Identification of a deep intronic POLR3A variant causing inclusion of a pseudo-exon derived from an Alu element in Pol III-related leukodystrophy

Brief Communications

Identification of a deep intronic *POLR3A* variant causing inclusion of a pseudo-exon derived from an *Alu* element in Pol III-related leukodystrophy

Takuya Hiraide[1], Mitsuko Nakashima[1], Takahiro Ikeda[2], Daisuke Tanaka[2], Hitoshi Osaka[2] and Hirotomo Saitsu[1]

[1]Department of Biochemistry, Hamamatsu University School of Medicine; [2]Department of Pediatrics, Jichi Medical School

Correspondence: Hirotomo Saitsu, MD, PhD.
Department of Biochemistry, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi-ku, Hamamatsu 431-3192, Japan
Tel/Fax: +81-53-435-2327
E-mail: hsaitsu@hama-med.ac.jp

Running title: A deep intronic variant in *POLR3A*

The authors declare that they have no conflicts of interest.

Abstract

Pseudo-exon inclusion caused by deep intronic variants is an important genetic cause for various disorders. Here, we present a case of a hypomyelinating leukodystrophy with developmental delay, intellectual disability, autism spectrum disorder and hypodontia, which are consistent with autosomal recessive POLR3-related leukodystrophy. Whole-exome sequencing identified only a heterozygous missense variant (c.1451G>A) in *POLR3A*. To explore possible involvement of a deep intronic variant in another allele, we performed whole-genome sequencing of the patient with variant annotation by SpliceAI, a deep learning-based splicing prediction tool. A deep intronic variant (c.645+312C>T) in *POLR3A*, which was predicted to cause inclusion of a pseudo-exon derived from an *Alu* element, was identified and confirmed by mRNA analysis. These results clearly showed that whole-genome sequencing, in combination with deep-learning based annotation tools such as SpliceAI, will bring us further benefits in detecting and evaluating possible pathogenic variants in deep intronic regions.

Keywords: Deep intronic variants, Pseudo-exon inclusion, *POLR3A*, whole-genome sequencing, SpliceAI

Introduction

Whole-genome sequencing (WGS) has become widely used to detect pathogenic variants in clinically heterogeneous diseases because the costs of WGS have decreased [1]. A recent review described that deep intronic variants more than 100 base pairs from exon–intron boundaries may be causative for multiple diseases [2]. These deep intronic variants could activate cryptic (non-canonical) splice sites or alter splicing enhancer or silencer elements in introns; however, it is difficult to accurately predict the effects of these variants on RNA splicing [3]. Recently, bioinformatic splicing prediction tools that use a deep learning network, such as SpliceAI, have allowed accurate prediction of abnormal splicing caused by noncoding variants [4].

POLR3-related leukodystrophy is a rare autosomal recessive disease characterized by hypomyelination, hypodontia, and hypogonadotropic hypogonadism, and caused by biallelic variants in *POLR3A*- and *POLR3B*-encoding RNA polymerase III subunits [5]. Here, we report a first case of POLR3-related leukodystrophy caused by a combination of missense and deep intronic variants in *POLR3A*.

Subject and Methods

Case report

A Japanese boy was born without asphyxia after 36 weeks of gestation. There was no family history of neurodevelopmental disorders. His birth weight, body length, and head circumference were 3130 g (2.0 standard deviation [SD]), 47.6 cm (0.46 SD), and 33.0 cm (0.47 SD), respectively. Laryngomalacia was recognized at two weeks of age. He has had recurrent respiratory infections since he was 1 year of age. He was able to walk

independently at 1 year and 3 months of age, but could not go up and down the stairs until he was 3 years of age. Although he understood simple instructions by the time that he was 3 years of age, he still had no meaningful speech. At 3 years of age, he was diagnosed with autism spectrum disorder, intellectual disability and failure to thrive. On examination at 10 years of age, his height, weight, and head circumference were 128.0 cm (−1.9 SD), 19.8 kg (−2.0 SD), and 50.2 cm (−2.1 SD), respectively. The each volume of both testes was 1–2 ml suggesting that he was prepubertal. Mild hypotonia and normal deep tendon reflexes were observed. No cerebellar features or extrapyramidal features were recognized. The gonadotropin releasing hormone stimulation test and the GH secretion stimulation test were normal. Orthopantomography revealed all second premolar defects. Brain magnetic resonance imaging (MRI) at 11 years of age showed diffuse hypomyelination, cerebellar atrophy and thinning of the corpus callosum (Figure 1).

