A novel missense mutation in *HSF4* causes autosomal dominant congenital lamellar cataract in a British family

Vanita Berry[1,*,+], Nikolas Pontikos[1,3,+], Anthony Moore[2,4], Alexander C W Ionides[2], Vincent Plagnol[3], Michael E Cheetham[1] and Michel Michaelides[1,2*]

[1]UCL Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK;

[2]Moorfields Eye Hospital, London EC1V 2PD, UK;

[3]UCL Genetics Institute, University College London, London WC1E 6BT, UK;

[4] Ophthalmology Department, University of California School of Medicine USA

[+]Authors contributed equally to this work.

Corresponding authors:
Dr Vanita Berry, Department of Genetics, Institute of Ophthalmology, University College London, 11–43 Bath Street, London EC1V 9EL, UK; Phone: +44 207 608 4041 FAX: +44 207 608 6863; email: v.berry@ucl.ac.uk

Professor Michel Michaelides, Department of Genetics, Institute of Ophthalmology, University College London, 11–43 Bath Street, London EC1V 9EL, UK; Phone: +44 207 608 6864, FAX: +44 207 608 6903; email: michel.michaelides@ucl.ac.uk

For proofs and payment please contact Vanita Berry.

Abstract

Background: Inherited cataract, opacification of the lens, is the most common worldwide cause of blindness in children. We aimed to identify the genetic cause of isolated autosomal dominant lamellar cataract in a five generation British family.

Methods: Whole exome sequencing (WES) was performed on two affected individuals of the family and further validated by direct sequencing in family members.

Results: A novel missense mutation NM_001040667.2:c.190A>G;p.K64E was identified in the DNA-binding-domain of heat shock transcription factor 4 (*HSF4*) and found to co-segregate with disease.

Conclusion: We have identified a novel mutation in *HSF4* in a large British pedigree causing dominant congenital lamellar cataract. This is the second mutation in this gene found in the British population. This mutation is likely to be dominant negative and affect the DNA binding affinity of HSF4.

Key Words: Congenital cataract, *HSF4*, Whole exome sequencing

Introduction

Cataract, the cloudiness of the lens is the most frequent cause of blindness world-wide, representing almost half of all causes of blindness globally: the WHO estimates that 18 million people are bilaterally blind from cataract[1]. Congenital cataracts are seen in 1-15 per 10,000 births in the UK and are a significant cause of childhood visual impairment[2,3]. Congenital cataract can occur in isolation or in association with other non-ocular manifestations, and is a predominant feature in more than 200 genetic disorders. Congenital cataract may be familial and display considerable genotypic and phenotypic heterogeneity[4]. Inheritance is most commonly autosomal dominant (AD), usually with complete penetrance but with highly variable expressivity. The phenotypic classification of the cataract depends on the position and type of the lens opacity such as: anterior polar, posterior polar, nuclear, lamellar, coralliform, blue-dot (cerulean), cortical, pulverulent, polymorphic, complete cataract and posterior nuclear cataract[5,6]. Significant progress has been made in identifying the molecular genetic basis of human cataract. More than 40 genes have been recruited including the large family of crystallins encoding transparent intracellular lens proteins, membrane gap junction proteins (connexins), water channel proteins (aquaporins), solute carrier protein, cytoskeletal proteins, transmembrane proteins, lens intrinsic membrane protein, chromatin modifying protein-4B, receptor tyrosine kinase gene EPH receptor A2[7], an endoplasmic reticulum membrane-embedded protein, Wolframin [8] and transcription factor genes (e.g. *PAX6, FOXE3*, *EYA1*, *MAF*, *PITX3 and HSF4*)[7].

Herein we report a novel mutation in the heat shock transcription factor 4 (*HSF4*) gene causing an isolated autosomal dominant lamellar cataract in a five generation British family.

Methods

Phenotyping: In this study the family was identified through the proband attending the Genetic Service at Moorfields Eye Hospital, London, UK. Local ethics committee approval was obtained and all of the participants gave written informed consent. All the family members underwent full ophthalmic examination, including slit lamp examination; all affected individuals were diagnosed as having isolated lamellar cataract.

Whole Exome Sequencing (WES) and Bioinformatics Analysis: Genomic DNA was extracted from EDTA sequestered blood samples taken with informed consent and local ethical approval using the Nucleon II DNA extraction kit (Scotlab Bioscience, Strathclyde, Scotland, UK). The DNA was sequenced at AROS Applied Biotechnology (Aarhus, Denmark). Exon capture and target enrichment was performed using the SureSelectXT Human All Exon V5 kit (Agilent, Santa Rosa, CA, USA). Paired-end sequencing was ran on an Illumina Hiseq 2500 high throughput sequencer (Illumina, San Diego, CA, USA) generating mean exome coverage of 50x. Raw data in fastq format was analysed using the Phenopolis platform[9]. The short read sequence data were aligned using novoalign (version 3.02.08). Variants and indels were called according to GATK best practices (joint variant calling followed by variant quality score recalibration)[10]. The variants were then annotated using

the Variant Effect Predictor (VEP)[11]. Variants were then filtered to only contain novel variants not present in public control databases Kaviar (Glusman et al. 2011) and gnomAD (http://gnomad.broadinstitute.org/), and predicted to be moderately or highly damaging according to the VEP. Cosegregation of the filtered variants in both affected individuals was then performed. Finally, the list of variants was further screened using Phenopolis, for genes associated with the Human Phenotype Ontology [15] term "lamellar cataract" (HP:0007971) according to OMIM [16].

