



# RNA Analysis and Clinical Characterization of a Novel Splice Variant in the *NSD1* Gene Causing Familial Sotos Syndrome

Olatz Villate<sup>1</sup>, Hiart Maortua<sup>2,3</sup>, María-Isabel Tejada<sup>3,4,5</sup> and Isabel Llano-Rivas<sup>3\*</sup>

## OPEN ACCESS

### Edited by:

Emiliano González Vioque,  
University Clinical Hospital of  
Santiago, Spain

### Reviewed by:

Gerald Bradley Schaefer,  
University of Arkansas for Medical  
Sciences, United States  
Laurel Willig,  
Children's Mercy Hospital,  
United States  
Kenneth Myers,  
McGill University, Canada

### \*Correspondence:

Isabel Llano-Rivas  
isabel.llanorivas@osakidetza.eus

### Specialty section:

This article was submitted to  
Genetics of Common and Rare  
Diseases,  
a section of the journal  
Frontiers in Pediatrics

**Received:** 02 December 2021

**Accepted:** 10 January 2022

**Published:** 02 February 2022

### Citation:

Villate O, Maortua H, Tejada MI and  
Llano-Rivas I (2022) RNA Analysis and  
Clinical Characterization of a Novel  
Splice Variant in the *NSD1* Gene  
Causing Familial Sotos Syndrome.  
*Front. Pediatr.* 10:827802.  
doi: 10.3389/fped.2022.827802

<sup>1</sup> Pediatric Oncology Group, Biocruces Bizkaia Health Research Institute, Barakaldo, Spain, <sup>2</sup> Neurodegenerative Diseases Group, Biocruces Bizkaia Health Research Institute, Barakaldo, Spain, <sup>3</sup> Genetics Service, Hospital Universitario Cruces-Osakidetza, Barakaldo, Spain, <sup>4</sup> Biocruces Bizkaia Health Research Institute, Barakaldo, Spain, <sup>5</sup> Spanish Consortium for Research on Rare Diseases (CIBERER), Madrid, Spain

**Background:** Sotos syndrome is an autosomal dominant disorder characterized by overgrowth, macrocephaly, distinctive facial features and learning disabilities. Haploinsufficiency of the nuclear receptor SET domain-containing protein 1 (*NSD1*) gene located on chromosome 5q35 is the major cause of the syndrome. This syndrome shares characteristics with other overgrowth syndromes, which can complicate the differential diagnosis.

**Methods:** Genomic DNA was extracted from peripheral blood samples of members of the same family and targeted exome analysis was performed. *In silico* study of the variant found by next-generation sequencing was used to predict disruption/creation of splice sites and the identification of potential cryptic splice sites. RNA was extracted from peripheral blood samples of patients and functional analyses were performed to confirm the pathogenicity.

**Results:** We found a novel c.6463 + 5G>A heterozygous *NSD1* gene pathogenic variant in a son and his father. Molecular analyses revealed that part of the intron 22 of *NSD1* is retained due to the destruction of the splicing donor site, causing the appearance of a premature stop codon in the *NSD1* protein.

**Conclusions:** Our findings underline the importance of performing RNA functional assays in order to determine the clinical significance of intronic variants, and contribute to the genetic counseling and clinical management of patients and their relatives. Our work also highlights the relevance of using *in silico* prediction tools to detect a potential alteration in the splicing process.

**Keywords:** next-generation sequencing, *NSD1*, Sotos syndrome, molecular analysis, RNA, splicing mutation

## INTRODUCTION

Sotos syndrome (OMIM #117550) is a rare autosomal dominant disorder characterized by overgrowth (increased height, macrosomia, and macrocephaly), characteristic facial features and learning and intellectual disabilities (1). Other associated clinical features include scoliosis, seizures, renal and cardiac anomalies (2, 3).

Sotos Syndrome is caused by heterozygous pathogenic variants in the nuclear receptor SET domain-containing protein 1 (*NSD1*) gene located on 5q35. *NSD1* encodes a histone methyltransferase that catalyzes the transfer of methyl groups to lysine residues of histone tails (4). The overall prevalence of this syndrome is estimated at 1 in 14,000 (5). More than 95% of the cases arise from *de novo* mutations and familial cases of Sotos syndrome account for about 5% of the diagnoses (6–8).

