Clinical Epigenetics

BRIEF REPORT

Open Access

Reverse genotyping: unveiling *Alu* element insertion as a new cause of Kabuki syndrome using DNA methylation signature

Quentin Sabbagh[1,2], Nathalie Ruiz-Pallares[3], Cassandra Rastin[4], Jacques Puechberty[1], Thomas Guignard[3], Claire Jeandel[5], Fanny Merklen[6], Pascal Pujol[2,7], Jennifer Kerkhof[4], Bekim Sadikovic[4], Mouna Barat-Houari[3] and David Geneviève[1,2*]

Abstract

Kabuki syndrome type 1 (KS1) is a monogenic disorder arising from pathogenic variants within *KMT2D* and characterized by syndromic neurodevelopmental delay. We report the retrospective identification of a causative *Alu*Y insertion within *KMT2D* in a genetically unsolved individual with typical KS1 features, after identification of a DNA methylation signature. This is the first documentation of *Alu* insertion as a molecular mechanism responsible for KS1. This study emphasizes the need for reanalyzing inconclusive sequencing data in individuals with gene-specific phenotypes and reinforces episignature as a reliable diagnostic tool when NGS approaches fail to provide conclusive results in individuals with rare diseases.

Keywords Kabuki syndrome, *KMT2D*, Epigenetic signature, *Alu* element, Mobile element insertion

Introduction

Kabuki syndrome type 1 (KS1, MIM 147920) is a phenotypically recognizable autosomal dominant monogenic disorder arising from pathogenic variants within *KMT2D*. Clinical spectrum of KS1 usually encompasses syndromic intellectual disability in association with typical facial features, postnatal growth retardation, microcephaly, malformation syndrome, and autoimmune disorders [1].

KMT2D, also known as MLL4, is a ubiquitously expressed protein belonging to the family of histone H3 lysine 4 (H3K4) methyltransferases. More specifically, KMT2D interacts with the WRAD tetrameric module (WDR5, RbBP5, ASH2L, DPY30) along with four additional subunits, namely KDM6A, PTIP, PA1, and NCOA6, to form the KMT2D/MLL4-related complex of proteins associated with Set1 (COMPASS) histone methyltransferase complex [2]. In mammals, this KMT2D/MLL4-COMPASS complex is involved in H3K4 monomethylation (H3K4me1) at active enhancer

*Correspondence:
David Geneviève
d-genevieve@chu-montpellier.fr
[1] Department of Clinical Genetics, Centre de Référence « Anomalies du Développement et Syndromes Malformatifs », University Hospital of Montpellier, Inserm UMR1183, Montpellier University, Montpellier, France
[2] French Society for Predictive and Personalized Medicine (SFMPP), Montpellier, France
[3] Department of Molecular Genetics and Cytogenomics, University Hospital of Montpellier, Montpellier University, Montpellier, France
[4] Verspeeten Clinical Genome Centre, London Health Sciences Centre, London, ON N6A 5W9, Canada
[5] Department of Paediatric Endocrinology, University Hospital of Montpellier, Montpellier University, Montpellier, France
[6] Department of Otorhinolaryngology, University Hospital of Montpellier, Montpellier University, Montpellier, France
[7] Department of Cancer Genetics, University Hospital of Montpellier, UMR IRD 224-CNRS 5290, Montpellier University, Montpellier, France

regions where it further co-localizes with p300 and CREBBP acetyltransferases responsible for histone H3 lysine 27 acetylation (H3K27ac). KMT2D/MLL4-COMPASS-mediated H3K4 monomethylation has been demonstrated to dynamically orchestrate expression of developmental genes playing a critical role upon early embryonic development, particularly in the exit from ground-state pluripotency of embryonic stem cells [2].

Numerous rare genetic disorders including KS1 have been linked to unique DNA methylation profiles, known as "episignatures." In recent years, the so-called episignatures have emerged as robust and reliable biomarkers, playing a crucial role in diagnosing congenital genetic conditions and reclassifying variants of uncertain significance (VUS). Their application in clinical diagnostic laboratories has demonstrated significant utility in providing precision diagnoses for individuals with suspected rare monogenic disorders who previously lacked a clear molecular diagnosis, achieving an overall 18.7% diagnostic yield [3, 4].

