International Journal of
Molecular Sciences

MDPI

Article

Deep intronic ETFDH variants represent a recurrent pathogenic event in multiple acyl-CoA dehydrogenase deficiency.

Stefania Martino [1,†], Antonella Turchiano [1,†], Pietro D'Addabbo [2,†], Francesca Clementina Radio [3], Alessandro Bruselles [4], Viviana Cordeddu [4], Cecilia Mancini [3], ,Alessandro Stella [1], Nicola Laforgia [5], Donatella Capodiferro [5], Daniela Zuccarello [6], Simonetta Simonetti [7], Rita Fischetto [8], Rosanna Bagnulo [1], Orazio Palumbo [9], Flaviana Marzano [10], Cinzia Forleo [11], Marilidia Piglionica [12], Ornella Tabaku [1], Antonella Garganese [12], Michele Stasi [1], Marco Tartaglia [3], Graziano Pesole [2,10], Nicoletta Resta [1] and *

[1]Medical Genetics Unit, Department of Precision and Regenerative Medicine and Ionian Area (DiMePRe-J), University of Bari "Aldo Moro", 70124, Bari, Italy.
[2] Department of Biosciences, Biotechnologies & Environment, University of Bari "Aldo Moro", Via Edoardo Orabona 4, 70125 Bari, Italy.
[3]Molecular Genetics and Functional Genomics, Ospedale Pediatrico Bambino Gesù, IRCCS, Viale di San Paolo 15, 00146 Rome, Italy.
[4]Department of Oncology and Molecular Medicine, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy.
[5]Section of Neonatology and Neonatal Intensive Care Unit, Department of Interdisciplinary Medicine, "Aldo Moro" University of Bari, 70121 Bari, Italy.
[6]Clinical Genetics and Epidemiology Unit, University Hospital of Padova, Via Giustiniani 3, 35128 Padova, Italy.
[7]Clinical Pathology and Neonatal Screening, Hospital "Giovanni XXIII", University Hospital Consortium Corporation Polyclinics of Bari, Bari, Italy
[8]Clinical Genetics Unit, Department of Pediatric Medicine, XXIII Children's Hospital, Bari, Giovanni, Italy.
[9]Division of Medical Genetics, Fondazione IRCCS-Casa Sollievo della Sofferenza, San Giovanni Rotondo Foggia, Italy.
[10] Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, Consiglio Nazionale delle Ricerche, via Amendola 122/O, 70126 Bari, Italy
[11] University Cardiology Unit, Interdisciplinary Department of Medicine, University of Bari Aldo Moro, Polyclinic University Hospital, Bari, Italy.
[12] Medical Genetic Unit, University Hospital Consortium Corporation Policlinics of Bari, 70124 Bari, Italy
* Correspondence: nicoletta.resta@uniba.it; Tel.: +39-080-559-4247
†These authors contributed equally to this work.

Abstract: Multiple acyl-CoA dehydrogenase deficiency (MADD) is a rare inborn error of metabolism affecting fatty acid and amino acid oxidation with an incidence of 1 in 200,000 live births. MADD has three clinical phenotypes: severe neonatal-onset with or without congenital anomalies, and a milder late-onset form. Clinical diagnosis is supported by urinary organic acid and blood acylcarnitine analysis, detected using tandem mass spectrometry in newborn screening programs. MADD is an autosomal recessive trait caused by biallelic mutations in the ETFA, ETFB, and ETFDH genes, encoding the alpha and beta subunits of the electron transfer flavoprotein (ETF) and ETF-coenzyme Q oxidoreductase enzymes. Despite significant advancements in sequencing techniques, many patients remain undiagnosed, impacting their access to clinical care and genetic counseling. In this report, we achieved a definitive molecular diagnosis in a newborn by combining whole genome sequencing (WGS) with RNA sequencing (RNAseq). Whole exome sequencing or next-generation gene panels, fail to detect variants, possibly affecting splicing, in deep intronic regions. Here, we report a unique deep intronic mutation in intron 1 of the ETFDH gene, c.35-959A>G, in a patient with early onset lethal MADD, resulting in pseudo-exon inclusion. The identified variant, is the third mutation reported in this region, highlighting ETFDH intron 1 vulnerability. It cannot be excluded that these intronic sequence features may be more common in other genes than is currently believed. The study highlights the importance of incorporating RNA analysis into genome-wide testing to reveal the functional consequences of intronic mutations.

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

Keywords: ETFDH; deep intronic variant; MADD; genome sequencing; RNA sequencing; pseudo-exon; transcript processing; splicing

