RESEARCH ARTICLE

WILEY | HGVS HUMAN GENOME VARIATION SOCIETY

Severe neurocognitive and growth disorders due to variation in *THOC2*, an essential component of nuclear mRNA export machinery

Raman Kumar[1] (iD) | Alison Gardner[1] | Claire C. Homan[1] | Evelyn Douglas[2] |
Heather Mefford[3] | Dagmar Wieczorek[4,5] | Hermann-Josef Lüdecke[4,5] |
Zornitza Stark[6,7] | Simon Sadedin[6,8] | The Broad CMG[9] | Catherine Bearce Nowak[10] |
Jessica Douglas[10] | Gretchen Parsons[11] | Paul Mark[11] | Lourdes Loidi[12] |
Gail E. Herman[13] | Theresa Mihalic Mosher[13] | Meredith K. Gillespie[14] |
Lauren Brady[15] | Mark Tarnopolsky[15] | Irene Madrigal[16,17] | Jesús Eiris[18] |
Laura Domènech Salgado[19] | Raquel Rabionet[19] | Tim M. Strom[20] | Naoko Ishihara[21] |
Hidehito Inagaki[22] | Hiroki Kurahashi[22] | Tracy Dudding-Byth[23,24] (iD) |
Elizabeth E. Palmer[23,25] (iD) | Michael Field[23] | Jozef Gecz[1,26] (iD)

[1]Adelaide Medical School and the Robinson Research Institute, The University of Adelaide, Adelaide, Australia

[2]Genetics and Molecular Pathology, SA Pathology, Adelaide, Australia

[3]Division of Genetic Medicine, Department of Pediatrics, University of Washington & Seattle Children's Hospital, Seattle, Washington

[4]Heinrich-Heine-University, Medical Faculty, Institute of Human Genetics, Düsseldorf, Germany

[5]Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany

[6]Murdoch Children's Research Institute, Melbourne, Australia

[7]Department of Pediatrics, University of Melbourne, Melbourne, Australia

[8]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts

[9]Broad's Center for Mendelian Genomics, Cambridge, Massachusetts

[10]The Feingold Center for Children at the Department of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts

[11]Spectrum Health Medical Genetics, Grand Rapids, Michigan

[12]Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Spain

[13]Nationwide Children's Hospital and The Ohio State University, Columbus, Ohio

[14]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Canada

[15]Department of Pediatrics, McMaster University Medical Centre, Hamilton, Canada

[16]Biochemistry and Molecular Genetics Department, Hospital Clínic, IDIBAPS, Barcelona, Spain

[17]Centre for Biomedical Research on Rare Diseases (ISCIII), Barcelona, Spain

[18]Unidad de Neurología Pediátrica, Departamento de Pediatría, Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Spain

[19]Centre for Genomic Regulation (CRG), Universitat Pompeu Fabra and CIBERESP, Barcelona Institute for Science and Technology, Barcelona, Spain

[20]Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany

[21]Department of Pediatrics, Fujita Health University School of Medicine, Aichi, Japan

[22]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Aichi, Japan

[23]Genetics of Learning Disability Service, Hunter Genetics, Waratah, NSW, Australia

[24]University of Newcastle, Australia Grow-Up-Well Priority Research Center, Callaghan, Australia

[25]School of Women's and Children's Health, University of New South Wales, Randwick, NSW, Australia

[26]Healthy Mothers, Babies and Children, South Australian Health and Medical Research Institute, Adelaide, SA, Australia

Correspondence
Jozef Gecz, School of Medicine, The Robinson Research Institute, The University of Adelaide, North Adelaide, SA 5005, Australia.
Email: jozef.gecz@adelaide.edu.au

Abstract

Highly conserved TREX-mediated mRNA export is emerging as a key pathway in neuronal development and differentiation. TREX subunit variants cause neurodevelopmental disorders (NDDs) by interfering with mRNA export from the cell nucleus to the cytoplasm. Previously we implicated four missense variants in the X-linked *THOC2* gene in intellectual disability (ID). We now report an additional six affected individuals from five unrelated families with two *de novo* and three maternally inherited pathogenic or likely pathogenic variants in *THOC2* extending the genotypic and phenotypic spectrum. These comprise three rare missense *THOC2* variants that affect evolutionarily conserved amino acid residues and reduce protein stability and two with canonical splice-site *THOC2* variants that result in C-terminally truncated THOC2 proteins. We present detailed clinical assessment and functional studies on a *de novo* variant in a female with an epileptic encephalopathy and discuss an additional four families with rare variants in *THOC2* with supportive evidence for pathogenicity. Severe neurocognitive features, including movement and seizure disorders, were observed in this cohort. Taken together our data show that even subtle alterations to the canonical molecular pathways such as mRNA export, otherwise essential for cellular life, can be compatible with life, but lead to NDDs in humans.

KEYWORDS

mRNA export, partial loss-of-function variants, protein stability, THOC2, XLID

1 | INTRODUCTION

Intellectual disability (ID), characterized by substantial limitations in both intellectual functioning and adaptive behavior, affects 1%–3% of the population starting before the age of 18 years and has significant impact on individuals, families, and communities (Vissers, Gilissen, & Veltman, 2016). Individuals with ID are more likely than members of the general population to experience poor physical and mental health, have a lower life expectancy, experience inequalities accessing health care, and frequently have limited or no specific therapies for their core symptoms (Bittles et al., 2002; Hosking et al., 2016). Both genetic and environmental factors contribute to the development of ID (Milani, Ronzoni, & Esposito, 2015). Over 120 of the identified >800 ID genes are located on the X-chromosome (Chiurazzi & Pirozzi, 2016; Schwartz, 2015), and diagnosis of X-linked causes of ID remain critically important for accurate genetic counseling of families (Ropers & Hamel, 2005). Dramatic improvements in high-throughput DNA sequencing technologies and analyses software has led to identification of new ID genes and additional variants in the known ID genes (Dickinson et al., 2016; Vissers, et al., 2016). A systematic review of clinical data suggests that ID affected individuals frequently have comorbid neurological, psychiatric, and behavioral disorders (Oeseburg, Dijkstra, Groothoff, Reijneveld, & Jansen, 2011; Vissers et al., 2016), and disease variants in different parts of a gene can lead to a broad range of complex neurocognitive disorders (Palmer et al., 2017; Zhu, Need, Petrovski, & Goldstein, 2014). This complexity contributes to heterogeneity in clinical symptoms and indistinct boundaries between syndromic and nonsyndromic forms of neurodevelopmental disorder (NDD).

In 2015, we reported genetic, molecular, and protein structural data on four missense variants in an X-linked essential gene *THOC2* (MIM# 300957; NM_001081550.1; c.937C>T (p.Leu313Phe), c.1313T>C (p.Leu438Pro), c.2399T>C (p.Ile800Thr), and c.3034T>C

(p.Ser1012Pro), RNA not analyzed) (Kumar et al., 2015). The affected individuals had a syndromic NDD, characterized by borderline to severe ID, speech delay, short stature, and adult onset truncal obesity (Kumar et al., 2015). *THOC2* encodes for the THOC2 protein—the largest subunit of the highly conserved TREX (Transcription-Export) mRNA export complex essential for exporting mRNA from the cell nucleus to the cytoplasm (Heath, Viphakone, & Wilson, 2016). The TREX complex is composed of a THO sub-complex (THOC1, THOC2, THOC3, THOC5, THOC6, and THOC7) and accessory proteins (UAP56, UIF, Aly, CIP29, PDIP3, ZC11A, SRRT, Chtop) (Heath et al., 2016). The TREX complex, besides its canonical role in mRNA export in the mammalian cells, has been shown to play critical roles in gene expression, 3′ mRNA processing, stress responses, mitotic progression, and genome stability as well as developmental processes such as pluripotency maintenance and hematopoiesis (Yamazaki et al., 2010). We and others have recently demonstrated that subtle perturbations in mRNA export by gene variants or preferential cytoplasmic aggregation can lead to NDDs (Beaulieu et al., 2013; Coe et al., 2014; Kumar et al., 2015), neurodegeneration (Woerner et al., 2016), or cancer (Chinnam et al., 2014; Hautbergue, 2017; Liu et al., 2015; Viphakone et al., 2015). These alterations can have tissue-specific effects as TREX subunits are shown to have tissue-specific roles; for example, mouse *Thoc5* and *Thoc1* deficiency interferes with the maintenance of hematopoiesis (Guria et al., 2011; Mancini et al., 2010) and testis development (Wang et al., 2009). Taken together, altered TREX function can have diverse molecular and cellular consequences resulting in a range of diseases. Here, we present detailed information on the clinical presentations and functional investigations on an additional eight missense and two splice *THOC2* variants. These data reaffirm and extend our previous findings that *THOC2* variation plays a role in complex neurodevelopmental conditions with the core clinical presentation of ID.

2 | MATERIALS AND METHODS

2.1 | Molecular and cellular studies

RNA extraction, RT-qPCR (primers listed in Supp. Table S3), cyclo-heximide chase, and THOC2 immunofluorescence staining were performed as reported previously (Kumar et al., 2015). Molecular studies on the *THOC2* exon35:c.4450-2A>G variant were performed using blood DNA and skin fibroblasts derived from the affected individual and his heterozygous carrier mother. Genomic DNA or cDNA (generated by reverse transcribing the fibroblast total RNAs using Superscript III reverse transcriptase; Life Technologies, Scoresby, Victoria, Australia) was amplified with KAPA HiFi PCR Kit (Kapa Biosystems, Mulgrave, Victoria, Australia) using hTHOC2-4326F/hTHOC2-4519-R (Supp. Table S3) at 95°C for 3 min, 35 cycles of 98°C–15 sec, 59°C–15 sec, 72°C–30 sec, incubation at 72°C for 10 min, gel purified (MinElute Gel Extraction kit; Qiagen, Chadstone Centre, Victoria, Australia) and Sanger sequenced using the same primers. For the *THOC2* exon28:c.3503+4A>C, blood gDNA from unaffected father, carrier mother, and affected son was amplified with TaKaRa ExTaq (Clayton, Victoria, Australia) using THOC2-F/THOC2-R primers (Supp. Table S3) at 94°C for 2 min, 40 cycles of 94°C–30 sec, 60°C–30 sec, 72°C–30 sec, incubation at 72°C for 5 min. The cDNA was generated by reverse transcribing the white blood cell RNAs using Superscript III reverse transcriptase (Life Technologies) and amplified with TaKaRa ExTaq using THOC2-ex27F/THOC2-ex30R (Supp. Table S3) at 94°C for 2 min, 28 cycles of 94°C–30 sec, 60°C–30 sec, 72°C–30 sec, incubation at 72°C for 5 min. The amplified products were analyzed by Sanger sequencing.

2.2 | Generation of *THOC2* variant expression constructs

Generation of the wild-type Myc-tagged human THOC2 expression plasmid was reported earlier (Kumar et al., 2015). Briefly, the *THOC2* variants were introduced into the existing pCMV-Myc-THOC2 expression construct by overlap PCR method using the primers listed in Supp. Table S3. The variant plasmid sequences were confirmed by Sanger sequencing. Details relating generation of the *THOC2* variant expression constructs are available on request.

2.3 | Transient expression and Western blotting

For transient expression experiments, HEK293T and HeLa cells were transfected with expression constructs (400 ng pCMV-Myc-THOC2 plasmid and 400 ng pEGFP-C1 plasmid/transfection for stability and cycloheximide assays and 4 μg/transfection for immunofluorescence staining, IF) using Lipofectamine 3000 reagent according to manufacturer's protocol (Life Technologies). Twenty-four hours post-transfection, cells were either fixed with 4% formaldehyde for IF or collected and lysed in buffer containing 50 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.1% Triton-X 100, 1 mM EDTA, 50 mM NaF, 1× Protease inhibitor/no EDTA and 0.1 mM Na_3VO_4 for Western blot assay as reported previously (Kumar et al., 2015).

2.4 | *In silico* pathogenicity prediction

We used CADD v1.3 (includes Phylop, GERP++ & PolyPhen2) (Kircher et al., 2014), Provean (Choi & Chan, 2015), and ACMG (Richards et al., 2015) on-line tools for *in silico* prediction of the pathogenicity of different variants (Table 1 and Supp. Table S1).

3 | RESULTS

3.1 | Identification of *THOC2* variants

We previously implicated four missense *THOC2* variants in 25 individuals with ID and a range of other clinical features (Table 1 and Supp. Table S1 and 2) (Kumar et al., 2015). We identified an additional five *THOC2* variants (three missense; de novo c.2087C>T (p.Thr696Ile), de novo c.2138G>A (p.Gly713Asp), maternally inherited c.3559C>T (p.His1187Tyr), and two splicing-defective; maternally inherited chrX:122747561 exon35:c.4450-2A>G and chrX:122757634 exon28:c.3503+4A>C; GenBank: NM_001081550.1) variants in a further six affected individuals, including one pair of monozygotic twins (Table 1 and Figures 1 and 2). Whole exome (WES) or whole genome sequencing (WGS) of probands and parents was used to identify the variants that were confirmed by Sanger sequencing of the PCR amplified variant-carrying region of genomic DNA of the parents and affected individuals. The previously unreported *THOC2* variants affect amino acids that are highly conserved (Supp. Figure S1), are absent in gnomAD database and are predicted to be pathogenic based on a number of *in silico* analyses tools (Table 1). We included in our study a de novo missense p.Tyr517Cys variant in a female with moderate-severe ID, speech problems, epileptic encephalopathy, cortical visual impairment, and gait disturbances identified using WES as part of the Epi4K Consortium & Epilepsy Phenome/Genome Project (Epi et al., 2013) (Table 1). We have also collected further, rare and potentially pathogenic variants through international collaboration (Table 1, Supp. Table S1 and 2, and Figures 1 and 3; see Supplementary information for methods used for identifying the variants) and performed functional testing on several of these. The following three variants: c.229C>T (p.Arg77Cys), c.3034T>C (p.Ser1012Pro), and c.3781A>C (p.Asn1261His) showed no clear evidence of altered stability of variant THOC2 proteins in our cell-based assay. The reported variants have been submitted to ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/accession numbers SCV00680065-SCV000680074).

3.2 | Clinical presentations

The clinical features of the five previously unreported affected individuals with (likely) pathogenic *THOC2* variants, aged between 3 and 12 years, and the 10 year old female with de novo p.Tyr517Cys variant are summarized in Table 2 and photographs, when available, are shown in Figure 2. Detailed clinical information is available in the supplementary data. Each individual clinical center used local diagnostic criteria for determining degree of ID and diagnoses of comorbidities. ID was universal and at least moderate in severity: two out of seven were nonambulatory and three out of seven nonverbal. Behavioral problems

TABLE 1 Detailed description of the *THOC2* variants with supporting molecular evidence

Individual	From	Method of identification	Position hg19	NM_001081550	Mode of inheritance	CADD	Provean score	Provean prediction	GERP++	Phylop	gnomAD frequency	Polyphen2[a]	Variant plasmid tested	Reduced Protein Stability	ACMG pathogenicity classification[b]
Missense															
	Kumar et al. (2015) AJHG	X-Chr Exome sequencing	122799566	c.1313T>C;p.Leu438Pro	Maternal inheritance	28.1	−6.08	Deleterious	5.7	1.902	Absent	D	YES	YES	LP
1	Australia	Trio WES	122767853	c.2087C>T;p.Thr696Ile	De novo	27.4	−5.47	Deleterious	5.03	0.963	Absent	D	YES	YES	DP
2	USA	WES	122766890	c.2138G>A;p.Gly713Asp	De novo	31	−4.69	Deleterious	5.73	2.412	Absent	D	YES	YES	DP
	Kumar et al. (2015) AJHG	X-Chr Exome sequencing	122765621	c.2399T>C;p.Ile800Thr	Maternal inheritance	23.9	−4.01	Deleterious	5.97	2.016	Absent	P	YES	YES	LP
3-4	Canada/Germany/Russia Identical twins	WES	122757079	c.3559C>T;p.His1187Tyr	Maternal inheritance; Mother skewed (99.9:0.1%)	23.1	−5.07	Deleterious	6.07	2.571	Absent	P	YES	YES	LP
Splice															
5	Japan	Trio WES	122757634	Exon28:c.3503+4A>C	Maternal inheritance; Mother skewed 98:2%	10.8	N/A	N/A	5.57	1.86	Absent	N/A	N/A	ND	LP
6	Canada	WES	122747561	Exon35:c.4450-2A>G	Maternal inheritance; Mother skewed 94:6%	23.7	N/A	N/A	5.25	1.735	Absent	N/A	Fibroblasts of the affected male proband and carrier mother	NO	LP
Missense															
7	USA Epi4K Consortium & Epilepsy Phenome/Genome Project; Nature 501:217–221, 2013	WES	122778639	c.1550A>G;p.Tyr517Cys	De novo	26.6	−7.87	Deleterious	5.84	1.955	Absent	D	YES	YES	DP

[a] D, probably damaging; P, possibly damaging; N/A, not applicable; ND, not determined.
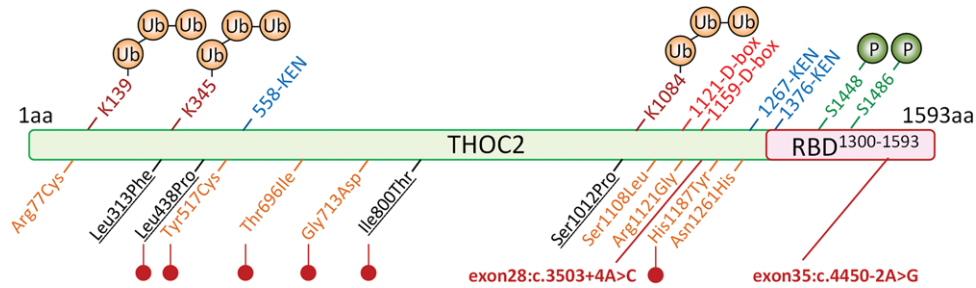[b] DP, de novo pathogenic; LP, likely pathogenic.