Here, we report the diagnostic uncovering of an *Alu*-mediated KS1 in an individual with highly compatible clinical features, but inconclusive next-generation sequencing (NGS) analyses, for whom we applied DNA methylation profiling followed by a "reverse genotyping" hunting strategy conducted on previously generated sequencing data. This study is, to the best of our knowledge, the first to document *Alu* element insertion in *KMT2D* as a molecular mechanism responsible for KS1.

Materials and methods

Gene panel and exome sequencing

Gene panel and exome sequencing were performed using NGS, targeting exonic regions, and flanking splice junctions. The gene panel included 98 chromatinopathy-related genes, including *KMT2D* and *KDM6A*. Both single-nucleotide variants (SNVs) and structural variants (SVs) were initially analyzed, and BAM files from the gene panel sequencing were subsequently used to retrieve the sequence of the *Alu*-mediated pathogenic allele in *KMT2D*.

DNA methylation analysis

Methylation analysis was conducted using the clinically validated EpiSign™ assay, following previously established methods [3, 4]. Methylated and unmethylated signal intensities generated from the EPIC array were imported into R 3.5.1 for normalization, background correction, and filtering. The classifier utilized the EpiSign™ Knowledge Database, which consists of over 10,000 methylation profiles from reference disorder-specific and unaffected control cohorts, to generate disorder-specific methylation variant pathogenicity (MVP) scores. These MVP scores are a measure of prediction confidence for each disorder and range from 0 (discordant) to 1 (highly concordant). The final matched EpiSign™ result is generated using these scores, along with the assessment of hierarchical clustering and multidimensional scaling.

PCR analysis and breakpoint sequencing

KMT2D exon 36 (NM_003482.4) was amplified in the proband and both healthy parents and analyzed by agarose gel electrophoresis using the following primers: forward (5′–3′): GGCGTGGTTGAAGTCAGGAT; reverse (5′–3′): TAGGGAGGGGAGCCAAGAAG. *Alu*-related upstream breakpoint (BP) analysis was performed by Sanger sequencing using the same primer pair. In contrast, downstream breakpoint analysis required a distinct set of primers, specifically designed based on the *Alu*-mediated alternative allele to selectively target the terminal poly(A) tail: forward (5′–3′): AAGAAAGTG ATGGCTCGGC; reverse (5′–3′): TAGGGAGGGGAG CCAAGAAG.

Results

Clinical description and molecular testing

Reported individual is the first of two children born from healthy and unrelated parents. Following an uneventful pregnancy, he was born at term with birth parameters lying within normal range. At birth, he presented with malformation syndrome associating atrial and ventricular septal defect, single umbilical artery, T7-to-T10 hemi-vertebrae, and diaphragmatic hernia. He further exhibited facial asymmetry and high arched palate, as well as feeding difficulties and neonatal hypoglycemia. At three years of age, diagnosis of KS1 was considered based on evocative facial features (Fig. 1A–C) and their association with global neurodevelopmental delay, postnatal growth retardation (−3 SD), and mixed bilateral hearing loss. The affected individual was therefore offered a targeted gene panel sequencing, which did not detect any pathogenic variants within *KMT2D* nor *KDM6A*. This was followed by a trio exome sequencing, which also yielded no molecular findings. Upon his last examination at sixteen years of age, he exhibited intellectual disability along with short stature (−2.5 SD) and surgically corrected *pes planus*. Immunoglobulins levels were within normal range, as well as immunophenotyping results.

DNA methylation profiling

EpiSign™ variant targeted analysis revealed a genome-wide DNA methylation profile consistent with Kabuki syndrome (Fig. 1D, E). This episignature was not only concordant with the methylation pattern observed