This syndrome shares characteristics with other overgrowth syndromes, which can complicate the differential diagnosis. The greatest phenotypic overlap is between Sotos and Weaver syndromes (5, 6). The clinical features of post-natal overgrowth, advanced bone age, and intellectual disability particularly in the context of mild facial dysmorphism can make it difficult to distinguish between these syndromes (6).

Targeted exome sequencing has become a powerful and useful method to diagnose patients with suspected overgrowth disorders (9). After performing next generation sequencing (NGS) studies, different novel variants of uncertain significance (VUS) can be detected in the studied patients so it is crucial to characterize their biological impact and determine their pathogenicity. In the case of variants that have been shown by bioinformatics programs to alter the splicing process, functional approaches such as RT-PCR analysis of patient-derived RNA or minigene splicing assays can be used to assess the effect of these variants on mRNA splicing (10, 11). The relevance of reclassification of splice variants classified as VUS lies in the fact that the patient needs to follow precise medical management, in accordance with the international guidelines (12, 13).

In this study, we report one novel intronic *NSD1* variant in two members of the same family. Clinical, familial and molecular data, together with our experimental RNA functional assays, have provided a pathogenicity characterization of the *NSD1* variant.

## MATERIALS AND METHODS

### Patients

A Spanish boy and his father were recruited. This study was conducted in full accordance with the World Medical Association Declaration of Helsinki (version 2008) and the additional requirements. Informed consent approved by the clinical ethical committee of Cruces University Hospital (Spain) was obtained prior to genetic testing.

### Next Generation Sequencing Studies and Sanger Validation

Genomic DNA was extracted from peripheral blood and targeted exome analysis was performed in the Institute of Genomic Medicine (IMEGEN, Valencia, Spain). The analysis of the genes

associated with overgrowth syndromes included: *AKT1*, *APC2*, *CDC45*, *CDC6*, *CDKN1C*, *CDT1*, *DICER1*, *DIS3L2*, *DNMT3A*, *EED*, *EZH2*, *FBN1*, *GMNN*, *GPC3*, *GPC4*, *H19*, *HERC1*, *IGF2*, *MCM5*, *MYO5A*, *NFIX*, *NSD1*, *OFD1*, *ORC1*, *ORC4*, *ORC6*, *PIK3CA*, *PTEN*, *RNF135*, *SETD2* and *SUZ12*. For this purpose, the exonic regions of interest were captured using the kit xGen Exome Panel v1.0 (IDT) and pair-end libraries were generated. Ultrasequencing was performed on the NextSeq 500 sequencing system platform (Illumina). Bioinformatics analysis of the data was done using Data Genomics Exon Pipeline software (version v1). Subsequently, the results were interpreted and prioritized. Changes with a number of readings >20 and with a frequency greater than 30% were considered variants. The variants detected were analyzed in four different databases (Exome Aggregation Consortium, 1000 Genomes, Human Gene Mutation Database and Exome Variant Server) and *in silico* prediction programs (SIFT, Polyphen-2, and MutationTaster-2). Conventional Sanger sequencing in genomic DNA validated candidate variant. Primers used were NSD1gF 5'-ACCCATTGCCACGGAAGAAA-3' and NSD1gR 5'-CAC CGCTGTCCCATCTTCA-3'.

### Splicing *in silico* Analyses

*In silico* study of the variant found by NGS was used to predict disruption/creation of splice sites and the identification of potential cryptic splice sites. The following online programs were used: NNsplice ([http://www.fruitfly.org/seq\\_tools/splice.html](http://www.fruitfly.org/seq_tools/splice.html)) and Human Splicing Finder (<http://www.umd.be/HSF/>).

