Journal of Pathology
J Pathol October 2021; **255:** 202–211
Published online 29 July 2021 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/path.5755

ORIGINAL PAPER

Optical genome mapping identifies a germline retrotransposon insertion in *SMARCB1* in two siblings with atypical teratoid rhabdoid tumors

Mariangela Sabatella[1†], Tuomo Mantere[2,3,4†], Esmé Waanders[5], Kornelia Neveling[2], Arjen R Mensenkamp[2,3], Freerk van Dijk[1], Jayne Y Hehir-Kwa[1], Ronnie Derks[2,3], Michael Kwint[2,3], Luke O'Gorman[2,3], Madalena Tropa Martins[1], Corrie EM Gidding[1], Maarten H Lequin[6], Benno Küsters[7], Pieter Wesseling[1,8], Marcel Nelen[2,3], Jacklyn A Biegel[9,10], Alexander Hoischen[2,3,11‡], Marjolijn C Jongmans[1,5‡] and Roland P Kuiper[1,2*‡]

[1] *Princess Máxima Centre for Pediatric Oncology, Utrecht, The Netherlands*
[2] *Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands*
[3] *Radboud Institute of Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands*
[4] *Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit and Biocenter Oulu, University of Oulu, Oulu, Finland*
[5] *Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands*
[6] *Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands*
[7] *Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands*
[8] *Department of Pathology, Amsterdam University Medical Centers, Location VUmc and Brain Tumor Center Amsterdam, Amsterdam, The Netherlands*
[9] *Department of Pathology and Laboratory Medicine, Children's Hospital, Los Angeles, Los Angeles, CA, USA*
[10] *Keck School of Medicine, University of Southern California, Los Angeles, CA, USA*
[11] *Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands*

**Correspondence to: RP Kuiper, Princess Máxima Center for Pediatric Oncology, Heidelberglaan 25, 3584 CS Utrecht, The Netherlands. E-mail: r.kuiper@prinsesmaximacentrum.nl*

[†]These authors contributed equally to this work.
[‡]These authors share senior authorship.

Abstract

In a subset of pediatric cancers, a germline cancer predisposition is highly suspected based on clinical and pathological findings, but genetic evidence is lacking, which hampers genetic counseling and predictive testing in the families involved. We describe a family with two siblings born from healthy parents who were both neonatally diagnosed with atypical teratoid rhabdoid tumor (ATRT). This rare and aggressive pediatric tumor is associated with biallelic inactivation of SMARCB1, and in 30% of the cases, a predisposing germline mutation is involved. Whereas the tumors of both siblings showed loss of expression of SMARCB1 and acquired homozygosity of the locus, whole exome and whole genome sequencing failed to identify germline or somatic *SMARCB1* pathogenic mutations. We therefore hypothesized that the insertion of a pathogenic repeat-rich structure might hamper its detection, and we performed optical genome mapping (OGM) as an alternative strategy to identify structural variation in this locus. Using this approach, an insertion of ~2.8 kb within intron 2 of *SMARCB1* was detected. Long-range PCR covering this region remained unsuccessful, but PacBio HiFi genome sequencing identified this insertion to be a SINE-VNTR-Alu, subfamily E (SVA-E) retrotransposon element, which was present in a mosaic state in the mother. This SVA-E insertion disrupts correct splicing of the gene, resulting in loss of a functional allele. This case demonstrates the power of OGM and long-read sequencing to identify genomic variations in high-risk cancer-predisposing genes that are refractory to detection with standard techniques, thereby completing the clinical and molecular diagnosis of such complex cases and greatly improving counseling and surveillance of the families involved.

Keywords: childhood cancer predisposition; rhabdoid tumors; SMARCB1; retrotransposon; optical imaging

Received 26 May 2021; Revised 28 June 2021; Accepted 4 July 2021

No conflicts of interest were declared.

Introduction

Extra-cranial malignant rhabdoid tumors [MRTs (MIM:609322)] and atypical teratoid rhabdoid tumors [ATRTs (MIM:609322)] are rare and aggressive pediatric tumors named after their 'rhabdoid' cells, typically characterized by eccentrically located, round nuclei with vesicular chromatin, prominent eosinophilic nucleoli, and extensive, homogeneously eosinophilic cytoplasm with occasional intracytoplasmic globular eosinophilic inclusions [1,2]. MRTs mostly occur in soft tissues [3], kidney [1], and liver [4]. ATRTs are located in the central nervous system (CNS) [5] and are mostly diagnosed in children younger than 3 years old [6–8]. Among all CNS tumors, ATRT is one of the most aggressive malignancies, with an extremely poor prognosis exacerbated by early age at diagnosis, high tumor stage, infratentorial location or metastasis [6,8]. Three distinct molecular subgroups of ATRT have been described [9,10]. ATRT-TYR is mostly found in the infratentorial compartment in infants and is characterized by overexpression of TYR and other melanosomal markers and DNA hypermethylation. ATRT-MYC tumors more often occur supratentorially in older children and are characterized by DNA hypomethylation and overexpression of MYC and HOX proteins. Finally, ATRT-SHH tumors have an infra- or supra-tentorial location with impairment of SHH and NOTCH factors, and DNA hypermethylation.

Despite the diverse pathogenesis of these three ATRT subtypes, they share loss of SMARCB1 expression in the tumor or, in rare cases, of SMARCA4 [11,12]. SMARCB1 is a bona fide tumor suppressor and one of the core subunits of the SWI/SNF chromatin remodeling complex, playing a role in epigenetic regulation, cell cycle progression, and crosstalk between signaling cascades [13]. *SMARCB1* [*INI1*, *BAF47* or *hSNF5* (MIM:601607)] is located on the long arm of chromosome 22 (22q11.23) [11] and usually harbors biallelic pathogenic mutations in the tumors of ATRT patients. About 30% of these ATRTs arise from germline pathogenic variants in one of the *SMARCB1* alleles [14–16]. Because of the high penetrance of pathogenic *SMARCB1* mutations and the extremely poor prognosis associated with this tumor, these germline variants often occur *de novo* [17], resulting in a lack of family history and low recurrence of this tumor type in other family members. Indeed, only a few cases of unaffected carriers of ATRT causative pathogenic variants have been described [16,18–22], suggesting the possibility of mosaicism or reduced penetrance of the disease in these subjects. Moreover, in some cases, causative *SMARCB1* variants are present in the non-coding region of the genome [23] and therefore not easily identified by standard diagnostic procedures.

We present a case of two siblings with congenital ATRT for which cancer predisposition was strongly suspected based on clinical and pathological analysis but no clear pathogenic *SMARCB1* variant was identified using routine diagnostics, including Sanger sequencing, microarray-based copy number analysis, whole exome sequencing (WES), and even (short-read) whole genome sequencing (WGS). An intronic insertion of an SVA-E retrotransposon element was eventually identified using optical genome mapping (OGM) and long-read sequencing.

