

# Identification of novel *BRCA1* large genomic rearrangements by a computational algorithm of amplicon-based Next-Generation Sequencing data

Arianna Nicolussi<sup>1</sup>, Francesca Belardinilli<sup>2</sup>, Valentina Silvestri<sup>2</sup>, Yasaman Mahdavian<sup>2</sup>, Virginia Valentini<sup>2</sup>, Sonia D'Inzeo<sup>3</sup>, Marialaura Petroni<sup>4</sup>, Massimo Zani<sup>2</sup>, Sergio Ferraro<sup>2</sup>, Stefano Di Giulio<sup>2</sup>, Francesca Fabretti<sup>2</sup>, Beatrice Fratini<sup>1</sup>, Angela Gradilone<sup>2</sup>, Laura Ottini<sup>2</sup>, Giuseppe Giannini<sup>Corresp., 2, 5</sup>, Anna Coppa<sup>Corresp., 1</sup>, Carlo Capalbo<sup>2</sup>

<sup>1</sup> Department of Experimental Medicine, University of Roma "La Sapienza", Roma, Italy

<sup>2</sup> Department of Molecular Medicine, University of Roma "La Sapienza", Roma, Italy

<sup>3</sup> U.O.C. Microbiology and Virology Laboratory, A.O. San Camillo Forlanini, Roma, Italy

<sup>4</sup> Istituto Italiano di Tecnologia, Center for Life Nano Science@Sapienza, Roma, Italy

<sup>5</sup> Istituto Pasteur-Fondazione Cenci Bolognetti, Roma, Italy

Corresponding Authors: Giuseppe Giannini, Anna Coppa

Email address: giuseppe.giannini@uniroma1.it, anna.coppa@uniroma1.it

**Background:** Genetic testing for *BRCA1/2* germline mutations in hereditary breast/ovarian cancer patients requires screening for single nucleotide variants, small insertions/deletions and large copy number variations (CNVs). These studies have long been run by Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA). The recent introduction of next-generation sequencing (NGS) platforms dramatically improved the speed and the efficiency of DNA testing for nucleotide variants, while the possibility to correctly detect CNVs by this mean is still debated. The purpose of this study was to establish whether and to which extent the development of an analytical algorithm could help us translating NGS sequencing via an Ion Torrent PGM platform into a tool suitable to identify LGRs in hereditary breast-ovarian cancer patients. **Methods:** We first used NGS data of a group of 3 patients (training set), previously screened in our laboratory by conventional methods, to develop an algorithm for the calculation of the dosage quotient (DQ) to be compared with the Ion Reporter (IR) analysis. Then, we tested the optimized pipeline with a consecutive cohort of 85 uncharacterized probands (validation set) also subjected to MLPA analysis. Characterization of the breakpoints of three novel *BRCA1* LGRs was obtained via long range PCR and direct sequencing of the DNA products. **Results:** In our cohort, the newly defined DQ-based algorithm detected 3/3 *BRCA1* LGRs, demonstrating 100% sensitivity and 100% negative predictive value (NPV) [95% CI:87.6-99.9]) compared to 2/3 cases detected by IR (66.7% sensitivity and 98.2% NPV [95% CI:85.6-99.9]). Interestingly, DQ and IR shared 12 positive results, but exons deletion calls matched only in 5 cases, two of which confirmed by MLPA. The breakpoints of the 3

novel *BRCA1* deletions, involving exons16-17, 21-22 and 20, have been characterized.

Conclusions: Our study defined a DQ-based algorithm to identify *BRCA1* LGRs using NGS data. Whether confirmed on larger data sets, this tool could guide the selection of samples to be subjected to MLPA analysis, leading to significant savings in time and money.

# Identification of novel *BRCAl* large genomic rearrangements by a computational algorithm of amplicon-based Next-Generation Sequencing data

Arianna Nicolussi<sup>1</sup>; Francesca Belardinilli<sup>2</sup>; Valentina Silvestri<sup>2</sup>; Yasaman Mahdavian<sup>2</sup>; Virginia Valentini<sup>2</sup>; Sonia D’Inzeo<sup>3</sup>; Marialaura Petroni<sup>4</sup>; Massimo Zani<sup>2</sup>; Sergio Ferraro<sup>2</sup>; Stefano Di Giulio<sup>2</sup>; Francesca Fabretti<sup>2</sup>; Beatrice Fratini<sup>1</sup>; Angela Gradilone<sup>2</sup>; Laura Ottini<sup>2</sup>; Giuseppe Giannini<sup>2,5</sup>; Anna Coppa<sup>1\*</sup> and Carlo Capalbo<sup>2\*</sup>

