







OPEN ACCESS

Short report

Adaptive nanopore sequencing to determine pathogenicity of *BRCA1* exonic duplication

Mathilde Filser ^{1,2}, Mathias Schwartz,^{1,2} Kevin Merchadou,^{2,3} Abderaouf Hamza,^{1,2} Marie-Charlotte Villy ^{4,5}, Antoine Decees,^{1,2} Eléonore Frouin,^{2,3} Elodie Girard,^{2,6} Sandrine M Caputo ^{1,2}, Victor Renault,^{2,3} Véronique Becette,^{2,7} Lisa Golmard ^{1,2}, Nicolas Servant,^{2,6} Dominique Stoppa-Lyonnet ^{1,5}, Olivier Delattre,^{1,8} Chrystelle Colas ^{2,4}, Julien Masliah-Planchon^{1,2}

¹Genetics Department, Institut Curie, Paris, France

²PSL Research University, Paris, France

³Clinical Bioinformatics Unit, Institut Curie, Paris, France

⁴Oncogenetic Clinic Unit, Institut Curie, Paris, France

⁵SIREDO Oncology Centre, Institut Curie, Paris, France

⁶INSERM U900, Institut Curie, Paris, France

⁷Anatomo- and Cyto-pathology, Institut Curie, Saint-Cloud, France

⁸Inserm U830, PSL University, Research Center, Institut Curie, Paris, France

Correspondence to

Dr Mathilde Filser, Genetics Department, Institut Curie, Paris 75005, France; mathilde.filser@curie.fr

Received 11 January 2023

Accepted 30 April 2023

Published Online First 1 June 2023

ABSTRACT

BRCA1 and *BRCA2* are tumour suppressor genes that have been characterised as predisposition genes for the development of hereditary breast and ovarian cancers among other malignancies. The molecular diagnosis of this predisposition syndrome is based on the detection of inactivating variants of any type in those genes. But in the case of structural variants, functional consequences can be difficult to assess using standard molecular methods, as the precise resolution of their sequence is often impossible with short-read next generation sequencing techniques. It has been recently demonstrated that Oxford Nanopore long-read sequencing technology can accurately and rapidly provide genetic diagnoses of Mendelian diseases, including those linked to pathogenic structural variants. Here, we report the accurate resolution of a germline duplication event of exons 18–20 of *BRCA1* using Nanopore sequencing with adaptive sampling target enrichment. This allowed us to classify this variant as pathogenic within a short timeframe of 10 days. This study provides a proof-of-concept that nanopore adaptive sampling is a highly efficient technique for the investigation of structural variants of tumour suppressor genes in a clinical context.

such as short-read sequencing, does not allow the differentiation between: (1) a tandem duplication that would be classified as pathogenic if it induces a premature STOP codon or if it were localised within a functional domain that would be destabilised and (2) an insertion of the extra copy somewhere else in the genome that would be classified of unknown significance. This distinction has a critical impact on the patient's management and guides the choice for the type of surgery and the type of treatment (PARP inhibitor therapy, chemotherapy, etc). Hence, a technology enabling the identification of tandem intragenic duplication at low-cost and high-speed would be of great interest in clinical routine molecular biology. The interest in nanopore sequencing is growing in the field of molecular pathology, notably because of its capacity to precisely resolve structural variants. Recent work showed that nanopore sequencing can rapidly and accurately provide genetic diagnoses of Mendelian diseases.⁷ Here, we accurately describe a germline tandem duplication of exons 18–20 of *BRCA1* using Oxford Nanopore Sequencing Technology, with adaptive sampling target enrichment. This allowed us to better characterise and classify this structural variant (SV) as pathogenic in a short timeframe.