Variant screening

This study was approved by the Institutional Review Board Committee at Jichi Medical School and Hamamatsu University School of Medicine. After receiving written informed consent, we performed case-only WES (whole-exome sequencing) and WGS. The patient's DNA was captured using an xGen Exome Research Panel kit (IDT, Coralville IA), and sequenced on a NextSeq500 (Illumina, San Diego, CA) with 75-bp paired-end reads. WGS was commissioned to Macrogen Japan Corp. (Kyoto, Japan). Data processing, variant calling, annotation, and filtering were performed as described in the Supplementary methods.

Reverse transcription polymerase chain reaction

Peripheral blood mononuclear cells (PBMCs) from the venous blood of the patient and one control were isolated by Lymphoprep (Axis-Shield, Oslo, Norway). Total RNA was extracted from the PBMC using a RNeasy Mini Kit (QIAGEN, Hilden, Germany), and subjected to reverse transcription using the PrimeScript RT reagent kit (TAKARA BIO, Kusatsu, Shiga, Japan) according to the manufacturer's protocol. We designed target specific primers for *POLR3A* on exon 5 and 6 (Figure 2B) to confirm aberrant splicing, and on exon 5 and 11 to estimate efficiency of pseudo-exon inclusion caused by the intronic variant (Supplemental Figure S1). TA cloning of PCR products was performed as described previously [6].

Results

First, we searched for the candidate variants by WES and found one candidate variant c.1451G>A, p.(Arg484Gln) in *POLR3A* (NM_007055.3) (Figure 2A). This variant was found in one of the 251,270 alleles in gnomAD v2.1.1 (http://gnomad.broadinstitute.org/; accessed April 2020) but was absent in ToMMo 4.7KJPN Allele Frequency Panel (v20190826) [7]. *In silico* pathogenicity prediction tools suggested this variant to be deleterious (Supplemental Table S1). However, we could find neither the second candidate variant in the coding regions or adjacent introns nor the copy number variant in *POLR3A* by WES. Given the phenotypic similarity to POLR3-related leukodystrophy, we performed WGS to explore the involvement of non-coding variants and identified a deep intronic variant (c.645+312C>T), which was predicted to cause exonization of *Alu* element in intron 5 by SpliceAI (Figure 2A, B, and Table 1). Sanger sequencing confirmed that the c.1451G>A variant and the

c.645+312C>T variant were transmitted from his father and mother, respectively (Figure 2C). In RT-PCR, three different sized products were amplified in the patient (Figure 2D). The longest product was digested with T7 endonuclease I, indicating heteroduplex formation (Figure 2D) [8]. The mutant transcript had 129 nucleotide insertions between exons 5 and 6, which generated a premature stop codon at the 254th residue (p.Leu216Met*39) (Figure 2E). These findings suggested that these variants were likely to be pathogenic (Supplemental Table S2). To investigate what percentage of the transcripts from the maternal allele undergo pseudo-exon inclusion, we performed TA cloning of long RT-PCR products spanning the c.1451G>A variant site to distinguish paternal and maternal alleles. Three of 27 clones without pseudo-exon inclusion were derived from the maternal allele (Supplemental Figure S1). If we assume that the paternal and maternal alleles have the same expression level, it could be estimated that 87.5% (21 of 24) of transcripts from the maternal allele have splicing abnormalities. However, we also found that one of 7 clones with pseudo-exon inclusion was derived from the paternal allele (14.2%). If we assume that this aberrant transcript of the paternal allele is artificially generated by PCR-mediated recombination [9], same recombination could occur during amplification of normal transcript.

