996863 (*CRYBB2*) in congenital cataracts and normal healthy controls. Although association studies using these markers have never been reported in congenital cataracts, studies on rs2070894 concerning colorectal and oral cancer[40,41] and rs2071861 concerning high myopia[42,43] have been reported.

In the present study, a higher distribution of the *CRYAA*-rs3788059 "AG" genotype in congenital cataracts is observed. The dominant model also showed that "A" allele is positively associated with an increased risk. HSF analyses for this SNP showed alteration of auxiliary sequences, whereas TRANSFAC analyses revealed loss of REV-ErbA and gain of HNF-1 (hepatocyte nuclear factor–1) and T3R transcription factor binding site. The *CRYAB*-rs2070894 "AG" genotype frequency was found to be more in controls than in cases, and the dominant model showed that *CRYAB*-rs2070894 "A" allele is protective. In two separate studies, Bau *et al*. (2011)[40] and Wu *et al*. (2018)[41] evaluated the association of *CRYAB*-rs2070894 polymorphism with colorectal and oral cancer, respectively, and did not report any significant association of the allele or genotype with the disease. The distribution of the "AC" genotype of *CRYBB2*-rs5752083 was found to be significantly less in cases than in controls. The dominant model for *CRYBB2*-rs5752083 showed that the "A" allele is protective. This SNP showed alteration of auxiliary sequences in HSF analyses and revealed loss of Sp1, Rar-alph, Rev-ErbA, RAR-beta, and ER and gain of YY1 transcription factor binding site in TRANSFAC analyses.

Haplotype analysis of polymorphisms rs2071861, rs5752083, and rs5996863 revealed the association of A-C-T haplotype with the risk of developing congenital cataract. In this study, the association of allele or genotypes of *CRYBA4*-rs2071861 and *CRYBB2*-rs5996863 SNPs with congenital cataract was not established. However, in two separate studies, Kawagoe *et al*. (2017)[42] showed a marginal association and Ho *et al*. (2012)[43] showed a significant association of *CRYBA4*-rs2071861 with high myopia. These observations made a presumption that apart from candidate gene mutations and genetic makeup of an individual, there are additional factors such as environmental factors and gene–gene interactions that might contribute toward the onset and progression of congenital cataract. Nevertheless, in the present study, it is too

Table 3: Allele and genotype distribution of selected SNPs in congenital cataract cases

Gene (SNP)	Allele/Genotype	Cases (*n*=141)	Controls (*n*=107)	OR [95% CI]	χ^2	*P*	*P* (HWE)
CRYAA (rs3788059)	G	244 (0.865)	205 (0.958)	-	-	-	C=0.69; CT=0.65
	A	38 (0.135)	9 (0.042)	3.55 [1.68, 7.51]	12.19	0.0005	
	GG	1.5 (0.745)	98 (0.916)	-	-	-	
	AG	34 (0.241)	9 (0.084)	3.53 [1.61, 7.73]	10.8	0.001	
	AA	2 (0.014)	0 (0.000)	4.67 [0.22, 98.45]	1.85	0.170	
CRYAB (rs2070894)	G	202 (0.716)	137 (0.640)	-	-	-	C=0.0005; CT=7.04-008
	A	80 (0.284)	77 (0.360)	0.7 [0.48, 1.03]	3.26	0.071	
	GG	64 (0.454)	31 (0.290)	-	-	-	
	AG	74 (0.525)	75 (0.701)	0.48 [0.28, 0.82]	7.4	0.007	
	AA	3 (0.021)	1 (0.009)	1.45 [0.15, 14.54]	0.1	0.750	
CRYBA4 (rs2071861)	A	167 (0.592)	136 (0.636)	-	-	-	C=0.02; CT=0.02
	G	115 (0.408)	78 (0.364)	1.2 (0.83, 1.73]	0.96	0.327	
	AA	56 (0.397)	49 (0.458)	-	-	-	
	AG	55 (0.390)	38 (0.355)	1.27 [0.72, 2.23]	0.68	0.411	
	GG	30 (0.213)	20 (0.187)	1.31 [0.66, 2.60]	0.61	0.440	
CRYBB2 (rs5752083)	C	140 (0.496)	104 (0.486)	-	-	-	C=1.14e−007; CT=1.94e−016
	A	142 (0.504)	110 (0.514)	0.96 [0.67, 1.37]	0.05	0.817	
	CC	19 (0.135)	4 (0.135)	-	-	-	
	AC	102 (0.723)	96 (0.897)	0.22 [0.07, 0.68]	8.04	0.005	
	AA	20 (0.142)	7 (0.065)	0.60 [0.15, 2.40]	0.53	0.470	
CRYBB2 (rs5996863)	C	120 (0.426)	100 (0.467)	-	-	-	C=1.24e−013; CT=1.03e−012
	T	162 (0.574)	114 (0.533)	1.18 [0.83, 1.69]	0.86	0.354	
	CC	4 (0.028)	5 (0.047)	-	-	-	
	CT	112 (0.794)	90 (0.841)	1.56 [0.41, 5.96]	0.42	0.520	
	TT	25 (0.177)	12 (0.112)	2.60 [0.59, 11.49]	1.66	0.200	