Structural Bioinformatics: The protein structure of HSF4 was analysed using SWISS-MODEL[12] (https://swissmodel.expasy.org/repository/uniprot/Q9ULV5). The best PDB[13] match, with a match of 75%, was the structure of the DNA-binding-domain (DBD) HSF1 (PDB ID 2LDU, solved with NMR by the Northeast Structural Genomics Consortium Target HR3023C 2011). An X-Ray solved structure of HSF1 DBD interacting with DNA was also found in PDB (PDB ID 5D5U (ref. 14)). All structures were downloaded in PDB format and analysed using Pymol (version 1.8) locally. The Pymol Mutagenesis tool was used to visualise the effect of the mutation on the HSF1-DNA complex.

Sanger sequencing: Bi-directional direct Sanger sequencing was performed to validate the variant identified by next-generation sequencing. Genomic DNA was amplified by PCR using GoTaq 2X master mix (AB gene; Thermo Scientific, Epsom, UK) and *HSF4*-specifc primers designed with Primer3 http://bioinfo.ut.ee/primer3-0.4.0/primer3/. PCR conditions

were followed as: 94°C for 10 minutes of initial denaturation followed by 30 cycles of amplification of 30 seconds at 94°C, 30 seconds at 60°C, and 45 seconds at 72°C. After the PCR products were reacted with BigDye Terminator v3.1, they were run on ABI 3730 Genetic Analyzer (both from Applied Biosystems) and analyzed using SeqMan Pro (version 8.0.2 from DNASTAR) sequence analysis. After validating the variant, family segregation was performed in all the individuals. Further, 96 unrelated patients from our autosomal dominant congenital cataract (ADCC) panel were screened for this variant by bi-directional Sanger sequencing to look for any other occurrence of the mutation.

Results

A large five-generation pedigree comprising 16 members, including 9 affected individuals, 3 unaffected individuals, and 4 spouses were examined and all affected members had evidence of lamellar cataract (Figure 1).

WES was undertaken in two affected individuals (IV-1, IV-6). Variant annotation and filtering was performed using the Phenopolis platform. From a total of 149,133 variants in the family, 43,441 were found to co-segregate in the two affected individuals. A filter for novel variants which were moderately or highly damaging was then applied and returned 79 variants. After gene panel screening, a rare heterozygous damaging variant, NM_001040667.2:c.190A>G, in *HSF4* on chromosome 16q22.1 was identified. Direct sequencing confirmed that the missense mutation c.190A>G in exon 4 of *HSF4* co-segregated with all affected members of the family (Figure 2).

This single base change is predicted to result in a lysine (K) to glutamic acid (E) amino acid substitution (p.K64E) in the highly conserved DNA Binding Domain, as shown on the protein structure of HSF4 (Figure 3.A). The structural change on HSF4 induced by the mutation was estimated to be negligible according to the difference in the PDB models returned by SWISS-MODEL (root mean square deviation of 0.009). However, the homologous amino acid has a hydrogen bond to the DNA molecule, in the structure of DNA binding complex of the homologous protein HSF1 (Figure 3.B). The mutation from a lysine to a glutamic acid is likely to affect that bond, as the lysine is positively charged whereas glutamic acid is negatively charged, which could have some effect on the binding affinity to the DNA molecule (Figure 3.C). Further in-silico analysis and experimental work is required to verify this claim.

Discussion

Here we report a missense mutation NM_001040667.2:c.190 A>G in the heat shock transcription factor 4 (*HSF4)* gene on chromosome 16q22.1 in a five generation English pedigree with congenital lamellar cataract. All the affected family members had bilateral congenital cataract and age at onset varied from birth to age 5-7 years. One of the patients IV-6 also had glaucoma.

Lamellar cataract is also referred to as zonular cataract and is one of the commonest phenotypes of ADCC. The inner fetal nucleus is made up of a clear lens surrounded by an opacified shell that is in turn surrounded by clear cortex, which may contain opacities referred to as “riders”

or “cortical spokes”. Lamellar cataract represents a disturbance in the lens development at a particular time and the cataractous “shell” varies in size according to the stage of fetal development at which the disturbance occurs[5]. Heat shock factor 4 protein (HSF4) is expressed in both cell types (epithelial cells and fiber cells) at an early stage of lens development and is required for normal cell growth and differentiation of these two cell types [17] .