FIGURE 1 Location of variant amino acids and structural features in THOC2 protein. Ubiquitinated (K139 and K1084 (Kim et al., 2011) and K345 (Lopitz-Otsoa et al., 2012; Wagner et al., 2011)), phosphorylated (S1448 and 1486 (Olsen et al., 2006)) amino acid residues, potential RNA binding domain (RBD) and destruction box (D-box) and KEN box sequences that interact with the Anaphase Promoting Complex/Cyclosome (APC/C) for protein ubiquitination and subsequent destruction by the proteasome (Morgan, 2013) are shown. Unreported (orange) and published (black: (Kumar et al., 2015)) missense variants effecting THOC2 protein stability are marked with red lollipops. The positions of two splice variants are shown in red
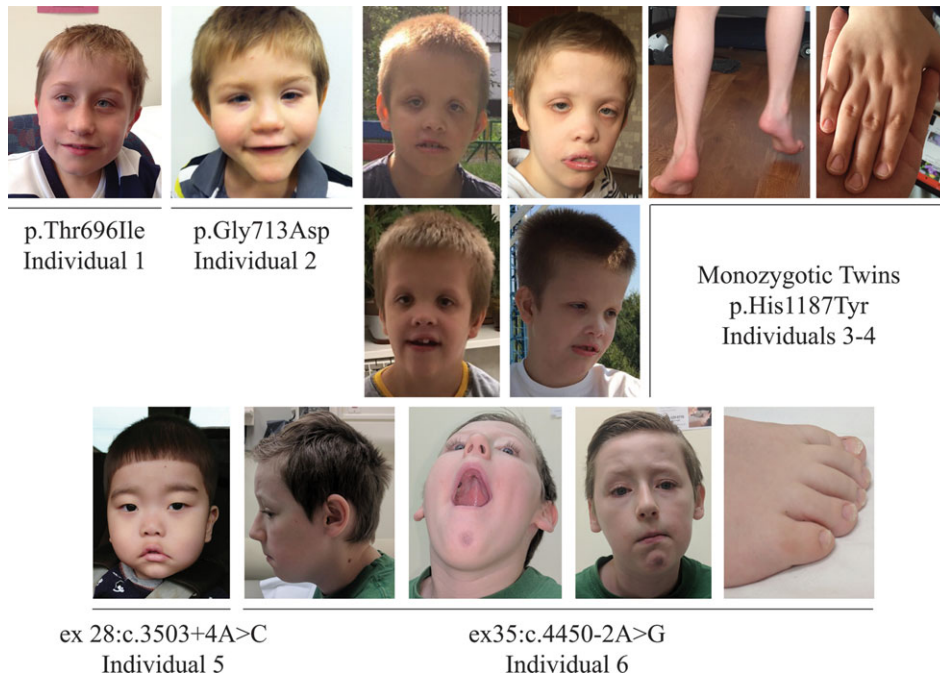

FIGURE 2 Front and side facial views of the affected individuals with *THOC2* variants

were reported in four individuals, with one meeting diagnostic criteria for autism spectrum disorder. Additional neurological features were common. Four of the seven had infantile hypotonia and two of the seven had tremor. The monozygotic twins (individuals 3 and 4) had a tendency to toe walking, which was considered behavioral as it was not associated with neurological signs of lower limb spasticity. Confirmed seizure disorder was only present in the affected female (individual 7) but suspected in individual 2. Neuroradiological studies were performed in five individuals: this was within normal limits for three individuals, with neuroanatomical differences reported in two. Individual 2 had complex neuroanatomical findings (see supplementary clinical description and Supp. Figure S2A) including changes in cortical gyral morphology, which in the inferior temporal lobes appeared finely nodular, as well as hypoplasia of the corpus callosum and reduced brainstem volume and individual 5 had mild dilatation of the lateral ventricles, mildly delayed myelination and an abnormal white matter lesion in the periventricular area close to the anterior horn (Supp. Figure S2B). Growth abnormalities were common

including low birth weight (three out of seven), microcephaly (two out of seven), and short stature (two out of seven). Facial features are shown in Figure 2. Appropriate consent for reporting variants, clinical data and photographs of the affected individuals was obtained from their parents or legal guardians. The research has been approved by the Women's and Children's Health Network Human Research Ethics Committee in Adelaide, Australia.

3.3 | THOC2 variant protein localization and stability

Without access to affected individuals' derived cells, we generated Myc-tagged THOC2 missense variant expression constructs to determine protein stability and localization. The THOC2 protein stability was determined in HEK293T cells and localization in both the HEK293T and HeLa cells. Total protein lysates of HEK293T cells ectopically-expressing the wild-type or variant Myc-THOC2 proteins were Western blotted for THOC2, EGFP, and β-Tubulin. We used HEK293T cells expressing Myc-p.Ile800Thr THOC2 as a control for
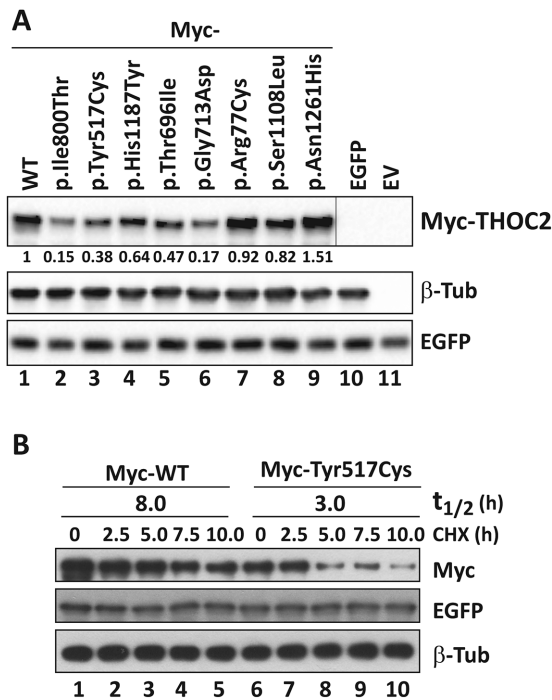
A

FIGURE 3 Functional testing of THOC2 missense variants. **A**: THOC2 variant protein stability is reduced in HEK293T cells. pCMV-Myc-THOC2 wild-type or variant expression constructs and pEGFP-C1 plasmid (transfection control) were transfected into HEK293T cells. Total protein lysates of cells 24 hr post-transfection were analyzed by Western blotting with mouse anti-Myc (clone 9E10; Sigma), mouse anti-EGFP (clones 7.1 and 13.1; Roche) and rabbit anti-β-tubulin (loading control; Abcam) antibodies. pCMV-Myc-THOC2 p.Ile800Thr construct expressing the p.Ile800Thr protein shown to have reduced stability was used as a control (Kumar et al., 2015). Western blot signals were quantified using ImageJ software. Averages of the Myc-THOC2 proteins normalized to the housekeeping β-tubulin signal from two independent runs are shown. **B**: Myc-p.Tyr517Cys THOC2 protein half-life is substantially reduced in HEK293T cells. pCMV-Myc-THOC2 or pCMV-Myc-THOC2-p.Tyr517Cys expression constructs and pEGFP-C1 plasmid (transfection control) were transfected into HEK293T cells. Next day the cells were cultured in the presence of 100 μg/ml translation inhibitor cycloheximide and harvested at the time points shown. Total protein lysates were analyzed by Western blotting with mouse anti-Myc, mouse anti-EGFP, and rabbit anti-β-tubulin (loading control) antibodies

protein stability assay as this variant is shown to cause reduced protein stability (Kumar et al., 2015). The results showed reduced stability of p.Tyr517Cys, p.Thr696Ile, p.Gly713Asp, and p.His1187Tyr THOC2 compared with the wild-type protein (Figure 3A). Presence of comparable levels of EGFP in the cells transfected with different expression constructs indicated that the reduced levels of THOC2 protein were not due to difference in transfection efficiency (Figure 3A). We also determined the turnover rate of Myc-p.Tyr517Cys THOC2 protein by cycloheximide chase. For this assay, the HEK293T cells transfected with pCMV-Myc-WT or pCMV-Myc-p.Tyr517Cys THOC2 and pEGFP-C1 transfection control plasmids were cultured in presence of translation inhibitor cycloheximide for different durations and Western blotted for THOC2, EGFP, and β-Tubulin. The results showed that p.Tyr517Cys THOC2 turnover rate was 3 hr compared with 8 hr

for the wild-type protein (Figure 3B). THOC2 variant proteins, similar to the wild-type, were mainly localized to the nucleus in both the HEK293T and HeLa cells (Supp. Figure S3).

3.4 | THOC2 splice variant: exon35:c.4450-2A>G, p.Arg1483fs52*

Sanger sequencing of amplified target region from affected son and mother's blood genomic DNA showed that the affected boy inherited chrX:122747561 exon35:c.4450-2A>G variant from his unaffected heterozygous carrier mother (Figure 4C). A -2 A>G change in the intron–exon splicing site boundary (acceptor AG) is predicted to abolish splicing (Ohno, Takeda, & Masuda, 2018). To validate this possibility, we generated skin fibroblast cultures from the heterozygous carrier mother and the affected son. We PCR amplified their fibroblast cDNAs using primers with binding sequences located within exon 34 and 35. Amplification of a 194 bp DNA fragment from the mother indicated normal splicing but a 537 bp product from the affected son indicated retention of the intron located between these exons (Figure 4A and B). We confirmed this result by Sanger sequencing of the PCR products generated from genomic DNA that showed presence of A/G nucleotides in the carrier mother but only G (A>G) nucleotide in the affected son (Figure 4C). The cDNA sequence showed presence of normally-spliced mRNA in the mother but retention of intronic sequence upstream of the exon 35 in the affected son indicating defective splicing due to presence of -2 G variant at the intron-exon 35 junction sequence (Figure 4C). The presence of normally spliced mRNA in the unaffected mother is consistent with X-inactivation (94% skewing) of the variant allele in her fibroblasts. We predicted that a retention of intron between exon 34–35 in the affected fibroblasts would result in loss of 110 C-terminal amino acids of the 1,593 wild-type THOC2 protein (that is, 1,483 amino acids); however, overall the variant protein would be 58 amino acid smaller as it would now be a 1,535 amino acid protein comprised of 1,483 amino acids of wild-type THOC2 and 52 translated from intronic sequence in the defective mRNA (Figure 4A). Consistent with our prediction, the Western blot data showed presence of a slightly smaller THOC2 protein band in the affected son's fibroblasts than his unaffected mother. Many independent Western blot runs showed presence of two closely located THOC2 bands—similar to the fruit fly THO2 (Rehwinkel et al., 2004)—in the unaffected mother but a single highly intense band in the affected son's fibroblasts (Figure 4D). The observed difference in levels of THOC2 protein was post-translational as we found comparable amounts of THOC2 mRNA, as assayed by real time RT-qPCR, in the mother and son (Figure. 4E). Finally, we observed no difference in THOC2 localization in fibroblasts of the affected son and his unaffected mother (Figure 4F).

3.5 | THOC2 splice variant: exon28:c.3503+4A>C, p.Gly1168fs7*, and normal 1,593 aa protein

For the second splice variant chrX:122757634 exon28:c.3503+4A>C, molecular studies were performed on the white blood cells of the unaffected father, carrier mother, and the male proband. Sanger sequencing of target region amplified from the unaffected father and

TABLE 2 Summary of clinical data of *THOC2* variants with supporting molecular evidence

	(Likely) pathogenic						
Individual	1	2	3	4	5	6	7
Variant details	c.2087C>T: p.Thr696Ile	c.2138G>A: p.Gly713Asp	c.3559C>T: p.His1187Tyr Twin 1	c.3559C>T: p.His1187Tyr Twin 2	Exon28: c.3503+4A>C	Exon35:c.4450-2A>G	c.1550A>G: p.Tyr517Cys
Gender	Male	Male	Male	Male	Male	Male	Female
Age (years)	12	5	7	7	3	10	10
Perinatal features							
Gestation (weeks)	36	37	37	37	37	41	NA
Low birth weight (<2.5kg)	Yes	No	Yes	Yes	No	No	NA
Birth weight (g)	2,000	2,650	1,990	2,420	3,018	4,365	NA
Neurologic features							
Intellectual disability	Severe	Mod+	Mod+	Mod+	Severe	Severe	Profound
Speech delay	Yes, single words, signs	Yes, nonverbal	Yes	Yes	Yes, nonverbal	Yes, nasal dysarthria	Yes, non-verbal
Hypotonia	No	Yes	NA	NA	Yes	Yes, central hypotonia	Yes
Spasticity	No	No	No	No	No	Yes- appendicular spasticity	No
Hyperkinesia	No	No	Yes	Yes	Yes	No	No
Tremor	No	Yes, intermittent	No	No	No	Yes	No
Epilepsy	No	Suspected	No	No	No	No	Yes, epileptic encephalopathy
Gait disturbances	No	Yes, gait/balance problems	Yes, toe walking[a]	Yes, toe walking[a]	Non ambulatory	Yes, ataxia/broad based gait	Non ambulatory
Behavior problems	No	Yes	Yes	Yes	Yes, ASD	No	NR
Anxiety	No	No	No	No	No	No	NR
Depression	No	No	No	No	No	No	NR
Brain MRI/CT	MRI normal	Thin corpus callosum, low brainstem volume, variability in gyral pattern.	ND	ND	Ventricular dilatation, delayed myelination, periventricular white matter lesion	MRI within normal limits	MRI normal
Growth parameters							
Microcephaly (≤3%)	Yes	Yes, <1%	No	No	No	Yes,2%	No, 5%
Short stature (≤3%)	Yes	Yes	No	No	No	No	No
Overweight (BMI≥25)	No	No	No	No	No	No	No
Broad high forehead	Yes	Yes	Yes	Yes	No	No	NR
Other features		Mild joint laxity, subluxed hips, disordered sleep, feeding difficulties (g-tube dependency), laryngomalacia, micrognathia, abnormal palmar creases	Noonan facies, pes planus, hypospadias	Noonan facies, pes planus, hypospadias		Clinodactyly, nystagmus, abnormality soft palate	Cortical visual impairment

Abbreviations: %, centile; ASD, autism spectrum disorder; CT, computerized tomography scan; g-tube, gastrostomy tube; mod+, at least moderate severity; MRI, magnetic resonance imaging; ND, not done; NA, not available; NR, not reported; NICU, neonatal intensive care unit; VOUS, variant of uncertain significance.

[a]Toe walking in absence of neurological signs of lower limb spasticity, therefore considered a behavioral manifestation.
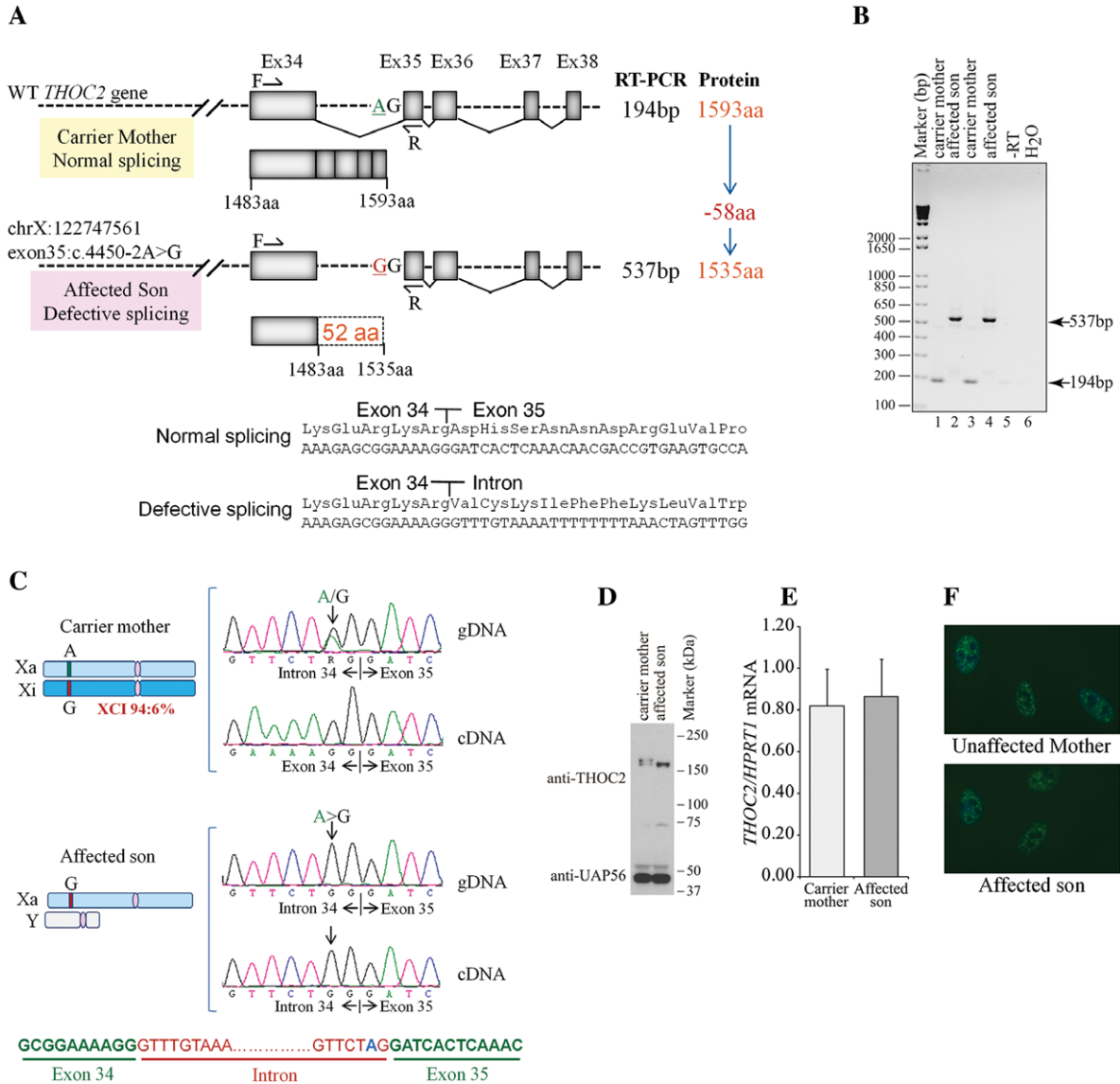
FIGURE 4 Exon35:c.4450-2A>G variant abolishes splicing of intron between exons 34–35. **A**: Part of the *THOC2* gene showing location of the A/G nucleotide in the heterozygous carrier mother and A>G splice variant in the affected son. The C-terminal part of the 1,593 amino acid wild-type and 1,535 amino acid (that contains 1,483 normal and 52 amino acids coded by the unspliced intron) THOC2 protein in the affected boy are also shown. **B**: Gel showing a 194 bp RT-PCR product from the normally spliced mRNA of the heterozygous carrier mother and a 537 bp product from defective splicing of mRNA causing retention of an intron between exon 34–35 in the affected son. RT-PCR products from total RNA isolated from passage 3 (lanes 1–2) and 5 (lanes 3–4) fibroblasts. Location of the forward and reverse primers within exons 34 and 35 is shown. **C**: Sanger sequencing chromatograms of PCR products amplified from genomic and cDNA of the affected son and his heterozygous carrier mother using primers located within exon 34 and 35. Genomic DNA around the Exon34-Intron-Exon35 region is shown. **D**: Western blot showing THOC2 protein in the affected son and his carrier mother's skin fibroblasts. TREX subunit UAP56 was used as a loading control. **E**: RT-qPCR showing levels of the THOC2 mRNA in the affected son and his carrier mother's skin fibroblasts. **F**: Immunofluorescence detection of THOC2 in skin fibroblasts of the unaffected mother and affected son

mother, and affected son's genomic DNA showed that the affected son inherited the A>C change from his unaffected carrier mother who had A/C nucleotides at this position (Figure 5). The intronic nucleotide change A>C at +4 position of the 5′ exon–intron donor splicing site sequence is predicted to cause aberrant splicing (https://www.med.nagoya-u.ac.jp/neurogenetics/SD_Score/sd_score.html). To confirm this possibility, we amplified cDNA generated by reverse transcribing blood RNA of the father, mother, and the affected son using primers located within exon 27 and exon 30 (Figure 5A and Supp.