1. Introduction

Multiple acyl-CoA dehydrogenase deficiency (MADD, MIM 231680), also known as glutaric acidemia II, is a rare recessive genetically heterogeneous combined disorder of fatty acid and amino acid oxidation, affecting approximately 1 in 200,000 live births with ethnic variations. Three clinical phenotypes with differences in presentation and age of onset have been recognized, including two MADD-severe (MADD-S) neonatal-onset forms, with or without congenital anomalies, and a MADD-mild (MADD-M) late-onset form [1]. MADD-S generally presents with non-ketotic hypoglycemia, hypotonia, hepatomegaly and severe metabolic acidosis within the first 24 hours of life, evolving in death early after birth. The associated congenital anomalies usually include dysplastic kidneys with multiple cysts and facial dysmorphism (e.g., low-set ears, high forehead, hypertelorism and hypoplastic midface), rocker-bottom feet and anomalies of external genitalia. MADD-S without congenital anomalies generally has a neonatal onset with hypotonia, tachypnea, hepatomegaly, metabolic acidosis and hypoketotic hypoglycemia. The majority of individuals dies early after onset. Those who survive show an evolutive cardiomyopathy. MADD-M has a broader clinical spectrum, ranging from intermittent episodes of vomiting, metabolic acidosis and hypoketotic hypoglycemia with or without cardiac involvement in infancy, to acute ketoacidosis and lipid storage myopathy in adolescents/adults. In all forms, urinary organic acid analysis typically reveals various combinations of increased dicarboxylic acids, glutaric acid, ethylmalonic acid, 2-hydroxyglutarate, and glycine conjugates. Blood acylcarnitines show increased C4-C18 species, although patients may be severely carnitine depleted, which can limit the extent of these abnormalities.

MADD can be screened using tandem mass spectrometry, which is an informative tool used in newborn screening programs. The identification of abnormalities, in particular in acylcarnitines levels, may also confirm the diagnosis, and plasma acylcarnitines profiling may suggest a block in fatty acid oxidation before symptoms appear [2]. First-line evaluation is generally provided by newborn screening, whose positive result is confirmed by second generation sequencing approaches. MADD is caused by pathogenic variants in the ETFA (MIM *608053; glutaric acidemia IIA), ETFB (MIM *130410; glutaric acidemia IIB), and ETFDH (MIM *231675; glutaric acidemia IIC) genes, which encode the alpha and beta subunits of the electron transfer flavoprotein (ETF) and ETF-coenzyme Q oxidoreductase [3,4].

Up to date, more than 900 ETFDH variants have been reported in the ClinVar database (last accessed 13th May 2024), including nearly 320 variants that have been classified as either pathogenic or likely pathogenic. Given that the condition is inherited as an autosomal recessive trait, establishing the molecular diagnosis requires the identification of biallelic disease-causing variants. MADD-S is caused by loss-of-function (LoF) variants, including those resulting in truncated forms of the encoded protein or causing aberrant mRNA expression, processing, or stability. On the other hand, missense mutations are generally linked to the milder late-onset presentation of MADD.

The implementation of novel high-throughput technologies, such as second generation sequencing, has greatly improved the diagnostic yield of gene testing in MADD. The present report investigates a patient clinically diagnosed with early-onset MADD where clinical exome sequencing (CES) revealed a single likely pathogenic variant in ETFDH predicting a stop codon loss. Subsequently, chromosome microarray analysis (CMA) and whole genome sequencing (WGS) analysis were carried out. CMA resulted negative for the presence of copy number variations (CNVs) involving the ETFDH gene, while WGS

disclosed the presence of a previously unreported deep intronic variant that was predicted to affect transcript processing, within a region in which pathogenic MADD variants had previously been reported [5,6] . In silico and functional validation analyses confirmed the clinical relevance of the identified intronic variant highlighting the significance of RNA-based investigations in obtaining a conclusive genetic diagnosis, allowing a correct genetic counseling and patient care.

2. Materials and Methods

2.1. Patient

The patient was born at 32 weeks of gestation by urgent caesarean section due to cardiotocographic alterations in a primigravida with a pregnancy complicated by oligohydramnios and intrauterine fetal growth retardation from the 31st week. Family history was negative, and parents were not related. Birth weight was 1470 g (24th centile), length 40 cm (21st centile), head circumference 31 cm (84th centile). Apgar 9-9. Non-invasive ventilation with n-CPAP and parenteral nutrition via umbilical venous catheter were applied. From the third day of life, progressive deterioration of her clinical conditions with hyporeactivity prompted the need of mechanical ventilation. Hyperammonemia (450 mMol/L) and metabolic acidosis with increased anion gap (27 mMol/L) were observed; parenteral solutions were substituted with only 10% glucose and treatment with carglumic acid, bioarginine and sodium bicarbonate was started. She was transferred on the fourth day of life to our NICU, where elevated levels of proline, valine, aminoisobutyrate, ornithine and lysine were found, both in plasma and urine: Additionally, there was a significant increase of glutaric, phenylacetic, adipic and lactic acids and the presence of isovalerylglycine and butyrylglycine in urine. Intravenous carnitine was added to previous treatment. Despite the reduction of ammonemia with a slight improvement of metabolic acidosis, clinical conditions did not improve with evident deep comatose state. Elevated transaminases (GOT 1241 U/L, GPT 227 U/L), CPK 3301 U/L and LDH 5445 U/L were also found. Severe hypoxemia and hypotension were unresponsive to oscillatory ventilation and high doses of intravenous dopamine and dobutamine, leading to her death on the eighth day of life.