HSF4 belongs to the family of heat shock transcription factors that regulate the expression of heat shock proteins in response to various stresses, such as high temperature, infection, and free radicals, and also in development. HSF4 expression is tissue-specific and has two splice forms, i) HSF4a which inhibits transcription of other heat shock factor genes by binding directly to the heat shock element and ii) HSF4b which contains an additional 30 amino acids, acts as an activator of transcription. *HSF4* comprises 15 exons and encodes a heat shock transcriptional protein of 493 amino acid residues with a DNA binding domain (DBD), an amino-terminal hydrophobic repeat (HR/A-B), an isoform specific region and downstream of hydrophobic repeat (DHR) [18,19]

As described in Table 1, sixteen different mutations in *HSF4* have been identified in different ethnicities, for both for autosomal dominant and autosomal recessive congenital cataract, as well as age related cataract[20,21]. Eight of these have been found in the Chinese population, three autosomal recessive mutations in Pakistani families[22] and one splice site mutation in a large Tunisian family[23]. Recently Benham et al[24], reported a novel homozygous mutation in an Iranian family causing congenital cataract. So far, only two mutations have been

found in Europeans; in Danish families, p.R119C was a recurrent mutation; and a novel mutation p.H35Y in a British family with no description of the cataract phenotype[25]. Each of these mutations were found in autosomal dominant cataract including our novel heterozygous mutation p.K64E in a large British pedigree with lamellar cataract.

Interestingly, all known autosomal dominant mutations in *HSF4* lie within the α-helical DBD, whereas the recessive mutations lie outside this highly conserved functional domain[26].

These observations support that mutations in the DBD are dominant negative rather than loss-of-function. The binding affinity of HSF4 to the DNA might be perturbed which could interfere with the wildtype molecule. This could potentially disrupt the transcription factor "orchestra" in eye development, leading to abnormal lens cells conducive to protein aggregates and hence the onset of cataract. It has also been confirmed that, *HSF4* regulates lens fiber cell differentiation by modulating expression of certain lens structural proteins, such as lens specific crystallins, beaded filament proteins and fibroblast growth factors [17]. Further studies have shown that *HSF4* plays a crucial part in the de-nucleation of lens fiber cells through regulating DNase 2β expression level and DNase activity[27]. Merath et al[28] have demonstrated that the functional loss associated with autosomal recessive *HSF4* mutations is due to a loss of regulatory domains situated at the C-terminal end, indicating that the transcriptional activation of *HSF4* is mediated by interactions between activator and repressor domains within the C-terminal end. Furthermore, the *Hsf4* knockout mice which demonstrate a cataract phenotype, have

abnormal lens development with nondegenerated nuclei in the secondary fiber cells, suggesting a critical role of *HSF4* in the lens fiber cell differentiation[29].

Acknowledgments

Supported by grants from the National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital National Health Service Foundation Trust and UCL Institute of Ophthalmology (UK), Moorfields Eye Hospital Special Trustees (UK), Moorfields Eye Charity (UK), and the Foundation Fighting Blindness (USA). Michel Michaelides is supported by an FFB Career Development Award. We would like to thank the members of the family for taking part in this study. We wish to also thank Manav Ponnekanti who helped in the formatting of the document.

Conflict of Interest

The authors report no conflict of Interest.

References:

1. Gilbert C, Foster A. Childhood blindness in the context of VISION 2020--the right to sight. *Bull World Health Organ* 2001; 79: 227–232.

2. Reddy MA, Francis PJ, Berry V, et al. Molecular genetic basis of inherited cataract and associated phenotypes. *Surv Ophthalmol* 2004; 49: 300–315.

3. Krumpaszky HG, Klauss V. Epidemiology of blindness and eye disease. *Ophthalmologica* 1996; 210: 1–84.

4. Churchill A, Graw J. Clinical and experimental advances in congenital and paediatric cataracts. *Philos Trans R Soc Lond B Biol Sci* 2011; 366: 1234–1249.

5. Ionides A, Francis P, Berry V, et al. Clinical and genetic heterogeneity in autosomal dominant cataract. *Br J Ophthalmol* 1999; 83: 802–808.

6. Francis PJ, Berry V, Bhattacharya SS, et al. The genetics of childhood cataract. *J Med Genet* 2000; 37: 481–488.

7. Shiels A, Bennett TM, Hejtmancik JF. Cat-Map: putting cataract on the map. *Mol Vis* 2010; 16: 2007–2015.

8. Berry V, Gregory-Evans C, Emmett W, et al. Wolfram gene (WFS1) mutation causes autosomal dominant congenital nuclear cataract in humans. *Eur J Hum Genet* 2013; 21: 1356–1360.

9. Pontikos N, Yu J, Moghul I, et al. Phenopolis: an open platform for harmonization and analysis of genetic and phenotypic data. *Bioinformatics*. Epub ahead of print 15 March 2017. DOI: 10.1093/bioinformatics/btx147.

10. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20: 1297–1303.

11. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol* 2016; 17: 122.

12. Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 2003; 31: 3381–3385.

13. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res* 2000; 28: 235–242.

14. Neudegger T, Verghese J, Hayer-Hartl M, et al. Structure of human heat-shock transcription factor 1 in complex with DNA. *Nat Struct Mol Biol* 2016; 23: 140–146.

15. Köhler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. *Nucleic Acids Res*. Epub ahead of print 28 November 2016. DOI: 10.1093/nar/gkw1039.

