A Novel Bioinformatic Approach for Large Deletion Detection in Multiplex Pcr-Based Ngs Assays

Fatemeh Abbaszadeh*, Shaza Abu Sirriya, Bincy Mathew, Ramin Badii, Vasiliki Chini, Zafar Nawaz & Susanna Akiki

Diagnostic Genomic Division, Department of Laboratory Medicine & Pathology, Qatar Rehabilitation Institute (QRI), Hamad Medical Corporation (HMC), Hamad Bin Khalifa Medical City, Doha-Qatar

Correspondence to: Dr. Fatemeh Abbaszadeh, Diagnostic Genomic Division, Department of Laboratory Medicine & Pathology, Qatar Rehabilitation Institute (QRI), Hamad Medical Corporation (HMC), Hamad Bin Khalifa Medical City, Doha-Qatar.

Copyright

Abstract

Background

Next-generation sequencing (NGS) has changed genetic diagnostics due to its high throughput, speed, and cost-effectiveness. This paper describes the challenges that were faced in detection of the frameshift variant c.1175_1214del in the *BRCA1* gene, undetectable by CNV analysis, using Oncomine BRCA Research Assay (OBRA) and Ion Torrent platform. This paper also describes the bioinformatic strategy implemented to successfully detect and annotate this large deletion. Frameshift variants comprise the highest proportion (>60%) of pathogenic variants identified in *BRCA1/2* genes and a false negative *BRCA1/2* result leads to the provision of inaccurate genetic counseling and has negative impacts on medical management and risk assessment for *BRCA*-related cancer patients and their at-risk relatives.

The *BRCA1* heterozygous c.1175_1214del was initially undetectable by the Ion Reporter software because the deletion coincides with the primer binding sites of two overlapping amplicons in the OBRA kit. This resulted in generation of low frequency super-amplicons during multiplex-PCR. To overcome this assay's primer design problem bioinformatically, we removed the default genomic coordinates for the relevant amplicons in the designed bed file for OBRA and replaced it with the corresponding coordinates for a super-amplicon. This resulted in overall reduction of read depth coverage as shorter reads derived from the normal allele were not counted in read depth calculation by the software. This read depth reduction, indirectly increased frequency of the deletion and facilitated its detection.

To further increase the software sensitivity for detection of this deletion without introducing false positives, we added the c.1175_1214del as a hotspot in the analysis workflow.

This bioinformatic approach is not limited to detection of deletions in BRCA1/2 genes using Ion Torrent platform. We believe it can be applied for detection of any frameshift variants in any genes that are sequenced by any NGS platforms that are compatible with multiplex-PCR library preparation. This strategy avoids the need for supplementary test to detect missed variants, redesigning of primers for the NGS assay, the wait for availability of the improved version of assay design in the market and reduces the risk of false negative results.

List of Abbreviations

CNV   copy number variant
HBOC   hereditary breast and ovarian cancer syndrome
Indel   insertion/deletion
NGS   next-generation sequencing
OBRA   Oncomine BRCA Research Assay
PCR   polymerase chain reaction
TAT   turnaround time

Introduction

Hereditary breast and ovarian cancer (HBOC) is an inherited cancer predisposition syndrome. The features suggestive of a hereditary cancer predisposition include: early onset of cancer, multiple primary cancers in a single individual, bilateral tumors, and a family history of cancer of the same spectrum in multiple generations along the same lineage. HBOC is characterized by an increased risk of breast cancer in both female and male, ovarian cancer in females, and to a lesser degree other cancer including prostate cancer, melanoma and pancreatic cancer.

Pathogenic variants in BRCA1 and BRCA2 highly penetrant genes account for 20-25% of hereditary breast cancers [1], 5-10% of all breast cancers [2], and around 15% of all ovarian cancers [3]. The demand for *BRCA* diagnostic genetic testing has steadily increased due to elevated public awareness of breast cancer, significantly increased risk of developing breast and/or ovarian cancer in individuals carrying *BRCA*

pathogenic variants compared to general population, and the availability of established management guidelines including routine surveillance, prophylactic measures and targeted therapies. This increasing demand for *BRCA* testing places a strain on diagnostic laboratories. To accommodate this pressure, and owing to the large size of the *BRCA* genes and no-mutation hot spots more diagnostic laboratories take advantage of advances in sequencing technologies, namely next generation sequencing (NGS), which allows for fast, scalable, and cost-efficient BRCA gene panel testing compared to traditional Sanger sequencing. However, transitioning from Sanger sequencing to NGS for *BRCA* in a clinical diagnostic laboratory is not an straight forward process and requires a significant validation to demonstrate the NGS assay including the wet procedure and bioinformatics analysis are precise, sensitive, specific and fit for the purpose prior to its adoption.