SNP: Single nucleotide polymorphism, C: Cases; CT: Controls, HWE: Hardy-Weinberg equilibrium, OR: Odds ratio, CI: Confidence interval

Table 4: Dominant and recessive models for the selected SNPs

Gene (SNP)	Genetic model		Cases (*n*=141)	Controls (*n*=107)	OR [95% CI]	χ^2	*P*
CRYAA (rs378805)	Dominant	AA + AG	36 (0.255)	9 (0.084)	3.73 [1.71, 8.15]	12.01	0.001
		GG	105 (0.745)	98 (0.916)	-	-	-
	Recessive	AA	2 (0.014)	0 (0.000)	3.85 [0.18, 81.10]	1.53	0.220
		AG + GG	139 (0.986)	107 (1.000)	-	-	-
CRYAB (rs2070894)	Dominant	AA + AG	77 (0.546)	76 (0.710)	0.49 [0.29, 0.84]	6.94	0.008
		GG	64 (0.454)	31 (0.290)	-	-	-
	Recessive	AA	3 (0.021)	1 (0.009)	2.3 [0.24, 22.47]	0.55	0.460
		AG + GG	138 (0.979)	106 (0.991)	-	-	-
CRYBA4 (rs2071861)	Dominant	GG + AA	85 (0.603)	58 (0.542)	1.28 [0.77, 1.13]	0.92	0.340
		AA	56 (0.397)	49 (0.458)	-	-	-
	Recessive	GG	30 (0.213)	20 (0.187)	1.18 [0.63, 2.21]	0.25	0.620
		AG + AA	111 (0.787)	87 (0.813)	-	-	-
CRYBB2 (rs5752083)	Dominant	AA + AC	122 (0.865)	103 (0.963)	0.25 [0.08, 0.76]	6.85	0.009
		CC	19 (0.135)	4 (0.037)	-	-	-
	Recessive	AA	20 (0.142)	7 (0.065)	2.36 [0.96, 5.81]	3.66	0.060
		AC + CC	121 (0.858)	100 (0.935)	-	-	-
CRYBB2 (rs5996863)	Dominant	TT + CT	137 (0.972)	102 (0.953)	1.68 [0.44, 6.41]	0.59	0.440
		CC	4 (0.028)	5 (0.047)	-	-	-
	Recessive	TT	25 (0.177)	12 (0.112)	1.71 [0.81, 3.57]	2.03	0.150
		CT + CC	116 (0.823)	95 (0.888)	-	-	-

SNP: Single nucleotide polymorphism, OR: Odds ratio, CI: Confidence interval

Table 5: Haplotype distribution for SNPs-rs2071861 (*CRYBA4*), rs5752083 (*CRYBB2*), and rs5996863 (*CRYBB2*)

Haplotype pair	Cases (*n*=282)	Controls (*n*=214)	OR [95% CI]	χ^2	*P*
AAC	13.91 (0.049)	9.56 (0.045)	1.11 [0.48, 2.58]	0.06	0.810
AAT	63.90 (0.227)	60.85 (0.284)	0.74 [0.49, 1.11]	2.16	0.142
ACC	67.00 (0.238)	58.92 (0.275)	0.82 [0.55, 1.23]	0.91	0.340
ACT	22.19 (0.079)	6.67 (0.031)	2.66 [1.09, 6.43]	5.01	0.025
GAC	8.67 (0.031)	3.46 (0.016)	1.93 [0.55, 6.81]	1.09	0.297
GAT	55.52 (0.197)	36.13 (0.169)	1.21 [0.76, 1.92]	0.64	0.425
GCC	30.43 (0.108)	28.07 (0.131)	0.80 [0.46, 1.38]	0.63	0.426
GCT	20.38 (0.072)	10.35 (0.048)	1.53 [0.71, 3.31]	1.20	0.274