2.2. Molecular analyses

Genomic DNA (gDNA) was extracted from proband and parents' peripheral blood (PB) samples by using QIAamp Mini Kit (Qiagen, Hilden Germany), following the manufacturer's instruction. Total RNA was extracted from PB specimens taken from both parents and unaffected sex- and age-matched individuals, collected in PAXgene Blood RNA tubes (PreAnalytiX, QIAGEN), and purified by PAXgene Blood RNA kit (PreAnalytiX, QIAGEN) according to the manufacturer's indications. gDNA and RNA were quantified with a BioSpectrometer Plus instrument (Eppendorf, Hamburg, Germany).

Trio-based CES was performed using the TruSight One Sequencing Panel kit (Illumina, San Diego, CA, USA). gDNA concentration and quality were evaluated by using Qubit dsDNA HS Assay Kit on a Qubit 2.0 Fluorimeter (Invitrogen, Carlsbad, CA, USA), following the manufacturer's instructions. Libraries were prepared by utilizing the NextEra Flex for Enrichment protocol (Illumina) and were sequenced on NextSeq550Dx Illumina platform (Illumina).

Data analysis was performed by using the NextSeq control software and Local Run Manager software, both provided by Illumina (Illumina). Reads were aligned against the human genome reference (GRCh38) by the BWA Aligner software [7]. Variant calling was performed using the Genome Analysis Toolkit (GATK) [8]. A mean coverage depth of 229X was obtained. Variant calling data were analyzed with Geneyx analysis software (Geneyx, Herzliya, Israel). Variants were filtered and prioritized by utilizing HPO terms (hyperammonemia, glutaric aciduria, hyperlactacidemia) [9], using in silico tools (Alamut Visual Plus, MaxEntScan, SpliceAI) and public databases (ClinVar, LOVD, Varsome,

Franklin by genoox, OMIM), and classified following the ACMG criteria [10]. BAM files were visually inspected by the Integrative Genome Viewer software (IGV) and Alamut Visual Plus Genome Viewer (Sophia Genetics, Lausanne, Switzerland), and variants were reported according to the Human Genome Variants Society (HGVS) recommendations [11].

CMA was performed using the CytoScan XON array (Thermo Fisher Scientific, Waltham, MA, USA). The CytoScan XON assay was performed according to the manufacturer's protocol, using 100 ng of DNA. XON array data were analyzed for the presence of intragenic microdeletions/duplications using the Affymetrix Chromosomal Analysis Suite (ChAS) software v.4.3. CNV pathogenicity was assessed using published literature and public databases (Database of Genomic Variant, http://dgv.tcag.ca/dgv/app/home; Clingen, https://clinicalgenome.org/; DECIPHER, https://www.deciphergenomics.org/; OMIM, https://www.omim.org/). Genomic positions, functional annotation on genomic regions and genes affected by CNVs and/or ROHs were derived from the University of California Santa Cruz Genome Browser tracks (http://genome.ucsc.edu/cgi-bin/hgGateway). The clinical significance of each rearrangement was assessed according to the ACMG and Clinical Genome Resource 2020 guidelines [12].

WGS was performed on a NovaSeq 6000 platform (Illumina) as per recommended protocols. Base calling and data analysis were performed using Bcl2FASTQ (Illumina). Paired-end reads mapping to the GRCh38 reference sequence, variant calling and joint genotyping were run using Sentieon v.2023-08 (https://www.sentieon.com). SNPs and short insertions/deletions (InDels) hard filtering were applied using GATK, version 3.8.0 (Broad Institute, Cambridge, England). High quality variants were first filtered by frequency ≤5% in the in-house WGS population matched database. Variants were annotated and filtered against public (gnomAD v.2.1.1, https://gnomad.broadinstitute.org) and in-house (>3,100 population-matched exomes) databases to retain private and rare (MAF < 0.1%) variants with any effect on the coding sequence, and within splice site regions. The predicted functional impact of variants was analyzed by Combined Annotation Dependent Depletion (CADD) v.1.6 [13], M-CAP v.1.3 [14] and InterVar v.2.2.2 algorithms[15] Clinical interpretation followed the ACMG 2015 guidelines[10]. Variants within non-coding regions were annotated and prioritized using Genomiser [16](phenotype data version 2302).

Structural variants were detected using DELLY v.1.1.6[17] and prioritized using AnnotSV v.3.3.2 [18]. WGS metrics and sequencing output are reported in Supplementary Table 1(S1).

RNAseq was performed using a NovaSeq6000 platform (Illumina). Raw sequences were inspected using FASTQC [19] and trimmed using FASTP [20] . High quality reads were aligned onto the GRCh38 assembly of the human genome, using STAR 2.7.11a and providing a list of known gene annotations from GENCODE v43. Detected splicing junctions were classified based on their annotation status in the human transcriptomes. Data were also graphically inspected using IGV and custom tracks on the UCSC genome browser. High quality reads were analyzed by Salmon v1.10.0, also providing the GENCODE v43 annotations, to estimate transcript level abundance from the RNAseq data.

All the clinically relevant variants identified by CES, WGS, and RNAseq were confirmed by Sanger sequencing (Applied Biosystem, Waltham, MA, USA). To validate the identified aberrant transcripts, targeted cDNA assay was carried out. cDNA was synthetized by using the Thermo Scientific Maxima Reverse Transcriptase kit (Thermo Fisher Scientific), following manufacturer's recommendations. cDNA PCRs were performed as previously described[6]. Primer sequences are available upon request. The obtained PCR amplicons were separated via agarose gel electrophoresis to visually inspect the different amplification patterns. The specific primer pairs used to detect RNA isoforms (designed according to RNAseq data) and to validate disease causing exon and intronic variants are listed in Supplementary Table 2 (S2).

