ISSN 0120-4157

Biomédica

Revista del Instituto Nacional de Salud

Citación provisional:

Gouvêa Moreira TC, da Silva Spínola P, Campos Rezende M, Moreira de Freitas CS, Borges Mury F, Rodrigues Bonvicino C. Correlation between number of false positive variants and quality of results in Ion Torrent PGM Sequencing to screen BRCA genes. Biomédica. 2021;41 (4).

Recibido: 31-07-20

Aceptado: 02-08-21

Publicación en línea: 18-08-21

Correlation between number of false positive variants and quality of results in Ion Torrent PGM Sequencing to screen *BRCA* genes

Correlación entre el número de variantes de falsos positivos y la calidad de los resultados en la secuenciación de Ion Torrent PGM para seleccionar genes *BRCA*

Analysis of different protocols with Ion Torrent PGM sequencing

Tiago César Gouvêa Moreira [1,2], Pricila da Silva Spínola [3,4], Micheline Campos Rezende [1,5], Carla Simone Moreira de Freitas [1,5], Fábio Borges Mury [6], Cibele Rodrigues Bonvicino [3,4], Luciana de Andrade Agostinho [1,2,7]

[1] Hospital do Câncer de Muriaé, Fundação Cristiano Varella, Muriaé, Brazil

[2] Centro Universitário UNIFAMINAS, Muriaé, Brazil

[3] Divisão de Genética, Instituto Nacional do Câncer, Rio de Janeiro, Brazil

[4] Programa de Pós Graduação em Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

[5] Instituto de Ensino e Pesquisa Santa Casa BH, Belo Horizonte, Brazil

[6] Thermo Fisher Scientific, São Paulo, Brazil

[7] Programa de Pós Graduação em Neurologia, Universidade Federal do Estado do Rio de Janeiro Rio de Janeiro, Brazil

Corresponding author:

Luciana de Andrade Agostinho, Av. Cristiano Ferreira Varella, 555 - Universitário, Muriaé, Brazil

Phone: 055 32 998181209

polucita@yahoo.com.br

Author contributions

Fábio Borges Mury: methodology, conceptualization and data curation

Tiago Cesar G. Moreira: resources, investigation, formal analysis; methodology, conceptualization, data curation.

Luciana de Andrade Agostinho participated in all stages of the research.

All authors participated in validation of results, writing and approval of the paper.

Introduction: Next Generation Sequencing (NGS) is cost-effective, capable to investigate the genes faster but the protocol has challenges throughout its steps.

Objective: We investigated different adjustments of protocol to screen the *BRCA* genes using Ion Torrent PGM sequencing and correlate the results with number of False Positive (FP) variants.

Material and methods: Library preparation process, number of FP InDels, library concentration, number of cycles in amplify targets step, purity of nucleic acid, input, and number of samples/Ion 314 chip were analyzed in association with the results obtained by NGS.

Results: 51 reactions and 9 adjustments of protocols were done and 8 FP InDels were observed in homopolymer regions. No FP Single-Nucleotide Polymorphism variant was observed. Protocol variables jointly are associated in 67.5% with the quality of results obtained($p<0.05$). The number of FP InDels decreased when the quality of results increased.

Conclusion: Ion AmpliSeq BRCA1/BRCA2 Community Panel had better performance with 4 samples per Ion-314 chip instead of 8 and the number of cycles in the amplification step, even using DNA with high quality, was better with 23. We observed better results when the manual equalization process was done, without the Ion Library Equalizer kit. These adjustments provided higher coverage of the variants and fewer artifacts (6.7-fold). Laboratories must perform the internal validation because FP InDel variants can vary according to the quality of results. NGS assay must be validated with Sanger.

Keywords: Sequence analysis; DNA; high-throughput nucleotide sequencing; genes, BRCA1; genes, BRCA2.

Introducción. La secuenciación de nueva generación (NGS) es rentable, capaz de investigar los genes más rápido, sin embargo, el protocolo tiene desafíos a lo largo de sus pasos.

Objetivo. Investigar diferentes ajustes de protocolo para seleccionar los genes *BRCA* usando secuenciación de Ion Torrent PGM y correlacionar los resultados con el número de variantes de falso positivo (FP).

Materiales y métodos. El proceso de preparación de la biblioteca, el número de FP InDels, la concentración de la biblioteca, el número de ciclos en el paso de amplificación de objetivos, la pureza del ácido nucleico, la entrada y el número de muestras/chip del Ion-314 se analizaron en asociación con los resultados obtenidos por NGS.

Resultados. Se realizaron 51 reacciones y 9 ajustes de protocolos y se observaron 8 FP InDels en regiones de homopolímeros. No se observó ninguna variante de polimorfismo de nucleótido simple FP. Las variables de protocolo en conjunto se asocian en 67.5% con la calidad de los resultados obtenidos($p<0.05$). El número de FP InDels disminuyó cuando la calidad de los resultados aumentó.

Conclusiones. El panel comunitario Ion AmpliSeq BRCA1/BRCA2 tuvo un mejor rendimiento con 4 muestras/chip Ion-314 en lugar de 8 y el número de ciclos en el paso de amplificación, incluso con ADN de alta calidad, fue mejor con 23. Observamos mejores resultados cuando se realizó el proceso de ecualización manual, sin el kit Ion Library Equalizer. Estos ajustes proporcionaron una mayor cobertura de las variantes y menos artefactos. Los laboratorios deben realizar la validación interna porque las variantes de FP InDel pueden variar según la calidad de los resultados. El ensayo NGS debe validarse con Sanger.

Palabras clave: análisis de secuencia, ADN, secuenciación de nucleótidos de alto rendimiento; genes BRCA1; BRCA2

Mutations in *BRCA1/BRCA2* genes are associated with breast, ovarian, prostate and pancreatic cancer (1). It is not clear whether these genes increase the risk of Colorectal Cancer (2).

The increasing demand for genetic testing diagnostic in a clinical setting has created the need for an alternative technology than Sanger sequencing. Next Generation Sequencing (NGS) is cost-effective, capable to investigate the genes faster (3). However, the NGS protocol is still expensive and presents challenges throughout its steps (4).