INTRODUCTION

Breast cancer susceptibility genes 1 and 2 (*BRCA1* and *BRCA2*) are tumour suppressor genes, exerting their functions by contributing to DNA repair mechanisms and transcriptional regulation in response to DNA damage.¹ Loss-of-function variants in one of these genes trigger the accumulation of genetic alterations and contribute to cancer development. These genes have been characterised as predisposition genes for the development of hereditary breast and ovarian cancers mainly (HBOC syndrome) and also other cancer types such as pancreatic and prostate cancers.^{2,3} The lifetime risk of developing a cancer for a carrier of a pathogenic *BRCA1* variant by the age of 80 is about 72% for breast cancer and 44% for ovarian cancer.^{4,5} Since loss of function of *BRCA1* increases cancer risk, every type of genomic inactivating alterations has been described in this gene, including intragenic duplication.⁶ However, the identification of an intragenic duplication in the *BRCA1* gene with classical molecular techniques,

PATIENT AND METHODS

Patient and family history

A woman under age 61 years was referred for a genetics consultation because of a triple negative (TN) ductal carcinoma of the breast. Her medical records reported among others tobacco and alcohol consumptions and an ectopic pregnancy. Family was not suggestive of HBOC syndrome. She had two healthy sons under age 25 years. Her two siblings (one brother and one sister, both under age 61 years) were cancer-free. Her father died from cirrhosis before age 60 years; no other information concerning the paternal side of the family was available. The patient's mother was cancer free after age 61 years. A maternal aunt developed a breast cancer before age 61 years, a maternal uncle died from a lung cancer, and a maternal cousin was diagnosed with a colorectal adenocarcinoma before age 40 years. The patient's maternal grandfather was diagnosed before age 61 years with a pharyngeal carcinoma.



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Filser M, Schwartz M, Merchadou K, et al. *J Med Genet* 2023;**60**:1206–1209.

As the patient was affected with a TN breast cancer before age 61 years, a HBOC multigene panel was prescribed whose results would have a direct and short-term impact on the patient's management. The patient gave written informed consent for genetic testing and research studies.

Initial molecular analysis

Germline DNA was extracted from blood cells using the QiaSymphony DSP DNA Midikit according to the manufacturer's instructions. Genic regions of interest were enriched with an Agilent custom SureSelect QXT kit. Short-read next generation sequencing (NGS) was performed on an Illumina NextSeq 500. Bowtie2, VarScan2 and DESeq were used for read mapping on the hg19/GRCh37 reference genome, variant calling and copy number variation (CNV) detection, respectively. Data analysis was restricted to 13 high-penetrance HBOC genes, including *BRCA1* (NM_007294.3), according to national guidelines.⁸ Germline CNV were confirmed by Multiplex Ligation Probe Amplification (MLPA, MRC Holland probe mix P002-D1) on DNA extracted from a buccal swab.

Nanopore sequencing and adaptive sampling

Library preparation was performed in 2 hours using Oxford Nanopore Technologies Ligation Sequencing Kit SQK-LSK110 on 2 µg of genomic germline DNA. Half of the library was then injected into a Minion Flow Cell R9 for 24 hours. After 24 hours of sequencing, the flow cell was washed with a nuclease mix (Oxford Nanopore technology Flow Cell Wash Kit EXP-WSH004), and the second half of the library was injected and sequenced for additional 24 hours. Adaptive sequencing enables target enrichment without additional library preparation steps;⁹ we targeted regions encompassing the entire loci (including 5'-UTR, 3'-UTR and intronic regions) and 5 kb flanking regions upstream and downstream of 120 genes, including *BRCA1*, totalling 49 Mb (1.58% of the human genome). Downstream bioinformatic analysis was performed using NanoCliD, a custom bioinformatic pipeline (<https://github.com/InstituteCurieClinic/alBioinformatics/NanoCliD>); NanoCliD relies on guppy for basecalling, minimap2¹⁰ for alignment, Clair3¹¹ for Single Nucleotide Variation (SNV) calling and SVIM,¹² Sniffles,¹³ cuteSV¹⁴ and NaNovar¹⁵ for SV calling. The data processing toolkit called 'guppy' allowed to convert FAST5 files to FASTQ files. Alignments were manually reviewed on Integrative Genomic Viewer.

The breakpoints of the *BRCA1* exons 18–20 duplication were subsequently confirmed by Sanger sequencing on genomic DNA, using the forward primer in *BRCA1* intron 20 5'-TCGGAAG-GCTGAGTTGAGAG-3' and the reverse primer in *BRCA1* intron 17 5'-TCCCAGTGTTCACAAAGGCC-3'.