<sup>1</sup>Department of Experimental Medicine, University of Roma “La Sapienza”, Italy

<sup>2</sup>Department of Molecular Medicine, University of Roma “La Sapienza”, Italy

<sup>3</sup>A.O. San Camillo Forlanini, U.O.C. Microbiology and Virology Laboratory, Roma, Italy

<sup>4</sup>Center for Life Nano Science@Sapienza, Istituto Italiano di Tecnologia, Roma, Italy

<sup>5</sup>Istituto Pasteur-Fondazione Cenci Bolognetti, Roma, Italy

\* These authors equally contributed to this work

Corresponding authors:

Dr. Anna Coppa, Department of Experimental Medicine, University of Roma “La Sapienza”, Italy

Prof. Giuseppe Giannini, Department of Molecular Medicine, University of Roma “La Sapienza”, Italy

Email address: [anna.coppa@uniroma1.it](mailto:anna.coppa@uniroma1.it); [giuseppe.giannini@uniroma1.it](mailto:giuseppe.giannini@uniroma1.it)

## 22 Abstract

23 **Background:** Genetic testing for *BRCA1/2* germline mutations in hereditary breast/ovarian cancer  
 24 patients requires screening for single nucleotide variants, small insertions/deletions and large copy  
 25 number variations (CNVs). These studies have long been run by Sanger sequencing and multiplex  
 26 ligation-dependent probe amplification (MLPA). The recent introduction of next-generation  
 27 sequencing (NGS) platforms dramatically improved the speed and the efficiency of DNA testing  
 28 for nucleotide variants, while the possibility to correctly detect CNVs by this mean is still debated.  
 29 The purpose of this study was to establish whether and to which extent the development of an  
 30 analytical algorithm could help us translating NGS sequencing via an Ion Torrent PGM platform  
 31 into a tool suitable to identify LGRs in hereditary breast-ovarian cancer patients.

32 **Methods:** We first used NGS data of a group of 3 patients (training set), previously screened in  
 33 our laboratory by conventional methods, to develop an algorithm for the calculation of the dosage  
 34 quotient (DQ) to be compared with the Ion Reporter (IR) analysis. Then, we tested the optimized  
 35 pipeline with a consecutive cohort of 85 uncharacterized probands (validation set) also subjected  
 36 to MLPA analysis. Characterization of the breakpoints of three novel *BRCA1* LGRs was obtained  
 37 via long range PCR and direct sequencing of the DNA products.

38 **Results:** In our cohort, the newly defined DQ-based algorithm detected 3/3 *BRCA1* LGRs,  
 39 demonstrating 100% sensitivity and 100% negative predictive value (NPV) [95% CI:87.6-99.9])  
 40 compared to 2/3 cases detected by IR (66.7% sensitivity and 98.2% NPV [95% CI:85.6-99.9]).  
 41 Interestingly, DQ and IR shared 12 positive results, but exons deletion calls matched only in 5  
 42 cases, two of which confirmed by MLPA. The breakpoints of the 3 novel *BRCA1* deletions,  
 43 involving exons16-17, 21-22 and 20, have been characterized.

**Conclusions:** Our study defined a DQ-based algorithm to identify *BRCA1* LGRs using NGS data. Whether confirmed on larger data sets, this tool could guide the selection of samples to be subjected to MLPA analysis, leading to significant savings in time and money.

# Keywords

*BRCA1* LGRs; NGS; Deep coverage; Analytical validation; DQ analysis; MLPA

# Background

Hereditary breast and ovarian cancer syndrome, caused by germline pathogenic mutations in the *BRCA1* (MIM#113705) or *BRCA2* (MIM#600185) genes, is characterized by an increased risk for breast, ovarian, pancreatic and other cancers (Palma et al., 2006). It has been recently estimated that the cumulative risks of breast cancer to age 80 years was 72% for *BRCA1* and 69% for *BRCA2* carriers (Kuchenbaecker et al., 2017). Differences in mutation type and site may at least partially impact on cancer risk definition (Rebbeck et al., 2015; Coppa et al., 2018; Rebbeck et al., 2018). *BRCA1* and *BRCA2* gene mutations are typically found in 25–30% of the breast cancer families subjected to genetic testing (Giannini et al., 2006; Economopoulou, Dimitriadis & Psyrri, 2015). The relatively low rate of success in finding relevant pathogenic mutations in this settings is likely due to the contribution of other moderate-to-high penetrance breast cancer susceptibility genes (i.e., *PALB2*, *ATM*, *CHK2*) (Economopoulou, Dimitriadis & Psyrri, 2015; Coppa et al., 2018), or to the influence of low penetrance and risk-modifying alleles (Couch et al., 2012; Ottini et al., 2013; Kuchenbaecker et al., 2014; Peterlongo et al., 2015), all of which needs to be taken into account for a more appropriate assessment of individual cancer risk. For quite some time, the use of classical qualitative PCR-based techniques incapable of detecting large genomic