Table S3). Interestingly, whereas a 491 bp PCR product was observed in highly skewed carrier mother (98:2%) and normal father, 491 and 634 bp PCR products were detected in the affected son. A 491 bp amplified product indicated normal splicing in the mother and father, and 491 and 634 bp bands suggested partially defective splicing in the affected son. Amplification of a 634 bp instead of a 994 bp fragment that would have resulted from a complete retention of intron between exon 28–29 indicated aberrant splicing event in the affected son (Figure 5). Sanger sequencing of 491 and 634 bp PCR products from
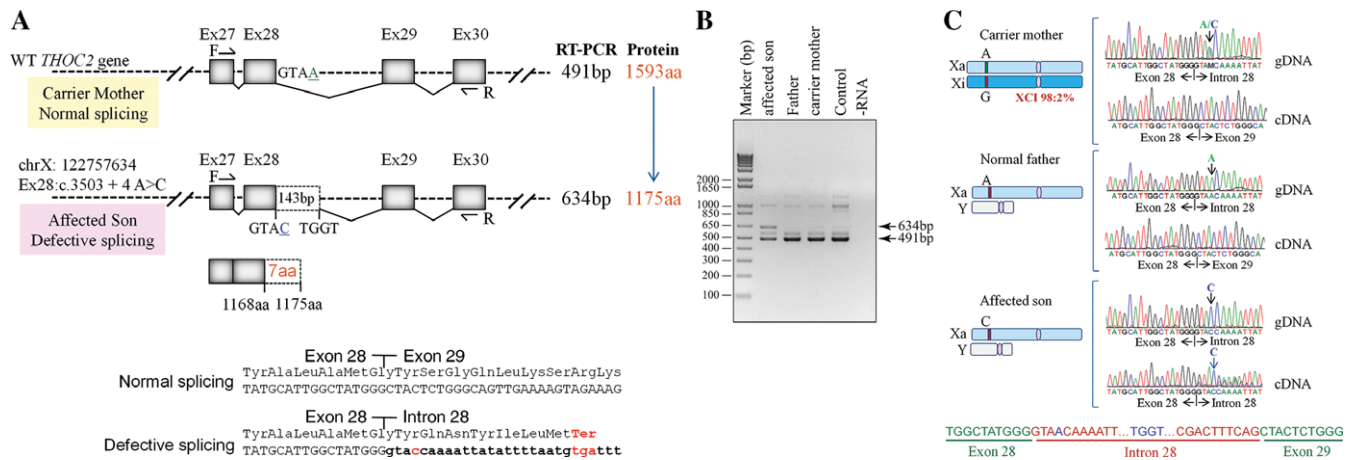
FIGURE 5 Exon28:c.3503+4A>C variant causes aberrant splicing of intron between exons 28–29. **A**: Part of the *THOC2* gene showing location of the A/C nucleotide in the heterozygous carrier mother and A>C splice variant in the affected son. The C-terminal part of the 1,593 amino acid wild-type and 1,175 amino acid (that contains 1,168 normal and 7 amino acids coded by the unspliced intron) THOC2 protein in the affected son are also shown. **B**: Gel showing a 491 bp RT- PCR product from the normally-spliced heterozygous carrier mother and unaffected father, and 491 and 634 bp (retaining 143 bp of the 503 bp intron between exons 28–29) RT-PCR products derived from the normally and aberrantly spliced mRNAs, respectively, in the affected son. Location of the forward and reverse primers within exons 27 and 30 is shown. **C**: Sanger sequencing chromatograms of PCR products amplified using primers located within exons 27 and 30 from genomic and cDNA of unaffected father and mother and the affected son

the mother, father and son confirmed normal splicing in the mother and father and aberrant splicing in the affected son. The sequence showed retention of a 143 bp instead of complete 503 bp fragment due to activation of a cryptic splice site within the intron between exon 28–29 in the son (Figure 5). Retention of 143 bps from intron between exon 28–29 in the mRNA is predicted to result in a 1175 amino acid truncated THOC2 protein containing 1,168 wild-type amino acids and 7 amino acids from the translation of the intronic sequence retained in the defective mRNA. This aberrant product would be present in addition to the wild-type 1,593 amino acid protein from the normally spliced mRNA in the affected son.

4 | DISCUSSION

Here, we present detailed clinical information, and molecular and functional studies, on five previously unreported *THOC2* variants in six affected males (two *de novo* variants and one maternally inherited variant in monozygotic twins) and on one affected female with a previously reported *de novo* p.Tyr517Cys variant. We present evidence that extends the genotypic spectrum beyond the four *THOC2* missense variants that we reported previously (Kumar et al., 2015) by including two intronic variants that affect splicing, and four missense variants that affect protein stability in a cell-based assay system. According to ACMG criteria, they were classified as pathogenic or likely pathogenic (Table 1) (Richards et al., 2015). These findings, along with the four missense variants reported earlier (Kumar et al., 2015), add to the existing evidence that alterations in essential mRNA export pathway cause NDDs (Amos et al., 2017; Beaulieu et al., 2013; Kumar et al., 2015).

We confirm that the core clinical feature of *THOC2*-related disorder in hemizygous males is ID, with several individuals having additional features including behavioral disorders, hypotonia, gait disturbance,

tremor, low birth weight, short stature, microcephaly, and variable neuroimaging findings. Although the range of neurodevelopmental features is similar, our original cohort contained males with ID in the mild or borderline range of intellectual functioning (Kumar et al., 2015), whereas all individuals in this cohort have ID which is at least in the moderate range. Individuals 2 and 6 had neurological signs that could be consistent with cerebellar dysfunction including tremor and a broad-based gait for individual 2 and nystagmus, tremor, and an ataxic broad-based gait for individual 6, in the absence of significant cerebellar abnormalities on MRI. This is interesting given the female patient with knockdown of THOC2 function due to a *de novo* X;8 translocation that created a *PTK2-THOC2* fusion had congenital cerebellar hypoplasia and prominent cerebellar signs with mild ID (Di Gregorio et al., 2013). We used computerized face-matching technology to specifically evaluate the cohort to assess if a characteristic facial gestalt was evident across individuals with pathogenic or likely pathogenic variants across our original and this expanded clinical cohort (Supp. Figure S4) (Dudding-Byth et al., 2017). Although a clearly recognizable facial gestalt was not obvious, there are some similarities. The facial gestalt spectrum associated with *THOC2* pathogenic variants will continue to emerge as more individuals are reported.

As was the case in our original cohort, heterozygous mothers were clinically unaffected, and, where available, X-chromosome inactivation (XCI) was highly skewed (Table 1). In contrast individual 7, with a *de novo* missense variant (p.Tyr517Cys) is a female with a particularly severe neurocognitive presentation. This is consistent with other reported severely affected females with *de novo* variants in X linked genes (de Lange et al., 2016; Palmer et al., 2016; Snijders Blok et al., 2015; Zweier et al., 2014). Unfortunately, we did not have access to individual 7's genomic DNA to test XCI status.

A range of protein–protein interactions are required for mRNA export (Chi et al., 2013). Proteins with altered stability (Hirayama et al.,

2008), localization (Beaulieu et al., 2013) (e.g., THOC6 p.Gly46Arg implicated in syndromic ID), or interaction (Chi et al., 2013) can impact mRNA export and consequently disrupt normal cell function. We did not observe mislocalization of the THOC2 variant proteins in cultured cells and did not test alterations in their interaction with the other known or unknown TREX proteins. However, reduced levels of a number of new (p.Tyr517Cys, p.His1187Tyr, p.Thr696Ile, p.Gly713Asp) and published (p.Leu438Pro, p.Ile800Thr; Kumar et al., 2015) missense THOC2 variant proteins are due to impaired protein stability or reduced levels of normal mRNA due to aberrant splicing (exon28:c.3503+4A>C). We also noted increased stability of p.Asn1261His THOC2 protein. We and others have shown that THOC2 controls TREX function by maintaining the stability of THOC1, 3, 5, and 7 subunits (Chi et al., 2013; Kumar et al., 2015). Reduced levels of THOC2 missense variant proteins are most likely due to enhanced proteasome-mediated degradation as THOC2 is ubiquitinated (Lopitz-Otsoa et al., 2012). THOC2 depletion has been reported to have different consequences in diverse organisms. For example, shRNA-mediated Thoc2 knockdown leads to significant increase in length of neurites in cultured rat primary hippocampal neurons (Di Gregorio et al., 2013) although effects on neurons with persistently reduced THOC2 variant proteins in the affected individuals may be different and *Caenorhabditis elegans thoc2* knockouts, that are completely immobile, slow-growing, sterile, have functional defects in specific sensory neurons and die prematurely (Di Gregorio et al., 2013). *Danio rerio Thoc2* is essential for embryonic development (Amsterdam et al., 2004) and in *Drosophila melanogaster* S2 cells *Thoc2* knockdown inhibits mRNA export and cell proliferation (Rehwinkel et al., 2004). THOC2 depletion also results in chromosome alignment, mitotic progression, and genomic stability in human HeLa cells (Yamazaki et al., 2010). Finally, *Thoc2* and *Thoc5* knockdown experiments have shown their role in regulation of embryonic stem cell (ESC) self-renewal (Wang et al., 2013).

Both the affected individuals carrying the splice-variants presented with severe neurocognitive features. The exon35:c.4450-2A>G and exon28:c.3503+4A>C THOC2 splice variants present interesting biological scenarios; the former resulting in a 1,535 amino acid truncated protein that is present at higher level and the latter with both normal (albeit potentially much reduced) and a 1,175 amino acid truncated THOC2 protein. We postulate that the clinical outcomes in the exon35:c.4450-2A>G individual are caused by partial loss of function due to loss of 110 amino acid C-terminal region and accumulation of the truncated THOC2 protein. However, pathogenicity in exon28:c.3503+4A>C affected individual is most likely caused by reduced levels of normal and potential dominant-negative effects of the C-terminally truncated THOC2 protein. That reduced THOC2 protein levels are associated with ID and other clinical symptoms is emerging as a frequent theme; for example, due to reduced THOC2 protein stability caused by missense variants (see above) or aberrant splicing. Indeed reduced THOC2 levels are shown to destabilize the TREX complex in humans (Chi et al., 2013; Kumar et al., 2015) and removal of any THO subunit causes destabilization of other TREX components in yeast (Pena et al., 2012).

Systematic functional analysis of the Tho2 C-terminal RNA binding region in yeast provides interesting explanation as to how the truncated THOC2 protein can perturb normal mRNA export function in human cells (Pena et al., 2012) (Figure 6). The data showed that whereas ΔTho2 yeast strain does not grow at 37°C (restrictive temperature), Tho2$_{1,408-1,597}$ and ThoΔ$_{1,271-1,597}$ growth is considerably reduced, suggesting that C-terminal 1,271–1,597 amino acids are required for cell survival at restrictive temperature (Pena et al., 2012). If the exon28:c.3503+A>C variant caused complete splicing defect retaining intron between exon 28–29 in all mRNAs, the cells would translate only 1,175 amino acids (with 1,168 normal) THOC2 protein; essentially lacking the C-terminal region encompassing the RNA binding domain (RBD) that when deleted in yeast ThoΔ$_{1,271-1,597}$ strain restricts its growth at 37°C. However, the affected boy carrying a single allele of the exon28:c.3503+A>C THOC2 variant, although with severe clinical symptoms, is alive. This could be explained by presence of reduced levels of THOC2 protein produced from translation of about 2/3rd normally spliced mRNA in the affected white blood cells. Taken together, clinical outcomes in the affected boy may be due to perturbed mRNA export caused by reduced levels of THOC2 protein and perhaps also C-terminally truncated THOC2 protein translated from about 1/3rd aberrantly spliced mRNAs that retain a part of intron sequence between exon 28–29.

We also identified a set of previously unreported THOC2 missense variants that, according to ACMG criteria, are variants of uncertain clinical significance (VOUS) (Supp. Table S1): namely, p.Arg77Cys, p.Ser1108Leu, p.Arg1121Gly, and p.Asn1261His. Nevertheless, these variants have supportive evidence pointing toward potential pathogenicity as they are rare (absent from ExAC/gnomAD databases of reference individuals) (Lek et al., 2016), affect highly evolutionarily conserved amino acid residues, are predicted to be pathogenic by *in silico* analyses and are within the clinical presentations spectrum of those seen in individuals with confirmed THOC2-related ID. However, they lack supportive evidence from our existing functional assays. These variants may still have a detrimental effect on THOC2 function due to altered protein structure impacting protein–RNA and/or protein–protein interactions with known or unknown TREX subunits [e.g., Boehringer et al., 2017]. The challenge of proving causality for previously unreported missense variants in NDD genes is well recognized and speaks to the need for ongoing intertwined clinical and research efforts to clarify causality of VOUS (Wright et al., 2018). We therefore report detailed variant and clinical data (see Supp. materials, Supp. Table S1 and S2, and Supp. Figure S4) with the intention of alerting researchers and clinicians to these variants, as future studies, for example identification of their recurrence in affected individuals with overlapping clinical phenotypes or pathophysiological investigations, may help clarify their clinical significance.

THOC2 is ubiquitously expressed in all human tissues (Thul et al., 2017) and more specifically in the developing and mature human brain (Johnson et al., 2009; Kumar et al., 2015; Uhlen et al., 2015) and mouse brain, with higher abundance in frontal cortex and cerebellum (Di Gregorio et al., 2013; Kumar et al., 2015). THOC2 is an essential mRNA export factor as its siRNA-mediated depletion results in almost complete retention of mRNAs in the cell nucleus (Chi et al., 2013), potentially toxic to the cell. These data are consistent with the findings that THOC2 is a highly constrained gene (Samocha et al., 2014)
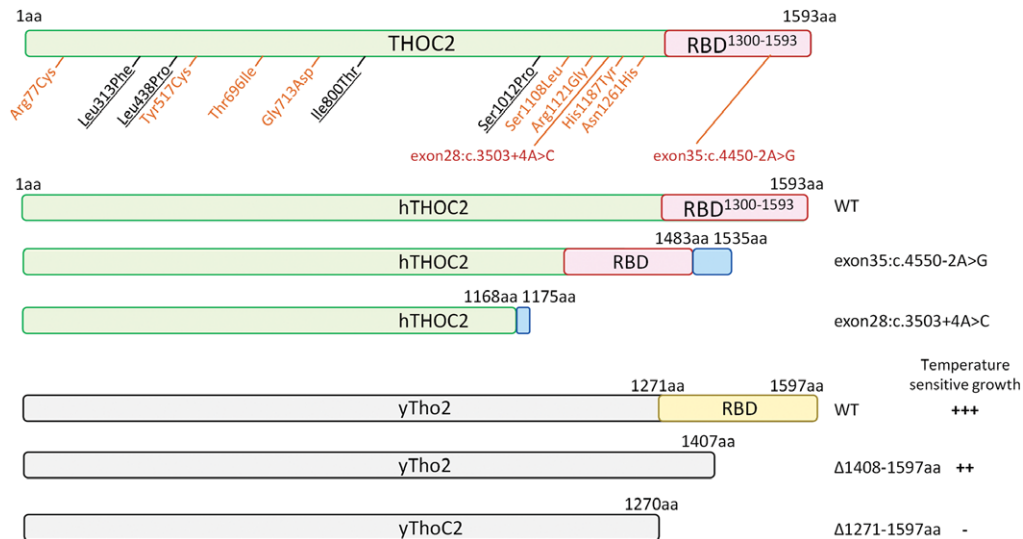
FIGURE 6 Summary of truncated human THOC2 proteins translated from aberrantly spliced mRNAs and functional outcomes of yeast C-terminal Tho2 deletion strains (Pena et al., 2012). Blue boxes depict the 52 and 7 amino acids coded by unspliced intron sequences of exon35:c.4550-2T>C and exon28:c.3503+4A>C variants, respectively. WT, +++ = normal, Δ1,408–1,597aa, ++ and Δ1,271–1,597aa = reduced growth at restrictive temperature

and *THOC* (e.g., *THOC1, 3, 5, 6*, and *7*) genes are essential for cell survival (Blomen et al., 2015). Taken together, as *THOC2* knockout cells will not survive due to complete mRNA nuclear retention, we predict that the identified *THOC2* variants represent partial loss-of-function that disrupt normal mRNA export in neuronal and possibly other cell types, potentially causing variable clinical presentations.