In our diagnostic laboratory, we validated *BRCA* NGS by Oncomine BRCA Research Assay (Life Technologies) using the Ion Torrent platform (Chef and Ion S5 sequencer) due to its locked down workflow and rapid turnaround time (TAT). This paper covers the challenges that were faced in detection of a heterozygous 40bp deletion in BRCA1 (c.1175_1214del (p.Leu392fs); (rs80359874)) in a known positive control sample during validation of this assay and describes the bioinformatic method that was adopted to detect and annotate this large deletion. It is worthwhile mentioning that this deletion is not long enough to be detected by CNV algorithms of the analysis software, further demonstrating the importance of the bioinformatic changes that result in detection of this deletion. Frameshift variants account for >60% of pathogenic variants identified in each of the *BRCA1/2* genes [4] and identification of these variants is essential for the provision of genetic counselling, medical management, and establishment of risk reduction strategies for *BRCA*-related cancer patients and their at risk family members.

we successfully identified the *BRCA1* c.1175_1214del in two patients with HBOC with the help of this bioinformatic method. This pathogenic variant has a rare population frequency (0.00001 in ExAC database) [5] but it is the tenth most frequent pathogenic variant identified in North American breast cancer families, accounting for 1% of such families [6].

This bioinformatic approach is not limited to detection of deletions in BRCA1/2 genes using Ion Torrent platform. We believe that the same or a similar approach can be applied for detection of any large frameshift variant below the amplicon size (typically <100 bp) in any genes that are missed detection by the analysis software due to problems in amplicon-based NGS assay designs, low depth of read coverage, or low variant frequency. As a substantial number of diagnostic laboratories use Illumina platform for diagnostic testing of BRCA1/2 genes, we checked the designed bed file for the AmpliSeq for Illumina BRCA Panel [7] and found out that the genomic coordinates for amplicons encompassing the *BRCA1* c.1175_1214del are the same as coordinates in OBRA bed file. Therefore, it is important to highlight the risk of a false negative result for the BRCA1 c.1175_1214del using this Illumine panel and that the described bioinformatic method has the potential for detection of this variant in samples sequenced on Illumina platform as well.

Material and Methods

A total of 43 samples were used to establish and validate a processing and analytical workflow for next generation sequencing (NGS) of BRCA genes using the Oncomine BRCA Research Assay and Ion S5

sequencer. These included 28 reference DNA samples with known pathogenic variants in BRCA genes and 15 negative samples. Amongst the positive reference samples was a sample with 40bp deletion in BRCA1 (c.1175_1214del), which was purchased from Coriell (NA14094) (https://www.coriell.org/0/Sections/Search/Sample_Detail.aspx?Ref=GM14094&PgId=166).

Genomic DNA (gDNA) for Ion Torrent sequencing was extracted from clinical samples using Maxwell® 16 blood DNA purification kit (Promega, USA) following manufacturer's instructions. DNA was suspended in 220µL of elution buffer. DNA quality was estimated using NanoDrop Spectrophotometer and quantity was assessed using the Qubit Fluorometer (Life Technologies) and the Qubit dsDNA HS (High Sensitivity) Assay Kit following kit instructions. According to Oncomine™ BRCA Research Assay kit protocol, 10ng of gDNA for each sample was used for library preparation with the Ion AmpliSeq™ Library Kit Plus (ThermoFisher Scientific, USA) and Oncomine™ BRCA Research Assay Chef-ready Primer Panel (ThermoFisher Scientific, USA). This two premixed primer pool panel amplifies a total of 265 amplicons in BRCA1 & BRCA2 genes, providing a sequence coverage performance of at least 600X for 32 samples in one Ion 520 chip, containing 12,530,194 wells. Coverage uniformity of >99.5%, and mean depth coverage of >600 reads were achieved, providing at least 50x read depth coverage for all BRCA exons as well as 64 bases of flanking intronic regions. Ion Chef System (Thermo Fisher Scientific, USA) was used for library preparation following manufacturer's instructions. Barcoded libraries were generated using Ion AmpliSeq Chef Solutions DL8 Kit (Thermo Fisher Scientific, USA) and an Oncomine™ BRCA Research Assay panel (Thermo Fisher Scientific, USA). Barcoded equimolar (~100pM) libraries were diluted and pooled to a final concentration of 33pM. Using an Ion 510™ & Ion 520™ & Ion 530™ Kit – Chef, clonal amplification of the libraries was carried out by emulsion PCR (Thermo Fisher Scientific, USA). Prepared & loaded libraries were then sequenced over an Ion S5 Prime Sequencer using an Ion 520 Chip and its corresponding Ion sequencing Kit (Thermo Fisher Scientific, USA).