rearrangements (LGRs) also contributed to failures in the identification of *BRCA* mutation carriers. Interestingly, the prevalence of *BRCA1/BRCA2* LGRs varies greatly among different populations ranging from 0 to 27% of mutation positive families in Iranian/French, Canadian, Dutch, Spanish, German, French and South Africa populations (Gad et al., 2002; Hogervorst et al., 2003; Hartmann et al., 2004; Pietschmann et al., 2005; Moisan et al., 2006; la Hoya et al., 2006; Sluiter & van Rensburg, 2011). Relevant differences in the frequency of *BRCA1* LGRs have also been reported within the Italian population (Montagna et al., 2003; Buffone et al., 2007). In general, *BRCA2* LGRs are less frequent (Woodward et al., 2005; Agata et al., 2005; Buffone et al., 2007), probably due to the lower density of Alu sequences compared to *BRCA1*, which are involved in the genesis of LGRs (Smith et al., 1996).

Multiplex ligation-dependent probe amplification (MLPA) is the most commonly used technique for the detection of large deletions/duplications in *BRCA1/2* genes.

The recent advances in sequencing technologies have increased the speed and efficiency of DNA testing and the emergence of benchtop next-generation sequencing (NGS) instruments are becoming the standard in molecular genetic diagnosis (Feliubadalo et al., 2013; Trujillano et al., 2015). NGS is capable of sensitive detection of sequence variants, but may also be used for detection of copy number variants (CNVs) (Tarabeux et al., 2014; Enyedi et al., 2016; Schenkel et al., 2016; Schmidt et al., 2017). The CNVs assessment is mainly performed using the sequencing read depth (RD) assessment approach, whose assumption is that the RD signal is proportional to the number of copies of chromosomal segments present in that specimen (Tan et al., 2014). The ability to detect LGRs from NGS multigene panel largely, but not uniquely, depends on the library preparation, and target enrichment approaches based on hybridization and capture seem to have better performances compared to amplicon-based methods. In general, NGS data are not routinely

used for CNVs detection in clinical settings for *BRCA* mutation screenings, due to concerns related to library preparation protocols, normalization procedures and employed software (Feliubadalo et al., 2013; Wallace, 2016). Recently, we adopted the NGS Ion AmpliSeq™ *BRCA1* and *BRCA2* Panel to perform routine *BRCA1/2* mutation screening on the Ion PGM platform (Nicolussi et al., 2019). Here, we aimed at establishing whether sequencing data generated by this approach could be processed by a computational algorithm to efficiently predict the presence of LGRs, based on the dosage quotient (DQ) calculation and the Ion Reporter (IR) analysis.

## Methods

### Patients and DNA

Families putatively affected by hereditary breast/ovarian cancer syndrome were recruited at the Hereditary Tumors section of Policlinico Umberto I, University La Sapienza, between July 2015 and September 2017 and selected as previously described (Capalbo et al., 2006b,a; Coppa et al., 2014). Comprehensive pre-test counseling was offered to all probands and their family members and informed consent was obtained. For each study participant, samples of blood or DNA from peripheral blood leukocytes were collected. DNA from blood samples was extracted and quantified as described by Nicolussi et al. (Nicolussi et al., 2019). All investigations were approved by Ethics Committee of the University of Roma “La Sapienza” (Prot.: 88/18; RIF.CE:4903, 31-01-2018) and conducted according to the principles outlined in the declaration of Helsinki.

A retrospective group of 3 DNA samples, previously found positive for *BRCA1* large genomic rearrangements (LGRs) by MLPA was used as a training set (TS). LGRs in the TS were as follows: sample BR59, *BRCA1* exon 23-24 deletion (c.5407-?(1\_?)del); sample BR328, *BRCA1* exon 18-19 deletion (c.5075-?\_5193+?del)(Buffone et al., 2007) and sample BR409, *NBR2* exon1 and

*BRCA1* exon 1-2 deletion (NBR2del EX1\_BRCA1 delEX1-2) (Coppa et al., 2018) (Table 1).

