Article

Evaluation of the Available Variant Calling Tools for Oxford Nanopore Sequencing in Breast Cancer

Asmaa A. Helal [1], Bishoy T. Saad [1,*], Mina T. Saad [1], Gamal S. Mosaad [1] and Khaled M. Aboshanab [2,*]

[1] Department of Bioinformatics, HITS Solutions Co., Cairo 11765, Egypt
[2] Department of Microbiology and Immunology, Faculty of Pharmacy, Ain Shams University, Organization of African Unity St., Cairo 11566, Egypt
* Correspondence: bishoyth@hitssolutions.com (B.T.S.); aboshanab2012@pharma.asu.edu.eg (K.M.A.)

Abstract: The goal of biomarker testing, in the field of personalized medicine, is to guide treatments to achieve the best possible results for each patient. The accurate and reliable identification of everyone's genome variants is essential for the success of clinical genomics, employing third-generation sequencing. Different variant calling techniques have been used and recommended by both Oxford Nanopore Technologies (ONT) and Nanopore communities. A thorough examination of the variant callers might give critical guidance for third-generation sequencing-based clinical genomics. In this study, two reference genome sample datasets (NA12878) and (NA24385) and the set of high-confidence variant calls provided by the Genome in a Bottle (GIAB) were used to allow the evaluation of the performance of six variant calling tools, including Human-SNP-wf, Clair3, Clair, NanoCaller, Longshot, and Medaka, as an integral step in the in-house variant detection workflow. Out of the six variant callers understudy, Clair3 and Human-SNP-wf that has Clair3 incorporated into it achieved the highest performance rates in comparison to the other variant callers. Evaluation of the results for the tool was expressed in terms of Precision, Recall, and F1-score using Hap.py tools for the comparison. In conclusion, our findings give important insights for identifying accurate variants from third-generation sequencing of personal genomes using different variant detection tools available for long-read sequencing.

Keywords: nanopore; variant detection; human-SNP-wf; Clair3; Clair; NanoCaller; Longshot; Medaka

1. Introduction

Over time, the field of genetic testing for many cancer biomarkers, such as breast cancer driver genes BRCA1 and BRCA2, improved, starting from single gene sequencing on sanger sequencing technology, followed by multigene panels, which were created as a result of developments in next-generation sequencing technology (NGS), allowing for a broader genetic assessment, a faster testing method, and better throughput, without being cost prohibitive but constrained by the generation of short reads [1,2]. MinION, the first long-read Nanopore-based sequencer, was released by Oxford Nanopore Technologies (ONT), overcoming the primary limitations of short-read sequence creation [3] by introducing long-read sequencing technology that was adapted by both ONT and Pacific Biosciences (PacBio) [4]. These technologies proved that new long-read, single-molecule sequencing technologies could reliably be able to identify small variants, indel, and structural variants (SVs), with significant improvements in both sensitivity and specificity [3,5].

In human genomes, single-nucleotide polymorphisms (SNPs) and short insertions and/or deletions (indel) are two forms of genetic variants [6,7]. They contribute to genetic diversity and have the ability to affect phenotypic differences, such as human disease susceptibility. Detecting SNPs and indel is challenging in studying genomic variants and functions using new generations of high-throughput sequencing data [5]. Many different variant (SNP/indel) callers were introduced by the Nanopore community and recommended by ONTs for accurate variant detection based on data from long-read sequencing.

Some variant callers implemented variant calling methods using deep learning, such as "Clair" [8], the successor of "Clairvoyant" [9]. "Longshot " [10] calls SNPs on long-read data using a Pair-Hidden Markov Model (pair-HMM) for a small local window surrounding candidate sites. Medaka [11], an SNP/indel caller based on deep learning on long-read data, was recently launched by ONTs [11]. Medaka predicts SNPs from unphased long reads before phasing them. For each set of phased reads, Medaka ends up making SNP and indel calling. Nanocaller [12] is a deep convolutional neural network that incorporates a long-range haplotype structure to improve variant detection on long-read sequencing data. "Clair3" [13] combines the greatest characteristics of two key method categories: pile-up calling, which handles most variant candidates fast, and full alignment, which tackles complicated candidates with precision and recall in an account. Accordingly, in this article, the development of a workflow for detecting disease-causing variants, starting from the sample to the variant call format (VCF) with annotated variants, was proposed where different variant calling tools were tested on reference genome samples to evaluate the output of each tool against "Truth" set of variants. The proposed pipeline for targeted sequencing of the data generated from long-read sequencing technology, where the two genes BRCA1 and BRCA2, which are recurrently mutated in breast cancer, were analyzed as an example of this workflow and an examination of its performance was described for future testing and implementation.

2. Materials and Methods

2.1. Targeted Sequencing Data Analysis Pipeline

The target amplicons' reads were aligned to reference sequences based on the public human genome build GRCh38/UCSC hg38 using Minimap2 Aligner trained on long reads generated by ONT-MinIon sequencer (https://github.com/lh3/minimap2 (accessed on 8 August 2022) [14]. After Minimap2 finishes the alignment, it generates a SAM file that is converted afterward to BAM format using Samtools (https://github.com/samtools/ (accessed on 8 August 2022) [15]. The resultant BAM file was sorted and indexed using Samtools to be ready for variant calling. The minimum sequencing depth value was found to never be below 50 X using Bedtools "coverage" (https://github.com/ryanlayer/bedtool (accessed on 8 August 2022) [16]. Afterward, the PCR duplicate removal was performed on the reads that have identical external coordinates, retaining only the reads with the highest mapping quality using Samtools rmdup (with s option) that removes the single-end reads from the sorted and indexed Bam file (https://github.com/samtools/ (accessed on 8 August 2022)) [15]. Regarding the variant calling step, six variant callers were tested in parallel on the MinIon sequencing data: (1) Medaka (https://github.com/nanoporetech/medaka (accessed on 8 August 2022)) [11], (2) epi2me-labs/wf-human-snp (https://github.com/epi2me-labs/wf-human-snp (accessed on 8 August 2022)) [17], (3) Clair3 https://github.com/HKU-BAL/Clair3 (accessed on 8 August 2022) [13], (4) Clair (https://github.com/HKU-BAL/Clair (accessed on 8 August 2022)) [8], (5) Longshot (https://github.com/pjedge/longshot (accessed on 8 August 2022)) [10], (6) Nanocaller (https://github.com/WGLab/NanoCaller (accessed on 8 August 2022)) [12]. A custom-made BED file was created to target the region of the BRCA1 and BRCA2 genes for the variant callers to call only variants in our target regions. The 'SNV' (single-nucleotide variant) and 'INDEL' (insertion–deletion) files were filtered by removal of non "PASS" variants and with Quality "QUAL" below 20. The filtered VCF of variants was then annotated using the Genetic variant annotation and functional effect prediction toolbox SnpEff (https://pcingola.github.io/SnpEff/ (accessed on 8 August 2022)) [18], which predicts the effects of the resultant variants on genes and amino acid changes. The ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/ (accessed on 8 August 2022)) [19] database was used to check for the clinical significance of the annotated variants. The database is strongly linked to the databases dbSNP and dbVar, which keep track of the site of variations in human assembly. ClinVar is based on the phenotypic descriptions kept in MedGen (http://www.ncbi.nlm.nih.gov/medgen (accessed on 8 August 2022)) [20] as well. The

SNV and INDEL variants that were clinically significant are reported and stored in the in-house database (Table 1).