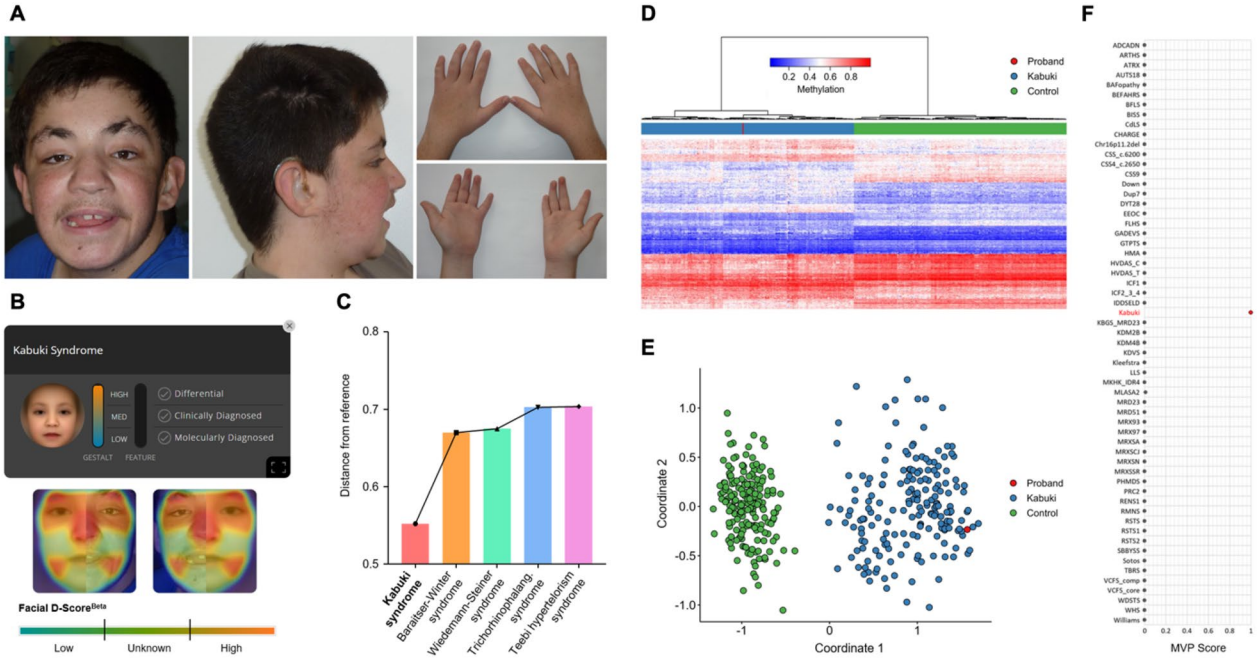

Fig. 1 Assessment of affected individual's craniofacial morphology and peripheral blood DNA methylation profile. **A** Affected individual displays morphological features evocative of Kabuki syndrome (KS) type 1 including arched and interrupted eyebrows, long and wide-apart palpebral fissures, long eyelashes, flat facial profile, persistence of finger pads, and vanished distal interphalangeal crease of third and fourth fingers. **B** *Face2Gene* (https://www.face2gene.com/) and **C** *GestaltMatcher* (https://www.gestaltmatcher.org/) artificial intelligence-based next-generation phenotyping tools highlighting KS as the most likely diagnosis based on individual's frontal photograph. **D** and **E** Hierarchical clustering and multidimensional scaling plots indicate the proband (red) has a DNA methylation profile similar to individuals with confirmed KS episignature (blue) and distinct from controls (green). **F** MVP score, a multiclass supervised classification system capable of discerning between multiple episignatures, showing a methylation signature similar to the KS reference

in individuals with pathogenic variants in *KMT2D* or *KDM6A*, but also distinct from other established episignatures associated with genetic disorders (Fig. 1F), as demonstrated by Euclidean clustering, multidimensional scaling, and an elevated MVP score (1.0).

Reverse genotyping approach

The previously conducted gene panel sequencing was retrospectively re-evaluated with a focus on *KMT2D* and *KDM6A*. Meticulous visual reanalysis using the *Integrative Genomics Viewer* (*IGV*) software uncovered a small heterozygous insertion in *KMT2D* exon 36, characterized by abnormal soft-clipped reads and poly(T) tracks associated with increased covering depth in this genomic region (Fig. 2A). The alternative allele sequence was then retrieved in silico from BAM files (Supplemental Fig. 1A) and processed with *RepeatMasker* (https://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) tool which revealed a 100% match to an *Alu* element spanning 281 bp (Fig. 2B, Supplemental Fig. 1B). This in silico prediction was validated by PCR, confirming the de novo heterozygous insertion in *KMT2D* exon 36 (Fig. 2C). The diagnostic loop was finally closed by breakpoint analysis which characterized the insertion as a 335 bp sequence within *KMT2D*, encompassing the *Alu* element's core (281 bp), its poly(A) tail (38 bp), and a 16 bp target site duplication (TSD) (Fig. 2D-E, Supplemental Fig. 1A).