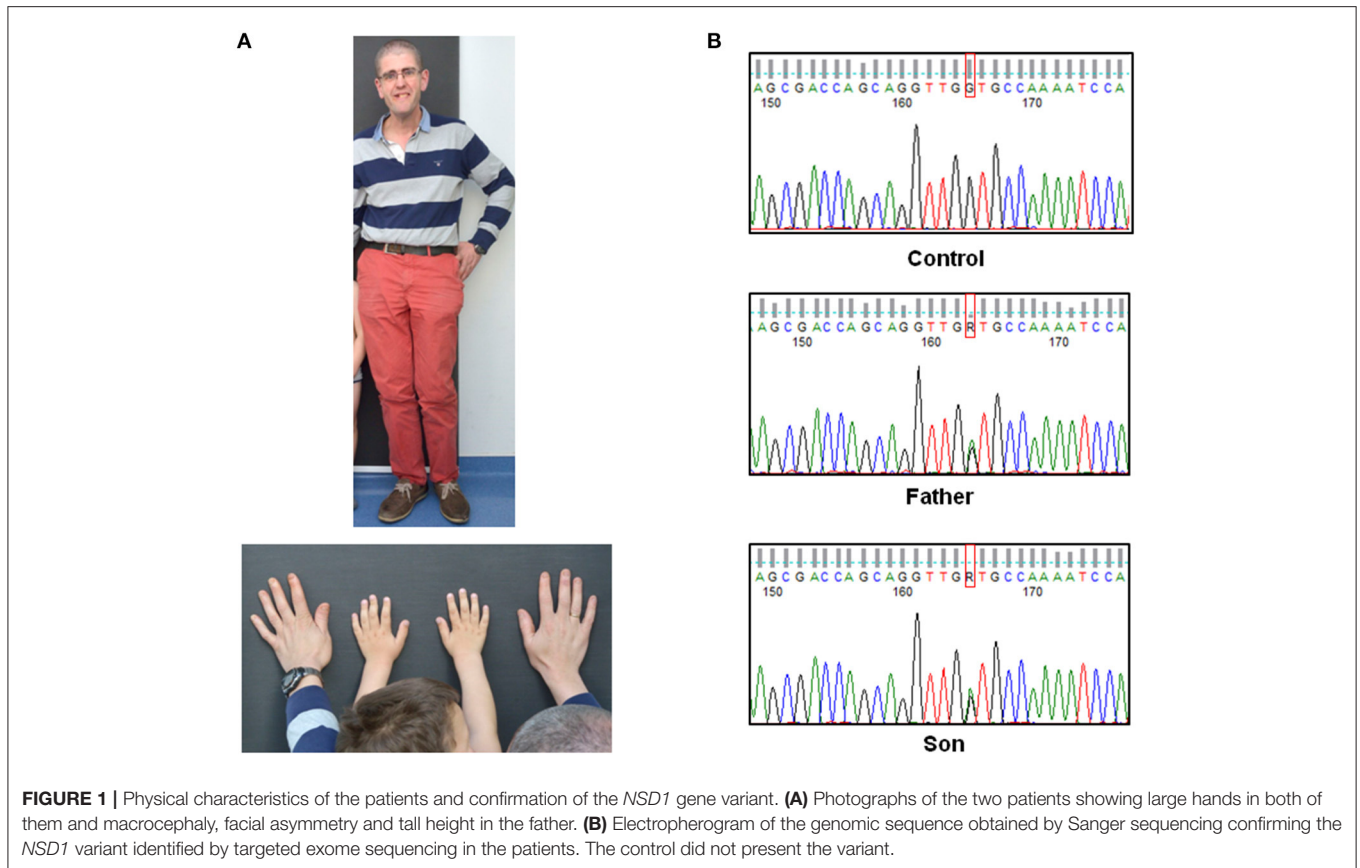
### RT-PCR and cDNA Analysis

cDNA was obtained using Superscript RT II enzyme (Invitrogen, Carlsbad, CA, USA) from 500 ng of total RNA extracted from peripheral blood in a volume of 20  $\mu$ l. cDNA was amplified and sequenced to identify potential splicing variants. Primers used for cDNA amplification and identification of splice variants were: NSD1cF 5'-CCAAAGAATCAACCCATTGC-3' and NSD1cR 5'-GATTGCTCTGCCAGGTGAGT-3'. Primers for new splice variant sequencing were: NSD1svF 5'-AGCAGGTTGATGCCA AAATC-3' and NSD1svR 5'-CTCTGCCAGGTGAGTGCTT-3'.

