





RESEARCH ARTICLE OPEN ACCESS

# RNA Sequencing Resolves Cryptic Pathogenic Variants in Mitochondrial Disease

Zhimei Liu<sup>1</sup>  | Xin Duan<sup>1</sup> | Fatemeh Peymani<sup>2,3</sup> | Jia Wang<sup>4</sup> | Chengjia Bao<sup>4</sup> | Chaolong Xu<sup>1</sup> | Ying Zou<sup>1</sup> | Zixuan Zhang<sup>1</sup> | Yunxi Zhang<sup>1</sup> | Tongyue Li<sup>1</sup> | Martin Pavlov<sup>2,3</sup> | Junling Wang<sup>5</sup> | Minhan Song<sup>1</sup> | Tianyu Song<sup>1</sup> | Xiaodi Han<sup>1</sup> | Mingxi Sun<sup>1</sup> | Danmin Shen<sup>1</sup> | Ruoyu Duan<sup>1</sup> | Huafang Jiang<sup>6</sup> | Manting Xu<sup>1</sup>  | Holger Prokisch<sup>2,3</sup>  | Fang Fang<sup>1</sup> 

<sup>1</sup>Department of Neurology, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China | <sup>2</sup>Institute of Human Genetics, School of Medicine and Health, Technical University of Munich, Munich, Germany | <sup>3</sup>Institute of Neurogenomics, Computational Health Center, Helmholtz Munich, Neuherberg, Germany | <sup>4</sup>Cipher Gene Ltd, Beijing, China | <sup>5</sup>Department of Pediatrics, Third Affiliated Hospital of Zhengzhou University, Zhengzhou, China | <sup>6</sup>Department of Pediatrics, Weifang Maternal and Child Health Hospital, Weifang, China

**Correspondence:** Holger Prokisch ([holger.prokisch@helmholtz-munich.de](mailto:holger.prokisch@helmholtz-munich.de)) | Fang Fang ([fangfang@bch.com.cn](mailto:fangfang@bch.com.cn))

**Received:** 14 February 2026 | **Accepted:** 4 March 2026

**Keywords:** mitochondrial diseases | pediatric | RNA sequencing | whole-exome sequencing | whole-genome sequencing

## ABSTRACT

**Objective:** Mitochondrial diseases are the most common inherited metabolic disorders, characterized by pronounced clinical and genetic heterogeneity that complicates molecular diagnosis. Although DNA-based sequencing approaches have become standard in genetic testing, up to half of patients remain without a definitive diagnosis. We aimed to perform RNA sequencing (RNA-seq) of patient-derived skin fibroblasts to enhance the molecular diagnostic efficacy of mitochondrial disease in undiagnosed cases in China.

**Methods:** We performed RNA-seq on skin fibroblasts from 140 pediatric patients with suspected mitochondrial disease who remained genetically undiagnosed after whole exome sequencing (WES). Aberrant RNA expression and splicing were identified using the detection of RNA outliers pipeline (DROP). Based on WES findings, patients were stratified into a candidate group ( $n=28$ ), in which RNA-seq evaluated the pathogenicity of WES-identified variants of uncertain significance and an unsolved group ( $n=112$ ), in which RNA-seq was used to pinpoint candidate genes. In six cases where RNA-seq identified the aberrant RNA event but WES did not detect the causative variants, whole genome sequencing (WGS) was performed.

**Results:** Integrative RNA-seq, WES, and WGS analysis resulted in a genetic diagnosis in 25% of patients overall (20/28 [71%] in the candidate group; 15/112 [13%] in the unsolved group). Aberrant splicing explained most candidate-group diagnoses, including variants misclassified by in silico predictors such as SpliceAI. 14% of protein-truncating variants predicted to undergo nonsense-mediated decay (NMD) escaped degradation, highlighting the functional limits of current predictions. The variants identified in the unsolved cohort included synonymous, missense, deep intronic, near-splice-site variants, and large deletions. The most frequent among them was a recurrent synonymous East Asian founder mutation in *ECHS1*, accounting for seven cases. Interestingly, across 233 pathogenic variants associated with aberrant RNA phenotypes compiled from this study and prior reports, half were noncoding and half were coding variants.

**Conclusion:** RNA-seq substantially enhances molecular diagnosis in mitochondrial disease by exposing cryptic splicing, regulatory, and NMD-escape events invisible to DNA sequencing alone. These data advocate transcriptome analysis as an essential component of comprehensive genomic diagnostics in neurometabolic disease.

Zhimei Liu, Xin Duan, and Fatemeh Peymani contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2026 The Author(s). *Annals of Clinical and Translational Neurology* published by Wiley Periodicals LLC on behalf of American Neurological Association.

## 1 | Introduction

Mitochondrial dysfunction, predominantly characterized by defects in oxidative phosphorylation and impaired energy metabolism, gives rise to a spectrum of clinically diverse disorders collectively known as mitochondrial diseases. As a group, mitochondrial diseases constitute the most common class of inherited metabolic disorders, with an estimated incidence of approximately 1 in 5000 individuals. Pathogenic variants in more than 400 genes in both nuclear and mitochondrial DNA have been implicated, giving rise to marked clinical and genetic heterogeneity. This variability, combined with extensive phenotypic overlap with other Mendelian disorders, renders molecular diagnosis particularly challenging.

The routine clinical implementation of whole exome sequencing (WES) has revolutionized the genetic diagnosis of mitochondrial diseases and facilitated the discovery of more than 200 novel disease genes [1]. Nevertheless, despite comprehensive genomic analyses, the diagnostic yield in suspected cases rarely exceeds 50%, leaving a substantial proportion of patients without a molecular diagnosis [2]. WES, by design, captures only exonic regions and therefore misses noncoding variants, structural rearrangements, and repeat expansions. While whole-genome sequencing (WGS), especially in its long-read form, can detect these variant types, its incremental diagnostic yield remains modest [3]. The discovery of millions of variants in the human genome resulted in an explosion of *in silico* prediction tools interpreting protein-coding variants [4] and splice site variants [5] with increasing accuracy; however, already the prediction of functional effects of near splice site variants becomes challenging, and the accuracy in predicting the consequences of deep intronic variants is still very low.

Current research indicates that approximately 30% of pathogenic variants are located in noncoding regions of the genome [6, 7]. Variants in the coding region may also affect gene expression or splicing. Multiplex splicing assays have revealed that around 10% of pathogenic exonic variants disrupt splicing [8, 9], another category of alterations that remains challenging to predict accurately. Moreover, most predicted protein-truncating variants (PTVs) induce nonsense-mediated mRNA decay (NMD). A large number of truncated proteins would remain functional, but NMD causes a loss-of-function. Roughly 30% of PTVs escape predicted NMD with a fundamental consequence on the clinical interpretation [10–13]. Consequently, without experimental evidence from patient biopsy material, many rare variants remain variants of uncertain significance (VUS) [14].

RNA sequencing (RNA-seq) enables a comprehensive analysis of transcript abundance, facilitating the systematic detection of RNA phenotypes and providing allelic information across the entire transcriptome. These phenotypes include aberrant gene expression, abnormal splicing patterns, and monoallelic expression of rare variants. Identifying such anomalies can provide functional evidence for VUS in both coding and noncoding regions, facilitating their correct classification. Moreover, the analysis of RNA-seq data may also discover aberrant RNA phenotypes in genes not prioritized in the genomic evaluation. To date, RNA-seq has been applied to

various patient cohorts and improved the diagnostic yield on average by 15% [14].

Here, we report the implementation of RNA-seq within a national diagnostic program for mitochondrial disease at Beijing Children's Hospital. We performed RNA-seq on fibroblasts from 140 pediatric patients who remained undiagnosed after WES and achieved a molecular diagnosis in 25% of cases, demonstrating that functional transcriptomics is a powerful addition to the diagnostic workflow. Our findings further highlight the limitations of current prediction tools for NMD and splicing and emphasize the necessity of empirical transcriptomic evaluation in the clinical interpretation of rare variants.

## 2 | Materials and Methods

### 2.1 | Study Cohort

Patients were recruited between November 2017 and August 2023 from the Department of Neurology, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health (China) through the Chinese Mitochondrial Disease Network (mitoC-Network; ChiCTR1900028101). Inclusion required a Mitochondrial Disease Criteria (MDC) score  $\geq 2$ , or MDC  $< 2$  with a candidate mitochondrial gene [15, 16], and no molecular diagnosis after whole-exome sequencing (WES). Skin biopsies were obtained from 140 pediatric patients (median age at biopsy 3.4 years; IQR 1.7–7.1; male: female  $\approx 1.5:1$ ). Written informed consent for genetic testing and research was obtained from legal guardians in accordance with the Declaration of Helsinki. The study was approved by the institutional ethics committee (2017-k-60).

Based on WES findings, patients were assigned to two groups: A candidate group ( $n = 28$ ), in which RNA-seq evaluated the pathogenicity of WES-identified variants of uncertain significance, and an unsolved group ( $n = 112$ ), in which RNA-seq was used to pinpoint candidate genes.

### 2.2 | WGS and WES Reanalysis

Raw WES and WES/WGS data of each participant were reanalyzed. Variants were classified according to the guidelines of the American College of Medical Genetics and Genomics and the Association of Molecular Pathology (ACMG/AMP) [17]. For the SpliceAI prediction, we used the conservative thresholds to ensure the specificity. A  $\Delta$  score  $\leq 0.01$  was considered benign, and a  $\Delta$  score  $\geq 0.8$  was considered pathogenic effects on splicing [18, 19].

### 2.3 | Cell Culture

Primary fibroblast cell lines obtained from patient skin biopsies (diameter 2–3 mm) were cultured in 1 mL culture medium with high-glucose DMEM (Life Technologies) supplemented with 20% FBS and 1% penicillin/streptomycin at 37°C and 5% CO<sub>2</sub>. Culture media were changed every 2 days, and the skin flap was

passaged approximately 10 days after implantation once confluence was reached.

## 2.4 | RNA Isolation, Library Preparation, and Sequencing

Fibroblast cells ( $> 1 \times 10^6$ ) were harvested in 1 mL TRIzol reagent (Thermo Fisher Scientific). Total RNA was extracted using an AllPrep RNA Kit (Qiagen). RNA quality and integrity were assessed using an Agilent 2100 BioAnalyzer. PolyA-tailed mRNA was enriched with oligo (dT) magnetic beads, fragmented, and reverse-transcribed. Purified cDNA was subsequently subjected to end repair, A-tailing, and adapter ligation. 350–450-bp cDNA libraries were selected, PCR-amplified, and purified. High-quality libraries were sequenced as 150-bp paired-end runs on an Illumina NovaSeq 6000 platform (Illumina) to generate more than 40 million reads (12 GB raw data).

## 2.5 | RNA-Seq Alignment and Variant Calling

Low-quality bases and sequencing adapters were removed, and reads were aligned to the human GRCh38/hg38 genome assembly using STAR 2.7.6a [20] with the twopassMode set to “Basic” to enable sensitive junction discovery. Following the removal of duplicate reads, base quality score recalibration, and read splitting at the junction, variants from RNA-seq data were called using the Sentieon software suite (202010.02) Haplotyper algorithm with default parameters.

## 2.6 | Detection of Aberrant RNA Phenotypes

STAR-aligned BAM files were used to calculate gene-level counts based on NCBI Refseq (hg38.refseq.gff: GRCh38\_20200919), which consisted of 39,225 genes by HTseq-count [21, 22]. Genes with 95th percentile fragments per kilobase of transcript per million mapped reads (FPKM) less than 1 were considered insufficiently expressed and were excluded from further analysis.