A custom workflow in Ion Reporter analysis software (version 5.10, Thermo Fisher Scientific, USA) was designed to align raw reads to hg19 human reference to call variants through variant caller (version 5.10) plugin and to annotate variants based on different databases as described in Table 1. A custom filter chain to filter benign and likely benign variants was established. Integrative Genomics Viewer was used to manually review all detected variants.

**Table 1:** *Databases used for variant annotation from Ion Reporter software*

Databases	Version
ExAC	31
5000 Exomes	20161108
phylop	20160919
cosmic	82
refgeneScores	7
clinvar	20180225
refgene	83

dbsnp	150
pfam	31

List of databases that were used for annotation of germline variants in BRCA1/2 genes.

Results

Sample NA14094 is the positive reference DNA, which contains the heterozygous c.1175_1214del in the *BRCA1* gene. Testing this sample with OBRA kit achieved 100x depth of coverage for all targeted amplicons and nucleotides in *BRCA* genes (Figure.1).
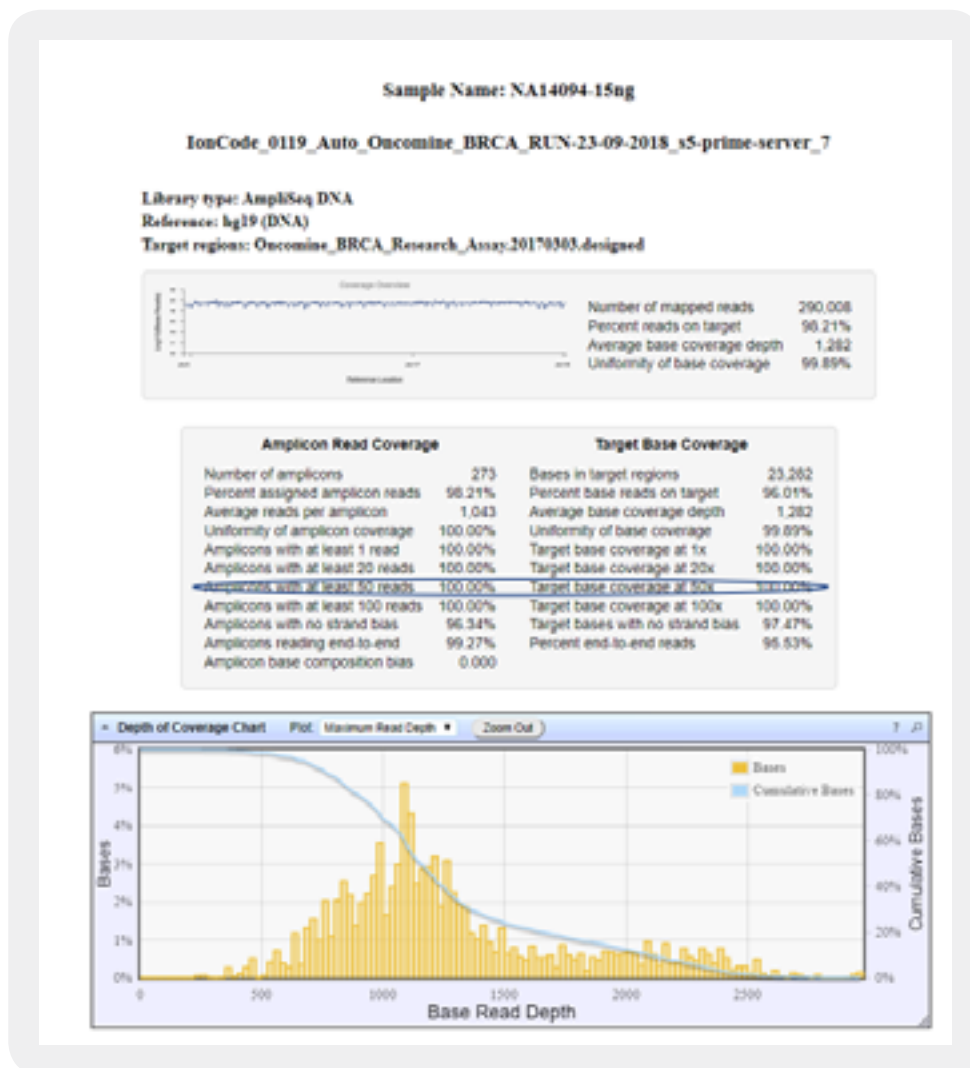