Table 1. Summary of the tools used in both SNP and indel detection.

Tool	Version	Function
Guppy	v5.0.16	data processing toolkit that contains Oxford Nanopore's base-calling algorithms. Guppy is integrated into MinKNOW and is also available as a standalone version.
Minimap2	v2.22	A sequence alignment tool that aligns DNA or mRNA sequences to a vast library of reference sequences.
Samtools	v.1.14	a collection of programs for manipulating alignments in the SAM, BAM, and CRAM formats. It converts between formats, sorts, merges, and indexes data, it can quickly remove PCR duplicates and calculate the mean coverage for a target region
Medaka	v1.4.4	a program that uses Nanopore sequencing data to generate consensus sequences and calling of variants.
Clair	v2.11	a tool that uses single molecule sequencing data to call germline small variants quickly and accurately.
Longshot	v0.4.1	a tool for detecting variants in diploid genomes using long error-prone reads. It takes an aligned BAM/CRAM file as input and outputs a phased VCF file containing variant and haplotype information.
NanoCaller	v2.1.2	a computational method for detecting SNPs/indels in long-read sequencing data that integrates long reads in a deep convolutional neural network and generates predictions for each SNP candidate variant site by considering pileup information from other candidate sites that share reads.
Clair3	v0.1-r11	a long-read germline small variant caller excels in two major method categories: pileup calling, which handles most variant candidates quickly, and full alignment, which tackles complex candidates to maximize precision and recall.
Hap.py	v0.3.15	To compare a VCF with a gold standard dataset vcf
SnpEff	v5.1	Toolbox for genetic variant annotation and functional effect prediction. It describes and estimates the effects of genetic variants on genes and proteins (such as amino acid changes)
Epi2me-labs/wf-human-SNP	v0.3.1	includes a nextflow workflow for calling diploid variants in whole genome data. Clair3 is used in this workflow to identify small variants in long reads.

SAM: Sequence Alignment Map, BAM: Binary Alignment Map, CRAM: Compressed Reference-oriented Alignment Map, VCF: Variant call format

2.2. Classification of the Pathogenicity of Variants

The information deposited in the ClinVar database and the recommendations of the American College of Medical Genetics and Genomics (ACMG) were used to classify the detected mutations [21,22]. The results of the BRCA1/2 gene variant detection were classified as wild type (no harmful variants), variant of unknown significance (VUS), pathogenic variants (PV), and likely pathogenic variants (LPV); not all the benign variants were reported [23].

2.3. Validation Data Set

To ensure the pipeline's usefulness and readiness, two long-read datasets based on publicly accessible human reference samples HG001 (NA12878) (https://www.ncbi.nlm.nih.gov/popset/?term=NA12878 (accessed on 8 August 2022)) and HG002 (NA24385) (https://www.ncbi.nlm.nih.gov/genome/?term=NA24385 (accessed on 8 August 2022) were provided by the ONT-open-data registry that is provided to support: (1) exploration of the properties of Nanopore sequence data; (2) performance evaluation and replication; (3) tool and method development. These are two of the most used reference samples. The Fastq files provided along with the bam files for this sample were used as input to test the validity of different tools' output (https://registry.opendata.aws/ont-open-data/ (accessed

on 8 August 2022)) using the benchmarking tool (https://github.com/Illumina/hap.py (accessed on 8 August 2022)) [24].

3. Results

3.1. Data Analysis Workflow Outcome

Data analysis workflow for the HG001 and HG002 reference genomes started with the read sequence aligner Minimap2, which aligns DNA sequences against the GRCh38 human reference genome with a SAM file as an output. Samtools "View" was used to convert the SAM file to a BAM file, followed by Samtools "Sort" and "Index" to generate a sorted and indexed BAM file ready for variant calling. As a part of the workflow pipeline, a step of PCR duplicate removal from the aligned reads of the two reference samples was included, to avoid overestimation of the coverage and overestimated variants resulting from PCR duplication with Samtools "rmdup". The mean coverage was calculated by bedtools "coverage". For the sample HG001, the mean coverage for the reads before PCR-duplicate removal was 32.62 X and 36.89 X for BRCA1 and BRCA2, respectively, while after removal of PCR duplicates, the mean coverage for BRCA1 and BRCA2 was found to be the same. The mean coverage for HG002, before the PCR-duplicate removal, was 53.85 X for BRCA1 and 70.06 X for BRCA2. After removing the duplicates, the mean coverage of BRCA1 and BRCA2 was found to be the same, which suggested that the published reference samples previously underwent the step of PCR-duplicate removal or it was sequenced as a whole-genome sequencing sample, which is more logical (Table 2).

Table 2. The coverage difference before removing duplicates and after removing duplicates.

Sample	Before Removing Duplicates		After Removing Duplicates	
	BRCA1	BRCA2	BRCA1	BRCA2
HG001	32.62 X	36.89 X	32.55 X	36.89 X
HG002	53.85 X	70.06 X	53.85 X	70.06 X

3.2. Primary Filtering Outcomes

The BAM files were ready for the next step, which was the variant calling step. Six tools were used to call variants in the BRCA genes in HG001 and HG002; some of these tools were recommended by ONT and some by the ONT community for variant calling, such as Medaka, Clair, Nanocaller, Longshot, Clair3, and wf-human-snp workflow, which is the workflow provided by ONT employing Clair3 with pre-adjusted parameters for accurate variant calling. All of the generated output VCFs were filtered, including the variants with "PASS" and QUAL > 20 as a threshold for the comparison of the output of the tools. Long-read sequencing data aligned to a reference genome are taken as an input along with a BED file designed to target the two genes' coordination, which restricts the variants called in the target location into different variant callers, which output a VCF file with predicted SNPs and indel. The output after the primary filtering for the three samples is described in Tables 3 and 4.