SNP: Single nucleotide polymorphism, OR: Odds ratio, CI: Confidence interval. Note: Frequencies <0.03 in both cases and controls have been dropped. Global χ^2=10.33, degrees of freedom=7, *P*=0.17

early to predict how the genotype that showed association with congenital cataract can influence the gene to cause congenital cataract. But it is anticipated that these markers might present near other disease-causing functional SNPs that need to be scrutinized further.

In the present study, the SNP *CRYAA*-rs3788059 alone was in HWE in both cases and controls, whereas the other SNPs *CRYAB*-rs2070894, *CRYBA4*-rs2071861, and *CRYBB2*-rs5752083 and rs5996863 were not. Deviations from HWE can occur due to several reasons such as genotyping error, copy number variation, purifying selection, inbreeding, or population substructure.[44-46] To eliminate potential genotyping error, genotyping was performed thrice by three different observers who were masked for the sample details. Turner *et al*. (2011)[47] reported a consistent deviation of many SNPs from HWE at any given significant threshold. They suggested that such SNPs should never be eliminated from further evaluations; instead, they should be flagged for advanced analysis once the association analysis has been performed.

Conclusion

In conclusion, the intronic SNPs *CRYAA*-rs3788059, *CRYAB*-rs2070894, and *CRYBB2*-rs5752083 were significantly associated with congenital cataract. However, this study has a limitation of small sample size, and hence the present finding needs to be replicated in large cohorts and in different populations to confirm the association.

Financial support and sponsorship

This study was supported by grants from the Indian Council of Medical Research, Government of India (file nos. 5/4/6/10/Oph. 11-NCD II and 5/4/6/2012-RMC).

Conflicts of interest

There are no conflicts of interest.

