tvst

Article

Noncanonical Splice Site and Deep Intronic *FRMD7* Variants Activate Cryptic Exons in X-linked Infantile Nystagmus

Junwon Lee[1,*], Han Jeong[2,3,*], Dongju Won[4], Saeam Shin[4], Seung-Tae Lee[4,5], Jong Rak Choi[4,5], Suk Ho Byeon[2,3], Helen J. Kuht[6], Mervyn G. Thomas[6,#], and Jinu Han[1,#]

[1] Institute of Vision Research, Department of Ophthalmology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, South Korea
[2] Brain Korea 21 Project for Medical Science, Yonsei University, Seoul, South Korea
[3] Institute of Vision Research, Department of Ophthalmology, Severance Hospital, Yonsei University College of Medicine, Seoul, South Korea
[4] Department of Laboratory Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, South Korea
[5] Dxome Co., Ltd. Seongnam-si, Gyeonggi-do, South Korea
[6] The University of Leicester Ulverscroft Eye Unit, Department of Neuroscience, Psychology and Behaviour, University of Leicester, RKCSB, PO Box 65, Leicester LE2 7LX, UK

Correspondence: Jinu Han, Institute of Vision Research, Department of Ophthalmology, Yonsei University College of Medicine, 211 Eonju-ro, Gangnamgu, Seoul 06273, Korea.
e-mail: jinuhan@yuhs.ac
Mervyn G. Thomas, Ulverscroft Eye Unit, Department of Neuroscience, Psychology and Behavior, University of Leicester, Robert Kilpatrick Clinical Sciences Building (Room 516, Level 5), Leicester, LE2 7LX, UK.
e-mail: mt350@le.ac.uk

Received: February 10, 2022
Accepted: May 29, 2022
Published: June 28, 2022

Keywords: infantile nystagmus syndrome; intronic variants; noncoding variants; *FRMD7*; exonic splicing enhancer; genome sequencing

Citation: Lee J, Jeong H, Won D, Shin S, Lee ST, Choi JR, Byeon SH, Kuht HJ, Thomas MG, Han J. Noncanonical splice site and deep intronic *FRMD7* variants activate cryptic exons in X-linked infantile nystagmus. Transl Vis Sci Technol. 2022;11(6):25, https://doi.org/10.1167/tvst.11.6.25

Purpose: We aim to report noncoding pathogenic variants in patients with *FRMD7*-related infantile nystagmus (FIN).

Methods: Genome sequencing ($n = 2$ families) and reanalysis of targeted panel next generation sequencing ($n = 2$ families) was performed in genetically unsolved cases of suspected FIN. Previous sequence analysis showed no pathogenic coding variants in genes associated with infantile nystagmus. SpliceAI, SpliceRover, and Alamut consensus programs were used to annotate noncoding variants. Minigene splicing assay was performed to confirm aberrant splicing. In silico analysis of exonic splicing enhancer and silencer was also performed.

Results: *FRMD7* intronic variants were identified based on genome sequencing and targeted next-generation sequencing analysis. These included c.285-12A>G (pedigree 1), c.284+63T>A (pedigrees 2 and 3), and c. 383-1368A>G (pedigree 4). All variants were absent in gnomAD, and the both c.285-12A>G and c.284+63T>A variants were predicted to enhance new splicing acceptor gains with SpliceAI, SpliceRover, and Alamut consensus approaches. However, the c.383-1368 A>G variant only had a significant impact score on the SpliceRover program. The c.383-1368A>G variant was predicted to promote pseudoexon inclusion by binding of exonic splicing enhancer. Aberrant exonizations were validated through minigene constructs, and all variants were segregated in the families.

Conclusions: Deep learning–based annotation of noncoding variants facilitates the discovery of hidden genetic variations in patients with FIN. This study provides evidence of effectiveness of combined deep learning–based splicing tools to identify hidden pathogenic variants in previously unsolved patients with infantile nystagmus.

Translational Relevance: These results demonstrate robust analysis using two deep learning splicing predictions and in vitro functional study can lead to finding hidden genetic variations in unsolved patients.

Introduction

FRMD7-related infantile nystagmus (FIN) is characterized by infantile-onset horizontal conjugate nystagmus associated with pathogenic variants in *FRMD7* (located at Xq26.2). FIN is fully penetrant in males with hemizygous *FRMD7* pathogenic variants; however, the penetrance is variable in heterozygous females.[1] In FIN, visual acuity is typically better than 20/40, with a normal retinal structure or only grade 1 foveal hypoplasia.[2–4] Single gene testing or massively parallel sequencing such as targeted panel next-generation sequencing (NGS) or exome sequencing (ES) have been used widely to identify causative variants in infantile nystagmus syndrome.[5–8] A prompt genetic diagnosis enables clinicians to provide visual prognosis and genetic counselling for future family planning.[9] Moreover, the identification of pathogenic variants in infantile nystagmus syndrome could prevent further unnecessary investigations such as electrodiagnostic testing and brain imaging in affected children.[5,6,10] In FIN, causative variants were detected in approximately 83% to 94% of familial cases.[11,12] To date, copy number variations and large structural variants were only identified in four families including exon 2 to 4 deletion, exon 2 to 12 deletion, and a large deletion including the entire *FRMD7*.[5,13–15] In addition, a deep intronic variant (c.285-118C>T) was reported, but to date, no functional assay has been conducted.[3]

Identifying hidden genetic variants in unsolved patients after targeted panel NGS or ES remains challenging for ophthalmologists and geneticists. Although genome sequencing (GS) enables us to search entire genomic regions in humans, there is no effective method to effectively interpret numerous noncoding variants. GS with complementary RNA sequencing increased the molecular diagnostic rate in Mendelian disorder.[16] However, obtaining proxy tissue for RNA sequencing is almost impossible for inherited eye diseases. Therefore, we searched causative variants via GS in unsolved families of presumed FIN after targeted panel sequencing and conducted minigene splicing assay to confirm aberrant splicing in these putative pathogenic variants.

Methods

This project was approved by the Ethics Review Board at Gangnam Severance Hospital, Seoul, South Korea, and met the Tenets of the Declaration of Helsinki. This study included four unsolved Korean families with presumed FIN. Our targeted gene panel (n = 595 genes) (Supplementary Table S1) also covered deep intronic regions known to harbor pathogenic variants associated with infantile nystagmus syndrome. These include variants such as c.659-113T>G, c.885+748G>A in *GPR143* and c.285-118C>T in *FRMD7* (Supplementary Fig. S1). However, no pathogenic variants were found in the coding regions and previously reported deep intronic regions in the four families with presumed FIN. In family 1 (II-1), the noncanonical splice site c.285-12A>G variant in *FRMD7* was detected using targeted panel sequencing; therefore, no further GS was undertaken. GS was performed in the other unsolved probands with X-linked infantile nystagmus after targeted panel NGS. Patients underwent complete ophthalmological examinations, which included the measurement of visual acuity and refractive errors, slit-lamp examination, and dilated fundus examination. Spectral domain optical coherence tomography (Heidelberg Engineering, Heidelberg, Germany) was performed when feasible. Eye movement recordings were obtained using video nystagmography (SLMED, Seoul, South Korea). Written informed consent was obtained from all participants. The study was approved by the Institutional Review Board at Gangnam Severance Hospital.