Gene count matrices from two batches (88 samples and 56 samples) of cultured skin fibroblast cells and related data from Genotype-Tissue Expression (GTEx) (cultured fibroblasts, skin not sun-exposed (suprapubic), skin sun-exposed (lower leg), and whole blood) were normalized and transformed for variance stabilization.

Aberrant expression (AE), aberrant splicing (AS), and monoallelic expression (MAE) were analyzed using the core module of DROP [21]. OUTRIDER (1.12.0) [23] was used to identify aberrant candidate gene expression events using  $p$  values ( $< 0.005$ ) and  $|Z\text{-scores}| (> 3)$ . All disease-related candidate events were subsequently validated using Integrative Genome Viewer [24].

FRASER (1.7.0) was used to detect splicing outliers, which is based on count distribution and multiple testing corrections to reduce the number of calls by two orders of magnitude over the commonly applied Z-score cutoffs [21, 25]. Exon–exon and

exon–intron junctions with more than 10 reads in at least one sample were analyzed. Outlier junctions were defined as those in splicing outlier genes, with  $p$  values  $< 0.05$  and  $|Z\text{ scores}| > 2$  and count  $\geq 5$ .

For the detection of monoallelic expression, allelic RNA counts were extracted from the gVCF file based on the corresponding genomic variants identified from the WES data. Monoallelic expression was determined using a negative binomial test applied to read counts using the tMAE package with the DESeq4MAE function [21, 26, 27]. Benjamini-Hochberg adjusted  $p$  values  $< 0.05$  indicated significant MAE.

## 2.7 | Systematic Literature Review

We searched PubMed, ISI Web of Science, and Scopus (January 2017–September 2024) using: (“Rare disease” OR “Mendelian disease” OR “Mendelian disorders”) AND (“transcriptome” OR “transcriptome sequencing” OR “RNA-seq” OR “RNA sequencing”). Inclusion criteria were original cohort studies applying RNA-seq for Mendelian disease diagnostics. Reviews, meta-analyses, case reports, and unrelated articles were excluded. Data were extracted on cohort size, tissue, analytical pipeline, and diagnostic yield. The analysis was patient-based.

## 3 | Results

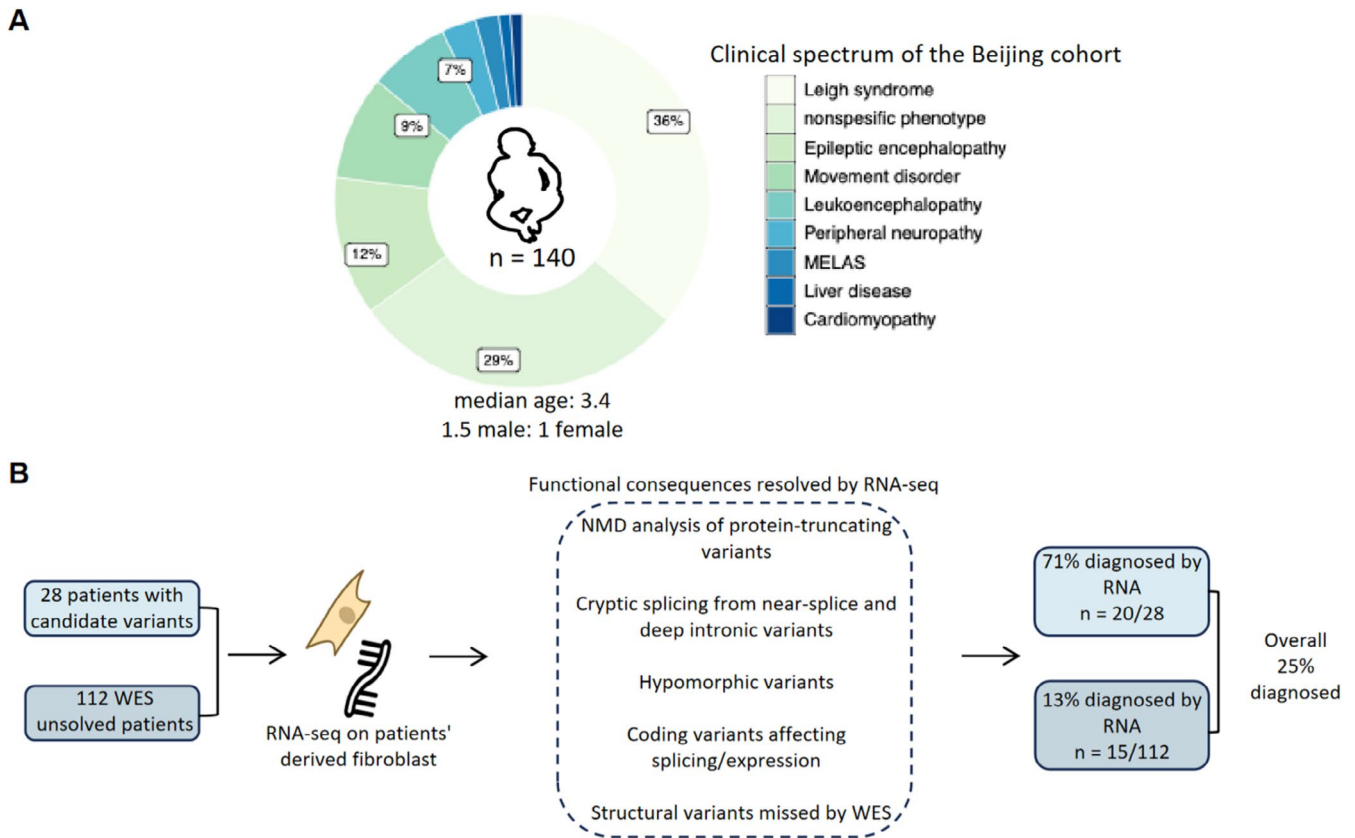
### 3.1 | Study Design and Patient Characteristics

Within the last 5 years, the mitoC-Network recruited 1405 patients suspected of having mitochondrial disease and performed genome-wide molecular screening to identify disease-causing variants. This was achieved in approximately 45% of cases, and hence no genetic diagnosis could be made for the majority of patients. To support routine diagnostic testing, we established fibroblast cell lines from 140 genetically unsolved patients for functional studies. The most frequent clinical presentation in this cohort of patients was Leigh syndrome (36%), followed by nonspecific phenotypes (29%) and movement disorders (9%) (Figure 1A).

RNA-seq was conducted on a per-individual basis. Each sample was subjected to a single assay at a median sequencing depth of 94.5 million reads, ranging from 74.5 to 133.4 million reads, using a nonstrand-specific protocol that utilized automated protocols to minimize sample handling and ensure high reproducibility of results (see Methods). After alignment, RNA-seq data were analyzed using the computational workflow DROP to search for aberrant expression, aberrant splicing, or monoallelic expression of a rare variant. RNA-seq results were then integrated with genomic and clinical data to guide variant interpretation (Figure 1B).

### 3.2 | Evaluation of Variants of Uncertain Significance

Among the 140 patients, 28 harbored candidate variants of uncertain significance in genes consistent with their clinical



**FIGURE 1** | RNA-based diagnostic workflow for WES inconclusive pediatric patients with suspected mitochondrial disorders. (A) Clinical characteristics of 140 pediatric patients in the cohort. (B) Study design and outcome. Patients were categorized into two groups: Candidate ( $n = 28$ ), in whom WES had prioritized variants and RNA-seq was used to assess the functional effect (molecular diagnosis in 20/28, 71.4%); and unsolved ( $n = 112$ ), with no prioritized variants, in whom RNA-seq was used to identify the candidate genes with aberrant RNA phenotypes followed by WES reanalysis or WGS (diagnosis in 15/112, 13.4%). The overall diagnostic yield of 25% (35/140) was achieved. WES, whole-exome sequencing; NMD, nonsense-mediated mRNA decay.

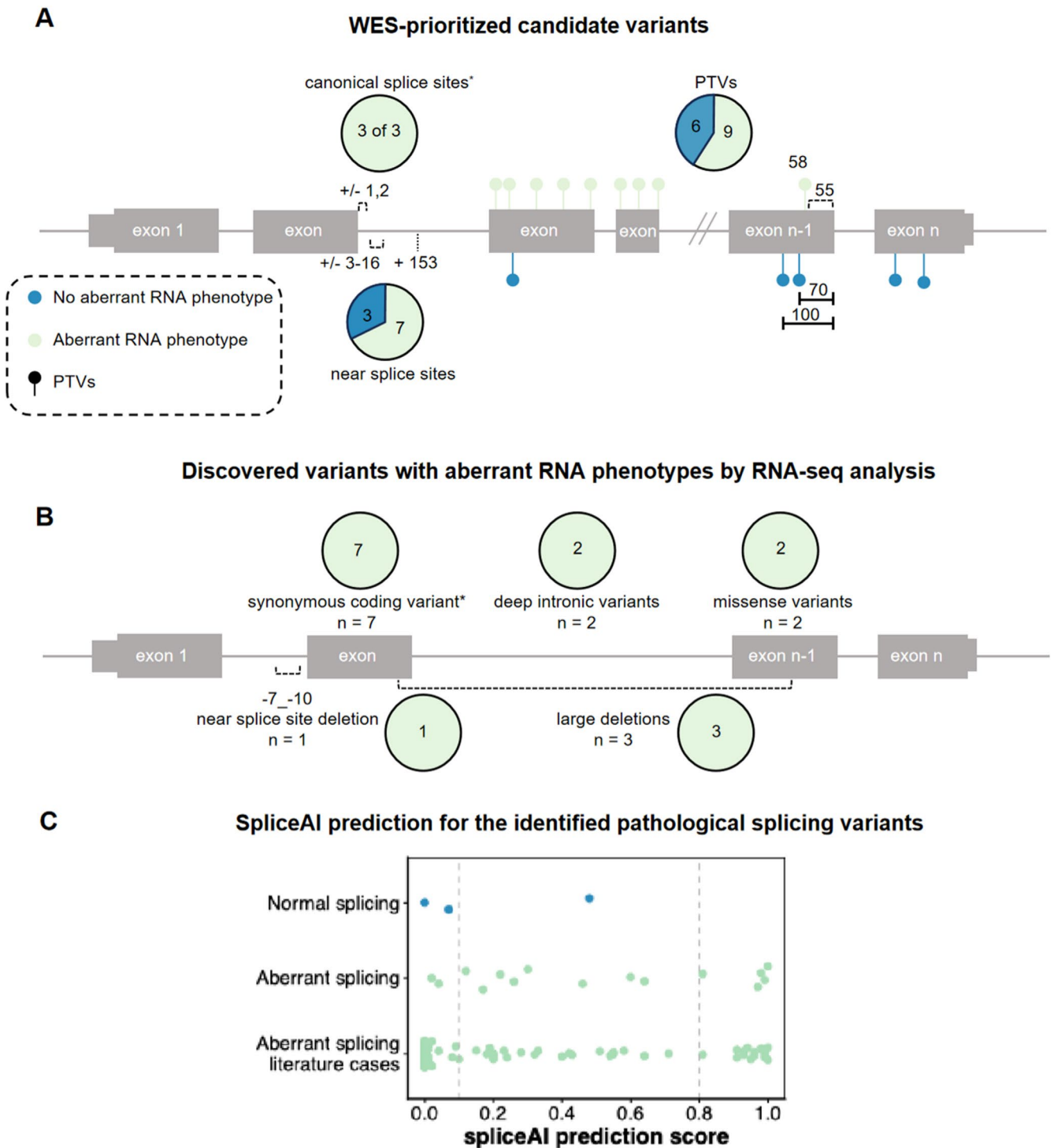
presentation. These included near splice-site ( $n = 10$ ,  $\pm 3$  to 153) and protein-truncating variants (PTVs;  $n = 15$ ) (Figure 2A; Table S1).