16. Hamosh A, Scott AF, Amberger JS, et al. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2005; 33: D514–7.

17. Fujimoto M, Izu H, Seki K, et al. HSF4 is required for normal cell growth and differentiation during mouse lens development. *EMBO J* 2004; 23: 4297–4306.

18. Nakai A, Tanabe M, Kawazoe Y, et al. HSF4, a new member of the human heat shock factor family which lacks properties of a transcriptional activator. *Mol Cell Biol* 1997; 17: 469–481.

19. Tanabe M, Sasai N, Nagata K, et al. The mammalian HSF4 gene generates both an activator and a repressor of heat shock genes by alternative splicing. *J Biol Chem* 1999; 274: 27845–27856.

20. Bu L, Jin Y, Shi Y, et al. Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. *Nat Genet* 2002; 31: 276–278.

21. Enoki Y, Mukoda Y, Furutani C, et al. DNA-binding and transcriptional activities of human HSF4 containing mutations that associate with congenital and age-related cataracts. *Biochim Biophys Acta* 2010; 1802: 749–753.

22. Sajjad N, Goebel I, Kakar N, et al. 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.

23. 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.

24. Behnam M, Imagawa E, Chaleshtori ARS, et al. A novel homozygous mutation in HSF4 causing autosomal recessive congenital cataract. *J Hum Genet* 2016; 61: 177–179.

25. Gillespie RL, O'Sullivan J, Ashworth J, et al. Personalized diagnosis and management of congenital cataract by next-generation sequencing. *Ophthalmology* 2014; 121: 2124–37.e1–2.

26. Nakai A. Molecular basis of HSF regulation. *Nat Struct Mol Biol* 2016; 23: 93–95.

27. Cui X, Wang L, Zhang J, et al. HSF4 regulates DLAD expression and promotes lens de-nucleation. *Biochim Biophys Acta* 2013; 1832: 1167–1172.

28. Merath K, Ronchetti A, Sidjanin DJ. Functional analysis of HSF4 mutations found in patients with autosomal recessive congenital cataracts. *Invest Ophthalmol Vis Sci* 2013; 54: 6646–6654.

29. Shi X, Cui B, Wang Z, et al. Removal of Hsf4 leads to cataract development in mice through down-regulation of gamma S-crystallin and Bfsp expression. *BMC Mol Biol* 2009; 10: 10.

30. Lv H, Huang C, Zhang J, et al. A novel HSF4 gene mutation causes autosomal-dominant cataracts in a Chinese family. *G3* 2014; 4: 823–828.

31. Li D, Wang S, Ye H, et al. Distribution of gene mutations in sporadic congenital cataract in a Han Chinese population. *Mol Vis* 2016; 22: 589–598.

32. Shi Y, Shi X, Jin Y, et al. Mutation screening of HSF4 in 150 age-related cataract patients. *Mol Vis* 2008; 14: 1850–1855.

33. Ke T, Wang QK, Ji B, et al. Novel HSF4 mutation causes congenital total white cataract in a Chinese family. *Am J Ophthalmol* 2006; 142: 298–303.

34. Liu L, Zhang Q, Zhou L-X, et al. A novel HSF4 mutation in a Chinese family with autosomal dominant congenital cataract. *J Huazhong Univ Sci Technolog Med Sci* 2015; 35: 316–318.

35. Hansen L, Mikkelsen A, Nürnberg P, et al. Comprehensive mutational screening in a cohort of Danish families with hereditary congenital cataract. *Invest Ophthalmol Vis Sci* 2009; 50: 3291–3303.

36. 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.