Materials and methods

Clinical ascertainment and consent

The patients were identified in a diagnostic setting at Radboud University Medical Center in Nijmegen (The Netherlands). The parents provided written informed consent for publication of their data and those of their daughters.

Short-read sequencing

WES was performed in a diagnostic setting at the Department of Human Genetics of the Radboud University Medical Center. Germline DNA was extracted from peripheral blood lymphocytes of the two siblings (II-1 and II-2) and their parents (I-1 and I-2), and libraries were prepared using SureSelect Human All Exon enrichment kits v5 (Agilent Technologies, Santa Clara, CA, USA), which were sequenced on Illumina HiSeq4000 sequencing platforms (2 × 150 bp paired end). Reads were aligned to the reference genome build GRCh37/hg19 using BWA [24]. Genome Analysis Toolkit (GATK) [25] software, implemented following the GATK best practices [26], was used to identify single nucleotide variants (SNVs) and indels.

WGS solely on sibling II-2 and both parents was performed by the Hartwig Medical Foundation (HMF). These data were analyzed according to a standard HMF variant calling pipeline as described previously [27].

For the structural variant calling on WGS data, discordant paired-end alignments and split-read alignments were extracted from the aligned BAM file using Samtools (v1.3) [24]. Afterwards, Lumpy express (v0.2.13) [28] was used for structural variant (SV) calling. In addition, as a second SV detection method, the BAM files were analyzed using Delly Call (v0.7.7) [29], with default settings, and Mobster was used to search for mobile element insertions [30]. All outputs were then combined, and results were assessed. The suspected regions in the BAM files were visually inspected.

Single nucleotide polymorphism (SNP) array

SNP array analysis was performed on tumor DNA (extracted from fresh tissue for II-1 and FFPE for II-2) using the Affymetrix Oncoscan arrays (Affymetrix, Santa Clara, CA, USA) as previously described [31]. Data were then analyzed using Nexus Copy Number software, Edition 7 (Biodiscovery, El Segundo, CA,

USA) with reference genome GRCh37/hg19 and SNP-FASST2 Segmentation Algorithm. Loss of heterozygosity (LOH) was scored when larger than 15 Mb.

SMARCB1 transcript analysis

A standardized *SMARCB1* transcript analysis was performed as described previously [14]. In brief, cDNA was generated by reverse transcription of total RNA extracted from peripheral blood lymphocytes. Exons 1–3 of *SMARCB1* were amplified using PCR primers (cD1F1: CAGCCCTCCTGATCCCT; cD3R1: AACAGGGTCACACTGGTGG) developed by Children's Hospital, Los Angeles, based on the published sequence of *SMARCB1* [32]. The PCR conditions used were as follows: initial activation of enzyme at 95 °C for 5 min, 35 cycles of denaturation (95 °C for 30 s), annealing (60 °C for 30 s), extension (72 °C for 1 min), and final extension for 10 min at 72 °C. Sanger sequencing was performed using the BigDye Terminator v1.1 (Life Technologies, Carlsbad, CA, USA) on an automated fluorescent sequencer (ABI 3730 Genetic Analyzer; Applied Biosystems, Inc, Foster City, CA, USA). Sequences were compared with the normal reference control sequence (NM_003073.3) for *SMARCB1*.

DNA amplification for genotype analysis

To determine the parental origin of the genetic defect in the *SMARCB1* allele, WGS data from sibling II-2 and both parents were used to select ten variants in the *SMARCB1* locus that were heterozygous in the patient and discordant homozygous in the parents. These variants were subsequently Sanger-sequenced in the tumor of both siblings to determine the retained (duplicated) allele. The primers used for these analyses are listed in supplementary material, Table S1.

Optical genome mapping

OGM was performed as previously described [33,34]. In brief, high-molecular-weight germline DNA was extracted from an Epstein–Barr virus (EBV)-transformed lymphoblastoid cell line from II-2, using the SP Blood & Cell Culture DNA Isolation Kit according to the manufacturer's instructions (Bionano Genomics®, San Diego, CA, USA). The cell pellet (1.5 million cells) was resuspended in proteinase K and RNAse and mixed with LBB-lysis buffer to release genomic DNA (gDNA). The sample was treated with phenyl-methyl-sulfonyl-fluoride (PMSF) (Sigma-Aldrich, St Louis, MO, USA), and isopropanol was added after placing a Nanobind disk on each sample. After washing, gDNA was eluted from the disks and equilibrated overnight to facilitate DNA homogeneity. Next, gDNA was labeled according to the manufacturer's instructions using the Bionano Prep Direct Label and Stain (DLS) Protocol. The sample was subsequently stained overnight using the Bionano DNA stain reagent and loaded on a 3 × 1300 Gb Saphyr chip (G2.3) to be imaged using the Saphyr instrument. Bionano Solve v3.4 was used to perform the *de novo* genome assembly, variant calling, and annotation with default settings. Annotated variants were filtered for rare events as described previously [34].

PacBio HiFi genome sequencing

PacBio HiFi long-read genome sequencing was performed on high-molecular-weight DNA from the EBV-transformed lymphoblastoid cell line of II-2 that was extracted for Bionano OGM. SMRTbell libraries were prepared according to Procedure & Checklist – Preparing HiFi SMRTbell Libraries using SMRTbell® Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA). Size selection was performed using the BluePippin system (Sage Science, Beverly, MA, USA) to target fragments of ±18–22 kb. Sequence primer V2 and Polymerase 2.0 were used for binding. Sequencing was performed on the Sequel II System using a 68 pM on-plate loading concentration on an 8M SMRTcell. Sequencing was performed according to the manufacturer's instructions. Following sequencing, circular consensus reads were analyzed and mapped in SMRTlink 8.0 against human reference hg19. Structural variants were called using pbsv v2.2.2 (SMRTLink v8.0.0) and annotation was performed using an in-house pipeline with public databases including Decipher [35], Wellderly (http://genomics.scripps.edu), Genome of the Netherlands (GoNL) [36], 1000 Genomes Project [37,38], Exome Aggregation Consortium (ExAC) [39], and Database of Genomic Variants (http://dgv.tcag.ca).

Sanger validation of the SVA insertion

The identified insertion was validated by performing breakpoint-spanning PCRs using Taq Polymerase (Roche, Basel, Switzerland; 5' side breakpoint) or One Taq polymerase (New England Biolabs, Ipswich, MA, USA; 3' breakpoint). The primers used for amplification and sequencing are listed in supplementary material, Table S2.