Discussion

Recent reports estimate that variants affecting splicing comprise about 10% of the pathogenic variants in rare genetic disorders [4]. Recently, deep learning as a class of machine learning methods has been widely used for genome analysis [10]. SpliceAI, a deep residual neural network, predicts splice junctions [4], is more accurate than other tools in predicting splicing abnormalities [11]. Therefore, detection of cryptic

(non-canonical) splicing events in the deep intronic region could improve by adopting SpliceAI in WGS data analysis. In fact, SpliceAI predicted that the c.645+312C>T variant creates a new splice donor site leading to pseudo-exon inclusion by activating a pre-existing cryptic acceptor splice site. The cryptic acceptor splice site is predicted by other tools (NetGene2, ESEfinder, and Human Splicing Finder) in the wild-type sequence, but their prediction scores are not increased by the variant in two programs (Table 1). We also explored how many variants can be scored 0.2 or more. SpliceAI annotates variants within genes defined by an annotation file (grch38.txt) and does not annotate variants located in intergenic region. When we analyzed 141,609 variants within genes on chromosome 1, SpliceAI predicted 19 variants as high precision (Supplemental Table S3). Notably, intronic variants except for canonical splice sites account for 11 variants, suggesting potentials for detecting possible pathogenic intronic variants.

Pseudo-exons often derive from transposable elements like *Alu* [12]. Exonization of a silent intronic *Alu* tends to occur predominantly from the right arm on their antisense orientation relative to the sense orientation [13, 14]. Indeed, the c.645+312C>T variant was located at the most frequently selected 5′ splice site in exonization of the inverted *Alu* sequence [14] (Figure 2B). Our data further support that deep intronic variants in the *Alu* sequence are important candidates for genetic disorders.

In conclusion, we identified compound heterozygous *POLR3A* variants in exonic and deep intronic regions using WES and WGS. WGS is an important diagnostic tool in patients with only one variant found in autosomal recessive disease.

Acknowledgments

We would like to thank the patient's family for participating in this work. This work was supported by Grant-in-Aid from the Ministry of Health, Labour and Welfare of Japan; the Takeda Science Foundation, and a HUSM Grant-in-Aid from Hamamatsu University School of Medicine.

Conflict of Interest

The authors declare that they have no conflicts of interest.

References

1. Schwarze K, Buchanan J, Taylor JC, Wordsworth S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genet Med. 2018;20:1122−1130.
2. Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. Hum Genet. 2017;136:1093−1111.
3. Jian X, Boerwinkle E, Liu X. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. Genet Med. 2014;16:497–503.
4. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019;176:535−548.e24.
5. Bernard G, Vanderver A. POLR3-Related Leukodystrophy. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993−2019. 2012 Aug 2 [updated 2017 May 11]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK99167/
6. Miyamoto S, Nakashima M, Ohashi T, Hiraide T, Kurosawa K, Yamamoto T, et al. A case of de novo splice site variant in *SLC35A2* showing developmental delays, spastic paraplegia, and delayed myelination. Mol Genet Genomic Med. 2019;7:e814.

7. Tadaka S, Katsuoka F, Ueki M, Kojima K, Makino S, Saito S, et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. Hum Genome Var. 2019;6:28.
8. Lowell JL, Klein DA. Heteroduplex resolution using T7 endonuclease I in microbial community analyses. Biotechniques. 2000;28:676–8, 680, 681.
9. Potapov V, Ong JL. Examining Sources of Error in PCR by Single-Molecule Sequencing. PLoS One. 2017;12:e0169774.
10. Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. A primer on deep learning in genomics. Nat Genet. 2019;51:12−18.
11. Wai HA, Lord J, Lyon M, Gunning A, Kelly H, Cibin P, et al. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. Genet Med. 2020. https://doi.org/10.1038/s41436-020-0766-9.
12. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet. 2010;11:345–55.
13. Lev-Maor G, Sorek R, Shomron N, Ast G. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. Science. 2003;300:1288–91.
14. Ram O, Schwartz S, Ast G. Multifactorial interplay controls the splicing profile of Alu-derived exons. Mol Cell Biol. 2008;28:3513–25.

Figure legends

Figure 1. Brain MRI findings at 11 years of age.