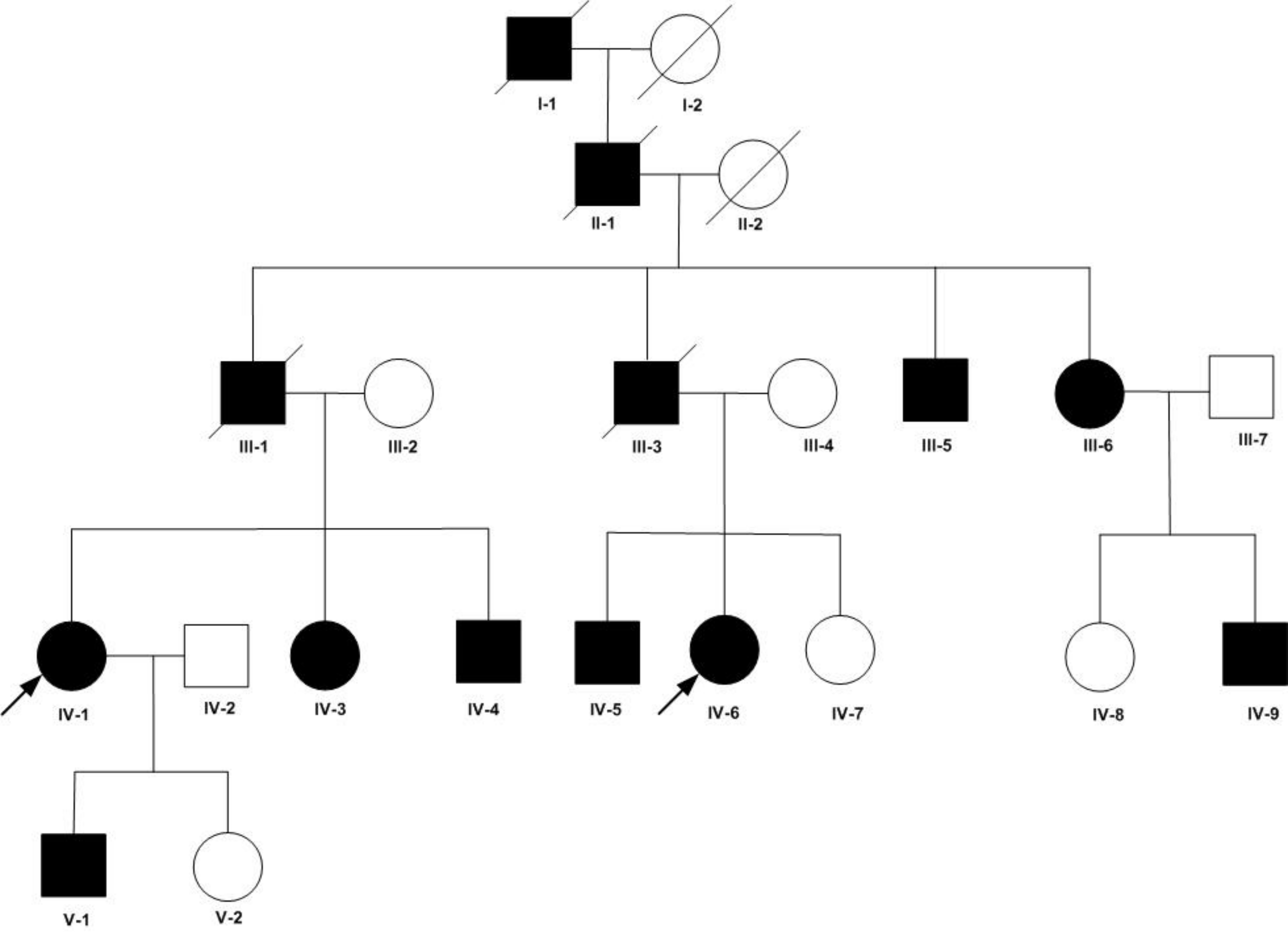
Titles and legends to figures

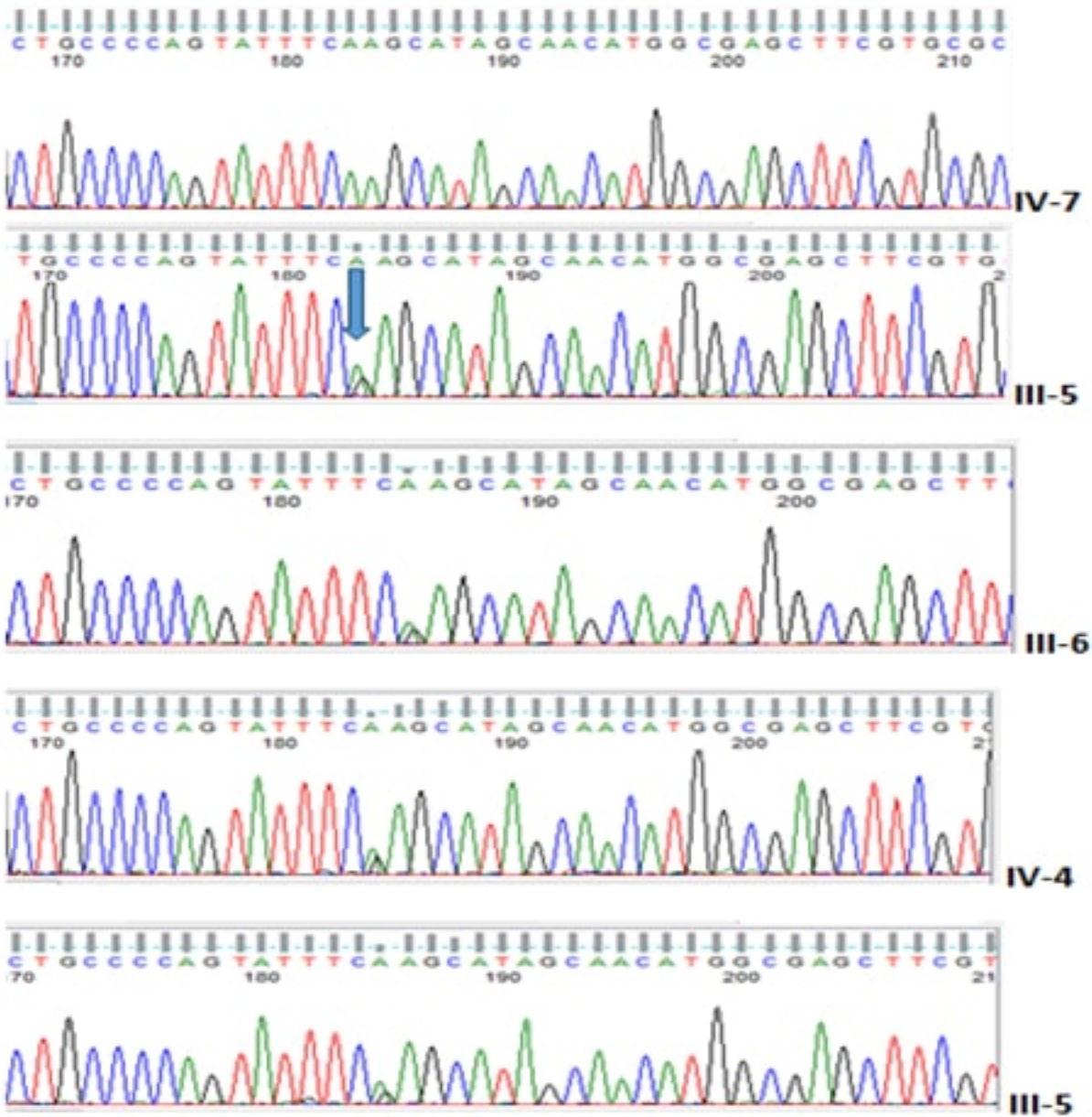
Figure 1: Abridged pedigree of the British family with lamellar cataract. Squares and circles symbolize males and females respectively. Open and filled symbols indicate unaffected and affected individuals.