GS, Data Processing, and Variant Analysis

Genomic DNA was isolated from venous blood using a Qiagen Mini Blood Kit (Qiagen, Hilden, Germany). Samples were prepared by Library Preparation Enzymatic Fragmentation Kit 2.0 (Twist, San Francisco, CA), and paired-end sequencing was performed using Illumina NovaSeq 6000 platform. The quality metrics of GS are presented in Supplementary Table S2. Briefly, GS reads were aligned to Genome Reference Consortium human genome build 38 (GRCh38 or hg38) using the Burrows–Wheeler Aligner algorithms v0.7.17, followed by removal of duplicate reads, realignment of insertions and deletions, base quality recalibration, and variant calling using the Genome Analysis Tool Kit best practice workflow (version 4.1.2.0).[17] Split-read based detection of large insertions and deletions was conducted using Delly and Manta algorithms v1.6.0[18,19]; both results were finally crosschecked. Read-depth–based detection of copy number variation was conducted using cn.Mops.[20] A mobile transposable element (TE) analysis was performed using the Mobile Element Locator Tool algorithm v2.2.2.[21] From the pedigree information and clinical diagnoses, an X-linked mode of inheritance was assumed in each

family. Variant analyses were performed in candidate genes located in chromosome X. Therefore, we specifically searched for copy number variation, structural variants, mobile element insertion, deep intronic, and regulatory variants in genes associated with nystagmus in chromosome X. Supplementary Figure S2 details the candidate gene filtering approach, and variant prioritization in each pedigree. Two in silico tools (spliceAI and SpliceRover) were selected specifically to annotate all rare noncanonical or deep intronic variation in *FRMD7* for predicted splicing effects.[22,23] Briefly, spliceAI uses 32-layer deep neural network, predicts splicing from 10,000 nucleotides of flanking sequence. SpliceRover also uses convolutional neural network to predict aberrant splicing. The input for SpliceAI is vcf format, and input for SpliceRover is fasta format with minimum sequence length of 398 for humans. Additional splicing prediction tools (GeneSplicer, MaxEntScan, Splice Site Prediction by Neural Network, Splice Site Finder) were accessed through Alamut Visual 2.15 (Interactive Biosoftware, France). All variants were uploaded to LOVD3 (https://databases.lovd.nl/shared/genes/FRMD7).[24]

In Silico Analyses of the Exonic Splicing Enhancer (ESE) and Exonic Splicing Silencer (ESS) Motifs

The ESE finder 3.0 (http://rulai.cshl.edu/tools/ESE/) was used for ESE motif prediction and EX-SKIP (https://ex-skip.img.cas.cz) program was used to calculate the total number of ESSs, ESEs and their ratio. ESE finder and RESCUE-ESE were visualized using exon skipping in Alamut Visual v.2.15 (Interactive Biosoftware, Rouen, France).

Splicing Minigene or Midigene Construct, ARPE19 Cell Transfection, and Reverse Transcription Polymerase Chain Reaction (RT-PCR)

To assess the splicing patterns for the *FRMD7* intronic variants (c.285-12A>G, c.284+63T>A and c.383-1368A>G), three types of vectors were constructed for each variant. First, the positive control vector expressing the canonical transcript has the sequences of exons without introns. The second vector, a wild-type splicing minigene vector, has the sequence of normal exons and introns, and it is predicted to exhibit a similar expression pattern to the control vector by canonical splicing. Further, the third vector, a splicing minigene vector with variant, was created through the site-directed mutagenesis from the second vector and contained the sequences consisting of exons and introns with the variant of interest; this is a major vector of interest that can be used for estimating how the splicing pattern is altered by the variant. All vectors were expressed under the cytomegalovirus promoter.

To construct the positive control vectors, normal sequences of exons (exon 4–exon 5, exon 3–exon 4–exon 5, and exon 5–exon 6, respectively) of *FRMD7* were amplified from the complementary DNA (cDNA) of adult retinal pigmented epithelium 19 (ARPE19) cells with primer pairs possessing the target sequences of the restriction enzyme. Amplified products were subcloned into pcDNA3.1(+) vector (ThermoFisher Scientific, Waltham, MA) using restriction enzymes, NheI-XhoI or NheI-HindIII (New England BioLabs, Ipswich, MA), and T4 DNA ligase (Takara Bio, Shiga, Japan). The splicing minigene vectors were constructed by amplifying normal sequences of exons and introns (exon 4–intron 4–exon 5, exon 3–intron 3–exon 4–intron 4–exon 5, and exon 5–intron 5–exon 6, respectively) of *FRMD7* from genomic DNA of ARPE19 cells with primer pairs having target sequences of the restriction enzyme. In the case of the minigene with the exon 5–intron 5–exon 6 sequences for the c.383-1368A>G variant, the intron 5 sequence was extremely large (8007 bp), and there were technical difficulties in amplifying and creating a minigene construct. Therefore, the middle region of the intron (c.382+903∼6223 [5321 bp]) was excluded in the final minigene construct. It was confirmed that the splicing pattern was not altered by the exclusion of partial intron sequences. Amplified products were subcloned into pcDNA3.1(+) vector (ThermoFisher Scientific) using the restriction enzyme, NheI-XhoI or NheI-HindIII- XhoI (New England BioLabs) and T4 DNA ligase (Takara Bio). The intronic variants of interest (c.285-12A>G, c.284+63T>A and c.383-1368A>G) were introduced using the QuickChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instruction. Oligonucleotide sequences used for subcloning and site-directed mutagenesis are listed in Supplementary Table S3.

To analyze the splicing pattern of each intronic variant, RT-PCR was performed after transfection of the three types of vectors for each variant into ARPE19 cells. The expression of FRMD7 was confirmed sufficiently in ARPE19 cells (Supplementary Fig. S3). ARPE19 cells were cultured in Dulbecco's Modified Eagle Medium/F12 (1:1) medium (ThermoFisher Scientific), supplemented with 10% fetal bovine serum (ThermoFisher Scientific) and 1% penicillin/streptomycin (ThermoFisher Scientific). ARPE19 cells were seeded at 2×10^5 cells on a

35-mm culture plate. After 24 hours, 3 µg of each vector was transfected for 48 hours using Lipofectamine 2000 (ThermoFisher Scientific). The total RNA was purified using TRIzol (Invitrogen, Waltham, MA) with RNA Clean & Concentrator-5 including Dnase I (ZYMO RESEARCH, Irvine, CA) and reversed-transcribed using PrimeScript first strand cDNA Synthesis Kit (Takara Bio) for cDNA synthesis. RT-PCR was performed using Pfu-X DNA polymerase (Solgent, Daejeon, Korea) with specific primers. Amplified products were electrophoretically separated using a 1%, 3%, or 4% agarose gel, and the excised bands were analyzed via Sanger sequencing after extraction with MEGAquick-spin Plus Total Fragment DNA purification Kit (iNtRON Biotechnology, Seongnam, Korea) according to the manufacturer's instructions. All band sequences were confirmed through Sanger sequencing, and glyceraldehyde 3-phosphate dehydrogenase was used as a loading control. Oligonucleotide sequences used for RT-PCR are listed in Supplementary Table S3.