The predicted functional consequences of the identified deep intronic variants were inspected using the splicing module of Alamut Visual Plus version 1.6.1, (Sophia Genetics). Splicing donor site (SDS) and splicing acceptor site (SAS) scores were calculated via the Maximum entropy (MaxEnt) of the Burge Laboratory's MaxEntScan web-tool [21]. Finally, SpliceAI (https://spliceailookup.broadinstitute.org) was used to explore gain and/or loss of splicing sites.

3. Results

Trio-based CES revealed the heterozygous and paternally inherited variant c.1852T>C (p.*618Glnext*13) in the ETFDH gene (NM_004453.4), which was subsequently validated by Sanger sequencing (Fig. 1). The variant was classified as likely pathogenic according to the ACMG criteria (PP5, PM2, PM4) and to the LOVD database(https://www.lovd.nl/).

Since the clinical features of the proband well fitted the recessive condition associated with LoF variants in ETFDH, complementary approaches were performed to identify a possible second disease-causing variant in trans. Since XON testing disclosed no CNVs involving the ETFDH gene, trio-based WGS analysis was carried out, identifying a deep intronic variant (c.35-959A>G) within the ETFDH intron 1. Sanger sequencing validated the variant, documenting its maternal transmission (Fig. 1).

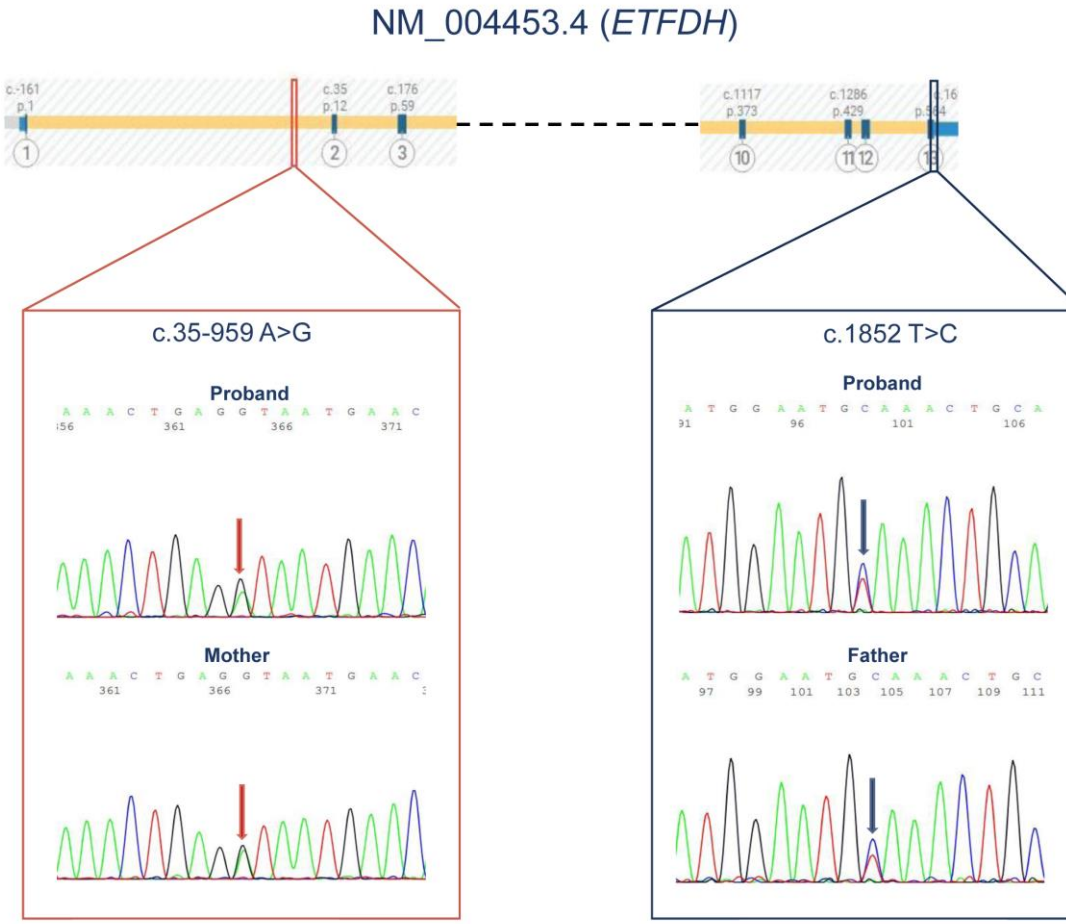

Figure 1. Sanger sequencing results of confirmatory and segregation analysis.

On the top the Alamut Visual Plus (V 1.8.1) screenshot of the first three exons and last four exons the ETFDH gene (NM_004453.4) is depicted. The deep intronic variant found in the patient is in the first intron. The paternally inherited mutation occurred in exon 13, the last of the ETFDH gene. For each variant the electropherograms results are

displayed. The left panel shows the intronic variant both in the proband and in her mother. In the right panel, the presence of the exonic variant, detected in the proband and in the father PB specimen, is shown.

