npj | genomic medicine

Case report

Published in partnership with CEGMR, King Abdulaziz University

https://doi.org/10.1038/s41525-025-00517-0

A founder *BRCA1* exonic duplication involving breakpoint in T2T reference genome-specific region results in constitutional fusion transcript

Mathias Schwartz[1,2,3] ✉, Mathilde Filser[1,3], Kevin Merchadou[3,4], Elisa Lemaitre[1,3], Khadija Abidallah[1,3], Henrique Tenreiro[1,3], Catherine Dubois D'enghien[1,3], Audrey Rapinat[3,5], Elise Pierre-Noel[1,3], Voreak Suybeng[1,3], Marion Espenel[3,6], Sylvain Baulande[3,6], Séverine Adams[1,3], Audrey Remenieras[7], Crystal Renaud[8,9], Camille Aucouturier[8,9], Capucine Delnatte[10], Céline Garrec[10], Victor Renault[3,4], Lisa Golmard[1,3], Emmanuelle Fourme[1,3], Julien Masliah-Planchon[1,3,11] & Sandrine M. Caputo[1,3,11]

Pathogenicity assessment of genetic variants is the cornerstone of genetic counselling. Copy gains of exons are challenging, as pathogenicity depends on the localization of the additional exons. Eight patients form six families carried copy gains of *BRCA1* exons 8–20. For appropriate characterization, long-read sequencing aligned on three distinct reference genome assemblies, optical genomic mapping, short-read and long-read RNA sequencing were performed. All patients shared the same pathogenic structural variant, involving a large segment located downstream in the genome. One breakpoint occurred in a region incorrectly annotated in GRCh37/hg19 and GRCh38/hg38. Alignment to the T2T-CHM13/hs1 assembly was therefore necessary for accurate characterization. This rearrangement caused various *BRCA1* transcriptomic abnormalities: back-splicing, forward genomic strand transcription by insertion of an ectopic promoter, fusion transcripts with the "Next to *BRCA1*" gene 1 (*NBR1*). Our findings underscore the need to combine advanced technologies with the latest genome references to resolve complex rearrangements with significant medical implications.

Constitutional pathogenic genetic variants in the *BRCA1* gene confer a lifetime risk of breast and ovarian cancer of approximately 70% and 45%, respectively[1]. In tumour cells, *BRCA1* inactivation leads to homologous recombination deficiency that can be targeted therapeutically with PARP inhibitors (PARPi)[2]. Identifying *BRCA1* pathogenic variants is therefore of major importance to guide genetic counselling and therapeutic strategies. However, many genetic variants remain of uncertain significance, with insufficient evidence to be classified as benign (with no clinical consequences) or pathogenic (leading to cancer predisposition and PARPi sensitivity)[3,4]. Copy gains of one or several exons are challenging, as pathogenicity depends on the frame, size, and localization of the additional exons[5–9]. As most duplications are intragenic, either deep-intronic DNA sequencing or RNA sequencing (RNAseq) usually allows the precise succession of exons to be determined[6,8]. In most cases, duplicated exons are in tandem[6]. However, the *BRCA1* sequence can also be disrupted by more complex structural variants (SVs) that cannot be properly described by standard short-read sequencing[10]. Long-read sequencing with Nanopore adaptive sampling has been shown to be a useful approach for accurate description of cancer-predisposing SVs[8,9].

[1]Department of Genetics, Institut Curie, Paris, France. [2]Dynamics of epigenetic plasticity in cancer, UMR3244, Institut Curie, Paris, France. [3]Paris Sciences Lettres (PSL) Research University, Paris, France. [4]Clinical Bioinformatics Unit, Institut Curie, Paris, France. [5]Genomics Core Facility, Department of Translational Research, Institut Curie, Paris, France. [6]Institut Curie Genomics of Excellence (ICGex) Next-Generation Sequencing Core Facility, Institut Curie, Paris, France. [7]Laboratoire d'Oncogénétique Moléculaire, Département de Biologie du Cancer, Institut Paoli Calmettes, Marseille, France. [8]Laboratoire de biologie et de génétique du cancer, Centre François Baclesse, Caen, France. [9]Inserm U1245, Cancer and Brain Genomics, Rouen, France. [10]Laboratoire de Génétique Moléculaire, Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) de Nantes, Nantes, France. [11]These authors contributed equally: Julien Masliah-Planchon, Sandrine M. Caputo. ✉e-mail: mathias.schwartz@curie.fr

Description of genetic variants always relies on comparison between patient sequence and a reference genome assembly. The Genome Reference Consortium (GRC) released 14 versions of the GRCh37/hg19 reference genome assembly between 2009 and 2013, and 15 versions of the GRCh38/hg38 assembly between 2013 and 2022. These assemblies are widely used in medical genetic practice. In 2022, the Telomere-to-Telomere (T2T) Consortium released the first complete human genome assembly (T2T-CHM13/hs1) using long-read sequencing technologies[11]. This assembly includes an unprecedented description of complex regions such as centromeres (6% of the human genome) and segmental duplications (7% of the human genome)[12,13]. These latter are of particular interest as they promote genomic rearrangements that may have clinical consequences[13]. However, in actual clinical practice, no cases of T2T "*Homo Sapiens* 1" reference genome input have been reported.

We describe a founder pathogenic SV in six distinct French families, resulting in duplication of exons 8 to 20 of the *BRCA1* gene and cancer predisposition in carriers. This event involved a highly complex region, located downstream in the genome (approximately 110 kb away from *BRCA1*) and incorrectly annotated in genome assemblies prior to T2T-CHM13/hs1. Characterization and pathogenicity assessment of this SV was therefore impossible when using GRCh37/hg19 or GRCh38/hg38 assemblies as reference genomes. This SV provoked various *BRCA1* transcriptomic abnormalities.