Table 3. The total no. of the output variants (SNPs, INDELs, and MNPs) of the six variant callers in comparison to both BRCA1 and BRCA2 genes in the HG001.

Tool Name	Total No. of BRCA1 Variants	Total No. of BRCA2 Variants	Total
Clair	482	348	830
Longshot	124	108	232
NanoCaller	121	97	218
Medaka	221	221	442
Clair3	225	172	397
Epi2me-labs/wf-human-SNP	370	285	655

Table 4. The total no. of the output variants (SNPs, INDELs, and MNPs) of the six variant callers in comparison to both BRCA1 and BRCA2 genes in the HG002.

Tool Name	Total No. of BRCA1 Variants	Total No. of BRCA2 Variants	Total
Clair	482	372	854
Longshot	124	108	232
NanoCaller	121	97	218
Medaka	111	98	209
Clair3	370	172	542
Epi2me-labs/wf-human-SNP	370	285	655

3.3. Comparison of the Variant Caller's Performance

For a comparison of the variant caller's performance, the traditional binary classification performance assessment paradigm of simply determining true and false "positives" and "negatives" lends itself well to evaluating the performance of variant callers [6]. By comparing the results to the truth sets for the NA24385 sample or NA12878 sample using the Hap.py tool that enumerates the variants between a "truth" VCF file containing the truth set of variants and a "query" VCF file, which contains the set of output variants of the variant caller along with a BED file that restricts the comparison to variants in the specified target location to determine the reliability of the variant calling conducted. The hap.py tool outputs a summary with the true positive "TP", false positive "FP", false negative "FN", Precision, Recall, or sensitivity, and finally, F1-score, which is an indication and a representation of both precision and recall. The data generated from the comparison tool "Happy" were summarized to include important metrics, such as Recall, Precision, F1-score, and the time taken by the tool to call the variants in both genes (Tables 5 and 6). With respect to the time taken for the tools to perform the variant calling on only the coordination of BRCA1 and BRCA2 genes for each sample, Nanocaller proved to be faster in this aspect where the time taken for Nanocaller was the lowest and Clair was proved to take the longest time in two samples HG001 and HG002 (Tables 5 and 6).

Table 5. Summary for the benchmarking output for HG001 with 6 different variant callers, highlighting the recall, precision, and F1-score.

HG001 (NA12878)		Recall	Precision	F1 Score	Total Time Taken
1. Human-SNP-wf	BRCA1-SNP	98.04%	95.24%	96.62%	1 h
	BRCA1-INDEL	94.12%	80.00%	86.49%	
	BRCA2-SNP	95.24%	96.15%	95.69%	
	BRCA2-INDEL	94.74%	75.00%	83.72%	
2. Clair3	BRCA1-SNP	99.02%	96.19%	97.58%	1 h 22 min
	BRCA1-INDEL	94.12%	80.00%	86.49%	
	BRCA2-SNP	96.19%	97.12%	96.65%	
	BRCA2-INDEL	94.74%	81.82%	87.80%	
3. Medaka	BRCA1-SNP	92.16%	89.52%	90.82%	1 h 29 min
	BRCA1-INDEL	58.82%	50.00%	54.05%	
	BRCA2-SNP	94.29%	95.19%	94.74%	
	BRCA2-INDEL	57.89%	50.00%	53.66%	
4. Nanocaller	BRCA1-SNP	96.08%	93.33%	94.69%	42 min
	BRCA1-INDEL	76.47%	65.00%	70.27%	
	BRCA2-SNP	95.24%	96.15%	95.69%	
	BRCA2-INDEL	80.00%	54.55%	64.86%	
5. Longshot	BRCA1-SNP	95.10%	92.38%	93.72%	48 min
	BRCA1-INDEL	70.59%	60.00%	64.86%	
	BRCA2-SNP	93.33%	94.23%	93.78%	
	BRCA2-INDEL	68.42%	59.09%	63.41%	
6. Clair	BRCA1-SNP	96.08%	93.33%	94.69%	2 h
	BRCA1-INDEL	64.71%	55.00%	59.46%	
	BRCA2-SNP	93.33%	94.23%	93.78%	
	BRCA2-INDEL	63.16%	54.55%	58.54%	

Table 6. Summary for the benchmarking output for HG002 with 6 different variant callers, highlighting the recall, precision, and F1-score.

	HG002 (NA24385)	Recall	Precision	F1-Score	Total Time Taken
1. wf-Human-SNP	BRCA1-SNP	97.20%	99.05%	98.11%	
	BRCA1-INDEL	93.33%	70.00%	80.00%	
	BRCA2-SNP	97.06%	98.02%	97.54%	43 min
	BRCA2-INDEL	95.00%	90.48%	92.68%	
2. Clair3	BRCA1-SNP	96.26%	98.10%	97.17%	
	BRCA1-INDEL	86.67%	65.00%	74.29%	
	BRCA2-SNP	95.10%	96.04%	95.57%	1 h 7 min
	BRCA2-INDEL	85.00%	80.95%	82.93%	
3. Medaka	BRCA1-SNP	91.59%	93.33%	92.45%	
	BRCA1-INDEL	60.00%	45.00%	51.43%	
	BRCA2-SNP	90.20%	91.09%	90.64%	39 min
	BRCA2-INDEL	60.00%	57.14%	58.54%	
4. Nanocaller	BRCA1-SNP	95.33%	97.14%	96.23%	
	BRCA1-INDEL	80.00%	60.00%	68.57%	
	BRCA2-SNP	94.12%	95.05%	94.58%	28 min
	BRCA2-INDEL	85.00%	80.95%	82.93%	
5. Longshot	BRCA1-SNP	94.39%	96.19%	95.28%	
	BRCA1-INDEL	73.33%	55.00%	62.86%	
	BRCA2-SNP	92.16%	93.07%	92.61%	38 min
	BRCA2-INDEL	75.00%	71.43%	73.17%	
6. Clair	BRCA1-SNP	93.46%	95.24%	94.34%	
	BRCA1-INDEL	66.67%	50.00%	57.14%	
	BRCA2-SNP	91.18%	92.08%	91.63%	1 h 11 min
	BRCA2-INDEL	65.00%	61.90%	63.41%	

4. Discussion

Evaluation of BRCA1/2 molecular status has become the standard of care in the treatment of individuals with breast cancer. Precision medicine has made significant progress against this type of cancer, which accounts for one-third of all new female cancers every year. Female breast cancer is the sixth biggest cause of mortality worldwide, with an estimate of 685,000 deaths in 2020 [25]. One example is the development and clinical application of PARP-inhibitor (PARPi); Poly (adenosine diphosphate-ribose) polymerase inhibitors (PARPi) are a key arrow in the oncologist's quiver among new therapeutics [26,27]. Indeed, PARPi has been found to enhance the clinical outcomes of breast cancer patients with BRCA1/2 germline or somatic mutations, which have been found to improve survival and quality of life [28–32]. As a result, current worldwide guidelines strongly advise BRCA1/2 testing in all patients. Rapid and dependable genetic screening for BRCA1/2 germline or somatic mutations has become critical in identifying individuals who would most likely benefit from these treatments [3,33,34].