## RESULTS

### Patients

The proband was a 3-year-old male patient being the second child of a non-consanguineous Spanish couple. He presented with plagiocephaly, asymmetric facies, high palate with decreased posterior palatal movement and uvula slightly deviated to the left. Prominent chin and prominent forehead were also present. Ear pinnae were well implanted. Chest and abdomen did not show alterations and no scoliosis was observed. Size was 111.5 cm (>p99 +3.3 SDS), height 23 kg (p99, +2.3 SDS) and head circumference 54 cm (>p99, +2.5 DS). Full limbs, symmetrical, with large hands and feet and thumbs of the four extremities wide (**Figure 1A**). Smooth skin, no spots, no asymmetric overgrowths and no alterations of subcutaneous tissue were noticed. Psychomotor development considered within normality except for delayed speech, language articulation problems, with



communicative intention. No seizures. Features were shared with his 43-year-old father presenting with tall stature (184 cm, p84 +1.02 SDS), macrocephaly, facial asymmetry (**Figure 1A**) and velopalatine insufficiency. Intelligence was normal-low and he had no seizures. Paternal grandfather, deceased, seemed to present with macrocephaly and facial features similar to the father.

### Targeted Exome Analysis

Father's DNA was used for Next Generation Sequencing studies. In total 94.8% of the region of interest (coding region of the selected genes) was covered at least 20X. A heterozygous change was detected in intron 22 of the *NSD1* gene c.6463 + 5G>A (NM\_022455.4). This change was not previously described and it was considered a variant of uncertain significance. In total 100% of the region of interest of the *NSD1* gene was covered with a depth greater than 20X. The coding region of this gene was covered with an average depth of 93.4X. The presence of this variant was validated in the father and his son by Sanger Sequencing (**Figure 1B**) and was not identified in the healthy sister of the proband.

### In silico Analyses

The Human Splicing Finder program predicted an alteration of the donor site indicating a probable disruption of splicing in intron 22. The NNSplice program predicted that the score for the

donor splice site changes from a high value of 0.80 in the wild type sequence to a low value of 0.23 in the mutated sequence indicating that the donor consensus sequence is weakened. The next donor splice site in the mutated sequence in the intron 22 was predicted to be 120 bp far from the first one indicating that the variant could induce the retention of part of the intron (**Figure 2A**).

### Molecular Characterization of the New Variant

To perform the functional studies, RNA samples were obtained from the father, the son and two healthy controls. *NSD1* transcripts were first analyzed by PCR in two controls. Only one band corresponding to the size of the normal transcript was observed and it was confirmed by sequencing (**Figure 2B**). Subsequently, the same study was performed on cDNA obtained from RNA from the father and the son and the results showed two bands, the band corresponding to the normal transcript and an additional one 120 bp longer than the normal one (**Figure 2B**). The two bands obtained in each case were sequenced. Sequencing confirmed that 120 bp of the intron were retained as predicted by *in silico* analysis. The retention of 120 bp has a potential impact on the protein as a premature stop codon is produced and a PHD finger and a C5HCH domain are affected in the mutated protein (**Figure 3**). This alteration confirms the pathogenicity of the variant.