Results

Patients

Eight patients from six unrelated families (numbered F1 to F6) carried a gain of copy of *BRCA1* exons 8 to 20 identified in routine diagnostic setting. Extended analysis was performed in Family 1 – Patient 2 (F1-P2). She was an unaffected young woman (<30 years-old) carrying the exonic copy gain of *BRCA1* previously identified in her grandmother (F1-P1) diagnosed with early-onset ovarian cancer.

Genomic characterization

Long-read sequencing in F1-P2, revealed fusion reads between *BRCA1* intron 20 and *NBR1* gene (breakpoint 1 [BP1]), and between *BRCA1* intron 7 and *CCDC200* gene (breakpoint 2 [BP2]) (Fig. 1). However, the orientations of BP1 and BP2 were inconsistent relative to each other when reads were aligned on GRCh37/hg19 or GRCh38/hg38 reference genomes (Fig. 1A). It was therefore impossible to localize the supplementary exons and to assess the pathogenicity of this SV. This was resolved when aligning on the T2T-CHM13v2.0/hs1 reference genome: both BP1 (involving *BRCA1* intron 20) and BP2 (involving *BRCA1* intron 7) orientations were consistent with a large intragenic insertion between *BRCA1* duplicated exons (Fig. 1B). At both breakpoints, this insertion matched the reversed sequence of a large genomic segment containing *NBR1* and *CCDC200* genes. This conclusion was confirmed by OGM which showed that the genomic segment containing *NBR1* and *CCDC200* genes was duplicated, inverted, and inserted between *BRCA1* duplicated exons (Fig. 1C). LiftOver annotations between T2T-CHM13v2.0/hs1 and previous reference genome assemblies showed that the complex region containing BP2 is mostly absent from GRCh37/hg19 and is inverted in GRCh38/hg38 (Fig. 1D). The BP1 and BP2 sequences were then confirmed using Sanger sequencing (Fig. S1-A).

We then compared these results to other patients carrying the copy gains of exons 8 to 20 in the *BRCA1* gene. All eight patients, from six families (Fig. S2), unrelated to each other, shared the same breakpoints and a common *BRCA1* haplotype (Fig. S1-B and S1-C). Clinical and histopathological family histories provided a combined likelihood ratio (LR) for pathogenicity of 5.96, cosegregation analysis provided a LR for pathogenicity of 326.33. The multifactorial combined LR was therefore 1943.83. Thus, with a prior probability of pathogenicity set to 50%, estimated posterior probability of pathogenicity was over 99.9% for the SV-carrying haplotype. With a prior probability of pathogenicity set to 10%, posterior odds reached 215.98, yielding an estimated posterior probability of pathogenicity over 99.5%.

Transcriptomic characterization

We additionally characterized the transcriptional consequences of this SV in F1-P2 (Figs. 2, 3). To this end, we analysed allelic ratios of 9 single nucleotide polymorphisms (SNPs) she carried on *BRCA1* exonic sequences (Fig. 2A). All those SNPs were benign, with minor allele frequencies ranging from 0.33 to 0.38 in the gnomAD database. Two of these SNPs were located in exon 24, after the duplicated exons. The allelic ratio for these two SNPs was approximately 50% in genomic DNA and 100% in RNA-seq, indicating that only one allele of the end of the *BRCA1* gene was transcribed (Fig. 2A). Seven other SNPs, located in the duplicated exons, had an allelic ratio of approximately 33% in genomic DNA, suggesting they were present only on the non-duplicated allele. The RNA-seq results showed an allelic ratio of 50%, indicating that only one set of duplicated exons was transcribed (Fig. 2A).

Moreover, short-read RNA-seq and RNA Sanger sequencing revealed an abnormal splicing junction between *BRCA1* exons 20 and 2 (Fig. 2B) suggestive of a back-splicing event, likely resulting in a circular RNA (circRNA), which was not observed in the 12 control samples.

The *BRCA1* gene is physiologically transcribed from the reverse strand. In F1-P2, strand-specific RNA-seq showed that exons 1 to 20 underwent a significant increase of forward-strand transcription (Fig. 2C). Long-read RNA-seq detected two reads linking *NBR1* antisense non-coding RNA with forward strand *BRCA1* sequences (Fig. 2D), suggesting that forward strand *BRCA1* transcription occurred from the *NBR1* bi-directional promoter.

Short-read RNA-seq also detected abnormal junctions between *BRCA1* exon 20 and *NBR1* exon 2 (both with physiological transcription orientation). Long-read RNA-seq confirmed the existence of *BRCA1*::*NBR1* fusion transcripts with various alternative splicing (Fig. 2D). Among 65 reads spanning *BRCA1* exon 20 in long-read RNA-seq, 21 reads (32%) supported fusion with *NBR1* exon1 (1 read), exon 2 (19 reads), or exon 3 (1 read) (Fig. 2D).

Altogether, RNA-seq revealed various *BRCA1* transcriptomic abnormalities caused by this complex SV: back-splicing, forward strand transcription, and fusion transcript with *NBR1* (Fig. 3). Mostly, analysing allelic ratios of several exonic SNPs demonstrated the mono-allelic expression of the final part of *BRCA1*, giving additional argument for the pathogenicity of this complex SV.