TREX complex couples transcription and mRNA biogenesis with nuclear mRNA export, and has emerged as an essential pathway in embryogenesis, organogenesis and differentiation (Heath et al., 2016). For example, Thoc2 and Thoc5 selectively bind and regulate export of mRNAs (e.g., *Nanog, Sox2, Esrrb*, and *Klf4* mRNAs) involved in maintenance of pluripotency of mouse ESCs (Wang et al., 2013) and Thoc5 in maintenance of hematopoiesis and HSP70 mRNA export (Katahira, Inoue, Hurt, & Yoneda, 2009; Mancini et al., 2010). Mouse modeling shows that both *Thoc1* and *Thoc5* knockouts are embryonic lethal (Mancini et al., 2010; Wang, Chang, Li, Zhang, & Goodrich, 2006). However, *Thoc1* and *Thoc5* expression in a range of developing and adult tissues may indicate that the two genes have a more essential role in early embryonic development compared to less stringent requirement during later stages of embryonic or adult development (Mancini et al., 2010; Wang et al., 2006); a functional pattern most likely followed by the *THOC2* gene. Essentiality of *THOC2* gene indicates that *THOC2* knockout will also be lethal. However, reduced levels or perturbed functionality can lead to a range of NDD phenotypes as observed for a cohort of THOC2 variants identified by us. It is now well established that development of brain depends on tightly regulated and complex sequence of events involving neuronal and glial cell proliferation, migration, and maturation (Chiurazzi & Pirozzi, 2016). Therefore, it is not surprising that our *THOC2* variant data and published work (Dickinson et al., 2016) provides strong evidence that even subtle alterations to the canonical molecular pathways such as mRNA export, otherwise essential for cellular life, can be tolerated but at a cost of a NDD.

In summary, we present detailed clinical data on seven individuals with *THOC2*-associated ID caused by both missense and splice variants that meet ACMG criteria for (likely) pathogenicity. They have a core phenotype of ID, and common findings of behavioural disorders, infantile hypotonia, gait disturbance and growth impairment, similar to the affected males with THOC2-associated ID we previously reported (Kumar et al., 2015). Other than the affected female with a *de novo* missense variant, heterozygote carrier females are typically unaffected. We also present data on five individuals with four previously unreported rare missense variants that show clinical overlap with our core group, but where convincing evidence for causality is still required. The significance of these variants may be clarified as additional individuals with *THOC2* variants are reported. We have also "adopted" *THOC2* on the Human Disease Gene (HDG) Website Series (https://humandiseasegenes.nl/thoc2) in an effort to continue to explore the phenotypic-genotypic spectrum for THOC2-related ID.

DISCLOSURE STATEMENT

The authors declare no conflict of interest.

ORCID

Raman Kumar (iD) http://orcid.org/0000-0001-7976-8386

Tracy Dudding-Byth (iD) http://orcid.org/0000-0002-9551-1107

Elizabeth E. Palmer (iD) http://orcid.org/0000-0003-1844-215X

Jozef Gecz (iD) http://orcid.org/0000-0002-7884-6861

REFERENCES

Allen, A. S., Berkovic, S. F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E. E., … Winawer, M. R. (2013). De novo mutations in epileptic encephalopathies. *Nature*, *501*(7466), 217–221.

Amos, J. S., Huang, L., Thevenon, J., Kariminedjad, A., Beaulieu, C. L., Masurel-Paulet, A., … Boycott, K.M. (2017). Autosomal recessive mutations in THOC6 cause intellectual disability: Syndrome delineation requiring forward and reverse phenotyping. *Clinical Genetics*, *91*(1), 92–99.

Amsterdam, A., Nissen, R. M., Sun, Z., Swindell, E. C., Farrington, S., & Hopkins, N. (2004). Identification of 315 genes essential for early zebrafish development. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 12792–12797.

Beaulieu, C. L., Huang, L., Innes, A. M., Akimenko, M. A., Puffenberger, E. G., Schwartz, C., … Boycott, K. M. (2013). Intellectual disability associated with a homozygous missense mutation in THOC6. *Orphanet Journal of Rare Diseases*, *8*(1), 62.

Bittles, A. H., Petterson, B. A., Sullivan, S. G., Hussain, R., Glasson, E. J., & Montgomery, P. D. (2002). The influence of intellectual disability on life expectancy. *Journal of Gerontology A Biological Sciences and Medical Sciences*, *57*(7), M470–2.

Blomen, V. A., Majek, P., Jae, L. T., Bigenzahn, J. W., Nieuwenhuis, J., Staring, J., … Brummelkamp, T. R. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science*, *350*(6264), 1092–1096.

Boehringer, A., Garcia-Mansfield, K., Singh, G., Bakkar, N., Pirrotte, P., & Bowser, R. (2017). ALS associated mutations in Matrin 3 alter protein-protein interactions and impede mRNA nuclear export. *Science Reports*, *7*(1), 14529.

Chi, B., Wang, Q., Wu, G., Tan, M., Wang, L., Shi, M., … Cheng, H. (2013). Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Research*, *41*(2), 1294–1306.

Chinnam, M., Wang, Y., Zhang, X., Gold, D. L., Khoury, T., Nikitin, A. Y., … Goodrich, D. W. (2014). The Thoc1 ribonucleoprotein and prostate cancer progression. *Journal of the National Cancer Institute*, *106*(11), dju306.

Chiurazzi, P., & Pirozzi, F. (2016). Advances in understanding - genetic basis of intellectual disability. *F1000Res*, *5*, 599.

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, *31*(16), 2745–2747.

Coe, B. P., Witherspoon, K., Rosenfeld, J. A., van Bon, B. W., Vulto-van Silfhout, A. T., Bosco, P., … Eichler, E. E. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nature Genetics*, *46*(10), 1063–1071.

de Lange, I. M., Helbig, K. L., Weckhuysen, S., Moller, R. S., Velinov, M., Dolzhanskaya, N., … Koeleman, B. P. C. (2016). De novo mutations of KIAA2022 in females cause intellectual disability and intractable epilepsy. *Journal of Medical Genetics*, *53*(12), 850–858.

Di Gregorio, E., Bianchi, F. T., Schiavi, A., Chiotto, A. M., Rolando, M., Verdun di Cantogno, L., … Brusco, A. (2013). A de novo X;8 translocation creates a PTK2-THOC2 gene fusion with THOC2 expression knockdown in a patient with psychomotor retardation and congenital cerebellar hypoplasia. *Journal of Medical Genetics*, *50*(8), 543–551.

Dickinson, M. E., Flenniken, A. M., Ji, X., Teboul, L., Wong, M. D., White, J. K., … Murray, S. A. (2016). High-throughput discovery of novel developmental phenotypes. *Nature*, *537*(7621), 508–514.

Dudding-Byth, T., Baxter, A., Holliday, E. G., Hackett, A., O'Donnell, S., White, S. M., … Lovell, B. C. (2017). Computer face-matching technology using two-dimensional photographs accurately matches the facial gestalt of unrelated individuals with the same syndromic form of intellectual disability. *BMC Biotechnology*, *17*(1), 90.

Guria, A., Tran, D. D., Ramachandran, S., Koch, A., El Bounkari, O., Dutta, P., … Tamura, T. (2011). Identification of mRNAs that are spliced but not exported to the cytoplasm in the absence of THOC5 in mouse embryo fibroblasts. *RNA*, *17*(6), 1048–1056.

Hautbergue, G. M. (2017). RNA Nuclear Export: From neurological disorders to cancer. *Advances in Experimental Medicine and Biology*, *1007*, 89–109.

Heath, C. G., Viphakone, N., & Wilson, S. A. (2016). The role of TREX in gene expression and disease. *Biochemical Journal*, *473*(19), 2911–2935.

Hirayama, S., Yamazaki, Y., Kitamura, A., Oda, Y., Morito, D., Okawa, K., … Nagata, K. (2008). MKKS is a centrosome-shuttling protein degraded by disease-causing mutations via CHIP-mediated ubiquitination. *Molecular Biology of the Cell*, *19*(3), 899–911.

Hosking, F. J., Carey, I. M., Shah, S. M., Harris, T., DeWilde, S., Beighton, C., & Cook, D. G. (2016). Mortality among adults with intellectual disability in England: Comparisons with the general population. *American Journal of Public Health*, *106*(8), 1483–1490.

Johnson, M. B., Kawasawa, Y. I., Mason, C. E., Krsnik, Z., Coppola, G., Bogdanovic, D., … Sestan, N. (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*, *62*(4), 494–509.

Katahira, J., Inoue, H., Hurt, E., & Yoneda, Y. (2009). Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. *EMBO Journal*, *28*(5), 556–567.

Kim, W., Bennett, E. J., Huttlin, E. L., Guo, A., Li, J., Possemato, A., … Gygi, S. P. (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. *Molecular Cell*, *44*(2), 325–340.

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315.

Kumar, R., Corbett, M. A., van Bon, B. W., Woenig, J. A., Weir, L., Douglas, E., … Gecz, J. (2015). THOC2 mutations implicate mRNA-export pathway in X-linked intellectual disability. *American Journal of Human Genetics*, *97*(2), 302–310.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., … MacArthur, D. G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285–291.

Liu, C., Yue, B., Yuan, C., Zhao, S., Fang, C., Yu, Y., & Yan, D. (2015). Elevated expression of Thoc1 is associated with aggressive phenotype and poor prognosis in colorectal cancer. *Biochemical and Biophysical Research Communication*, *468*(1-2), 53–58.

Lopitz-Otsoa, F., Rodriguez-Suarez, E., Aillet, F., Casado-Vela, J., Lang, V., Matthiesen, R., … Rodriguez, M. S. (2012). Integrative analysis of the ubiquitin proteome isolated using Tandem Ubiquitin Binding Entities (TUBEs). *Journal of Proteomics*, *75*(10), 2998–3014.

Mancini, A., Niemann-Seyde, S. C., Pankow, R., El Bounkari, O., Klebba-Farber, S., Koch, A., … Tamura, T. (2010). THOC5/FMIP, an mRNA export TREX complex protein, is essential for hematopoietic primitive cell survival in vivo. *BMC Biology*, *8*, 1.

Milani, D., Ronzoni, L., & Esposito, S. (2015). Genetic advances in intellectual disability. *Journal of Pediatric Genetics*, *4*(3), 125–127.

Morgan, D. O. (2013). The D box meets its match. *Molecular Cell*, *50*(5), 609–610.

Oeseburg, B., Dijkstra, G. J., Groothoff, J. W., Reijneveld, S. A., & Jansen, D. E. (2011). Prevalence of chronic health conditions in children with intellectual disability: A systematic literature review. *Intellectual and Developmental Disabilities*, *49*(2), 59–85.

Ohno, K., Takeda, J. I., & Masuda, A. (2018). Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdisciplinary Reviews RNA*, *9*(1), e1451.

Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., & Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, 127(3), 635–648.

Palmer, E. E., Kumar, R., Gordon, C. T., Shaw, M., Hubert, L., Carroll, R., … Gecz, J. (2017). A recurrent de novo nonsense variant in ZSWIM6 results in severe intellectual disability without frontonasal or limb malformations. *American Journal of Human Genetics*, 101(6), 995–1005.

Palmer, E. E., Stuhlmann, T., Weinert, S., Haan, E., Van Esch, H., Holvoet, M., … Kalscheuer, V. M. (2016). De novo and inherited mutations in the X-linked gene CLCN4 are associated with syndromic intellectual disability and behavior and seizure disorders in males and females. *Molecular Psychiatry*, 23(2), 222–230.

Pena, A., Gewartowski, K., Mroczek, S., Cuellar, J., Szykowska, A., Prokop, A., … Dziembowski, A. (2012). Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *EMBO Journal*, 31(6), 1605–1616.

Rehwinkel, J., Herold, A., Gari, K., Kocher, T., Rode, M., Ciccarelli, F. L., … Izaurralde, E. (2004). Genome-wide analysis of mRNAs regulated by the THO complex in *Drosophila melanogaster*. *Nature Structural and Molecular Biology*, 11(6), 558–566.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., … Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424.

Ropers, H. H., & Hamel, B. C. (2005). X-linked mental retardation. *Nature Reviews Genetics*, 6(1), 46–57.

Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., … Daly, M. J. (2014). A framework for the interpretation of de novo mutation in human disease. *Nature Genetics*, 46(9), 944–950.

Schwartz, C. E. (2015). *X-linked intellectual disability genetics.* Wiley Online Library; Chichester, UK: John Wiley & Sons, Ltd.

Snijders Blok, L., Madsen, E., Juusola, J., Gilissen, C., Baralle, D., Reijnders, M. R., … Kleefstra, T. (2015). Mutations in DDX3X are a common cause of unexplained intellectual disability with gender-specific effects on Wnt signaling. *American Journal of Human Genetics*, 97(2), 343–352.

Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., … Lundberg, E. (2017). A subcellular map of the human proteome. *Science*, 356(6340)

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., … Ponten, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science*, 347(6220), 1260419.

Viphakone, N., Cumberbatch, M. G., Livingstone, M. J., Heath, P. R., Dickman, M. J., Catto, J. W., & Wilson, S. A. (2015). Luzp4 defines a new mRNA export pathway in cancer cells. *Nucleic Acids Research*, 43(4), 2353–2366.

Vissers, L. E., Gilissen, C., & Veltman, J. A. (2016). Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics*, 17(1), 9–18.

Wagner, S. A., Beli, P., Weinert, B. T., Nielsen, M. L., Cox, J., Mann, M., & Choudhary, C. (2011). A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Molecular and Cellular Proteomics*, 10(10). M111.013284.

Wang, L., Miao, Y. L., Zheng, X., Lackford, B., Zhou, B., Han, L., … Hu, G. (2013). The THO complex regulates pluripotency gene mRNA export and controls embryonic stem cell self-renewal and somatic cell reprogramming. *Cell Stem Cell*, 13(6), 676–690.

Wang, X., Chang, Y., Li, Y., Zhang, X., & Goodrich, D. W. (2006). Thoc1/Hpr1/p84 is essential for early embryonic development in the mouse. *Molecular and Cellular Biology*, 26(11), 4362–4367.

Wang, X., Chinnam, M., Wang, J., Wang, Y., Zhang, X., Marcon, E., … Goodrich, D. W. (2009). Thoc1 deficiency compromises gene expression necessary for normal testis development in the mouse. *Molecular and Cellular Biology*, 29(10), 2794–2803.

Woerner, A. C., Frottin, F., Hornburg, D., Feng, L. R., Meissner, F., Patra, M., … Hipp, M. S. (2016). Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. *Science*, 351(6269), 173–176.

Wright, C. F., McRae, J. F., Clayton, S., Gallone, G., Aitken, S., FitzGerald, T. W., … Firth, H. V. (2018). Making new genetic diagnoses with old data: Iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genetics in Medicine*, https://doi.org/10.1038/gim.2017.246

Yamazaki, T., Fujiwara, N., Yukinaga, H., Ebisuya, M., Shiki, T., Kurihara, T., … Weis, K. (2010). The closely related RNA helicases, UAP56 and URH49, preferentially form distinct mRNA export machineries and coordinately regulate mitotic progression. *Molecular Biology of Cell*, 21(16), 2953–2965.

Zhu, X., Need, A. C., Petrovski, S., & Goldstein, D. B. (2014). One gene, many neuropsychiatric disorders: Lessons from Mendelian diseases. *Nature Neuroscience*, 17(6), 773–781.

Zweier, C., Rittinger, O., Bader, I., Berland, S., Cole, T., Degenhardt, F., … Wieczorek, D. (2014). Females with de novo aberrations in PHF6: Clinical overlap of Borjeson-Forssman-Lehmann with Coffin-Siris syndrome. *American Journal of Medical Genetics C: Seminars in Medical Genetics*, 166C(3), 290–301.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

FOXA2 gene mutation in a patient with congenital complex pituitary hormone deficiency

Hiroko Boda[a], Masafumi Miyata[a], Hidehito Inagaki[b, c], Yasuko Shinkai[c], Takema Kato[b, c], Tetsushi Yoshikawa[a], Hiroki Kurahashi[b, c, *]

[a] Department of Pediatrics, Fujita Health University School of Medicine, Japan
[b] Division of Molecular Genetics, Institute for Comprehensive Medical Science, Japan
[c] Genome and Transcriptome Analysis Center, Fujita Health University, Toyoake, Japan

ARTICLE INFO

ABSTRACT

We report a patient with congenital complex pituitary hormone deficiency (CPHD) with intestinal malrotation and anal atresia. We identified a *de novo* heterozygous mutation, c.664T > G (p.Cys222Gly), in the *FOXA2* gene in this individual. This missense mutation had the potential to affect the DNA binding properties of the FOXA2 protein based on a protein structure prediction. Since a CPHD patient with another missense mutation and one other case with an entire gene deletion have also been reported, we speculated that a haploinsufficiency of the *FOXA2* gene might be a genetic etiology for this disorder. Phenotypic similarities and differences among these three cases are also discussed.

1. Introduction

Congenital complex pituitary hormone deficiency (CPHD) is a disorder in which the anterior pituitary hormones (growth hormone, GH; prolactin, PRL; thyroid stimulating hormone, TSH; adrenocorticotropic hormone, ACTH; luteinizing hormone, LH; and follicle stimulating hormone, FSH) are deficient due to pituitary dysplasia [de Graaff, 2000]. The severity and phenotype of CPHD vary due to the combination of defective hormones. The onset of this condition may be noticed early in the newborn period, but can also be diagnosed in later infancy or childhood or adulthood. Occasionally, CPHD manifests as a syndromic disease with some extra-pituitary symptoms.