The technology used in BRCA 1/2 gene testing held an important impact on getting the full picture of the two genes. Traditional Sanger sequencing is expensive and takes a long turn-around time (TAT). Next-generation sequencing (NGS) is a game-changing high-throughput nucleotide sequencing approach that produces rapid, cheap, and accurate genomic data. NGS developed the clinical methodology for genetic examination across various fields of medicine [34]. NGS can massively sequence millions of DNA reads, allowing for accurate characterization of the "status" of multiple genes; in this context, NGS-targeted gene sequencing enables the detection of driver mutations, which are responsible for progression and relapse and might be employed as predictive or prognostic biomarkers in breast cancer [33,34]. When compared to Sanger sequencing, NGS can offer doctors comparable genetic information at a cheaper cost and shorter time to results [2,34], yet the NGS limitations are the small read size and the difficulty in analyzing large alterations as structural variants. Many studies employed the NGS as a technology in the detection of BRCA1/2 gene variants in various ethnic groups to implement the detection of gene variants using NGS in the routine line of diagnostics and may allow doctors to make more prompt and informed decisions about surgery or neo-adjuvant chemotherapy in breast cancer patients [35–42].

However, the use of NGS technologies in clinical diagnostics necessitates a large initial investment in the sequencer, which is a barrier for local research institutions in underdeveloped nations, as well as small research institutes and hospitals. MinION, the first commercially available sequencer based on Nanopore technology, might be a viable alternative [43,44]. MinION has previously been utilized effectively to identify mutations in TP53 and ABL1 genes in CLL and CML patients [45–48], respectively. Furthermore, the cheap cost, ease of use, and length of the reads make MinION a perfect instrument for targeted gene sequencing; the long read can enable researchers to detect and phase genetic variants, as well as thoroughly define new isoforms and fusion transcripts, using Nanopore technology. Nanopore technology sheds new light on health and disease, ranging from cancer to immunology and neurology [48].

In the current study, the main focus was on the data analysis of data generated using Nanopore technology, as there are many proposed tools by the Nanopore community, a hub for all the Nanopore technology users (https://community.nanoporetech.com/ (accessed on 8 August 2022)) for every step along the way in data analysis. The in-house targeted gene sequencing workflow was divided into two parts: (1) design a data analysis pipeline for SNV/INDEL/SV detection and how to validate this pipeline and (2) design an in-house primer panel for BRCA1/2 genes as a prototype for future implementation. The pipeline design started with a set of tools designed and trained on long-read data generated from the MinIon ONT sequencer; the reference samples used as the input data for validation of this workflow are the publicly published "NA12878" (HG001) reference sample [49] and "NA24385" (HG002) dataset that contain whole-genome sequencing of well-known human cell lines, sequenced using Nanopore technology [50]. Each, therefore, serves as a helpful benchmark sample. The HG002 cell line was used as a "seen" sample in the current (PrecisionFDA Truth Challenge V2) competition [51]. The method of validating the performance of workflows and especially the variant callers is called "Benchmarking", where a reference sample is used either as DNA to be sequenced and undergo the workflow or using the data for this reference sample from the public repository in-silico for a data analysis step, a method that was recommended by Global Alliance for Genomics and Health (GA4GH) [52].

The pipeline went as follows: (1) mapping for the reads stored in the fastq file that outputs the reads into a SAM file format using "Minimap2" mapper for long reads against reference sequences based on the GRCh38/UCSC hg38 public human genome build, (2) sorting and indexing using Samtools as a versatile tool as it was heavily used in many pipelines proposed by other studies, used to convert a SAM file to BAM file, sort and index the BAM output, (3) removing the PCR duplicates even though the reference data samples used to validate this workflow were whole-genome sequencing, not including a PCR step but were included in the workflow as this workflow will be used on targeted gene sequencing data, (4) calculating the mean coverage of the targeted genes using Bedtools, (5) variant calling step, which is the main event in the workflow and the focus of our study; there are many variant callers both recommended by ONT and the Nanopore community, so the output variants were filtered based on "PASS" and QUAL > 20 as a threshold for the comparison of the tools output, (6) annotating the variants using SnpEff as an annotation tool, and (7) checking the clinical significance of the annotated variants using ClinVar clinical database.

The focus of the current study was to evaluate this workflow as well as compare the performance of the commonly used software pipelines for variant calling, which is another key element in variant discovery. The comparison is based on how well the tool calls the "True" variants when compared to the benchmarking VCF file; the tools analyzed in this study are Medaka, Clair, Nanocaller, Longshot, Clair3, and ONT's wf-human-snp workflow for variant calling, which employs Clair3 with pre-adjusted parameters for the accurate calling of variants.

Recent studies attempted to enhance variant calling by using phasing information from long-read sequencing data. Longshot calls SNPs on long-read data using a pair-

hidden Markov Model (pair-HMM) for a small local window surrounding candidate sites and then improves genotyping of identified SNPs using Hap-CUT2 [53] based on the most probable pair of haplotypes given the present variant genotypes, but on the other hand, is incapable of detecting indel. Medaka was provided by ONT, an SNP/indel caller that uses deep learning on long-read data. Medaka predicts SNPs from unphased long readings before using WhatsHap [54] to phase the data Medaka eventually makes SNP and indel calls for each phased read group. Clair, the successor of Clairvoyante, is a tool for detecting germline minor variants quickly and accurately using single-molecule sequencing data. Clair outperforms several competing systems for ONT data, including Clairvoyante, Longshot, and Medaka, in terms of precision, recall, and speed. As a deep learning approach, Nanocaller detects SNPs using long-range haplotype information, then phases long reads with identified SNPs and calls indels using local realignment.