Discussion

We report a complex rearrangement of the *BRCA1* gene involving a duplication of exons 8–20 coupled with a duplicated and inverted insertion of a large segment, located downstream in the genome and containing *NBR1* and *CCDC200* genes (Fig. 3). This SV was identified in eight patients from six families and produced abnormal transcripts, leading to *BRCA1* loss of function and increased cancer risk. The accurate characterization of this SV was made possible using advanced molecular technologies, including long-read Nanopore sequencing, and the alignment of sequencing data to the latest T2T-CHM13/hs1 version of the human genome.

Genomic profiling technologies are continually evolving to offer more powerful tools for genome exploration. Nanopore sequencing is a particularly valuable innovation. Its unique method — real-time detection of ionic current changes as nucleic acids pass through a nano-scale pore — allows for the reading of long to ultra-long sequences. This capability makes Nanopore sequencing especially useful for resolving complex SVs, as it can accurately reconstruct these regions whereas short-read sequencing struggles with repetitive or complex sequences. Consequently, interest in long-read DNA sequencing technologies, including Nanopore sequencing, has grown in molecular biology, with their applications in routine practice expanding accordingly[14].

In this study, the use of long-read DNA sequencing was decisive for precise breakpoints characterization and ensure the absence of additional breakpoints. By confirming the global structure of the SV, OGM also