Figure 1: *Coverage analysis report of BRCA genes for sample NA14094. Amplicon Depth of coverage for sample NA14094. All amplicons in BRCA genes achieved depth of coverage of 50X. The base coverage uniformity for all the amplicons was 99.89%.*

However, Ion Reporter failed to detect and annotate the c.1175_1214del in BRCA1 gene in this sample. Furthermore, this variant was not present in the list of filtered out variants. To rule out the possibility of a sample swap this sample was tested by Sanger sequencing and the variant was detected correctly (Figure. 2).

Figure 2: *BRCA1 c.1175_1214del by Sanger Sequencing. Electropherogram showing detection of BRCA1 c.1175_1214del in sample NA14094 by Sanger Sequencing.*

To investigate the reasons that this variant was not detected by Ion reporter we browsed the chromosomal location (chr17:41,246,334-41,246,373) (GRCh37) of this variant in Integrative Genomic Viewer (IGV V.2.1, Broad Institute, Cambridge, Massachusetts, USA) [8]. As shown in Figure.3, the deletion coincides with the binding sites of primers for two of the assay's overlapping amplicons (OBRA_BRCA1_81 and OBRA_BRCA1_82) and creates two types of super- amplicons consisting of OBRA_BRCA1_81 and OBRA_BRCA1_82 and OBRA_BRCA1_83 amplicons or OBRA_BRCA1_81 and OBRA_BRCA1_82). However, due to their large size, the frequency of these super amplicons is low (approx. 11%). This is while the default minimum frequency required for indel detection in Ion Reporter for this assay is 25%.
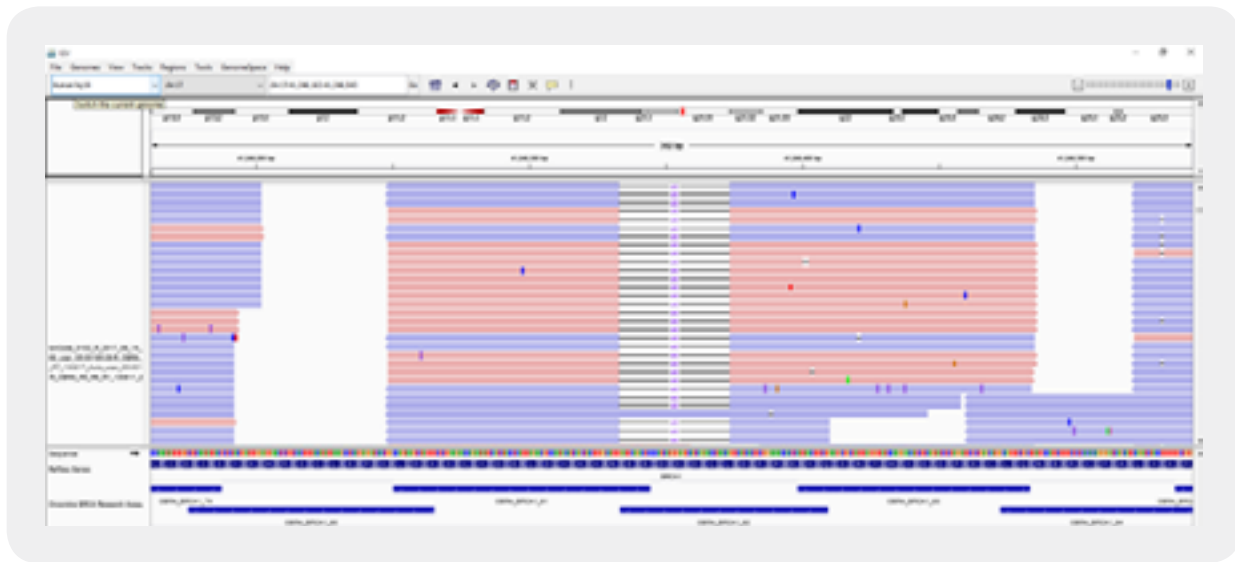
Figure 3: *The c.1175_1214del variant coincides with primer binding site of overlapping amplicons in OBRA. The c.1175_1214del deletes the primer binding sites of two of the OBRA's amplicons (OBRA_BRCA1_81 and OBRA_BRCA1_82), creating two types of super amplicons consisting of OBRA_BRCA1_81 and OBRA_ BRCA1_82 and OBRA_BRCA1_83 amplicons or OBRA_BRCA1_81 and OBRA_BRCA1_82).*