RESULTS

The HBOC NGS panel showed the presence of a heterozygous germline duplication encompassing *BRCA1* exons 18–20, subsequently confirmed by MLPA on an independent sample. No other variant of interest was detected. Based on the American College of Medical Genetics (ACMG) classification, this variant was located in a functional domain (PM1) and the exons 18–20 duplication had been reported in nine French families, without cosegregation data within the same family (PP1).¹⁶ The *BRCA1* exonic duplication was therefore classified as a 'variant of unknown significance' (class 3), as NGS data analysis could not demonstrate that the reading frame of the *BRCA1* transcript was altered (ie, that the event was a tandem duplication). Further cDNA analysis was required to confirm pathogenicity. However,

RNA is not routinely available and the technique is very time-consuming (ie, requiring ~2 months for analysis).

In the meantime, we were able to confirm the presence of the duplication and demonstrate that it was indeed a tandem duplication in less than 10 days using nanopore sequencing with adaptive sampling (figure 1). The mean genomic depth of coverage was 2.77× whereas it was about 10 times higher for the targeted regions with a mean coverage of 22.55×. The mean depth of coverage over the *BRCA1* locus was 24.63× and eight reads encompassed the breakpoints of the SV. The mean read length was 3.7 Kb. The precise breakpoint coordinates (Chr 17 (NC_000017.10:g.41,208,234_41,216,908dup ; NM_007300.4:c.5138–940_5340+835dup) collocated with Alu repetitive elements in intron 17 (AluY ; chr17:41,216,845–41,217,145) and intron 20 (AluJb ; chr17:41,208,192–41,208,495) of the *BRCA1* gene. This *BRCA1* tandem duplication of exons 18–20 alters the reading frame and introduces a premature stop codon. Sanger sequencing (figure 1) subsequently confirmed for clinical purposes the breakpoint, allowing a fast return of results to the patient. The demonstration of the tandem nature of this duplication and the disruption of the reading frame allowing the addition of the PVS1 argument of the ACMG classification,¹⁷ the duplication was therefore classified as pathogenic (class 5) within a timeframe compatible with optimal patient management; with this new classification, the patient was now eligible for surgery, and clinical doctors could discuss an annexectomy and complete mastectomy.

DISCUSSION

Genomic SV include deletions, duplications, inversions and triplications of varying sizes and are an important cause of genetic diseases, including SV affecting the *BRCA1* gene.^{18 19} CNV can indeed contribute to human genetic diseases by either influencing the copy number of dosage-sensitive genes or disturbing the gene sequence as a result of intragenic CNV.

The development of microarray technologies, such as Comparative Genomic Hybridization (CGH) and Single Nucleotide Polymorphism (SNP) arrays, and their implementation in the genetic laboratories since the 2000s have progressively supplanted the karyotyping technology with copy number profiling, thus improving the resolution and hence the identification of small CNV of few Kb. More recently, the application of bioinformatic tools to short-read NGS data has emerged as another powerful method to identify SV including fine intragenic CNV that previously remained beyond the resolution limit of conventional microarrays.

A study led on 184 germline duplications throughout the genome found that the most frequent mechanism is tandem duplication with direct orientation (83% in that study), while others result in the creation of gene fusion at breakpoints, triplications or adjacent duplications, insertional translocations and complex rearrangements.^{20 21} Intragenic duplications might consequently cause a loss of function if the duplicated exons are organised head-to-tail and either not in-frame, of a large size or in a functional domain of importance in the protein. However, this type of SV can be indistinguishable in short-read NGS data from an insertion of an extra copy somewhere else in the genome. In this case, the SV might not disrupt the gene and might not be pathogenic because the extra copy is not supposed to disrupt the sequence of the gene.