NGS analysis has good application in Single-Nucleotide Polymorphism (SNP) investigation, InDel and genomic rearrangements (5). Homopolymer regions may have imprecise results such as strand bias and low quality in determining the variants investigated (6). Sanger sequencing is considered a gold standard to validate mutations in these regions (7).

The *BRCA1/BRCA2* are considered high-penetrance dominant autosomal genes for breast cancer (BC) susceptibility responsible for 25% of the risk for familial BC (8-10). The miscegenation of the Brazilian population makes important the investigation of these genes (10).

NGS technology is a process that requires configurations to improve analytical efficiency, as well as its sensitivity and specificity (3). This study investigated different adjustments of protocol to screen the BRCA genes using the Ion Torrent PGM NGS technology and correlated the results with the number of FP variants.

Material and Methods

This is a transversal and observational study.

Patients

Patients were from the Cancer Hospital from Muriaé city, Brazil. The inclusion criteria were the presence of breast (n=29), colorectal and rectum (n=4 and n=2, respectively), prostate (n=3) and/or ovary (n=1) cancer at any age with familial or sporadic history. One negative control sample was also used. Although the association between colorectal cancer and the *BRCA1* and *BRCA2* genes is weak, we did not exclude these individuals (2).

This project was approved by the FAMINAS-Muriaé Ethics Committee (62262416.3.00005105) and the exclusion criteria were individuals younger than 18 years.

Sample selection and DNA extraction

Forty peripheral blood samples were collected in two EDTA tubes. DNA extraction was performed with QIAamp® DNA Mini Kit (Qiagen). DNA quantification was performed on Qubit 3.0 fluorometer using dsDNA BR Assay kit and DNA qualification on Nanodrop spectrophotometer (Thermo Fisher Scientific). A sample with a pathogenic variant as the positive control was used and three samples were validated by the National Cancer Institute (INCA) in Brazil, Division of Genetics.

Sample preparation and Ion Torrent PGM Sequencing

DNA inputs were tested with 10-30 ng to construct the libraries manually using the Ion Ampliseq Library kit 2.0. The number of cycles to amplify the targets was 19, 21 and 23 and *Ion AmpliSeq BRCA1 and BRCA2 Community Panel (Thermo Fisher Scientific)* were used with 167 primer pairs (3 primer pools).

Then, the amplicons were partially digested and the barcodes were inserted and the sample purified.

The libraries were purified using Agencourt Ampure XP Beads (Beckman Coulter), quantified and equalized with the Ion Library Equalizer kit or manually diluted with ultrapure water. All libraries were quantified by qPCR using the Ion Library Quantitation Kit and some samples were quantified by fluorimetry using Qubit dsDNA HS Assay Kit.

Multiplexed barcoded libraries were amplified by PCR emulsion using the PGM Hi-Q View Chef kit and Ion 314 chip. The PCR emulsion was performed in Ion Chef equipment with consecutive analysis in the PGM, for the sequencing step.

Variants selection and quality parameters

Variants were called by the Variant Caller plugin version 5.6 and compared to the genome version GRCh37/hg19 in Torrent Suite and Ion Reporter algorithms.

We tested different adjustments of protocols to screen the *BRCA1/BRCA2* genes by NGS, we tried different factors as the input, the library preparation process, the number of samples/ Ion 314 chip and the number of cycles in the *Amplify Targets* step. The parameters evaluated at the final runs were: On target (%), Mean depth (%), Uniformity and Mapped reads (%) to analyze the results of protocols performed.

All targeted coding exons and exon–intron boundaries 20 base pairs (bp) of *BRCA1*/*BRCA2* genes were analyzed. NGS assay used is not recommended to detect large deletions and duplications variants and MLPA assay is suggested to investigate CNV variants. In this study, all pathogenic, some benign and variants of uncertain significance (VUS) were confirmed by Sanger.

Variant data analysis

Torrent Suite software version 5.2 was used to analyze the amplicons and to generate run metrics and quality. We used Generic-PGM-Germline Low stringency as a parameter to variant calling (cutoff with 15 reads for InDels and 6 for SNPs). Only variants with a minimum average base quality with 30 Phred Score were selected. The variants selected must have a number ≥25 when the allele coverage is multiplied by the zygosity, according to recommendation by Lih et al. (2017) and the NCI-MATCH NGS assay to ensure the confidence in variant calls (9).

The homozygous alleles must have a minimum value of 25 reads when the allelic frequency was 90-100%, and the heterozygous alleles, 50 reads when the frequency was 40-60% to be selected as a valid variant in this study. All the variants with coverage lower than 6 reads were also analyzed by Sanger.

Statistical analysis

The descriptive statistical analysis was performed by SPSS software version 17. To compare the quality factor among the different protocols tested, a non-parametric test for independent samples was performed. Multiple Regression analysis was performed to associate the quality value of each protocol tested with the Library preparation process, number of FP InDels, library concentration in pM, number of cycles in amplify targets step, 260/230 and 260/280 ratios of DNA, DNA Input and number of samples/Ion 314 chip. Spearman's test was used to correlate the number of InDel variants, the number of extension cycles and the number of samples/Ion 314 chip with the quality of results. Independent-samples median test compared the results obtained between the protocols. The results were considered statistically significant when $p<0.05$.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the Brazilian ethical standards as Resolution number 466/2012 and the CEP-Faminas research committee (CAAE number 62262416.3.00005105) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Results

Library and Template preparation for NGS Sequencing

Nine different adjustments of protocols were performed in *BRCA1/BRCA2* genes. In the first protocol performed (n=7) the sequencing chemistry used in PGM was the PGM Hi-Q Chef kit (*Thermo Fisher Scientific*). The Amplify targets step was applied with 19-cycle extension considering the high quality of DNA samples. In 4/7 samples we observed two amplicons with coverage <16 reads. The low coverage amplicons were AMPL225505032 and AMPL223390724, in exon 23 and exon 20, respectively.