Two key designs differ greatly in terms of performance and speed either employing pileup or full alignment as the input of the decision-making neural network. Clair and Nanocaller are pileup-based calling networks that aggregate read alignments into features and counts before sending them into a variant calling network. PEPPER-Margin-DeepVariant5 (PEPPER) [55] is fully alignment based. The DeepVariant variant calling network input is retained with spatial information in the full alignment method and is tens of times greater in size than the pileup method. Medaka is consensus based, using pileup input to generate a diploid consensus in the first iteration and two haploid consensuses in the second. Variants are formed by identifying and combining differences between the reference and consensus. To fill the void, Clair3 was created, which combines the best of both designs. It is as quick as pileup-based callers and performs just as well as full alignment callers. First, the pileup calling network goes through all the variant candidates that met a coverage and alternative allele frequency criterion. The high-quality pileup calls are then used to phase the alignments and generate the final output. Then, for each low-quality pileup call for full-alignment calling, the alignments phased by WhatsHap are utilized to create full-alignment input that is 23-times greater in size than the pileup input. Finally, as the final output, the full-alignment calls are combined with the high-quality pileup calls.

For performance validation of the pipeline along with the variant callers, the process started with the genome in a bottle (GIAB) reference samples HG001 and the Ashkenazi son sample HG002 ONT reads that were used as an input for mapping with Minimap2, sorting and indexing with Samtools, calculating the mean coverage within the BRCA1/2 gene bed file with coordination. for the variant calling step, the default parameters were used for all the variant callers to ensure uniformity in the output variants. The benchmarking variant VCF "Truth set" used was the GIAB v.4.2.1 for each reference genome sample to compare the output of different variant callers. The hap.py [24] tool was used for benchmarking, which is a reference implementation of the GA4GH recommendations for variant caller benchmarking with the "vcfeval" engine for comparison; it generated metrics as "False positive", "False negative", "True positive", "Precision", "Recall", and "F1 score". It was found that three metrics are the most important for variant caller performance evaluations, which are "Precision", "Recall", and, most importantly, "F1 score", which is the mean of precision and recall and is commonly used to test the performance of the callers [56–58].

Based on the metrics obtained in our results, it is suggested that Clair3 as a standalone or incorporated into a workflow as Human-SNP-wf by ONT, was found to be outperforming other variant callers concerning performance. The Clair3 method's efficiency is based on its ability to effectively distinguish between true and false calls during pileup calling, allowing only essential candidates to be transferred to the considerably more computationally costly full alignment calling. Following that comes Nanocaller, which performed in a better way than the rest of the variant callers, Longshot, Clair, and Medaka, respectively, agree with the findings of another study. Even though Clair is supposed to outperform Longshot, it was found to have lower F1 scores in both reference samples and that may be because Clair was outdated and was succeeded by Clair3 in May 2021

(https://github.com/HKU-BAL/Clair (accessed on 8 August 2022)). Although Medaka was, up until the release of Clair3, the recommended variant caller for SNP calling using the "medaka_variant" argument, which was formerly implemented inside the medaka package, it has been exceeded in accuracy and computing performance by alternative approaches and is, thus, deprecated and it is advised to utilize Clair3 either directly or through the Oxford Nanopore Technologies offered Nextflow implementation (Human-SNP-wf) (https://github.com/nanoporetech/medaka (accessed on 8 August 2022)) and that may explain the low performance. It was intentional not to test Nanopolish [59], which is also capable of variant calling on ONT data since it requires fast5 raw signals file as input, which are not publicly accessible for HG002, so it was excluded from the variant callers' comparison.

Targeted gene panels are one of the most frequent ways of enriching the genomic areas to be sequenced and they are widely utilized in NGS technology. Using Nanopore technology, we were able to enrich all the gene areas of interest without being limited by the read length. MinION real-time sequencing allows reads to be evaluated as they are produced, considerably speeding up analysis and allowing for the modification of experimental conditions as needed. Another benefit of MinION over second-generation sequencers is its mobility and ease of use for library preparation and sequencing, as well as its low cost. There are currently many custom/academic or commercial BRCA1/2 target panels that have been established in recent years because of investigations on the use and impact of NGS in breast/ovarian cancer [56,60–62], the majority of which are based on the amplicon sequencing technique. There are currently many commercial short-read amplicon-based BRCA gene panels available that detect SNV and/or copy number variation. Nonetheless, efforts to create a complete gene panel useful for BRCA prognosis and medication impact prediction are ongoing. The design of a primer panel targeting different oncology biomarkers will be incorporated into our future plan for trial on different cancer sample types.

5. Conclusions

In this study, six variant calling tools, including Human-SNP-wf, Clair3, Clair, NanoCaller, Longshot, and Medaka, were evaluated regarding their performance and accuracy in the detection of genetic variants. The tested genetic variants were single-nucleotide polymorphisms (SNPs) and short insertions and/or deletions (indel) of BRCA1 and BRCA2 genes, where two reference genome sample datasets (NA12878) and (NA24385) were used. The set of high-confidence variant calls provided by Genome in a Bottle (GIAB) was used to allow for the evaluation of the performance of six variant calling tools. The obtained results provide important insights for identifying accurate variants from third-generation sequencing of personal genomes using different variant detection tools available for long-read sequencing. The evaluation of the results was expressed in terms of Precision, Recall, and F1-score using Hap.py tools for the comparison. Both Clair3 and Human-SNP-wf tools accomplished the highest performance rates and should be implemented for evaluating the prognosis of breast cancer in humans.

Conflicts of Interest: The authors declare no conflict of interest as well as the funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

ACMG: American College of Medical Genetics and Genomics; BAM, Binary Alignment Map; BED, Browser Extensible Data; BRCA1, Breast cancer gene 1; BRCA2, Breast cancer gene 2; ClinVar, Database that aggregates information about genomic variation and its relationship to human health; CRAM, Compressed Reference-oriented Alignment Map; dbSNP, Database of Single Nucleotide Polymorphisms; dbVar, Database of human genomic structural variation; DSB, Double-strand breaks; GA4GH, Global Alliance for Genomics and Health; GIAB, Genome in a Bottle; LPV, Likely Pathogenic Variant; NGS, Next-generation sequencing; ONT, Oxford Nanopore Technologies; PacBio, Pacific Biosciences; PCR, Polymerase chain reaction; PV, Pathogenic variant; SAM, Sequence Alignment Map; SNP, Single-nucleotide polymorphisms; SNV, Single-nucleotide variant; SV, Structural variants; VCF, Variant call format; VUS, Variant of unknown significance.