(A–C) Axial T2-weighted images show a diffusely elevated signal in the white matter. Relative T2 hypointensity of the optic radiations (white arrow) and the ventrolateral thalamus (white arrowheads) are observed (B). The cerebellar white matter signal is mild T2 hyperintense; the dentate nucleus appears T2 hypointense (black arrowhead) (C). (D) Sagittal T2-weighted image shows atrophy of cerebellar vermis and hypoplastic corpus callosum.

Figure 2. *POLR3A* variants in the patient and its effects on splicing.

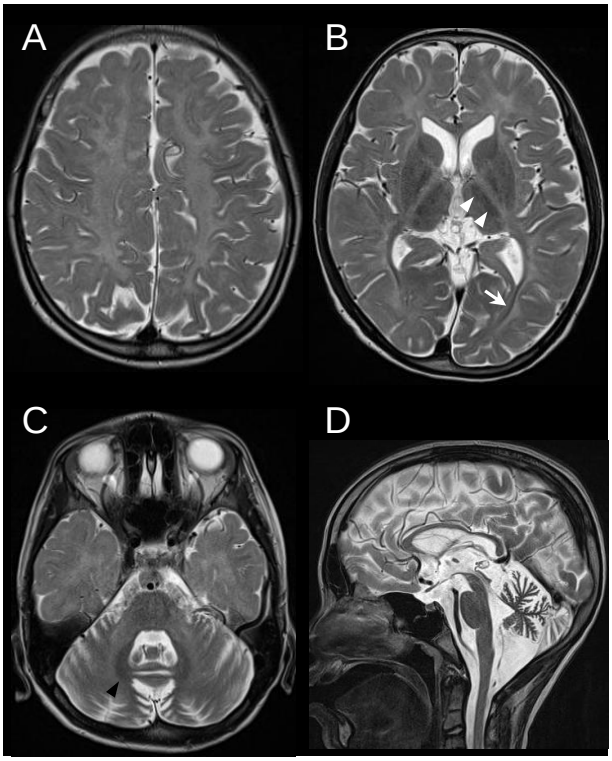
(A) Schematic presentation of the *POLR3A* gene (upper) and RPC1 protein structure encoded by *POLR3A* (lower). The *POLR3A* variants identified in our case are shown. The c.1451G>A variant is highly evolutionarily conserved. Multiple amino acid sequences of RPC1 were aligned using the ClustalW tool (http://www.genome.jp/tools/clustalw). (B) SpliceAI predicted that the c.645+312C>T variant created a novel donor splice site and activated an upstream acceptor splice site; this led to inclusion of a 129-nucleotide inverted *Alu* sequence in intron 5. RT-PCR was performed using target-specific primers designed at exons 5 and 6 (arrows). (C) Sanger sequencing of the *POLR3A* (NM_007055.3) show compound heterozygous variants. The c.1451G>A variant is inherited from the boy's father, and the c.645+312C>T variant is from his mother. (D) cDNA amplicons of the patient and a healthy control showed three different-sized PCR products in the patient sample. The upper product