References

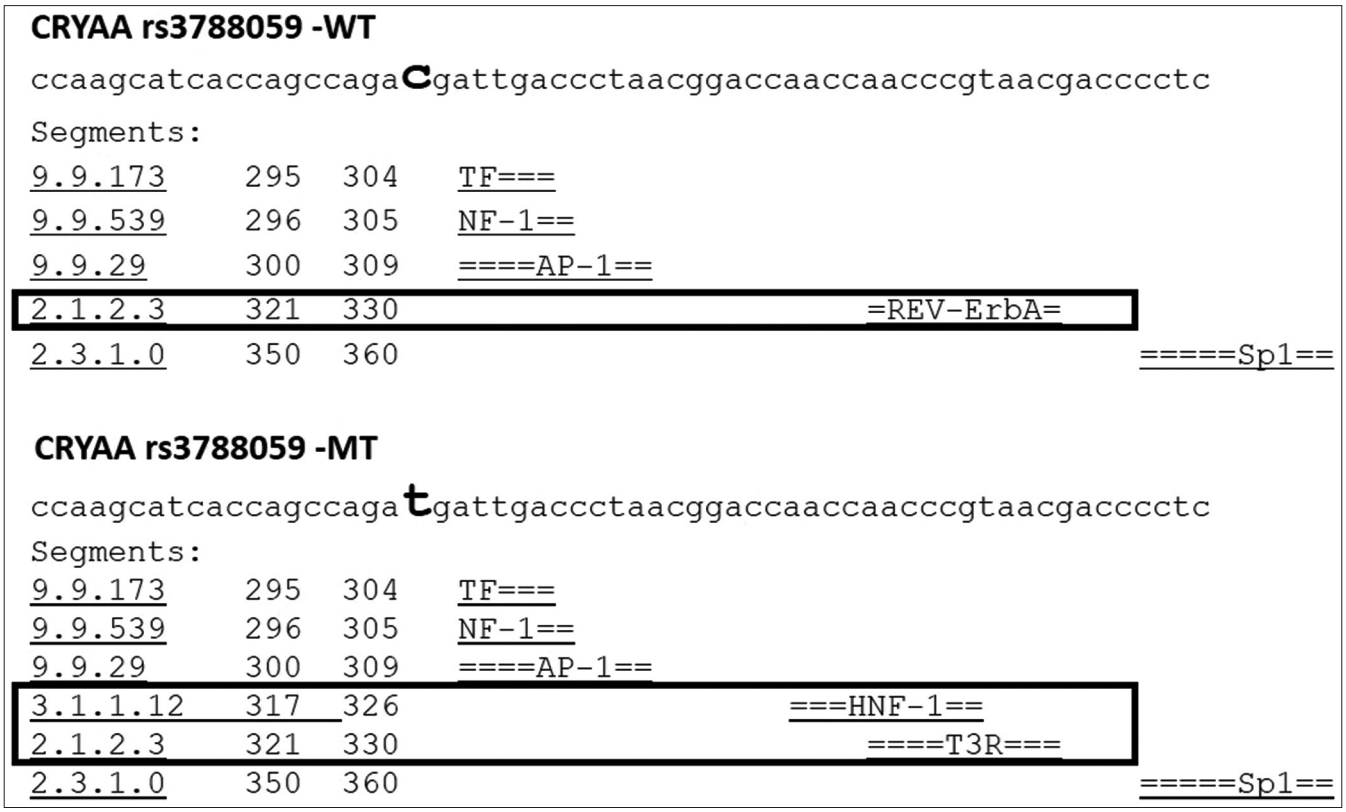
1. Sullivan JC, Reitzel AM, Finnerty JR. A high percentage of introns in human genes were present early in animal evolution from the basal metazoan Nematostella vectensis. Genome Inform 2006;17:219-29.
2. Yang YF, Zhu T, Niu DK. Association of intron loss with high mutation rate in Arabidopsis: Implications for genome size evolution. Genome Biol Evol 2013;5:723-33.
3. Maniatis T, Reed R. An extensive network of coupling among gene expression machines. Nature 2002;416:499-506.
4. Kalsotra A, Cooper TA. Functional consequences of developmentally regulated alternative splicing. Nat Rev Genet 2011;12:715-29.
5. Gorlova O, Fedorov A, Logothetis C, Amos C, Gorlov I. Genes with a large intronic burden show greater evolutionary conservation on the protein level. BMC Evol Biol 2014;14:50.
6. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al*. Finding the missing heritability of complex diseases. Nature 2009;461:747-53.
7. Giral H, Landmesser U, Kratzer A. Into the wild: GWAS exploration of non-coding RNAs. Front Cardiovasc Med 2018;5:181.
8. Ferreiro-Iglesias A, Montes A, Perez-Pampin E, Cañete JD, Raya E, Magro-Checa C, *et al*. Evaluation of 12 GWAS-drawn SNPs as biomarkers of rheumatoid arthritis response to TNF inhibitors. A potential SNP association with response to etanercept. PLoS One 2019;14:e0213073.
9. Cirillo E, Kutmon M, Hernandez MG, Hooimeijer T, Adriaens ME, Eijssen LMT, *et al*. From SNPs to pathways: Biological interpretation of type-2 diabetes (T2DM) genome wide association study (GWAS) results. PLoS One 2018;13:e0193515.
