RESEARCH ARTICLE OPEN ACCESS

Identification of Intronic Variants in *NDUFA3* as a Cause of Leigh Syndrome by Whole Genome Sequencing and RNA Sequencing

Kohta Nakamura,[1] Yoshihito Kishita,[1,2] Ayumu Sugiura,[1] Kokoro Ozaki,[1,3] Yukiko Yatsuka,[1] Naoyuki Matsumoto,[1] Atsuko Okazaki,[1] Holger Prokisch,[4,5] Koichi Maruyama,[6] Hiroyasu Iwasa,[7] Kei Murayama,[1,8] Hiroshi Matsumoto,[9] Akira Ohtake,[9,10] Yuichi Shiraishi,[11] and Yasushi Okazaki[1,3]

Correspondence
Prof. Okazaki
ya-okazaki@juntendo.ac.jp

Neurol Genet 2026;12:e200330. doi:10.1212/NXG.0000000000200330

Abstract

Background and Objectives

Leigh syndrome is an important manifestation of childhood-onset primary mitochondrial disease. Panel sequencing and whole exome sequencing are cost-effective for diagnosing mitochondrial diseases; however, more than half of mitochondrial disease cases remain genetically undiagnosed. This study aimed to demonstrate that combining whole genome sequencing (WGS) and RNA sequencing (RNA-seq) analyses can identify disease-causing variants that would otherwise be missed.

Methods

We performed WGS and RNA-seq on a patient with Leigh syndrome. Chromosomal phasing using Sanger sequencing of parental and patient blood samples was conducted to confirm compound heterozygous variants. RNA-seq data were analyzed for splicing abnormalities. Overexpression studies of wild-type *NDUFA3* in patient-derived fibroblasts were performed to assess restoration of mitochondrial function.

Results

We discovered compound heterozygous intronic variants (c.86-16_86-15del in intron2 and c.164-362G>A in intron3) of the *NDUFA3* gene. RNA-seq data analysis revealed intron retention and exonization in *NDUFA3*. Exonization was related to a variant involving the mobile element Alu that resulted in complex abnormal splicing events. Overexpression of wild-type *NDUFA3* restored mitochondrial dysfunction in patient-derived fibroblasts, confirming *NDUFA3* as a Leigh syndrome causative gene.

Discussion

This study highlights the importance of combining WGS and RNA-seq and provides new insights into detecting abnormalities in deep intronic regions, particularly those involving mobile elements, such as Alu. This approach can play a crucial role in identifying genetic variations and elucidating transcriptional control mechanisms that are not readily achieved by conventional methods, especially in the context of mobile element-induced complexities.

MORE ONLINE

Supplementary Material

Introduction

Mitochondrial diseases are among the most common congenital metabolic disorders and, in many cases, are intractable systemic diseases.[1,2] Despite advances in genetic diagnostics, the

[1]Diagnostics and Therapeutics of Intractable Disease, Juntendo University, Tokyo, Japan; [2]Department of Life Science, Faculty of Science and Engineering, Kindai University, Osaka, Japan; [3]Laboratory for Comprehensive Genomic Analysis, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; [4]Institute of Neurogenomics, Computational Health Center, Helmholtz Zentrum München, Neuherberg, Germany; [5]School of Medicine, Institute of Human Genetics, Technical University of Munich, Germany; [6]Department of Pediatric Neurology, Aichi Developmental Disability Center Central Hospital, Aichi, Japan; [7]Department of General Medicine, Aichi Developmental Disability Center Central Hospital, Aichi, Japan; [8]Department of Metabolism, Chiba Children's Hospital, Chiba, Japan; [9]Department of Pediatrics, Saitama Medical University, Saitama, Japan; [10]Department of Clinical Genomics, Saitama Medical University, Saitama, Japan; and [11]Division of Genome Analysis Platform Development, National Cancer Center Research Institute, Tokyo, Japan.

The Article Processing Charge was funded by the authors.

Glossary

AD = autosomal dominant; **CS** = citrate synthase; **FRASER** = find rare splicing events in RNA-seq; **GFP** = green fluorescent protein; **MAF** = minor allele frequency; **OCR** = oxygen consumption rate; **ORF** = open reading frame; **RNA-seq** = RNA sequencing; **WES** = whole exome sequencing; **WGS** = whole genome sequencing.

diagnostic rate for mitochondrial diseases remains relatively low, ranging from 30% to 40%.[3] This is partly due to limitations in detecting deep intronic variants that can cause splicing abnormalities, as well as other factors such as structural variants and copy number variations.[4,5] Leigh syndrome is one of the most common mitochondrial diseases and is a severe, chronic, progressive neurologic disorder that typically presents in infancy and early childhood. It is associated with diverse genotypes and complex multisystem clinical symptoms that result in a poor prognosis with death occurring within a few years after onset.[6] Leigh syndrome can be caused by genetic defects in either mitochondrial or nuclear DNA, with over 110 genes implicated thus far.[7] Genetic analysis of 166 Japanese Leigh syndrome patients diagnosed between 2007 and 2017 showed that *NDUFAF6*, *ECHS1*, and *SURF1* are common causative nuclear genes.[8] Mitochondrial respiratory chain complex I (NADH ubiquinone oxidoreductase) functions in electron transfer from NADH to ubiquinone through proton pumping, thereby establishing the electrochemical gradient across the inner mitochondrial membrane during ATP synthesis. Mitochondrial respiratory chain complex I composed of 44 subunits and can be divided into several structural modules.[9,10] The ND1 module is located at the heel of the L-shaped complex[11] and is formed from MT-ND1, *NDUFA3*, *NDUFA8*, and *NDUFA13*.[12] Of the genes that encode these proteins, pathogenic variants in *MT-ND1*, *NDUFA8*, and *NDUFA13* are known to cause mitochondrial disease.[13-15] Recently, Li et al. identified compound heterozygous variants in *NDUFA3* (NM_004542: c.10+1G>T and c.66_68del) in 3 siblings with Leigh syndrome, providing the first evidence linking this gene to the disease. Their study included a minigene assay demonstrating a splicing defect for the c.10+1G>T variant. However, the pathogenicity of the c.66_68del in-frame deletion was not fully elucidated, and the study did not definitively establish *NDUFA3* as a bona fide Leigh syndrome susceptibility gene according to stringent ClinGen criteria. This study builds on this foundation by providing comprehensive RNA sequencing (RNA-seq), robust functional rescue experiments, and detailed biochemical analyses, thereby strengthening the evidence supporting *NDUFA3* crucial role in mitochondrial function and Leigh syndrome pathogenesis.[16] Panel sequencing and whole exome sequencing (WES) are commonly used for clinical diagnosis. However, these methods do not comprehensively cover intronic regions, potentially leading to the oversight of pathogenic intronic variants. Such variants can affect normal gene expression and splicing patterns through mechanisms such as aberrant splicing or structural variants. A subset of these intronic regions involves Alu elements. Alu elements are short interspersed nuclear elements about 300 base pairs in length and are the most abundant repetitive elements in the human genome.[17] Because Alu elements are short interspersed repetitive sequences approximately 300 base pairs in length and are the most abundant in the human genome, frequently located within intronic and intergenic regions, Alu elements and variants are often missed by standard WES.[18] The application of whole genome sequencing (WGS) for genetic diagnosis is becoming more widespread and has led to the discovery of large or medium-sized structural variant that would have been missed by panel sequencing or WES. WGS can identify variants in the intergenic and intron regions. However, the number of variants in these regions is so large that it is not easy to narrow down those that cause disease. Moreover, in clinical settings, WGS is often performed using targeted panels or WES, which may only cover the intron-exon boundaries and a scant amount of flanking intronic sequence. Tools, such as SpliceAI, have made it possible to prioritize variants with putative effects on splicing.[19-21] There are various types of splicing aberrations, such as exon skipping, intron retention, and exonization; therefore, analysis using multifaceted methodologies is required. Bioinformatic tools, such as Find Rare Splicing Events in RNA-seq (FRASER) and Splicing Associated Variants Net (SAVNet), have been developed to efficiently detect splicing aberrations in RNA-seq data.[22-24]