## DISCUSSION

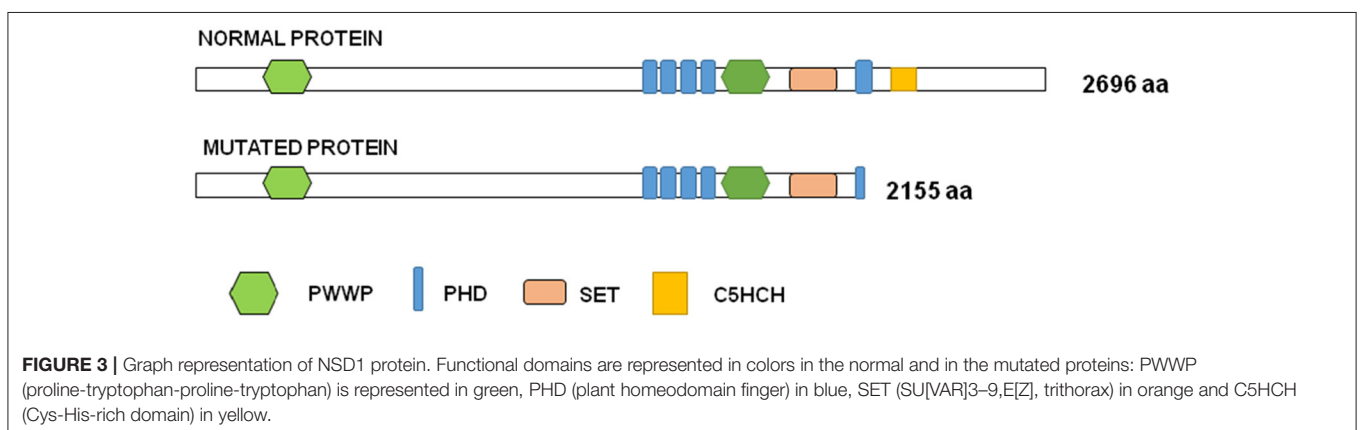
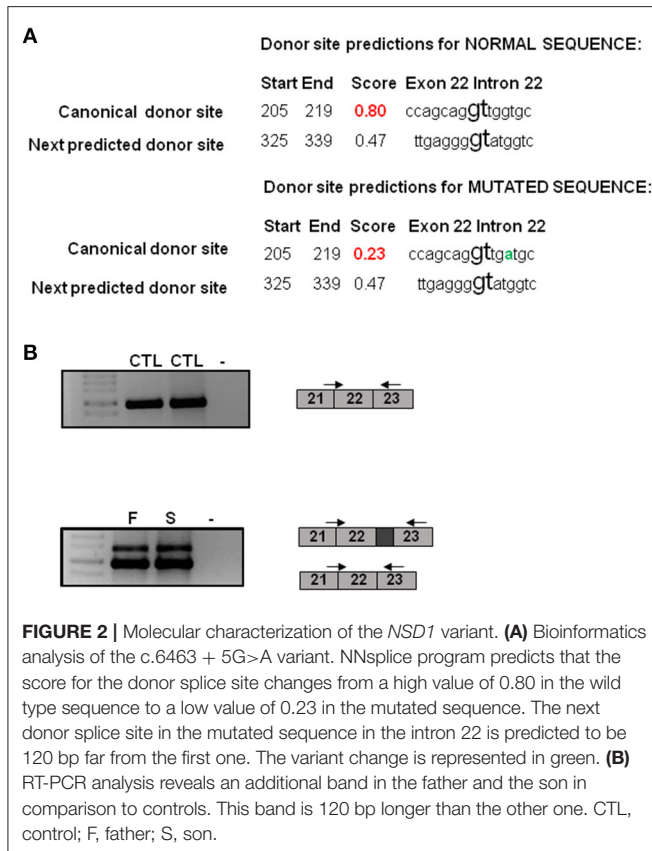
In this study, we have combined molecular, clinical and functional analysis of two patients of the same family suspected of having an overgrowth syndrome. Our main aim was to contribute to the molecular and clinical classification of the variant of uncertain significance detected in *NSD1*.

The variant found in *NSD1* by targeted exome sequencing was mapped in the position +5 of the intron 22. It is well known that the GT dinucleotide at the 5' end of an intron (positions

+ 1 and + 2) is highly conserved in human genes (> 98%) and critical for RNA splicing (14). These alterations are nearly systematically classified as pathogenic or probably pathogenic. In the position +5 of the 5' splice sites the guanine is the most frequently nucleotide found, reaching 78% conservation among over 180,000 human 5' splice sites (14) and it has been shown that disease-associated alterations are very often detected at this nucleotide (15). The variants at this position are usually classified as VUS so functional studies are required to determine their pathogenicity. Our results support the relevant role of intronic position + 5 in normal splicing and agree with previous studies showing that this position seems particularly prone to aberrant splicing when altered (10, 16). Moreover, we observed a good concordance between *in silico* programs used to predict splicing alterations and the results of the molecular assays for this variant located at the + 5 position.

To identify more disease causing variants in overgrowth syndromes there is a need to extend the studies to intronic variants. Due to the size of intronic regions, identifying deep intronic variants that affect splicing is challenging. The recent applications of whole-genome sequencing (WGS) to clinical screening studies enable the investigation of non-coding variation and identification of pathogenic deep intronic variants that lie >100 bp away from the nearest canonical splice sites (17).