To investigate the mosaicism level in germline DNA of I-2, a 1:2 serial dilution of peripheral blood-derived DNA (10 ng/μl) from II-2 or I-2 in control DNA (10 ng/μl) was performed with up to ten dilutions. Both the 5' and the 3' breakpoint-spanning PCRs were included in this analysis.

cDNA amplification to investigate the effect of the SVA insertion on splicing

To investigate the effect of the SVA insertion on splicing, RNA was extracted from EBV-transformed lymphoblasts of II-2 using the AllPrep-DNA/RNA/miRNA Universal kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. RNA was reverse-transcribed and cDNA was amplified using the SuperScript IV One-Step RT-PCR System (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. The primers used for amplification and sequencing are listed in supplementary material, Table S2.

Results

A hidden germline pathogenic aberration in *SMARCB1*

The first daughter (II-1) of a healthy Dutch couple was diagnosed with ATRT at birth, with negative immunohistochemical staining for SMARCB1 (supplementary material, Case description and Figure S1). Because of the high prevalence of germline *SMARCB1* pathogenic variants in congenital ATRT, II-1 was genetically tested for germline variants in this gene. However, targeted sequencing and WES as well as copy number aberration analysis were negative for pathogenic variants affecting exons or splice sites in *SMARCB1*. Unfortunately, a few years later, a second daughter (II-2) was also born with congenital ATRT, again negative for SMARCB1 immunostaining (supplementary material, Case description and Figure S1), strongly suggesting that a germline pathogenic variant was previously missed (Figure 1A). Both tumors were classified as ATRT-SHH subtype through methylome profiling (supplementary material, Case description). Whereas routine diagnostic tests also did not reveal a pathogenic germline aberration in sibling II-2, the tumors of both siblings did show loss of SMARCB1 expression (Figure 1B) and hypermethylation of the *SMARCB1* locus (supplementary material, Case description). Furthermore, SNP array analysis revealed an acquired uniparental disomy of chromosome 22, including the *SMARCB1* locus, with slightly different breakpoints between the tumors (Figure 1C). These results supported the hypothesis of a hidden pathogenic variant in the *SMARCB1* locus.

Next, standard clinically validated *SMARCB1* transcript analysis on blood-derived RNA from II-2 revealed the presence of a low-level alternatively spliced transcript in which exon 2 was skipped (Figure 1D). Exon 2 skipping results in a frameshift that may induce nonsense-mediated decay but since all *SMARCB1* splice sites were intact, the origin of this low-level mis-splicing remained obscure. Therefore, we performed WGS of germline DNA from II-2 to interrogate the intronic sequences of *SMARCB1*. Again, no pathogenic variant or copy number aberration was identified, but alignment of the sequencing reads showed heterozygous discordant mapping in intron 2 at the *SMARCB1* locus (Figure 2A). Unfortunately, various attempts using (long-range) PCR encompassing this intronic region amplified only the wild-type allele. We therefore hypothesized the presence of a causative germline aberration in intron 2 of the *SMARCB1* locus that impaired splicing, and likely involved a repetitive region that impeded proper alignment of WGS reads and DNA amplification.

Maternal inheritance

The diagnosis of ATRT in two siblings strongly suggested that the underlying germline aberration was inherited from one of the parents. WGS of germline DNA from the father showed no abnormalities, but we did observe discordant mapping in intron 2 in the mother, albeit at much lower levels, at the same position in which it was observed in II-2 (Figure 2A). These findings suggest that the mother is a mosaic carrier of the aberration, which could explain why she did not develop ATRT herself. To confirm that the aberration was located on one of the maternal alleles, we used WGS data from the parents to select ten SNPs in a region covering about 16 kb of the *SMARCB1* locus, which were homozygous but discordant in the two parents. Subsequently, we genotyped these SNPs in the tumors of both siblings, which revealed that, at all positions, the maternal allele was retained in homozygosity, thus confirming that the affected allele was of maternal origin (supplementary material, Table S1).

Optical genome mapping identified a structural variation at the *SMARCB1* locus

We hypothesized that the causative germline aberration in the *SMARCB1* locus could be an insertion and/or inversion that was missed due to the repetitive nature of the sequence or its breakpoints, which hampered amplification and mapping.

Therefore, we applied OGM using the Bionano Saphyr system (Bionano, San Diego, CA, USA) as an alternative strategy for structural variation detection in this locus. With this technique, ultra-high-molecular-weight single DNA molecules are fluorescently labeled on 6-nucleotide consensus sequences, creating recognizable locus-specific fingerprints that can be mapped to the genome [40]. Structural variants in a specific locus become discernible by spatial differences in the labeling pattern between sample and reference genome, even if these involve repeat-rich sequences or regions difficult to amplify. High-molecular-weight germline DNA from II-2 was extracted from available EBV-transformed lymphoblast cells and processed for OGM. Among the rare 36 SVs called in II-2 by Bionano (supplementary material, Figure S2), one was localized to the *SMARCB1* locus. In this locus, we revealed a shift in the labeling pattern corresponding to an insertion of ~2.8 kb with the insertion site in a region spanning ~5 kb from intron 1 to intron 3 of *SMARCB1* (Figure 2B). Since long-range PCR using several primer combinations around the predicted insertion region remained fruitless, we performed PacBio HiFi long-read genome sequencing to reveal the sequence and the exact position of the insertion. This technique reaches high accuracy through repetitive sequencing of long circularized fragments from which consensus reads are generated, a technique referred to as circular consensus sequencing (CCS) [41]. The long high-fidelity (HiFi) reads have an accuracy of 99.8% and provide unbiased coverage also for GC-rich regions [41,42]. Using the high-molecular-weight DNA from sibling II-2, we sequenced the genome with a coverage of 8x. Analysis of the PacBio HiFi reads by CCS mapping and SV calling revealed an insertion of 2763 bp, 33 bp downstream of exon 2 of *SMARCB1* (Figure 2C). The inserted sequence was identified as a