References

1. Guan, Y.F.; Li, G.R.; Wang, R.J.; Yi, Y.T.; Yang, L.; Jiang, D.; Zhang, X.P.; Peng, Y. Application of Next-Generation Sequencing in Clinical Oncology to Advance Personalized Treatment of Cancer. *Chin. J. Cancer* **2012**, *31*, 463–470. [CrossRef] [PubMed]
2. Park, H.S.; Park, S.J.; Kim, J.Y.; Kim, S.; Ryu, J.; Sohn, J.; Park, S.; Kim, G.M.; Hwang, I.S.; Choi, J.R.; et al. Next-Generation Sequencing of BRCA1/2 in Breast Cancer Patients: Potential Effects on Clinical Decision-Making Using Rapid, High-Accuracy Genetic Results. *Ann. Surg. Treat. Res.* **2017**, *92*, 331. [CrossRef] [PubMed]
3. Bevers, T.B.; Anderson, B.O.; Bonaccio, E.; Buys, S.; Daly, M.B.; Dempsey, P.J.; Farrar, W.B.; Fleming, I.; Garber, J.E.; Harris, R.E.; et al. Breast Cancer Screening and Diagnosis. *J. Nat. Comprehen. Cancer Netw.* **2009**, *7*, 1060–1096. [CrossRef] [PubMed]
4. Jain, M.; Koren, S.; Miga, K.H.; Quick, J.; Rand, A.C.; Sasani, T.A.; Tyson, J.R.; Beggs, A.D.; Dilthey, A.T.; Fiddes, I.T.; et al. Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads. *Nat. Biotechnol.* **2018**, *36*, 338–345. [CrossRef] [PubMed]
5. Aganezov, S.; Goodwin, S.; Sherman, R.M.; Sedlazeck, F.J.; Arun, G.; Bhatia, S.; Lee, I.; Kirsche, M.; Wappel, R.; Kramer, M.; et al. Comprehensive Analysis of Structural Variants in Breast Cancer Genomes Using Single-Molecule Sequencing. *Genome Res.* **2020**, *30*, 1258–1273. [CrossRef] [PubMed]
6. Krusche, P.; Trigg, L.; Boutros, P.C.; Mason, C.E.; De La Vega, F.M.; Moore, B.L.; Gonzalez-Porta, M.; Eberle, M.A.; Tezak, Z.; Lababidi, S.; et al. Best Practices for Benchmarking Germline Small-Variant Calls in Human Genomes. *Nat. Biotechnol.* **2019**, *37*, 555–560. [CrossRef]
7. Karami, F.; Mehdipour, P. A Comprehensive Focus on Global Spectrum of BRCA1 and BRCA2 Mutations in Breast Cancer. *BioMed Res. Inter.* **2013**, *2013*, 928562. [CrossRef]
8. Luo, R.; Wong, C.-L.; Wong, Y.-S.; Tang, C.-I.; Liu, C.-M.; Leung, C.-M.; Lam, T.-W. Exploring the Limit of Using a Deep Neural Network on Pileup Data for Germline Variant Calling. *Nat. Mach. Intell.* **2020**, *2*, 220–227. [CrossRef]
9. Luo, R.; Sedlazeck, F.J.; Lam, T.-W.; Schatz, M.C. Clairvoyante: A Multi-Task Convolutional Deep Neural Network for Variant Calling in Single Molecule Sequencing. *Nat. Commun.* **2019**, *10*, 998. [CrossRef]
10. Edge, P.; Bansal, V. Longshot Enables Accurate Variant Calling in Diploid Genomes from Single-Molecule Long Read Sequencing. *Nat. Commun.* **2019**, *10*, 4660. [CrossRef]
11. GitHub—Nanoporetech/Medaka: Sequence Correction Provided by ONT Research. Available online: https://github.com/nanoporetech/medaka (accessed on 10 August 2022).
12. Ahsan, M.U.; Liu, Q.; Fang, L.; Wang, K. NanoCaller for Accurate Detection of SNPs and Indels in Difficult-to-Map Regions from Long-Read Sequencing by Haplotype-Aware Deep Neural Networks. *Genome Biol.* **2021**, *22*, 261. [CrossRef] [PubMed]
13. Luo, J.; Ding, H.; Shen, J.; Zhai, H.; Wu, Z.; Yan, C.; Luo, H. BreakNet: Detecting deletions using long reads and a deep learning approach. *BMC Bioinform.* **2021**, *22*, 577. [CrossRef] [PubMed]
14. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef] [PubMed]
15. Samtools·GitHub. GitHub. Available online: https://github.com/samtools/ (accessed on 10 August 2022).
16. GitHub—Ryanlayer/Bedtools. Available online: https://github.com/ryanlayer/bedtools (accessed on 10 August 2022).
17. GitHub—Epi2me-Labs/Wf-Human-Snp: Small Variant Calling for Human Samples. Available online: https://github.com/epi2me-labs/wf-human-snp (accessed on 10 August 2022).
18. SnpEff and SnpSift. Available online: http://pcingola.github.io/SnpEff/ (accessed on 10 August 2022).
19. ClinVar. Available online: https://www.ncbi.nlm.nih.gov/clinvar/ (accessed on 8 August 2022).
20. Home—MedGen—NCBI. Available online: https://www.ncbi.nlm.nih.gov/medgen/ (accessed on 8 August 2022).

21. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; 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*, 405–423. [CrossRef] [PubMed]

22. Kearney, H.M.; Thorland, E.C.; Brown, K.K.; Quintero-Rivera, F.; South, S.T. American College of Medical Genetics Standards and Guidelines for Interpretation and Reporting of Postnatal Constitutional Copy Number Variants. *Genet. Med.* **2011**, *13*, 680–685. [CrossRef]

23. Plon, S.E.; Eccles, D.M.; Easton, D.; Foulkes, W.D.; Genuardi, M.; Greenblatt, M.S.; Hogervorst, F.B.L.; Hoogerbrugge, N.; Spurdle, A.B.; Tavtigian, S.V. Sequence Variant Classification and Reporting: Recommendations for Improving the Interpretation of Cancer Susceptibility Genetic Test Results. *Hum. Mutat.* **2008**, *29*, 1282–1291. [CrossRef]

24. GitHub—Illumina/Hap.Py: Haplotype VCF Comparison Tools. Available online: https://github.com/Illumina/hap.py. (accessed on 10 August 2022).

25. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries—Sung—2021—CA: A Cancer Journal for Clinicians—Wiley Online Library. Available online: https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21660 (accessed on 8 August 2022).