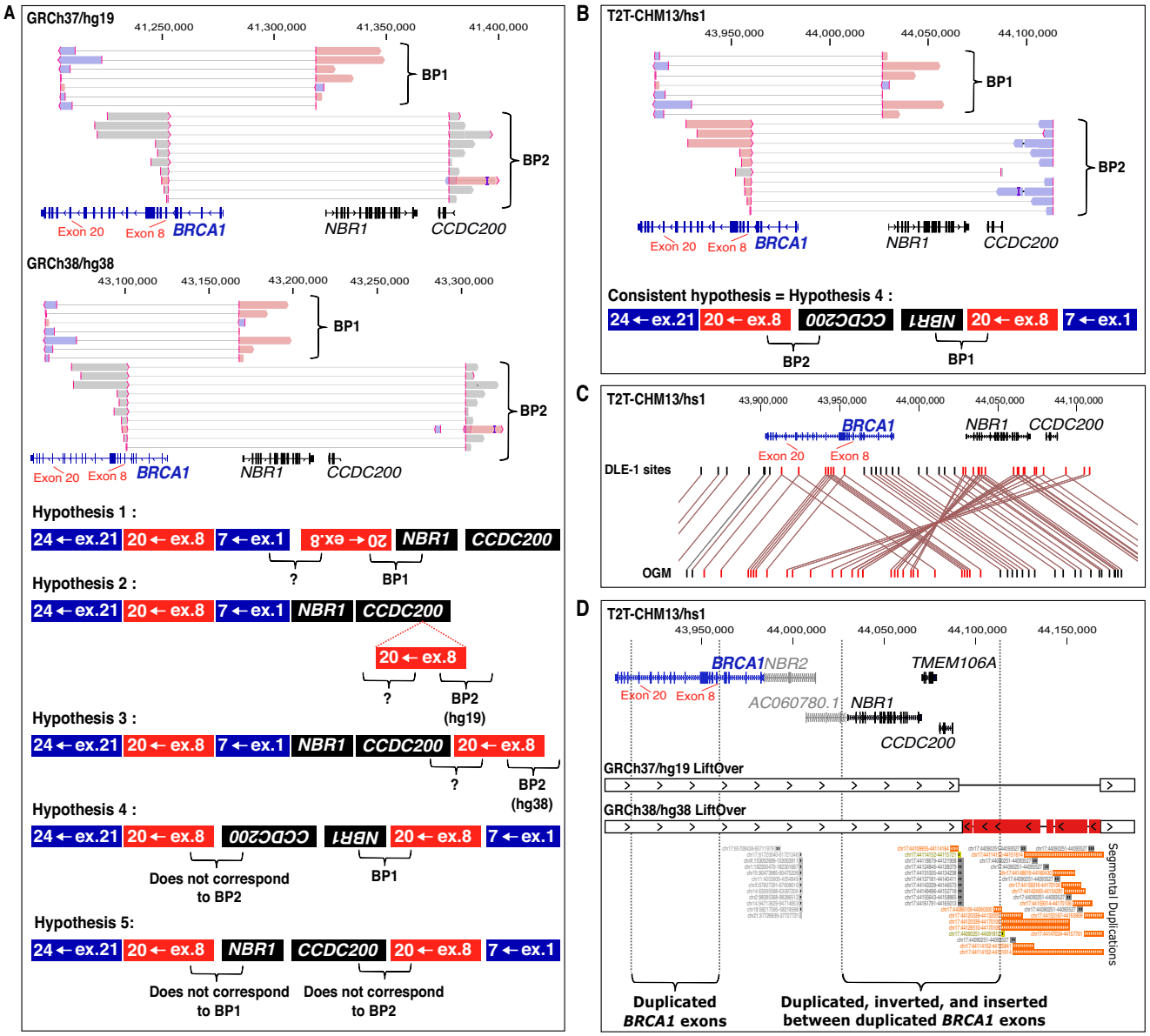

Fig. 1 | Genomic characterization in Family 1 – Patient 2 (F1-P2). A Long-read DNA sequencing aligned on GRCh37/hg19 and GRCh38/hg38 reference genome assemblies, and diagrams representing various structural hypothesis. Blue boxes represent non-duplicated *BRCA1* exons (ex.), red boxes represent *BRCA1* duplicated exons (ex.8 to ex.20). None of the hypothesis were totally consistent with long-read DNA sequencing data. BP: Breakpoint. **B** Long-read DNA sequencing aligned on T2T-CHM13/hs1 reference genome assembly, and diagram representing consistent structural hypothesis. **C** Optical Genomic Mapping (OGM) confirming general structure of the structural variant. **D** Rearranged region (T2T-CHM13/hs1) compared to GRCh37/hg19 and GRCh38/hg38 assemblies. Breakpoint 1 (BP1) occurred in *BRCA1* intron 20 and first intron of *NBR1* antisense transcript. Breakpoint 2 (BP2) occurred in *BRCA1* intron 7 and in a region containing many segmental duplications annotated by Vollger and colleagues[13]. LiftOver with prior reference genome assemblies showed this region is totally absent from GRCh37/hg19, and incomplete and inverted in GRCh38/hg38. Figures 1-A and 1-B were prepared from BAM file visualization in Integrative Genome Viewer (IGV) software[26]. Annotation data in 2D Figure 1-D were extracted from UCSC Genome Browser (http://genome.ucsc.edu)[27].

provided decisive insights. Then, Sanger sequencing allowed to easily assess that all patients from all families carried the same SV. The comparison between DNA-seq and RNA-seq supported the pathogenicity of this SV by showing that, on the recombined allele, the final *BRCA1* exons were not transcribed.

RNA-seq also revealed several unusual transcriptomic features. First, we discovered that this SV causes constitutional fusion transcripts between *BRCA1* and the neighbouring gene *NBR1*. Long-read RNA-seq enabled a detailed characterization of these fusion transcripts. More than 90% of detected *BRCA1*::*NBR1* fusion transcripts had a junction between *BRCA1* exon 20 and *NBR1* exon 2 (19 reads among 21 reads spanning both *BRCA1* and *NBR1*). *BRCA1* exon 20 ends with a full Lysine codon (AAG) while *NBR1* exon 2 starts with a short untranslated region (UTR) containing 9 nucleotides (CCTCACAGC) before the initiating Methionine codon (ATG). Therefore, it is likely that the majority of detected *BRCA1*::*NBR1* fusion transcripts encode an in-frame fusion protein containing: (1) amino-acids encoded by *BRCA1* from its initiating codon to the end of exon 20, then (2) three amino-acids corresponding to the final part of *NBR1* 5'UTR (Proline-Histidine-Serine), and finally (3) the full in-frame coding sequence of *NBR1* gene. This fusion protein could retain some function of the wild type NBR1 protein or, more hypothetically, gain additional function. However, it is very unlikely that it retains significant function of the wild type BRCA1 protein. Indeed, a single amino acid change in the BRCT domains of the terminal part of BRCA1 can be considered as pathogenic by the