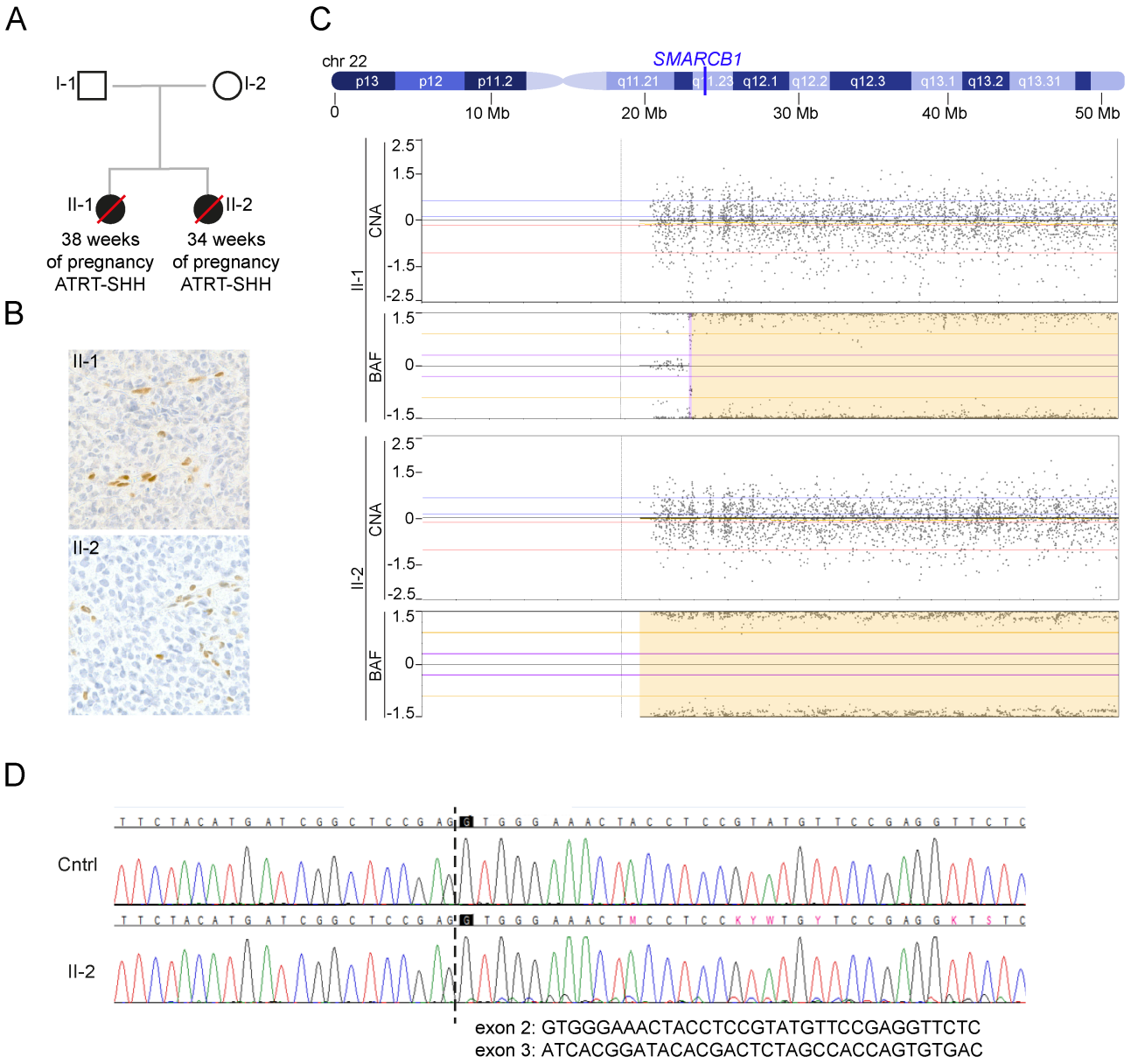

Figure 1. Germline pathogenic aberration in *SMARCB1* was not identified by routine diagnostics. (A) Pedigree of the family described in this study. (B) Formalin-fixed, paraffin-embedded (FFPE) tumor sections from the two siblings were used for immunohistochemistry analysis. SMARCB1 staining was absent in tumor cells. (C) Ideogram of chromosome 22 with the location of the *SMARCB1* locus (22q11.23) and SNP array analyses performed on tumor DNA from the two siblings. For each sibling, the copy number plot (CNA) and biallelic frequency plot (BAF) are shown. Both tumors showed loss of heterozygosity of the q-arm of chromosome 22 (marked with yellow in the BAF plots) with a breakpoint around 16 and 20 Mb for II-1 and II-2, respectively. (D) *SMARCB1* transcript analysis on blood-derived RNA from the second sibling (II-2) showing low-level exon 2 skipping. The analysis revealed low-level exon 3 sequence at the junction with exon 1 (indicated by the black dashed line) beside the presence of the exon 2 sequence. The result was compared with RNA sequencing of a control sample (Cntrl) that only showed the exon 2 sequence joined to exon 1.

SINE-VNTR-Alu (SVA) retrotransposon element, consisting of a CCCT repeat, an Alu-like repeat, a variable number of GC-rich tandem repeats (VNTRs), a short interspersed nuclear element region (SINE), and a poly-A tail [43]. The insertion showed hallmarks typical for the L1-mediated target-primed reverse transcription (TPRT) system as it (1) occurs in a DNA site resembling the L1 endonuclease consensus (for instance, 5'–CTTT/A–3'), (2) ends with a long poly-A stretch, and (3) is flanked by a 10 bp target-site duplication (TSD) present at both sites of the insertion that originates from the native *SMARCB1* sequence at the insertion site [44–46] (Figure 3 and supplementary material, Figure S3A,B). SVA elements are classified into six subfamilies (A to F) based on the analysis of their full-length SINE region [43]. Analysis of the SINE region of the identified SVA insertion, using the Repeat Masker Web server, revealed strongest resemblance to SVA-E, one of the youngest and most active human-specific subfamilies.

To confirm that the mother was indeed a mosaic carrier of this SVA-E insertion, we developed unique breakpoint-spanning PCRs for each of the insertion