To enable detection of this variant we modified the designed bed file that was provided by Life Technologies for OBRA assay. Table 2 shows the original chromosomal coordinates for amplicons OBRA_BRCA_81 to OBRA_BRCA_83. We created a super-amplicon id (OBRA_BRCA1_81_superamplicon) that covers amplicons 81-83 with coordinates from 41246250-41246483 and replaced the original chromosomal coordinates for amplicons 81-83 in the designed bed file with the coordinates for this super amplicon. This led to the successful detection of this heterozygous 40bp deletion (Figure. 4).

Table 2: *Chromosomal coordinates for OBRA amplicons 81-83 for BRCA1*

Amplicon ID	Chromosome	Position start	Position end
OBRA_BRCA1_81	17	41246250	41246344
OBRA_BRCA1_82	17	41246333	41246409
OBRA_BRCA1_83	17	41246398	41246483

The original chromosomal coordinates for amplicons OBRA_BRCA_81 to OBRA_BRCA_83 in the designed bed file by Life Technologies for OBRA assay.

Figure 4A
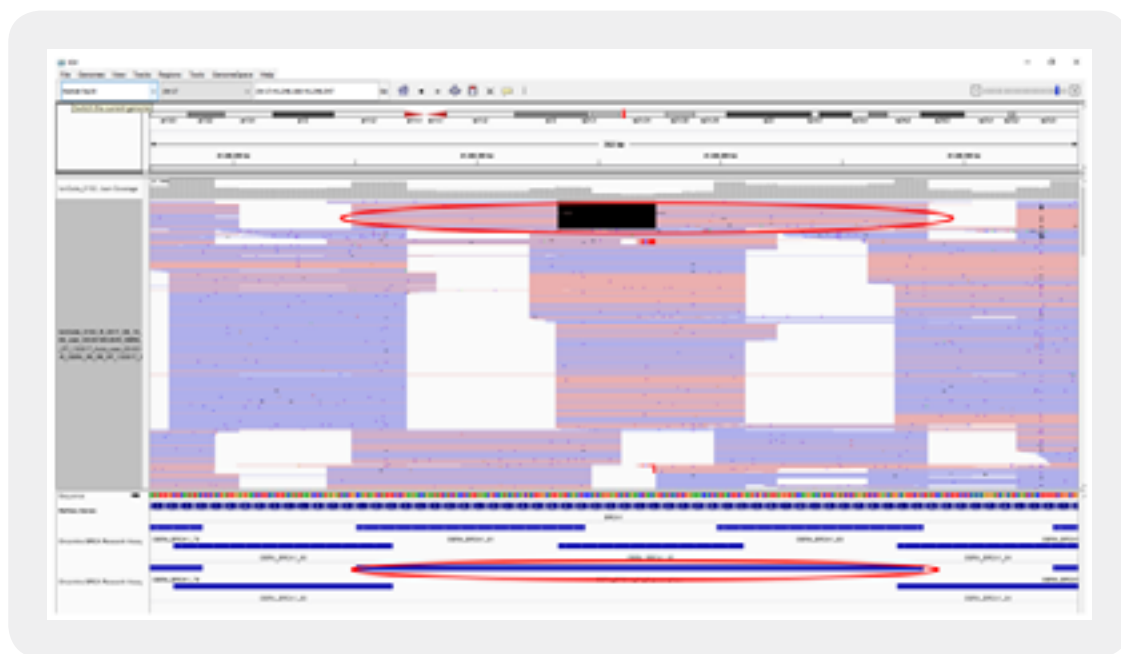
Figure 4B

Figure 4: *Creation of super amplicon consisting of amplicons 81–83 and detection of c.1175_1214del variant . A) Demonstrates successful detection of c.1175_1214del40 variant following creation of super Amplicon consisting of amplicons 81–83 in the bed file. B) Shows the reduced depth for super amplicon coverage as shorter reads are excluded by Ion reporter due to low quality.*