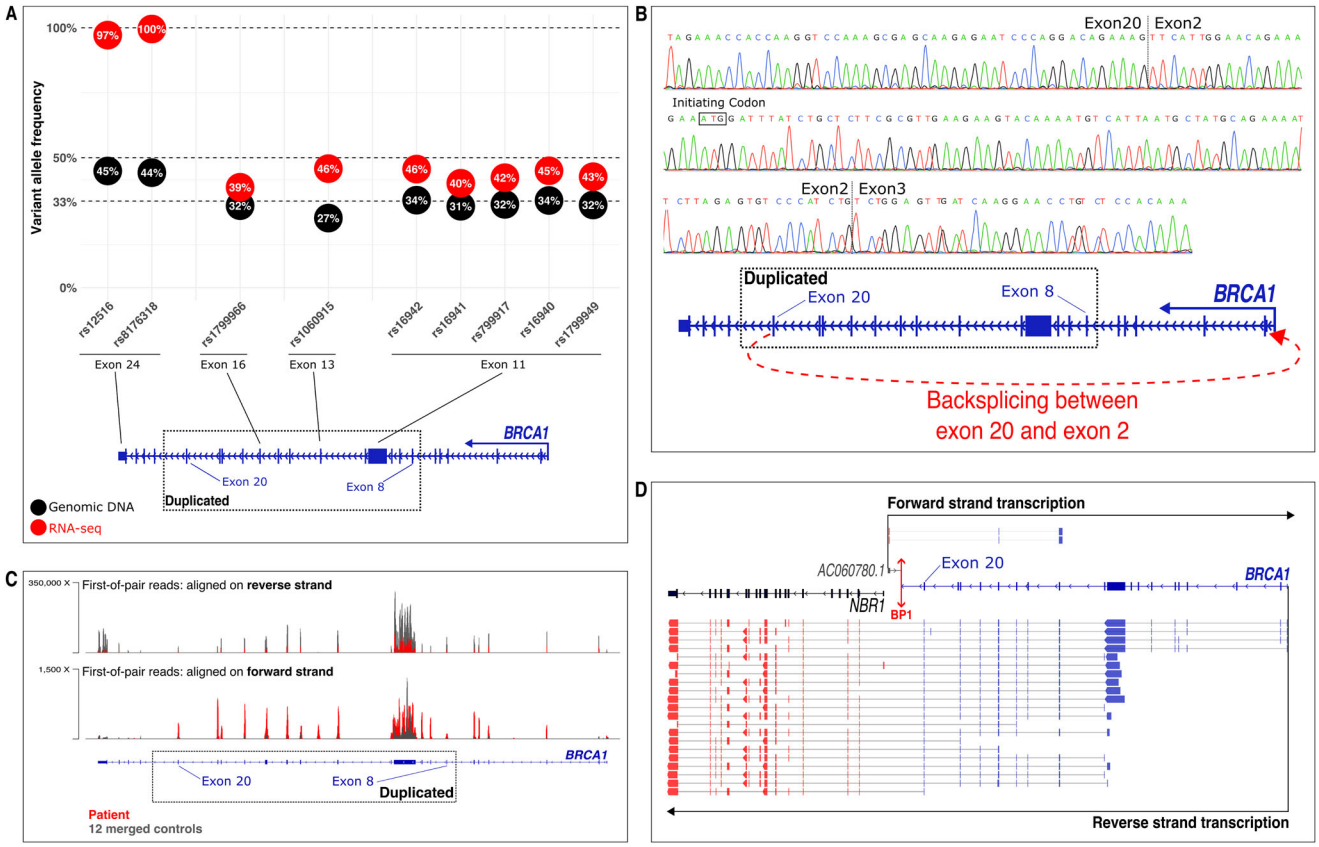

Fig. 2 | Transcriptomic characterization. A Variant allele frequencies of Single Nucleotide Polymorphisms (SNPs) carried in *BRCA1* by Family 1 – Patient 2 (F1-P2). Seven SNPs were detected in the duplicated region, all with an allelic ratio of approximately 33% in genomic DNA and approximately 50% in RNA-seq, indicating that only one set of duplicated exons was transcribed. Two SNPs were detected downstream of duplicated exons, with allelic ratios of approximately 50% in genomic DNA and approximately 100% in RNA-seq, indicating that only one allele of the end of the gene is transcribed. **B** Coding DNA (cDNA) Sanger sequencing showing back-splicing between *BRCA1* exon 20 and exon 2. **C** *BRCA1* strand-specific short-read RNA sequencing (RNA-seq) in Family 1 – Patient 2 (F1-P2) and 12 merged controls. Depth of coverage scales are indicated on the left. **D** Fusion reads involving *BRCA1* in long-read direct RNA sequencing (RNA-seq). Represented on rearranged breakpoint 1 (BP1). The two reads aligning on the forward strand of the rearranged region are shown at the top: both started in *NBR1* antisense transcript first exon (red) and included *BRCA1* exonic sequences (blue). Twenty-one reads aligned on the reverse strand of the rearranged region are shown at the bottom. They all started in *BRCA1* (blue) and continued with full *NBR1* gene (red) with various alternative splicing. Figures 2-C and 2-D were prepared from BAM file visualization in IGV software[26].

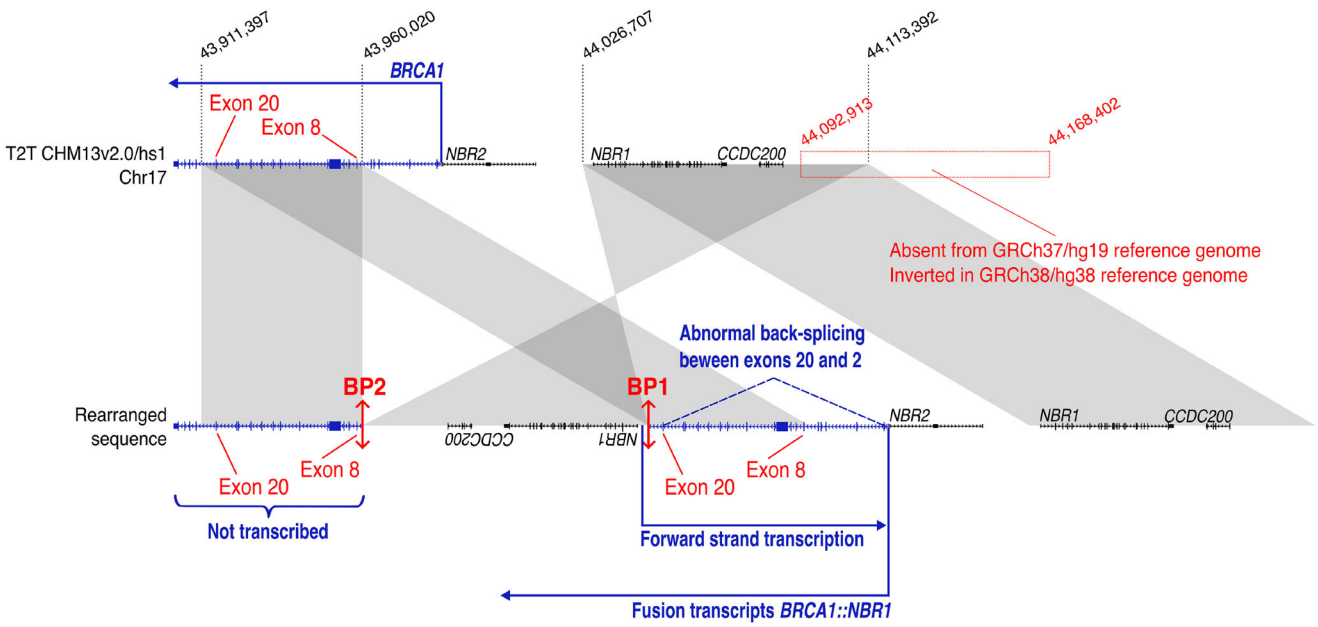

Fig. 3 | Founder *BRCA1* rearrangement and its transcriptomic consequences.

international ENIGMA consortium (for instance: c.5516 T > C / p.(Leu1839Ser), c.5513 T > A / p.(Val1838Glu), c.5509 T > C / p.(Trp1837Arg), c.5363 G > T / p.(Gly1788Val), *etc.*).

Second, strand-specific RNA-seq revealed a marked increase in forward-strand transcription at the *BRCA1* locus. Long-read RNA-seq linked this atypical transcriptional orientation to the bi-directional promoter of *NBR1*.