In the second protocol, the extension cycle in the amplify target step (with +2) and the concentration of library in pM were increased in an attempt to improve the quality of results and the mean read length. The chemistry used was the Ion PGM Hi-Q View Chef kit. Among the samples analyzed in this protocol (n=29), only one had their amplicons covered properly. 26 samples had low coverage (<16 reads) in amplicon AMPL225505032 and 27 in AMPL223390724. Three samples had low coverage also in amplicon AMPL224626553 and one of them in AMPL225316548.

The third protocol had a cycle extension with 23 units, used to amplify DNA with low quality. In 2/3 of samples analyzed, a higher mean read length was observed.

All samples analyzed (n=3) had amplicons with low coverage. In 2/3 samples, the coverage kept low in both amplicons: AMPL225505032 and AMPL223390724.

In the fourth protocol onwards, we loaded four samples in the Ion 314 chip, instead of eight samples. All samples (n=3) had low coverage in both AMPL225505032 and AMPL223390724 amplicons. One of them was a positive control sample with a pathogenic SNP mutation, reanalyzed in one external laboratory as reported in our methodology.

In the fifth protocol, three samples used in protocol 2 were reanalyzed with four samples per chip and all of them had low coverage in both AMPL225505032 and AMPL223390724 amplicons.

The library preparation of protocols 1 to 5 was performed using the Ion Ampliseq Library kit 2.0 and Ion Library Equalizer kit. It is important to mention that the number of samples analyzed per chip is a different number of samples analyzed per protocol, the first one is only a run configuration.

In protocol number six, only one sample already analyzed in protocol 2, was tested. The library preparation was performed with all steps of the Ion Ampliseq Library kit 2.0, including library enrichment step with Platinum PCR SuperMix Hifi and Equalizer Primers, different from previous protocols. The final library concentration was approximately 77 pM, and the library equalization process was manually performed with dilutions with ultrapure water.

In the seventh protocol, one sample was used, and the library preparation was performed using Ion Ampliseq Library kit 2.0. The enrichment step with Platinum PCR SuperMix Hifi and the Equalizer Primers was not performed. The sample was diluted to approximately 73 pM and had low coverage in both AMPL225505032

and AMPL223390724 amplicons. Mapped reads and Mean Depth decreased despite maintaining the same number of variants observed.

In the eighth protocol, two samples already investigated in protocol 2 were used, with a larger input (20 ng/reaction) and the library preparation was according to protocol 6. Only one sample had low coverage in the amplicon AMPL223390724 and the quality factor increased.

In the ninth protocol, the same samples of protocol eight were analyzed, however, with higher input and library concentration with 30 pM. Both samples analyzed had low coverage in the amplicon AMPL223390724 and the mean read length increased. Protocols 8 and 9 were considered to be routinely applied in our Molecular Biology laboratory, taking into account their higher quality, absence of false positive variants and higher mean read length of amplicons.

All adjusted protocols analyzed were presented in the table 1.

Although the same variants were observed between protocols 2 and 6, using the same sample, all quality parameters improved, except the uniformity, with approximately 95% in both runs. This sample had low coverage in both amplicons, AMPL225505032 and AMPL223390724 and the mean read length increased.

The Ion NGS workflow failed to obtain an average coverage depth >20X in a small number of amplicons, but in several samples (mainly AMPL22550532 in exon 23 and AMPL223392219 in exon 20 of *BRCA2*). IGV visual inspection and Sanger sequencing of those regions confirmed the base calls.

All protocols had amplicons lower than 50bp (figure 1). The amplicons smaller than 50bp decreased when we increased the input and the number of extension cycles, and when 4 samples/chip were applied instead of 8 samples.

It was observed a 6.7-fold increase in the quality factor in protocol 9 (p<0.05) when compared to protocol 1, and between 1 and 8, it was 11.3 times higher in 8 (p=0.03).

Quality parameters of protocols

The protocols tested had their quality of results evaluated according to four variables: On target (%), Mean depth (%), Uniformity and Mapped reads (%) (figure 2). Base coverage with 20x was constant in all protocols from 96,6 to 99,4% (median 99,2%) and 100x base coverage varied from 70,3 to 99,3% (median 96,2%).

In order to make the statistical analysis easier, the variables On target and Uniformity were divided by 100 and, to obtain the quality factor, the four parameters were multiplied and divided by 1000 to generate a single quality value per protocol (table 2).

All protocols had the parameter AQ17 with 100% and the AQ20 ranged from 95.5 to 97.9%. According the base coverage, at least 96% of the bases were observed with a coverage of 20x.

Variants

Among the 40 investigated samples (39 affected patients and 1 negative control) distributed in 51 reactions and 9 protocols, the last two presented better quality and absence of FP InDel variants (table 3). The number of samples tested per protocol adjusted was not standard or larger because these protocols were performed to validate the *BRCA1* and *BRCA2* test using NGS in our laboratory. The use of these approaches for the genetic testing is complex, time consuming, expensive and requires extensive technical labor. 75 different variants (results not shown) in 40 samples were observed, in the total there were 567, also accounting for the

reproducibility test. The results obtained by Sanger detected 8 FP InDels, including 5 different variants, all of them in homopolymer regions. It is important to mention that all variants observed were not described in this paper as we had focused on the technical field of NGS assay. We reported variants suspected as false positives and associated with possible artifacts focused on the necessity to improve our results and to decrease the number of FP InDels in NGS assay.

Three samples were validated in INCA laboratory (Genetics Division), through the full reanalysis of these samples. There was no contradiction between the results. All variants determined as false positives were selected by the inclusion criteria and were confirmed by Sanger.

In the first protocol, two InDel variants were observed in homopolymer regions and all of them were excluded by the Sanger sequencing as FP. The chemistry used in PGM was the PGM Hi-Q Ion Chef kit.

In the second protocol, five InDel variants were observed, five of them were in homopolymer regions and all of them were considered FP. In the third protocol, one FP InDel was observed in the homopolymer region.