Results

Clinical Phenotypes

In all four families exhibiting infantile nystagmus, X-linked inheritance with incomplete female penetrance was observed. Convergence-induced dampening of nystagmus was noted in all patients, and all affected patients displayed no other neurological signs and symptoms. All affected patients had infantile-onset horizontal nystagmus. No transillumination defect of the iris or iris abnormalities were noted. Dilated fundus examinations revealed normal fundus pigmentation and spectral domain optical coherence tomography indicated normal foveal morphology. The measurement of visual acuity was possible in four patients, and the best-corrected visual acuity was 20/30 or greater in all patients, which was consistent with FIN. Brain magnetic resonance imaging was performed in one patient, and the result was normal. Details of clinical features are presented in Table 1.

In pedigree 1, putative pathogenic noncanonical splice site variant (NM_194277.3:c.285-12A>G) in *FRMD7* was identified after targeted panel NGS. Alamut 3/4 consensus approach (GeneSplice, MaxEntScan, Splice Site Prediction by Neural Network, and Splice Site Finder) revealed a loss of the canonical splice site acceptor site (Supplementary Fig. S4); however, it did not predict the gain of a new splice site acceptor. Human Splice Funder v3.1 revealed that there was no significant impact on splicing. However, both SpliceAI and SpliceRover predicted the gain of a splice acceptor site 1 bp downstream (c.285-11). This variant was not found in gnomAD v.3.1.2 (hg38) and v.2.1.1 (hg19). Although this variant was present as 2/132,345 minor allele frequency in BRAVO (https://bravo.sph.umich.edu/freeze5/hg38/), it was only reported as a heterozygous variant in females.

In pedigree 2, no candidate pathogenic variant was identified after targeted panel NGS. Therefore, GS was performed, and the GS analysis of the candidate genes revealed putative pathogenic variants in *FRMD7*. In pedigree 2, a novel deep intronic variant in intron 4 (NM_194277.3:c.284+63T>A) was identified. This variant was not observed in any GS control database (gnomAD v.2.1.1, v.3.1.2 and BRAVO). All

Table 1. Phenotype Summary of Patients With X-Linked Infantile Nystagmus

	Pedigree 1 (Proband, II-1)	Pedigree 2 (Proband, I-3)	Pedigree 2 (II-1)	Pedigree 3 (Proband, II-4)	Pedigree 4 (Proband, II-1)	Pedigree 4 (I-2)
Gene	*FRMD7*	*FRMD7*	*FRMD7*	*FRMD7*	*FRMD7*	*FRMD7*
Variant[a]	c.285-12A>G	c.284+63T>A	c.284+63T>A	c.284+63T>A	c.383-1368A>G	c.383-1368A>G
Age (y)	2	40	1	5	6	33
Nystagmus	1–2 Hz pendular nystagmus	Periodic alternating nystagmus	2 Hz RBJ nystagmus	Periodic alternating nystagmus	3 Hz LBJ nystagmus	Periodic alternating nystagmus
Face turn	No	No	No	No	Yes	R face turn 15°
Convergence dampening	Yes	Yes	Yes	Yes	Yes	Yes
Strabismus	Exotropia	Exotropia	Orthotropia	Exotropia	Orthotropia	Orthotropia
Visual Acuity	RE: CSM LE: CSM	RE: 20/25 LE: 20/30	RE: CSM LE: CSM	RE: 20/30 LE: 20/30	RE: 20/30 LE: 20/30	RE: 20/25 LE: 20/25
Refractive error	RE: –2.00 LE: –2.00	RE: –1.50 –3.00 180 LE: –1.50 –3.25 180	RE: +0.50 LE: +0.50	RE: plano LE: plano	RE: +0.50 –1.00 180 LE: +1.25 –0.75 180	RE: +1.25 –1.00 180 LE: –1.25 –1.00 180
Fundus	Normal	Normal	Normal	Normal	Normal	Normal
OCT	Not performed	Normal fovea	Not performed	Normal fovea	Normal fovea	Normal fovea
Systemic features	None	None	None	None	None	None
Brain MRI	Normal	Not performed	Not performed	Normal	Not performed	Not performed

CSM, constant, steady, and maintained fixation; LBJ, left beating jerk; LE, left eye; OCT, optical coherence tomography; RBJ, right beating jerk; RE, right eye.

[a] *FRMD7* NM_194277.3 transcript was used.

Table 2. Minor Allele Frequency, Conservation Scores, and In Silico Predictions of Presumed Pathogenic Noncoding Variants in *FRMD7*

Gene	*FRMD7*	*FMRD7*	*FRMD7*
Variant	c.285-12A>G	c.284+63T>A	c.383-1368A>G
Genomic position (hg19)	ChrX(GRCh37):g.131228179T>C	ChrX(GRCh37):g.131231231A>T	ChrX(GRCh37):g.131221430T>C
Genomic position (hg38)	ChrX(GRCh38):g.132094151T>C	ChrX(GRCh38):g. 132097203A>T	ChrX(GRCh38):g.132087402T>C
phyloMam[a]	0.044	0.644	0.346
phyloVert[b]	0.391	0.693	0.354
CADD[c]	22.1	8.884	9.095
FATHMM[d]	0.268	0.194	0.129
GnomAD v.2.1.1[e]	Not found	Not found	Not found
GnomAD v3.1.2[f]	Not found	Not found	Not found
BRAVO[g]	2/132345	Not found	Not found
KRGDB[h]	Not found	Not found	Not found
GeneSplicer	Acc-loss (–12 bp) (9.5 → 5.1)	Acc-gain 5.0 (–2 bp)	No impact
MaxEntScan	Acc-loss (–12 bp) (11.4 → 8.1)	Acc-gain 9.1 (–2 bp)	No impact
NNSPLICE	No impact	Acc-gain 1.0 (–2 bp)	No impact
SpliceSiteFinder-like	Acc-loss (–12 bp) (93.9 → 0)	Acc-gain 86.9 (–2 bp)	No impact
Human Splice Finder 3.1	No impact	Acc-gain (–2 bp) 56.95 → 84.82 (48.94%)	No impact
SpliceAI[i]	Acc-loss 0.55 (–12 bp) Acc-gain 0.17 (–1 bp) Don-loss 0.10 (–109 bp)	Acc-gain 0.60 (–2 bp) Don-loss 0.27 (63 bp) Don-gain 0.57 (–143 bp)	Acc-gain 0.19 (49 bp) Don-gain 0.19 (–30 bp)
SpliceRover[j]	Acc-gain 0.684 (–1 bp)	Acc-gain 0.953 (–2 bp)	Acc-gain 0.926 (49 bp) Don-gain 0.679 (–30 bp)

Acc, splice acceptor; Don, splice donor; NNSPLICE, Splice Site Prediction by Neural Network.

[a]Phylogenetic *P* value mammals (PhyloPMam_avg) values of >2.3 are predicted pathogenic.

[b]Phylogenetic *P* value vertebrates (PhyloVert_avg) values of >4 are predicted pathogenic.

[c]Combined Annotation-Dependent Depletion (CADD phred) v1.6 values of ≥15 are predicted pathogenic.

[d]Functional Analysis through Hidden Markov Models (FATHMM) of >0.5 are predicted pathogenic.

[e]Genome Aggregation Database v.2.1.1 (hg19).

[f]Genome Aggregation Database v3.1.2 (hg38).