26. Cortesi, L.; Rugo, H.S.; Jackisch, C. An Overview of PARP Inhibitors for the Treatment of Breast Cancer. *Target. Oncol.* **2021**, *16*, 255–282. [CrossRef]

27. McCann, K.E.; Hurvitz, S.A. Advances in the Use of PARP Inhibitor Therapy for Breast Cancer. *Drugs Context* **2018**, *7*, 212540. [CrossRef]

28. Tung, N.; Garber, J.E. PARP Inhibition in Breast Cancer: Progress Made and Future Hopes. *NPJ Breast Cancer* **2022**, *8*, 47. [CrossRef]

29. Demir Cetinkaya, B.; Biray Avci, C. Molecular Perspective on Targeted Therapy in Breast Cancer: A Review of Current Status. *Med. Oncol.* **2022**, *39*, 149. [CrossRef]

30. Pop, L.; Suciu, I.; Ionescu, O.; Bacalbasa, N.; Ionescu, P. The Role of Novel Poly (ADP-Ribose) Inhibitors in the Treatment of Locally Advanced and Metastatic Her-2/Neu Negative Breast Cancer with Inherited Germline BRCA1/2 Mutations. A Review of the Literature. *J. Med. Life* **2021**, *14*, 17–20. [CrossRef] [PubMed]

31. Liu, X.; Wu, K.; Zheng, D.; Luo, C.; Fan, Y.; Zhong, X.; Zheng, H. Efficacy and Safety of PARP Inhibitors in Advanced or Metastatic Triple-Negative Breast Cancer: A Systematic Review and Meta-Analysis. *Front. Oncol.* **2021**, *11*, 4363. [CrossRef] [PubMed]

32. Taylor, A.M.; Chan, D.L.H.; Tio, M.; Patil, S.M.; Traina, T.A.; Robson, M.E.; Khasraw, M. PARP (Poly ADP-Ribose Polymerase) Inhibitors for Locally Advanced or Metastatic Breast Cancer. *Cochrane Database Syst. Rev.* **2021**, *2021*, CD011395. [CrossRef]

33. Cardoso, F.; Kyriakides, S.; Ohno, S.; Penault-Llorca, F.; Poortmans, P.; Rubio, I.T.; Zackrisson, S.; Senkus, E. Early Breast Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann. Oncol.* **2019**, *30*, 1194–1220. [CrossRef]

34. Pepe, F.; Pisapia, P.; Russo, G.; Nacchio, M.; Pallante, P.; Vigliar, E.; De Angelis, C.; Insabato, L.; Bellevicine, C.; De Placido, S.; et al. BRCA1/2 NGS Somatic Testing in Clinical Practice: A Short Report. *Genes* **2021**, *12*, 1917. [CrossRef]

35. Concolino, P.; Rizza, R.; Mignone, F.; Costella, A.; Guarino, D.; Carboni, I.; Capoluongo, E.; Santonocito, C.; Urbani, A.; Minucci, A. A Comprehensive BRCA1/2 NGS Pipeline for an Immediate Copy Number Variation (CNV) Detection in Breast and Ovarian Cancer Molecular Diagnosis. *Clin. Chim. Acta* **2018**, *480*, 173–179. [CrossRef]

36. Fumagalli, C.; Rappa, A.; Casadio, C.; Betella, I.; Colombo, N.; Barberis, M.; Guerini-Rocco, E. Next-Generation Sequencing-Based BRCA Testing on Cytological Specimens from Ovarian Cancer Ascites Reveals High Concordance with Tumour Tissue Analysis. *J. Clin. Pathol.* **2020**, *73*, 168–171. [CrossRef]

37. Grafodatskaya, D.; O'Rielly, D.D.; Bedard, K.; Butcher, D.T.; Howlett, C.J.; Lytwyn, A.; McCready, E.; Parboosingh, J.; Spriggs, E.L.; Vaags, A.K.; et al. Practice Guidelines for BRCA1/2 Tumour Testing in Ovarian Cancer. *J. Med. Genet.* **2022**, *59*, 727–736. [CrossRef]

38. Trujillano, D.; Weiss, M.E.R.; Schneider, J.; Köster, J.; Papachristos, E.B.; Saviouk, V.; Zakharkina, T.; Nahavandi, N.; Kovacevic, L.; Rolfs, A. Next-Generation Sequencing of the BRCA1 and BRCA2 Genes for the Genetic Diagnostics of Hereditary Breast and/or Ovarian Cancer. *J. Mol. Diagn.* **2015**, *17*, 162–170. [CrossRef]

39. Solano, A.R.; Palmero, E.I.; Delgado, L.; Carraro, D.M.; Ortíz-López, R.; Carranza, C.L.; Santamaria, C.; Cifuentes, L.; Jara Sosa, L.E.; Toland, A.E. Sequencing Technology Status of BRCA1/2 Testing in Latin American Countries. *NPJ Genom. Med.* **2020**, *5*, 22. [CrossRef]

40. Sultova, E.; Westphalen, C.B.; Jung, A.; Kumbrink, J.; Kirchner, T.; Mayr, D.; Rudelius, M.; Ormanns, S.; Heinemann, V.; Metzeler, K.H.; et al. NGS-Guided Precision Oncology in Metastatic Breast and Gynecological Cancer: First Experiences at the CCC Munich LMU. *Arch. Gynecol. Obstet.* **2021**, *303*, 1331–1345. [CrossRef] [PubMed]

41. Hempel, D.; Ebner, F.; Garg, A.; Trepotec, Z.; Both, A.; Stein, W.; Gaumann, A.; Güttler, L.; Janni, W.; DeGregorio, A.; et al. Real World Data Analysis of next Generation Sequencing and Protein Expression in Metastatic Breast Cancer Patients. *Sci. Rep.* **2020**, *10*, 10459. [CrossRef] [PubMed]

42. Burris III, H.A.; Schlauch, D.; McKenzie, A.; Sharma, Y.; Spigel, D.R.; Jones, S.F.; Dilks, H.H. Adoption and Utilization of NGS-Based Molecular Profiling in Community-Based Oncology Practices: Insights from Sarah Cannon. *J. Glob. Oncol.* **2019**, *5*, 34. [CrossRef]

43. Sturgill, E.G.; Misch, A.; Lachs, R.; Jones, C.C.; Schlauch, D.; Jones, S.F.; Shastry, M.; Yardley, D.A.; Burris, H.A.; Spigel, D.R.; et al. Next-Generation Sequencing of Patients With Breast Cancer in Community Oncology Clinics. *JCO Precis. Oncol.* **2021**, *5*, 1297–1311. [CrossRef] [PubMed]