Here, we introduce a novel approach combining short-read and long-read sequencing for both WGS and RNA-seq to evaluate splicing abnormalities caused by variants in deep intronic regions. We performed both WGS and RNA-seq on a patient with Leigh syndrome and detected splicing defects in *NDUFA3*. From these analyses, we identified a variant that causes intron retention and another variant that causes exonization. Haplotype phasing by Sanger sequencing of family members confirmed that the 2 splice variants were inherited in a recessive manner. The overexpression of wild-type *NDUFA3* in patient-derived fibroblast cells complemented mitochondrial dysfunction. From these observations, we conclude that *NDUFA3* is a novel causative gene for Leigh syndrome with recessive inheritance. Furthermore, our approach can enhance diagnostic rates by accurately detecting complex splicing abnormalities.

Methods

Ethics Statement

This study was approved by the ethics boards of Saitama Medical University, Chiba Children's Hospital, and

Juntendo University. Informed consent obtained from family members.

Mitochondrial Assessments

The activity of mitochondrial respiratory chain complexes I, III, and IV was measured using crude supernatant obtained following centrifugation at 600×*g* from mitochondria isolated from patients' cultured skin fibroblasts. The enzyme activity of each complex was shown as the percentage of the normal control mean relative to citrate synthase (CS).[25,26]

WES

The detailed protocol for WES has been described previously.[27-29] Indexed genomic DNA libraries were prepared from genomic DNA of the patient's fibroblast cells, and exomes were captured using the TruSeq Exome Enrichment Kit (Agilent Technologies) according to the manufactures' protocols. Sequencing was performed using 150-bp paired-end reads on a Hiseq2500 (Illumina). Total reads yielded 7.5 Gb of raw sequencing data. Quality of raw data was checked by FASTQC. After removing the low quality reads and adaptors, reads were mapped to the reference genome (GRCh38/hg38) with the Burrows-Wheeler Aligner, Picard, and SAMtools. GATK was also used for insertion and deletion realignment, quality recalibration, and variant calling. Detected variants were annotated using both ANNOVAR and custom Ruby scripts. A bioinformatics pipeline was used to process and analyze the sequencing data. Following quality control and adapter trimming, reads were aligned to the GRCh38/hg38 reference genome using BWA, and variants were called using GATK. Variants were annotated using ANNOVAR, incorporating multiple databases, including RefSeq, gnomAD, ExAC, 1000 Genomes, ClinVar, HGMD, MitoCarta, OMIM, dbSNP, and the 8.3KJPN database of Japanese Multi Omics Reference Panel (jMorp) from the Tohoku Medical Megabank Organization (ToMMo). To prioritize potentially pathogenic variants, we used a multistep filtering approach. First, variants with minor allele frequencies exceeding 0.5% in population databases (e.g., ExAC [AFR, AMR, EAS, FIN, NFE], GA100K, gnomAD, dbSNP, 1000 Genomes, and 8.3KJPN) were removed. For autosomal recessive inheritance patterns, we focused on variants present in affected individuals as either 2 heterozygous alleles (het/het) or 1 homozygous allele (hom). For autosomal dominant (AD) inheritance patterns, we retained heterozygous variants with a minor allele frequency (MAF) of 0.

Short-Read RNA Sequencing

RNA was purified from patient-derived skin fibroblasts by the Maxwell RSC simplyRNA Cells Kit and a Maxwell RSC Instrument (Promega). mRNA was enriched using oligo (dT) beads and rRNA removed using the Illumina Ribo-Zero kit. Isolated mRNA was fragmented randomly by adding fragmentation buffer from the Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus kit.[30] cDNA was synthesized from the mRNA template by random hexamers primers, followed by addition of a custom second-strand synthesis buffer (Illumina), dNTPs, RNase H, and DNA polymerase I to initiate second-strand synthesis. After a terminal repair, A ligation, and sequencing adaptor ligation, the double-stranded cDNA library was completed through size selection and PCR enrichment. Sequencing was performed using 150-bp paired-end reads on a NovaSeq6000 (Illumina). Total reads yielded 3.3 Gb(50 Mb) of raw sequencing data. FASTQ files were aligned to the GRCh38/hg38 genome by STAR. The outlier mRNA expression analysis was performed using OUTRIDER.[31] OUTRIDER was detected by comparing 99 suspected mitochondrial disease cases for which RNA-seq analysis was performed on the same platform. SAVNet, an existing bioinformatics tool for aberrant splicing, and FRASER were performed using RNA-seq data.[23,24] SAVNet is an intron retention file calculated from RNA-seq BAM files, variant call format and Splice Junction file (SJ.out) that describes variant information, In FRASER, RNA-seq BAM files were prepared and analyzed. From each tool, we attempted to detect Intron Retention in SAVNet and exonization or exon skipping in FRASER. Exonization is the formation of novel exons through splicing abnormalities. We also checked the number of reads in each sample from the BAM file, SJ.out.tab, using samtool, and confirmed that the results obtained with SAVNet and FRASER were not false positives. Variants and splicing aberrations were also evaluated by using SpliceAI and Pathogenicity predictor for Deep Intronic Variants causing Aberrant Splicing.[20,32]

Short-Read WGS

Genomic DNA was isolated from the patient derived fibroblast cells by phenol-chloroform extraction according to the standard protocol. The detailed protocol for WGS has been described previously.[33] WGS libraries were prepared from 100 ng of genomic DNA using an MGIEasy FS DNA Library Prep Kit v2.1 (MGItech) following the manufacturer's instructions. Paired-end 150-bp sequencing was performed on a DNBSEQ-T7 sequencer (MGItech) using a DNBSEQ-T7 High Throughput Sequencing set (PE150) v1.0 (MGItech). Total reads yielded 90 Gb of raw sequencing data. A bioinformatics pipeline was used, as described previously. SpliceAI was used to predict the impact of variants on splicing, including intronic regions, retaining those with a Δ-score of 0.2 or higher. A bioinformatics pipeline is the same as WES.