[g]BRAVO variant browser (705 million variants observed in 132,345 deeply sequenced [>38×] genomes from the TOPMed [Trans-Omics for Precision Medicine] data freeze 8).

[h]Korean Reference Genome Database consisted of 1722 Korean individuals.

[i]SpliceAI v1.3 was used for annotation.

[j]SpliceRover: human donor and acceptor splice site prediction was used (approximately 820 nucleotide sequence around the variant was used for input).

in silico predictions strongly predicted the gain of a splice acceptor site 2 bp downstream (c.284+65), and SpliceAI predicted the inclusion of an aberrant exon of 142 bp. Segregation analysis revealed that the variant was found in the affected daughter. We re-analyzed the targeted panel NGS in previously unsolved presumed FIN, the same c.284+63T>A variant was identified in an additional unrelated family (pedigree 3).

GS in proband 4 also revealed a rare deep intronic variant in intron 5 of *FRMD7* (NM_194277.3:c.383-1368A>G). This deep intronic variant was also not found in any GS control database. In silico splice prediction scores using the Alamut 3/4 consensus approach did not flag this deep intronic variant as a pathogenic acceptor gain (Supplementary Fig. S3). SpliceAI predicted the gain of a splice acceptor site 49 bp upstream and the gain of a splice donor site 30 bp downstream, but both scores (0.19) were below the threshold (0.20, high recall; 0.50, recommended; and 0.80, high precision). However, SpliceRover strongly predicted the gain of a splice acceptor site 49 bp upstream and the gain of a splice donor site 30 bp downstream. All details of conservation scores, other in silico predictions, minor allele frequencies of the putative pathogenic noncoding variants are presented in Table 2. All predicted splicing models are depicted in Supplementary Figure S5.

In Vitro Minigene Splicing Assays

In the c.285-12A>G variant, the wild-type splicing minigene vector displayed canonical expression, with a 177 bp-sized transcript composed of spliced exons 4 and 5 (Figs. 1C, 1D, product B) similar to the positive control vector (Figs. 1C, 1D, product A). In contrast, the mutant splicing minigene vector expressed a slightly larger transcript (Fig. 1C, product C), containing an additional 11-bp sequence (TTTCTTTGCAG) located at the end of intron 4 (Fig. 1D, product C). The

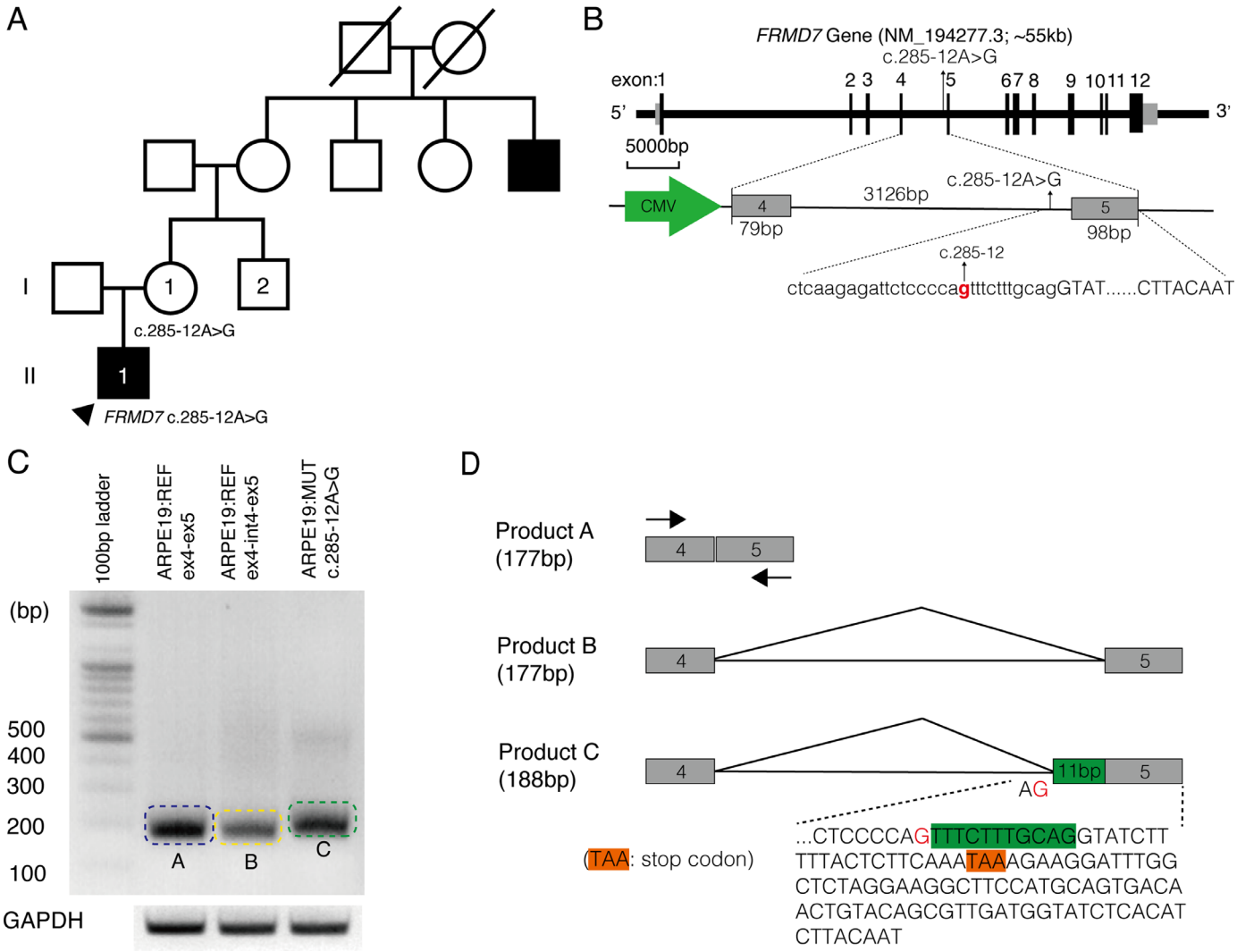

Figure 1. Segregation analysis in pedigree 1 and the functional validation of the *FRMD7* noncanonical splice site variant (c.285-12A>G). (A) Pedigree structure and segregation analysis for the c.285-12A>G variant. (B) Schematic representation of *FRMD7* and the location of the intronic variant (c.285-12A>G) (*top*). (*Bottom*) Schema includes splicing minigene construct. (C) RT-PCR of ARPE19 cells transfected with positive control vector, wild-type splicing minigene vector, and splicing minigene vector with variant (c.285-12A>G) of *FRMD7* vectors. *Blue boxes* denote control transcripts and normally spliced transcripts. *Red box* indicates aberrant transcript. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a loading control. (D) Schematic representation of the cDNA products obtained in minigene splicing assay by Sanger sequencing. *Green box* indicates the location and sequence of retained intron (11 bp) in Product C, the aberrant transcripts. *Orange box* indicates the predicted stop codon (TAA) generated by the frameshift. Black arrows indicate the location of RT-PCR primer pairs. CMV, cytomegalovirus promoter; bp, base pair.

c.285-12A>G variant created a new AG sequence, which is presumed to act as a new splicing acceptor site (Fig. 1D).