Finally, both short-read RNA-seq and Sanger RNA sequencing identified a back-splicing event between *BRCA1* exon 20 and exon 2, likely resulting in a circRNA[15,16]. As this class of long non-coding RNAs lacks polyadenylation (poly-A), the back-splicing event was not detected by long-read RNA sequencing, which was performed after poly-A capture. To our knowledge, this represents the first documented case of a SV inducing a back-splicing event. Globally, the integration of all techniques mentioned above underscores the importance of a multimodal approach in genomic research.

In routine clinical genetic testing, time and cost constraints, driven by the large number of patients and the need for rapid diagnostic answers, make it challenging to adopt such multimodal approach. Fortunately, the vast majority of exonic duplications can be resolved using more straightforward strategies, such as deep-intronic DNA sequencing or RNA sequencing alone. For more complex SVs, optimal strategy will depend on each case. In any case, databases and literature reviews can be decisive. For example, the SV reported here can now be readily identified in any patient through targeted Sanger sequencing of the breakpoints we have characterized. This underscores the importance of sharing such findings and experiences with the broader community. In our effort to describe this SV, the main challenge we encountered was the limited reliability of the GRCh37/hg19 and GRCh38/hg38 reference genome assemblies.

The Human Genome Project remains one of the greatest scientific achievements in the history of biology. This international project was launched in 1990 to comprehensively determine all of the base pairs composing the human genome. The commonly used GRCh37/hg19 and GRCh38/hg38 reference genomes still have limitations that can impair proper genetic diagnosis for genes located in complex regions[17]. Our work demonstrates that incorrectly annotated regions can also impair genetic diagnosis for distant genes located in genomic regions with no particular complexity, such as *BRCA1*. The T2T-CHM13/hs1 assembly provides a more complete description of challenging genomic regions[11–13]. This was crucial for accurate mapping of the structural variant reported here. Although limitations remain, which can impair, for example, the detection of alterations involving CFHR-Factor H cluster genes involved in complement disorders[18], our findings emphasize the need for the latest genome builds in clinical diagnostics.

Methods

Patients

Copy gain of *BRCA1* exons 8–20 were identified in routine diagnostic setting in each family by Next-Generation Sequencing performed in four distinct medical centres. Exonic copy gains were confirmed by Multiplex Ligation-dependant Probe Amplification (MLPA, MRC Holland probe mix P002-D1) on DNA extracted from blood.

All patients provided informed consent and were included in the COVAR (COsegregation of VARiants) study (NCT01689584), authorized by Ethics Committee in 2011 (Comité de protection des personnes Ile de France III, Am5677-1-2940). All procedures involving human participants were conducted in accordance with the Declaration of Helsinki and its amendments.

DNA and RNA extraction

Blood DNA was extracted with the QiaSymphony DSP DNA Midikit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

RNA was extracted from B lymphoblastoid cell lines established by in vitro infection with Epstein Barr Virus. Cells were treated by puromycin to inhibit nonsense mediated decay. After storage in 1 mL Trizol (Invitrogen, ref. 15596026), RNA was extracted using the standard chloroform/isopropanol procedure.

Family analysis

Haplotype determination was performed by amplifying five microsatellite regions surrounding *BRCA1* gene: D17S1327 downstream *BRCA1* in the genome (5'-mCTAAGGAGGTTTCTCTGGAC-3', 5'-TTCACAACTCAA GGTAAGATAGG-3'), D17S1323 in intron 12 (5'-mTAGGAGATGGATT ATTGGTG-3', 5'-AAGCAACTTTGCAATGAGTG-3'), D17S1322 in intron 19 (5'-mGCAGGAAGCAGGAATGGAAC-3', 5'-CTAGCCTGGG CAACAGAACGA-3'), D17S855 in intron 20 (5'-mACACAGACTTGTCC TACTGCC-3', 5'-GGATGGCCTTTTAGAAAGTGG-3') and D17S1185 upstream (5'-mGGTGACAGAACAAGACTCCATC-3', GGGCACTGC-TATGGTTTAGA-3'). PCR was performed with AmpliTaq GOLD DNA Polymerase according to the manufacturer's recommendations (Applied Biosystems, ref. 4311818) for 30 cycles (Hybridization: 55 °C). Amplified DNA (2µL) was then mixed with 0.5µL 500 LIZ dye size standard (Applied Biosystem, ref. 4322682) for fragment size determination by capillary electrophoresis.

As previously described[4], likelihood ratios (LR) for pathogenicity were computed from clinical and histopathological family histories[19,20], as well as on cosegregation analysis using a Bayesian statistical model described by Thompson et al. and updated by Belman et al.[21,22]. Probability for pathogenicity was computed using the multifactorial model defined by Goldgar et al.[23]. As prior probabilities of pathogenicity have not been calibrated for complex structural variants[3,24], we tested prior probabilities ranging from 0.5 (prior odds = 1, so the posterior odds equal the LR, making the posterior probability depend only on the LR[25]) to 0.1 (a conservative prior).

Short-read and long-read DNA sequencing (DNA-seq)

Short-read DNA-seq was performed on a NextSeq 500 (Illumina) after enrichment using a custom SureSelect QXT kit (Agilent), as described previously[8]. Mapping on GRCh37/hg19 was performed with Bowtie2.