Then, we investigated the potential impact of the deep intronic variant on splicing using a variety of in silico prediction tools. Notably, two others deep intronic variants, c.35-768A>G and c.35-1008T>G, which are located in the same intronic region, were previously shown to cause abnormal mRNA processing [6] by including a cryptic pseudo-exon in the final mRNA coding sequence. Therefore, we first verified the creation/abrogation/modification of ESE sites with the splicing module of Alamut Visual. Similarly to what was observed for the two previously reported intronic variants, the c.35-959A>G variant was predicted to create a novel SF2/ASF ESE and perturb the strength of an SF2/ASF IgM-BRCA1 ESE (prediction score increased from 2.38 to 4.16) (Fig. 2a) Next, we used MaxEnt scan to assess if novel splicing donor or acceptor sites were generated or abolished. The three tested variants similarly created novel SDSs, which outscored the wild-type SDS of intron 1 in all cases (Fig. 2b). The pseudo-exon generated by the three variants was predicted to share a common SAS, which again outscores the wild-type SAS at the end of intron 1. Finally, wild-type and mutated sequences were inspected with Splice AI. Once more, for all the three variants a significant increase of the delta score for the donor site of pseudo-exons (≥0.80) was observed [22](Fig. 2c). Overall, these in silico predictions consistently indicated a putative impact of the identified deep intronic variant on transcript processing.

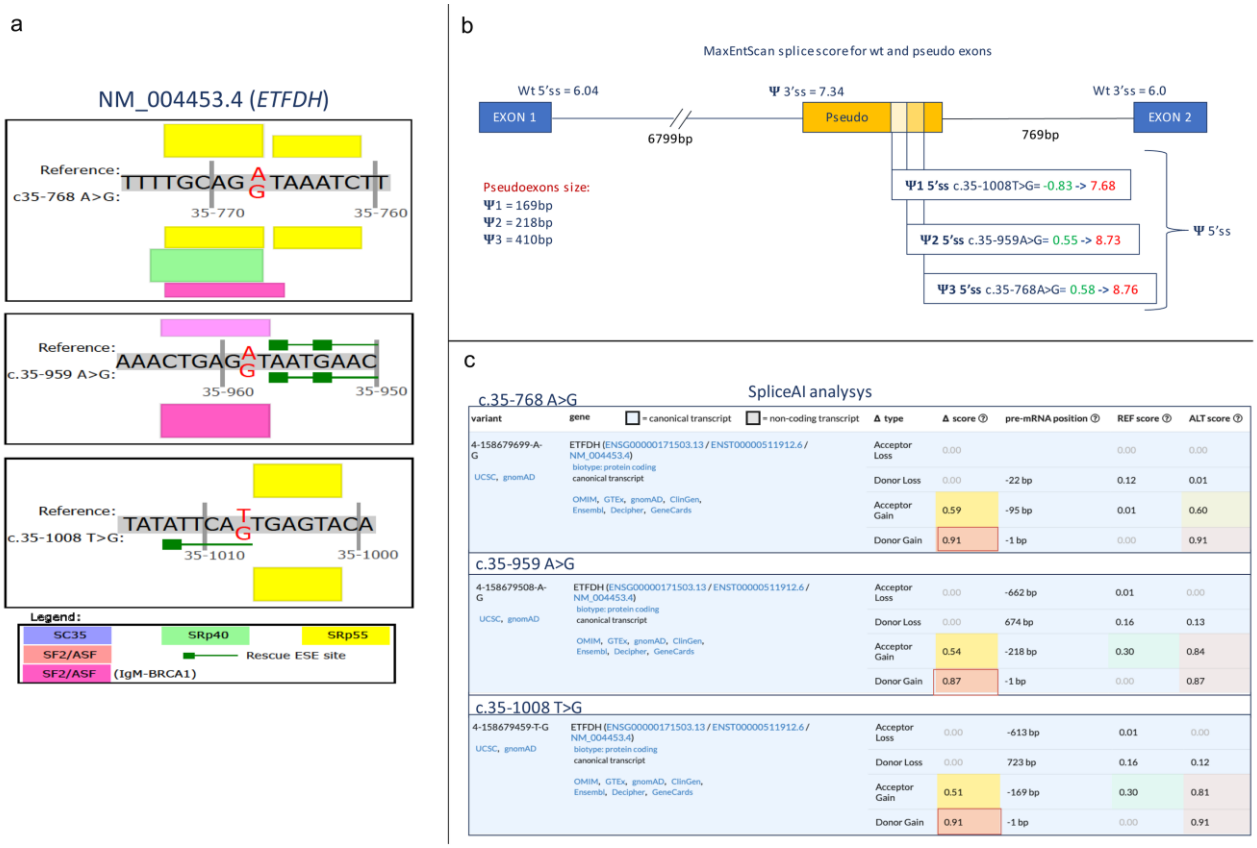

Figure 2. In silico prediction of potential impact of the reported deep intronic variant on splicing. a) Outline of ESE changes determined by the three deep ETFDH intron 1 mutations. Box height is proportional to ESE score as computed by the Alamut splicing module. b) MaxEntScan analysis of 5′ and 3′ splicing site scores for both wild type exon 1 and exon 2 junctions and the pseudo-exon. In green the score of the normal sequence, in red the score of the mutated sequence for the pseudo-exon. c) SpliceAI analysis. The scores for the wild type and mutated sequences are indicated for all the three mutations in ETFDH intron 1. Boxed in red is the difference (i.e. Δ score) between normal and mutated sequence scores for the donor site. The light red color indicates that

the increase of SpliceAI scores for donor sites is highly significant (for all three mutations > 0.80) as reported by Sagakuchi et al.[22].