In the c.284+63T>A variant, the wild-type splicing minigene vector expressed two different transcripts. The smaller transcript comprised 220 bp and was composed of spliced exons 3, 4, and 5 (Figs. 2D, 2E, product B) similar to the positive control vector (Figs. 2D, 2E, product A); the larger transcript was 2343 bp in size and consisted of exon 3, intron 3, exon 4, and exon 5 without splicing out of intron 3, which was also observed in the case of the splicing minigene vector with the variant (Figs. 2D, 2E, product C). The splicing minigene vector with the variant expressed a slightly larger transcript (Fig. 2D, product D) than the 220-bp canonical transcript consisting of exons 3, 4, and 5. Sanger sequencing confirmed that the transcript was 283 bp in size and consisted of exon 3 and a 142-bp cryptic exon occurring in the intron 4 region and exon 5. Presumably, the c.284+63T>A variant formed a new splicing acceptor (AG) site together with the subsequent G and caused the skipping of exon 4, generation

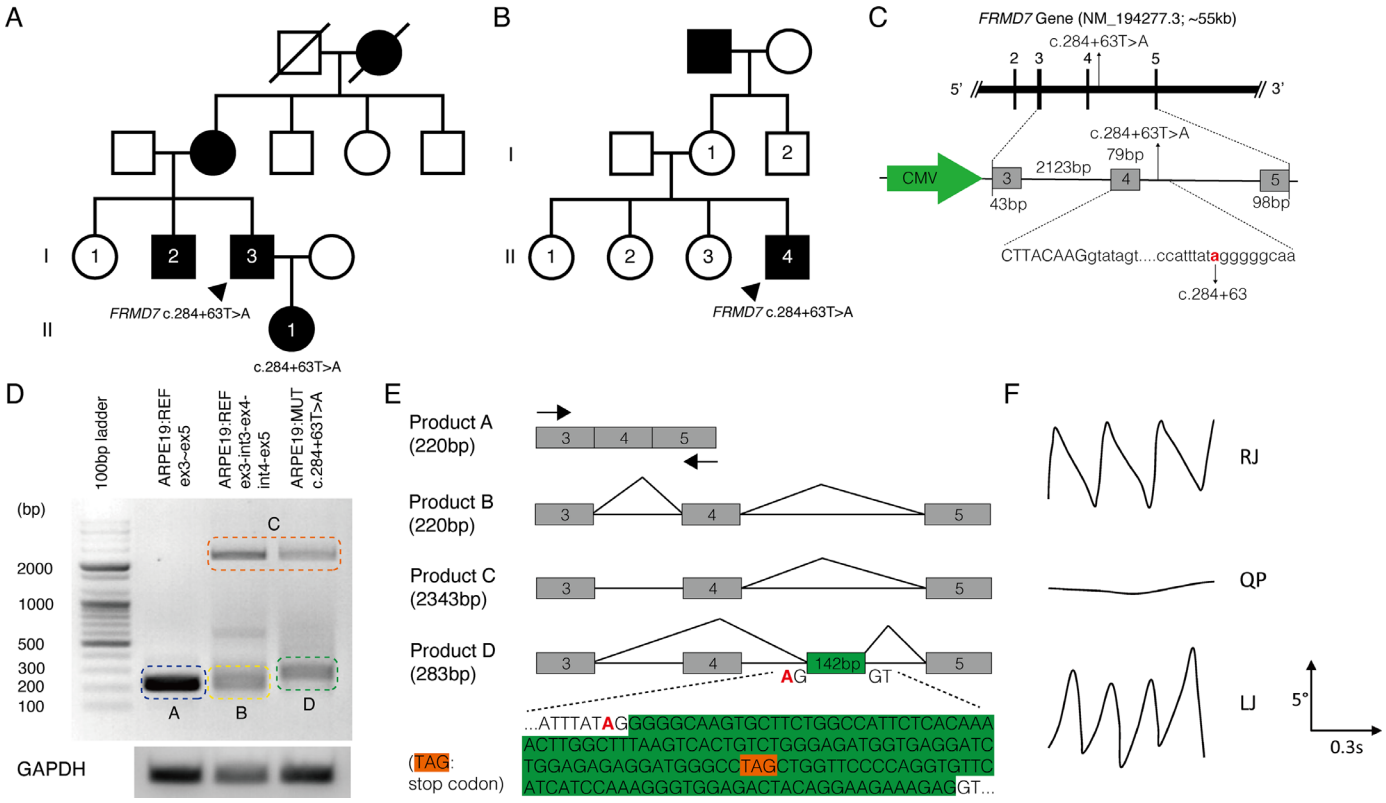

Figure 2. Segregation analysis in pedigrees 2 and 3, the functional validation of the *FRMD7* deep intronic variant (c.284+63T>A), and eye movement recordings in the proband. (A) Pedigree showed that the c.284+64 variant was detected in the proband and the affected daughter. (B) In unrelated family (pedigree 3), c.284+63T>A *FRMD7* variant was detected. (C) Schematic representation of *FRMD7* and the location of the intronic variant (c.284+63T>A) (*top*). (*Bottom*) Schema includes splicing minigene construct. (D) RT-PCR of ARPE19 cells transfected with positive control vector, wild-type splicing minigene vector, and splicing minigene vector with variant (c.284+63T>A) of *FRMD7* vectors. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a loading control. *Blue boxes* denote control transcripts and normally spliced transcripts. *Red box* indicates aberrant transcript. (E) Schematic representation of the cDNA products obtained in minigene splicing assay by Sanger sequencing. *Green box* indicates the location and sequence of cryptic exon (142 bp) in product D, the aberrant transcripts. *Orange box* indicates the predicted stop codon (TAG) generated by the frameshift. Black arrows indicate the location of RT-PCR primer pairs. (F) Traces for the horizontal eye movement recordings demonstrating the different phases seen in periodic alternating nystagmus (PAN). One cycle of PAN consists of right jerk (RJ) followed by a quiet phase (QP), left jerk (LJ) and another QP. Upward deflection of the horizontal position trace represents right-beating nystagmus and downward deflection represents left-beating nystagmus. CMV, cytomegalovirus promoter.

of the cryptic exon, and early termination of translation by the stop codon (TAG) within the cryptic exon (Fig. 2E).

In the c.383-1368A>G variant, the splicing minigene vector sufficiently expressed the canonical transcript composed of exons 5 and 6, the same as that of the positive control vector (Fig. 3C, products A and C). The splicing minigene vector with the variant, expressed a slightly larger transcript (Fig. 3C, product D) than the 213-bp canonical transcript consisting of exons 5 and 6 (Fig. 3C, products A and C). Sanger sequencing confirmed that the transcript was 293 bp in size, which consisted of exon 5 (98 bp), cryptic exon (80 bp), and exon 6 (115 bp) (Fig. 3D, product D).