Long-Read Genome Sequencing

The adaptive sampling enables real-time selection and sequencing of DNA molecules in Oxford Nanopore sequencers.[34] Genomic DNA (1 ug) from the patient skin fibroblasts was fragmented with Covaris g-TUBE (Covaris) with a target range of 5–10 kb length, and further library preparation by SQK-LSK109 Kit was conducted according to the manufacturer's instructions. The library was sequenced on 1 R9.4.1 Rev D 106 flow cell on the Oxford Nanopore MINION sequencer. A target region with *NDUFA3* gene and its bilateral 100 kb margins (chr19:57080000-57289000, T2T) was used for enrichment. Successfully sequenced data were base-called using Guppy (version 6.4.6) with the super-accuracy mode

and then mapped to the reference genome (GRCh38/hg38) using Minimap2 (version 2.24-r1122),[35] variant-called with DeepVariant (version 1.3.0), and haplotyped with PEPPER-Margin-DeepVariant (version r0.8).[36]

Western Blotting Analysis

SDS-PAGE and Western blot were performed as previously described. Cells were directly harvested by 1× SDS sample buffer (62.5 mM Tris-HCl pH6.8, 2% SDS, 5% sucrose) containing protease inhibitor cocktail (Nacalai Tesque). After sonication, 2-mercaptoethanol was added to the samples to a final concentration of 5%. Prepared samples were denatured for 5 minutes at 95°C and separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis on the 5%–20% gradient polyacrylamide gel (2331730, ATTO). Proteins were transferred to PVDF membrane and subjected to Western blotting. Each antibody was obtained as follows: *NDUFA3* (17257-1-AP, Proteintech), Green Fluorescent Protein (GFP) (598, MBL International), and glyceraldehyde 3-phosphate dehydrogenase (G9545, Sigma-Aldrich).

Lentiviral Expression

Cells were cultured at 37°C and 5% CO2 in Dulbecco's modified Eagle's medium (DMEM 4.5 g/L glucose, Nacalai Tesque) supplemented with 10% fetal bovine serum. The open reading frame (ORF) of the *NDUFA3* gene (NM_004542.4) was amplified from cDNA of NHDF cells by PCR using Ex Premier DNA polymerase (TAKARA). *NDUFA3* or GFP was cloned into the CS-CA-MCS lentiviral vector with CAG promoter for mammalian cell expression and blasticidin resistance using the In-Fusion HD Cloning Kit (Clontech Laboratories, Inc). Lentiviral supernatants were prepared from HEK293FT cells transfected with ViraPower Packaging vectors (pLP1, pLP2, pLP/VSVG; Invitrogen) and a pCS-CA-ORF (candidate gene) vector using Lipofectamine 2000 (Invitrogen). Collected supernatants were infected to patient derived fibroblast cells in antibiotics-free DMEM supplemented with 10 μg/mL polybrene for constitutive expression of *NDUFA3* or *GFP*.

Oxygen Consumption Rate

One day before assay, cells were seeded in an XF24 cell culture microplate (100882-004, Agilent) at 3×10^4 cells/well. During assay, cells were treated with 1 μM oligomycin, 1 μM FCCP, and 0.5 μM rotenone/antimycin A. At the end of run, cells were fixed with 4% paraformaldehyde containing Hoechst33342 (19172-51, Nacalai Tesuque) for 15 minutes at 37 °C. The number of cells were counted using an all-in-one fluorescence microscope (BZ-X800, KEYENCE) equipped with an image cytometer module (BZ-H4XI, KEYENCE). Oxygen consumption rate (OCR) was normalized with the number of cells.

Data Availability

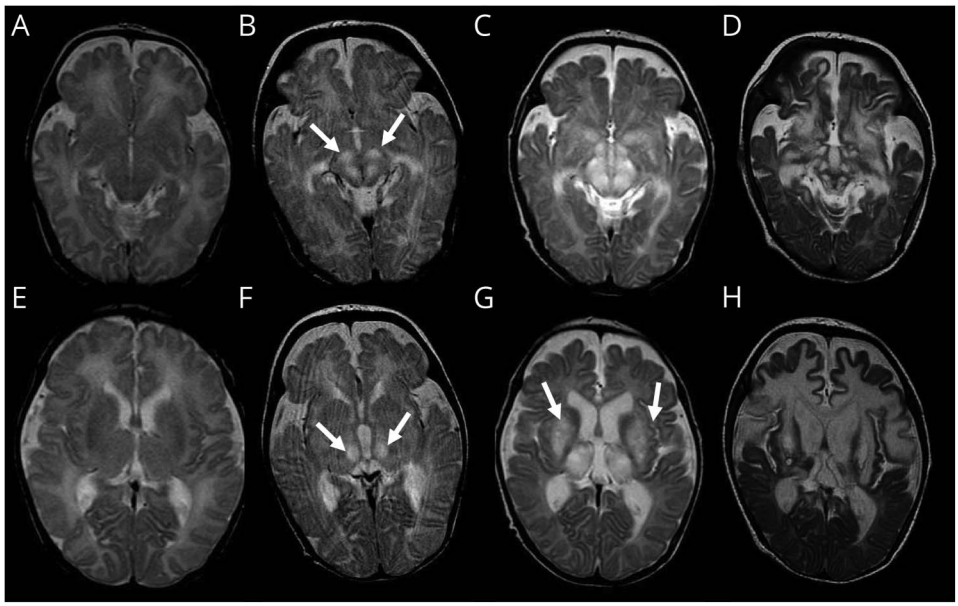
The data sets generated and/or analyzed in this study are available from the corresponding author on reasonable request.

Results

Case Report

This male patient was born weighing 2,900 g at 40 weeks of gestation. After experiencing neonatal asphyxia with an Apgar score of 2 at 5 minutes, the patient resumed breathing spontaneously. He had difficulty in oral feeding and required tube feeding when he left the hospital at 23 days after birth. Laboratory findings on admission revealed elevated blood lactate (29.9 mg/dL; reference range 4–14 mg/dL) and pyruvate (1.16 mg/dL; reference range 0.3–0.9 mg/dL) levels, with a lactate-to-pyruvate (L/P) ratio of 25.8 (reference range 7–20), suggesting a potential mitochondrial respiratory chain disorder. Very long chain fatty acids and arylsulfatase A levels were within normal ranges. Urinary organic acid analysis did not demonstrate patterns specifically associated with metabolic disorders. Blood amino acid analysis showed a mild elevation in alanine levels, not exceeding 3 times the upper limit of normal. Initially, the patient's clinical presentation, including dysphagia, respiratory distress, and movements suspicious for seizures, was attributed to neonatal asphyxia. However, the biochemical profile, particularly the elevated L/P ratio, prompted further investigation into possible mitochondrial disease and the patient's course was closely monitored with this consideration in mind. At the time of discharge, an MRI scan of the brain showed no obvious abnormality (Figure 1, A and E). After discharge from a hospital, the patient presented abnormal eye movement such as sunset phenomenon and seizures of leg twitching and body stiffness. Six weeks after discharge, he was readmitted to the intensive care unit due to prominent respiratory acidosis and hyponatremia (125 mEq/L; reference range 136–145 mEq/L). He was intubated and mechanical ventilation was started. MRI of the brain showed abnormalities in the bilateral midbrain tegmentum and thalamus (Figure 1, B and F). Sequential brain MRI presented rapidly progressive multiple brain lesions containing brain stem, basal ganglia, thalamus, and subcortical white matter of frontal lobes (Figure 1, C, D, G, H). At 3 months, these brain lesions on MRI showed liquefaction degeneration, and thereafter, prominent cerebral atrophy was proceeded. After that, he showed severe neurologic and multiple organ disorders with spastic quadriplegia, severe intellectual disability, neurogenic bladder, chronic type 2 respiratory failure, dilated cardiomyopathy, and type 2 diabetes mellitus. At 16 years, the patient showed cardiac involvement on preoperative echocardiography during tracheostomy tube replacement: reduced left ventricular ejection fraction (LVEF 34%) and ventricular septal hypertrophy were observed, with normal NT-proBNP (11 pg/mL; ref: 0–55). Based on these findings, he was diagnosed with transition from hypertrophic to dilated phase cardiomyopathy associated with mitochondrial disorder. His anthropometrics remained stable (height 134.0 cm, weight 32.8 kg; BMI 18.3 kg/m^2). The final hospitalization for septic shock required vasopressor support, but acute cardiac functional assessment was not performed, with a last recorded weight of 30.9 kg. He died of cardiac arrest at 17 years. Analysis of mitochondrial respiratory chain complex