Discussion

Alu elements are retrotransposons belonging to the family of short interspersed nuclear elements (SINEs) which propagate throughout the genome using a mechanism of target-primed reverse transcription (TPRT). These ~300 bp sequences are ubiquitous within the human genome, with an estimated copy number of 1.1 million, and have significantly contributed to primate genome diversity across evolution. *Alu* element's structure is composed of two monomers (right and left) deriving from 7SL RNA, associated with a 5′ internal RNA polymerase III promoter (A and B boxes), acting as a transcription initiating site, and a 3' poly(A) tail [5]. Following its transcription, *Alu* element is provided with both endonuclease and reverse transcriptase by the L1 enzyme machinery. 5′-TTT/AA-3′ consensus site is firstly recognized by L1 endonuclease, resulting in

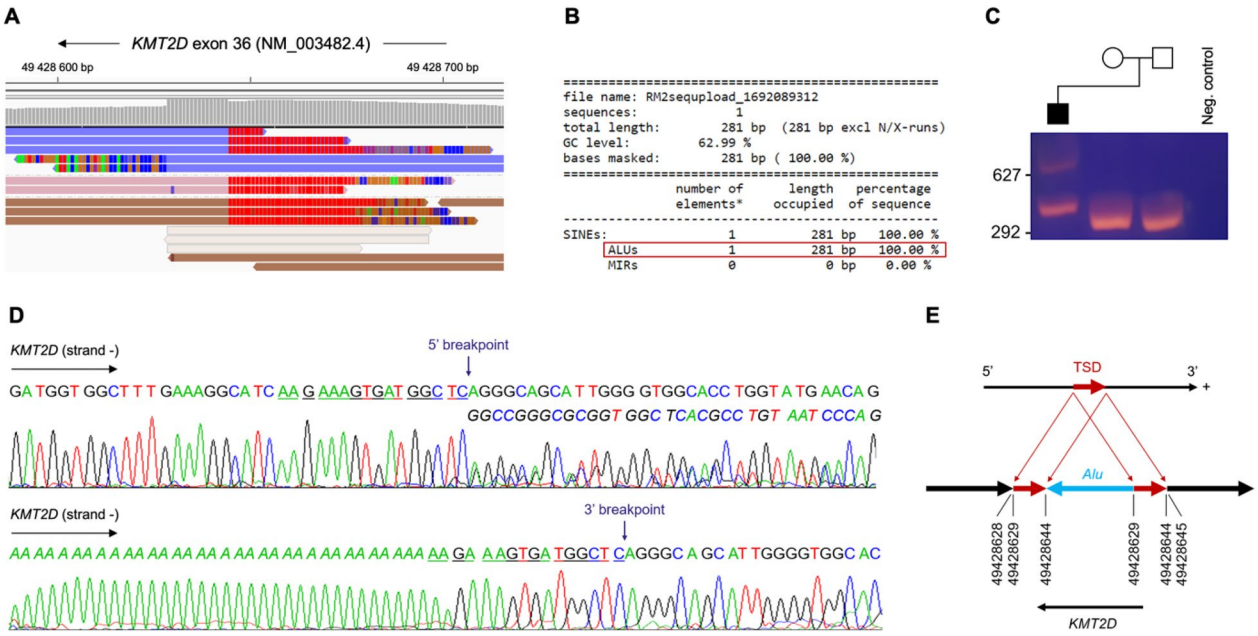

Fig. 2 Unmasking of *Alu* element insertion in *KMT2D* exon 36 through reverse genotyping approach. **A** Visual reassessment of the previously performed gene panel sequencing using *Integrative Genomics Viewer* (*IGV*) software showing soft-clipped reads mapping on *KMT2D* exon 36 along with increased covering depth. **B** Alignment data of the in silico-retrieved inserted sequence consistent with an *Alu* element insertion, as per *RepeatMasker* tool. **C** PCR analysis ascertaining the heterozygous *KMT2D* insertion in the proband, resulting in an abnormal 627 bp amplicon absent in both parents and the negative control. **D** Breakpoint analysis corroborating the in silico-expected *Alu* element insertion within *KMT2D*. *Alu* element sequence is shown in italic while TSD pattern is underlined. **E** Summarized diagram of the reported *Alu* element exonic insertion in *KMT2D* as a novel molecular mechanism responsible for Kabuki syndrome type 1 (GRCh37/hg19)

the cleavage of the related genomic region. *Alu*-related RNA then undergoes reverse transcription as a result of its binding to the released consensus region, while complementary sequence is subsequently synthetized by host DNA polymerase using *Alu* element's single-stranded DNA as a template. Upon *Alu* element insertion into such new genomic region, DNA target oligomer is further duplicated, generating two flanking repeat sequences of 7–20 bp, coined as target site duplications (TSD) [5].