To validate the in-silico predictions, transcriptomic analyses were conducted using mRNA from parental PB samples. Consistently, splicing junction detections by STAR allowed the identification of an aberrant processing of the maternal ETFDH transcript, which was not present in the GENCODE v43 database. The variant ETFDH mRNA included an intronic sequence downstream the first exon that was retained due to the use of a cryptic SAS located in the first intron of the gene (Fig. 3a). The introduced sequence resulted in a shift of the reading frame and premature termination of the encoded protein (Fig. 3a).

Targeted RNA analysis was conducted using maternal PB-derived cDNA to confirm this finding, documenting heterozygosity for the wild-type isoform (expected amplicon size of 206 bp) and an aberrant amplicon (amplicon size of 424 nucleotides) (Fig. 3b), the latter resulting from the insertion of the 218 nucleotide-long retained intronic region between the canonical exons 1 and 2 (Fig. 3b and 3c).

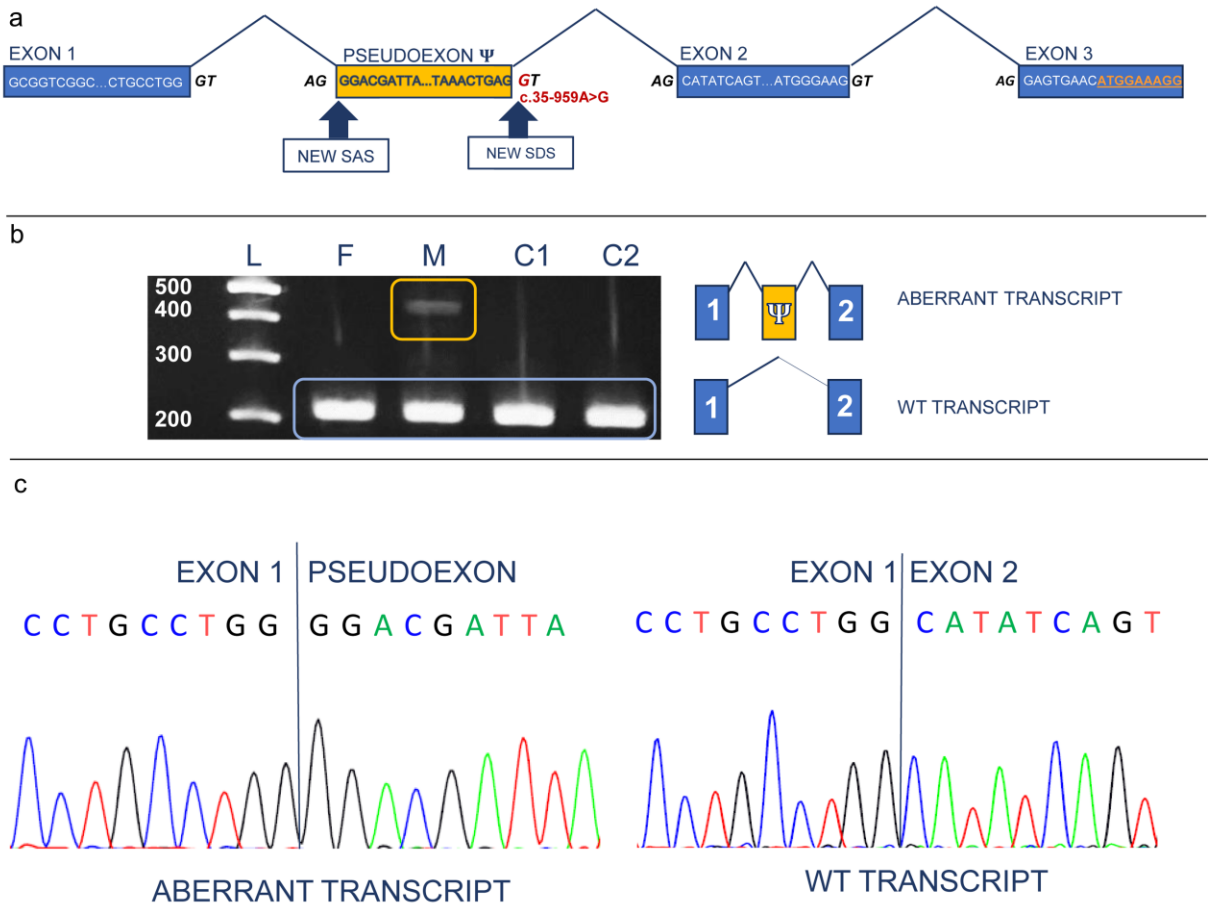

Figure 3. Confirmatory Targeted RNA analysis on maternal PB-derived cDNA. a) The blue rectangles represent the first three exons of the ETFDH gene present in both the normal and the aberrant transcript. In yellow the pseudo-exon included in the abnormal mRNA. For each exon and for the pseudo-exon the sequence of first last nucleotides is indicated. The splicing donor and acceptor sites are also depicted. The additional SAS and SDS created by the intronic mutation (in red) are indicated by the arrows. The aberrant isoform has a start codon within exon 3, with the novel coding sequence in orange, underlined. b) RT-PCR experiment on ETFDH cDNA. On the left, the gel electrophoresis image is shown. In the first lane is present the ladder (L) with 200, 300, 400, and 500 bp long DNA fragments. The second lane (F) and the third (M) are the PCR products from father and mother cDNA, respectively. The fourth and fifth lane (C1, C2) indicates the negative

control PCR products. WT isoforms are boxed in light blue while the novel aberrant transcript, present only in rectangle. In the M line an additional isoform was disclosed (the M lane is boxed in yellow. On the right, the schematic representations of both the normal and the aberrant isoforms with exon 1 and 2 in blue, and the pseudo-exon (**Ψ**) in orange. c) Sanger sequencing electropherograms of reference (right panel) and abnormal PCR products (left panel) showing the splicing junction between exon1 and exon2 and between exon1 and the pseudo-exon, respectively.

4. Discussion

Despite the remarkable progress in second-generation sequencing techniques, a significant number of patients still lack a definitive diagnosis, which inevitably impacts their access to clinical care and proper genetic counseling [6]. Here, by coupling WGS to RNAseq analysis, we reached a definitive molecular diagnosis for a rare disease compatible with the clinical symptoms occurring in the affected newborn, for which CES had allowed only the identification of a single pathogenic variant associated with the suspected condition.

In fact, methods based on whole exome sequencing or next generation gene panels do not allow to identify variants in deep intronic regions. Thus, it is crucial to employ combined approaches that include patients' RNA sequencing analysis to identify the plethora of aberrations affecting the splicing process. Mutations within intronic regions can disrupt this finely tuned mechanism, resulting in the inclusion of pseudo-exons, segments of intronic DNA erroneously recognized as exons, into mRNA by the splicing machinery. To date, numerous cases of deep intronic mutations have been identified across various genetic diseases, that with distinct pathogenetic mechanisms