10. Wang X, Wang K, Radovich M, Wang Y, Wang G, Feng W, *et al*. Genome-wide prediction of cis-acting RNA elements regulating tissue-specific pre-mRNA alternative splicing. BMC Genomics 2009;10(Suppl 1):S4.
11. Tress ML, Martelli PL, Frankish A, Reeves GA, Wesselink JJ, Yeats C, *et al*. The impications of alternative splicing in the ENCODE protein complement. Proc Natl Acad Sci U S A 2007;104:5495-500.
12. Gingeras TR. Implications of chimeric non-co-linear transcripts. Nature 2009;461:206-11.
13. Solis AS, Shariat N, Patton JG. Splicing fidelity, enhancers, and disease. Front Biosci 2008;13:1926-42.
14. Cooper DN. Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. Hum Genomics 2010;4:284-8.
15. Santhiya AT, Shyam Manohar M, Rawlley D, Vijayalakshmi P, Namperumalsamy P, *et al*. Novel mutations in the gamma-crystallin genes causes autosomal dominant congenital cataracts. J Med Genet 2002;39:352-8.
16. Javadiyan S, Craig JE, Sharma S, Lower KM, Casey T, Haan E, *et al*. Novel missense mutation in the bZIP transcription factor, MAF, associated with congenital cataract, developmental delay, seizures and hearing loss (Aymé-Gripp syndrome). BMC Med Genet 2017;18:52.
17. Hejtmancik JF, Smaoui N. Molecular genetics of cataract. In: Wissinger B, Kohl S, Langenbeneck U, editors. Genetics in Ophthalmology. Basel: Karger; 2003. p. 67-82.
18. Khan AO, Aldahmesh MA, Alkuraya FS. Phenotypes of recessive pediatric cataract in a cohort of children with identified homozygous gene mutations (An American Ophthalmological Society Thesis). Trans Am Ophthamol Soc 2015;113:T7.