Over the past 30 years, *Alu* elements have been demonstrated to be involved in a wide spectrum of monogenic conditions through distinct molecular mechanisms, both at the genomic and transcript level. As a matter of fact, *Alu* recombination-mediated deletions (ARMD) are estimated to be responsible for approximately 0.3% of human genetic disorders and arise from *Alu*-induced structural rearrangements, either occurring upon meiosis I through *Alu*-*Alu*-mediated rearrangements (AAMR) and non-homologous end joining (NHEJ), or during DNA replication by fork stalling and template switching/microhomology-mediated break-induced repair (FoSTeS/MMBIR) [6]. At the transcript level, intronic *Alu* elements might impair gene expression by activating a cryptic alternative splicing site, which may be accompanied by the creation of a novel exon (*i.e.*, exonization), or by inducing an exon skipping phenomenon. On the other hand, *Alu* insertion in mature mRNA might disrupt the open reading frame (ORF) by the introduction of a frameshift variation resulting in a premature termination codon (PTC), but also alter protein 3D structure, or foster the production of double-stranded RNA subsequently processed through mRNA decay and/or adenosine-to-inosine (A-to-I) editing [7].

Strikingly, despite individual's phenotype being highly compatible with KS1, both targeted gene panel and trio exome sequencing performed in 2010 and 2013, respectively, failed to detect any pathogenic variants in *KMT2D* and *KDM6A*. Such discrepancy prompted us to analyze individual's peripheral blood methylation profile using the EpiSign™ assay which provided a high confidence KS episignature, among more than 100 other episignatures detectable by the assay [3, 4]. This crucial finding led us to revisit individual's sequencing data through a reverse genotyping approach, ultimately unveiling an *Alu* element insertion within *KMT2D* exon 36 initially missed by both previous NGS approaches.

This *Alu* element is considered as being part of the *Alu*Y subtype as per *RepeatMasker* tool (Supplemental

Fig. 1B) and estimated to be 319 bp (*i.e.*, 281 bp main core; 38 bp poly(A) tail). According to *VariantValidator*, such heterozygous de novo insertion is anticipated to result in *KMT2D* loss-of-function by inducing the following truncating variant: NM_003482.4 (NP_003473.3):p.(Lys3436Glyfs*40) (GRCh37/hg19). This frameshift variant is considered as pathogenic (class 5) in compliance with the ACMG/AMP guidelines as it meets the following criteria: PVS1 (pLI = 1, o/e ratio = 0.11, ClinGen HI score = 3), PS2, PM2, and PP4. Nevertheless, only additional in vitro mRNA and/or protein functional studies might definitely ascertain KMT2D reduced expression as a result of this *Alu*Y exonic insertion.

At the time both gene panel and exome sequencing were conducted, bioinformatics pipelines and variant callers were likely not sufficiently optimized to detect atypical structural variations, including mobile element insertions (MEIs). Hence, this limitation may have contributed to the inconclusive results obtained in 2010 and 2013. Ever since, tailored algorithms like the Mobile Element Locator Tool (MELT) have been specifically developed to foster the identification such MEIs that might otherwise remain undetected [8]. Looking ahead, long-read sequencing (LRS) technologies are also expected to offer unparalleled advantages in identifying novel atypical rearrangements, as evidenced by recent studies demonstrating LRS's ability to detect disease-causing MEIs initially overlooked by short-read approaches [9]. In the same vein, optical genome mapping (OGM) has proved to be a valuable tool for uncovering MEIs resistant to conventional techniques, facilitating the resolution of complex molecular diagnoses [10]. As these technologies continue to evolve, LRS and OGM are anticipated to become a *sine qua non* for the identification of rare structural rearrangements, thereby opening the door to precision medicine and personalized follow-up for individuals with unsolved rare diseases.

Overall, this study underscores the importance of reanalyzing inconclusive sequencing data in individuals presenting gene-specific phenotypes, as it may uncover atypical and scarcely reported pathogenic mechanisms, such as MEIs. Additionally, it refines the genetic landscape associated with KS1, while reiterating episignature analysis robustness as a valuable diagnostic tool for rapidly expanding repertoire of rare diseases.