RNA-seq provided functional validation for 20 of these 28 cases. Among the 15 predicted PTVs, NMD was confirmed in nine, while six showed no evidence of NMD. In one case, the affected gene (*PIGM*) contained only one exon. One affected gene (*SCN1A*) was not expressed in fibroblasts. In another case, premature termination codons (PTCs) occurred within the final 55 nucleotides of the penultimate exon of the gene (*COA6*); it was not predicted to cause NMD. However, in two cases with variants located 70 and 100 nucleotides upstream of the last splice junction in the penultimate exon of the genes (*IBA57*, *GLRX5*), NMD was predicted but not observed. In other cases, involving variants in the same region, we observed NMD. This highlights that predicting NMD in this region is inaccurate and that RNA-seq is needed to determine the functional consequences of such variants. In one patient, a frameshift variant occurs in the second of nine exons in the main *NDUFAF6* transcript, which should result in NMD. However, RNA-seq revealed that *NDUFAF6* expression levels were within the normal physiological range. Manual inspection revealed a high number of well-expressed *NDUFAF6* transcript isoforms and showed that the variant is outside the

coding region in the 5' UTR in 90% of them. It is not possible to conclude from the short-read data whether there is NMD of the isoform in which the variant is located within the coding region. Despite this, by leveraging established fibroblasts, studies of respiratory chain activities provided functional evidence for the pathogenicity of compound heterozygous PTVs in three patients (Tables S1 and S2). In conclusion, RNA-seq provided a loss-of-function characterization for the nine PTVs, thereby confirming their pathogenicity and aiding in diagnosis.

In 13 cases, the genome analysis identified variants that may affect splicing: three canonical splice site variants, and nine near splice site variants in positions  $\pm 3$ –16, plus one variant at position +153. RNA-seq analysis confirmed a clear splice defect in 10 of 13 cases. Importantly, five of seven near-splice-site variants were misclassified by SpliceAI, demonstrating the diagnostic value of empirical RNA-seq validation (Table S1).

In contrast, an in-frame deletion in *HSD17B4* and a missense variant in *POLG* showed no RNA abnormalities, while a large deletion in another gene was detected through reduced expression. In summary, RNA-seq analysis helped to establish a definitive molecular diagnosis in 20 out of 28 candidate cases, corresponding to a diagnostic rate of 71%.



**FIGURE 2** | Variant types causing aberrant RNA phenotypes and the added value of RNA-seq over prediction tools. (A) The RNA phenotypes in association with the WES-prioritized candidate variants are shown. The dataset comprises 30 variants across 28 patients. The number in the pie chart shows the number of cases. With the exception of the canonical splice site variants (2 variants observed in 3 cases), the count of the cases equals the count of the variants. The numbers on the gene structure represent the nucleotide distance of each variant to the nearest exon junction. One missense variant (no RNA phenotype), one in-frame deletion (no RNA phenotype), and one large deletion (aberrant expression) were not shown in the graph, given the presence of only one variant per category. (B) Shown are the variant types identified in 15 cases that were genetically solved after RNA-seq. The dataset includes 9 variants across 15 patients. Numbers indicate cases. For all classes except the synonymous variant (1 variant observed in 7 cases), the number of variants equals the number of cases. (C) SpliceAI prediction for variants with and without aberrant RNA splicing (based on RNA-seq). The dataset includes 18 variants from panels A and B: 11 near splice site variants (10 from panel A and 1 from panel B), 2 canonical splice, 1 synonymous variant, 2 missense variants, and 2 deep intronic variants, and 64 variants causing aberrant splicing reported in the literature. PTVs, protein-truncating variants; WES, whole-exome sequencing.

### 3.3 | Discovery of Disease-Causing Variants by RNA-Seq

In the second group of 112 unsolved cases for which no candidate genes could be identified, integrating genomic data with aberrant RNA phenotypes and their clinical interpretation resulted in a molecular diagnosis for 15 patients, which we will discuss in detail below. The variants causing aberrant RNA phenotypes included: a common synonymous variant causing a splice defect (in seven cases); three intragenic deletions that were either not covered or called by WES; two predicted missense variants causing aberrant splicing or expression; two deep intronic variants and a deletion of -7 to -10 near splice sites that were missed by WES (Table S1, Figure 2B). SpliceAI failed to predict most of the splice defects, underscoring the empirical power of RNA-seq in variant interpretation (Figure 2C).

#### 3.3.1 | A Synonymous *ECHS1* Variant Causes Aberrant Splicing in Seven Patients With Leigh Syndrome

RNA-seq detected underexpression of *ECHS1* in seven unrelated patients with Leigh syndrome (Figure 3), all heterozygous for the synonymous variant c.489G>A (p.Pro163=) in trans with known pathogenic variants (p.Met103Thr, p.Thr266Ile, p.Thr266Ala, p.Glu274Lys, p.Ala278Thr, and p.Phe279Leu). The variant, located 25bp from the donor and 74bp from the acceptor splice site, was not predicted to affect splicing and expression. However, we detected partial skipping of the exon containing the variant (supported by 4% of reads), leading to a frameshift (p.Phe139ValfsTer65). This is predicted to cause NMD, which explains the underexpression of *ECHS1*. Despite NMD, Exon 4 skipping was supported by 70 reads across all *ECHS1* patients. Overall, *ECHS1* expression was reduced by approximately 20% across patients, indicating that ~40% of transcripts derived from the allele carrying the synonymous variant underwent NMD. This, together with the splice defect, leaves only half of the transcribed synonymous allele functional. This observation is further supported by monoallelic expression of the missense variants in trans in all seven cases. The observed ratio of alleles with 1/3 of the synonymous variant and 2/3 of the missense variants perfectly fits with the observed NMD quantification. Altogether, these data show that 25% of intact *ECHS1* is not enough for preventing the manifestation of Leigh syndrome, while 50% of *ECHS1* expression level is tolerated as observed in homozygous carriers of the synonymous variant. The c.489G>A variant has a high allele frequency of 1.05% in the East Asian population (Figure 3C), suggesting a regional founder effect.

*ECHS1* encodes short-chain enoyl-CoA hydratase 1, which is involved in mitochondrial valine metabolism and short-chain fatty acid beta-oxidation. The clinical presentation of *ECHS1* deficiency (OMIM #616277) is heterogeneous [28, 29], but often includes severely delayed psychomotor development, neurodegeneration, increased lactic acid levels, and brain lesions in the basal ganglia consistent with Leigh syndrome. The clinical findings in these seven cases were consistent with a diagnosis of *ECHS1* deficiency. All individuals presented with similar clinical manifestations and neuroimaging features characteristic of Leigh syndrome (Figure 3A), alongside relatively stable conditions. None had echocardiographic signs of cardiomyopathy.

Metabolic testing performed on four patients revealed elevated urine concentrations of 2,3-dihydroxy-2-methylbutyric acid in all of them and an elevated concentration of S-(2-carboxypropyl) cysteamine in three (Table S2), and western blotting performed for three patients showed a reduction in the expression of *ECHS1* in the fibroblasts [30], further supporting the diagnosis. Treatment was modified accordingly with a valine-restricted diet and acetylcysteine supplementation, which were initiated after diagnosis [30, 31].

#### 3.3.2 | RNA-Seq-Guided Detection of Cryptic Deletions in *PANK2* and *SERAC1*

A 6-year-old boy exhibiting neurodevelopmental delay and regression, as well as dystonia, underwent a brain magnetic resonance imaging (MRI) scan, which revealed bilateral abnormalities in the globus pallidus, indicating the “eye of the tiger” sign (Figure 4B). This finding supported the diagnosis of neurodegeneration with brain iron accumulation (NBIA). Trio-WES identified a rare maternal heterozygous missense variant (p.Arg440Pro) in *PANK2* (Figure 4A), which has been previously reported in association with NBIA [32, 33]. RNA-seq revealed significantly reduced *PANK2* expression levels (Figure 4C). The maternal variant was monoallelically expressed, indicating the loss of expression of the paternal allele (Figure 4D). Trio-WGS identified a paternal intragenic deletion involving the terminal exon, consistent with compound heterozygosity and confirming *PANK2*-related NBIA (OMIM #234200, Figure 4A–D).

A 3-year-and-8-month-old boy had presented with liver dysfunction, failure to thrive, neurodevelopmental delay, and hearing impairment in infancy. At 2 years of age, epilepsy developed with a first status epilepticus, and since then, developmental regression has been observed. Metabolic evaluation revealed lactic acidosis in the blood and elevated urinary 3-methylglutaconic acid excretion. On brain MRI, abnormalities in the basal ganglia and thalamus were detected, which were interpreted as Leigh syndrome, together with generalized cerebral and cerebellar atrophy (Figure 4F). His younger sister had died from lactic acidosis 5 days after birth. Initial WES identified heterozygosity for a predicted pathogenic PTV (p.Arg306\*) in *SERAC1* (Figure 4E). RNA-seq analysis revealed strongly reduced *SERAC1* expression levels (Figure 4G), less than expected from monoallelic NMD. Consequently, trio-WGS revealed a large deletion–insertion structural variant spanning from Intron 6 to Intron 8 in the paternal genome (Figure 4H). The *SERAC1* protein is involved in the remodeling of phosphatidylglycerol 36:1 with the effect of abnormal cardiolipin species in mitochondrial membranes [34]. Pathogenic variants within *SERAC1* are associated with 3-methylglutaconic aciduria, hepatopathy, deafness, encephalopathy, and Leigh-like syndrome (MEGHDEL, OMIM #614739), which is consistent with the patient’s clinical presentation and in line with the genetic diagnosis.

#### 3.3.3 | Pseudoexon Creation by Deep Intronic Variants in *WARS2* and *SUCLG1*

A boy of 4 years and 8 months presented with nystagmus, ataxia, seizures, developmental regression, and elevated