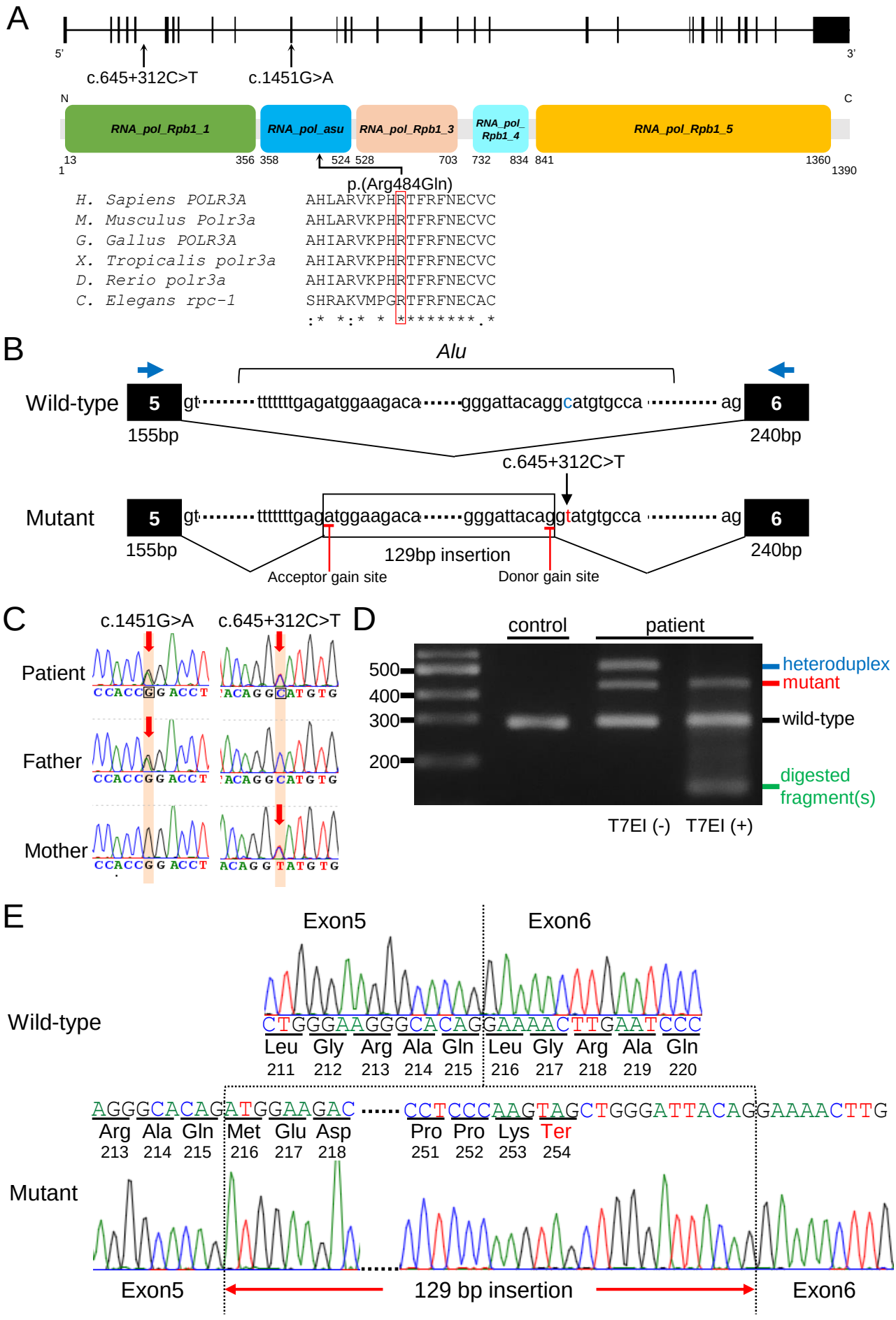
was digested with T7 endonuclease I (T7EI), and digested fragment(s) could be observed as the lower band(s), indicating heteroduplex formation. (E) The sequence of wild and mutant amplicons clearly shows inclusion of a 129-nucleotide intronic sequence between exons 5 and 6 (dashed lines) as predicted by SpliceAI resulting in a premature stop codon.

Table 1. Splicing predictions of the c.645+312C>T variant by SpliceAI and multiple bioinformatic splicing prediction tools.

Prediction tools	Acceptor gain site		Donor gain site	
	Wild-type score	Mutant score	Wild-type score	Mutant score
SpliceAI	N.A.	0.55	N.A.	0.77
NetGene2	0.17	0.23	-	Detected but its confidence score is 0.00
ESEfinder	7.06220	7.06220	-	10.88050
Human Splicing Finder	84.23	84.23	-	91.2

SpliceAI predicts both acceptor and donor gains. SpliceAI scores were evaluated as 0.2 (high recall), 0.5 (recommended), and 0.8 (high precision). Other prediction tools predicted a donor gain site in mutant sequences. Acceptor gain is also predicted in the wild-type sequence suggesting a cryptic splice acceptor site, but their scores are not increased by the variant in two programs. NetGene2: http://www.cbs.dtu.dk/services/NetGene2/, ESE finder: http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home, Human Splicing Finder: http://www.umd.be/HSF/, N.A., not available.