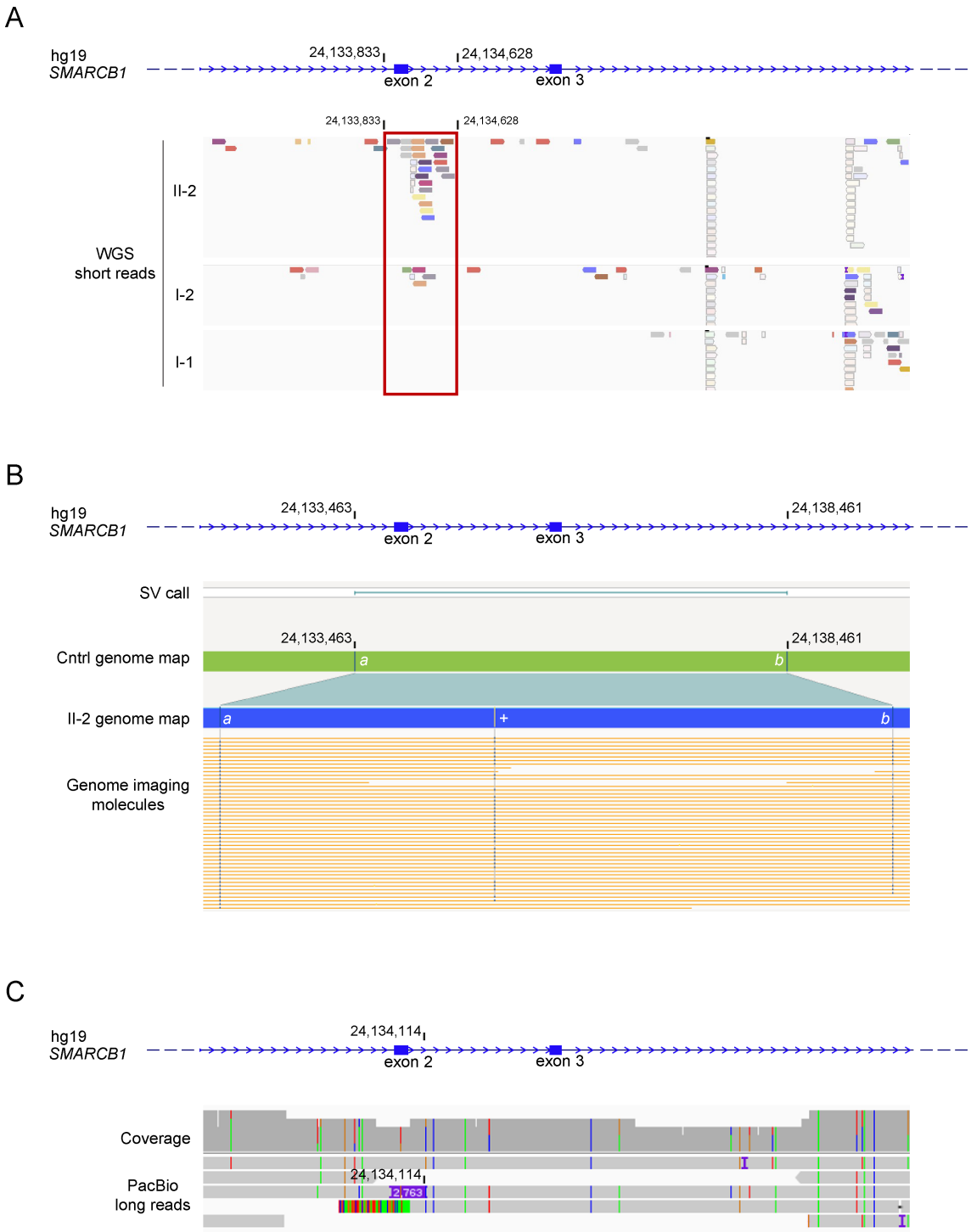

Figure 2. Structural variant identified by optical genome mapping and PacBio HiFi long-read sequencing. Schematic representation of the *SMARCB1* locus enlarged at position chr22: 24,130,251–24,139,993 (in blue). (A) WGS analysis of germline DNA from sibling II-2, mother (I-2), and father (I-1). The red rectangle highlights the heterozygous discordant reads mapped in the region chr22: 24,133,833–24,134,628 of the second sibling. Discordant reads mapping was also observed in the mother, albeit at a lower level, but not in the father. The different colors of the discordant reads indicate that they partially match with other chromosomal regions; the origin of this could, however, not be revealed. (B) OGM analysis of germline DNA from sibling II-2 showing the SV call enclosing the insertion at the *SMARCB1* locus. '*a*' and '*b*' mark the specific 6-nucleotide consensus sequences surrounding the SVs that are recognized by the labels (genomic positions 24,133,463 and 24,138,461) as called in a control DNA sample (Cntrl) and germline DNA from II-2. The blue bar represents the mutated *SMARCB1* allele with the ~2.8 kb insertion introducing an extra consensus sequence ('+') in II-2. Individual molecules supporting the allele with insertion are indicated by yellow lines. (C) PacBio HiFi sequencing analysis showing the coverage and the alignment of the long-read sequencing of germline DNA from sibling II-2. The alignment shows two reads representing the alternative allele: one read running into the insertion of 2,763 nucleotides (in purple) identified at genomic position 24,134,114 and one read spanning the entire insertion. The two other reads represent the wild-type allele in the germline DNA from II-2.

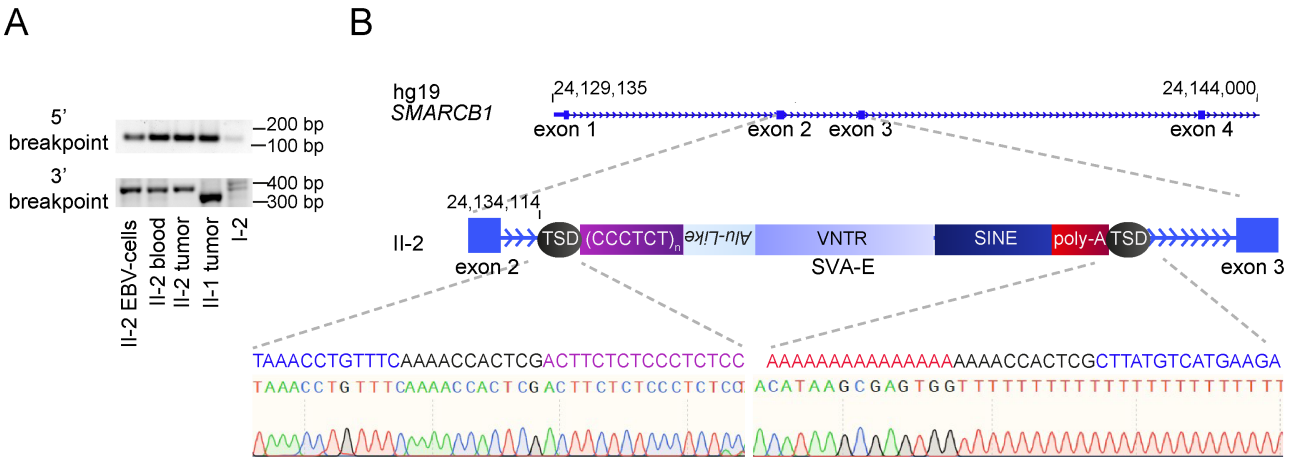

Figure 3. Molecular validation of the SVA retrotransposon element insertion at the *SMARCB1* locus. (A) Amplification of the *SMARCB1* regions spanning the PacBio HiFi-predicted 5' and 3' breakpoints in germline DNA (extracted from EBV-transformed lymphoblast – indicated as EBV-cells – and blood) and tumor DNA from the second sibling (II-2), tumor DNA from the first sibling (II-1), and germline DNA from the mother (I-2). All amplifications gave products of the size predicted by PacBio HiFi analysis (supplementary material, Table S2). Only the PCR spanning the 3' breakpoint in tumor II-1 gave a slightly smaller product. This suggests the presence of a shorter poly-A, likely due to genetic instability of the SVA-E. The primers used are listed in supplementary material, Table S2. (B) Schematic representation of the region spanning from exon 2 to exon 3 of the *SMARCB1* locus as in the hg19 reference genome and in sibling II-2. The colored blocks indicate the SVA-E insertion at position 24,134,114 in intron 2. The bottom panel displays the 5' and 3' insertion breakpoints as validated by Sanger sequencing (chromatogram): *SMARCB1* sequence in blue; TSD in black; beginning of the insertion in purple; poly-A in red. The 5' breakpoint was sequenced using the forward primer used for the amplification. The 3' breakpoint was sequenced using the reverse primer used for the amplification.

breakpoints. Both breakpoint regions could be amplified from tumor DNA of both siblings and from peripheral blood-derived DNA of sibling 2 (II-2) as well as the mother (I-2) but not in a control sample (Figure 3A and supplementary material, Figure S3A), indicating that the mother indeed carries the insertion in blood. Sanger sequencing confirmed the breakpoint-spanning sequences as identified by PacBio HiFi sequencing (Figure 3B and supplementary material, Figure S3B). Moreover, a dilution series of the breakpoint-spanning PCRs in blood-derived DNA from the mother and sibling II-2 indicated that amplification levels were lower in the mother, indeed suggesting that she is a mosaic carrier of the SVA-E insertion (supplementary material, Figure S3).