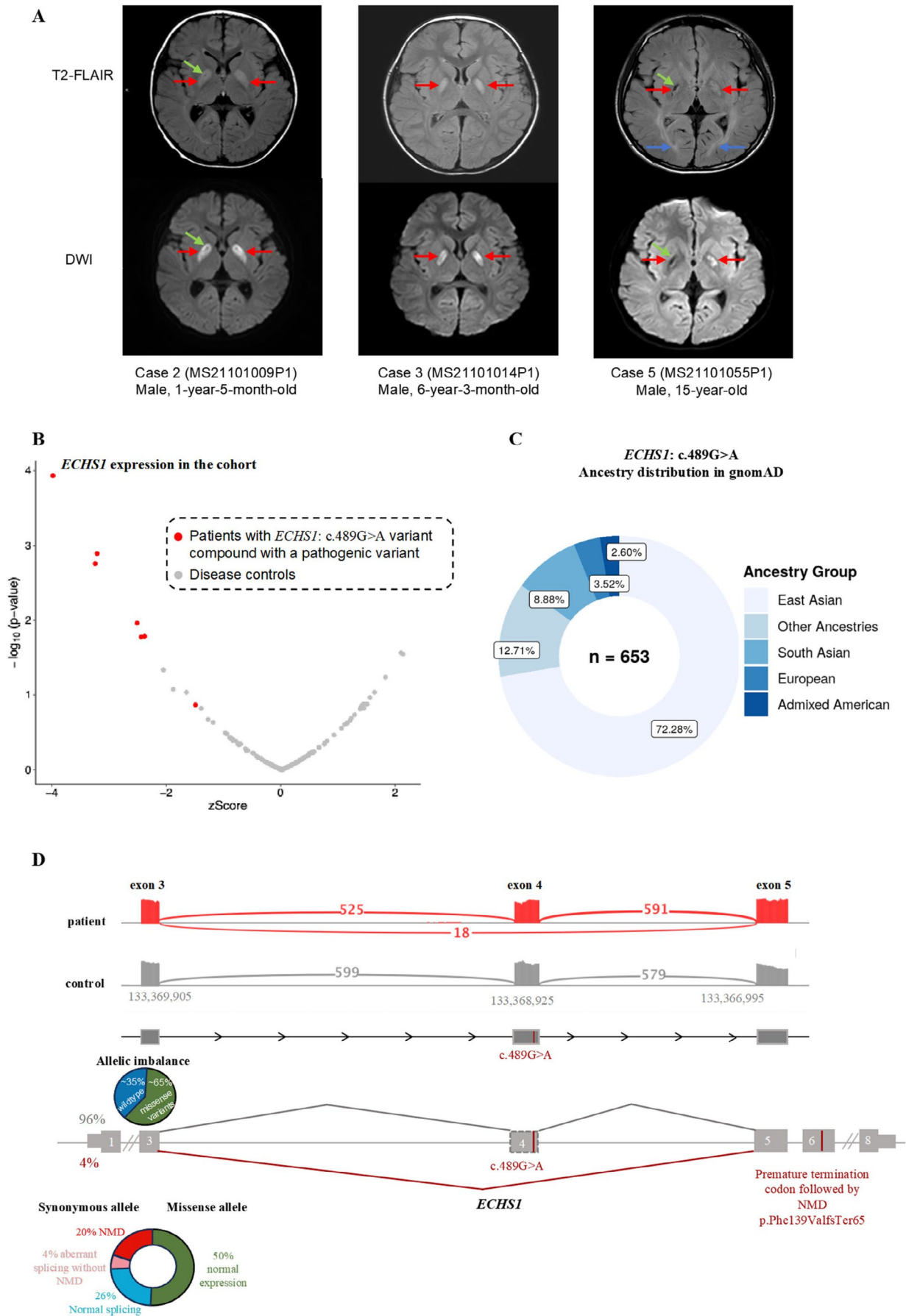


FIGURE 3 | Legend on next page.

**FIGURE 3** | Aberrant expression of *ECHS1* in seven unsolved Leigh syndrome cases harboring the common synonymous variant c.489G>A (p.Pro163=). MRI images of three of the seven patients. Arrows indicate the positions of abnormal signals. Bilateral globus pallidus (red arrows) involvement was a common feature; Case 2 showed punctate necrosis (green arrows), Case 5 showed patchy necrosis (green arrows), and necrotic lesions were not accompanied by diffusion restriction (DWI). In addition to globus pallidus involvement, the bilateral optic radiation was also affected in Case 5 (blue arrows). (B) Volcano plot of *ECHS1* gene expression across all the cases. Seven *ECHS1* diagnosed cases carrying the c.489G>A variant in trans with other pathogenic variants are labeled in red. (C) Ancestry distribution of *ECHS1*: C.489G>A variant based on the gnomAD database shows the high frequency of the variant in the East Asian population. (D) *ECHS1* sashimi plots of fibroblast-derived RNA-seq data from an *ECHS1* patient, and one control, in red and gray, respectively. The numbers on the exon-connecting lines show the number of unnormalized reads spanning the exons. Partial skipping of Exon 4 (in 4% of the reads) is observed in the patient. The position of the c.489G>A variant in Exon 4 is indicated. The schematic presentation of aberrant (red) and normal (gray) splicing is depicted at the bottom. The skipping of Exon 4 due to the c.489G>A variant results in a frameshift and premature termination codon (p.Phe139ValfsTer65) in Exon 6. The average allelic imbalance across all seven missense variants present in trans with the synonymous variant is shown in a pie chart. The relative allelic expression of missense and synonymous variants in *ECHS1* transcripts is depicted in a donut chart based on the mean value across all patients. gnomAD, genome aggregation database.

blood lactate. A brain MRI revealed mild cerebral atrophy (Figure 5B). Trio-WES did not identify any candidate gene. RNA-seq revealed aberrant splicing that created a pseudoexon in intron 1 of *WARS2* by activating a cryptic splice site (Figure 5C). RNA-seq variant calling detected a homozygous deep intronic variant (c.90+268C>T) located within the pseudoexon. Trio-WGS confirmed that this was the only rare variant in Intron 1 and was inherited from both parents (Figure 5A). It is predicted that the creation of the pseudoexon will lead to a stop-gain codon after eight amino acids and NMD. However, no NMD was detected. *WARS2* encodes mitochondrial tryptophanyl-tRNA synthase, which catalyzes the aminoacylation of mitochondrial tRNA with tryptophan. Biallelic mutations in *WARS2* are associated with neurodevelopmental disorders, abnormal movements, and lactic acidosis with or without seizures (NEMMLAS, OMIM #617710), symptoms in line with the patient's presentation.

A 32-month-old girl presented with hypotonia, developmental delay, and regression. She also had an elevated blood lactate concentration and showed mild but unequivocally increased urinary excretion of methylmalonic acid (MMA). Brain MRI showed symmetrical basal ganglia lesions and bilateral lateral ventricle broadening (Figure 5E). The patient was clinically classified to have Leigh syndrome. RNA-seq analysis revealed a significant decrease in *SUCLG1* expression (Figure 5F). A rare heterozygous predicted pathogenic missense variant (c.542G>A, p.Cys181Tyr) showed monoallelic expression (Figure 5G). Furthermore, aberrant splicing of a pseudoexon in Intron 5 was detected with very few reads, indicating NMD of this splice isoform. RNA-seq variant calling revealed a variant in the pseudoexon that likely activated the cryptic splice donor site (Figure 5G). Trio-WGS confirmed that the deep intronic variant c.590-545T>G is *in trans* with the missense variant (Figure 5D). Biallelic mutations in *SUCLG1*, encoding the alpha subunit of succinate-coenzyme A ligase, have been associated with Leigh syndrome and methylmalonic aciduria (OMIM #245400), a clinical presentation that is fully consistent with the symptoms of the patient.

### 3.3.4 | A Homozygous Missense Variant in *CARS2* Causes Aberrant Splicing

A 2-year-old boy presented with severe, progressive drug-therapy refractory epileptic seizures. He showed psychomotor

developmental regression since 7 months of age and was clinically diagnosed with developmental and epileptic encephalopathy. His older brother and sister had exhibited similar symptoms, with onset at 7 and 6 months and had died at 1.5 and 2 years, respectively (Figure 6A). Brain MRI of the index case revealed massive cerebral and cerebellar atrophy (Figure 6B), and electroencephalography (EEG) showed significant epileptic activity. Trio-WES identified a homozygous missense VUS in *CARS2*. RNA-seq analysis revealed aberrant splicing of *CARS2*, resulting in less than 5% of the primary transcript being retained (Figure 6C). The coding VUS located 10 base pairs downstream of the canonical Exon 14 acceptor splice site significantly reduced its recognition, resulting in alternative usage of a weak acceptor in Intron 13. This resulted in a final exon without further splicing, thereby omitting Exons 14 and 15. This alternative exon usage is predicted to create a termination codon after the insertion of 25 amino acids (p.Tyr473GlnfsTer26). Despite the splice defect, overall gene expression was unaffected. Biallelic mutations in *CARS2*, which encodes the mitochondrial cysteinyl-tRNA synthetase, cause combined oxidative phosphorylation deficiency type 27 (OMIM #616672), characterized by developmental regression and epileptic encephalopathy, thus an excellent match to the symptoms of the patient.

In summary, the integration of RNA-seq data from fibroblasts with the genomic and clinical information of 140 patients suspected of having a mitochondrial disorder resulted in a molecular diagnosis in 35 cases, yielding an overall diagnostic rate of 25% (Figure 1, Table S1). With the exception of three cases, all of the identified causal genes were already associated with mitochondrial diseases (Table S1).

## 3.4 | Comparison With Published RNA-Seq Studies

Including this cohort, 16 studies have reported RNA-seq implementation for Mendelian disorders (Table S3), encompassing 1049 patients and 227 molecular diagnoses, for an average diagnostic yield of 22% [14, 35–49]. Diagnostic success varied by tissue type and disorder: highest in muscle (34%–38%) in 138 patients with neuromuscular disorders [36, 37, 46], intermediate in fibroblasts (10%–33%) in 501 patients across six cohorts with diverse phenotypes, and lowest in blood (5%–8%) in 178 patients with suspected rare diseases [40, 49]. One study performed RNA-seq analysis of induced neurons derived from skin

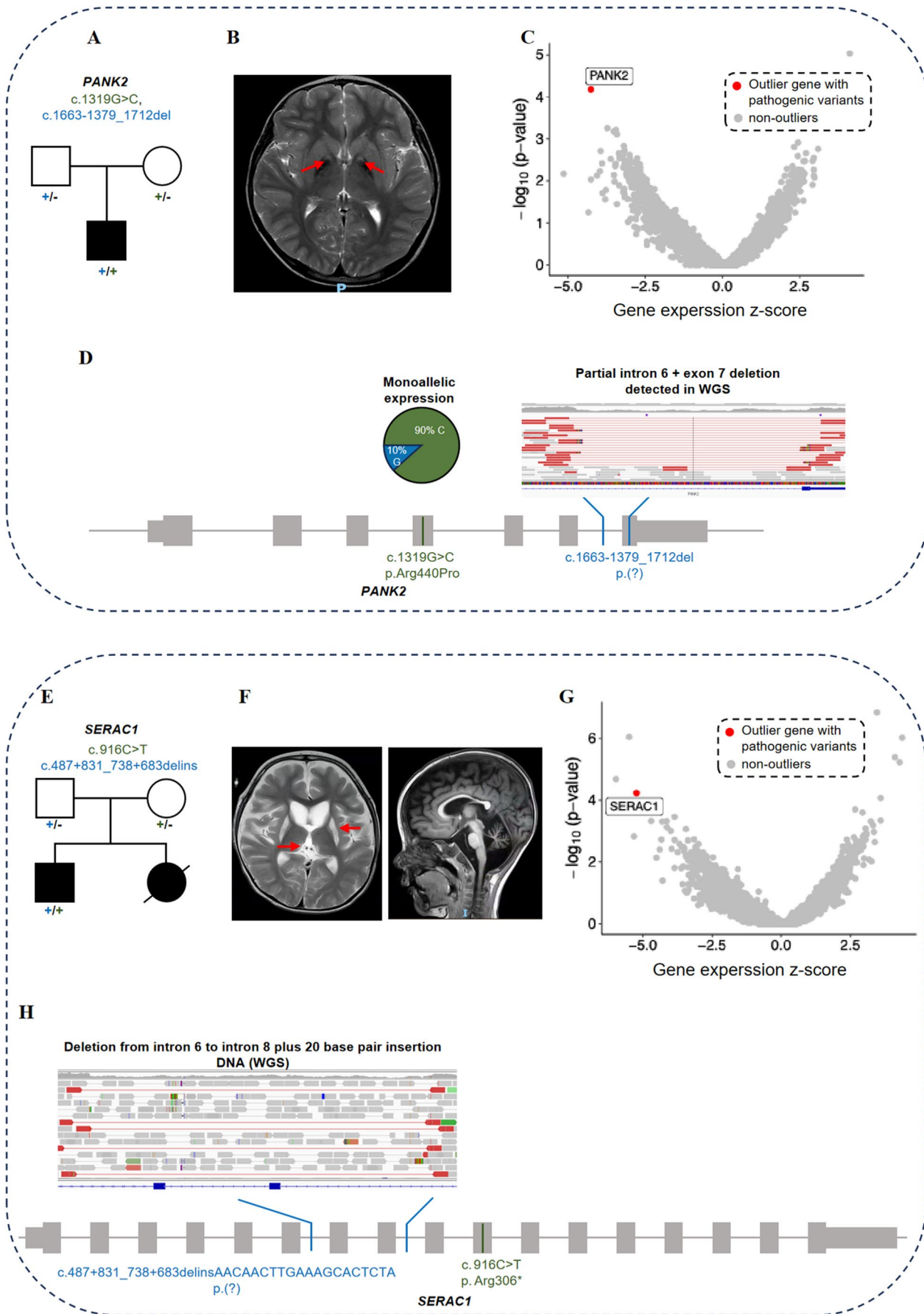


FIGURE 4 | Legend on next page.