Therefore, interpreting the pathogenic consequences of a duplication ideally requires breakpoint-level analysis as the subsequent clinical management of the patients may be totally

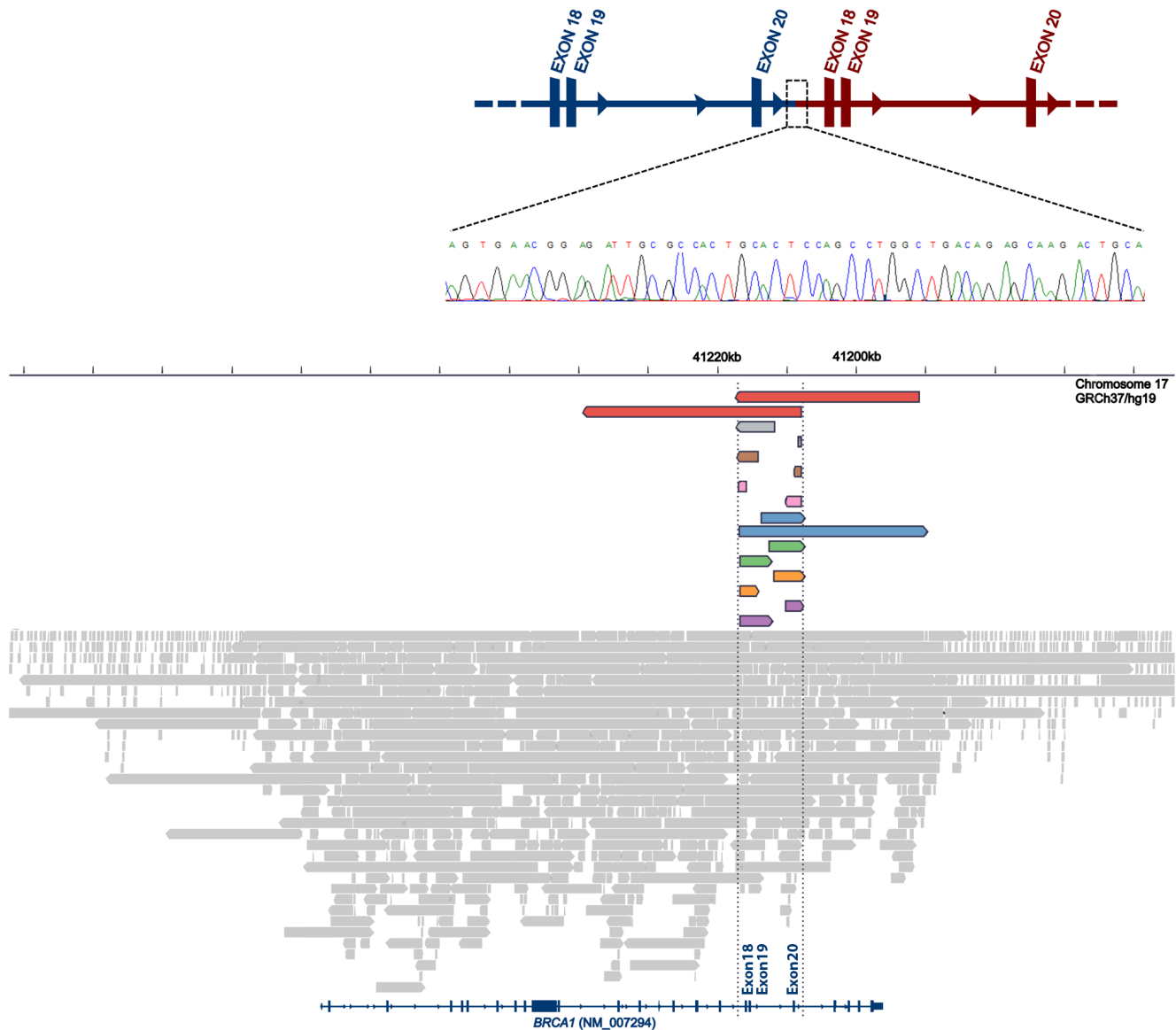


Figure 1 Duplication of exons 18–20 of *BRCA1* gene. Long-read sequences generated by nanopore sequencing visualised in IGV indicate the presence of the tandem duplication of exons 18–20 of *BRCA1* in our patient with hereditary breast and ovarian cancer syndrome. The grey reads represent all the reads generated for the patient's DNA at this localisation and the coloured reads above represent the reads spanning the duplication. Sanger sequencing chromatogram on the upper panel allows to precisely determine the sequence of the breakpoint: the normal sequence of intron 20 of *BRCA1* is framed in blue and the following sequence framed in red is abnormal and represents a sequence of intron 17 of *BRCA1* hence confirming the tandem duplication of exons 18–20. IGV, Integrative Genomic Viewer.