ultimately lead to altered gene expression [23]. Besides the inclusion of pseudo-exons, ablation of transcriptional regulatory motifs, genomic rearrangements, activation of cryptic splice sites as well as inactivation of intron-encoded RNA genes represent common events. The inclusion of pseudo-exons is usually due to mutations that create novel splice donor or acceptor sites or activate existing cryptic splice sites [24]. Numerous studies have reported genomic variants that lead to the inclusion of pseudo-exons, including those identified in the ETFDH gene [6,25] . Given the number of splicing-altering variations reported in this region, the identification of a third unique deep intronic mutation resulting in a pseudo-exon inclusion in the ETFDH gene in a patient with early-onset MADD emphasizes the natural vulnerability of intron 1 in the ETFDH gene. Generally, the inclusion of pseudo-exons in the mature mRNA transcript can lead to frameshifts, premature stop codons, or to the insertion of noncanonical amino acid sequences, generally generating nonfunctional proteins or proteins with a deleterious function. The deep intronic heterozygous variants c.35-959A>G described in this work is the third mutation reported in intron 1 of the ETFDH gene that leads to a pseudo-exon activation [6]. We demonstrate that these variants create novel exonic splicing enhancers (ESEs), leading to pseudo-exon inclusion through the formation of a new splicing donor site that successfully competes with the natural donor site, altering normal splicing. Creation of a novel SDS is the commonest mechanism of pseudo-exon activation in the human genome [24]. Indeed, the three variants determined an increase of the SD score from approximately 0 to well above the score of the natural SD site computed by the MaxEntScan (MES) algorithm. Further, both the c.35-959A>G herein described and the c.35-1008T>G activate a pseudo-exon whose size is in the range of those reviewed in Vaz-Drago et al. [23], while variant c.35-768A>G generates a 410 bp pseudo-exon that would represent one of the largest reported so far. The identification of three deep intronic variants in a stretch of 240 nucleotides is rather unusual. The ETFDH intron 1 pseudo-exon, even in the wild type sequence, possesses a strong acceptor site at its 5' end, as evidenced by its MES score that is higher than the normal acceptor in ETFDH exon 2. Thus, it is possible that any nucleotide variation, downstream of the pseudo-exon acceptor creating a strong novel donor site, could promote the rescue of a pseudo-exon. The identification of a mutation that promotes pseudo-exon inclusion highlights the necessity of extending genetic testing beyond conventional exonic mutations particularly when the

coding sequence screening had been negative. Also, our findings underscore the potential for intronic and splicing-related mutations to contribute to the phenotypic variability observed in MADD. Standard DNA sequencing methods may overlook deep intronic mutations and their consequent impact on splicing. Therefore, integrating RNA analysis into genome wide testing can provide valuable insights into the functional consequences of intronic mutations, revealing aberrant splicing events that contribute to disease pathology. This approach not only facilitates a more accurate genetic diagnosis, effective genetic counseling and patient management but also opens avenues for therapeutic interventions targeting splicing mechanisms, that can modulate splicing patterns and restore normal gene function.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Table S1: WGS metrics and sequencing output; Table S2: Forward/reverse specific primer pairs used for confirmatory analysis on ETFDH cDNA (ETFDH F1 AND ETFDH R3) and on genomic DNA.