Long-read DNA-seq was performed for the first run on a Oxford Nanopore Technologies Minion Flow Cell R9.4.1 (ref. FLO-MIN106D) after library preparation by manufacturer ligation kit (Oxford Nanopore Technologies, ref. SQK-LSK110) on 2 µg DNA, as described previously[8]. For the second run, a long-read DNA library was prepared using the new chemistry with the long sequencing kit SQK-LSK114 on 2 µg DNA, as per the supplier's recommendations. The DNA library was then injected into a Minion Flow Cell R10.4.1 (ref. FLO-MIN114). Computational enrichment was performed by adaptive sequencing (GRCh37/hg19). The first run targeted coding sequences of 120 genes including *BRCA1* (49 Mb)[8]. The second run targeted the whole long arm of chromosome 17 (84 Mb) including *BRCA1*. Bioinformatics analysis was performed with a custom NanoCliD pipeline (https://github.com/InstituteCurieClinicalBioinformatics/NanoCliD) including Minimap2 for alignment on GRCh37/hg19, GRCh38/hg38, or T2T-CHM13/hs1.

Optical Genomic Mapping (OGM)

Ultra-high-molecular weight DNA was isolated and purified using the Bionano Prep SP-G2 Blood and Cell Kit as per the manufacturer's instructions. Direct DNA labelling on CTTAAG sequence was conducted according to the DLS-G2 protocol with the DLE1 enzyme.

Labelled molecules were linearized into Saphyr chip G3.3 nanochannels to allow simultaneous direct imaging on the Saphyr instrument. A de novo assembly was carried out using the Bionano serve 3.7 and Access software version 1.7.

Genomic DNA and complementary DNA (cDNA) Sanger Sequencing

Breakpoints were confirmed by DNA Sanger sequencing after PCR amplification with primers 5'-GCTGTTTGCGTTGAAGAAGT-3' and 5'-CTGCCATTTCTTTTCACTCTGG-3' for breakpoint 1 (BP1); and 5'-ACC CCAGCACTCCTAAGAAC-3' and 5'-GGGACCACTATCAGCTGAC T-3' for breakpoint 2 (BP2).

For RNA Sanger sequencing, RNA was reverse-transcribed using SuperScript II reverse-transcriptase (Invitrogen, ref. 18064014) as per the manufacturer's instructions, with 1U/µL RNAse inhibitor (Applied Biosystems, ref. N8080119) and 2.5 µM Random Hexamer Primers (Invitrogen, ref. N8080127). cDNA was then amplified using a forward primer specific to *BRCA1* exon 20 (5'-AGAAACCACCAAGGTCCAAAG-3') and a reverse primer specific to *BRCA1* exon 9 (5'-GCCTTATTAACGGT ATCTTCAG-3').

PCR reactions were performed with Taq DNA Polymerase (VWR, ref. 733–1301) as per the manufacturer's instructions over 35 "touchdown" cycles (Hybridization: 58 °C x2; 57 °C x2; 56 °C x2; 55 °C x3; 54 °C x3; 53 °C x; 52 °C x4; 51 °C x5; 50 °C x10). Sequencing reactions were performed using Big Dye Terminator as per the manufacturer's instructions (ThermoFisher, ref. 4337452).

Strand-specific short read RNA sequencing (RNA-seq)

Strand-specific RNA-seq was performed on a NextSeq 500 (Illumina) after library preparation with custom SureSelect XT HS2 RNA probes (Agilent). We followed the manufacturer's protocol for strand-specific library preparation. Briefly, after initial preparation and fragmentation of 200 ng RNA, first-strand and second-strand cDNA were synthesized in two distinct steps with two distinct mixes. The second-strand cDNA mix contained dUTPs for specific second-strand marking. Reads were mapped on GRCh37/hg19 using STAR.

The sequencing depths of forward-strand and reverse-strand transcripts were compared to a merged bam file containing data from 12 distinct controls. These controls were patients suspected of carrying a genetic variant causing a splicing defect in a gene involved in paediatric cancer predisposition ($n = 2$), ataxia-telangiectasia or ataxia-telangiectasia-like disorders ($n = 3$), or digestive cancer predisposition ($n = 7$). All controls had provided informed consent for genetic analysis for diagnostic and research purposes.

Long-read direct RNA-seq

Long-read RNA-seq libraries were prepared from 1 µg of total RNA using the Oxford Nanopore Direct RNA Sequencing Kit (ref. SQK-RNA004). After an initial hybridization step, polyadenylated (polyA) messenger RNAs were captured, reverse-transcribed, and sequencing adaptors were ligated. Sequencing was performed using PromethION flowcell RNA (Oxford Nanopore Technologies, ref. FLO-PRO004RA) and reads were mapped to GRCh37/hg19 using Minimap2. For Fig. 2D design, alignments were displayed in Integrative Genomics Viewers (IGV 2.15.4) software, and fusion reads spanning *BRCA1* were exported for manual reconstruction on BP1 structure.

Reference transcripts

Represented transcripts correspond to NM_007294.4/ENST00000357654.9 for *BRCA1*, ENST00000657841.1 for *NBR2*, NM_005899.5/ ENST00000590996.6 for *NBR1*, NM_145041.4/ENST00000612339.4 for *TMEM160A*, NM_001363254.2/ENST00000636331.2 for *CCDC200*, and ENST00000635600.1 for *NBR1* antisense transcript *AC060780.1* (sharing the same first exon as NR_110868/ LOC101929767).

Data availability

Data available on reasonable request.