Until protocol 3, 12 samples tested showed the variant rs80359770 (c.956_957insA) at chr13:32906565 (GRCh37.p13). The same variant was investigated by Sanger and was excluded as FP InDel, despite the mean value of the Phred QUAL Score with 642.349 and 96 reads reporting an insertion of A (Ref:C, Observed Allele:CA/CA) (figure 3).

Quality parameters from protocol four onwards, no FP InDel was observed, and the quality parameters improved. Therefore, when we inserted 4 samples/Ion 314 chip, *BRCA2* variants as chr13:32906547 (c.937_938insT) and chr13:32906565

(c.956_957insA) were not observed again. Sample 9 was analyzed in two protocols (3 and 4), and in the last protocol the FP InDel variant disappeared.

Among the variants selected as true using the validation parameters, 29 SNPs showed coverage between 76 and 407 reads. These variants were also reanalyzed and confirmed by Sanger and none of them were in a homopolymer region.

In 8 InDels observed in protocols 1-3, with coverage of 51 to 124 reads, all of them were in homopolymer regions. All variants were excluded as FP by Sanger and IGV analysis.

All 44 genetic regions (SNP) validated by Sanger, no variant was found as False Negative when compared with NGS results.

Variants observed as *NO CALL* (considering quality score, Coverage <3 and realignment errors) were validated by Sanger to confirm the validation parameters applied in our laboratory. *No call* variants were observed in chr13:32907304, 32945109, 32945115 and 32945129, chr17:41222985 and 41223001. These excluded variants, with strand bias and low coverage, appeared with high frequency in our sample and were not observed when analyzed by Sanger.

The parameters applied to include/exclude a variant can increase sensitivity and decrease the number of FP variants called, mainly in homopolymer regions. To solve the low coverage of some amplicons (<16) using the *BRCA1/BRCA2* panel, all samples were fully sequenced in exons 20 and 23 in *BRCA2* gene.

Regarding the FP InDel variants included by the validation parameters and excluded by Sanger, we increased the cutoff for these variants. We analyzed by Sanger more than one sample bearing these InDel FP variants to find out a true threshold value considering low and high quality runs (table 3). The variants InDel

FP excluded by the validation parameters were not counted as FP variants (according Variant Data Analysis described in our Material and Methods).

The zygosity must be also considered to exclude a variant as FP InDel when compared in world scientific data available.

Correlation between the variants and the Quality parameters

The library preparation process, number of FP InDels, library concentration in pM, number of cycles in amplify targets step, 260/230 and 260/280 ratios of DNA, DNA Input and number of samples/Ion 314 chip were analyzed in association with the quality of results. We observed that these variants jointly are correlated to (67.5%) the quality of results obtained (R^2= 0.67 and p<0.05).

There was a statistically significant difference between the quality of results obtained in all protocols tested (p=0.03). Protocols 8 and 9 showed high quality parameters as Mapped reads, Uniformity, On target and Mean depth (2 samples tested in each one). Protocol 9 (input with 30 ng) had more variation in Mapped reads and Mean depth than protocol 8 (input with 20 ng). Protocols 6 and 7 were performed with 1 sample each one to test different Equalizaton processes. In protocol 6 (sample 25), after the Library purification step, the sample was amplified before quantification to enrich amplifiable material to obtain sufficient sample for accurate quantification. In protocol 7 (sample 16), after the Library purification step, we diluted the sample in 50 µL of TE and did not perform the enrichment. As the enrichment step showed better quality, we followed the next samples performing this sequence of steps. This step is described in the User guide (Thermo Fisher Scientific) in Option 3: Quantify the amplified library with the Qubit Fluorometer instrument.

There was no correlation between the number of FP InDel and the quality of results calculated and named as Quality factor (p=0.14), although the number of FP InDel variants decreased when the quality factor increased. No FP SNP variant was observed.

There was a weak inverse correlation of 33% between the quality of results (quality factor) and the number of samples/ Ion 314 chip (p=0.01) and a positive correlation between the quality factor and the number of extension cycles (r=0.50, p<0.05), even though the high quality of DNA used. The number of samples per chip is weakly correlated with the number of FP InDel (r=0.37, p=0.007).

Discussion

The panel used in this study proved to be efficient in covering all exons and a part of the introns, but there is great variability in amplification efficiency of the 167 targets. Thus, a high value of mean coverage is essential to ensure that even regions of lower efficiency in the PCR are represented in a minimum cutoff in the sequencing data. This presentation was especially evident in our clinical cohort and confirmation by Sanger is needed for the regions with poor coverage (<20X). The number of cycles in the amplification step, for samples with high quality, is 19 cycles (as the User guide) and even using DNA samples with high quality, the best performance was observed with 23 cycles. The online Chip Calculator (Ion Ampliseq designer) recommends inserting 8 samples per Ion 314 chip, in contrast, 4 samples per chip had better quality and fewer artifacts. We observed better quality of runs when we applied a manual equalization process without this kit without the Ion Library Equalizer kit.

The American College of Medical Genetics guidelines recommend to analyse the performance of different types of variants separately. We validated with Sanger, not only the variants with low quality parameters, but also variants with high quality, SNPs and InDels, separately. This step is important to the accuracy assay and to distinguish FP from genuine variants (11).

InDel variants are a challenge for NGS, mainly variants located in homopolymer regions (11,12), as we observed in our study. We confirmed FP InDel variants by Sanger observed with high quality in NGS.

According to (13), the quality of targeted NGS of a disease-specific subset of genes is equal to the quality of Sanger sequencing, although, our study and (11,12) observed some FP variants, mainly InDel, that need to be validated by Sanger. And different from others, this study reported all steps of validation process starting from the beginning. Different number of FP variants can be reported due to the settings used in each laboratory as number of samples per chip, use of Equalization kit, number of amplification cycles and the quality of samples, factors directly associated with the final quality of the NGS results. A detailed validation protocol, as our study, provides important information for other laboratories that are starting to use NGS sequencing decreasing technical errors, perform NGS validation in a shorter time interval with less cost and avoidance of FP variants as candidates to be inserted in the patient report. In our study, *BRCA2* variants ch13:32906547 (c.937_938insT) and chr13:32906565 (c.956_957insA) were confirmed as FP. Mehta *et al.* (2018) reported these variants as true germline mutations in Indian patients with breast cancer and the coverage data and Sanger validation were not reported (14).