**FIGURE 4** | Aberrant expression caused by large intragenic deletions, undetected by WES. (A–D) *PANK2* patient; (E–H) *SERAC1* patient. (A) Family pedigree: *PANK2* compound heterozygous pathogenic variants identified after RNA-seq and trio-WGS are indicated. (B) Brain MRI in the T2-weighted image. The red arrow indicates bilateral abnormalities in globus pallidus, also known as “eye of the tiger” sign. (C) Gene-level volcano plot, showing the significance versus z-score of the gene expression in the patient, with *PANK2* highlighted as the most significant underexpression outlier. (D) Schematic representation of the *PANK2* gene structure, showing the position of the missense variant (chr20:3912541G>C, c.1319G>C) and large deletion (chr20:3921864–3923293del) identified in the patient. The monoallelic expression of the missense variant (c.1319G>C) in RNA-seq data (RNA: Alt/total = 85/94 (90%)) indicates the reduced expression of the allele carrying the large deletion. IGV from the mapped reads of the patient’s WGS data shows the heterozygous intragenic deletion (chr20:3921864–3923293del). (E) Family pedigree; *SERAC1* compound heterozygous pathogenic variants identified after RNA-seq and trio-WGS are indicated. (F) Brain MRI in T2-weighted images showed bilateral basal ganglia and thalamus abnormalities (red arrows) and cerebral and cerebellar atrophy. (G) Gene-level volcano plot, showing the significance versus z-score of the gene expression in the patient. *SERAC1* is highlighted as the underexpressed outlier harboring two pathogenic variants. (H) Schematic representation of the *SERAC1* gene structure, showing the position of the large deletion–insertion (chr6:158142373–158145953delinsTAGAGTGCTTCAAGTGTGTTG, c.487 + 831\_738 + 683delinsAACAACTTGAAAGCACTCTA) and stop-gain variant (chr6:158128207G>A, c.916C>T) identified in the patient. IGV from the mapped reads of the patient’s WGS data shows the heterozygous chr6:158142373–158145953delinsTAGAGTGCTTCAAGTTGTTG. *SERAC1* is transcribed from the reverse strand but, for consistency, is depicted in the forward orientation in the schematic. WES, whole-exome sequencing; WGS, whole-genome sequencing; IGV, integrative genomics viewer.

fibroblasts and achieved a diagnostic efficiency of 25.4% (18/71) in patients with primary neurological phenotypes [48].

Across 233 curated pathogenic variants with aberrant RNA phenotypes, half resided in noncoding regions and half in coding regions (Figure 7, Table S4). The most common noncoding variants were deep intronic variants (19%), whereas missense (15%) and frameshift (12%) variants predominated among coding alterations.

## 4 | Discussion

This study demonstrates that integrating RNA sequencing into the genomic diagnostic workflow substantially enhances the resolution of genetic testing for neurometabolic diseases. Unlike the genome, the transcriptome is cell-type specific. Fibroblast cell lines, used in this study, offer an optimal balance between accessibility and the breadth of expressed genes [14]. In our fibroblast-derived transcriptome dataset, 15,536 genes were detected, covering 67% of Mendelian disease-associated genes and 90% of genes implicated in mitochondrial disorders. By combining RNA-seq with WES and WGS in a large, clinically well-characterized pediatric cohort, we achieved a molecular diagnosis in 25% of previously unresolved cases. Beyond identifying additional pathogenic variants, our data provide functional insights into how diverse classes of genomic variants—coding, noncoding, and structural—disrupt gene expression, splicing, or transcript stability.

### 4.1 | Functional Validation of Variants of Uncertain Significance

Genome-wide sequencing has transformed rare disease diagnostics but has also revealed the complexity of variant interpretation. The large number of variants identified in each patient demands rigorous classification frameworks such as ACMG/AMP, supported by curated resources including ClinGen, ClinVar, and disease-specific databases. Such strict standards minimize false-positive diagnoses but also increase the number of candidate variants of uncertain significance (VUS) [50]. In

ClinVar, more than half of likely pathogenic variants are protein-truncating variants (PTVs), whereas missense variants account for approximately 25%–30% [51]. This discrepancy reflects the greater predictability of loss-of-function variants compared with missense changes rather than their true frequency in the population. Nevertheless, not all PTVs are pathogenic: fewer than 1000 PTVs are classified as benign, while more than 10,000 remain annotated as VUS or as having conflicting interpretations.

Predictions of NMD are also imperfect. The first ~150 nucleotides of the coding sequence, large exons (>400 nt), the last ~55 nucleotides of the penultimate exon, and the final exon are predicted to be NMD-insensitive; however, approximately 30% of PTVs outside these regions still escape NMD [10–12]. Establishing whether a PTV truly induces NMD can change the strength of ACMG/AMP PVS1 evidence [13] and the variant classification. RNA-seq provides direct, empirical evidence by measuring transcriptional consequences and confirming or refuting NMD activity. In our cohort, 15 PTVs were analyzed: Six showed no evidence of NMD, while nine clearly underwent decay. Notably, two PTVs located 70 and 100 nucleotides upstream of the final exon–exon junction, and one in the WARS2 pseudoexon, escaped NMD contrary to rule-based prediction. These findings indicate that a subset of PTVs in ClinVar may be misclassified due to erroneous NMD assumptions and should be re-evaluated using RNA-based evidence.

### 4.2 | Splicing Disruption Beyond Canonical Junctions

Classification of coding variants often relies on *in silico* predictions [52]. Although algorithms have improved and more than 40,000 missense variants are now annotated as (likely) pathogenic, many variants continue to be reclassified as new data emerge [50, 53]. Most coding variant predictors focus on protein structure or conservation but perform poorly for splicing effects, with average precision of only 1%–2% [54]. Splice prediction algorithms such as SpliceAI, Pangolin, and MMSplice perform well for canonical junctions (accuracy 75%–95%) but remain unreliable for near-splice, coding, or intronic regions [54–56]. Our analysis and previously published cohorts indicate

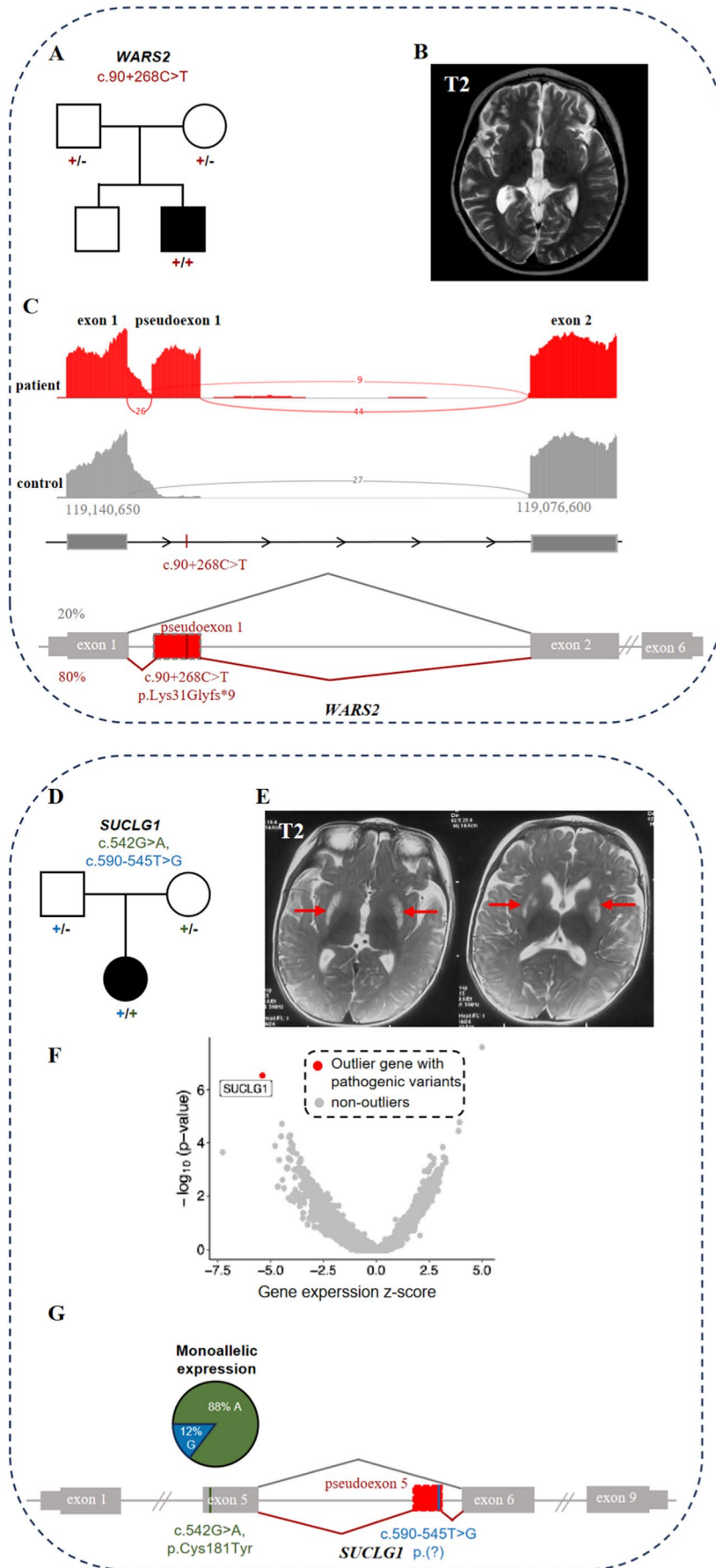


FIGURE 5 | Legend on next page.

**FIGURE 5** | Aberrant splicing due to deep intronic variants, not coerced by WES data. (A–C) *WARS2* patient; (D–G) *SUCLG1* patient. (A) Family pedigree; *WARS2* homozygous deep intronic variant (chr1:119140287G>A, c.90+268C>T) identified after RNA-seq and trio-WGS is indicated. (B) Brain MRI in T2-weighted image, showing mild cerebral atrophy-like changes. (C) *WARS2* sashimi plots of fibroblast-derived RNA-seq data from the *WARS2* patient, and one control, in red and gray, respectively. The numbers on the exon-connecting lines show the number of unnormalized reads spanning the exons. Creation of a pseudoexon in intron 1 due to a homozygous c.90+268C>T deep intronic variant is observed in the patient. The schematic presentation of aberrant (red; ~80% of the reads) and normal (gray; ~20% of the reads) splicing in the patient is depicted at the bottom. Inclusion of the pseudoexon introduces a frameshift with a premature termination codon within the pseudoexon (p.Lys31Glyfs\*9). *WARS2* is transcribed from the reverse strand but, for consistency, is depicted in the forward orientation in the schematic. (D) Family pedigree; *SUCLG1* compound heterozygous variants identified after RNA-seq and trio-WGS are indicated. (E) Brain MRI in T2-weighted images, showing symmetrical basal ganglia lesions (red arrows) and cerebral atrophy. (F) Gene-level volcano plot, showing the significance versus z-score of the gene expression in the patient, with *SUCLG1* highlighted as the most significant underexpressed outlier. (G) Schematic representation of the *SUCLG1* gene structure, showing the missense variant (chr2:84441094C>T, c.542G>A) and deep intronic variant (chr2:84433980A>C, c.590-545T>G). The monoallelic expression of the missense variant (c.542G>A) in RNA-seq data (RNA: Alt/total = 88%) indicates the reduced expression of the allele carrying the intronic variant. Aberrant splicing in association with the deep intronic variant results in the creation of a pseudoexon in intron 5 of the gene (shown in red). *SUCLG1* is transcribed from the reverse strand but, for consistency, is depicted in the forward orientation. WES, whole-exome sequencing; WGS, whole-genome sequencing.

that approximately 80% of pathogenic splice-altering variants are misclassified by such tools (Figure 2C). Thus, both coding and noncoding variants can perturb RNA processing, and transcriptome profiling remains indispensable for mechanistic validation.