19. Wistow G. The human crystalline gene families. Hum Genomics 2012;6:26.
20. Shiels A, Bennett TM, Hejtmancik JF. Cat-Map: Putting cataract on the map. Mol Vis 2010;6:2007-15.
21. Santhiya ST, Manisastry SM, Rawlley D, Malathi R, Anishetty S, Gopinath PM, *et al*. Mutation analysis of congenital cataracts in Indian families: Identification of SNPs and a new causative allele in CRYBB2 gene. Invest Ophthalmol Vis Sci 2004;45:3599-607.
22. Cui XJ, Lv FY, Li FH, Zheng K. Correlations of single nucleotide polymorphisms of CRYAA and CRYAB genes with the risk and clinicopathological features of children suffering from congenital cataract. Medicine (Baltimore) 2017;96:e7158.
23. Mehra S, Kapur S, Vasavada AR. Polymorphisms of the gamma crystalline A and B genes among Indian patients with pediatric cataract. J Postgrad Med 2011;57:201-5.
24. Griffin JR, Cotter SA. The Bruckner test: Evaluation of clinical usefulness. Optom Vis Sci 1986;63:957-61.
25. Mussavi M, Asadollahi K, Janbaz F, Mansoori E, Abbasi N. The evaluation of red reflex sensitivity and specificity test among neonates in different conditions. Iran J Pediatr 2014;24:697-702.
26. Anker S, Atkinson J, Braddick O, Ehrlich D, Hartley T, Nardini M, *et al*. Identification of infants with significant refractive error and strabismus in a population screening program using noncycloplegic videorefraction and orthoptic examination. Invest Ophthalmol Vis Sci 2003;44:497-504.
27. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.
28. Wangkumhang P, Chaichoompu K, Ngamphiw C, Ruangrit U, Chanprasert J, Assawamakin A, *et al*. WASP: A web-based allele-specific PCR assay designing tool for detecting SNPs and mutations. BMC Genomics 2007;8:275.
29. Wingender E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. Brief Bioinform 2008;9:326-32.
30. Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human splicing finder: An online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009;37:e67.
31. Shi YY, He L. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. Cell Res 2005;15:97-8.
32. Bland JM, Altman DG. Statistics notes. The odds ratio. BMJ 2003;320:1468.
33. Li L, Fan D, Zhao Y, Li Y, Kong D, Cai F, *et al*. Two novel mutations identified in ADCC families impair crystallin protein distribution and induce apoptosis in human lens epithelial cells. Sci Rep 2017;7:17848. doi: 10.1038/s41598-017-18222-z.
34. Andley UP, Goldman JW. Autophagy and UPR in alpha-crystallin mutant knock-in mouse models of hereditary cataracts. Biochim Biophys Acta 2016;1860:234-9.
35. Datiles MB, Ansari RR, Suh KI, Vitale S, Reed GF, Zigler JS Jr, *et al*. Clinical detection of precataractous lens protein changes using dynamic light scattering. Arch Ophthalmol 2008;126:1687-93.
36. Hejtmancik JF, Kaiser MI, Piatigorsky J. Molecular biology and inherited disorders of the eye lens. The metabolic and molecular basis of inherited disease 2001;8:6033-62.
37. Reddy MA, Bateman OA, Chakarova C, Ferris J, Berry V, Lomas E, *et al*. Characterization of the G91del CRYBA1/3-crystallin protein: A cause of human inherited cataract. Hum Mol Genet 2004;945-53.
38. Vidya NG, Vasavada AR, Rajkumar S. Evaluating the association of bone morphogenetic protein 4-V152A and SIX homeobox 6-H141N polymorphisms with congenital cataract and microphthalmia in Western Indian population. J Postgrad Med 2018;64:86-91.
39. Vidya NG, Ganatra D, Vasavada AR, Rajkumar S. Association of *FOXE3*-p.Ala170Ala and *PITX3*-p.Ile95Ile polymorphisms with congenital cataract and microphthalmia. J Ophthalmic Vis Res 2018;13:397-402.
40. Bau DT, Tsai CW, Lin CC, Tsai RY, Tsai MH. Association of alpha B-crystallin genotypes with oral cancer susceptibility, survival, and recurrence in Taiwan. PLoS One 2011;6:e16374.
41. Wu X, Zheng YZ, Han B, Wang K. Alpha B-crystallin C-802G polymorphism and colorectal cancer susceptibility and clinical outcome in Chinese population. Sci Rep 2018;8:11731.
42. Kawagoe T, Ota M, Meguro A, Takeuchi M, Yamane T, Shimazaki H, *et al*. Associations between CRYBA4 gene variants and high myopia in a Japanese population. Clin Ophthalmol 2017;11:2151-6.
43. Ho DW, Yap MK, Ng PW, Fung WY, Yip SP. Association of high myopia with crystalline beta A4 (CRYBA4) gene polymorphisms in the linkage-identified MYP6 locus. PLoS One 2012;7:e40238.
44. Lee S, Kasif S, Weng Z, Cantor CR. Quantitative analysis of single nucleotide polymorphisms within copy number variation. PLoS One 2008;3:e3906.
45. Wang J, Shete S. Testing departure from Hardy-Weinberg proportions. Methods Mol Biol 2012;850:77-102.
46. Graffelman J, Jain D, Weir B. A genome-wide study of Hardy-Weinberg equilibrium with next generation sequence data. Hum Genet 2017;136:727-41.
47. Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, *et al*. Quality control procedures for genome wide association studies. Curr Protoc Hum Genet 2011;Chapter 1:Unit1.19. doi: 10.1002/0471142905.hg0119s68.