Next, we aimed to reveal how this SVA-E insertion leads to aberrant splicing. Previous studies have shown that SVA elements carry multiple cryptic splice acceptor sites [47]. To verify whether the SVA-E insertion in intron 2 interferes with normal splicing between exons 2 and 3, we amplified cDNA of II-2 using primers targeting the first three exons of *SMARCB1* and different regions of the SVA-E insertion. Our results indeed verified the presence of splicing from exon 2 into the Alu-like domain of SVA-E, one of the previously recognized splice acceptor sites (Figure 4). This finding indicates that insertion of the SVA-E leads to aberrant splicing of the *SMARCB1* transcript, thereby causing disruption of the allele.

Discussion

We present a case of two siblings with congenital ATRT due to a germline SVA-E retrotransposon insertion into intron 2 of the *SMARCB1* allele. The insertion severely impairs normal splicing, resulting in functional loss of the allele. Because of the intronic location and the repeat-rich nature of the insertion, this germline abnormality was missed with conventional genetic procedures after diagnosis of the first child. Consequently, the high recurrence risk in the second pregnancy was unfortunately not recognized. Identification of the underlying genomic aberration now makes genetic testing possible, which provides options for the parents to consider a new pregnancy. Furthermore, this finding illustrates how standard procedures generally applied in clinical genetics, including WGS, can fail in identifying disease gene mutations.

SVA retrotransposons are evolutionary young and active non-long terminal repeat elements that are expanding in the human genome with an estimated mobilization rate of 1:916 births [48]. Integration of SVAs in the genome occurs through an RNA intermediate using the TPRT system, mediated *in trans* by the L1 retrotransposon protein machinery [49,50]. Although the exact mechanism of an SVA insertion is not completely understood, it is clear that such structural variants, including other transposing events, can have negative effects on genomic integrity and can cause disease, including cancer [51,52]. The insertion of an SVA into a gene locus introduces several functional units enclosed in its sequence (such as splice acceptor/donor sites, SP1 binding elements, and hormone-responsive elements) that can cause gene disruption and deregulate gene expression at DNA, RNA or epigenetic level, leading to disease [43,47,53–57]. We have shown that, in the case presented in this study, an SVA-E insertion introduces splice acceptor sites that mediate aberrant splicing,

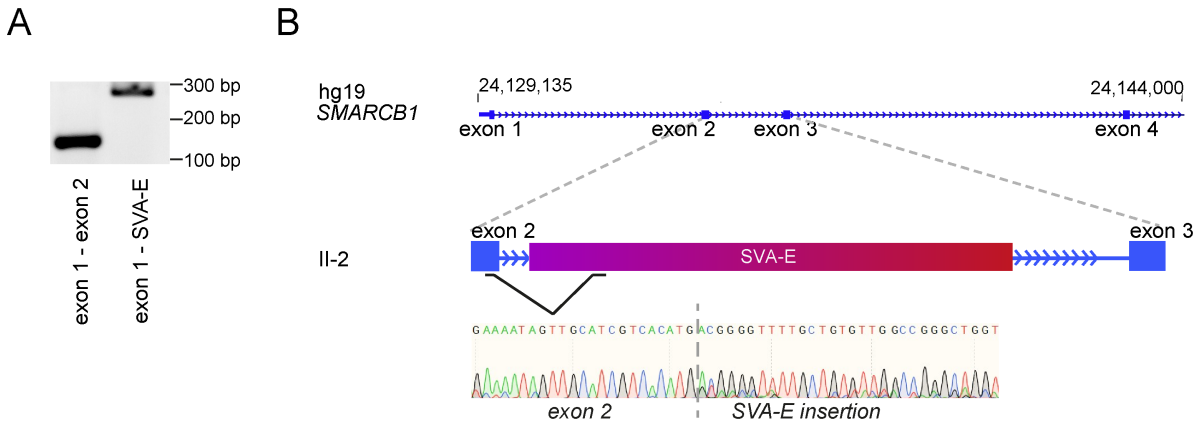

Figure 4. Analysis of the effect of the SVA-E insertion on *SMARCB1* transcription. (A) Transcript analysis of the cDNA of II-2 using primers targeting exon 1 and exon 2 or the SVA-E insertion. Amplification using primers targeting exon 1 and SVA-E showed a clear band indicating the presence of a transcript including (part of) the SVA-E insertion. The primers used are listed in supplementary material, Table S2. (B) Schematic representation of the *SMARCB1* locus with an enlarged view of the region spanning from exon 2 to exon 3 of II-2. The black line indicates the alternative splicing between a premature splicing donor site in exon 2 and a splicing acceptor site in the SVA-E insertion as demonstrated by Sanger sequencing (chromatogram) of the exon 1–SVA-E product shown in A.

preventing normal splicing between exons 2 and 3. We speculate that the exon 2 skipping, which was observed in an earlier stage and results in a truncated open reading frame, is an indirect consequence of the dysregulated splicing process in intron 2.

The congenital ATRT in two siblings implies a heritable mutation, and mosaicism in one of the parents was therefore suspected. Indeed, results from WGS and PCR-based validations strongly suggested that the mother (I-2) is a mosaic carrier of the SVA-E insertion, which will have decreased the risk of developing the disease herself considerably. Only a few cases of parental mosaicism for *SMARCB1* mutations have been described in ATRT. Of note, half of the known unaffected *SMARCB1* mutation carriers carry a splice-site mutation [2] that may lead to only partial loss of function, thus reducing the penetrance for ATRT. Based on the congenital presentation of ATRT in the two siblings, the SVA-E insertion appears to result in complete disruption of the allele and high penetrance, which may only be compatible with mosaicism in unaffected carriers.

The fact that both children in this family inherited the SVA-E insertion suggests that the level of mosaicism of this aberration in the oocytes of the mother may be considerable, thus creating a high risk of recurrence in future pregnancies. Before the SVA-E insertion was discovered, we had already selected SNPs in and around the *SMARCB1* locus that were associated with the affected allele to enable prenatal or preimplantation genetic testing. However, in the case of mosaicism, it is important to specifically test for the presence of the causative aberration itself. Therefore, the identification of the SVA-E insertion as the causative germline aberration and the ability to test its presence using breakpoint-spanning primers is extremely relevant and avoids uncertainties in a test result.