### 4.3 | A Founder Synonymous *ECHS1* Variant in East Asian Populations

We identified a recurrent synonymous *ECHS1* variant (c.489G>A, p.Pro163=) that induces partial Exon 4 skipping, leading to decreased transcript levels and functional haploinsufficiency. This allele, enriched in East Asian populations (frequency 1.05%), represents a regional founder mutation. Its hypomorphic nature, still detected by RNA-seq, explains the mild or late-onset presentations in compound heterozygotes and the apparently benign homozygous state observed in reference databases [57]. Parallel studies have reported this variant in 17 unrelated families from Japan, Samoa, and other Pacific populations, supporting a local founder effect [57–59]. Screening of our in-house Leigh syndrome cohort identified 13 additional cases, accounting for 42% of *ECHS1*-related diagnoses in Beijing. This finding reinforces the pathogenicity of the variant in trans with a known disease allele and highlights the importance of population-specific reference data in variant interpretation. Homozygous carriers of the *ECHS1* synonymous variant are functionally similar to individuals heterozygous for loss-of-function variants (LOF). Low constraint of *ECHS1* against LOF variants according to gnomAD explains the high frequency of homozygous carriers of this synonymous variant in East Asian populations. Importantly, *ECHS1* deficiency is a potentially treatable cause of Leigh syndrome, responsive to valine restriction and acetylcysteine supplementation [30, 31].

### 4.4 | Deep Intronic and Structural Variants Revealed by RNA-Seq

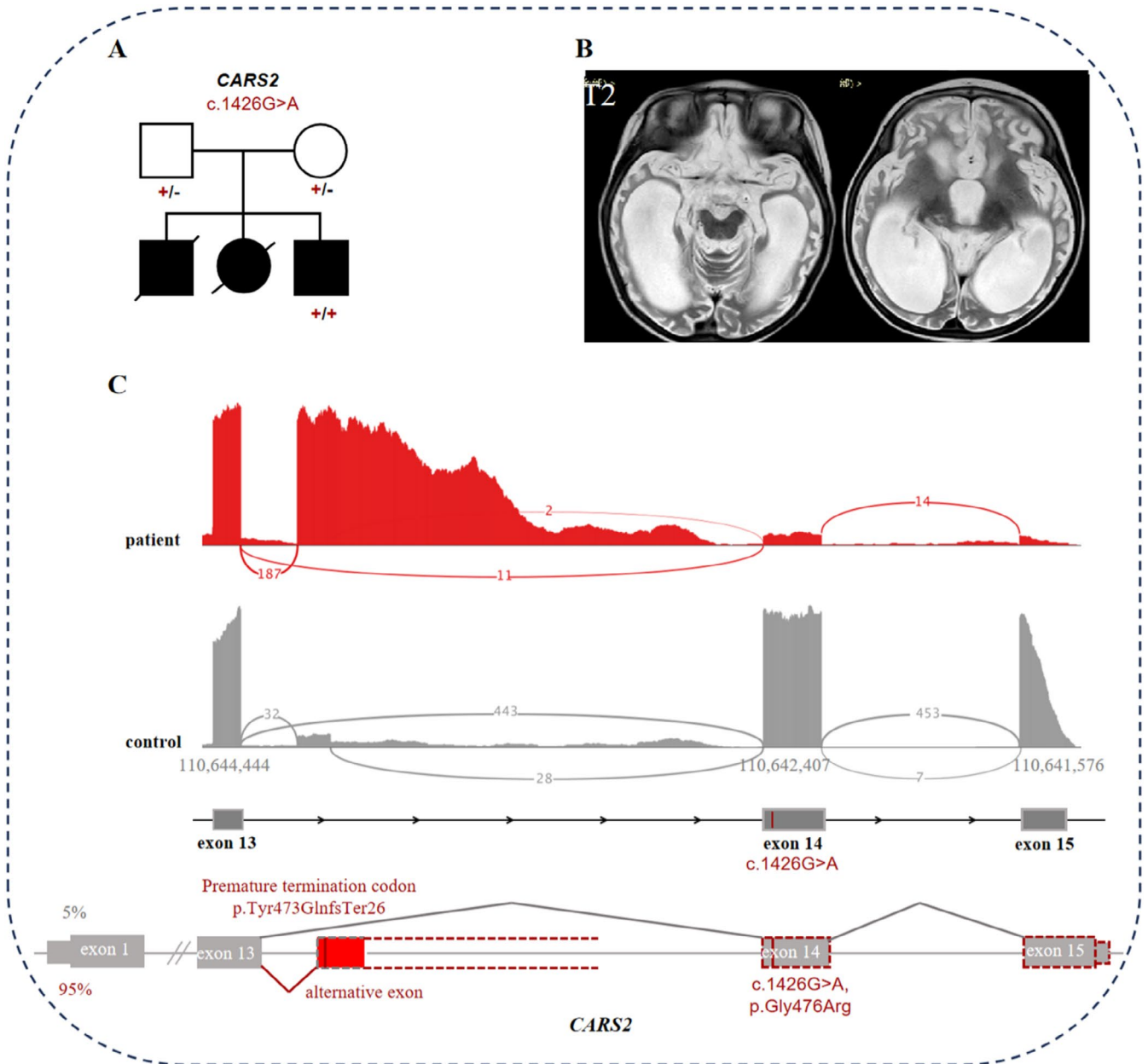
RNA-seq proved particularly effective in detecting pathogenic deep intronic variants invisible to WES and difficult to interpret from WGS. In two cases, pseudoexon creation in *WARS2* and

*SUCLG1* produced premature stop codons and loss of function, illustrating how cryptic splice activation contributes to disease. RNA-seq-guided WGS also uncovered intragenic deletions in *PANK2* and *SERAC1* missed by WES, confirming that aberrant expression outliers can pinpoint underlying structural variants. Together, these examples highlight RNA-seq not only as a confirmatory tool but also as a discovery instrument that directs targeted reanalysis of genome data.

When splice defects lead to transcript degradation via NMD, they may manifest as expression outliers. Treatment of cultured cells with NMD inhibitors, such as cycloheximide or puromycin, can stabilize aberrant transcripts and facilitate their detection. However, a recent study by De Cock et al. (2025) [60] demonstrated that while such treatment increases transcript coverage and refines splice characterization, it does not necessarily increase diagnostic yield.

### 4.5 | Broader Implications and Limitations

Integrating RNA-seq into clinical diagnostics shifts the focus from variant detection to variant interpretation. In our fibroblast dataset, 15,536 genes were expressed, representing two-thirds of all Mendelian disease genes, 90% of mitochondrial genes, and 92% of Leigh syndrome-associated genes, demonstrating that fibroblasts are a robust model for functional testing. However, gene expression is tissue-specific; genes restricted to the brain or muscle remain inaccessible in fibroblasts. Combining fibroblast transcriptomes with induced pluripotent stem cell-derived or neuron-specific RNA-seq may help overcome this limitation. To address tissue-specific expression, Hölzlwimmer et al. (2025) analyzed the GTEx dataset, which comprises transcriptomic and genomic data from 633 individuals across 49 tissues. They developed AbExp, a machine-learning-based approach that predicts tissue-specific gene expression by integrating genomic data with transcriptomes from clinically accessible tissues [54]. In a dystonia cohort, AbExp-based prediction of brain gene expression using genomic data and fibroblast RNA-seq confirmed downregulation in the brain for genes already identified as expression outliers in fibroblasts. However, this approach did not reveal additional aberrant RNA events beyond those detectable



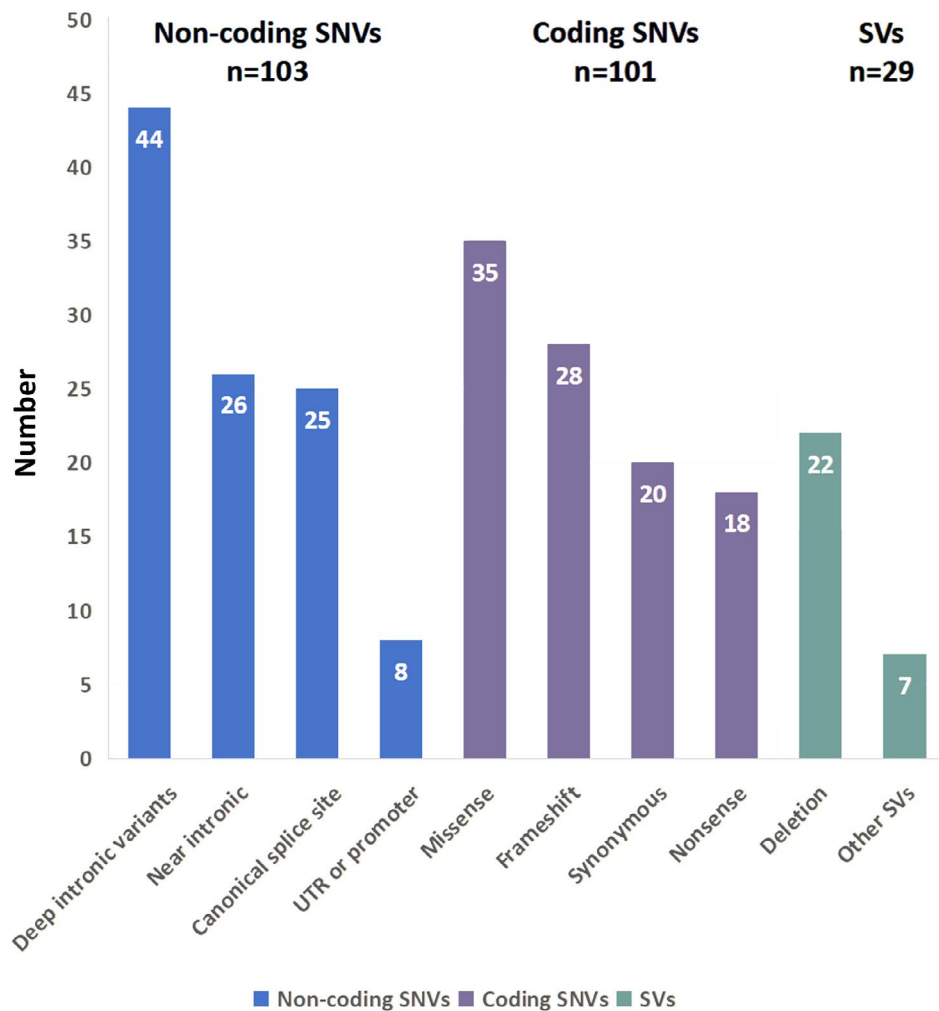
**FIGURE 6** | Homozygous missense variants in *CARS2* cause aberrant splicing. (A) Family pedigree; *CARS2* homozygous missense variant (chr13:110642512C>T, c.1426G>A) identified after RNA-seq and WES analysis is indicated. (B) Brain MRI shows severe cerebral and cerebellar atrophy. (C) *CARS2* sashimi plots of fibroblast-derived RNA-seq data from the *CARS2* patient, and one control, in red and gray, respectively. The numbers on the exon-connecting lines show the number of unnormalized reads spanning the exons. Activation of the alternative splice site and the intron retention is observed in the patient. The schematic presentation of aberrant (red; ~95% of the reads) and normal (gray; ~5% of the reads) splicing in the patient is depicted at the bottom. The c.1426G>A variant in exon 14 reduced the usage of the Exon 14 acceptor splice site and activated the alternative splice site within the intron, resulting in a premature termination codon and reduced expression of Exons 14 and 15. The total expression of the gene is not affected. *CARS2* is transcribed from the reverse strand but, for consistency, is depicted in the forward orientation. WES, whole-exome sequencing.

in fibroblast cell lines, supporting the utility of fibroblast cell lines for functional studies in neurological disorders [61].