For NGS-based LGR analysis, a consecutive group of 127 NGS/MPLA negative samples have been used to create a baseline and a prospective consecutive cohort of 85 uncharacterized probands, validation set (VS), was studied.

### **Ion Torrent PGM sequencing**

The target regions in the *BRCA1* and *BRCA2* genes were amplified using the Ion AmpliSeq™ *BRCA1* and *BRCA2* Panel (Life Technologies) according to the manufacturers' procedures and processed as previously described (Belardinilli et al., 2015; Nicolussi et al., 2019).

### **Sanger sequencing**

All clinical samples were sequenced for the entire coding regions by Sanger sequencing, using an ABI PRISM DyeDeoxy Terminator Cycle Sequencing Kit and an ABI 3100 Genetic Analyzer (Applied Biosystems, Warrington, UK). Reference sequence for *BRCA1* was Genbank, NM\_007294.3, and reference sequence for *BRCA2* was Genebank, NM\_000059.3.

### **MLPA analyses**

Multiple-Ligation-dependent-Probe-Amplification (MLPA) methodology (Schouten et al., 2002) was performed, according to the manufacturer's instructions (MRC–Holland, Amsterdam, the Netherlands), to identify *BRCA1/2* genomic rearrangements. For the statistical analysis we transferred the size and the peak areas of each sample to an Excel file. The peak areas of the expected MLPA products were evaluated by comparison with a normal control and by cumulative comparison of all samples within the same experiment (Buffone et al., 2007; Coppa et al., 2018).



# NGS-based Large Genomic Rearrangement (LGRs) Analysis

Large Genomic rearrangements in *BRCA1* gene were studied by two distinct approaches: the manual calculation of the dosage quotient (DQ) and the Ion Reporter (IR) platform. In the manual approach, DQ for each sample was calculated as follows: amplicon read count normalized on the *BRCA1* and *BRCA2* total reads/average of normalized amplicon read counts obtained from all samples. Specifically, we referred to DQA when amplicon counts were normalized vs the coverage data of all samples run on the same single chip, and to DQB when amplicon counts were normalized vs coverage data obtained from a baseline built from 127 LGRs negative samples. In addition, DQB has been alternatively obtained either considering together all amplicons of the Ion AmpliSeq™ BRCA1 and BRCA2 Panel (DQB1) or by separately considering the three different pools of amplicons (DQB2). DQ value higher than mean plus 2 standard deviations (SD) was considered indicative of a duplication; DQ value lower than mean minus 2 SD was considered indicative of a deletion. Particular attention has been also paid to reduction of multiple consecutive amplicons, even when they failed to trespass the above defined thresholds.

In the IR approach, we create a user-defined CNV detection workflow by a tunable Ion Reporter™ Software algorithm based on Hidden Markov Model (HMM), that utilize normalized read coverage across amplicons to predict the copy number or ploidy (<https://assets.thermofisher.com/TFS-Assets/LSG/brochures/CNV-Detection-by-Ion.pdf>). The data coverage of 20 mutation-negative patients has been used as CNV baseline to analyze the samples of both TS and VS. We detected no *BRCA2* LGR in both the TS and VS. Thus, our analysis is necessarily limited to *BRCA1* LGRs.

## DNA breakpoint analysis

Newly discovered *BRCA1* large deletions were validated by characterization of the genomic breakpoints. Long-range PCR was performed according to the manufacturer's instructions using primers sitting on closer undeleted exons, whose sequences are available upon request. PCR products were purified with ExoSAP-IT (USB Corp., Cleveland, USA) according to the manufacturer's instructions and sequenced using the ABI PRISM DyeDeoxy Terminator Cycle Sequencing Kit and an ABI 3100 Genetic Analyzer (Applied Biosystems, Warrington, UK). Reference sequences for *BRCA1* and *BRCA2* were Genebank NM\_007294.3 and NM\_000059.3, respectively.

## Statistical analysis

Validation metrics were defined as: Accuracy =  $(TP + TN)/(TP + FP + TN + FN)$ ; Sensitivity =  $TP/(TP + FN)$ ; Specificity =  $TN/(TN + FP)$ ; FDR =  $FP/(TP + FP)$ ; Negative Predictive Value =  $TN/(TN + FN)$ , where TP = true positives, TN = true negatives, FP = false positives, FN = false negatives. The confidence intervals (CIs) were calculated by the method of Wilson (1927)(EB, 2019).