Supplementary Table 1: Effect of SNPs on splicing and transcription factor binding

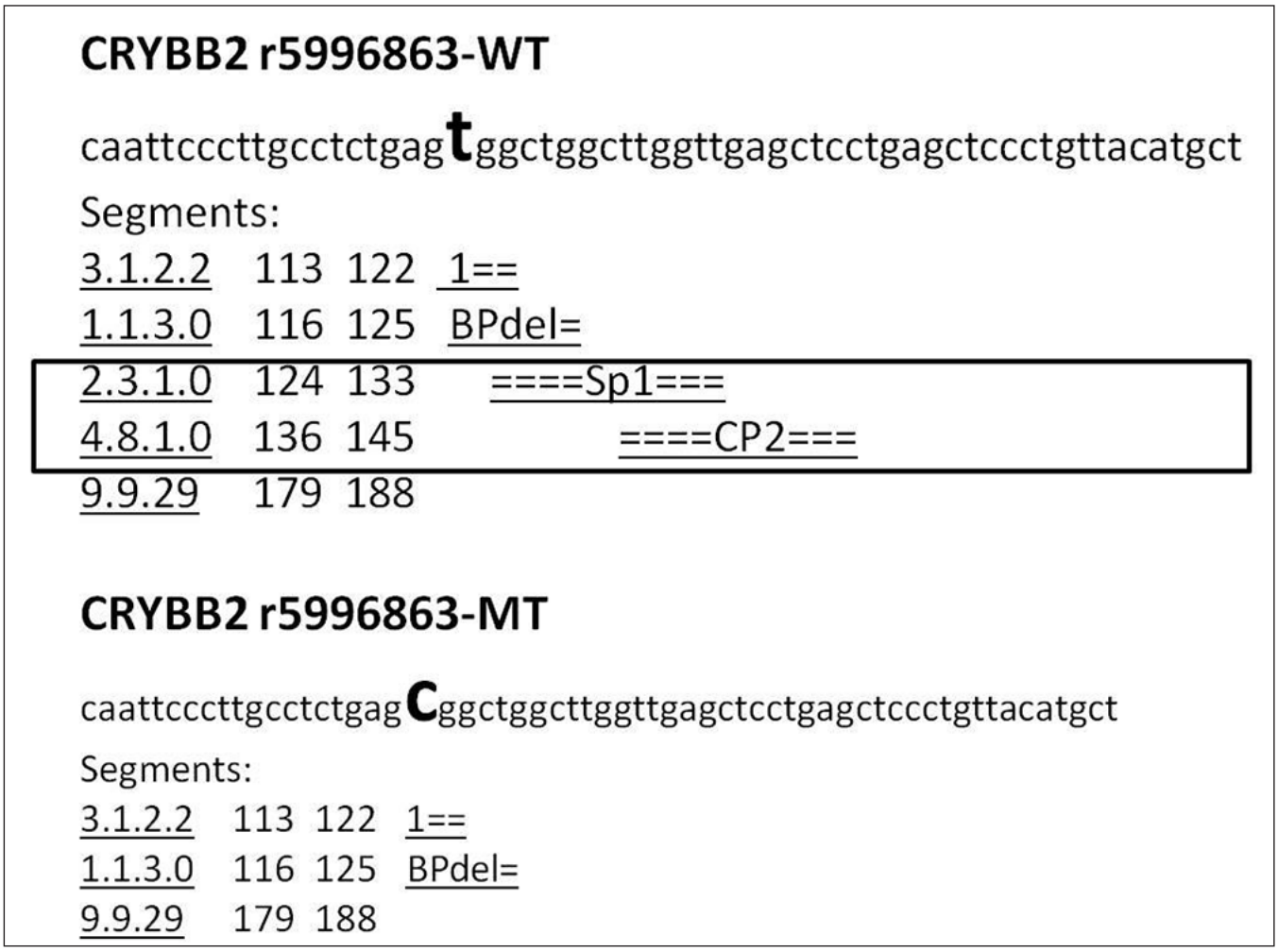
Gene and SNP ID	Human Splicing Finder Analyses		TRANSFAC Analyses
CRYAA (rs3788059)	Signal	Interpretation	Loss of REV-ErbA and gain of HNF-1 and T3R transcription factor binding site
	Alteration of auxiliary sequences	Significant alteration of ESE/ESS motifs ratio (10)	
	New acceptor splice site	Activation of a cryptic acceptor site. Potential alteration of splicing	
	HSF acceptor site (matrix AG)	ACCAGCCAGACGAT > ACCAGCCAGAAGAT (chr21:43170137; 39.44 > 67.31 (70.66%))	
CRYAB (rs2070894)	No significant impact on splicing signals.	No alteration in splicing	No changes in TF binding sites
CRYBA4 (rs2071861)	No significant impact on splicing signals.	No alteration in splicing	No changes in TF binding sites
CRYBB2 (rs5752083)	Alteration of auxiliary sequences	Significant alteration of ESE/ESS motifs ratio (6)	Loss of Sp1, Rar-alph, Rev-ErbA, RAR-beta and ER and gain of YY1 transcription factor binding site
CRYBB2 (rs5996863)	Upstream variant	Not applicable	Loss of TF Sp1 and CP2 transcription factor binding site

SNP: Single nucleotide polymorphism, HSF: Human Splicing Finder, HNF: Hepatocyte nuclear factor, TF: Transcription factor, ESS: Exon splicing silencer, ESE: Exon splicing enhancer



Supplementary Figure 1: TRANSFAC analysis for *CRYAA* rs3788059 wild type and mutant shows the loss of binding site for transcription factor REV-ErbA and gain of binding site for transcription factor HNF-1 and T3R in the mutant

Supplementary Figure 2: TRANSFAC analysis for CRYBB2 rs5752083 wild type and mutant shows the loss of binding site for transcription factor for Sp1, Rar-alph, Rev-ErbA, RAR-beta and ER and gain of binding site for transcription factor YY1



Supplementary Figure 3: TRANSFAC analysis for CRYBB2 rs5996863 wild type and mutant shows the loss of binding site for transcription factor Sp1 and CP2 in the mutant