Since the first report of CPHD with GH, PRL, and TSH deficiencies due to a *POU1F1/PIT1* abnormality, the genes involved in the development of the pituitary-hypothalamus have been clarified consecutively. Abnormalities in many transcription factors such as *POU1F1*, *PROP1*, *HESX1*, *LHX3*, *LHX4*, *PITX2*, *SOX3*, *SOX2*, and *OTX2* have been identified [Tatsumi et al., 1992; Mehta and Dattani, 2008; Giordano, 2016]. The symptoms caused by these abnormalities are diverse, and the mode of inheritance also varies. Although many etiological genes are now known, the proportion of CPHD patients

with mutations in these known genes is not high. This suggests that there are many causative genes for this disorder that are yet to be identified.

Recent advances in next generation sequencing have facilitated the identification of novel disease-causing genes. By means of whole exome sequencing, we previously identified a *de novo* missense mutation, c.664T > G (p.Cys222Gly), in *FOXA2* in a girl with CPHD accompanied by gastrointestinal malformations (anal atresia, intestinal malrotation). In the current literature, there are two reports of a *FOXA2* mutation in patients with syndromic CPHD: an entire gene deletion and a c.505T > C mutation (p.Ser169Pro) [Tsai et al., 2015; Giri et al., 2017]. *FOXA2* has thus been identified as another causative gene in CPHD.

2. Patient data

A Japanese female infant was born by vaginal delivery at 40 weeks of gestation. Her Apgar score were 8 and 9 at 1 and 5 min after birth, respectively. At birth, her height was 52.5 cm (+1.5 S.D.), weight was 3574g (+0.5 S.D.), and head circumference was 35.0 cm (+1.1 S.D.). She was admitted to our neonatal intensive care unit one day after birth due to hypoglycemia (19 mg/dl at 3 h after birth)

* Corresponding author. Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi, 470-1192, Japan.
  Email address: kura@fujita-hu.ac.jp (H. Kurahashi).

which was not improved by breastfeeding. The patient did not have a dysmorphic face. Her mother had hydramnios during the pregnancy. However, her 25-year old mother, 28-year old father and 6-year old sister were all healthy.

After hospitalization, the patient was administered a glucose infusion but she still often had hypoglycemia without hyperinsulinemia. She had no growth hormone (GH) secretion response to an arginine stimulation test (GH < 0.03 ng/ml). She also had a low level of free thyroxine (fT4) (0.75 ng/ml) without thyroid stimulating hormone (TSH) elevation (3.78 µl/ml). There were no fT4 and TSH secretion responses to a thyrotropin-releasing hormone (TRH) stimulation test. Additionally, she had a low level of cortisol (0.1 µg/dl) and adenocorticotrophic hormone (ACTH) was undetectable (< 1.6 pg/ml). There was no cortisol secretion response to corticotropin releasing hormone (CRH) stimulation. An MRI scan of the patient's brain revealed no sella turcica, and neither the anterior pituitary nor the posterior pituitary was found in its normal position. Additionally, no ectopic posterior pituitary tissue was detected. Based on these results, we made a diagnosis of CPHD and administered thyroxine (10 µg/kg), hydrocortisone (40 mg/m^2), and somatropin (0.175 mg/kg/week).

This patient also had low anal atresia with anocutaneous fistula diagnosed by a lower gastrointestinal series. Defecation events were well controlled using an enema. However, as her bilious vomiting continued, we performed an upper gastrointestinal series from which we diagnosed an intestinal malrotation for which a radical operation was performed at 24 days after birth.

She underwent a further radical operation for low anal atresia at 9 months, but developed hypernatremia (158–165 mEq/L), high plasma osmolality (326 mOsm) and polyuria (max 14.9 ml/kg/hour) at five days post-surgery. We diagnosed central diabetes insipidus and commenced desmopressin treatment (0.5 µg/day). At the age of 1 year and 4 months, her growth was inadequate (height 68 cm (−3.3 S.D.), weight 6815 g (−3.2 S.D.), and head circumference 42.8 cm (−1.9 S.D.)) but her psychomotor development was normal.

3. Methods

3.1. Subjects

Peripheral blood samples were obtained from the patient and her parents. The research protocol for this study was approved by the local ethics committees of Fujita Health University, Japan. Written informed consent for the participation in this study was obtained from the parents.

3.2. Cytogenetic microarray

High-resolution chromosomal microarray analysis was conducted using the CytoScan HD array and Chromosome Analysis Suite 3.0 (Affymetrix, Santa Clara, CA) with a threshold level of 20 probes for a 50 kb region.

3.3. Whole exome sequencing

Fifty nanograms of the DNA samples were used to generate a whole exome library using SureSelect QXT Reagent and SureSelect Clinical Research Exome (Agilent Technologies, Santa Clara, CA). Sequencing was carried out using a HiSeq 1500 (Illumina, San Diego, CA). After demultiplexing from other sample data, the reads were mapped onto the human reference hg19 using BWA 0.7.15 [Li and Durbin, 2009]. Sorting and recalibration of the mapped reads, and variants were called into a VCF file using Picard tools 2.8.0 [Broad

Institute, 2017] and GATK 3.7 [Van der Auwera et al., 2013]. Annotations were added using Variant Studio 2.3 (Illumina). To identify disease causing mutations, we excluded known variants listed in the public databases (dbSNP147, 1000 Genomes Project, NHLBI ESP6500, and Exome Aggregation Consortium [ExAC]) and a control in-house database, except for those also identified as pathogenic mutations in the NCBI ClinVar and HGMD databases. Variants consistent with the phenotype in the pedigree (autosomal recessive or *de novo* dominant variants) were extracted. We focused on non-synonymous single nucleotide variants, insertions and deletions (indels), and splice site variants. The mutation was confirmed by Sanger sequencing of a PCR amplicon of the corresponding region.

3.4. Analysis of protein structure

Protein structure predictions were made using UCSF Chimera software [Pettersen et al., 2004].

4. Results

Since the current study patient was found to carry multiple congenital anomalies, we initially hypothesized that her phenotype was generated by a contiguous gene deletion syndrome. We performed cytogenetic microarray analysis to identify the responsible variant but no possibly pathogenic copy number variations were identified. We also performed targeted exome sequencing for the capture of 4813 disease-associated genes, including some known CPHD genes. We screened these data using CPHD as a keyword, but no mutation was identified among the candidate genes.

We next obtained parental samples and conducted trio-whole exome sequencing. We identified two *de novo* mutations, one of which was the missense mutation NM_021784.4: c.664T > G; p.(Cys222Gly) in *FOXA2* (Fig. 1A). This was confirmed by Sanger sequencing (Fig. 1B). This variant is not present in databases of normal healthy populations such as ExAC or 3554 Japanese database iJGVD [Yamaguchi-Kabata et al., 2015]. *In silico* functional analysis suggested that the mutation was damaging when using Polyphen2 (1.000) and deleterious with SIFT (0.000). Another missense mutation, NM_002095.5: c.346A > G; p.(Lys116Glu), was also identified in *GTF2E2*, which was not found in the databases either. However, *in silico* analysis predicted it as benign by Polyphen2 (0.153) and tolerated by SIFT (0.123). The classifications of *FOXA2* and *GTF2E2* mutations using the ACMG guideline [Richards et al., 2015] are likely pathogenic and uncertain significance, respectively. Thus, we considered that the c.664T > G; p.(Cys222Gly) mutation in *FOXA2* was the strong candidate as the cause of our patient's phenotype.

The Cys222 residue is located within the forkhead domain of FOXA2 and is highly conserved among various specie (human, mouse, dog, elephant, chicken and zebrafish) (Fig. 1A). Forkhead domains are DNA binding domains that are commonly found in transcription factors. We predicted the effect of *FOXA2* mutation in our current patient by protein structure analysis using Protein Data Bank Japan (PDBj). Recently, the c.505T > C (p.Ser169Pro) mutation in *FOXA2* was reported in a patient with CPHD (Giri et al., 2017). The Ser169 residue in FOXA2 (Fig. 2) is located at a position that is predicted to directly contact the DNA backbone, suggesting that the p.Ser169Pro mutation would affect the DNA binding ability of FOXA2. On the other hand, the Cys222 is located between the α-helix of this protein, at residues Gln206-Phe219, that is predicted to associate with the DNA major groove and the Lys225 that also contacts DNA (Fig. 2), suggesting that the p.Cys222Gly substitution might also affect the DNA binding ability of FOXA2.
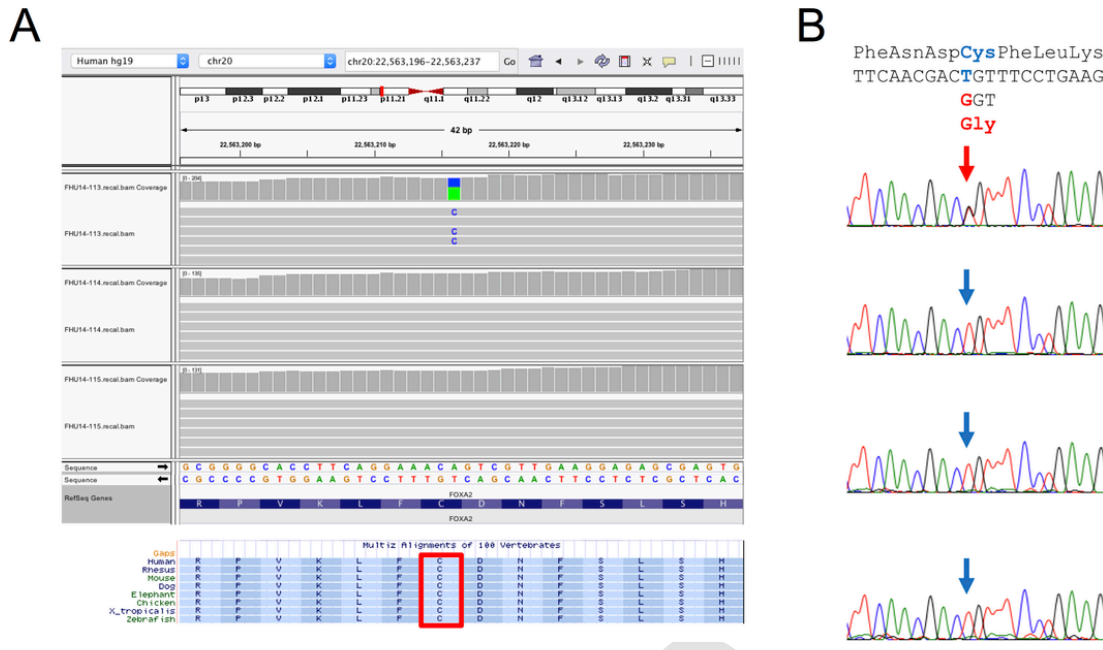
Fig. 1. Mutational analysis of the current study patient with CPHD. **A.** Trio analysis of whole exome sequences obtained using a next generation sequencer; patient (top), father (middle), and mother (bottom). A *de novo* heterozygous c.664T > G mutation was identified in *FOXA2*. The complementary strand is shown. Levels of conservation of the forkhead domain in FOXA2 among different species is shown at the bottom. **B.** Confirmation of the c.664T > G mutation by Sanger sequencing. This is a missense mutation that causes a TGT(Cys) to GGT(Gly) substitution.
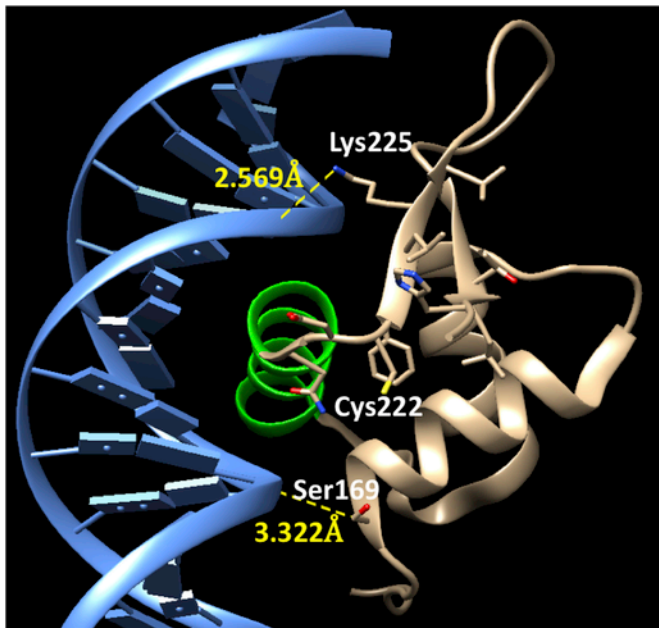

Fig. 2. Prediction of FOXA2 structures using PDB ID:5X07. The Ser169, Cys222, and Lys225 residues are indicated. The α-helix of Gln206-Phe219 is colored in green. Blue indicates DNA. The c.505T > C (p.Ser169Pro) mutation may directly affect the DNA binding affinity of the protein. The c.664T > G (p.Cys222Gly) mutation might also impact on the DNA binding ability of FOXA2. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

5. Discussion

In this study, we have identified a novel *de novo* missense mutation in *FOXA2* in a CPHD patient. This mutation is not listed in databases for the normal population and was predicted to affect the DNA binding ability of FOXA2. Further, *FOXA2* encoding this transcription factor is abundantly expressed in the hypothalamic-pituitary axis in the developing mouse embryo [Giri et al., 2017]. Two other CPHD patients have been reported in the present literature with a *FOXA2* mutation [Tsai et al., 2015; Giri et al., 2017]. The evidence from these two prior reports and our current analyses thus indicate that *FOXA2* mutations are causative for CPHD.

Our current case and one previous CPHD patient have been found to carry a missense mutation in *FOXA2* [Giri et al., 2017], located within the highly conserved forkhead domain in both instances. Protein structural analysis enabled us to predict the effect of these mutations on the DNA binding propensity of the protein products. Our analysis indicates that these mutations may cause a loss-of-function. Further, as another reported case showing an entire *FOXA2* deletion also manifested CPHD, we speculate that a haploinsufficiency of *FOXA2* might play an etiologic role in this disorder [Tsai et al., 2015]. *FOXA2* is a key transcriptional factor during embryonic development, and *FOXA2* null mice die during the early embryonic stage [Ang and Rossant, 1994; Weinstein et al., 1994]. Individuals with a *FOXA2* haploinsufficiency could possibly survive this stage but might show a combination of congenital disorders of organs whose development is regulated by *FOXA2*.

The disease phenotype of the three CPHD patients reported to date with *FOXA2* mutations, including our current case, are compared in Table 1. The main clinical findings of our present patient were CPHD, intestinal malrotation and anal atresia. The prior case with p.Ser169Pro reported by Giri et al. presented with CPHD but no gastrointestinal malformation other than gastroesophageal reflux [Giri et al., 2017]. Interestingly, a previously reported female proband with an entire *FOXA2* deletion manifested an intestinal malrotation similar to our present patient, but did not have CPHD. However, her father who also carries the deletion suffered from this disorder. Such phenotypic variations, even among cases with the same mutation, are often observed in dominant diseases caused by a haploinsufficiency in a transcriptional regulator that functions in development. Recently, another patient with *de novo FOXA2* mutation, c.770G > T, p.(Arg257Leu) was reported [Vajravelu et al., 2018]. Similar to our case, the mutation was located within the DNA bind-

Table 1
Summary of phenotypes of cases with *FOXA2* mutations.

	Case 1[a]	Case 2[b]	Case 3 (father of Case 2)[b]	Case 4[c]	Present case
Age	5 years	2 months	unknown	7 h of life	1 day
Sex	Female	Female	Male	Female	Female
Genetic findings	*FOXA2*: c.505T > C; p.Ser169Pro	277 kb deletion of 20p11.21 (*FOXA2* involved)	277 kb deletion of 20p11.21 (*FOXA2* involved)	*FOXA2*: c.770G > T; p.R257L	*FOXA2*: c.664T > G, p.Cys222Gly
Pituitary function	Panhypopituitarism	Normal	Panhypopituitarism	Panhypopituitarism	Panhypopituitarism
MRI findings	Thin pituitary stalk, hypoplastic anterior pituitary, thin corpus callosum	N/A	N/A	Small pituitary gland, shallow sella turcica with diminutive pituitary tissue, an ectopic posterior pituitary bright spot along the tuber cinereum, nonvisualization/absence of the infundibulum	Defect of sella turcica, anterior and posterior pituitary
Other complications	Hyperinsulinism, single median maxillary central incisor, congenital nasal pyriform aperture stenosis, choroidal coloboma, supra-valvular pulmonary stenosis, feed intolerance, severe gastro-esophageal reflex disease, portal-portal bridging fibrosis in liver, persistent oxygen requirement of unknown etiology, speech and motor developmental delay	Abdominal heterotaxy, biliary atresia, enlarged spleen, intestinal malrotation, interrupted inferior vena cava, cholangitis	Situs inversus, polysplenia, mildly dysmorphic facial features	Hyperinsulinism, coarse facial features, hypertelorism, thin upper lip, low-set ears, widely spaced nipples	Hypoglycemia without hyperinsulinism, anal atresia, intestinal malrotation

[a] Giri et al., [2017].
[b] Tsai et al., [2015].
[c] Vajravelu et al., [2018].

ing domain. Remarkably, all of the patients with *de novo* FOXA2 mutation exhibited dysfunction of the glucose regulation (Table 1). Further studies are needed to better understand what affects phenotype penetrance among individuals who have a haploinsufficiency in *FOXA2*.