To further increase the detection rate of this large deletion we introduced the c.1175_1214del variant as a hotspot in the analysis workflow. The default minimum variant frequency for hotspot indels is 10%. This approach increased the software detection sensitivity for this variant only and minimized the risk of introducing any false positive results that may have been detected if the frequency for all indels in the workflow was reduced.

Discussion

We investigated the reason for lack of detection of *BRCA1* c.1175_1214del heterozygous variant in the Coriell NA14094 reference sample that was used for validation of Oncomine *BRCA* Research Assay (OBRA) using Ion Chef and Ion S5 sequencer in our diagnostic laboratory.

OBRA uses a massively multiplex PCR for library preparation. We demonstrated through IGV that the *BRCA1* c.1175_1214del deletes the primer binding sites of two of the assay amplicons (OBRA_BRCA1_81 and OBRA_BRCA1_82) and therefore these two amplicons do not individually amplify efficiently on the allele containing the deletion. However, this deletion causes the forward primers for amplicon 81 and the reverse primers for amplicon 82 to amplify a super-amplicon. Also, the forward primers of amplicon 81 and the reverse primers of amplicon 83 amplify another super-amplicon on the deleted allele. The frequency of the generated super-amplicons (11%) is below the default minimum required variant frequency for detection of indels, which is 25%.

We tried to overcome this primer design problem of the assay, bioinformatically by removing the default genomic coordinates for amplicons 81, 82, and 83 in the designed bed file that was provided by Life Technologies for OBRA and replacing it with the custom-made coordinates for a super-amplicon consisting of these three amplicons. This resulted in reduction of read depth coverage as all shorter reads created by the normal allele were not recognized by the software and not counted in read depth of coverage calculation. Because of this read depth reduction, the deletion frequency automatically increased, elevating the chance of its detection by the software.

To further increase the chance of detection of this large deletion we introduced the c.1175_1214del variant as a hotspot in the analysis workflow. The default minimum variant frequency for hotspot indels is 10%. This approach increased the detection sensitivity of the Ion Reporter software for this particular variant without the risk of introducing many unnecessary false positives that may have been detected by overall reduction of frequency of all indel in the workflow.

Conclusion

The described approach in this paper can be adopted when encountering technical design problems in any multiplex PCR-based NGS assays resulting from large indels that coincide with primer binding sites of some amplicons. Such assay design problems may lead to low depth of read coverage or low variant frequency and false negative results. This approach minimizes the risk of false negative results and the need for redesigning of the assay, the wait for the next version of improved assay to become available in the market, or the need for a supplementary test to detect missed variants.

Acknowledgements

Conflict of Interests

None declared

Bibliography

1. Easton, D. F. (1999). How many more breast cancer predisposition genes are there? *Breast Cancer Res., 1*(1), 14-17.

2. Campeau, P. M., Foulkes, W. D. & Tischkowitz, M. D. (2008). Hereditary breast cancer: new genetic developments, new therapeutic avenues. *Human Genetics, 124*(1), 31-42.

3. Pal, T., Permuth-Wey, J., Betts, J. A., Krischer, J. P., Fiorica, J., Arango, H., *et al.* (2005). BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. *Cancer, 104*(12), 2807-2816.

4. Tsiolkas, C., Kouris, A., Chapple, C. E., Aguilera, M. A., Meyer, R. & Massouras, A. (2019). VarSome: The Human Genomic Variant Search Engine. *Bioinformatics, 35*(11), 1978-1980.

5. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res., 29*(1), 308-311.

6. Rebbeck, T. R., Friebel, T. M., Friedman, E., Hamann, U., Huo, D., Kwong, A., *et al.* (2018). Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations. *Hum Mutat., 39*(5), 593-620.

7. AmpliSeq for Illumina BRCA Panel Product Files.

8. Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., *et al.* (2011). Integrative genomics viewer. *Nat Biotechnol., 29*(1), 24-26.