different depending on the variant classification.^{6 18 19} The identification of a duplication in the *BRCA1* gene with classic molecular techniques such as short-read sequencing techniques does not differentiate between a tandem duplication that would be classified as pathogenic or a duplication of unknown significance. Usually, in our laboratory, we would have investigated the pathogenicity of such a SV by a cDNA analysis. Yet, such analysis requires RNA, is very time-consuming and the average turn-around time is 2 months.

As an alternative, nanopore whole genome sequencing (WGS) can rapidly and accurately provide genetic diagnoses of mendelian diseases.⁷ Long reads do not suffer from the same mapping issues as short reads in repetitive regions and permit the detection of SV at a base-pair resolution. However, although WGS provides a more comprehensive genomic analysis, an advantage of targeted sequencing is its greater depth of coverage in

regions of interest and lower cost. In the clinical setting, a gene panel approach restricting the analysis to only actionable genes at greater depth of coverage (providing more confidence in the results) is common. Interestingly, a highly flexible targeted approach, called adaptive sampling, has recently been developed for nanopore sequencing.^{9 22} This strictly computational method targets the sequencing of specific regions of interest with nanopore technology without the need of effortful nucleic acid preparation. During the sequencing steps, fragments of interest are selected in real-time as they migrate through proteic nanopores by converting the electrical current into nucleotide sequence and comparing it to the provided regions of interest. Off-target DNA fragments are ejected from the pores, preserving sequencing capacity for DNA fragments of interest from on-target regions. Hence, by targeting genomic regions of interests, adaptive sequencing can enrich the sequenced data and

improve the read depth yield. In our case, we were able to reach a depth of coverage 10 times higher in the targeted genomic regions and to observe a tandem duplication of exons 18–20 of *BRCA1* gene. Although the 24× depth obtained by Adaptive sequencing is lower than those possibly obtained by choosing a Crispr enrichment approach, the alteration was unquestionably visible on the long reads. Moreover, the adaptive approach is extremely flexible since the targeted genes can be adapted by a simple change of the bed file, without designing new guides as is the case with the Crispr sequencing. In addition, this method allowed us to precisely identify the SV breakpoints, located in two Alu repetitive elements sharing 74% of identity. This result supports the hypothesis that this SV was mediated by non-allelic homologous recombination, ultimately gleaming further information about a possible mechanism leading to the SV of the *BRCA1* gene.

Conventional methods of enrichment typically involve cumbersome and expensive DNA processing steps prior to sequencing such as PCR-based amplification, hybridisation capture or Cas-mediated enrichment. By computationally selecting molecules during the sequencing process, these steps are avoided, making the process simpler, faster and easily adaptable. The possibilities offered by adaptive sampling contribute to the current trend of continuous improvement in applications of molecular biology for human diseases.

Nanopore sequencing coupled with adaptive sampling was demonstrated to be an effective, reliable and fast long-read sequencing technique. In the case reported here, it enabled the accurate resolution of an intragenic duplications of *BRCA1* and its classification as a pathogenic variant—ultimately guiding the clinician's decision and thus improving the clinical management of the patient and her relatives. This result obtained on one case is promising, and protocol optimisations as well as studies on more cases are to be conducted in order to improve yields and enhance the potential of nanopore sequencing.

Twitter Sandrine M Caputo @sandmarie77

Acknowledgements The author would like to thank Trenton Dailey-Chwalibóg for providing writing support and proofreading the manuscript.