Laboratories must perform the internal validation because FP variants can vary according to the run quality parameters and the presence of homopolymer regions. The selection criteria of variants must be chosen in accordance with the internal validation process of each laboratory (15). Lih et al. (2017) strongly recommend the reprocessing from template preparation of samples with low sequencing quality observed. A good quality of nucleic acid samples investigated by NGS is required to obtain true results and the DNA quantification is essential (9).

Ion Ampliseq DNA library preparation at amplify targets step (manufacturer's instructions) recommends an amplification cycle with 19 to high DNA quality and 22 cycles to low quality samples. Our study used high quality samples, although, we observed greater value in quality parameters when we used 23 cycles.

The manufacturer's instructions also suggest the use of Ion 314 chip with *BRCA1/BRCA2* panel with 16 samples, to obtain a coverage >95% of bases at 30x. We observed that the analysis of eight samples/ Ion 314 chip was not enough to perform a high coverage in some amplicons. Zanella *et al.* (2017) also observed the AMPL22550532 in exon 23 with an average coverage depth <20x (6).

Studies with NGS in germline variants generally use a 20x of minimum coverage/base, sequenced on both strands and accounting for at least ≥20% of total reads (15,16). To solve the low coverage of some amplicons (<16 reads), we analyzed 4 samples/chip and the mean read length turned around 151.8±1.3bp in protocols eight and nine. However, when eight samples/chip were analyzed, the mean read length was 110.1±11.2. Because of the low coverage in some

amplicons as reported in our study, all samples had exons 20 and 23 sequenced by Sanger in *BRCA2* gene.

All pathogenic variants must be confirmed by Sanger (17). We validated variants with low coverage and some variants with coverage greater than 60 reads to confirm variants. Some variants with coverage >60 reads were also validated, because an uncommon high frequency, mainly in protocols with low quality, was observed. After the protocol four, samples were reanalyzed and the FP InDel variants were not observed again.

Vendrell *et al.* (2018) reported that some FP variants, such as *BRCA2* c.2175dup, c.1689del and c.9739del and *BRCA1* c.5289del, had a variant allelic frequency of approximately 50% and this could be attributed to systematic artifacts (18). Some FP variants were also described in this study (table 3) and to distinguish them from the true variants, we increased the coverage cutoff. Each variant suspected of FP was checked by Sanger, in low quality and high quality runs and in more than one sample, to estimate the coverage cutoff mainly in homopolymer regions. In addition, the frequency of polymorphisms observed in the world population and the zygosity (example, as the world database Exome Aggregation Consortium - ExAC) should be considered to indicate a FP variant. If a clinical significance variant, suspected as FP, is observed in homozygous (by NGS) and, in the world databases, this same variant is in heterozygous, it is important to confirm with Sanger.

Jennings et al. (2017) recommend Sanger sequencing coupled with targeted mutation analysis when the allele burden is expected to be low (19).

In order to obtain greater statistical power, a larger number of samples per protocol should be analyzed. In some protocols only one sample was tested because NGS sequencing is still an expensive technique.

Beck et al.(2016) analyzed 5.660 variants, 19 were identified by NGS but not by Sanger sequencing, representing 13 unique single nucleotide variants, they observed a minimum of 99.96% accuracy ratio for NGS compared to Sanger sequencing (20).

Buzolin *et al.*(2017) observed quality parameters as number of mapped reads with 75.330, on target with 93.16%, average base depth coverage with 429.2 and 96.40% of uniformity (21). They used an input with 20 ng/ reaction and 4 samples per Ion 314 chip or 8 samples in Ion 316 chip. Our eighth protocol with 20 ng/ reaction input had 189.216 mapped reads, with 98.07%, average base depth coverage with 1.203.000 and 98.87% of uniformity. They identified 587 variants with 35 FP (5.9%) in 26 samples analyzed. Our study observed 1.4% (8/567) false positives variants considering all protocols performed, after applying the validation parameters developed.

Even with the optimization of the bioinformatics parameters used in our pipeline, which improved the quality of mapping and variant calling, we had false positive variants (1.4%). Allele Frequency, zygosity, number of studies that found out similar data and the clinical interpretation reported in databases must be also evaluated beyond the Ion Reporter.

NGS assay offers higher throughput and less cost when compared with Sanger (22). NGS shows high power estimation than Sanger (23) and is now being widely

adopted in clinical settings (15). NGS assay has a high complexity in data analysis and challenges surrounding the technical regarding the whole method (24,25). The parameters investigated do not include any novelty over NGS assay as seen in original articles, but as a methodological paper it can collaborate with researchers involved in this complex and manual work of sequencing in the laboratory bench. The analysis of the parameters carried out in this study is not found in scientific articles in a critical way and the technicians have only the manuals of the machines to follow as a guide for use. This study helps beginners in NGS sequencing to avoid wasting time and reducing costs with the errors generally made when using this technology for the first time.

This analytical validation study met the expected performance requirement for the intended used as recommended by manufacturer, but some important specifications would be done. The number of cycles in the amplification step, even using DNA with high quality, was better with 23. Ion AmpliSeq BRCA1/BRCA2 Community Panel had better performance with 4 samples per Ion-314 chip instead of 8. We observed better results when manual equalization process was done, without Ion Library Equalizer kit. These adjustments provided higher coverage of the variants and fewer artifacts (6.7-fold in the quality of results). Laboratories must perform the internal validation because FP InDel variants can vary according to the quality of results. Finally, NGS assay must be validated with Sanger in the first stage of validation of this technique in laboratories, and then, for new and low quality coverage variants observed in the laboratory routine (mainly InDels).

Acknowledgements

We are grateful to Gerson Moura Ferreira from National Cancer Institute-INCA, Fernanda de Paris and Gabriel de Souza Macedo from Clinical Hospital of Porto Alegre city for their help in our validation process sharing their NGS experiences.

Conflicts of interest

The authors declare that they have no conflict of interest.