Figure 2: Sequence analysis of *HSF4.* An unaffected individual (upper chromatogram illustrates a normal control and a missense mutation c.190A>G shown in few affected members of the family with lamellar cataract.

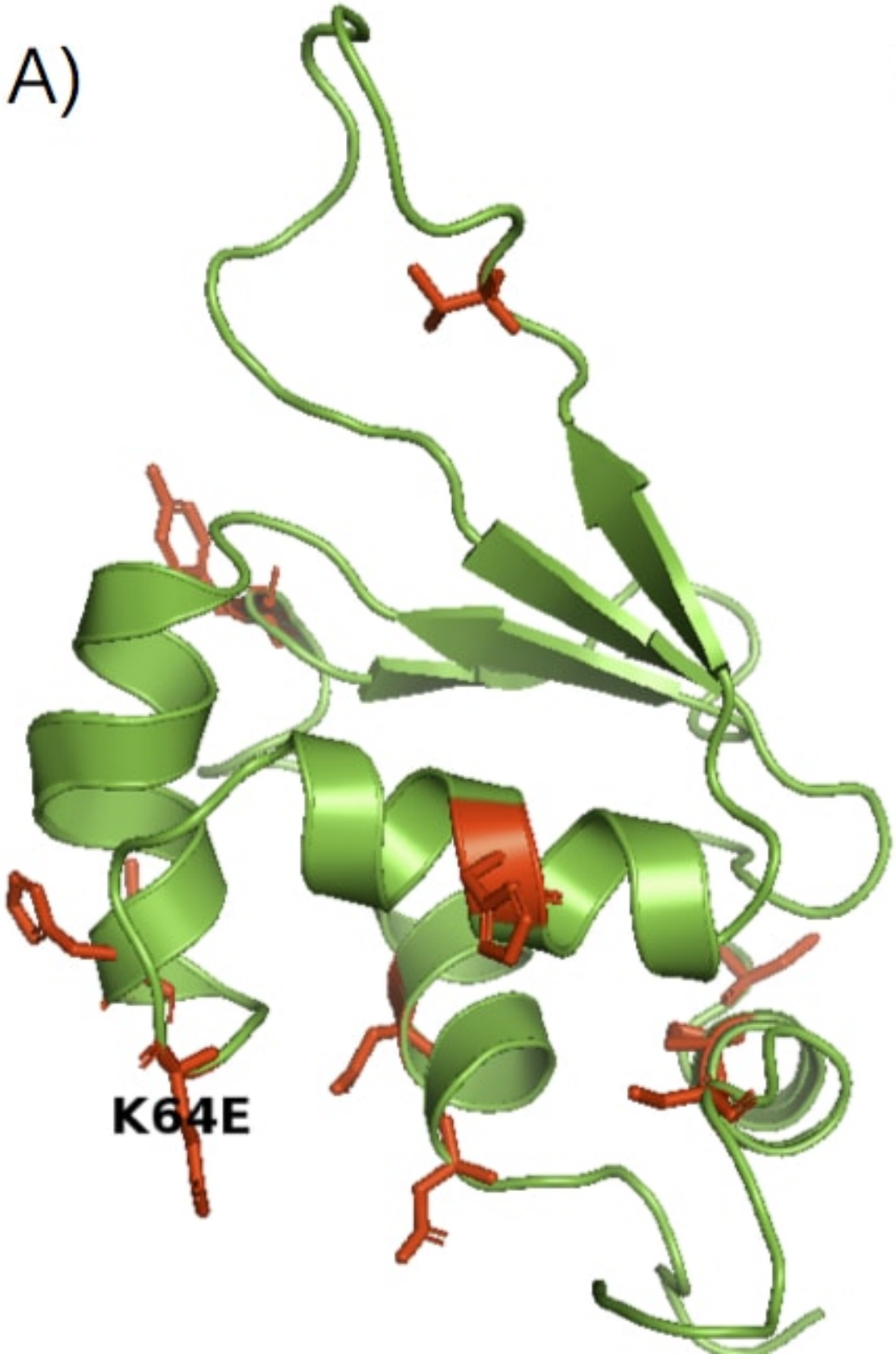
Figure 3: Structure of the DNA binding domain of HSF4. A) Known mutations and their sidechains are displayed in red. The K64E mutation is just downstream of an alpha helix and is a point of contact with the DNA molecule. **B) and C)** Point of contact of the amino acid (red) in the DBD interacting with the DNA molecule (orange and blue). **B)** In the wildtype protein the amino acid is lysine. **C)** In the mutant protein the amino acid is a glutamic acid.