Despite the diagnostic efficacy of RNA-seq, many patients remained genetically unsolved, suggesting additional pathogenic mechanisms. In some mitochondrial disorders, *de novo* missense variants can be causative, as recently shown for a number of ATP synthase subunits [62, 63]; they can be inherited dominantly and with reduced penetrance or cause Leigh syndrome

through digenic interactions between mtDNA and nuclear genes [64]. Rare combinations, such as biallelic *DNAJC30* variants plus additional pathogenic Complex I variants, illustrate the diversity of inheritance models [65, 66]. However, such events are extremely rare, and their identification requires statistical analysis of a large cohort of patients.

Future integration of long-read sequencing and complementary omics, including proteomics and methylomics,



**FIGURE 7** | Variant types underlying the aberrant RNA phenotypes curated from this study and previous reports. The figure is patient based (the same variant in X patients counted X times). A total of 233 variants with aberrant RNA-seq events, including noncoding ( $n=103$ , 44%), coding single-nucleotide variants (SNVs) ( $n=101$ , 44%), and structural variants (SVs) ( $n=29$ , 12%), were identified. Noncoding SNVs included deep intronic variants ( $n=44$ ), near splice sites ( $\pm 3$  to  $\pm 10$  bp) ( $n=26$ ), canonical splice sites ( $n=25$ ), UTR ( $n=6$ ), and enhancer ( $n=2$ ). Coding SNVs included missense ( $n=35$ ), frameshift ( $n=28$ ), synonymous ( $n=20$ ), and nonsense ( $n=18$ ). Among the structural variants (SVs), deletion was the most common ( $n=22$ ), followed by complex SV ( $n=2$ ), inversion–deletion, inversion, duplication, insertion, and insertion–deletion.

promises to further enhance diagnostic power. Proteomic studies have already demonstrated the ability to capture protein-destabilizing missense effects and reveal pathogenic variants not prioritized genomically [67, 68]. Understanding underexpression outliers can be challenging. Ozaki et al. (2024) required long-read WGS to identify a repeat expansion in the *NAXE* promoter as the cause of reduced expression in a patient with mitochondrial encephalopathy [69]. Moreover, epigenetic mechanisms can contribute to gene expression outliers, as exemplified by the discovery of promoter hypermethylation of *NDUFAF2* and its epigenetic silencing due to *MORC2* mutations in *trans*, associated with Leigh syndrome [70]. Indeed, aberrant expression of *NDUFAF2* in two patients with Leigh syndrome within our cohort has been identified, which was caused by *MORC2* de novo mutations. A more nonspecific *trans* effect on gene expression was recently reported in a patient with mitochondrial disease and pathogenic variants in *RNU4-2*. *RNU4-2* encodes a small nuclear RNA that plays a crucial role in the major spliceosome, which is

the cellular machinery responsible for removing introns from pre-mRNA [71]. RNA-seq can also reveal systematic *trans*-splicing defects. By analyzing excess intron retention outliers, Arriaga et al. (2025) [72] identified four patients with biallelic variants in another RNA gene *RNU4ATAC*. Such examples highlight how RNA-seq captures both cis- and trans-acting transcriptomic perturbations and variant-to-function interpretation, broadening its diagnostic reach.

## 5 | Conclusion

Our results establish RNA-seq as a decisive functional complement to exome and genome sequencing in mitochondrial disease diagnostics. By revealing cryptic splicing defects, NMD escape, and noncoding variant effects, transcriptome profiling bridges the gap between genotype and phenotype, providing direct functional validation within a clinically relevant timeframe. Beyond its immediate diagnostic impact, the transcriptome analysis

deepens mechanistic understanding of disease pathology and paves the way toward comprehensive, multiomic precision medicine in Mendelian disorders.

### Author Contributions

F.F., H.P., Z.L., X.D., and F.P. contributed to the conception and design of the study. Z.L., X.D., J.W., C.B., C.X., and Y.Z. performed the experiments and analyzed the data. Z.L., X.D., F.P., C.X., Y.Z., Z.Z., Y.Z., T.L., M.P., J.W., M.S., T.S., X.H., M.S., D.S., R.D., H.J., M.X., H.P., and F.F. contributed to the acquisition and analysis of data. F.F., H.P., Z.L., X.D., and F.P. contributed to drafting the text or preparing the figures.

### Acknowledgments

We thank patients and their families for participation in this study. We thank René Santer for the critical reading of the manuscript.

This study was supported by the National Natural Science Foundation of China (Grant No. 82271493), the R&D Program of Beijing Municipal Education Commission (Grant No. KZ202210025033), and the Chinese Institutes for Medical Research, Beijing (Grant No. CX24PY27). H.P., F.P., and M.P. were supported by the BMBF (German Federal Ministry of Education and Research) through the mitoNET German Network for Mitochondrial Diseases (grant number 01GM1906B), PerMiM Personalized Mitochondrial Medicine (grant number 01KU2016A), the EJP RD project GENOMIT (01GM1920A, genomit.eu), and the German Center for Child and Adolescent Health (DZKJ) under the funding code 01GL2406B.

### Funding

This work was supported by National Natural Science Foundation of China; Beijing Municipal Education Commission; Ministry of Education, Republic of China (Taiwan); German Federal Ministry for Research, Technology and Space.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The data needed to reproduce the figures are available upon request. Anonymized data not provided in the article may be shared at the request of any qualified investigator to replicate the procedures and results.

### References

1. S. L. Stenton and H. Prokisch, "Genetics of Mitochondrial Diseases: Identifying Mutations to Help Diagnosis," *eBioMedicine* 56 (2020): 102784.
2. S. L. Stenton, M. Shimura, D. Piekutowska-Abramczuk, et al., "Diagnosing Pediatric Mitochondrial Disease: Lessons From 2000 Exomes," *medRxiv* (2021): 2021.2006.2021.21259171.
3. M. M. Clark, Z. Stark, L. Farnaes, et al., "Meta-Analysis of the Diagnostic and Clinical Utility of Genome and Exome Sequencing and Chromosomal Microarray in Children With Suspected Genetic Diseases," *NPJ Genomic Medicine* 3 (2018): 16.
4. B. J. Livesey and J. A. Marsh, "Updated Benchmarking of Variant Effect Predictors Using Deep Mutational Scanning," *Molecular Systems Biology* 19, no. 8 (2023): e11474.
5. T. V. Riepe, M. Khan, S. Roosing, F. P. M. Cremers, and P. A. C. 't Hoen, "Benchmarking Deep Learning Splice Prediction Tools Using Functional Splice Assays," *Human Mutation* 42, no. 7 (2021): 799–810.

6. P. D. Stenson, M. Mort, E. V. Ball, et al., "The Human Gene Mutation Database: Towards a Comprehensive Repository of Inherited Mutation Data for Medical Research, Genetic Diagnosis and Next-Generation Sequencing Studies," *Human Genetics* 136, no. 6 (2017): 665–677.
7. M. Ma, Y. Ru, L. S. Chuang, et al., "Disease-Associated Variants in Different Categories of Disease Located in Distinct Regulatory Elements," *BMC Genomics* 16 Suppl 8, no. Suppl 8 (2015): S3.
8. R. Chong, K. D. Insigne, D. Yao, et al., "A Multiplexed Assay for Exon Recognition Reveals That an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions," *Molecular Cell* 73, no. 1 (2019): 183–194.e188.
9. R. Soemedi, K. J. Cygan, C. L. Rhine, et al., "Pathogenic Variants That Alter Protein Code Often Disrupt Splicing," *Nature Genetics* 49, no. 6 (2017): 848–855.
10. T. Lappalainen, M. Sammeth, M. R. Friedlander, et al., "Transcriptome and Genome Sequencing Uncovers Functional Variation in Humans," *Nature* 501, no. 7468 (2013): 506–511.
11. R. I. Torene, M. J. Guillen Sacoto, F. Millan, et al., "Systematic Analysis of Variants Escaping Nonsense-Mediated Decay Uncovers Candidate Mendelian Diseases," *American Journal of Human Genetics* 111, no. 1 (2024): 70–81.
12. M. A. Rivas, M. Pirinen, D. F. Conrad, et al., "Human Genomics. Effect of Predicted Protein-Truncating Genetic Variants on the Human Transcriptome," *Science* 348, no. 6235 (2015): 666–669.
13. A. N. Abou Tayoun, T. Pesaran, M. T. DiStefano, et al., "Recommendations for Interpreting the Loss of Function PVS1 ACMG/AMP Variant Criterion," *Human Mutation* 39, no. 11 (2018): 1517–1524.
14. V. A. Yepez, M. Gusic, R. Kopajtich, et al., "Clinical Implementation of RNA Sequencing for Mendelian Disease Diagnostics," *Genome Medicine* 14, no. 1 (2022): 38.
15. P. Witters, A. Saada, T. Honzik, et al., "Revisiting Mitochondrial Diagnostic Criteria in the New Era of Genomics," *Genetics in Medicine* 20, no. 4 (2018): 444–451.
16. E. Morava, L. van den Heuvel, F. Hol, et al., "Mitochondrial Disease Criteria: Diagnostic Applications in Children," *Neurology* 67, no. 10 (2006): 1823–1826.
17. S. Richards, N. Aziz, S. Bale, et al., "Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology," *Genetics in Medicine* 17, no. 5 (2015): 405–424.
18. K. Jaganathan, S. Kyriazopoulou Panagiotopoulou, J. F. McRae, et al., "Predicting Splicing From Primary Sequence With Deep Learning," *Cell* 176, no. 3 (2019): 535–548.e524.
19. L. C. Walker, M. Hoya, G. A. R. Wiggins, et al., "Using the ACMG/AMP Framework to Capture Evidence Related to Predicted and Observed Impact on Splicing: Recommendations From the ClinGen SVI Splicing Subgroup," *American Journal of Human Genetics* 110, no. 7 (2023): 1046–1067.
20. A. Dobin, C. A. Davis, F. Schlesinger, et al., "STAR: Ultrafast Universal RNA-Seq Aligner," *Bioinformatics* 29, no. 1 (2013): 15–21.
21. V. A. Yepez, C. Mertes, M. F. Muller, et al., "Detection of Aberrant Gene Expression Events in RNA Sequencing Data," *Nature Protocols* 16, no. 2 (2021): 1276–1296.
22. S. Anders, P. T. Pyl, and W. Huber, "HTSeq—A Python Framework to Work With High-Throughput Sequencing Data," *Bioinformatics* 31, no. 2 (2015): 166–169.
23. F. Brechtmann, C. Mertes, A. Matuseviciute, et al., "OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data," *American Journal of Human Genetics* 103, no. 6 (2018): 907–917.