Figure 1 MRI Imaging Results



T2-weighted MRI images at the levels of and midbrain and superior colliculus(A–D) and thalamus(E–H). There was no obvious abnormal intensity in the brain parenchyma at day-14 (A and E). At 1 month, symmetrical high intensity areas (HIA) were seen in midbrain tegmentum and thalamus (B and F). At 2 months, bilateral midbrain and thalamic lesions became extensive(C) and new lesions appeared in lens nucleus (G, arrows). At 3 months, previous lesions in thalamus and midbrain exhibited cystic changes, and new lesions appeared in bilateral putamen. Subcortical white matter in frontal lobes shows markedly extensive cystic changes (D and H).

enzyme activity revealed complex I defects in skin fibroblasts (I/CS: 0.0395, normal range: 0.267–0.792).

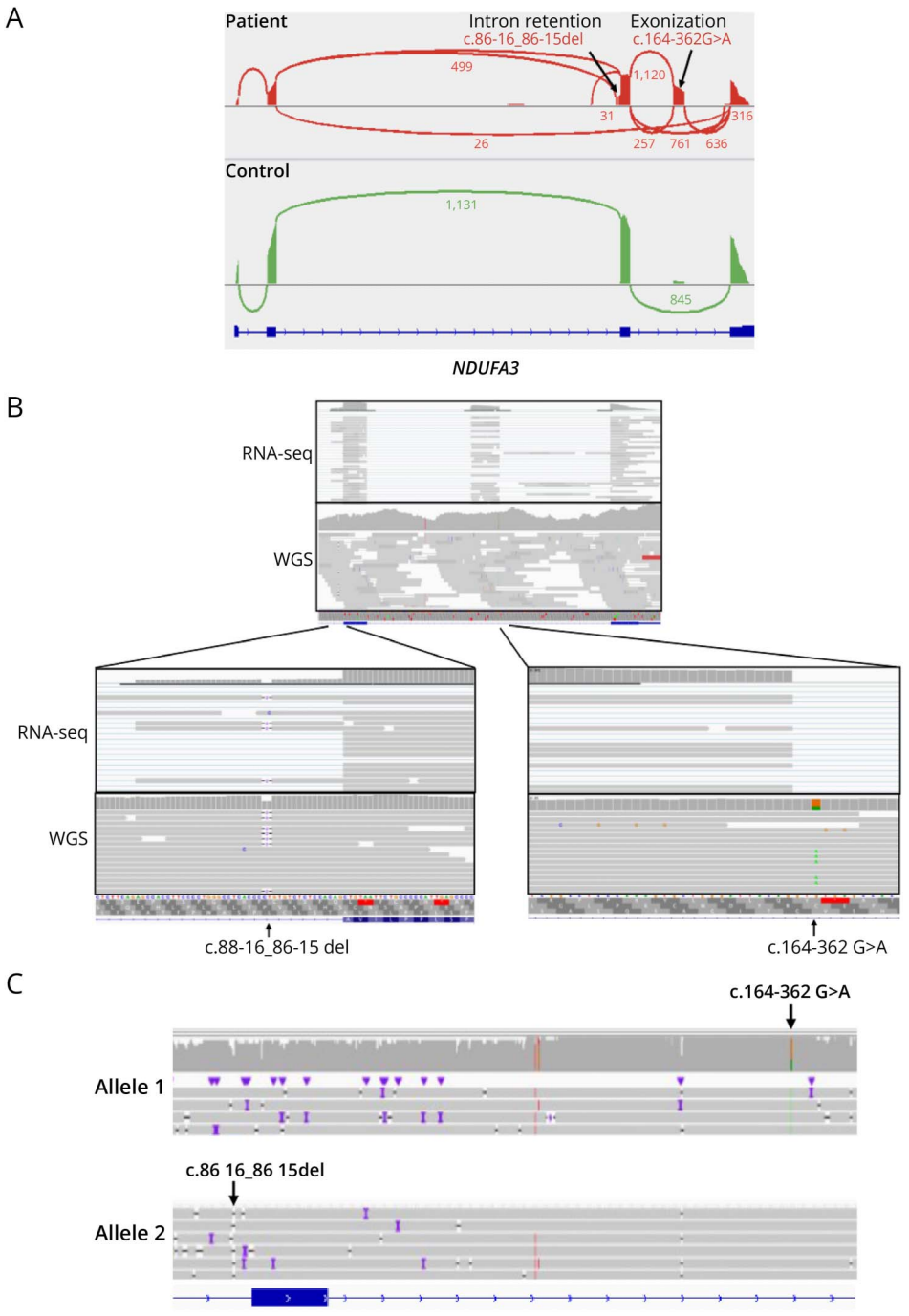
Identification of Splicing Abnormalities in Intronic Regions and Elongated Transcripts by Short-Read RNA Sequencing

We performed WES on the patient, but no variants associated with Leigh syndrome or other neurodevelopmental disorders were identified (eTable 1). We then attempted to detect aberrant gene expression by performing short-read RNA sequencing (SR RNA-seq) using skin fibroblasts isolated from the patient. First, OUTRIDER, which can detect gene expression outliers in a patient population, was used to search for candidate causative genes. Analysis of a population, including 99 cases of mitochondrial disease and suspected mitochondrial disease by OUTRIDER, did not identify any genes with significantly decreased expression in the patient. It is important to note that OUTRIDER detects statistical outliers by considering all genes in the data set simultaneously, rather than analyzing specific gene sets individually. Next, we searched for RNA splicing abnormalities in the patient using FRASER and SAVNet, which can detect various types of splicing changes from RNA-seq data of a patient population. Among the 99 cases, FRASER and SAVNet detected splicing abnormalities in 1230 and 335 genes, respectively (eFigure 1). Deep intronic regions were defined as intronic regions that are at least 50 bp away from the donor and acceptor splice sites.[32] Focusing on these deep intronic regions, our analysis identified 121 genes with splice abnormalities across all cases. FRASER and SAVNet analysis identified 2 splicing abnormalities in *NDUFA3*, which is listed in MitoCarta as a gene encoding a component protein of mitochondrial respiratory chain complex I. FRASER predicted the splicing abnormality in *NDUFA3* to cause exonization, in which a portion of the deep intronic region was exonized, while SAVNet predicted intron retention (Figure 2A). FRASER analysis for the patient identified a cryptic exon between exons 3 and 4 of *NDUFA3*. SAVNet analysis identified an intron extension on the 5′ side of exon 3 in *NDUFA3*. In addition to these splicing abnormalities, several other minor splicing abnormalities were observed in the vicinity of exons 3 and 4.