Contributors JM-P: conceptualisation. MF, MS, AD, VB: data curation. KM, EF, EG, VR: formal analysis. JM-P: funding acquisition. MS, JM-P: methodology. AH, EF, VR, NS: software. LG, DS-L, OD, CC: supervision. MF, MS, JM-P: validation. MF, MS, JM-P: Roles/writing—original draft. SMC, M-CV, OD, DS-L: Writing—review and editing.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Consent obtained directly from patient(s)

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Mathilde Filser <http://orcid.org/0000-0002-5074-8100>
 Marie-Charlotte Villy <http://orcid.org/0000-0002-5714-6612>
 Sandrine M Caputo <http://orcid.org/0000-0001-5338-9388>
 Lisa Golmard <http://orcid.org/0000-0002-5158-8284>
 Dominique Stoppa-Lyonnet <http://orcid.org/0000-0002-5438-8309>
 Chrystelle Colas <http://orcid.org/0000-0001-9146-9346>

REFERENCES

- Paul A. The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers. *Front Biosci* 2014;19:605.
- Orban TI, Olah E. Emerging roles of BRCA1 alternative splicing. *Mol Pathol* 2003;56:191–7.
- Li S, Silvestri V, Leslie G, et al. Cancer risks associated with BRCA1 and BRCA2 pathogenic variants. *JCO* 2022;40:1529–41.
- van der Kolk DM, de Bock GH, Leegte BK, et al. Penetrance of breast cancer, ovarian cancer and contralateral breast cancer in BRCA1 and BRCA2 families: high cancer incidence at older age. *Breast Cancer Res Treat* 2010;124:643–51.
- Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 Mutation carriers. *JAMA* 2017;317:2402–16.
- Rouleau E, Jesson B, Briaux A, et al. Rare Germline large rearrangements in the BRCA1/2 genes and eight candidate genes in 472 patients with breast cancer predisposition. *Breast Cancer Res Treat* 2012;133:1179–90.
- Gozdzinski JE, Goenka SD, Shafiq K, et al. Ultrarapid Nanopore genome sequencing in a critical care setting. *N Engl J Med* 2022;386:700–2.
- Moretta J, Berthet P, Bonadona V, et al. Recommandations Françaises Du Groupe Génétique et cancer pour L'Analyse en panel de Gènes Dans LES Prédipositions Hérititaires au cancer Du Sein Ou de L'Ovaire | Elsevier enhanced reader. *Bulletin Du Cancer* 2018;105:907–17.
- Kovaka S, Fan Y, Ni B, et al. Targeted Nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nat Biotechnol* 2021;39:431–41.
- Li H, Birol I. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–100.
- Zheng Z, Li S, Su J, et al. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Bioinformatics* [Preprint] 2021.
- Heller D, Vingron M, Birol I. SVIM: structural variant identification using mapped long reads. *Bioinformatics* 2019;35:2907–15.
- Sedlazeck FJ, Rescheneder P, Smolka M, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 2018;15:461–8.
- Jiang T, Liu Y, Jiang Y, et al. Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol* 2020;21:189.
- Tham CY, Tirado-Magallanes R, Goh Y, et al. Nanovar: accurate characterization of patients' genomic structural variants using low-depth Nanopore sequencing. *Genome Biol* 2020;21:56.
- Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet Med* 2015;17:405–24.
- Xiang J, Peng J, Baxter S, et al. AutoPVS1: an automatic classification tool for PVS1 interpretation of null variants. *Hum Mutat* 2020;41:1488–98.
- Gad S, Caux-Moncoutier V, Pagès-Berhouet S, et al. Significant contribution of large BRCA1 Gene rearrangements in 120 French breast and ovarian cancer families. *Oncogene* 2002;21:6841–7.
- Caputo S, Telly D, Briaux A, 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* 2021;13:3171.
- Newman S, Hermetz KE, Weckselblatt B, et al. Next-generation sequencing of duplication CNVs reveals that most are Tandem and some create fusion genes at Breakpoints | Elsevier enhanced reader. *Am J Hum Genet* 2015;96:208–20.
- Richardson ME, Chong H, Mu W, et al. DNA Breakpoint assay reveals a majority of gross duplications occur in tandem reducing VUS classifications in breast cancer predisposition genes. *Genet Med* 2019;21:1669.
- Payne A, Holmes N, Clarke T, et al. Readfish enables targeted Nanopore sequencing of Gigabase-sized genomes. *Nat Biotechnol* 2021;39:442–50.