Abbreviations

AAMR	*Alu*–*Alu*-mediated rearrangements
BP	Breakpoint
COMPASS	Complex of proteins associated with Set1
FoSTeS/MMBIR	Fork stalling and template switching/microhomology-mediated break-induced repair
KS1	Kabuki syndrome type 1
LRS	Long-read sequencing
MEI	Mobile element insertion
MELT	Mobile Element Locator Tool
MVP	Methylation variant pathogenicity
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
OGM	Optical genome mapping
ORF	Open reading frame
PTC	Premature termination codon
SINEs	Short interspersed nuclear elements
SNVs	Single-nucleotide variants
SVs	Structural variants
TPRT	Target-primed reverse transcription
TSD	Target site duplication
VUS	Variant of uncertain significance

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13148-025-01879-z.

Supplementary Material 1. *Alu* element-related in silico data. **A** In silico-retrieved sequence of the *Alu*-mediated *KMT2D* pathogenic allele from gene panel's BAM files. **B** *Alu* sequence alignment using *RepeatMasker*.

Acknowledgements
We deeply thank the family for participating in the study, along with the French Society for Predictive and Personalized Medicine (SFMPP) for supporting this work.

Author contributions
Conception and design of the study were performed by QS, NRP, CR, BS, MBH, and DG. Acquisition of data, analysis and/or interpretation of data, revising the manuscript, and approval of the version of the manuscript to be published were done by QS, NRP, CR, JP, TG, CJ, FM, PP, JK, BS, MBH, and DG. Drafting the manuscript was done by QS and DG.

Funding
Funding for this study was provided, in part, by the London Health Sciences Molecular Diagnostics Innovation and Development Fund and Genome Canada Genomic Applications Partnership Program Grant (Beyond Genomics: Assessing the Improvement in Diagnosis of Rare Diseases using Clinical Epigenomics in Canada, EpiSign-CAN) awarded to BS. This work was further supported by a national grant from the French Ministry of Health (PHRC AOM-09-070) awarded to DG. Article processing charges (APC) were covered by the French Society for Predictive and Personalized Medicine (SFMPP).

Availability of data and materials
No datasets were generated or analyzed during the current study.

Declarations

Ethical approval and consent to participate
This research protocol has been validated by the ethics committee (CPP) of the French national research program (PHRC AOM-09-070; ClinicalTrial.gov ID: NCT01314534). Written consent was further received from the family for participation in the study and publication of photographs. Clinical examination, molecular analyses, and DNA methylation profiling were performed as part of routine care.

Consent for publication
Please see above.

Competing interests
Bekim Sadikovic is a shareholder in EpiSign Inc., a biotech company involved in commercial uses of EpiSignTM technologies.

Received: 19 November 2024 Accepted: 9 April 2025
Published online: 29 April 2025

References

1. Adam MP, Banka S, Bjornsson HT, Bodamer O, Chudley AE, Harris J, et al. Kabuki syndrome: international consensus diagnostic criteria. J Med Genet. 2019;56(2):89–95.
2. Froimchuk E, Jang Y, Ge K. Histone H3 lysine 4 methyltransferase KMT2D. Gene. 2017;627:337–42.
3. Kerkhof J, Rastin C, Levy MA, Relator R, McConkey H, Demain L, et al. Diagnostic utility and reporting recommendations for clinical DNA methylation episignature testing in genetically undiagnosed rare diseases. Genet Med. 2024;26(5): 101075.
4. Levy MA, McConkey H, Kerkhof J, Barat-Houari M, Bargiacchi S, Biamino E, et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. Human Genet Genomics Adv. 2022;3(1): 100075.
5. Deininger P. Alu elements: know the SINEs. Genome Biol. 2011;12(12):236.
6. Callinan PA, Wang J, Herke SW, Garber RK, Liang P, Batzer MA. Alu retrotransposition-mediated deletion. J Mol Biol. 2005;348(4):791–800.
7. Gussakovsky D, McKenna SA. Alu RNA and their roles in human disease states. RNA Biol. 2021;18(sup2):574–85.
8. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, et al. The mobile element locator tool (MELT): population-scale mobile element discovery and biology. Genome Res. 2017;27(11):1916–29.
9. Watson CM, Crinnion LA, Lindsay H, Mitchell R, Camm N, Robinson R, et al. Assessing the utility of long-read nanopore sequencing for rapid and efficient characterization of mobile element insertions. Lab Invest. 2021;101(4):442–9.
10. Sabatella M, Mantere T, Waanders E, Neveling K, Mensenkamp AR, van Dijk F, et al. Optical genome mapping identifies a germline retrotransposon insertion in SMARCB1 in two siblings with atypical teratoid rhabdoid tumors. J Pathol. 2021;255(2):202–11.

Publisher's Note