Received: 14 February 2025; Accepted: 17 July 2025;
Published online: 31 July 2025

References

1. Kuchenbaecker, K. B. et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA* **317**, 2402–2416 (2017).
2. Tutt, A. N. J. et al. Adjuvant Olaparib for patients with *BRCA1* - or *BRCA2* -mutated breast cancer. *N. Engl. J. Med.* **384**, 2394–2405 (2021).
3. Parsons, M. T. et al. Large scale multifactorial likelihood quantitative analysis of *BRCA1* and *BRCA2* variants: An ENIGMA resource to support clinical variant classification. *Hum. Mutat.* **40**, 1557–1578 (2019).
4. Caputo, S. M. et al. Classification of 101 BRCA1 and BRCA2 variants of uncertain significance by cosegregation study: A powerful approach. *Am. J. Hum. Genet.* **108**, 1907–1923 (2021).
5. Brandt, T. et al. Adapting ACMG/AMP sequence variant classification guidelines for single-gene copy number variants. *Genet. Med.* **22**, 336–344 (2020).
6. Richardson, M. E. et al. DNA breakpoint assay reveals a majority of gross duplications occur in tandem reducing VUS classifications in breast cancer predisposition genes. *Genet. Med.* **21**, 683–693 (2019).
7. Caputo, S. et al. 5′ Region Large Genomic Rearrangements in the BRCA1 Gene in French Families: Identification of a Tandem Triplication and Nine Distinct Deletions with Five Recurrent Breakpoints. *Cancers* **13**, 3171 (2021).
8. Filser, M. et al. Adaptive nanopore sequencing to determine pathogenicity of BRCA1 exonic duplication. *J. Med Genet.* https://doi.org/10.1136/jmg-2023-109155 (2023).
9. Chevrier, S., Richard, C., Mille, M., Bertrand, D. & Boidot, R. Nanopore adaptive sampling accurately detects nucleotide variants and improves the characterization of large-scale rearrangement for the diagnosis of cancer predisposition. *Clin. Transl. Med.* **15**, e70138 (2025).
10. Jones, M. A. et al. The landscape of BRCA1 and BRCA2 large rearrangements in an international cohort of over 20 000 ovarian tumors identified using next-generation sequencing. *Genes Chromosomes Cancer* **62**, 589–596 (2023).
11. Nurk, S. et al. The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
12. Altemose, N. et al. Complete genomic and epigenetic maps of human centromeres. *Science* **376**, eabl4178 (2022).
13. Vollger, M. R. et al. Segmental duplications and their variation in a complete human genome. *Science* **376**, eabj6965 (2022).
14. Oehler, J. B., Wright, H., Stark, Z., Mallett, A. J. & Schmitz, U. The application of long-read sequencing in clinical settings. *Hum. Genomics* **17**, 73 (2023).
15. Yu, C.-Y. & Kuo, H.-C. The emerging roles and functions of circular RNAs and their generation. *J. Biomed. Sci.* **26**, 29 (2019).
16. Wilusz, J. E. A 360° view of circular RNAs: From biogenesis to functions. *Wiley Interdiscip. Rev. RNA* **9**, e1478 (2018).
17. Wagner, J. et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat. Biotechnol.* **40**, 672–680 (2022).
18. Hamza, A. et al. The absence of CFHR3 and CFHR1 genes from the T2T-CHM13 assembly can limit the molecular diagnosis of complement-related diseases. *Eur. J. Hum. Genet.* **31**, 730–732 (2023).
19. Spurdle, A. B. et al. Refined histopathological predictors of BRCA1 and BRCA2mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. *Breast Cancer Res.* **16**, 3419 (2014).
20. O'Mahony, D. G. et al. Ovarian cancer pathology characteristics as predictors of variant pathogenicity in BRCA1 and BRCA2. *Br. J. Cancer* **128**, 2283–2294 (2023).
21. Belman, S., Parsons, M. T., Spurdle, A. B., Goldgar, D. E. & Feng, B.-J. Considerations in assessing germline variant pathogenicity using cosegregation analysis. *Genet Med.* **22**, 2052–2059 (2020).
22. Thompson, D., Easton, D. F. & Goldgar, D. E. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am. J. Hum. Genet.* **73**, 652–655 (2003).
23. Goldgar, D. E. et al. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum. Mutat.* **29**, 1265–1272 (2008).

24. Vallée, M. P. et al. Adding in silico assessment of potential splice aberration to the integrated evaluation of *BRCA* gene unclassified variants. *Hum. Mutat.* **37**, 627–639 (2016).
25. French COVAR group collaborators et al. Full in-frame exon 3 skipping of *BRCA2* confers high risk of breast and/or ovarian cancer. *Oncotarget* **9**, 17334–17348 (2018).
26. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinforma.* **14**, 178–192 (2013).
27. Perez, G. et al. The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res.* **53**, D1243–D1249 (2025).

Acknowledgements

We gratefully acknowledge all patients who participated in this study for their valuable contribution.

Author contributions

M.S., M.F., V.S., C.R., C.A., and J.M.P. analysed long-read sequencing and optical genomic mapping data. M.S., K.A., E.P.N., V.S., and L.G. analysed transcriptomic data. M.S., K.A., E.P.N., V.S., A.Re., C.G., and L.G. analysed short-read and Sanger sequencing data. M.F. and E.L. performed long-read sequencing and RNA sequencing. K.M. and V.R. performed bioinformatics analysis. K.A., H.T., and C.D.D.E. performed short-reads and Sanger sequencing, multiplex-ligation probe amplification, and microsatellites analysis. A.Ra. performed optical genomic mapping. M.E. and S.B. performed long-read RNA sequencing. S.A. and S.M.C. centralized patients' information and performed cosegregation analysis. C.D. and E.F. performed medical consultations. M.S., M.F., L.G., E.F., J.M.P., and S.M.C. were major contributors in writing the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-025-00517-0.

Correspondence and requests for materials should be addressed to Mathias Schwartz.