Identification of Splice Variants by Short-Read WGS and Long-Read Genome Sequencing

We did not detect any exon variants in *NDUFA3* by WES; therefore, we assumed that the causative variant was present in a deep intronic region. To search for single-nucleotide variants and small indels and also for larger structural variations, we performed short-read and long-read genome sequencing (SR WGS and LR WGS). We performed LR WGS sequencing by adaptive sampling on nanopore technology to focus on the *NDUFA3* gene. LR WGS did not detect any structural variations in *NDUFA3*. However, both SR WGS and LR WGS identified c.86-16_86-15del in the intronic region near exon 3 and c.164-362G>A at a cryptic exon (Figure 2, B and C). The minor allele frequencies for c.86-16_86-15del were 0.000033 in jMorp (60KJPN) and 0.000526 in gnomAD (v4.1.0), and for c.164-362G>A, they were 0.000450 in jMorp and not reported in gnomAD. Splicing abnormalities for each variant were predicted using SpliceAI, with c.86-16_86-15del having a Δ-score of 0.24 for acceptor gain, and c.164-362G>A having a Δ-score of 0.40 for donor gain. Analysis of Sashimi plots confirmed that the c.86-16_86-

Figure 2 Splicing Aberrations in *NDUFA3* Gene and Their Relationship to Identified Variants



(A) Sashimi plot of SR RNA-seq data indicates patient-specific splicing aberrations. Intron retention around exon3 and exonization in deep intronic regions between exons 3 and 4 were observed. (B) Diagram of the relationship between variants and splicing abnormalities, showing SR RNA-seq and SR WGS side by side. An intron retention is related with c.86-16_86-15del in exon3 and an exonization with c.164-362G>A. (C) Nanopore adaptive sampling of LR WGS revealed that the variants responsible for each splicing abnormality were different alleles and shows no structural variations around these variants. Purple arrowheads indicate insertions, while purple bars represent deletions in the sequencing output.

15del variant resulted in intron retention, while the c.164-362G >A variant led to the formation of a novel exon. These findings, combined with the MAF data, support the prioritization of these variants as potential causes of splicing abnormalities. These scores, combined with allele frequency data, supported the prioritization of these variants as potential causes of splicing abnormalities. Given the challenges in evaluating splicing abnormalities in deep intronic regions, we used SpliceAI scores to identify potential causative variants. We considered variants with a SpliceAI score of 0.2 or higher as candidates for further investigation. The use of SpliceAI for prioritizing deep intronic variants is crucial for detecting splicing alterations that may not be apparent through conventional variant analysis methods. LR WGS showed each variant to be located in a different allele (Figure 2C). This result is consistent with RNA-seq, which also showed that exon 3 retention and the cryptic exon were present in different transcripts. For c.86-16_86-15del, a two-base deletion in the

elongated exon was identified in the RNA-seq data, resulting in a frameshift and the introduction of a premature stop codon that truncates the protein at 58 amino acids, compared with the full-length *NDUFA3* protein of 84 amino acids. For c.164-362G>A, which is located 3 bp from the 3′ side of the cryptic exon and was thought to be involved in exonization, the inclusion of the cryptic exon also caused a frameshift, leading to a truncated protein of 71 amino acids. These changes indicate that both variants are likely to cause loss-of-function of the *NDUFA3* protein. It is conceivable that each variant would introduce a premature termination codon for each allele. Detailed analysis of SR RNA-seq results from the patient revealed complex splicing patterns, as shown in Figure. 3A. Detailed analysis of the exonization caused by c.164-362G>A revealed the involvement of AluY (Figure 3B). The insertion of an Alu sequence is closely related with this kind of exonization event.[37] In particular, Alu sequences that are inserted into introns in the opposite orientation to that of the gene are prone to exonization. The addition of a single-nucleotide substitution to the normal inverted Alu sequence increases its splicing activity. Exonization was likely caused by this phenomenon in this patient. This discovery is an example of how deep intronic variations can cause complex splicing abnormalities. To confirm the generality of this finding, we re-examined the data from fibroblasts derived from the 99 patients of our analysis cohort. We identified splicing abnormalities involving AluY in the exonization of the *NDUFA3* gene (eFigure 2). We confirmed segregation of the variants by Sanger sequencing of parental DNA samples, demonstrating that the c.86-16_87-15 deletion was inherited from the mother, while the c.164-362G>A variant was inherited from the father (Figure 4A). Both variants produce splicing abnormalities, resulting in frame shifts in their respective alleles (Figure 3A).

Functional Experiments Confirm *NDUFA3* to Be the Causative Gene of Leigh Syndrome

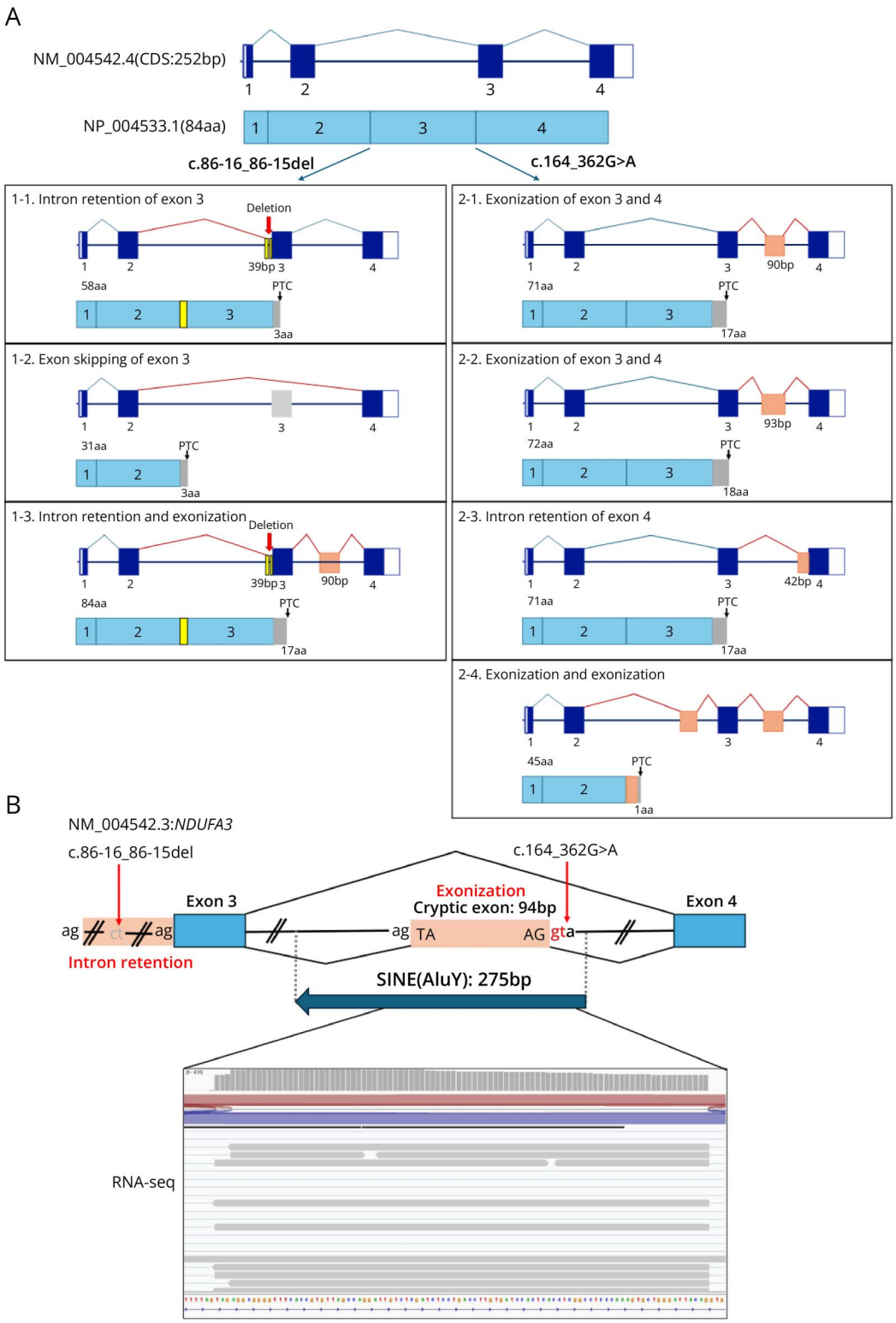
Given that only bioinformatics analysis could not be fully explained, we validated the pathogenicity of variant by experiments. OCR measurements showed that lower respiration of patient-derived fibroblasts (F78) than control skin fibroblasts (NHDF), indicating the deleterious effect of variant on mitochondrial functions. In addition, owing to respiratory chain defects in patient fibroblasts, we evaluated if *NDUFA3* expression through lentiviral vector could restore mitochondrial function (Figure 4B, eFigure 3).To further confirm the pathogenicity of variant, we conducted the rescue experiments where patient-derived fibroblasts were infected with lentiviruses expressing *NDUFA3* or GFP as a control. Western blot analysis showed that the level of *NDUFA3* protein was significantly decreased in patient cells (Figure 4C, lane 1 and 2) and restored by expressing *NDUFA3* (Figure 4C, lane 1 and 3). Consistently, OCR was also significantly improved in patient-derived fibroblasts infected with lentivirus expressing *NDUFA3* (Figure 4D).