References

Ang, S.L., Rossant, J., 1994. HNF-3 beta is essential for node and notochord formation in mouse development. Cell 78, 561–574.
Broad Institute, 2017. Picard [program]. Retrieved from https://broadinstitute.github.io/picard/.
de Graaff, L.C.G., 2000. PROP1-Related Combined Pituitary Hormone Deficiency. GeneReviews®[Internet], Seattle, WA.
Giordano, M., 2016. Genetic causes of isolated and combined pituitary hormone deficiency. Best Practice & Research. Clin. Endocrinol. Metabol. 30, 679–691. https://doi.org/10.1016/j.beem.2016.09.005.
Giri, D., Vignola, M.L., Gualtieri, A., Scagliotti, V., McNamara, P., Peak, M., Didi, M., Gaston-Massuet, C., Senniappan, S., 2017. Novel FOXA2 mutation causes hyper-

insulinism, hypopituitarism with craniofacial and endoderm-derived organ abnormalities. Hum. Mol. Genet. 26, 4315–4326. https://doi.org/10.1093/hmg/ddx318.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., Scherer, S.W., 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 42, D986–D992. https://doi.org/10.1534/g3.113.008797.
Mehta, A., Dattani, M.T., 2008. Developmental disorders of the hypothalamus and pituitary gland associated with congenital hypopituitarism. Best Practice & Research. Clin. Endocrinol. Metabol. 22, 191–206. https://doi.org/10.1016/j.beem.2007.07.007.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084.
Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. Genet. Med. 17, 405–423. https://doi.org/10.1038/gim.2015.30.
Tatsumi, K., Miyai, K., Notomi, T., Kaibe, K., Amino, N., Mizuno, Y., Kohno, H., 1992. Cretinism with combined hormone deficiency caused by a mutation in the PIT1 gene. Nat. Genet. 1, 56–58. https://doi.org/10.1038/ng0492-56.
Tsai, E.A., Grochowski, C.M., Falsey, A.M., Rajagopalan, R., Wendel, D., Devoto, M., Krantz, I.D., Loomes, K.M., Spinner, N.B., 2015. Heterozygous deletion of FOXA2 segregates with disease in a family with heterotaxy, panhypopituitarism, and biliary atresia. Hum. Mutat. 36, 631–637. https://doi.org/10.1002/humu.22786.
Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinf. 43, https://doi.org/10.1002/0471250953.bi1110s43, 11.10.1-11.10.33.
Vajravelu, M.E., Chai, J., Krock, B., Baker, S., Langdon, D., Alter, C., De León, D.D., 2018. Congenital hyperinsulinism and hypopituitarism attributable to a mutation in FOXA2. J. Clin. Endocrinol. Metabol. 103, 1042–1047. https://doi.org/10.1210/jc.2017-02157.
Weinstein, D.C., Ruiz i Altaba, A., Chen, W.S., Hoodless, P., Prezioso, V.R., Jessell, T.M., Darnell Jr., J.E., 1994. The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. Cell 78, 575–588.
Yamaguchi-Kabata, Y., Nariai, N., Kawai, Y., Sato, Y., Kojima, K., Tateno, M., Katsuoka, F., Yasuda, J., Yamamoto, M., Nagasaki, M., 2015. iJGVD: an integrative Japanese genome variation database based on whole-genome sequencing. Hum. Genome Var. 2, 15050. https://doi.org/10.1038/hgv.2015.50.

SCIENTIFIC REPORTS

Genotype determination of the *OPN1LW/OPN1MW* genes: novel disease-causing mechanisms in Japanese patients with blue cone monochromacy

Satoshi Katagiri [1], Maki Iwasa[2], Takaaki Hayashi[1,3], Katsuhiro Hosono[4], Takahiro Yamashita[5], Kazuki Kuniyoshi[6], Shinji Ueno[7], Mineo Kondo[8], Hisao Ueyama[9], Hisakazu Ogita[9], Yoshinori Shichida[5], Hidehito Inagaki[10], Hiroki Kurahashi[10], Hiroyuki Kondo [11], Masahito Ohji[2], Yoshihiro Hotta[4] & Tadashi Nakano[1]

Blue cone monochromacy (BCM) is characterized by loss of function of both *OPN1LW* (the first) and *OPN1MW* (the downstream) genes on the X chromosome. The purpose of this study was to investigate the first and downstream genes in the *OPN1LW/OPN1MW* array in four unrelated Japanese males with BCM. In Case 1, only one gene was present. Abnormalities were found in the promoter, which had a mixed unique profile of first and downstream gene promoters and a −71A > C substitution. As the promoter was active in the reporter assay, the cause of BCM remains unclear. In Case 2, the same novel mutation, M273K, was present in exon 5 of both genes in a two-gene array. The mutant pigments showed no absorbance at any of the wavelengths tested, suggesting that the mutation causes pigment dysfunction. Case 3 had a large deletion including the locus control region and entire first gene. Case 4 also had a large deletion involving exons 2–6 of the first gene. As an intact LCR was present upstream and one apparently normal downstream gene was present, BCM in Case 4 was not ascribed solely to the deletion. The deletions in Cases 3 and 4 were considered to have been caused by non-homologous recombination.

The human retina contains three types of cone photoreceptors: long-wavelength sensitive cones (L cones), medium-wavelength sensitive cones (M cones), and short-wavelength sensitive cones (S cones). These cone photoreceptors express respective visual pigments, L, M, and S opsins. Among these, the genes encoding L opsin (*OPN1LW*, OMIM; *300822) and M opsin (*OPN1MW*, OMIM; *300821) are present in tandem on the human X chromosome[1,2], forming an L/M pigment gene array. In individuals with normal color vision, the first gene in the array is an L gene, and the downstream (the second and later) gene(s) is/are M genes. Abnormalities in the array are reportedly associated with protan and deutan color vision deficiencies[3], blue cone monochromacy (BCM)[4], and Bornholm eye disease[5].

[1]Department of Ophthalmology, The Jikei University School of Medicine, Tokyo, Japan. [2]Department of Ophthalmology, Shiga University of Medical Science, Shiga, Japan. [3]Department of Ophthalmology, Katsushika Medical Center, The Jikei University School of Medicine, Tokyo, Japan. [4]Department of Ophthalmology, Hamamatsu University School of Medicine, Shizuoka, Japan. [5]Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto, Japan. [6]Department of Ophthalmology, Kindai University Faculty of Medicine, Osaka, Japan. [7]Department of Ophthalmology, Nagoya University Graduate School of Medicine, Aichi, Japan. [8]Department of Ophthalmology, Mie University Graduate School of Medicine, Mie, Japan. [9]Department of Biochemistry and Molecular Biology, Shiga University of Medical Science, Shiga, Japan. [10]Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Aichi, Japan. [11]Department of Ophthalmology, University of Occupational and Environmental Health, Fukuoka, Japan. Satoshi Katagiri and Maki Iwasa contributed equally to this work. Correspondence and requests for materials should be addressed to T.H. (email: taka@jikei.ac.jp) or H.U. (email: datt@belle.shiga-med.ac.jp)

BCM (OMIM; #303700) is a rare congenital color vision deficiency with an X-linked inheritance pattern[4,6]. Cases of BCM typically present with severely impaired color discrimination, reduced visual acuity, nystagmus, photophobia, and diminished L/M cone function despite retention of rod and blue cone function[6,7]. The dysfunction in both L and M cones in BCM is reportedly caused by one of the three genotypes. One genotype involves deletion of the locus control region (LCR)[4,8–12], which is located upstream of the L/M gene array (−3,681 to −3,021 from the cap site of the first gene) and believed to be involved in the mutually exclusive expression of L and M genes[13,14]. Therefore, neither gene is expressed in the absence of the LCR. Another genotype involves a deleterious mutation in a single-gene array (either the L or M gene present alone in the array). The derivation of this genotype has two obvious steps: first, non-homologous recombination between the L and M genes to form a single-gene array followed by an inactivating mutation in the single gene (reverse order is also possible). The most common mutation is C203R[4,8,15–17], but other mutations, such as P307L[8], R247X[8], and deletion of exon 2[16], have also been documented. The LIAVA haplotype in exon 3, which affects splicing[18], was also reported in a single L gene[17]. The third genotype involves inactivating mutations in both the L and M genes. Although the C203R mutation has been documented in this genotype[8,16,19], the LIAVA haplotype (or a very similar haplotype) in exon 3 of both genes seems to be frequent[17,20].

Although little is known about the prevalence of BCM in the Japanese population, to date, only two BCM families have been described in the literature, demonstrating the mechanism of deletion of the LCR in both families[11,21].

In the current study, the L/M pigment gene arrays in four unrelated Japanese males with BCM were analyzed. The purpose of this study was to investigate their genotypes in the L/M pigment gene array, which could be categorized into one of the three above-mentioned genotypes, but others were unreported mechanisms and differed from each other.

Results

Case 1 (JU#1299).
Long-range polymerase chain reaction (PCR) was successful for the first gene but not for downstream gene(s) (Fig. 1A). Promoter analysis of gene number, by contrast, showed that the first gene promoter was absent (only the downstream gene promoter was detected) (Fig. 1B). From the results of repeated long-range PCR analysis of downstream genes, we concluded that this subject had a single gene (no downstream genes) in the array. The LCR was present upstream of the single gene, and its sequence had no aberrations. The curious result that this subject had only the downstream gene promoter (Fig. 1B) was later found to be due to the unique sequence of the promoter. The first and downstream gene promoters differ by 14 nucleotides, but in the promoter region of the single gene, 8 upstream sites were associated with the first gene, whereas the other 6 sites were a random mixture of these nucleotides (Fig. 2). Moreover, the promoter had a −71 A > C substitution, which has been reported to be associated with deutan color vision deficiency[22] (Fig. 2). The G at the −30 position in the $^{-34}$GCCGGT^{-29} sequence (the number is from the cap site of the first gene) in the promoter analysis indicated that the first gene promoter was absent (Fig. 1B). The −30 (A or G) site discriminates the promoters of the first and downstream genes by *Cfr*10I (recognition sequence = RCCGGY). Our conclusion was that Case 1 had a single gene array because the sequencing of the promoter showed only one curious pattern as shown in Fig. 2. The luciferase activity of the promoter of the single gene was more than twice that of the usual first gene promoter in the reporter assay (Supplementary Fig. S1). No abnormalities were found in exons 1–6 and their adjacent introns in nucleotide sequencing. Exons 2–5 were M type, and the haplotype of exon 3 was MVAIA rather than LIAVA (Table 1). The curious promoter found in Case 1 has not been reported previously.

Case 2 (JU#1311, KINKI-125-70).
Products of both the first and downstream genes were obtained using long-range PCR (Fig. 1A). Promoter analysis showed that the subject had a 2-gene array (Fig. 1B). Both genes had the same missense mutation (c.818 T > A, M273K) in exon 5 (Fig. 3A,B). The chromosome positions (GRCh38. p7) of the mutation are 154,156,367 (L gene) and 154,193,481 (M gene). The M273K mutation has not been reported previously, and not found in the Single Nucleotide Polymorphism Database (https://www.ncbi.nlm.nih.gov/projects/SNP/), Genome Aggregation Database (http://gnomad.broadinstitute.org/), Exome Aggregation Consortium (http://exac.broadinstitute.org/) and Human Genetic Variation Database (http://www.hgvd.genome.med.kyoto-u.ac.jp/). The analysis of the recombinant proteins of M273K mutants revealed that the opsin with the M273K mutation was significantly detectable in the Western blot and cultured cells (Supplementary Fig. S2) but showed no absorbance at any of the wavelengths tested after reconstitution with 11-*cis*-retinal (Fig. 3C). These results indicated that the M273K mutation in both genes results in dysfunctional opsin protein, probably because of a lack of the ability to bind to 11-*cis*-retinal. We therefore ascribed the BCM phenotype in this subject to the mutation. In the first gene, exons 2, 3, and 4 were L type, exon 5 was M type, and the haplotype of exon 3 was LVAIS. In the downstream gene, exons 2–5 were M type, and the haplotype of exon 3 was MVAIA (Table 1).

Case 3 (JU#1318, MIE-050-0071).
Long-range PCR was unsuccessful for both the first and downstream genes (Fig. 1A). Promoter analysis for determining gene number showed that the first gene promoter was absent (only the downstream gene promoter was detected) (Fig. 1B). Amplification of the LCR was also unsuccessful, indicating a deletion including both the LCR and first gene. To determine the exact deletion breakpoints, 11 sets of PCR primers were designed to cover the sequence −53,930 to −9,320 (number is from the cap site of the first gene) (NT_025965.12: 707,760 to 752,370). PCR products were obtained when using the UP8F and UP8R pair (Supplementary Table S1) but not the UP9F and UP9R pair (Supplementary Table S1), suggesting that the upstream breakpoint of the deletion was between −32,015 and −28,150 (NT_025965.12: 729,665 to 733,530). Long-range PCR using the primer sets UP8F/E5R and UP8F/E6R was successful in this subject but not in a color-normal control subject (Fig. 1C). According to the human genome database (NT_025965.12), the distance between UP8F- and E5R-corresponding regions and UP8F- and E6R-corresponding regions were 81,131 bp and
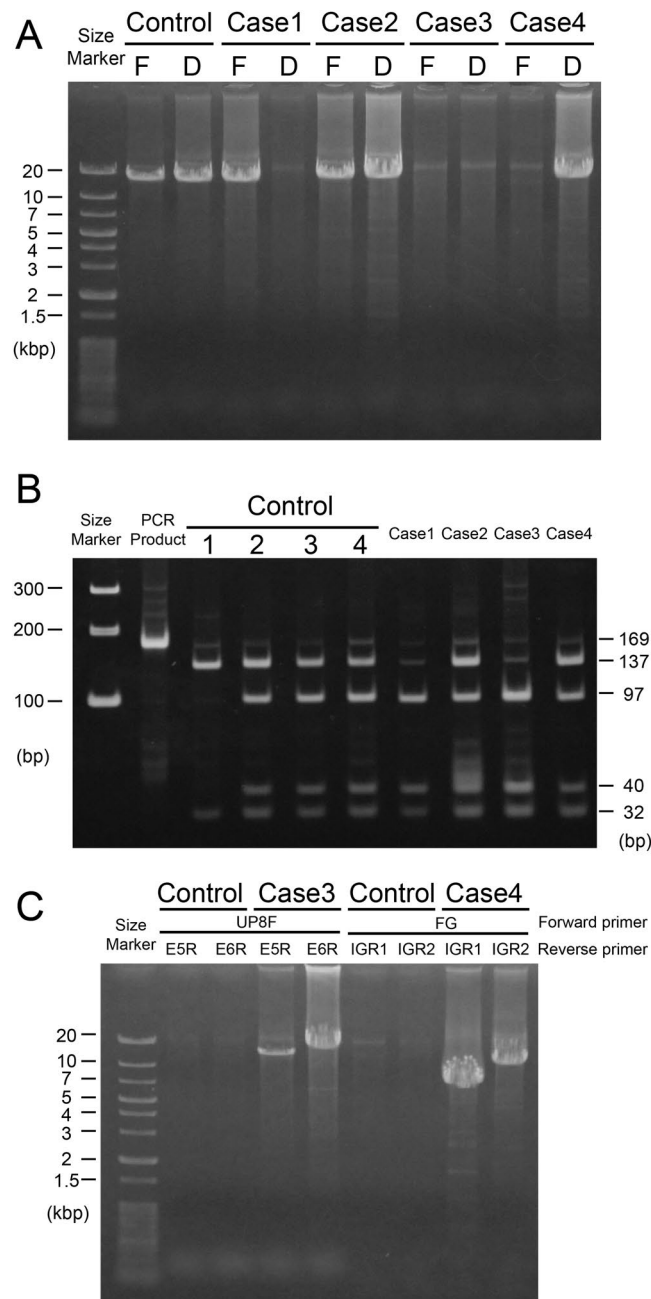
Figure 1. Long-range PCR and promoter analysis. (**A**) First and downstream genes in the L/M gene array were amplified separately by long-range PCR. The control was a color-normal subject having both the first and downstream genes. F, first gene; D, downstream gene(s). Thin bands of approximately 20 kb are not amplified products but the templates (genomic DNA, usually approximately 100 ng per reaction). (**B**) Promoter analysis of gene number. Promoters were amplified by PCR using primers common to the first and downstream genes. PCR products (169 bp) digested with *Cfr*10I were loaded onto a polyacrylamide gel. Controls 1–4 have gene numbers 1–4, respectively[28]. (**C**) Long-range PCR beyond the deletion. Combinations of the primers UP8F/E5R and UP8F/E6R were used for long-range PCR in the control and Case 3. Combinations of the primers FG/IGR1 and FG/IGR2 were used for long-range PCR in the control and Case 4.

83,925 bp, respectively (the E5R- and E6R-corresponding regions are those of the downstream gene), which were too far for long-range PCR. In Case 3, however, due to the large deletion including the first gene, the distances between the regions had been reduced to about 12 kbp and 15 kbp, respectively, and therefore, long-range PCR products were obtained in this case (Fig. 1C). The 15-kbp product contained not only exons 1–6 (exons 2, 4, and 5 were L type, exon 3 was M type, with the haplotype of MVAIA) (Table 1) but also the downstream gene promoter. By sequencing the 15-kbp product using the UP12F primer (Supplementary Table S1), the upstream breakpoint of the deletion was determined to be somewhere in the sequence $^{-31,241}$GAACTCCTGACCTCAGG$^{-31,225}$ (the number is from the cap site of the first gene) (NT_025965.12: 730,439 to 730,455), and the downstream breakpoint
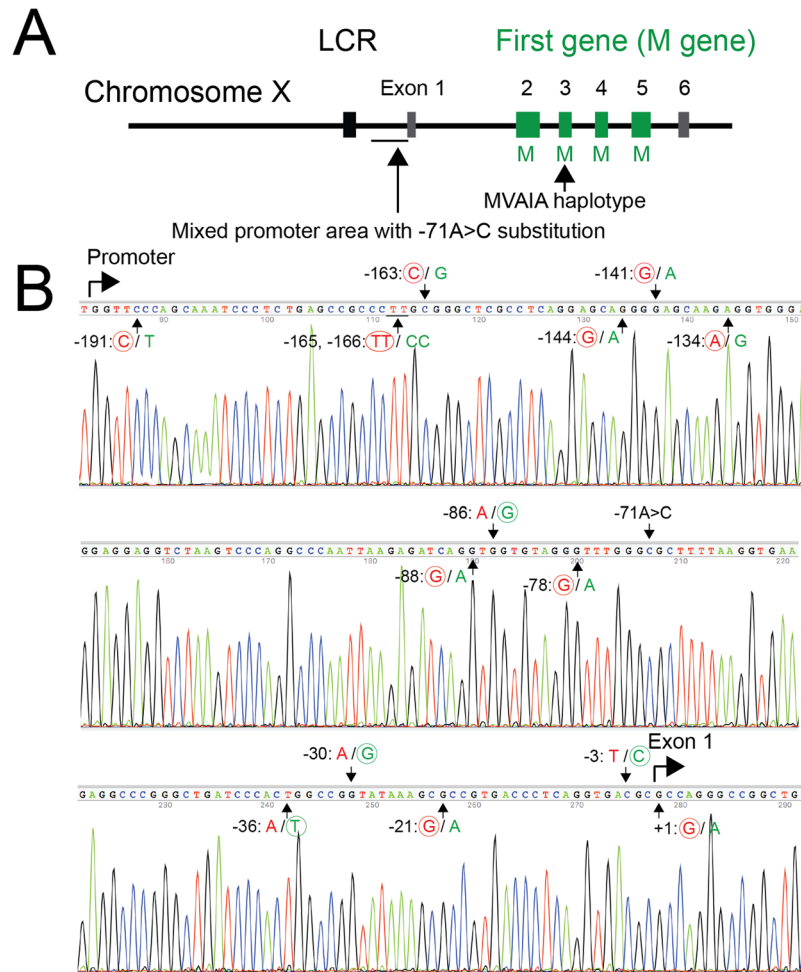
Figure 2. Genotype of Case 1. (**A**) Overview of the genotype of Case 1. Case 1 had an intact LCR and a single M gene array in which no aberrations were found. The promoter regions had a unique profile including a $-71\,A>C$ substitution. (**B**) The promoter of the single M-gene array. Black arrows indicate the 14 nucleotides differing between first and downstream genes and the $-71\,A>C$ substitution. At each position, the usual nucleotide of the first gene promoter is shown on the left side in red and that of the downstream gene promoter is shown on the right side in green. The nucleotides in Case 1 are circled.

was determined to be somewhere in the sequence $^{-407}$GAACTCCTGACCTCAGG^{-391} (the number is from the cap site of the downstream gene) (NT_025965.12: 799,682 to 799,698) (Fig. 4). The reason why long-range PCR was unsuccessful for the downstream gene (Fig. 1A) is that the deletion includes the region corresponding to the forward primer for long-range PCR, DG (-748 to -728 from the cap site of the downstream gene). The estimated size of the deletion was 69,243 bp. The long-range PCR products in Fig. 1C were calculated to be exactly 11,888 bp and 14,682 bp. As a LCR is reportedly necessary for the expression of L/M genes[13,14], the BCM phenotype in this subject was ascribed to the deletion.