44. Eltokhy, M.A.; Saad, B.T.; Eltayeb, W.N.; El-Ansary, M.R.; Aboshanab, K.M.; Ashour, M.S.E. A Metagenomic Nanopore Sequence Analysis Combined with Conventional Screening and Spectroscopic Methods for Deciphering the Antimicrobial Metabolites Produced by *Alcaligenes faecalis* Soil Isolate MZ921504. *Antibiotics* **2021**, *10*, 1382. [CrossRef] [PubMed]

45. Orsini, P.; Minervini, C.F.; Cumbo, C.; Anelli, L.; Zagaria, A.; Minervini, A.; Coccaro, N.; Tota, G.; Casieri, P.; Impera, L.; et al. Design and MinION Testing of a Nanopore Targeted Gene Sequencing Panel for Chronic Lymphocytic Leukemia. *Sci. Rep.* **2018**, *8*, 11798. [CrossRef]

46. Minervini, C.F.; Cumbo, C.; Orsini, P.; Anelli, L.; Zagaria, A.; Impera, L.; Coccaro, N.; Brunetti, C.; Minervini, A.; Casieri, P.; et al. Mutational Analysis in BCR-ABL1 Positive Leukemia by Deep Sequencing Based on Nanopore MinION Technology. *Exp. Mol. Pathol.* **2017**, *103*, 33–37. [CrossRef]

47. Minervini, C.F.; Cumbo, C.; Orsini, P.; Brunetti, C.; Anelli, L.; Zagaria, A.; Minervini, A.; Casieri, P.; Coccaro, N.; Tota, G.; et al. TP53 Gene Mutation Analysis in Chronic Lymphocytic Leukemia by Nanopore MinION Sequencing. *Diagn. Pathol.* **2016**, *11*, 96. [CrossRef]

48. Minervini, C.F.; Cumbo, C.; Orsini, P.; Anelli, L.; Zagaria, A.; Specchia, G.; Albano, F. Nanopore Sequencing in Blood Diseases: A Wide Range of Opportunities. *Front. Genet.* **2020**, *11*, 76. [CrossRef]

49. Index of /Giab/Ftp/Data/NA12878/Ultralong_OxfordNanopore. Available online: https://github.com/genome-in-a-bottle/giab_data_indexes (accessed on 10 August 2022).

50. Index of /Giab/Ftp/Data/AshkenazimTrio/HG002_NA24385_son/Ultralong_OxfordNanopore/Guppy-V3.2.4_2020-01-22. Available online: https://github.com/genome-in-a-bottle/giab_data_indexes/blob/master/AshkenazimTrio/sequence.index.AJtrio_UCSC_ONT_UL_guppy-V3.2.4_2020-01-22 (accessed on 10 August 2022).

51. Olson, N.D.; Wagner, J.; McDaniel, J.; Stephens, S.H.; Westreich, S.T.; Prasanna, A.G.; Johanson, E.; Boja, E.; Maier, E.J.; Serang, O.; et al. PrecisionFDA Truth Challenge V2: Calling variants from short and long reads in difficult-to-map regions. *Cell Genom.* **2022**, *11*, 100129. [CrossRef]

52. Global Alliance for Genomics and Health·GitHub. Available online: https://github.com/ga4gh (accessed on 10 August 2022).

53. Edge, P.; Bafna, V.; Bansal, V. HapCUT2: Robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res.* **2017**, *27*, 801–812. [CrossRef] [PubMed]

54. Martin, M.; Patterson, M.; Garg, S.; Fischer, S.O.; Pisanti, N.; Klau, G.W.; Schöenhuth, A.; Marschall, T. WhatsHap: Fast and Accurate Read-Based Phasing. *bioRxiv* **2016**, 085050. Available online: https://pub.uni-bielefeld.de/record/2941794 (accessed on 8 August 2022). [CrossRef]

55. Shafin, K.; Pesout, T.; Chang, P.C.; Nattestad, M.; Kolesnikov, A.; Goel, S.; Baid, G.; Kolmogorov, M.; Eizenga, J.M.; Miga, K.H.; et al. Haplotype-Aware Variant Calling with PEPPER-Margin-DeepVariant Enables High Accuracy in Nanopore Long-Reads. *Nat. Methods* **2021**, *18*, 1322. [CrossRef]

56. Barbitoff, Y.A.; Abasov, R.; Tvorogova, V.E.; Glotov, A.S.; Predeus, A.V. Systematic Benchmark of State-of-the-Art Variant Calling Pipelines Identifies Major Factors Affecting Accuracy of Coding Sequence Variant Discovery. *BMC Genom.* **2022**, *23*, 155. [CrossRef]

57. Koboldt, D.C. Best Practices for Variant Calling in Clinical Sequencing. *Genome Med.* **2020**, *12*, 91. [CrossRef] [PubMed]

58. Supernat, A.; Vidarsson, O.V.; Steen, V.M.; Stokowy, T. Comparison of Three Variant Callers for Human Whole Genome Sequencing. *Sci. Rep.* **2018**, *8*, 17851. [CrossRef]

59. GitHub—Jts/Nanopolish: Signal-Level Algorithms for MinION Data. Available online: https://github.com/jts/nanopolish (accessed on 8 August 2022).

60. Vendrell, J.A.; Vilquin, P.; Larrieux, M.; Van Goethem, C.; Solassol, J. Benchmarking of Amplicon-Based Next-Generation Sequencing Panels Combined with Bioinformatics Solutions for Germline BRCA1 and BRCA2 Alteration Detection. *J. Mol. Diagn.* **2018**, *20*, 754–764. [CrossRef]

61. Feliubadaló, L.; Lopez-Doriga, A.; Castellsagué, E.; Del Valle, J.; Menéndez, M.; Tornero, E.; Montes, E.; Cuesta, R.; Gómez, C.; Campos, O.; et al. Next-Generation Sequencing Meets Genetic Diagnostics: Development of a Comprehensive Workflow for the Analysis of BRCA1 and BRCA2 Genes. *Eur. J. Human Genet.* **2013**, *21*, 864–870. [CrossRef]

62. Park, K.; Kim, M.K.; Lee, T.; Hong, J.; Kim, H.K.; Ahn, S.; Lee, Y.J.; Kim, J.; Lee, S.W.; Lee, J.W.; et al. Performance Evaluation of an Amplicon-based Next-generation Sequencing Panel for BRCA1 and BRCA2 Variant Detection. *J. Clin. Lab. Anal.* **2020**, *34*, e23524. [CrossRef]