In the Human Gene Mutation Database Professional 2020.4, 564 mutations have been reported in *NSD1*, 461 of them associated to Sotos Syndrome. The c.6463 + 5G>A variant is not reported in HGMD and our patients are the first ones described with this change. It is worthy to note that 24 splice variants are described as responsible for Sotos Syndrome phenotype and only one of them is mapped in position +5 of the splice site (18). This could suggest that functional studies are not being performed on variants at positions other than +1 and +2, which are necessary to confirm the diagnosis and provide correct genetic counseling. Of the 24 splice variants described in HGMD, 21 of them affect positions +1, +2, -1 and -2. The Human Splicing Finder program predicts in these variants an alteration of the donor or acceptor site depending on the position, indicating that most probably





affect the splicing process. The variant previously described at position +5, c.4378 + 5G>C, according to bioinformatics analysis causes an alteration of the wild type donor site, most probably affecting splicing. There is also a splice variant described at position -5, c.6152-5T>G, and it is predicted to alter the wild type acceptor site, most probably affecting splicing. The last variant described that is not in the canonical positions is found at position +33, c.3796 + 33A>T, and in this case, the Human Splicing Finder prediction considers no significant impact on splicing signals. The NSD family members, consisting of NSD1, NSD2 and NSD3 are methyltransferases implicated in multiple diseases (19). We have shown that the variant found in *NSD1* has a potential impact on the protein causing a shortening and affecting C5HCH and PHD finger domains due to the appearance of a premature stop codon. NSD1 binds upstream of the bone morphogenetic protein 4 promoter, enforces H3K36 methylation levels within this region, and thus promotes bone morphogenetic protein 4 transcription (20). It has been shown that the PHD5-C5HCH domains of the NSD1 protein might have chromatin targeting ability and that Sotos syndrome mutations within these domains seem to alter the normal function of NSD1 (19). This would confirm the pathogenicity of the variant present in the two patients.

## CONCLUSIONS

Our findings underline the importance of performing RNA splicing assays in order to determine the clinical significance of intronic variants, and contribute to the genetic counseling and clinical management of patients and their relatives. Our work also contributes to highlight the relevance of using *in silico* prediction tools to detect a potential alteration in the splicing process.

## REFERENCES

- Sotos JF, Dodge PR, Muirhead D, Crawford JD, Talbot NB. Cerebral gigantism in childhood: a syndrome of excessively rapid growth with acromegalic features and a nonprogressive neurologic disorder. *N Engl J Med.* (1964) 271:109–16. doi: 10.1056/NEJM196407162710301
- Foster A, Zachariou A, Loveday C, Ashraf T, Blair E, Clayton-Smith J, et al. The phenotype of Sotos syndrome in adulthood: a review of 44 individuals. *Am J Med Genet.* (2019) 181:502–8. doi: 10.1002/ajmg.c.31738
- Tatton-Brown K, Douglas J, Coleman K, Baujat G, Cole TRP, Das S, et al. Genotype-phenotype associations in sotos syndrome: an analysis of 266 individuals with NSD1 aberrations. *Am J Med Genet.* (2005) 77:193–204. doi: 10.1086/432082
- Rayasam GV. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* (2003) 22:3153–63. doi: 10.1093/emboj/cdg288
- Tatton-Brown K, Rahman N. Sotos syndrome. *Eur J Hum Genet.* (2007) 15:264–71. doi: 10.1038/sj.ejhg.5201686
- Manor J, Lalani SR. Overgrowth syndromes—evaluation, diagnosis, and management. *Front Pediatr.* (2020) 8:574857. doi: 10.3389/fped.2020.574857
- Waggoner DJ, Raca G, Welch K, Dempsey M, Anderes E, Ostrovskaya I, et al. NSD1 analysis for Sotos syndrome: Insights and perspectives from the clinical laboratory. *Genet Med.* (2005) 7:524–33. doi: 10.1097/01.GIM.0000178503.15559.d3

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Clinical Ethical Committee of Cruces University Hospital (Spain). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

OV and HM performed the bioinformatics and molecular analyses. M-IT directed the project granted and reviewed the manuscript. IL-R carried out the clinical characterization of patients. OV and IL-R conceived the study, supervised it, and wrote the manuscript. All authors have read and approved the final manuscript.