Case 4 (JU#1368, Nagoya-140). Long-range PCR was successful for downstream gene(s) but not for the first gene (Fig. 1A). Promoter analysis of gene number showed a 1:1 ratio for the first and downstream genes (Fig. 1B). As the first gene promoter was shown to be present, the FG primer–corresponding region (upstream of the promoter) should also be present. PCR analysis using combinations of the FG primer and various intragenic reverse primers revealed that the upstream breakpoint of the deletion was within intron 1 (between the primer I1R1– and primer I1R2–corresponding regions) and that the deletion expanded beyond exon 6. The failure of long-range PCR for the first gene was ascribed to the absence of exon 6 (primer E6R corresponds to exon 6). To determine the downstream breakpoint of the deletion, 15 reverse primers specific to the intergenic region (between the first and downstream genes) were designed for long-range PCR. PCR products were obtained for two primer pairs (FG/IGR1, and FG/IGR2) in this subject but not in the control (Fig. 1C). According to the human genome database (NT_025965.12), the distances between the FG- and IGR1-corresponding regions and between the FG- and IGR2-corresponding regions were 30,137 bp and 34,286 bp, respectively, which were too long for long-range PCR. In Case 4, however, due to the large deletion, the distances had been reduced to approximately 7 kbp and 11 kbp, respectively, and therefore, long-range PCR products were obtained (Fig. 1C).

		Exon 2				Exon 3								Exon 4					Exon 5											Haplotype in Exon 3*
	Nucleotide position	194	300	331	347	453	457	465	511	513	521	532	538	689	697	698	699	706	820	823	825	828	830	835	849	853	888	892	926	
Reference	L gene	C	A	A	C	G	C	G	G/A	G/T	C/T	A/G	T	T	G	C	T	A	A	T	T	G	A	G	C	A	T	G	A	
	M gene	T	G	G	A	A	A	C	G/A	G/T	C/T	A/G	G	C	A	G	C	G	G	C	G	A	T	T	A	G	C	C	T	
	Amino acid Position	65	100	111	116	151	153	155	171–1	171–3	174	178	180	230	233–1	233–2	233–3	236	274	275–1	275–3	276	277	279	283	285	296	298	309	
	L gene	T	L	I	S	R	L	V	V/I		A/V	I/V	S	I	A			M	I	F		A	Y	V	P	T	G	A	Y	
	M gene	I	L	V	Y	R	M	V	V/I		A/V	I/V	A	T	S			V	V	L		A	F	F	P	A	G	P	F	
Case 1 First (single) gene	Nucleotide	T	G	G	A	A	A	C	G	G	C	A	G	C	A	G	C	G	G	C	G	A	T	T	A	G	C	C	T	MVAIA
	Amino acid	I	L	V	Y	R	M	V	V		A	I	A	T	S			V	V	L		A	F	F	P	A	G	P	F	
Case 2 First gene	Nucleotide	C	A	A	C	G	C	G	G	G	C	A	T	T	G	C	T	A	G	C	G	A	T	T	A	G	C	C	T	LVAIS
	Amino acid	T	L	I	S	R	L	V	V		A	I	S	I	A			M	V	L		A	F	F	P	A	G	P	F	
Case 2 Downstream gene	Nucleotide	T	G	G	A	A	A	C	G	G	C	A	G	C	A	G	C	G	G	C	G	A	T	T	A	G	C	C	T	MVAIA
	Amino acid	I	L	V	Y	R	M	V	V		A	I	A	T	S			V	V	L		A	F	F	P	A	G	P	F	
Case 3 Downstream gene (First gene was deleted)	Nucleotide	C	A	A	C	A	A	C	G	G	C	A	G	T	G	C	T	A	A	T	T	G	A	G	C	A	T	G	A	MVAIA
	Amino acid	T	L	I	S	R	M	V	V		A	I	A	I	A			M	I	F		A	Y	V	P	T	G	A	Y	
Case 4 Downstream gene (First gene was deleted)	Nucleotide	T	G	G	A	G	C	G	G	G	T	G	G	C	A	G	C	G	G	C	G	A	T	T	A	G	C	C	T	LVVVA
	Amino acid	I	L	V	Y	R	L	V	V		V	V	A	T	S			V	V	L		A	F	F	P	A	G	P	F	

Table 1. Nucleotides in each gene of the four cases. The positions of nucleotides different between wild-type L and M genes, and polymorphic nucleotide positions 511, 513, 522 and 532 as well, are shown. *Haplotype in Exon 3 was determined by amino acid residues at 153, 171, 174, 178, and 180.

The IGR1 primer corresponds to the region +14,867 to +14,887 (the number is from the stop codon in exon 6) (NT_025965.12: 791,231 to 791,251), and the IGR2 primer corresponds to the region +15,232 to +15,252 (NT_025965.12: 767,324 to 767,329). Using the 11-kbp PCR product and primer I1F (Supplementary Table S1), the upstream breakpoint of the deletion was determined to be somewhere in $^{+5,492}$TGAGCC$^{+5,497}$ (the number is from 5′ splice site of intron 1 of the first gene) (NT_025965.12: 767,324 to 767,329), and the downstream break-point was determined to be somewhere in $^{+14,714}$TGAGCC$^{+14,719}$ (the number is from the stop codon in exon 6) (NT_025965.12: 790,713 to 790,718) (Fig. 5). The estimated size of the deletion was 23,389 bp. The long-range PCR products shown in Fig. 1C were calculated to be exactly 6,748 bp and 10,897 bp. PCR confirmed the presence of the LCR, and its sequence had no aberrations. The downstream gene had no abnormalities in the promoter, exons 1–6, or their adjacent introns. Exons 2–5 were M type, and the haplotype of exon 3 was LVVVA (Table 1). As the downstream gene was apparently normal, the cause of BCM in this subject could not be confidently determined.

Discussion

In this study, we reported the results of gene analyses in four cases of BCM. Their genotypes were unreported and different from each other.

Case 1 had a single-gene array having a curious promoter sequences with a −71 A > C substitution. The −71 A > C substitution was reported to be associated with deutan color vision deficiency due to decreased promoter activity[22]. We hypothesized that the −71 A > C substitution causes dysfunction of the single gene. However, rather than being low, the activity of the promoter in Case 1 was more than twice that of the control in the reporter assay. It is reported that not only the LCR but also normal promoters in the L/M gene array were essential for expression of both L and M genes[13,14]. The LCR was not contained in the constructs for our reporter assay system. Although the dysfunction of the single opsin gene in Case 1 remains unclear, it might be possible that the curious promoter sequences (Fig. 2) might interfere with the LCR binding to the promoter.

Case 2 had a two-gene array, and both genes had a novel missense mutation (M273K) that would cause dysfunction of both gene products (Fig. 3C). The C203R mutation in exon 4 reportedly causes loss of function of the L/M gene products; this mutation causes deutan color vision deficiency when present in the M gene[23,24] and BCM when present in both the L and M genes[8,16,19]. The occurrence of C203R mutation in both the L and M genes was explained by gene conversion[19] (i.e., transfer of the C203R mutation present in the downstream M gene to the first L gene). The deleterious LIAVA haplotype in exon 3 of both the L and M genes was also explained by gene conversion[20]. Also, the occurrence of the M273K mutation in both the first and second genes (Fig. 3A) might be explained by the same mechanism of gene conversion seen in the C203R mutation[19]. Otherwise, because the first gene in this subject had L-type exons 2–4 and M-type exon 5 (with the mutation), we developed the
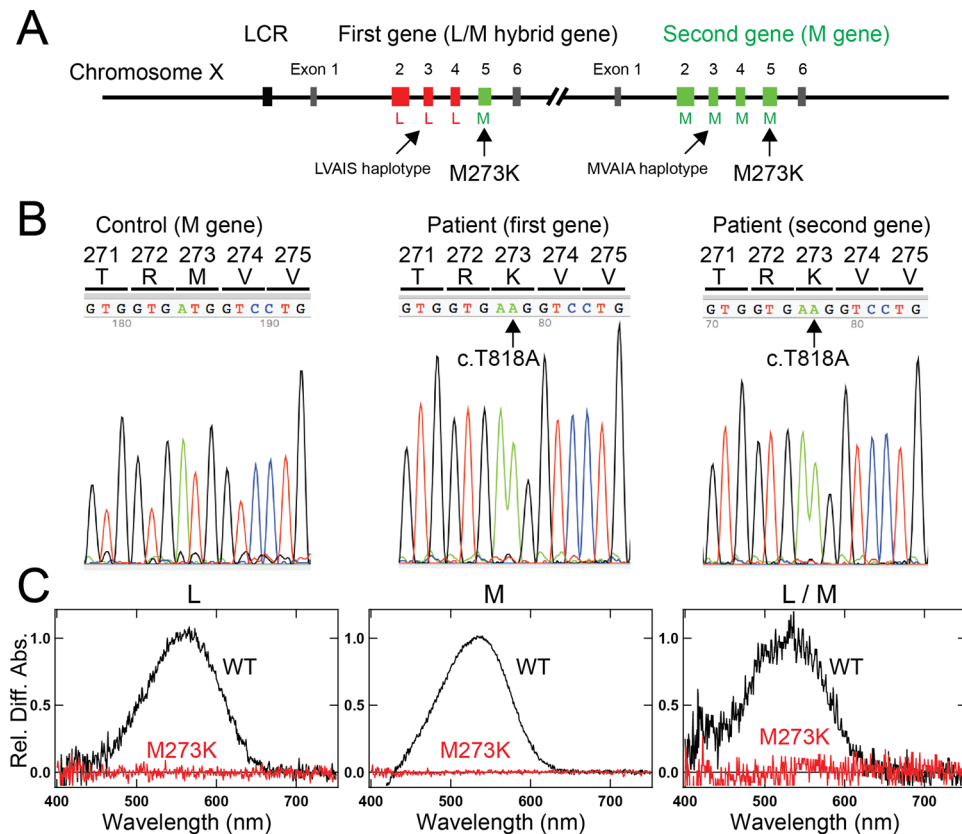
Figure 3. Genotype of Case 2. (**A**) Overview of the genotype of Case 2. Case 2 had an intact LCR and two genes. The first gene had exons 2–4 of L type, exon 3 with LVAIS haplotype, and exon 5 of M type. The second gene had exons 2–5 of M type and exon 3 with MVAIA haplotype. Both genes had the same missense mutation (c.818 T > A, M273K) in exon 5. (**B**) Partial sequence data around the missense mutation (c.818 T > A, M273K) in exon 5 in the control and two genes of Case 2. (**C**) Opsin reconstitution experiments. L, L opsin in which exons 2–5–derived amino acid sequences are all L type; M, M opsin in which exons 2–5–derived amino acid sequences are all M type, as in the product of the second gene of Case 2; L/M, M opsin in which exons 2–4–derived amino acid sequences are L type but exon 5–derived amino acid sequence is M type, as in the product of the first gene of Case 2. WT, wild-type opsin; M273K, mutant opsin with the M273K mutation. "Rel. Diff. Abs." indicates relative difference absorption.

following hypothesis as the other alternative possibility. The M273K mutation occurred in the downstream M gene (the array was L-M*; *denotes the mutation), duplication of the M gene occurred (L-M*-M*), followed by non-homologous recombination (L/M* hybrid-M*) as shown in Fig. 3A. Duplication of the second gene was supported by the fact that (i) many color-normal individuals have multiple downstream M genes with the same nucleotide sequence[25] and (ii) the result that three (or more) downstream M genes had the same 11-bp deletion in a protanopia subject[26]. Case 2 had a genotype with two unique profiles; the M273K mutation was novel and present in both L and M genes.

Cases 3 and 4 showed large deletions of 62,934 bp including the LCR and 23,389 bp not including the LCR, respectively. The genotype of Case 3 was consistent with the known genotype of BCM. In Case 4, the first gene was obviously non-functional due to the absence of exons 2–6, but the downstream gene seemed to be functional, as no deleterious mutations were found in the promoter and exons, including their adjacent introns. The clinical phenotype of BCM in Case 4 indicated that the downstream gene was non-functional. The gene array revealed that the promoter of the second gene was directly connected to intron 1 of the first gene with absence of exons 2–6 (23,389 deletion) (Fig. 5). Although we cannot explain reasonable mechanisms underlying BCM in Case 4, the residual sequences (exon 1 and partial intron 1) of the first gene might impact on the promoter activity of the second gene, resulting in suppression of the second gene expression. The breakpoints of the deletion were within *Alu* elements in both cases, as in previously reported BCM cases[11]. Many reports have described deletions in the L/M gene array[4,8–12,17]. However, few studies have determined the exact breakpoints of the deletion at the nucleotide level[4,11,16]; the breakpoints have resulted from simple breakage and fusion[4,16] outside the repetitive sequence and simple breakage and fusion (and insertion) in *Alu* elements[11]. The breakpoints we determined differed from these; non-equal crossing-over occurred between the two *Alu* elements in the region of the same sequence (Figs 4 and 5).

The various haplotypes of five amino acid residues at positions 153 (L/M), 171 (V/I), 174 (A/V), 178 (I/V), and 180 (S/A) in exon 3 have been reported in subjects with normal color vision and subjects with color vision deficiencies[18,20]. The haplotypes have been roughly classified into four groups in terms of the magnitude of the
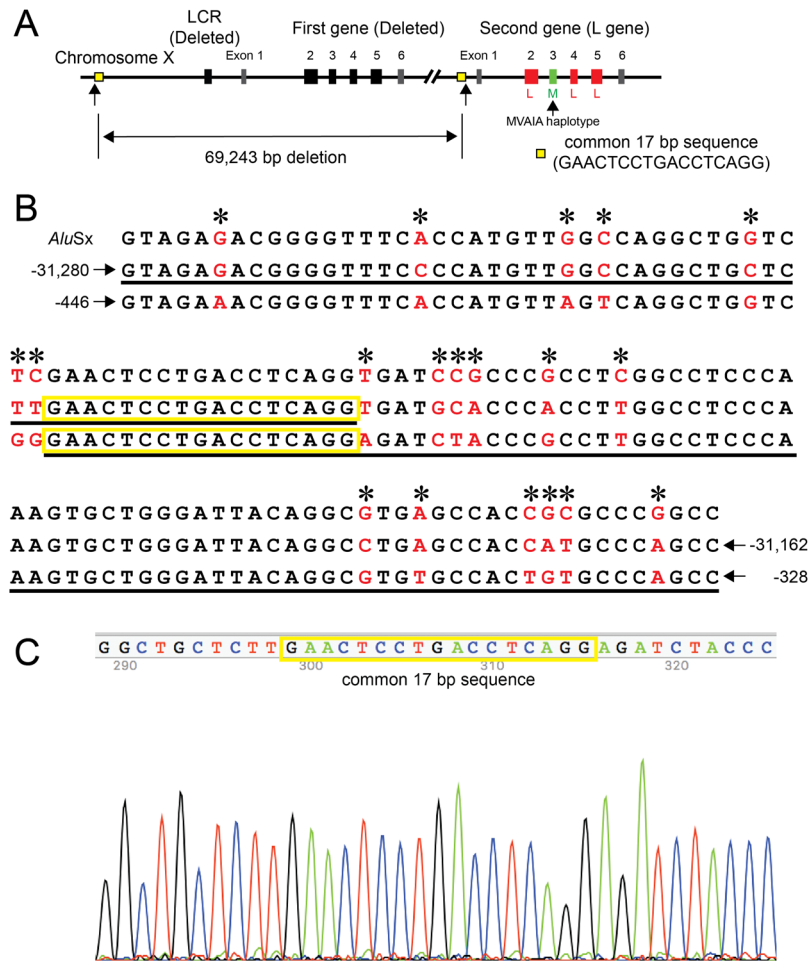
Figure 4. Genotype of Case 3. (**A**) Overview of the genotype of Case 3. Case 3 had a large deletion of 69,243 bp including the LCR and first gene. The remaining second gene was an L gene in which exons 2, 4, and 5 were L type. Exon 3 was M with MVAIA haplotype. (**B**) Upper row shows the consensus sequence for the left monomer of the *Alu* element (complementary sequence of AluSx). Middle row shows a part of the upstream region of the first gene (number is from the cap site of the first gene). Lower row shows a part of the upstream region of the downstream gene (number is from the cap site of the downstream gene). The nucleotides differing among the regions are shown in red and by asterisks. The actual sequence obtained in Case 3 is underlined. (**C**) Partial sequence data around the breakpoint of the deletion. The breakpoint is somewhere in the common 17-bp sequence.

splicing defect[20]; highly deleterious haplotypes include LIAVA, MIAVA, and LVAVA; intermediately deleterious haplotypes include LIAIA, LIAVS, and MVAVA, minor deleterious haplotypes include LVAIA, LVAIS, MVAIA, and MVVVA, and the MVAIS haplotype exhibits no splicing defect. According to our data, the MVAIA and LVAIS haplotypes in Cases 1–3 would be expected to produce essentially correct splicing; however, the LVVVA haplotype in Case 4 could not be classified, as this haplotype was not described in the above-mentioned study[20]. We[18] and other researchers[27] examined the LVVVA haplotype using a mini-gene system and observed that the opsin mRNA retaining exon 3 was in clearly greater abundance than that lacking exon 3. Moreover, we reported one color-normal subject in which the exon 3 haplotype was LVVVA[18]. Based on these results, the BCM phenotype in Case 4 could not be ascribed to the LVVVA haplotype in the downstream gene.