Deep Intronic Variant c.383-1368A>G Creates a New ESE

The c.383-1368A>G variant was predicted to activate a splicing acceptor site 49 bp upstream (c.383-1319). We hypothesized that c.383-1368A>G variant could act by creating a splicing enhancer or by disrupting a splicing silencer. We searched the wild-type and c.383-1368A>G pseudoexon sequence for ESE and ESSs using ESE finder 3.0 program.[25] The results indicated that the c.383-1368A>G nucleotide increases the score of an SF2/ASF binding motif (AAGAG**A**A>AAGAG**G**A) from below the threshold (0.656) to a high score

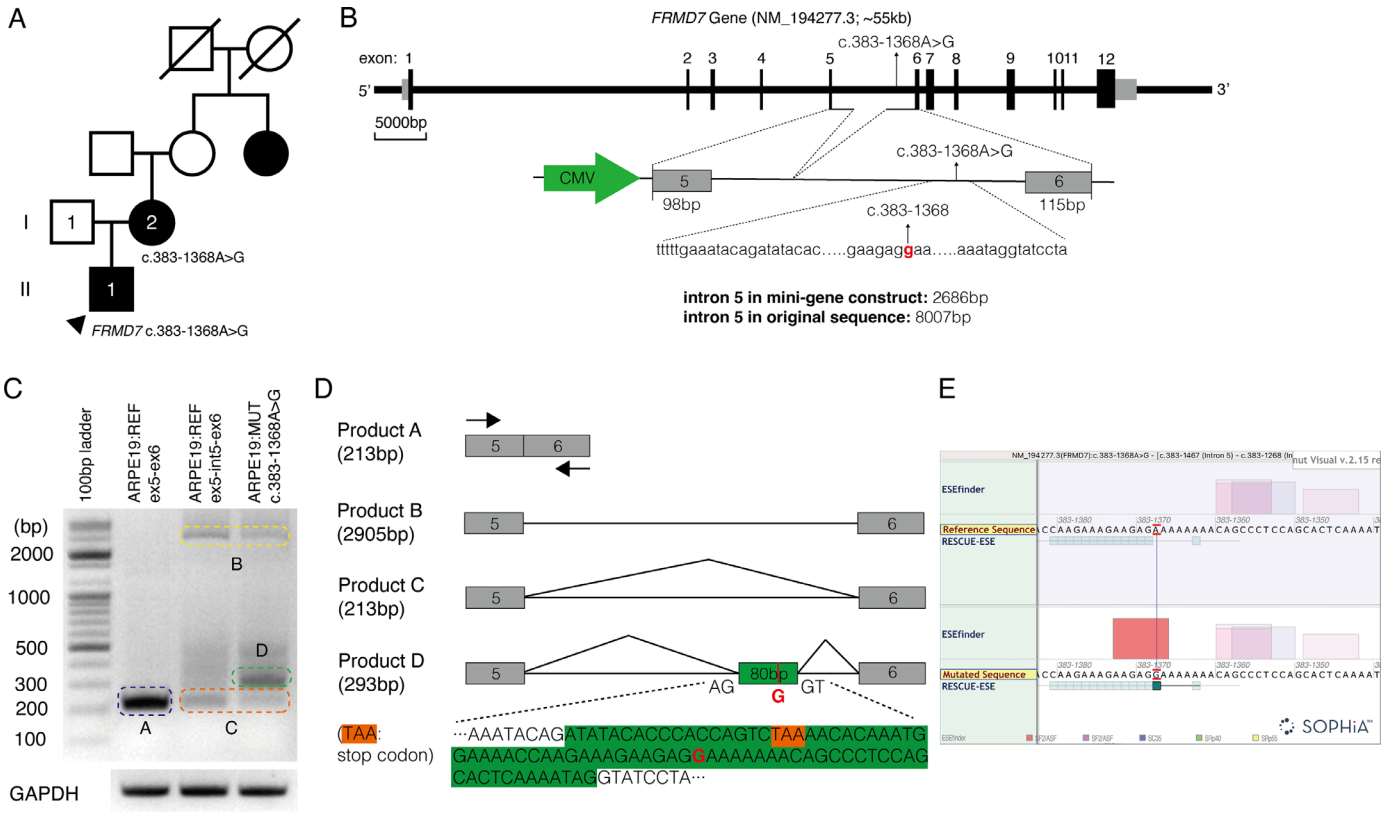

Figure 3. Segregation analysis in pedigree 4 and the functional validation of the *FRMD7* deep intronic variant (c.383-1368A>G). (A) Pedigree structure and segregation analysis for the c.383-1368A>G variant. (B) Schematic representation of *FRMD7* and the location of the deep intronic variant (c.383-1368A>G) (*top*). (*Bottom*) Schema includes splicing minigene construct. (C) RT-PCR of ARPE19 cells transfected with positive control vector, wild-type splicing minigene vector, and splicing minigene vector with variant (c.383-1368A>G) of *FRMD7* vectors. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a loading control. *Blue boxes* denote control transcripts and normally spliced transcripts. *Red box* indicates aberrant transcript. (D) Schematic representation of the cDNA products obtained in minigene splicing assay by Sanger sequencing**.** *Green box* indicates the location and sequence of cryptic exon (80 bp) in product D, the aberrant transcripts. *Orange box* indicates the predicted stop codon (TAA) generated by the frameshift. *Black arrows* indicate location of the RT-PCR primer pairs. (E) The c.383-1368A>G variant is predicted to create an ESE bound by SF2/ASF (SRSF1), which facilitates the recognition of a weak 5′ splice site 49 bp upstream to the variant site (c.383-1318), leading to pseudoexon inclusion. CMV, cytomegalovirus promoter.

(3.234) (Supplementary Table S4 and Supplementary Fig. S6). Other splicing enhancers such as SC35, SRp40, and SRp55 motif were not affected. Recognition of weak splice sites can either be facilitated by splicing enhancers or suppressed by splicing silencers; our data suggested that the c.383-1368A>G variant activates the pseudoexon by strengthening ESE, which potentially might act by binding of the splicing factor SF2/ASF.

Discussion

In this study, we identified three novel deep intronic and noncanonical splice site variants in patients with FIN. The noncanonical splice site variant (c.285-12A>G) was located in close proximity to the exon–intron boundary, and the deep intronic variant (c.284+63T>A) was also covered in the targeted panel NGS. Therefore, a reanalysis of existing sequencing data may uncover hidden genetic variations in unsolved but presumed FIN cases. However, the c.383-1368A>G was not covered in targeted panel NGS or ES. Further sequencing such as GS might be required to identify variants in these regions.

These noncoding variants are the first experimentally validated pathogenic noncanonical or deep intronic *FRMD7* variants. Precise phenotyping guided us to focus on genetic variations in *FRMD7* only, which led to discovery of novel noncoding variants in each of the four families. Interestingly, in silico prediction of the c.383-1368A>G variant is below the cut-off value of SpliceAI and no impact on Alamut

consensus approach. CADD score v1.6 (GRCh38) on the c.383-1368A>G variant, which uses MMSplice and SpliceAI, was also lower than the threshold.[26] However, SpliceRover strongly predicted an aberrant splicing gain at the 49 bp upstream from the mutation site. Although SpliceAI is regarded as the best single method to predict aberrant splicing, the combined use of different tools can further increase accuracy.[27,28] It is hypothesized that the c.383-1368A>G variant creates a novel binding site for the splicing enhancer, protein SRSF1 (SF2/ASF), leading to the inclusion of a pseudo-exon containing a premature stop codon. Although a minigene splicing assay with sequence analysis of the mutated transcript exactly matched aberrant splicing as predicted by combining deep learning splice prediction tools,[22,23] additional families would further support and validate our results.