## FUNDING

This work was supported by grant CIBERER17-TRASL7-02 from CIBERER (Spanish Consortium for Research on Rare Diseases).

## ACKNOWLEDGMENTS

We would like to thank the family for their cooperation and for consenting to the publication of this manuscript.

- Laccetta G, Moscuza F, Michelucci A, Guzzetta A, Lunardi S, Lorenzoni F, et al. Novel missense mutation of the NSD1 gene associated with overgrowth in three generations of an Italian family: case report, differential diagnosis, and review of mutations of NSD1 gene in familial sotos syndrome. *Front Pediatr.* (2017) 5:236. doi: 10.3389/fped.2017.00236
- Kim Y-M, Lee Y-J, Park JH, Lee H-D, Cheon CK, Kim S-Y, et al. High diagnostic yield of clinically unidentifiable syndromic growth disorders by targeted exome sequencing. *Clin Genet.* (2017) 92:594–605. doi: 10.1111/cge.13038
- Piñero TA, Soukariéh O, Rolain M, Alvarez K, López-Köstner F, Torrezan GT, et al. MLH1 intronic variants mapping to + 5 position of splice donor sites lead to deleterious effects on RNA splicing. *Fam Cancer.* (2020) 19:323–36. doi: 10.1007/s10689-020-00182-5
- Villate O, Ibarluzea N, Fraile-Bethencourt E, Valenzuela A, Velasco EA, Grozeva D, et al. Functional analyses of a novel splice variant in the CHD7 gene, found by next generation sequencing, confirm its pathogenicity in a Spanish patient and diagnose him with CHARGE syndrome. *Front Genet.* (2018) 9:7. doi: 10.3389/fgene.2018.00007
- Agianitopoulos K, Pepe G, Papadopoulou E, Tsaousis GN, Kampouri S, Maravelaki S, et al. Clinical utility of functional RNA analysis for the reclassification of splicing gene variants in hereditary cancer. *Cancer Genomics Proteomics.* (2021) 18:285–94. doi: 10.21873/cgp.20259
- Truty R, Ouyang K, Rojahn S, Garcia S, Colavin A, Hamlington B, et al. Spectrum of splicing variants in disease genes and the ability of RNA analysis

- to reduce uncertainty in clinical interpretation. *Am J Med Genet.* (2021) 108:696–708. doi: 10.1016/j.ajhg.2021.03.006
14. Sahashi K, Masuda A, Matsuura T, Shinmi J, Zhang Z, Takeshima Y, et al. *In vitro* and *in silico* analysis reveals an efficient algorithm to predict the splicing consequences of mutations at the 5' splice sites. *Nucleic Acids Res.* (2007) 35:5995–6003. doi: 10.1093/nar/gkm647
  15. Krawczak M, Thomas NST, Hundrieser B, Mort M, Wittig M, Hampe J, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum Mutat.* (2007) 28:150–8. doi: 10.1002/humu.20400
  16. Buratti E, Chivers M, Kráľovičová J, Romano M, Baralle M, Krainer AR, et al. Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res.* (2007) 35:4250–63. doi: 10.1093/nar/gkm402
  17. Qian X, Wang J, Wang M, Igelman AD, Jones KD, Li Y, Wang K, et al. Identification of deep-intronic splice mutations in a large cohort of patients with inherited retinal diseases. *Front Genet.* (2021) 12:647400. doi: 10.3389/fgene.2021.647400
  18. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, 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. *Hum Genet.* (2017) 136:665–77. doi: 10.1007/s00439-017-1779-6
  19. He C, Li F, Zhang J, Wu J, Shi Y. The methyltransferase NSD3 has chromatin-binding motifs, PHD5-C5HCH, that are distinct from other NSD (Nuclear Receptor SET Domain) family members in their histone H3 recognition. *J Biol Chem.* (2013) 288:4692–703. doi: 10.1074/jbc.M112.426148
  20. Lucio-Eterovic AK, Singh MM, Gardner JE, Veerappan CS, Rice JC, Carpenter PB. Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc Natl Acad Sci U S A.* (2010) 107:16952–7. doi: 10.1073/pnas.1002653107
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Villate, Maortua, Tejada and Llano-Rivas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.