24. J. T. Robinson, H. Thorvaldsdottir, W. Winckler, et al., "Integrative Genomics Viewer," *Nature Biotechnology* 29, no. 1 (2011): 24–26.
25. C. Mertes, I. F. Scheller, V. A. Yopez, et al., "Detection of Aberrant Splicing Events in RNA-Seq Data Using FRASER," *Nature Communications* 12, no. 1 (2021): 529.
26. S. Anders and W. Huber, "Differential Expression Analysis for Sequence Count Data," *Genome Biology* 11, no. 10 (2010): R106.
27. M. I. Love, W. Huber, and S. Anders, "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With DESeq2," *Genome Biology* 15, no. 12 (2014): 550.
28. H. Peters, N. Buck, R. Wanders, et al., "ECHS1 Mutations in Leigh Disease: A New Inborn Error of Metabolism Affecting Valine Metabolism," *Brain* 137, no. Pt 11 (2014): 2903–2908.
29. T. B. Haack, C. B. Jackson, K. Murayama, et al., "Deficiency of ECHS1 Causes Mitochondrial Encephalopathy With Cardiac Involvement," *Annals of Clinical Translational Neurology* 2, no. 5 (2015): 492–509.
30. Y. Li, T. Wu, Y. Li, et al., "Ectopic Protein Lysine Methacrylation Contributes to Defects Caused by Loss of HIBCH or ECHS1," *Cell Reports* 44, no. 3 (2025): 115379.
31. F. Distelmaier, T. B. Haack, S. B. Wortmann, J. A. Mayr, and H. Prokisch, "Treatable Mitochondrial Diseases: Cofactor Metabolism and Beyond," *Brain* 140, no. 2 (2017): e11.
32. F. Yang, J. Wang, Z. Yang, Z. Ren, and F. Zeng, "PANK2 p.A170fs:A Novel Pathogenic Mutation, Compound With PANK2 p.R440P, Causing Pantothenate Kinase Associated Neurodegeneration in a Chinese Family," *International Journal of Neuroscience* 132, no. 6 (2022): 582–588.
33. S. H. Kim, Y. H. Sung, K. H. Park, et al., "Novel Compound Heterozygous Mutations in the Pantothenate Kinase 2 Gene in a Korean Patient With Atypical Pantothenate Kinase Associated Neurodegeneration," *Journal of Movement Disorders* 2, no. 1 (2009): 45–47.
34. S. B. Wortmann, F. M. Vaz, T. Gardeitchik, et al., "Mutations in the Phospholipid Remodeling Gene SERAC1 Impair Mitochondrial Function and Intracellular Cholesterol Trafficking and Cause Dystonia and Deafness," *Nature Genetics* 44, no. 7 (2012): 797–802.
35. L. S. Kremer, D. M. Bader, C. Mertes, et al., "Genetic Diagnosis of Mendelian Disorders via RNA Sequencing," *Nature Communications* 8, no. 1 (2017): 15824.
36. B. B. Cummings, J. L. Marshall, T. Tukiainen, et al., "Improving Genetic Diagnosis in Mendelian Disease With Transcriptome Sequencing," *Science Translational Medicine* 9, no. 386 (2017): eaal5209.
37. H. D. Gonorazky, S. Naumenko, A. K. Ramani, et al., "Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease," *American Journal of Human Genetics* 104, no. 3 (2019): 466–483.
38. L. Youssefian, A. H. Saeidian, F. Palizban, et al., "Whole-Transcriptome Analysis by RNA Sequencing for Genetic Diagnosis of Mendelian Skin Disorders in the Context of Consanguinity," *Clinical Chemistry* 67, no. 6 (2021): 876–888.
39. J. Dekker, R. Schot, M. Bongaerts, et al., "Web-Accessible Application for Identifying Pathogenic Transcripts With RNA-Seq: Increased Sensitivity in Diagnosis of Neurodevelopmental Disorders," *American Journal of Human Genetics* 110, no. 2 (2023): 251–272.
40. L. Fresard, C. Smail, N. M. Ferraro, et al., "Identification of Rare-Disease Genes Using Blood Transcriptome Sequencing and Large Control Cohorts," *Nature Medicine* 25, no. 6 (2019): 911–919.
41. H. Lee, A. Y. Huang, L. K. Wang, et al., "Diagnostic Utility of Transcriptome Sequencing for Rare Mendelian Diseases," *Genetics in Medicine* 22, no. 3 (2020): 490–499.
42. S. Rentas, K. S. Rathi, M. Kaur, et al., "Diagnosing Cornelia de Lange Syndrome and Related Neurodevelopmental Disorders Using RNA Sequencing," *Genetics in Medicine* 22, no. 5 (2020): 927–936.
43. D. R. Murdock, H. Dai, L. C. Burrage, et al., "Transcriptome-Directed Analysis for Mendelian Disease Diagnosis Overcomes Limitations of Conventional Genomic Testing," *Journal of Clinical Investigation* 131, no. 1 (2021): e141500.
44. K. Riquin, B. Isidor, S. Mercier, et al., "Integrating RNA-Seq Into Genome Sequencing Workflow Enhances the Analysis of Structural Variants Causing Neurodevelopmental Disorders," *Journal of Medical Genetics* 61, no. 1 (2024): 47–56.
45. C. Jaramillo Oquendo, H. A. Wai, W. I. Rich, et al., "Identification of Diagnostic Candidates in Mendelian Disorders Using an RNA Sequencing-Centric Approach," *Genome Medicine* 16, no. 1 (2024): 110.
46. S. E. Hong, J. Kneissl, A. Cho, et al., "Transcriptome-Based Variant Calling and Aberrant mRNA Discovery Enhance Diagnostic Efficiency for Neuromuscular Diseases," *Journal of Medical Genetics* 59, no. 11 (2022): 1075–1081.
47. E. Colin, Y. Duffourd, M. Chevarin, et al., "Stepwise Use of Genomics and Transcriptomics Technologies Increases Diagnostic Yield in Mendelian Disorders," *Frontiers in Cell and Development Biology* 11 (2023): 1021920.
48. S. Li, S. Zhao, J. C. Sinson, et al., "The Clinical Utility and Diagnostic Implementation of Human Subject Cell Transdifferentiation Followed by RNA Sequencing," *American Journal of Human Genetics* 111, no. 5 (2024): 841–862.
49. A. R. Deshwar, K. E. Yuki, H. Hou, et al., "Trio RNA Sequencing in a Cohort of Medically Complex Children," *American Journal of Human Genetics* 110, no. 5 (2023): 895–900.
50. E. Bonetti, G. Tini, and L. Mazzarella, "Accuracy of Renovo Predictions on Variants Reclassified Over Time," *Journal of Translational Medicine* 22, no. 1 (2024): 713.
51. E. Perez-Palma, M. Gramm, P. Nurnberg, et al., "Simple ClinVar: An Interactive Web Server to Explore and Retrieve Gene and Disease Variants Aggregated in ClinVar Database," *Nucleic Acids Research* 47, no. W1 (2019): W99–W105.
52. E. H. Wilcox, M. Sarmady, B. Wulf, et al., "Evaluating the Impact of In Silico Predictors on Clinical Variant Classification," *Genetics in Medicine* 24, no. 4 (2022): 924–930.
53. Y. Kobayashi, E. Chen, F. M. Facio, et al., "Clinical Variant Reclassification in Hereditary Disease Genetic Testing," *JAMA Network Open* 7, no. 11 (2024): e2444526.
54. F. R. Holzlwimmer, J. Lindner, G. Tsitsiridis, et al., "Aberrant Gene Expression Prediction Across Human Tissues," *Nature Communications* 16, no. 1 (2025): 3061.
55. C. Capitanchik, O. G. Wilkins, N. Wagner, J. Gagneur, and J. Ule, "From Computational Models of the Splicing Code to Regulatory Mechanisms and Therapeutic Implications," *Nature Reviews. Genetics* 26, no. 3 (2025): 171–190.
56. C. Smith and J. O. Kitzman, "Benchmarking Splice Variant Prediction Algorithms Using Massively Parallel Splicing Assays," *Genome Biology* 24, no. 1 (2023): 294.
57. M. T. Simon, S. S. Eftekharian, S. Ferdinandusse, et al., "ECHS1 Disease in Two Unrelated Families of Samoan Descent: Common Variant - Rare Disorder," *American Journal of Medical Genetics. Part A* 185, no. 1 (2021): 157–167.
58. I. Bernhardt, L. E. Frajman, B. Ryder, et al., "Further Delineation of Short-Chain Enoyl-CoA Hydratase Deficiency in the Pacific Population," *Molecular Genetics and Metabolism* 142, no. 3 (2024): 108508.

59. Y. Kishita, A. Sugiura, T. Onuki, et al., “Strategic Validation of Variants of Uncertain Significance in ECHS1 Genetic Testing,” *Journal of Medical Genetics* 60, no. 10 (2023): 1006–1015.
60. L. De Cock, E. D’Haenens, L. Vantomme, et al., “Cracking Rare Disorders: A New Minimally Invasive RNA-Seq Protocol,” *NPJ Genomic Medicine* 10, no. 1 (2025): 45.
61. A. Saparov, I. Dzinovic, T. Brunet, et al., “Fibroblast Transcriptomics in Molecular Diagnostics of a Comprehensive Dystonia Cohort,” *Annals of Neurology* (2026).
62. S. M. Fielder, M. W. Friederich, D. H. Hock, et al., “Dominant Negative ATP5F1A Variants Disrupt Oxidative Phosphorylation Causing Neurological Disorders,” *EMBO Molecular Medicine* 17, no. 10 (2025): 2562–2585.
63. A. Nasca, N. E. Mencacci, F. Invernizzi, et al., “Variants in ATP5F1B Are Associated With Dominantly Inherited Dystonia,” *Brain* 146, no. 7 (2023): 2730–2738.
64. B. Blickhauser, S. L. Stenton, C. M. Neuhofer, et al., “Digenic Leigh Syndrome on the Background of the m.11778G>A Leber Hereditary Optic Neuropathy Variant,” *Brain* 147, no. 6 (2024): 1967–1974.
65. M. Zawadzka, M. Krygier, M. Pawlowicz, et al., “Expanding the Phenotype of DNAJC30-Associated Leigh Syndrome,” *Clinical Genetics* 102, no. 5 (2022): 438–443.
66. S. L. Stenton, M. Tesarova, N. L. Sheremet, et al., “DNAJC30 Defect: A Frequent Cause of Recessive Leber Hereditary Optic Neuropathy and Leigh Syndrome,” *Brain* 145, no. 5 (2022): 1624–1631.
67. S. P. Correia, M. F. Moedas, L. S. Taylor, et al., “Quantitative Proteomics of Patient Fibroblasts Reveal Biomarkers and Diagnostic Signatures of Mitochondrial Disease,” *JCI Insight* 9, no. 20 (2024): e178645.
68. D. H. Hock, N. J. Caruana, L. N. Semcesen, et al., “Untargeted Proteomics Enables Ultra-Rapid Variant Prioritisation in Mitochondrial and Other Rare Diseases,” *Genome Medicine* 17, no. 1 (2025): 58.
69. K. Ozaki, Y. Yatsuka, Y. Oyazato, et al., “Biallelic GGGCC Repeat Expansion Leading to NAXE-Related Mitochondrial Encephalopathy,” *NPJ Genomic Medicine* 9, no. 1 (2024): 48.
70. F. Peymani, T. Ebihara, D. Smirnov, et al., “Pleiotropic Effects of MORC2 Derive From Its Epigenetic Signature,” *Brain* 149, no. 1 (2025): 163–177.
71. C. Charenton, M. E. Wilkinson, and K. Nagai, “Mechanism of 5’ Splice Site Transfer for Human Spliceosome Activation,” *Science* 364, no. 6438 (2019): 362–367.
72. M. T. Arriaga, R. Mendez, R. A. Ungar, et al., “Transcriptome-Wide Outlier Approach Identifies Individuals With Minor Spliceopathies,” *medRxiv* 112 (2025): 2458–2475.

### Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Summary of cases diagnosed by RNA-seq and with candidate variants but no aberrant RNA phenotype. **Table S2:** Summary of the biochemical analysis. **Table S3:** Summary of clinical application of RNA-seq for rare disorders by others and this study. **Table S4:** Summary of variants causing aberrant RNA phenotypes identified in this study and reported in the literature.