On the basis of the RNA sequencing, WGS, and rescue experiment results, and the phenotype of the patient, we concluded that their mitochondrial disease resulted from splicing defects caused by variants in the deep intronic region of *NDUFA3*. We have reviewed the ACMG guidelines and determined that those 2 variants were Pathogenic because they meet the criteria for PS3 and PVS1. In addition, we applied PP4 as supporting evidence due to the biochemical defect observed in fibroblasts, which aligns with the phenotype reported by Li et al. and is further validated by our rescue experiments demonstrating a clear genotype-phenotype correlation. We also applied PM2 to the c.86-16_87-15 deletion due to its absence or extremely low frequency in control populations, consistent with a rare recessive disorder.

Discussion

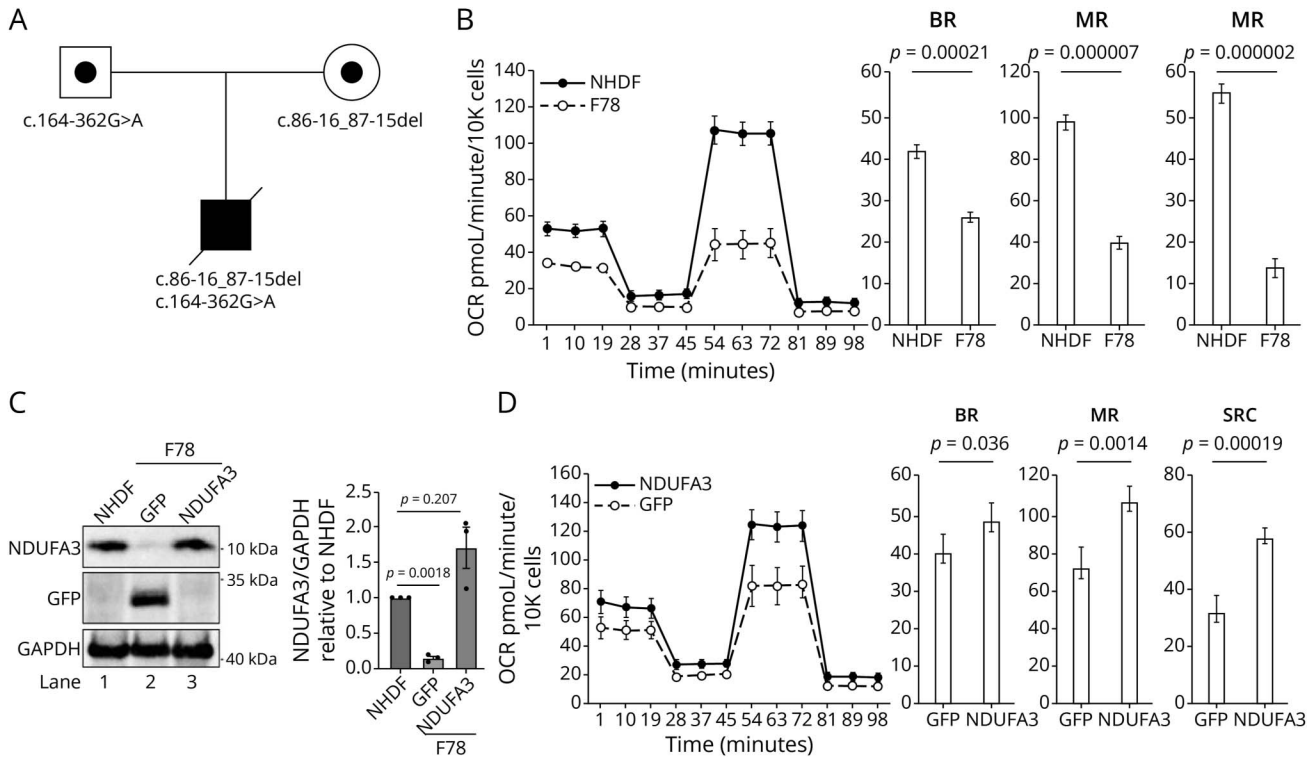
We have shown that pathogenic variant in *NDUFA3* is a cause of Leigh syndrome. The similarity between this case and previously reported cases of Leigh syndrome caused by complex I deficiency was confirmed by MRI findings, which revealed lesions in the brainstem and basal ganglia, prominent cystic degeneration of white matter lesions, and marked cerebral atrophy. The clinical findings in our patient were consistent with typical Leigh syndrome. Compared with the previously reported case by Li et al., our patient showed more severe multiorgan involvement, suggesting that different *NDUFA3* variants can influence phenotype. We identified biallelic intronic variants in *NDUFA3* (c.86–16_86–15del and c.164-362G>A) that lead to significant splicing abnormalities in a patient with Leigh syndrome. These variants align with findings from a study highlighting the importance of atypical splicing variants.[38] The c.86-16_87-15 deletion is located near a splice acceptor site and causes intron retention. This variant falls into the category of "near-splice positions," which can have a significant impact on the splicing mechanism. Such variants are often overlooked in traditional analyses, but systematic approaches have enabled their detection. By contrast, the c.164-362G>A variant is located deep within an intron and triggers exonization. This variant represents an example of an atypical splicing variant, having been identified through a combination of WGS and RNA-seq. This is an important example of how deep intronic variants can contribute to disease. Notably, this variant is associated with Alu elements.[39-41] Alu sequences are primarily found within introns and can potentially influence gene splicing. When Alu elements are inserted into deep introns, they can alter the surrounding splicing signals and create new splice junctions. Such changes in splicing can allow these normally weak junctions to create new exons, resulting in the production of new protein isoforms, modulation of gene regulation, frameshifts, or nonsense-mediated decay. Furthermore, Alu exonization is frequently observed throughout primate evolution, with many human genes expressing Alu exons. The exonization caused by the c.164-362G>A variant can be considered an example of splicing changes driven by Alu elements. Alu elements can act

Figure 3 Splicing Patterns and Alu-Related Exonization in NDUFA3



(A) Diagram of the possible splicing patterns by SR RNA-seq. (B) An exonization in NDUFA3 is related to an inverted AluY element. The variant c.164-362G>A is located on the 3′ side of the AluY element and likely creates a new donor splice site. This aberrant splice site leads to the inclusion of part of the AluY sequence in the mature mRNA, resulting in an abnormal transcript. Alu elements, including AluY, are known to contain cryptic splice sites, and their exonization can disrupt normal gene function, as seen in this case.