Table 1 : Published mutations in HSF4 that cause cataract. Mutations are ordered by amino acid position. The mutations fall in the following protein domains: DNA-binding-Domain (DBD), hydrophobic repeat (HR/A-B), and downstream of hydrophobic repeat (DHR). Genomic Evolutionary Rate Profiling (GERP) NR corresponds to the neutral rate conservation score of the site. The DBD mutation sites are more conserved than the HR sites. Combined Annotation Dependent Depletion (CADD) is score for the deleteriousness of a variant. A CADD score over 20 is considered damaging.

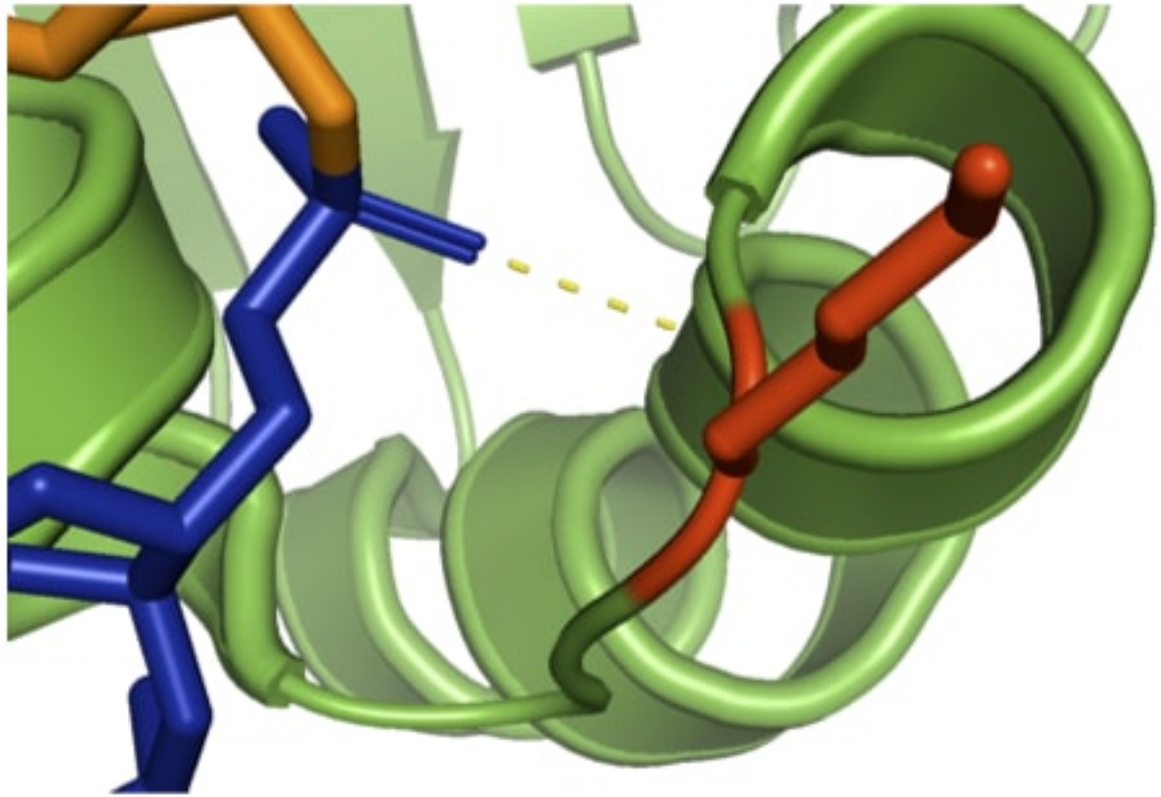

I-1
I-2
II-1
II-2
III-1
III-2
III-3
III-4
III-5
III-6
III-7
IV-1
IV-2
IV-3
IV-4
IV-5
IV-6
IV-7
IV-8
IV-9
V-1
V-2


170
180
190
200
210
IV-7
III-5
III-6
IV-4
III-5

A)



B)



C)