A
B
C
D


A
5'
3'
c.645+312C>T
c.1451G>A
N
C
RNA_pol_Rpb1_1
RNA_pol_asu
RNA_pol_Rpb1_3
RNA_pol_Rpb1_4
RNA_pol_Rpb1_5
13
356
358
524
528
703
732
834
841
1360
1
1390
p.(Arg484Gln)
H. Sapiens POLR3A AHLARVKPHRTFRFNECVC
M. Musculus Polr3a AHLARVKPHRTFRFNECVC
G. Gallus POLR3A AHIARVKPHRTFRFNECVC
X. Tropicalis polr3a AHIARVKPHRTFRFNECVC
D. Rerio polr3a AHIARVKPHRTFRFNECVC
C. Elegans rpc-1 SHRAKVMPGRTFRFNECAC
:* *:* * ********.*
B
Alu
Wild-type
5
gt
tttttttgagatggaagaca
gggattacaggcatgtgcca
ag
6
155bp
240bp
c.645+312C>T
Mutant
5
gt
tttttttgagatggaagaca
gggattacaggtatgtgcca
ag
6
155bp
129bp insertion
240bp
Acceptor gain site
Donor gain site
C
c.1451G>A
c.645+312C>T
Patient
Father
Mother
D
control
patient
500
400
300
200
heteroduplex
mutant
wild-type
digested fragment(s)
T7EI (-)
T7EI (+)
E
Exon5
Exon6
Wild-type
CTGGGAAGGGCACAGGAAAACTTGAATCCC
Leu Gly Arg Ala Gln Leu Gly Arg Ala Gln
211 212 213 214 215 216 217 218 219 220
AGGGCACAGATGGAAGAC ..... CCTCCCAAGTAGCTGGGATTACAGGAAAACTTG
Arg Ala Gln Met Glu Asp Pro Pro Lys Ter
213 214 215 216 217 218 251 252 253 254
Mutant
Exon5
129 bp insertion
Exon6

Supplementary Methods

In WES and WGS analysis, reads were aligned to the GRCh38 reference genome using BWA-mem (Version 0.7.17) with default parameters. Duplicated reads were removed by Picard (Version 2.20.3), and local realignment and base quality recalibration were performed by GATK Version 3.8. Variants were identified with the GATK HaplotypeCaller, and raw variants were filtered out when their parameters met any of the following values: QD < 2.0, MQ < 40.0, FS > 60.0, MQRankSum < −12.5, and ReadPosRankSum < −8.0 for single nucleotide variants; and QD < 2.0, ReadPosRankSum < −20.0, and FS > 200.0 for insertion/deletions. Final variants were annotated with Annovar (Wang et al. 2010) for the predictive value of the functional impact of the coding variants and assessing allele frequency: an in-house database of 218 control exomes, the Human genetic variation database (http://www.hgvd.genome.med.kyoto-u.ac.jp/) (Higasa et al. 2016), 4.7KJPN (https://ijgvd.megabank.tohoku.ac.jp/) (Nagasaki et al. 2015), and the gnomAD database (https://gnomad.broadinstitute.org/) (Lek et al. 2016). We defined “rare variants” as a minor allele frequency equal or less than 0.01 in the above four databases. Variant pathogenicity was predicted by SIFT, Polyphen-2, MutationTaster (Schwarz et al. 2014), CADD (Kircher et al. 2014) and M-CAP (Jagadeesh et al. 2016) (Table S1). Nucleotide conservation prediction was performed using GERP (http://mendel.stanford.edu/SidowLab/downloads/GERP/index.html) and PhastCons (http://compgen.cshl.edu/phast/). Splicing junction prediction was performed using SpliceAI (Jaganathan et al. 2019). Candidate variants were confirmed by Sanger sequencing using an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA).

Supplementary references

Wang K, Li M, Hakonarson H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164. doi: 10.1093/nar/gkq603.

Higasa K, Miyake N, Yoshimura J, Okamura K, Niihori T, Saitsu H, Doi K, Shimizu M,

Nakabayashi K, Aoki Y, Tsurusaki Y, Morishita S, Kawaguchi T, Migita O, Nakayama K, Nakashima M, Mitsui J, Narahara M, Hayashi K, Funayama R, Yamaguchi D, Ishiura H, Ko WY, Hata K, Nagashima T, Yamada R, Matsubara Y, Umezawa A, Tsuji S, Matsumoto N, Matsuda F (2016) Human genetic variation database, a reference database of genetic variations in the Japanese population. J Hum Genet 61:547-53. doi: 10.1038/jhg.2016.12.