Funding information

This work received funding from the National Program of Support to Cancer Oncology in Brazil (PRONON).

References

1. **Niell BL, Rennert G, Bonner JD, Almog R, Tomsho LP, Gruber SB.** BRCA1 and BRCA2 Founder Mutations and the Risk of Colorectal Cancer. J Natl Cancer Inst. 2004;96:15-21. https://doi.org/10.1093/jnci/djh008

2. **Garber JE, Syngal S.** One less thing to worry about: the shrinking spectrum of tumors in BRCA founder mutation carriers. J Natl Cancer Inst. 2004;96:2-310. https://doi.org/1093/jnci/djh021

3. **Wallace AJ.** New challenges for BRCA testing: a view from the diagnostic laboratory. Eur J Hum Genet. 2016;24 (Suppl 1):S10-18. https://doi.org/10.1038/ejhg.2016.94

4. **Loeffelholz M, Fofanov Y.** The main challenges that remain in applying high-throughput sequencing to clinical diagnostics. Expert Rev Mol Diagn. 2015;15:1405-8. https://doi.org/10.1586/14737159.2015.1088385

5. **Germani A, Libi F, Maggi S, Stanzani G, Lombardi A, Pellegrini P, et al.** Rapid detection of copy number variations and point mutations in BRCA1/2 genes

using a single workflow by ion semiconductor sequencing pipeline. Oncotarget. 2018;9:33648-55. https://doi.org/10.18632/oncotarget.26000

6. **Zanella I, Merola F, Biasiotto G, Archetti S, Spinelli E, Di Lorenzo D.** Evaluation of the Ion Torrent PGM sequencing workflow for the routine rapid detection of BRCA1 and BRCA2 germline mutations. Exp Mol Pathol. 2017;102:314-20. https://doi.org/10.1016/j.yexmp.2017.03.001

7. **Oosterwijk JC, de Vries J, Mourits MJ, de Bock GH.** Genetic testing and familial implications in breast-ovarian cancer families. Maturitas. 2014;78:252-7. https://doi.org/10.1016/j.maturitas.2014.05.002

8. **Claus EB, Schildkraut, Joellen M., Thompson Douglas, Risch, Neil J.** The genetic attributable risk of breast and ovarian cancer. Cancer. 1996;77:2318-24. https://doi.org/10.1002/(SICI)1097-0142(19960601)77:11<2318::AID-CNCR21>3.0.CO;2-Z

9. **Lih CJ, Harrington RD, Sims DJ, Harper KN, Bouk CH, Datta V, et al.** Analytical Validation of the Next-Generation Sequencing Assay for a Nationwide Signal-Finding Clinical Trial: Molecular Analysis for Therapy Choice Clinical Trial. J Mol Diagn. 2017;19:313-27. https://doi.org/10.1016/j.jmoldx.2016.10.007

10. **Maistro S, Teixeira N, Encinas G, Katayama ML, Niewiadonski VD, Cabral LG, et al.** Germline mutations in BRCA1 and BRCA2 in epithelial ovarian cancer patients in Brazil. BMC Cancer. 2016;16:934. https://doi.org/10.1186/s12885-016-2966-x

11. **Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW.** Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. PLoS Comput Biol. 2013;9:e1003031. https://doi.org/10.1371/journal.pcbi.1003031

12. **Kang HP, Maguire JR, Chu CS, Haque IS, Lai H, Mar-Heyming R, et al.** Design and validation of a next generation sequencing assay for hereditary BRCA1 and BRCA2 mutation testing. Peer J. 2016;4:e2162. https://doi.org/10.7717/peerj.2162

13. **Park J, Jang W, Chae H, Kim Y, Chi HY, Kim M.** Comparison of Targeted Next-Generation and Sanger Sequencing for the BRCA1 and BRCA2 Mutation Screening. Ann Lab Med. 2016;36:197-201.https://doi.org/201603197

14. **Mehta A, Vasudevan S, Sharma SK, Kumar D, Panigrahi M, Suryavanshi M, et al.** Germline BRCA1 and BRCA2 deleterious mutations and variants of unknown clinical significance associated with breast/ovarian cancer: a report from North India. Cancer Manag Res. 2018;10:6505-16. https://doi.org/10.2147/CMAR.S186563

15. **Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al.** ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013;15:733-747. https://doi.org/10.1038/gim.2013.92

16. **Trujillano D, Weiss ME, Schneider J, Koster J, Papachristos EB, Saviouk V, et al.** Next-generation sequencing of the BRCA1 and BRCA2 genes for the genetic diagnostics of hereditary breast and/or ovarian cancer. J Mol Diagn. 2015;17:162-70. https://doi.org/10.1016/j.jmoldx.2014.11.004

17. **Sanger F, Nicklen S, Coulson AR.** DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA. 1977;74:5463-7. https://doi.org/10.1073/pnas.74.12.5463

18. **Vendrell JA, 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;20:754-64. https://doi.org/S1525-1578(18)30072-2

19. **Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, et al.** Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. J Mol Diagn. 2017;19:341-65. https://doi.org/S1525-1578(17)30025-9

20. **Beck TF, Mullikin JC, Program NCS, Biesecker LG.** Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. Clin Chem. 2016;62:647-54. https://doi.org/10.1373/clinchem.2015.249623

21. **Buzolin AL, Moreira CM, Sacramento PR, Oku AY, Fornari A, Antonio DSM, et al.** Development and validation of a variant detection workflow for BRCA1 and BRCA2 genes and its clinical application based on the Ion Torrent technology. Hum Genomics. 2017;11:14. https://doi.org/10.1186/s40246-017-0110-x

22. **Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, et al.** Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol. 2009;10:R32. https://doi.org/10.1186/gb-2009-10-3-r32

23. **Strom CM, Rivera S, Elzinga C, Angeloni T, Rosenthal SH, Goos-Root D, et al.** Development and Validation of a Next-Generation Sequencing Assay for BRCA1 and BRCA2 Variants for the Clinical Laboratory. PLoS One. 2015;10:e0136419. https://doi.org/10.1371/journal.pone.0136419