In conclusion, we reported four novel and different genotypes in four unrelated Japanese patients with BCM. In two patients (Case 2 and Case 3), the genotypes were consistent with that of BCM (the same deleterious mutation in both opsin genes and deletion of the LCR), but in the other two cases (Case 1 and Case 4), the cause of BCM could not be clearly determined, although the patients exhibited very unique genotypes.

Methods

The protocol for this study was approved by the Institutional Review Boards of The Jikei University of Medicine, Shiga University of Medical Science, Hamamatsu University School of Medicine, Kyoto University, Kindai University Faculty of Medicine, Nagoya University Graduate School of Medicine, Mie University Graduate School of Medicine, and Fujita Health University. The protocol adhered to the tenets of the Declaration of Helsinki, and informed consent was obtained from all participants.
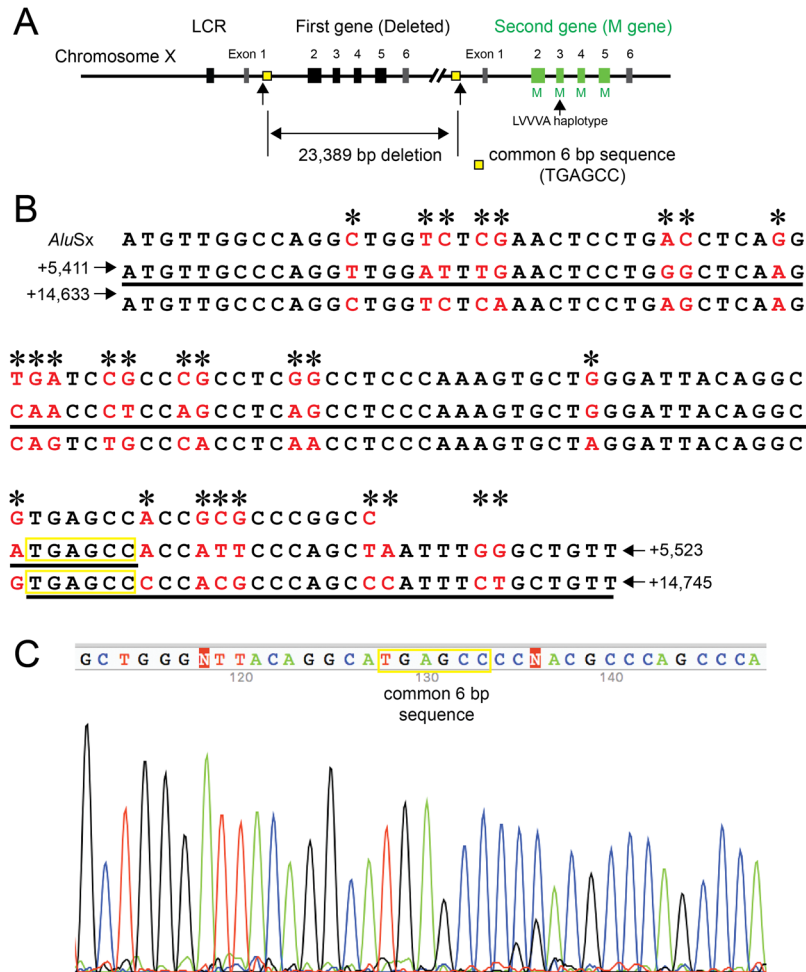
Figure 5. Genotype of Case 4. (**A**) Overview of the genotype of Case 4. Case 4 had a large deletion of 23,389 bp including exons 2–6 of the first gene. The intact LCR and second gene were present. The second gene had exons 2–5 of M type and exon 3 with LVVVA haplotype. (**B**) Upper row shows the consensus sequence for the left monomer of the *Alu* element (complementary sequence of AluSx). Middle row shows a part of intron 1 of the first gene (the number is from the 5′ splice site of intron 1). Lower row shows a part of the intergenic region (the number is from the stop codon in exon 6 of the first gene). The nucleotides differing among the regions are shown in red and by asterisks. The actual sequence obtained in Case 4 is underlined. (**C**) Partial sequence data around the breakpoint of the deletion. The breakpoint is somewhere in the common 6-bp sequence.

Participants. We recruited four unrelated Japanese male patients with BCM, whose diagnosis of BCM was made according to the findings reported[6,21]. In brief, all participants exhibited clinical findings of BCM, such as decreased visual acuity, severely impaired color discrimination in color vision tests, diminished L and M cone functions but retained S cone function in electroretinography, and an X-linked inheritance pattern in the family history. The detailed clinical findings are summarized in Supplementary Table S2.

Molecular genetic analysis. Genomic DNA was extracted from leucocytes in venous blood samples using a Gentra Puregene blood kit (Qiagen, Hilden, Germany). First and downstream genes of the L/M visual pigment gene array were separately amplified by long-range PCR using a QIAGEN LongRange PCR kit (Qiagen). Primers FG and E6R were used for the first gene, and primers DG and E6R were used for downstream gene(s) (Supplementary Table S1). The position of primers used in this study are schematically shown in Supplementary Fig. S3. Primer E6R was common to both genes, but primers FG and DG were designed specifically for the upstream region of each gene. The cycling parameters were: 93 °C for 3 min; 10 cycles of 93 °C for 30 s, 62 °C for 30 s, and 68 °C for 15 min; then 18 cycles of 93 °C for 30 s, 62 °C for 30 s, and 68 °C for 15 min, with a 20-s increment per cycle. The resulting PCR products were used as templates for sequencing the 'promoter + exon 1' and exons 2–6, including their adjacent introns, using a BigDye Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific, Waltham, MA, USA) and ABI 3130*xl* sequencer (Thermo Fisher Scientific). The primer pairs used for sequencing are listed in Supplementary Table S1.

The LCR, which is located about 3.5 kb upstream of the first gene, was amplified by PCR using primers LCRF and LCRR, and its nucleotide sequence was then determined. When deletion was suspected, multiple sets of primers were designed for PCR to determine the exact deletion breakpoints.

Array gene number was determined by promoter analysis, as previously described[22]. Briefly, the promoters were amplified by PCR using Takara *Taq* DNA polymerase (Takara Bio Inc., Kusatu, Japan) and the primer pair common to both genes (Supplementary Table S1). The 169-bp PCR product was digested with *Cfr*10I (Takara) and analyzed by electrophoresis on a polyacrylamide gel. The first gene promoter was expected to yield two DNA fragments (137 bp and 32 bp), whereas the downstream gene promoter was expected to yield three DNA fragments (97 bp, 40 bp, and 32 bp). Gene number was estimated from the fluorescence ratio ([137 bp + 97 bp]/137 bp). Gene Ladder Wide 2 was used as the size marker (Nippon Gene Co., Ltd., Toyama, Japan). Genomic DNAs from four subjects in which the L/M gene number was confirmed to be 1–4 by pulsed-field gel electrophoresis and Southern blotting[28] were used as control templates.

Promoter assay.   To evaluate the activity of the promoter region, a luciferase reporter assay was performed as previously described[29]. Briefly, the promoter region of interest (−190 to +41 from the cap site of the gene) was amplified by PCR using Phusion High-Fidelity DNA polymerase (New England BioLabs, Ipswich, MA, USA) and restriction site–tagged primers (*Nhe*I site upstream and *Hin*dIII site downstream). The PCR product was cloned between the *Nhe*I and *Hin*dIII sites of a luciferase reporter plasmid, pGL4.17 (Promega Corp., Fitchburg, WI, USA). The resulting plasmid was transfected into WERI Rb1 cells using X-tremeGENE 9 DNA transfection reagent (Sigma-Aldrich, St. Louis, MO, USA). Two days after transfection, the cells were collected and lysed using PicaGene Cell Culture Lysis Reagent Luc (Wako Chemicals, Osaka, Japan). Luciferase activity was measured using a luminometer (Lumicounter NU-2500, Microtech Co., Ltd., Funahashi, Japan) and PicaGene Luminescence kit (Wako). Transfection efficiency was monitored in cells co-transfected with a ß-galactosidase–encoding plasmid (Promega).

Analysis of M273K mutant pigments.   The cDNAs of human L and M pigments and respective hybrid pigment were tagged with the epitope sequence of the anti-bovine rhodopsin monoclonal antibody Rho1D4 (ETSQVAPA) at the C terminus and were inserted into the mammalian expression vector pMT4[30]. cDNAs harboring the mutation M273K were constructed using an In-Fusion cloning kit (Takara). For the spectral analysis, the plasmid DNA was transfected into HEK293 cells using the calcium-phosphate method[31]. After 2 days of incubation, the cells were collected by centrifugation and supplemented with 11-*cis*-retinal in buffer A (50 mM Hepes [pH 6.5] and 140 mM NaCl) to reconstitute the pigments. The reconstituted pigments were extracted using 0.75% CHAPS and 1 mg/mL phosphatidylcholine in buffer A. Absorption spectra of the extracted pigments were recorded at 0 °C using a Shimadzu UV-2450 spectrophotometer. The pigments were irradiated with orange light through an O58 cutoff filter (Toshiba, Tokyo, Japan) for 1 min. Difference spectra were calculated from spectra recorded before and after irradiation. For the western blot analysis, extracts from pigment-transfected or mock-transfected HEK293 cells were subjected to SDS-PAGE, transferred onto a polyvinylidene difluoride membrane, and probed with Rho1D4. Immunoreactive proteins were detected by ECL Western Blotting Detection Reagents (GE Healthcare, United Kingdom) and visualized by a luminescent image analyzer (LAS 4000mini, GE Healthcare). For the fluorescence microscopy analysis, pigment-transfected or mock-transfected HEK293 cells were seeded onto poly-L-lysine coated coverslips. After 24 h incubation, cells were fixed in cooled methanol for 5 min. After fixation, cells were washed three times in PBS and were incubated overnight with primary antibody, Rho1D4, in 10% normal goat serum at room temperature. Cells were washed three times in PBS and were incubated for 1 h with secondary antibody, Alexa Fluor 488 anti-mouse IgG, in 10% normal goat serum at room temperature. Cells were washed a final time and were mounted onto slides with home-made aqueous mounting media consisting of glycerol and polyvinyl alcohol.

References
1. Nathans, J., Thomas, D. & Hogness, D. S. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* **232**, 193–202 (1986).
2. Vollrath, D., Nathans, J. & Davis, R. W. Tandem array of human visual pigment genes at Xq28. *Science* **240**, 1669–1672 (1988).
3. Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B. & Hogness, D. S. Molecular genetics of inherited variation in human color vision. *Science* **232**, 203–210 (1986).
4. Nathans, J. *et al.* Molecular genetics of human blue cone monochromacy. *Science* **245**, 831–838 (1989).
5. McClements, M. *et al.* Variations in opsin coding sequences cause X-linked cone dysfunction syndrome with myopia and dichromacy. *Invest Ophthalmol Vis Sci* **54**, 1361–1369 (2013).
6. Berson, E. L., Sandberg, M. A., Rosner, B. & Sullivan, P. L. Color plates to help identify patients with blue cone monochromatism. *Am J Ophthalmol* **95**, 741–747 (1983).
7. Kellner, U. *et al.* Blue cone monochromatism: clinical findings in patients with mutations in the red/green opsin gene cluster. *Graefes Arch Clin Exp Ophthalmol* **242**, 729–735 (2004).
8. Nathans, J. *et al.* Genetic heterogeneity among blue-cone monochromats. *Am J Hum Genet* **53**, 987–1000 (1993).
9. Ayyagari, R. *et al.* Bilateral macular atrophy in blue cone monochromacy (BCM) with loss of the locus control region (LCR) and part of the red pigment gene. *Mol Vis* **5**, 13 (1999).
10. Ayyagari, R. *et al.* Spectrum of color gene deletions and phenotype in patients with blue cone monochromacy. *Hum Genet* **107**, 75–82 (2000).
11. Wang, C. *et al.* Novel *OPN1LW/OPN1MW* deletion mutations in 2 Japanese families with blue cone monochromacy. *Hum Genome Var* **3**, 16011 (2016).
12. Yatsenko, S. A. *et al.* High-resolution microarray analysis unravels complex Xq28 aberrations in patients and carriers affected by X-linked blue cone monochromacy. *Clin Genet* **89**, 82–87 (2016).
13. Wang, Y. *et al.* A locus control region adjacent to the human red and green visual pigment genes. *Neuron* **9**, 429–440 (1992).
14. Smallwood, P. M., Wang, Y. & Nathans, J. Role of a locus control region in the mutually exclusive expression of human red and green cone pigment genes. *Proc Natl Acad Sci USA* **99**, 1008–1011 (2002).

15. Michaelides, M. *et al.* Blue cone monochromatism: a phenotype and genotype assessment with evidence of progressive loss of cone function in older individuals. *Eye (Lond)* **19**, 2–10 (2005).
16. Gardner, J. C. *et al.* Blue cone monochromacy: causative mutations and associated phenotypes. *Mol Vis* **15**, 876–884 (2009).
17. Gardner, J. C. *et al.* Three different cone opsin gene array mutational mechanisms with genotype-phenotype correlation and functional investigation of cone opsin variants. *Hum Mutat* **35**, 1354–1362 (2014).
18. Ueyama, H. *et al.* Unique haplotype in exon 3 of cone opsin mRNA affects splicing of its precursor, leading to congenital color vision defect. *Biochem Biophys Res Commun* **424**, 152–157 (2012).
19. Reyniers, E. *et al.* Gene conversion between red and defective green opsin gene in blue cone monochromacy. *Genomics* **29**, 323–328 (1995).
20. Buena-Atienza, E. *et al. De novo* intrachromosomal gene conversion from *OPN1MW* to *OPN1LW* in the male germline results in Blue Cone Monochromacy. *Sci Rep* **6**, 28253 (2016).
21. Terasaki, H. & Miyake, Y. Association of acquired color vision defects in blue cone monochromatism. *Jpn J Ophthalmol* **39**, 55–59 (1995).
22. Ueyama, H. *et al.* An A-71C substitution in a green gene at the second position in the red/green visual-pigment gene array is associated with deutan color-vision deficiency. *Proc Natl Acad Sci USA* **100**, 3357–3362 (2003).
23. Winderickx, J. *et al.* Defective colour vision associated with a missense mutation in the human green visual pigment gene. *Nat Genet* **1**, 251–256 (1992).
24. Jagla, W. M., Jägle, H., Hayashi, T., Sharpe, L. T. & Deeb, S. S. The molecular basis of dichromatic color vision in males with multiple red and green visual pigment genes. *Hum Mol Genet* **11**, 23–32 (2002).
25. Hayashi, S., Ueyama, H., Tanabe, S., Yamade, S. & Kani, K. Number and variations of the red and green visual pigment genes in Japanese men with normal color vision. *Jpn J Ophthalmol* **45**, 60–67 (2001).
26. Ueyama, H. *et al.* Analysis of L-cone/M-cone visual pigment gene arrays in Japanese males with protan color-vision deficiency. *Vision Res* **44**, 2241–2252 (2004).
27. Greenwald, S. H., Kuchenbecker, J. A., Rowlan, J. S., Neitz, J. & Neitz, M. Role of a dual splicing and amino acid code in myopia, cone dysfunction and cone dystrophy associated with *L/M* opsin interchange mutations. *Transl Vis Sci Technol* **6**, 2 (2017).
28. Ueyama, H., Tanabe, S., Muraki-Oda, S., Yamade, S. & Ohkubo, I. Protan color vision deficiency with a unique order of green-red as the first two genes of a visual pigment array. *J Hum Genet* **51**, 686–694 (2006).
29. Ueyama, H. *et al.* Analysis of introns and promoters of L/M visual pigment genes in relation to deutan color-vision deficiency with an array of normal gene orders. *J Hum Genet* **54**, 525–530 (2009).
30. Oprian, D. D., Molday, R. S., Kaufman, R. J. & Khorana, H. G. Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proc Natl Acad Sci USA* **84**, 8874–8878 (1987).
31. Yamashita, T., Nakamura, S., Tsutsui, K., Morizumi, T. & Shichida, Y. Chloride-dependent spectral tuning mechanism of L-group cone visual pigments. *Biochemistry* **52**, 1192–1197 (2013).

Acknowledgements

Author Contributions

M.I. and H.U. performed the molecular genetic analyses in all four cases. K.H. and Y.H. performed part of the molecular genetic analysis in case 3. T.Y. and Y.S. performed analysis of M273K mutant pigments. S.K., M.I, T.H., and H.U. interpreted the data and wrote the manuscript. H.I., H. Kurahashi, H. Kondo, H.O., M.O., and T.N. assisted with data interpretation. S.K., T.H., K.K., S.U., and M.K. performed ophthalmic examinations at each institution. T.H. and H.U. designed and supervised the study. All authors have read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-29891-9.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.