Our findings should be an admonishment for pathologists, oncologists, and geneticists to realize that despitea negative family history and standard genetic test results, tumor data can still signify germline cancer predisposition. Indeed, the lack of a second hit or unexplained loss of protein expression in the tumor should push towards an analysis with deeper and more comprehensive approaches. Rearrangements involving repetitive regions, such as insertions of transposable elements, might be camouflaged in the genome and be missed with short-read sequencing. In addition to the case described here, transposable element insertions have been previously reported in MRT and ATRT cases [58,59]. RNA sequencing is likely to reveal aberrant transcripts that arise as a consequence of these insertions, but the exact genomic integration sites may remain undefined when transcripts splice into the inserted sequence, as we observed in the two siblings. Conversely, we show how, in such cases, OGM can be of highly additive value to clinical practice, because this methodology is independent of local sequence composition and quite accurately maps the rearrangement site. Furthermore, in most of the cases, subsequent PCR and Sanger sequencing validations are enough to support OGM and identify exact breakpoints and sequence of the aberration [33,34], making the approach highly cost-effective, offering added value to the gold-standard pathological and molecular genetic diagnosis.

In conclusion, we have described a case of an SVA-E retrotransposon element insertion in *SMARCB1* causing ATRT in a family with two affected siblings. We have shown the power of OGM to identify SVs refractory to detection with techniques commonly used in diagnostics. Wider implementation of this new technique will support the identification of causative SVs in other unsolved cases, enabling genetic counseling and family planning.

Acknowledgements

We are grateful to EJ Kamping for performing SNP arrays on the tumor samples and LT van der Veken for SNP

array analysis and interpretation. This research was funded by Stichting Kinderen Kankervrij (KiKa) and by NWO (Nederlandse organisatie voor Wetenschappelijk Onderzoek, the Dutch organisation of Scientific Research; project 184.034.019) as part of The Netherlands X-omics Initiative. TM was supported by the Sigrid Jusélius Foundation.

Author contributions statement

MS, TM, EW, KN, RD, MK, MTM, MN and JAB performed experiments. FvD, JYHK, LOG, TM and KN performed bioinformatics analysis. ARM, CEMG, BK, PW and MCJ performed clinical and pathological analysis. AH and RPK designed the study. MS, TM, KN and RPK wrote the manuscript. All the authors read and approved the final manuscript.

References

1. Haas JE, Palmer NF, Weinberg AG, *et al*. Ultrastructure of malignant rhabdoid tumor of the kidney. A distinctive renal tumor of children. *Hum Pathol* 1981; **12:** 646–657.
2. Masliah-Planchon J, Bièche I, Guinebretière JM, *et al*. SWI/SNF chromatin remodeling and human malignancies. *Annu Rev Pathol* 2015; **10:** 145–171.
3. Tsuneyoshi M, Daimaru Y, Hashimoto H, *et al*. Malignant soft tissue neoplasms with the histologic features of renal rhabdoid tumors: an ultrastructural and immunohistochemical study. *Hum Pathol* 1985; **16:** 1235–1242.
4. Trobaugh-Lotrario AD, Tomlinson GE, Finegold MJ, *et al*. Small cell undifferentiated variant of hepatoblastoma: adverse clinical and molecular features similar to rhabdoid tumors. *Pediatr Blood Cancer* 2009; **52:** 328–334.
5. Rorke LB, Packer RJ, Biegel JA. Central nervous system atypical teratoid/rhabdoid tumors of infancy and childhood: definition of an entity. *J Neurosurg* 1996; **85:** 56–65.
6. Mitchell SG, Pencheva B, Porter CC. Germline genetics and childhood cancer: emerging cancer predisposition syndromes and psychosocial impacts. *Curr Oncol Rep* 2019; **21:** 85.
7. Blessing MM, Alexandrescu S. Embryonal tumors of the central nervous system: an update. *Surg Pathol Clin* 2020; **13:** 235–247.
8. Nesvick CL, Nageswara Rao AA, Raghunathan A, *et al*. Case-based review: atypical teratoid/rhabdoid tumor. *Neurooncol Pract* 2019; **6:** 163–178.
9. Johann PD, Erkek S, Zapatka M, *et al*. Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. *Cancer Cell* 2016; **29:** 379–393.
10. Torchia J, Golbourn B, Feng S, *et al*. Integrated (epi)-genomic analyses identify subgroup-specific therapeutic targets in CNS rhabdoid tumors. *Cancer Cell* 2016; **30:** 891–908.
11. Versteege I, Sévenet N, Lange J, *et al*. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* 1998; **394:** 203–206.
12. Hasselblatt M, Nagel I, Oyen F, *et al*. *SMARCA4*-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. *Acta Neuropathol* 2014; **128:** 453–456.
13. Kalimuthu SN, Chetty R. Gene of the month: *SMARCB1*. *J Clin Pathol* 2016; **69:** 484–489.
14. Biegel JA, Zhou JY, Rorke LB, *et al*. Germ-line and acquired mutations of *INI1* in atypical teratoid and rhabdoid tumors. *Cancer Res* 1999; **59:** 74–79.
15. Sévenet N, Sheridan E, Amram D, *et al*. Constitutional mutations of the *hSNF5/INI1* gene predispose to a variety of cancers. *Am J Hum Genet* 1999; **65:** 1342–1348.
16. Eaton KW, Tooke LS, Wainwright LM, *et al*. Spectrum of *SMARCB1/INI1* mutations in familial and sporadic rhabdoid tumors. *Pediatr Blood Cancer* 2011; **56:** 7–15.
17. Plon SE, Lupo PJ. Genetic predisposition to childhood cancer in the genomic era. *Annu Rev Genomics Hum Genet* 2019; **20:** 241–263.
18. Bourdeaut F, Lequin D, Brugières L, *et al*. Frequent *hSNF5/INI1* germline mutations in patients with rhabdoid tumor. *Clin Cancer Res* 2011; **17:** 31–38.
19. Janson K, Nedzi LA, David O, *et al*. Predisposition to atypical teratoid/rhabdoid tumor due to an inherited *INI1* mutation. *Pediatr Blood Cancer* 2006; **47:** 279–284.
20. Bruggers CS, Bleyl SB, Pysher T, *et al*. Clinicopathologic comparison of familial versus sporadic atypical teratoid/rhabdoid tumors (AT/RT) of the central nervous system. *Pediatr Blood Cancer* 2011; **56:** 1026–1031.
21. Ammerlaan ACJ, Ararou A, Houben MPWA, *et al*. Long-term survival and transmission of *INI1*-mutation via nonpenetrant males in a family with rhabdoid tumour predisposition syndrome. *Br J Cancer* 2008; **98:** 474–479.
22. Taylor MD, Gokgoz N, Andrulis IL, *et al*. Familial posterior fossa brain tumors of infancy secondary to germline mutation of the *hSNF5* gene. *Am J Hum Genet* 2000; **66:** 1403–1406.
23. Tauziède-Espariat A, Masliah-Planchon J, Brugières L, *et al*. Deep intronic hotspot variant explaining rhabdoid tumor predisposition syndrome in two patients with atypical teratoid and rhabdoid tumor. *Eur J Hum Genet* 2017; **25:** 1170–1172.
24. Li H, Handsaker B, Wysoker A, *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; **25:** 2078–2079.
25. McKenna A, Hanna M, Banks E, *et al*. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20:** 1297–1303.
26. Van der Auwera GA, Carneiro MO, Hartl C, *et al*. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; **43:** 11.10.1–11.10.33.
27. Priestley P, Baber J, Lolkema M, *et al*. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* 2019; **575:** 210–216.
28. Layer RM, Chiang C, Quinlan AR, *et al*. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* 2014; **15:** R84.
29. Rausch T, Zichner T, Schlattl A, *et al*. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 2012; **28:** i333–i339.
30. Thung DT, de Ligt J, Vissers LE, *et al*. Mobster: accurate detection of mobile element insertions in next generation sequencing data. *Genome Biol* 2014; **15:** 488.
31. Pouwer AFW, van Den Einden LCG, van der Linden M, *et al*. Clonal relationship between lichen sclerosus, differentiated vulvar intraepithelial neoplasia and non HPV-related vulvar squamous cell carcinoma. *Cancer Genomics Proteomics* 2020; **17:** 151–160.
32. Kalpana GV, Marmon S, Wang W, *et al*. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* 1994; **266:** 2002–2006.
33. Neveling K, Mantere T, Vermeulen S, *et al*. Next generation cytogenetics: comprehensive assessment of 48 leukemia genomes by genome imaging. *bioRxiv* 2020. https://doi.org/10.1101/2020.02.06.935742 [Not peer reviewed].
34. Mantere T, Neveling K, Pebrel-Richard C, *et al*. Next generation cytogenetics: genome-imaging enables comprehensive structural variant detection for 100 constitutional chromosomal aberrations in 85 samples. *bioRxiv* 2020. https://doi.org/10.1101/2020.07.15.205245 [Not peer reviewed].