Figure 4 Confirmation of Variant Segregation and Functional Rescue of NDUFA3 Deficiency



(A) The segregation of the variants causing each splicing abnormalities were confirmed by Sanger sequencing of the family. (B) The Oxygen consumption rate (OCR) was measured using the Seahorse XFe24 Extracellular Flux Analyzer in normal human dermal fibroblasts (NHDF) and patient-derived fibroblasts (F78). Each line and bar represents the mean of 5 wells from a representative experiment; error bars indicate ± SD. Statistical significance between groups was evaluated using Welch's *t* test. Results were reproducible across 3 independent biological experiments. (C) Expression of *NDUFA3* was examined in normal (NHDF; normal neonatal human dermal fibroblast) and patient fibroblast cells (F78) by SDS-PAGE/Western blotting using indicated antibodies. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a loading control. The accompanying bar graph shows quantification from 3 independent biological replicates. *p* values were calculated using one-way ANOVA. (D) OCR analysis of F78 fibroblasts infected with lentiviruses expressing either NDUFA3 or GFP. Lines represent the mean of 5 wells from a representative experiment; error bars indicate ± SD. Results were reproducible in 3 independent biological experiments. BR = basal respiration; MR = maximal respiration; SRC = spare respiration capacity.

as regulatory factors of splicing, forming new splice sites that impact gene expression and function. The significant decrease in *NDUFA3* protein levels observed in patient-derived fibroblasts suggests that abnormal splicing substantially impairs levels of wild-type mRNA. This kind of variant involving Alu elements can result in the occurrence of specific genetic disorders. Alu element–mediated splicing alterations have been implicated in various genetic diseases.[42,43] These findings indicate that noncoding region variants, particularly those affecting splicing, can be responsible for complex inherited diseases, such as Leigh syndrome. This study highlights the value of WGS and RNA analysis for uncovering atypical splicing variants in undiagnosed rare diseases. The accumulation of such cases will improve our ability to predict variant pathogenicity and guide therapeutic development.

Combining WGS, RNA-seq, and multiomics analysis, which can indicate downregulation of protein levels, is very effective for detecting abnormalities involving deep intronic variants. However, without applying these methods, the final diagnosis of diseases caused by such splicing abnormalities is difficult. Recent studies suggest that cryptic splicing may account for approximately 10% of pathogenic variants in neurodevelopmental disorders. However, it is unclear how much of this involves splicing abnormalities in deep intronic regions. This number is only a prediction at this stage, and specific data on this issue are not yet available. Furthermore, for diseases with autosomal recessive forms of inheritance, whole genome long-read sequencing is very important to confirm the presence of variants in different alleles. However, it is unwise to rely solely on in silico analysis to identify the causative gene in a patient sample. It is essential to conduct validation experiments to confirm protein expression and OCR. We propose that combinations of these steps will provide a useful tool for identifying deep intronic variants.

Author Contributions

K. Nakamura: drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. Y. Kishita: drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. A. Sugiura: drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. K. Ozaki: drafting/

revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. Y. Yatsuka: drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. N. Matsumoto: analysis or interpretation of data. A. Okazaki: drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. H. Prokisch: major role in the acquisition of data. K. Maruyama: major role in the acquisition of data. H. Iwasa: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data. K. Murayama: drafting/revision of the manuscript for content, including medical writing for content; study concept or design. H. Matsumoto: drafting/revision of the manuscript for content, including medical writing for content; study concept or design. A. Ohtake: drafting/revision of the manuscript for content, including medical writing for content; study concept or design. Y. Shiraishi: drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. Y. Okazaki: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data.

Study Funding

This work was supported by JSPS KAKENHI Grant Numbers JP23K07488, JP22K15950, and JP23H00424; MEXT-Supported Program for the Private University Research Branding Project, the Practical Research Project (JP23ek0109625, and JP23ek0109672) for Rare/Intractable Diseases from the Japan Agency for Medical Research and Development, and grant for Moonshot Goal 7,AMED.

Disclosure

The authors report no relevant disclosures. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/NG.

Publication History

Received by *Neurology® Genetics* January 17, 2025. Accepted in final form September 24, 2025. Submitted and externally peer-reviewed. The handling editor was Deputy Editor Antonella Spinazzola, MD.