Nagasaki M, Yasuda J, Katsuoka F, Nariai N, Kojima K, Kawai Y, Yamaguchi-Kabata Y, Yokozawa J, Danjoh I, Saito S, Sato Y, Mimori T, Tsuda K, Saito R, Pan X, Nishikawa S, Ito S, Kuroki Y, Tanabe O, Fuse N, Kuriyama S, Kiyomoto H, Hozawa A, Minegishi N, Douglas Engel J, Kinoshita K, Kure S, Yaegashi N, To MJRPP, Yamamoto M (2015) Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat Commun 6: 8018. doi: 10.1038/ncomms9018

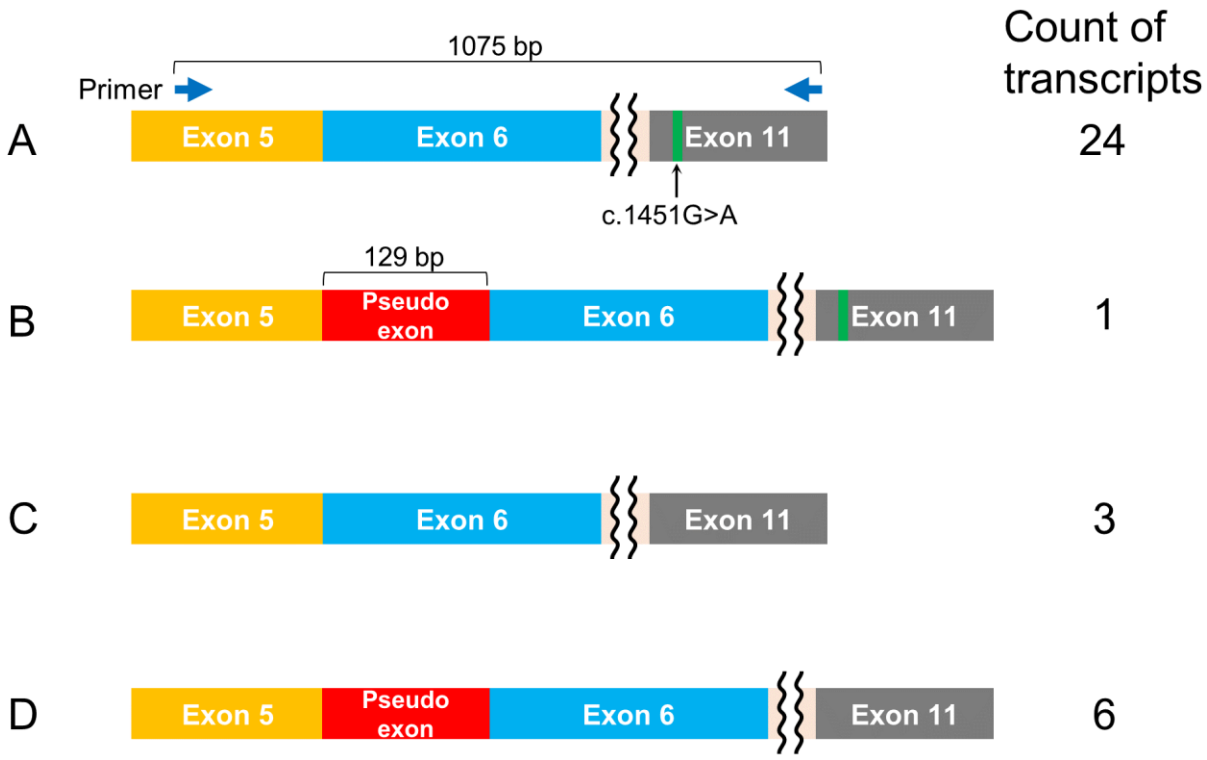
Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536: 285-91. doi: 10.1038/nature19057

Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014 Apr;11(4):361-2. doi: 10.1038/nmeth.2890.
Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46: 310-5. doi: 10.1038/ng.2892

Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, Bernstein JA, Bejerano G (2016) M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet 48: 1581-1586. doi: 10.1038/ng.3703

Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglou S, Sanders SJ, Farh KK. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019 Jan 24;176(3):535-548.e24. doi: 10.1016/j.cell.2018.12.015.