35. Firth HV, Richards SM, Bevan AP, *et al*. DECIPHER: *D*atabas*e* of *C*hromosomal *I*mbalance and *P*henotype in *H*umans Using *E*nsembl *R*esources. *Am J Hum Genet* 2009; **84:** 524–533.
36. The Genome of the Netherlands Consortium; Francioli LC, Menelaou A, Pulit SL, et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 2014; **46:** 818–825.
37. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, *et al*. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467:** 1061–1073.
38. 1000 Genomes Project Consortium, Abecasis GR, Auton A, *et al*. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491:** 56–65.
39. Lek M, Karczewski KJ, Minikel EV, *et al*. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536:** 285–291.
40. Barseghyan H, Tang W, Wang RT, *et al*. Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. *Genome Med* 2017; **9:** 90.
41. Wenger AM, Peluso P, Rowell WJ, *et al*. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 2019; **37:** 1155–1162.
42. Mizuguchi T, Suzuki T, Abe C, *et al*. A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing. *J Hum Genet* 2019; **64:** 359–368.
43. Wang H, Xing J, Grover D, *et al*. SVA elements: a hominid-specific retroposon family. *J Mol Biol* 2005; **354:** 994–1007.
44. Morrish TA, Gilbert N, Myers JS, *et al*. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* 2002; **31:** 159–165.
45. Hancks DC, Kazazian HH Jr. Roles for retrotransposon insertions in human disease. *Mob DNA* 2016; **7:** 9.
46. Stacey SN, Kehr B, Gudmundsson J, *et al*. Insertion of an SVA-E retrotransposon into the *CASP8* gene is associated with protection against prostate cancer. *Hum Mol Genet* 2016; **25:** 1008–1018.
47. Damert A, Raiz J, Horn AV, *et al*. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res* 2009; **19:** 1992–2008.
48. Vogt J, Bengesser K, Claes KB, *et al*. SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. *Genome Biol* 2014; **15:** R80.
49. Raiz J, Damert A, Chira S, *et al*. The non-autonomous retrotransposon SVA is *trans*-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res* 2012; **40:** 1666–1683.
50. Hancks DC, Goodier JL, Mandal PK, *et al*. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet* 2011; **20:** 3386–3400.
51. Takasu M, Hayashi R, Maruya E, *et al*. Deletion of entire *HLA-A* gene accompanied by an insertion of a retrotransposon. *Tissue Antigens* 2007; **70:** 144–150.
52. Lee E, Iskow R, Yang L, *et al*. Landscape of somatic retrotransposition in human cancers. *Science* 2012; **337:** 967–971.
53. Hancks DC, Kazazian HH Jr. SVA retrotransposons: evolution and genetic instability. *Semin Cancer Biol* 2010; **20:** 234–245.
54. Wilund KR, Yi M, Campagna F, *et al*. Molecular mechanisms of autosomal recessive hypercholesterolemia. *Hum Mol Genet* 2002; **11:** 3019–3030.
55. Mendell JT, Dietz HC. When the message goes awry: disease-producing mutations that influence mRNA content and performance. *Cell* 2001; **107:** 411–414.
56. Hassoun H, Coetzer TL, Vassiliadis JN, *et al*. A novel mobile element inserted in the alpha spectrin gene: spectrin dayton. A truncated alpha spectrin associated with hereditary elliptocytosis. *J Clin Invest* 1994; **94:** 643–648.
57. Makino S, Kaji R, Ando S, *et al*. Reduced neuron-specific expression of the *TAF1* gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet* 2007; **80:** 393–406.
58. Henssen AG, Koche R, Zhuang J, *et al*. PGBD5 promotes site-specific oncogenic mutations in human tumors. *Nat Genet* 2017; **49:** 1005–1014.
59. Thomas C, Oehl-Huber K, Bens S, *et al*. Transposable element insertion as a mechanism of *SMARCB1* inactivation in atypical teratoid/rhabdoid tumor. *Genes Chromosomes Cancer* 2021; **60:** 586–590.

SUPPLEMENTARY MATERIAL ONLINE

Case description

Figure S1. Radiological imaging of siblings II-1 (9 days old) and II-2 (1 day old)

Figure S2. Circos plot of the rare SVs identified with genome-wide OGM

Figure S3. Molecular validation of the SVA-E retrotransposon element insertion at the *SMARCB1* locus

Table S1. Genotype analysis of the tumors showing homozygosity for the maternal allele in the tumors of both children

Table S2. Primers used for the *SMARCB1*–SVA-E breakpoint-spanning PCRs and investigation of the *SMARCB1*–SVA-E transcript