Author Contributions: All authors contributed to the study conception and design. The project was coordinated and supervised by NR . Material preparation, data collection and analysis were performed by NR, AS, SM, AT, FCR, AB, VC, CM, PD, FM, RB, OP, MP and AG. Clinical data were collected by RF, DC, NL, SS, DZ, CF, OT and MS .The first draft of the manuscript was written by SM, AT, PD, AS, MT and GP; and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding: This work was supported by National Center for Gene Therapy and Drugs Based on RNA Technology- MUR (Project no. CN_00000041 to G.P) and by Ministry of Health project "Genoma mEdiciNa pERsonalizzata" (GENERA), T3-AN-04 (to G.P. and N.R.). This work was also supported by ELIXIRNextGenIT (Grant Code IR0000010 to G.P.).

Informed Consent Statement: All patients or their relatives signed informed consent for diagnostic and research analysis and specimen inclusion in a biobank**.**

Institutional Review Board Statement: All procedures in this study was conducted in accordance with the ethical standards of the 1964 Helsinki Declaration. The consent forms signed by our patients have been approved by the clinical risk management unit of the Policlinico of Bari.

Data Availability Statement: Data supporting the findings of this study are available within the main text of the article and in the supplementary files. Additional data are available from the corresponding author on reasonable request.

Acknowledgments: The authors acknowledge Dr Apollonia Tullo (ELIXIR-IT, the Italian Node of the European research infrastructure for life-science data, CUP B53C22000690005) who contributed to the final manuscript. The authors thank the family of the proband for the participation in this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Curcoy A, Olsen RKJ, Ribes A, Trenchs V, Vilaseca MA, Campistol J, et al. Late-onset form of beta-electron transfer flavoprotein deficiency. Mol Genet Metab. 2003;78(4):247–9.
2. van Rijt WJ, Ferdinandusse S, Giannopoulos P, Ruiter JPN, de Boer L, Bosch AM, et al. Prediction of disease severity in multiple acyl-CoA dehydrogenase deficiency: A retrospective and laboratory cohort study. J Inherit Metab Dis. 2019 May;42(5):878–89.
3. Prasun P. Multiple Acyl-CoA Dehydrogenase Deficiency. 1993.
4. Angle B, Burton BK. Risk of sudden death and acute life-threatening events in patients with glutaric acidemia type II. Mol Genet Metab. 2008;93(1):36–9.
5. Navarrete R, Leal F, Vega AI, Morais-López A, Garcia-Silva MT, Martín-Hernández E, et al. Value of genetic analysis for confirming inborn errors of metabolism detected through the Spanish neonatal screening program. Eur J Hum Genet. 2019;27(4):556–62.
6. Nogueira C, Silva L, Marcão A, Sousa C, Fonseca H, Rocha H, et al. Role of RNA in Molecular Diagnosis of MADD Patients. Biomedicines. 2021 May;9(5).
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 May;25(14):1754–60.

8. der Auwera GA Van, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra, 1st ed. O'Reilly Media, Inc.; 2020.
9. Gargano MA, Matentzoglu N, Coleman B, Addo-Lartey EB, Anagnostopoulos A V, Anderton J, et al. The Human Phenotype Ontology in 2024: phenotypes around the world. Nucleic Acids Res. 2024 May;52(D1):D1333–46.
10. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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. Genet Med. 2015;17(5):405–24.
11. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. Hum Mutat. 2016;37(6):564–9.
12. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2020;22(2):245–57.
13. Rentzsch P, Schubach M, Shendure J, Kircher M. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. Genome Med. 2021 May;13(1):31.
14. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016;48(12):1581–6.
15. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet. 2017 May;100(2):267–80.
16. Smedley D, Schubach M, Jacobsen JOB, Köhler S, Zemojtel T, Spielmann M, et al. A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease. Am J Hum Genet. 2016 Sep;99(3):595–606.
17. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333–9.
18. Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics. 2018 May;34(20):3572–4.
19. S. A. A quality control tool for high throughput sequence data. 2010.
20. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. In: Bioinformatics. Oxford University Press; 2018. p. i884–90.
21. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2–3):377–94.
22. Sakaguchi N, Suyama M. Pervasive occurrence of splice-site-creating mutations and their possible involvement in genetic disorders. NPJ genomic Med. 2022 May;7(1):22.
23. Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. Hum Genet. 2017;136(9):1093–111.
24. Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. J Med Genet. 2005 May;42(10):737–48.
25. Petersen USS, Doktor TK, Andresen BS. Pseudoexon activation in disease by non-splice site deep intronic sequence variation - wild type pseudoexons constitute high-risk sites in the human genome. Hum Mutat. 2022 May;43(2):103–27.