Supplemental Figure S1



Schematic representation of the cDNA sequences of the *POLR3A* from exon 5 to exon 11. RT-PCR was performed using primers designed at exon 5 and 11 (arrows). PCR products were cloned, and transcripts were classified and counted. We identified four types of transcripts: 24 transcripts with the c.1451G>A variant and no pseudo-exon (A), 1 transcript with the c.1451G>A variant and the pseudo-exon (B), 3 transcripts without the c.1451G>A variant and the pseudo-exon (C), and 6 transcripts with no c.1451G>A variant and the pseudo-exon (D).

Table S1. Candidate *POLR3A* variants identified by WES and WGS

Chr	Gene	Variant	Origin	gnomAD	SIFT	PP2 HVAR	CADD phred	M-CAP	GERP	Mutation Taster
10	*POLR3A*	NM_007055.3:c.1451G>A,p.(Arg484Gln)	paternal	0.000003980	0.001	0.935	29.7	0.1270	5.69	1
10	*POLR3A*	NM_007055.3:c.645+312C>T	maternal	−	N.A.	N.A.	5.845	N.A.	N.A.	N.A.

gnomAD (the Genome Aggregation Database): http://gnomad.broadinstitute.org/, SIFT (Sorting Intolerant From Tolerant): http://sift.jcvi.org/, Polyphen-2 Hum Var: http://genetics.bwh.harvard.edu/pph2/, CADD (Combined Annotation–Dependent Depletion): http://cadd.gs.washington.edu/score, M-CAP (Mendelian Clinically Applicable Pathogenicity): http://bejerano.stanford.edu/mcap/index.html, GERP (Genomic Evolutionary Rate Profiling): http://mendel.stanford.edu/SidowLab/downloads/gerp/, MutationTaster: http://www.mutationtaster.org/. N.A., not available.

Table S2. Variant classification following the ACMG guideline

Gene	Variant	Evidence of pathogenicity				Classification
		Very strong	Strong	Moderate	Supporting	
POLR3A	c.1451G>A,p.(Arg484Gln)	Not applicable	Not applicable	PM2, PM3	PP2, PP3	Likely pathogenic
POLR3A	c.645+312C>T	PSV1	Not applicable	PM2	Not applicable	Likely pathogenic

PSV1: Null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where loss of function is a known mechanism of disease.

PS2: *De novo* (both maternity and paternity confirmed) in a patient with the disease and no family history.

PM2: Absent from controls (or at extremely low frequency if recessive) in the Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium.

PM3: For recessive disorders, detected *in trans* with a pathogenic variant.

PP2: Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease.

PP3: Multiple lines of computational evidence support a deleterious effect on the gene or gene product (e.g., conservation, evolutionary, splicing impact).

Table S3. Numbers of variants predicted to affect splicing in chromosome 1 by SpliceAI

				High precision (>=0.8)			
	Total variants	High recall (>=0.2)	Recommended (>=0.5)	Total	Canonical splice sites	Intronic	Exonic including UTRs
All variants	141,609	383	44	19	5	11	3
Rare variants	2,500	10	1*	1	1*	0	0

Notes:
#1 SpliceAI only annotates variants within genes defined by the gene annotation file (https://github.com/Illumina/SpliceAI/blob/master/spliceai/annotations/grch38.txt). Thus we selected variants within these regions in chromosome 1, then performed SpliceAI analysis (version 1.3.1).
#2 Numbers of variants showing scores above each threshold in either acceptor gain, acceptor loss, donor gain, and donor loss were counted.
#3 All the canonical splice site variants were predicted as "High precision".
#4 See supplementary methods for filtering criteria of rare variants.
*This rare canonical variant (NM_001010969:c.898-1G>A) was the only variant that had score >=0.5.