24. **Bijwaard K, Dickey JS, Kelm K, Tezak Z.** The first FDA marketing authorizations of next-generation sequencing technology and tests: challenges,

solutions and impact for future assays. Expert Rev Mol Diagn. 2015;15:33-40. https://doi.org/10.1586/14737159.2015.979795

25. **Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al.** Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol. 2012;30:1033-6. https://doi.org/10.1038/nbt.2403

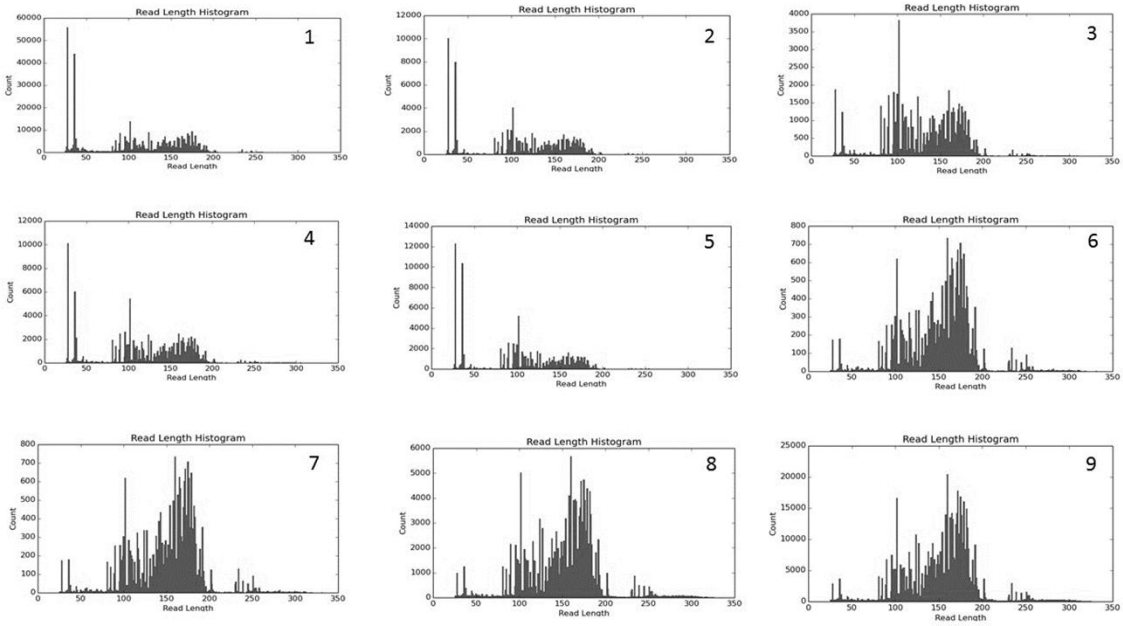

Figure 1: Amplicons obtained from one sample as representative of each protocol.

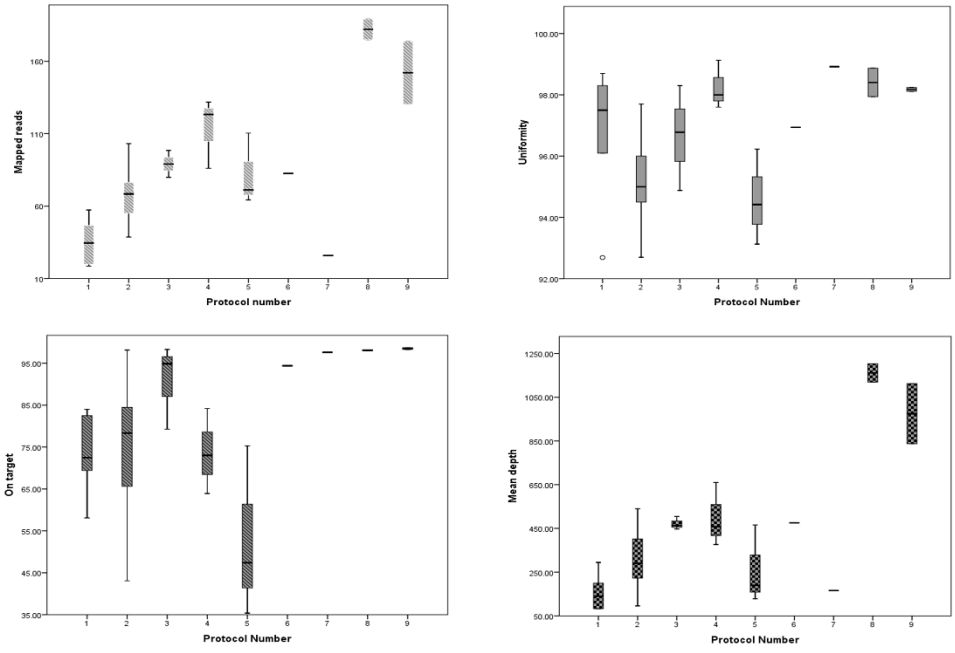

Figure 2 - Quality parameters of samples investigated in each protocol (median value).