The *FRMD7* gene is located on chromosome X, and its major isoform (NM_194277.3) contains 12 exons and 714 amino acids. The minor isoforms (NM_001306193.1) contains 699 amino acids and uses an alternate in-frame splice site (deletion of 45 bp) in the exon 4.[29] Although the precise role of alternative splicing has not been elucidated fully, the noncoding variants identified in this study likely activate aberrant splicing acceptor sites regardless of alternative splicing. Nonsense-mediated mRNA decay is likely to occur because all the variants remove more than 10% of the protein. The *FRMD7* knockout mouse model (*FRMD7^{tm}*) demonstrated a selective loss of horizontal direction selectivity and asymmetry of inhibitory inputs in horizontal direction selective ganglion cells.[30] It is hypothesized that impaired neuronal circuits and neurite outgrowths caused by *FRMD7* variants result in infantile onset horizontal nystagmus in our patients.[31]

Deep intronic variants and regulatory variants have accounted for hidden genetic variations in inherited retinal diseases.[32–35] The identification of pathogenic deep intronic variants has significant impact on precision medicine because deep intronic variants may be suitable candidates for CRISPR/Cas9 genome editing or antisense oligonucleotide therapy.[36,37] Although a previous study reported that analyzing at least 30 bp of the flanking exonic regions is recommended and targeted panel NGS and ES typically include up to 50 bp of the flanking exonic regions,[38] the c.284+63T>A variant can be detected using our targeted panel. Therefore, deep intronic variants can be identified by the reanalysis of sequencing data before submitting GS. If only one heterozygous variant was identified in the recessive inherited gene or the patient's phenotype is highly suggestive of a single genetic etiology in unsolved cases, searching rare variants in this flanking exonic region could be the appropriate subsequent step. Moreover, a TE analysis such as SCRAMble, which uses soft clipped reads aligned to TE sequences, can be performed in the targeted panel or exome.[39]

Targeted panel NGS and ES have been widely used in clinical practice. However, 30% to 50% of patients with inherited eye diseases remained unsolved after sequencing.[40,41] GS possesses some advantages over ES. Structural variants, deep intronic variants, regulatory variants, mitochondrial variations, variants in the GC-rich region, or TE insertions can be detected more easily via GS than targeted panel NGS.[42,43] A recent study evaluated the clinical usefulness of GS in rare Mendelian disorders in the UK 100,000 Genome Project, and pathogenic variants were identified in 25% of the 2183 probands.[43] Among 535 probands with positive results, noncoding single nucleotide variations or indel was identified in only 21 patients (4% of diagnoses), and 1% of diagnoses had variants in the mitochondrial genome. Forty-three patients (8% of diagnoses) had structural variants; however, most structural variants were exonic deletions except in four patients (noncoding exon 1 deletion or intronic deletions), which could not be detected by ES. Although we discovered noncoding pathogenic variants with deep learning–based annotation, it is possible with a focused approach that looked at only a few genes. Owing to the number of intronic variants generated from GS, it is not possible to perform splicing assays on all candidate variants. Our approach, although computationally demanding, uses a consensus-based approach to prioritize intronic variants for further study.

Both gene expression and splicing patterns are tissue specific; therefore, analyzing RNA from a clinically relevant tissue or cell type is essential.[33,44] Obtaining a clinically relevant tissue in inherited eye diseases is impossible because the retina and lens cannot be biopsied without functional compromise.[45] Therefore, a patient-derived retinal organoid transcriptome can be used to detect cryptic splice sites in genetically unsolved patients with inherited retinal diseases.[45] Dynamic retinal expression pattern of *FRMD7* during embryonic and fetal development has been reported using in situ hybridization techniques.[3,11] However, *FRMD7* expression is scant or absent in both human normal retina (https://ocular genomics.meei.harvard.edu/retinal-transcriptome/) and our in-house bulk RNA sequencing data of retinal organoids (Supplementary Fig. S7).[46] Therefore, we could not validate abnormal splicing patterns using this approach.

Instead, we used the conventional minigene splicing assay to confirm aberrant splicing. Our results may be limited by the difference in expression of splice factors and splice regulatory factors in different cell type and minigene construct length to a few exons (2 or 3). Moreover, splicing patterns in vivo may differ in the shorter size of minigene intron (2686 bp) compared with intron 5 of the genomic DNA (8007 bp). The splicing machinery is influenced by the genomic context where higher fidelity splicing outcomes are more likely with large genomic DNA[47,48]; therefore, we attempted to construct the minigene as large as possible.

In conclusion, to our knowledge, we report for the first time experimentally validated deep intronic variants contributing to FIN. This work expands the mutational landscape associated with this disorder and highlights a systematic approach in genetically evaluating unsolved cases of infantile nystagmus. Moreover, deep learning–based prioritization of noncoding variants and variation filtration using recent GS population databases enables improved analysis of GS and enhances our ability to detect causative variants in unsolved cases. GS requires more computational resources and bioinformatic post-processing to filter thousands of candidate variants, and typically fails to detect pathogenic variants in undiagnosed patients.[49] However, phenotype-driven GS analysis will enable us to focus on the variants of specific genes and will provide more opportunities to discover the missing heritability in inherited eye diseases.

Acknowledgments

The authors are grateful to the patients and the families who participated in this study.

Author Contributions:
JH and MGT contributed to the conception and the design of the study. JL, HJ, and SHB performed the experiments. DW, SS, S-TL, JRC, and JH analyzed and interpreted the data. JL and JH wrote the first draft of the manuscript. HJ, HJK, and MGT revised the manuscript. All authors contributed to the article and approved the submitted version.

Supported by intramural grants from the Korea Centers for Disease Control and Prevention (2018-ER6902-02), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1C1C1007965), and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2021R1I1A1A01045648). MGT is supported by the NIHR (CL-2017-11-003) and gratefully acknowledges the support from the Ulverscroft Foundation. HJK is supported by a Wellcome Trust Post-doctoral Fellowship.

Ethical Approval: This study was approved by the Institutional Review Board of Gangnam Severance Hospital and adhered to the tenets of the Declaration of Helsinki. Written informed consent was obtained for the participants.

Data Availability Statement: Data available on request from the authors.

Disclosure: **J. Lee**, None; **H. Jeong**, None; **D. Won**, None; **S. Shin**, None; **S.-T. Lee**, None; **J.R. Choi**, None; **S.H. Byeon**, None; **H.J. Kuht**, None; **M.G. Thomas**, None; **J. Han**, None

* JL and HJ contributed equally to the work and should be regarded as equivalent first authors.
# MGT and JH contributed equally to the work and should be regarded as equivalent corresponding authors.

References

1. Thomas MG, Maconachie G, Hisaund M, Gottlob I. *FRMD7-related infantile nystagmus. GeneReviews*. Seattle (WA): University of Washington, Seattle; 2009.
2. Han J, Lee T, Lee JB, Han SH. Retinal microstructures are altered in patients with idiopathic infantile nystagmus. *Graefes Arch Clin Exp Ophthalmol*. 2017;255(8):1661–1668.
3. Thomas MG, Crosier M, Lindsay S, et al. Abnormal retinal development associated with FRMD7 mutations. *Hum Mol Genet*. 2014;23(15):4086–4093.
4. Kuht HJ, Maconachie GDE, Han J, et al. Genotypic and phenotypic spectrum of foveal hypoplasia: a multicenter study. *Ophthalmology*. 2022;129(6):708–718.
5. Thomas MG, Maconachie G, Sheth V, McLean RJ, Gottlob I. Development and clinical utility of a novel diagnostic nystagmus gene panel using targeted next-generation sequencing. *Eur J Hum Genet*. 2017;25(6):725–734.
6. Rim JH, Lee ST, Gee HY, et al. Accuracy of next-generation sequencing for molecular diagnosis in patients with infantile nystagmus syndrome. *JAMA Ophthalmol*. 2017;135(12):1376–1385.