References

1. Gorman GS, Chinnery PF, DiMauro S, et al. Mitochondrial diseases. *Nat Rev Dis Primers.* 2016;2:16080. doi:10.1038/nrdp.2016.80
2. Tuppen HA, Blakely EL, Turnbull DM, Taylor RW. Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta.* 2010;1797(2):113-128. doi:10.1016/j.bbabio.2009.09.005
3. Forny P, Footitt E, Davison JE, et al. Diagnosing mitochondrial disorders remains challenging in the omics era. *Neurol Genet.* 2021;7(3):e597. doi:10.1212/NXG.0000000000000597
4. Schon KR, Horvath R, Wei W, et al. Use of whole genome sequencing to determine genetic basis of suspected mitochondrial disorders: cohort study. *BMJ.* 2021;375:e066288. doi:10.1136/bmj-2021-066288
5. Macken WL, Falabella M, McKittrick C, et al. Specialist multidisciplinary input maximises rare disease diagnoses from whole genome sequencing. *Nat Commun.* 2022;13(1):6324. doi:10.1038/s41467-022-32908-7
6. Stenton SL, Zou Y, Cheng H, Prokisch H, Fang F. Pediatric Leigh syndrome: neuroimaging features and genetic correlations. *Ann Neurol.* 2021;89(3):629-631. doi:10.1002/ana.25998
7. Schubert Baldo M, Vilarinho L. Molecular basis of Leigh syndrome: a current look. *Orphanet J Rare Dis.* 2020;15(1):31. doi:10.1186/s13023-020-1297-9
8. Ogawa E, Fushimi T, Ogawa-Tominaga M, et al. Mortality of Japanese patients with Leigh syndrome: effects of age at onset and genetic diagnosis. *J Inherit Metab Dis.* 2020;43(4):819-826. doi:10.1002/jimd.12218
9. Dang QCL, Phan DH, Johnson AN, et al. Analysis of human mutations in the supernumerary subunits of complex I. *Life (Basel).* 2020;10(11):296. doi:10.3390/life10110296
10. Barshad G, Zlotnikov-Poznianski N, Gal L, Schuldiner M, Mishmar D. Disease-causing mutations in subunits of OXPHOS complex I affect certain physical interactions. *Sci Rep.* 2019;9(1):9987. doi:10.1038/s41598-019-46446-8
11. Letts JA, Fiedorczuk K, Sazanov LA. The architecture of respiratory supercomplexes. *Nature.* 2016;537(7622):644-648. doi:10.1038/nature19774
12. Stroud DA, Surgenor EE, Formosa LE, et al. Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature.* 2016;538(7623):123-126. doi:10.1038/nature19754
13. Angebault C, Charif M, Guegen N, et al. Mutation in NDUFA13/GRIM19 leads to early onset hypotonia, dyskinesia and sensorial deficiencies, and mitochondrial complex I instability. *Hum Mol Genet.* 2015;24(14):3948-3955. doi:10.1093/hmg/ddv133
14. Fernandez-Moreira D, Ugalde C, Smeets R, et al. X-linked NDUFA1 gene mutations associated with mitochondrial encephalomyopathy. *Ann Neurol.* 2007;61(1):73-83. doi:10.1002/ana.21036
15. Yatsuka Y, Kishita Y, Formosa LE, et al. A homozygous variant in NDUFA8 is associated with developmental delay, microcephaly, and epilepsy due to mitochondrial complex I deficiency. *Clin Genet.* 2020;98(2):155-165. doi:10.1111/cge.13773
16. Li BG, Wu WJ, Wang LH, et al. Identification of a novel pathogenic gene, NDUFA3, in Leigh Syndrome through whole exome sequencing. *Neurogenetics.* 2024;26(1):13. doi:10.1007/s10048-024-00782-8
17. Deininger P. Alu elements: know the SINEs. *Genome Biol.* 2011;12:236. doi:10.1186/gb-2011-12-12-236
18. Funakoshi K, Bagheri M, Zhou M, Suzuki R, Abe H, Akashi H. Highly sensitive and specific Alu-based quantification of human cells among rodent cells. *Sci Rep.* 2017;7(1):13202. doi:10.1038/s41598-017-13402-3
19. Helman G, Compton AG, Hock DH, et al. Multiomic analysis elucidates Complex I deficiency caused by a deep intronic variant in NDUFB10. *Hum Mutat.* 2021;42(1):19-24. doi:10.1002/humu.24135
20. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell.* 2019;176(3):535-548.e24. doi:10.1016/j.cell.2018.12.015
21. Vaz-Drago R, Custodio N, Carmo-Fonseca M. Deep intronic mutations and human disease. *Hum Genet.* 2017;136(9):1093-1111. doi:10.1007/s00439-017-1809-4
22. Scheller IF, Lutz K, Mertes C, Yepez VA, Gagneur J. Improved detection of aberrant splicing with FRASER 2.0 and the intron Jaccard index. *Am J Hum Genet.* 2023;110(12):2056-2067. doi:10.1016/j.ajhg.2023.10.014
23. Mertes C, Scheller IF, Yepez VA, et al. Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun.* 2021;12(1):529. doi:10.1038/s41467-020-20573-7
24. Shiraishi Y, Kataoka K, Chiba K, et al. A comprehensive characterization of cis-acting splicing-associated variants in human cancer. *Genome Res.* 2018;28(8):1111-1125. doi:10.1101/gr.231951.117
25. Bernier FP, Boneh A, Dennett X, Chow CW, Cleary MA, Thorburn DR. Diagnostic criteria for respiratory chain disorders in adults and children. *Neurology.* 2002;59(9):1406-1411. doi:10.1212/01.wnl.0000033795.17156.00
26. Kirby DM, Thorburn DR, Turnbull DM, Taylor RW. Biochemical assays of respiratory chain complex activity. *Methods Cell Biol.* 2007;80:93-119. doi:10.1016/S0091-679X(06)80004-X
27. Kohda M, Tokuzawa Y, Kishita Y, et al. A comprehensive genomic analysis reveals the genetic landscape of mitochondrial respiratory chain complex deficiencies. *PLoS Genet.* 2016;12(1):e1005679. doi:10.1371/journal.pgen.1005679
28. Ogawa E, Shimura M, Fushimi T, et al. Clinical validity of biochemical and molecular analysis in diagnosing Leigh syndrome: a study of 106 Japanese patients. *J Inherit Metab Dis.* 2017;40(5):685-693. doi:10.1007/s10545-017-0042-6
29. Shimura M, Nozawa N, Ogawa-Tominaga M, et al. Effects of 5-aminolevulinic acid and sodium ferrous citrate on fibroblasts from individuals with mitochondrial diseases. *Sci Rep.* 2019;9(1):10549. doi:10.1038/s41598-019-46772-x
30. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5(7):621-628. doi:10.1038/nmeth.1226
31. Brechtmann F, Mertes C, Matuseviciute A, et al. OUTRIDER: a statistical method for detecting aberrantly expressed genes in RNA sequencing data. *Am J Hum Genet.* 2018;103(6):907-917. doi:10.1016/j.ajhg.2018.10.025
32. Kurosawa R, Iida K, Ajiro M, et al. PDIVAS: pathogenicity predictor for deep-intronic variants causing aberrant splicing. *BMC Genomics.* 2023;24(1):601. doi:10.1186/s12864-023-09645-2
33. Omichi N, Kishita Y, Nakama M, et al. Novel ITPA variants identified by whole genome sequencing and RNA sequencing. *J Hum Genet.* 2023;68(9):649-652. doi:10.1038/s10038-023-01156-y
34. Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, Leggett RM. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol.* 2022;23(1):11. doi:10.1186/s13059-021-02582-x

35. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018; 34(18):3094-3100. doi:10.1093/bioinformatics/bty191
36. Shafin K, Pesout T, Chang PC, et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat Methods*. 2021;18(11):1322-1332. doi:10.1038/s41592-021-01299-w
37. Daniel C, Behm M, Ohman M. The role of Alu elements in the cis-regulation of RNA processing. *Cell Mol Life Sci*. 2015;72(21):4063-4076. doi:10.1007/s00018-015-1990-3
38. Blakes AJM, Wai HA, Davies I, et al. A systematic analysis of splicing variants identifies new diagnoses in the 100,000 Genomes Project. *Genome Med*. 2022;14(1):79. doi:10.1186/s13073-022-01087-x
39. He Z, Chen O, Phillips N, Pasquesi GIM, Sabunciyan S, Florea L. Predicting Alu exonization in the human genome with a deep learning model. *bioRxiv*. 2024: 2024.01.03.574099. doi:10.1101/2024.01.03.574099
40. Sen SK, Han K, Wang J, et al. Human genomic deletions mediated by recombination between Alu elements. *Am J Hum Genet*. 2006;79(1):41-53. doi:10.1086/504600
41. Xing J, Zhang Y, Han K, et al. Mobile elements create structural variation: analysis of a complete human genome. *Genome Res*. 2009;19(9):1516-1526. doi:10.1101/gr.091827.109
42. Lev-Maor G, Ram O, Kim E, et al. Intronic Alus influence alternative splicing. *PLoS Genet*. 2008;4(9):e1000204. doi:10.1371/journal.pgen.1000204
43. Gussakovsky D, McKenna SA. Alu RNA and their roles in human disease states. *RNA Biol*. 2021;18(sup2):574-585. doi:10.1080/15476286.2021.1989201