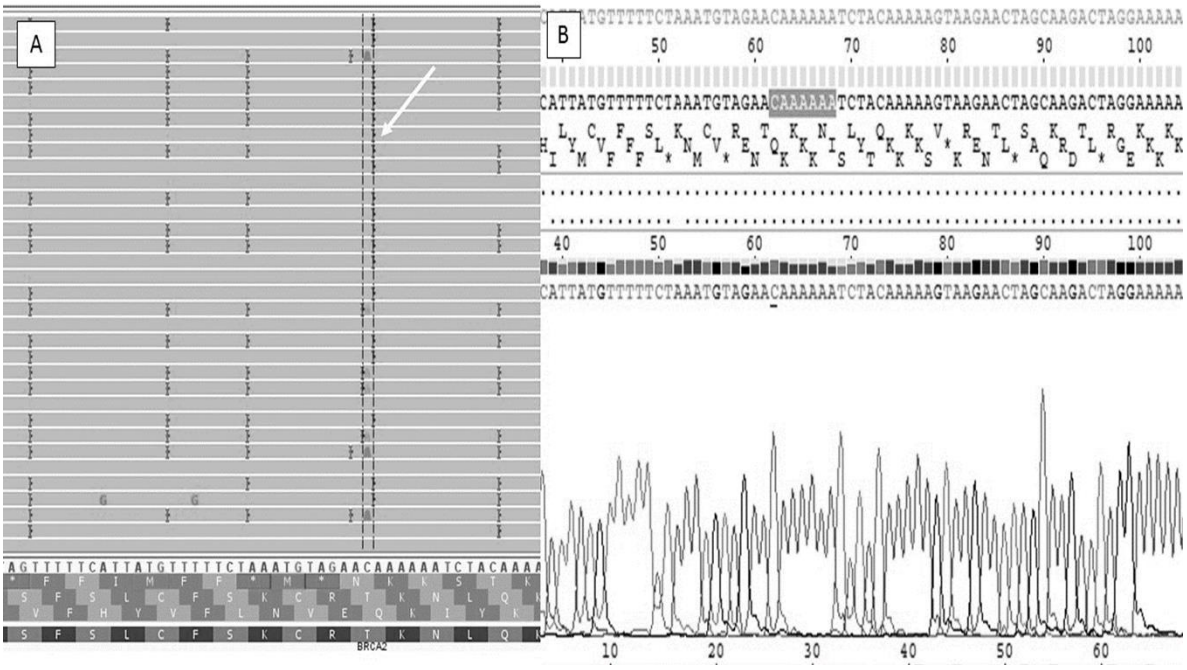

Figure 3 - Analysis by Sanger and by IGV (Integrative Genomics Viewer) of one FP variant ch13: 32906565 (GRCh37.p13). The insertion of A was observed by IGV (A) and no mutation was detected by Sanger (B).

Table 1: Protocols performed with varying factors and their quality of samples analyzed by spectrophotometry.

Protocol	260/230 ratio	260/280 ratio	Amplify targets (extension cycles)	Input (ng/reaction)	Library concentration (pM) Median (min-max)	Number of samples analyzed per chip
1	1.8 (1.8-1.8)	2 (1.4- 2.3)	19	20	52 (20-58)	8
2	1.8 (1.8-1.9)	2.1 (1.4- 2.3)	21	10	75 (36-111)	8
3	1.8 (1.8-1.8)	2.1 (1.9 - 2.2)	23	10	40 (35-44)	8
4[a]	1.8 (1.7-1.8)	2.1 (0.4 - 2.1)	23	10	25 (21-44)	4
5	1.8 (1.8-1.9)	2.2 (1.9 - 2.3)	21	10	69 (54-91)	4
6 [a b]	1.8	2.2	23	10	76.8	4
7 [a b]	1.8	2.1	23	10	73	4
8	1.8 (1.8- 1.9)	2 (1.9- 2.2)	23	20	24 (18-30)	4
9	1.8 (1.8- 1.9)	2 (1.9- 2.2)	23	30	28 (24-32)	4

[a] One sample tested
[b] Different manual equalization process considering the library enrichment step and the use of Ion Library Equalizer kit

Table 2 - Number and type of variants observed in each protocol (median value).

Number of protocol	N	SNP (absolute frequency)	InDel (absolute frequency)	InDel FP variants in homopolymer regions	Quality factor
1	7	11 (4-13)	2	2	12.5 (2.9-37.6)
2	29	11 (4-17)	5	5	13.5 (1.5-48)
3	3	14 (7-14)	1	1	35.5 (32.7-41.8)
4	3	8 (6-17)	0	0	35.2 (23.4-71.6)
5	3	16 (14-17)	0	0	6 (2.8-36.1)
6	1	13	0	0	35.9
7	1	12	0	0	4.1
8	2	15 (13-17)	0	0	204.4 (188.2-220.7)
9	2	15 (13-17)	0	0	146.1 (105.6-186.6)

Table 3: False Positive InDel variants included/excluded applying our validation parameters and confirmed by Sanger.

Quality of InDel FP variants included by the validation parameters and excluded by Sanger

Protocol	Variant	Samples	Coding (c.)	Protein (p.)	Mean Coverage	Phred Score	*p* value
1	32906547	4	c.937_938insT	p.Ser313fs	72	356.7	0.00001
	32906565	6	c.956_957insA	p.Asn319fs	99	718.5	0.00001
2	32906535	10 e 12	c.925_926insT	p.Ser309fs	54	406.8	0.00001
	32906547	10	c.937_938insT	p.Ser313fs	50	408.8	0.00001
	32906565	1, 8, 11, 13, 17, 23, 27, 28, 29 e 30	c.956_957insA	p.Asn319fs	64	429.5	0.00001
	32906576	10 e 28	c.966_967insA	p.Val323fs	83	583.2	0.00001
	32906602	10	c.994_995insA	p.Ile332fs	51	265.7	0.00001
3	32906565	9	c.956_957insA	p.Asn319fs	124	779	0.00001

Quality of InDel FP variants excluded by the validation parameters and excluded by Sanger

Protocol	Variant	Samples	Coding (c.)	Protein (p.)	Mean Coverage	Phred Score	*p* value
1	32893197	4	c.68-17AT>A	p.?	41	63.7	0.00001
	32907302	3, 4, 6 e 7	c.1689delG	p.Trp563fs	41	365.7	0.00001
2	32907302	1, 8, 10, 11, 12, 13, 17, 18, 19, 20, 21, 22, 23, 24, 26, 28, 29, 30 e 33	c.1689delG	p.Trp563fs	57	541.5	0.00001
	32907324	31	c.1711delT	p.Ser571fs	40	348.9	0.00001
3	32907302	9	c.1689delG	p.Trp563fs	113	1079.1	0.00001
4	32907302	sg49 e 9	c.1689delG	p.Trp563fs	102	925.3	0.00001
5	32907302	21	c.1689delG	p.Trp563fs	54	488.8	0.00001
6	32907302	25	c.1689delG	p.Trp563fs	78	721.1	0.00001
8	32907302	25	c.1689delG	p.Trp563fs	314	2996.1	0.00001