7. Jin S, Park SE, Won D, Lee ST, Han SH, Han J. TUBB3 M323V syndrome presents with infantile Nystagmus. *Genes (Basel)*. 2021;12(4):575.
8. Choi JH, Jung JH, Oh EH, et al. Genotype and Phenotype spectrum of FRMD7-associated infantile nystagmus syndrome. *Invest Ophthalmol Vis Sci*. 2018;59(7):3181–3188.
9. Moon D, Park HW, Surl D, et al. Precision medicine through next-generation sequencing in inherited eye diseases in a Korean cohort. *Genes (Basel)*. 2021;13(1):27.
10. Kuht HJ, Han J, Maconachie GDE, et al. SLC38A8 mutations result in arrested retinal development with loss of cone photoreceptor specialization. *Hum Mol Genet*. 2020;29(18):2989–3002.
11. Thomas MG, Crosier M, Lindsay S, et al. The clinical and molecular genetic features of idiopathic infantile periodic alternating nystagmus. *Brain*. 2011;134(Pt 3):892–902.
12. Dawar B, Kuht HJ, Han J, Maconachie GDE, Thomas MG. Clinical utility gene card for FRMD7-related infantile nystagmus. *Eur J Hum Genet*. 2021;29(10):1584–1588.
13. AlMoallem B, Bauwens M, Walraedt S, et al. Novel FRMD7 mutations and genomic rearrangement expand the molecular pathogenesis of X-linked idiopathic infantile nystagmus. *Invest Ophthalmol Vis Sci*. 2015;56(3):1701–1710.
14. Reynaert N, Braat E, de Zegher F. Congenital nystagmus and central hypothyroidism. *Int J Pediatr Endocrinol*. 2015;2015(1):7.
15. Fingert JH, Roos B, Eyestone ME, Pham JD, Mellot ML, Stone E. Novel intragenic FRMD7 deletion in a pedigree with congenital X-linked nystagmus. *Ophthalmic Genet*. 2010;31(2):77–80.
16. Lee H, Huang AY, Wang LK, et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med*. 2020;22(3):490–499.
17. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.
18. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865–2871.
19. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220–1222.
20. Klambauer G, Schwarzbauer K, Mayr A, et al. cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res*. 2012;40(9):e69.
21. Gardner EJ, Lam VK, Harris DN, et al. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res*. 2017;27(11):1916–1929.
22. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176(3):535–548.e524.
23. Zuallaert J, Godin F, Kim M, Soete A, Saeys Y, De Neve W. SpliceRover: interpretable convolutional neural networks for improved splice site prediction. *Bioinformatics*. 2018;34(24):4180–4188.
24. Fokkema I, Kroon M, López Hernández JA, et al. The LOVD3 platform: efficient genome-wide sharing of genetic variants. *Eur J Hum Genet*. 2021;29(12):1796–1803.
25. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res*. 2003;31(13):3568–3571.
26. Rentzsch P, Schubach M, Shendure J, Kircher M. CADD-splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med*. 2021;13(1):31.
27. Rowlands C, Thomas HB, Lord J, et al. Comparison of in silico strategies to prioritize rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. *Sci Rep*. 2021;11(1):20607.
28. Riepe TV, Khan M, Roosing S, Cremers FPM, t Hoen PAC. Benchmarking deep learning splice prediction tools using functional splice assays. *Hum Mutat*. 2021;42(7):799–810.
29. Li Y, Pu J, Liu Z, et al. Identification of a novel FRMD7 splice variant and functional analysis of two FRMD7 transcripts during human NT2 cell differentiation. *Mol Vis*. 2011;17:2986–2996.
30. Yonehara K, Fiscella M, Drinnenberg A, et al. Congenital nystagmus gene FRMD7 is necessary for establishing a neuronal circuit asymmetry for direction selectivity. *Neuron*. 2016;89(1):177–193.
31. Watkins RJ, Patil R, Goult BT, Thomas MG, Gottlob I, Shackleton S. A novel interaction between FRMD7 and CASK: evidence for a causal role in idiopathic infantile nystagmus. *Hum Mol Genet*. 2013;22(10):2105–2118.
32. Sangermano R, Bax NM, Bauwens M, et al. Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c.5461-10T→C mutation in Stargardt disease. *Ophthalmology*. 2016;123(6):1375–1385.

33. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med*. 2017;9(386):eaal5209.
34. Qian X, Wang J, Wang M, et al. Identification of deep-intronic splice mutations in a large cohort of patients with inherited retinal diseases. *Front Genet*. 2021;12:647400.
35. Sangermano R, Garanto A, Khan M, et al. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med*. 2019;21(8):1751–1760.
36. Kim J, Hu C, Moufawad El Achkar C, et al. Patient-customized oligonucleotide therapy for a rare genetic disease. *N Engl J Med*. 2019;381(17): 1644–1652.
37. Maeder ML, Stefanidakis M, Wilson CJ, et al. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat Med*. 2019;25(2):229–233.
38. Jamshidi F, Place EM, Mehrotra S, et al. Contribution of noncoding pathogenic variants to RPGRIP1-mediated inherited retinal degeneration. *Genet Med*. 2019;21(3):694–704.
39. Torene RI, Galens K, Liu S, et al. Mobile element insertion detection in 89,874 clinical exomes. *Genet Med*. 2020;22(5):974–978.
40. Zampaglione E, Kinde B, Place EM, et al. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. *Genet Med*. 2020;22(6):1079–1087.
41. Consugar MB, Navarro-Gomez D, Place EM, et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet Med*. 2015;17(4):253–261.
42. Carss KJ, Arno G, Erwood M, et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am J Hum Genet*. 2017;100(1):75–90.
43. Smedley D, Smith KR, Martin A, et al. 100,000 Genomes pilot on rare-disease diagnosis in health care - preliminary report. *N Engl J Med*. 2021;385(20):1868–1880.
44. Gonorazky HD, Naumenko S, Ramani AK, et al. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *Am J Hum Genet*. 2019;104(3):466–483.
45. Bronstein R, Capowski EE, Mehrotra S, et al. A combined RNA-seq and whole genome sequencing approach for identification of non-coding pathogenic variants in single families. *Hum Mol Genet*. 2020;29(6):967–979.
46. Farkas MH, Grant GR, White JA, Sousa ME, Consugar MB, Pierce EA. Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics*. 2013;14: 486.
47. Buratti E, Baralle M, Baralle FE. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res*. 2006;34(12):3494–3510.
48. Aparisi MJ, García-García G, Aller E, et al. Study of USH1 splicing variants through minigenes and transcript analysis from nasal epithelial cells. *PLoS One*. 2013;8(2):e57506.
49. Wu AC, McMahon P, Lu C. Ending the diagnostic odyssey—is whole-genome sequencing the answer? *JAMA Pediatr*. 2020;174(9):821–822.