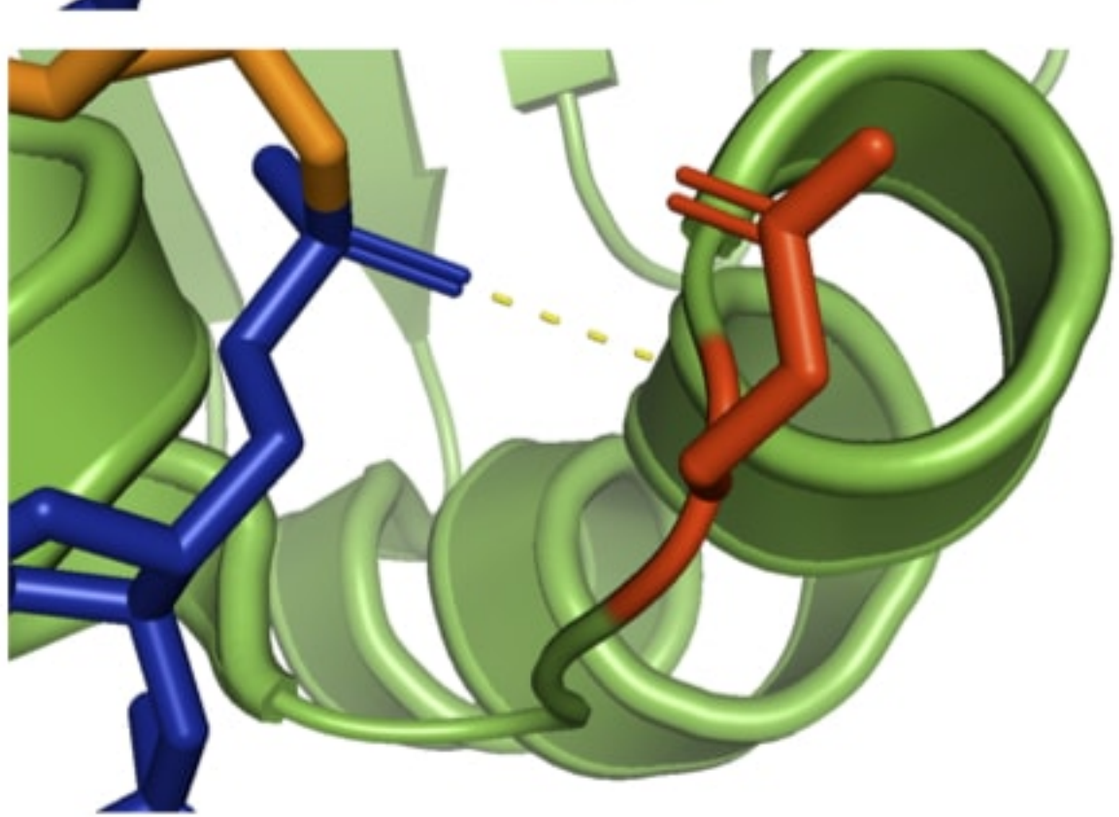


Table 1

	Genomic Position	Exon	HGVSc	HGVSp	Inheritance	Origin	Phenotype	Reference	Protein Domain	GERP NR	CADD	Gnomad Het count
1	16-67198770-C-A	3	c.56C>A	p.A19D	Dominant	China	?	20	DBD	5.22	26.2	0
2	16-67198783-G-T	3	c.69G>T	p.K23N	Dominant	China	Bilateral Cortical	30	DBD	5.31	29.2	0
3	16-67198817-C-T	3	c.103C>T	p.H35Y	-	UK	-	25	DBD	5.32	25.1	2
4	16-67199480-C-A	3	c.179C>A	p.P60H	Sporadic	China	Nuclear	31	DBD	5.37	33	0
5	16-67199483-A-G	4	c.182A>G	p.Q61R	-	China	Age-related cortical	32	DBD	5.37	22.2	0
6	**16-67199491-A-G**	**4**	**c.190A>G**	**p.K64E**	**Dominant**	**UK**	**Lamellar**	**Present Study**	**DBD**	**5.37**	**29.6**	**0**
7	16-67199519-G-A	4	c.218G>A	p.R73H	Dominant	China	Total	33	DBD	5.15	35	0
8	16-67199645-A-G	4	c.256A>G	p.I86V	Dominant	China	Cortical lamellar	20	DBD	4.86	15.4	0
9	16-67199717-C-T	5	c.331C>T	p.R110C	Dominant	China	Congenital	34	DBD	4.86	26.4	0
10	16-67199730-T-C	5	c.341T>C	p.L114P	Dominant	China, Denmark	Lamellar, Lamellar, sutural	20,35	DBD	4.86	31	0
11	16-67199744-C-T	5	c.355C>T	p.R119C	Dominant	China, Denmark	Zonular stellate & anterior polar	20,35	DBD	4.72	35	0
12	16-67200258-T-C	7	c.521T>C	p.L174P	Recessive	Iran	Congenital	24	HR-A/B	4.74	25.5	0
13	16-67200261-G-C	7	c.524G>C	p.R175P	Recessive	Pakistan	Nuclear & cortical	36	HR-A/B	4.74	27.7	0
14	16-67200494-GGGCC-	8	c.595_599del	p.G199EfsX15	Recessive	Pakistan	-	36	HR-A/B	4.43	34	not covered
15	16-67202963-C-T	11	c.1213C>T	p.R405X	Recessive	Pakistan	Congenital/infantile?	22	DHR	4.60	35	1
16	16-67203540-A-G	12	c.1327+4A>G	p.M419GfsX29	Recessive	Tunisia	Total cataract with Nystagmus	23	DHR	4.63	23.8	0