## Results

### NGS-dependent LGR analyses

To establish whether the data obtained by NGS via Ion AmpliSeq™ BRCA1 and BRCA2 Panel were suitable to identify copy number alterations in *BRCA1*, we used data from three samples (TS), already characterized in our laboratory for the presence of *BRCA1* LGRs by MLPA (Table 1). The sequencing data of the TS were analyzed by a locally devised algorithm for the calculation of the

dosage quotient (DQ) and by our custom modified Ion Reporter analysis (IR), as described in materials and methods. The intrarun DQ calculation (DQA), which includes normalization based on the coverage data of the samples sequenced in the same chip, was always included to monitor the variability eventually due to different batches of reagents or to time-related variables. In general, however, we thought we could get improved resolution and reduced numbers of CNV false calls by normalizing the coverage data of all amplicons of each sample vs those obtained from a reference set of 127 MLPA negative samples selected on the basis of their quality and uniformity of the coverage (DQB analysis). This baseline has been used to perform two DQB calculations, considering either all amplicons contained in the Ion AmpliSeq™ BRCA1 and BRCA2 Panel (DQB1) or dividing them into the three subsets identified by the amplification primer pools (DQB2).

As shown in Fig 1a, the DQA plot of the TS samples revealed the presence of peaks below the thresholds, in samples BR328 and BR409 (corresponding to deletions of *BRCA1* exons 18-19 and 1-2, respectively, in agreement with MLPA results). The DNA quality of BR59 sample was rather low, as evidenced by the many peaks out of the threshold. Nevertheless, the DQB1 analysis evidenced values below the threshold for 3 consecutive amplicons (AMPL223551867, 223530147 and 223954665), identifying *BRCA1* exon 23-24 deletion (Fig.1b), already discovered by MLPA analysis. Although they fail to trespass the threshold, the same consecutive amplicons showed strongly reduced values also at DQA evaluation (Fig. 1A). Hence, the careful examination of the two DQ calculations allowed us to identify all three *BRCA1* LGRs in the training set. Also, the analysis performed by IR software detected the presence of CNV (CNV=1) in the proper regions in all three TS samples (Table 2). On this basis, we extended DQA, DQB and IR analysis to a group of 85 consecutive samples (VS) negative for *BRCA1/2* pathogenic variants at NGS analysis

and compared it with MLPA results. Overall, DQA and DQB analysis resulted in detection of positive calls in 33/85 (39%) samples, while IR analysis detected CNVs in 29/85 (34%) (Table 3). Interestingly, DQ and IR evaluation only shared 12 positive results, with exon calls being not coincident in 7 of them and with a rather precise, although imperfect, indication of the exons involved in the remaining 5 (Table S1). MLPA confirmed *BRCAl* CNVs in 3/85 samples (Fig. 2): BR963 and BR1379, belonging to the small group of 5 DQ/IR double positive samples, and BR1154 resulted DQ positive-IR negative. Therefore, DQ calculation resulted 100% sensitive and displayed a 100% NPV (95% CI:87.6-99.9) (Table 3) in our VS, values not reached by IR analysis, which failed in the identification of BR1154 (Table 2). Within DQ analysis, the correct calls were more clearly defined by the DQB2 calculation (Fig. 3a, b, c). The appropriateness of the deletions calls of DQ, IR and MLPA evaluations were confirmed by the molecular characterization of the breakpoints, as described below.

Finally, we propose an operative algorithm which uses DQ calculation and IR analysis to select samples to be subjected to MLPA analysis, as indicated in Fig. 4. Indeed, all DQ positive samples should be subjected to MLPA, while DQ and IR double positive samples, sharing calls in the same regions, could be directly subjected to second level confirmation assay or directly to breakpoint characterization. In principle, all DQ negative samples (52 sample out of 85 in our VS) could be considered negative for LGRs, thus completing the analysis at this step.

## Characterization of LGRs

Identification of the breakpoints characterizing the LGRs is important for several reasons, including the possibility to develop diagnostic assays for segregation analyses in relatives. For different reasons DQ, IR and MLPA analyses are not able to provide such detailed molecular

characterization of LGR. To define the breakpoints of the newly identified *BRCA1* LGRs, PCR amplification of genomic DNA from the three samples and direct sequencing were performed. As shown in Fig. 5, PCR amplification of genomic DNA from the BR963 patient resulted in an aberrant fragment of approximately 1353 bp, whose direct sequencing confirmed loss of *BRCA1* exons 21 and 22, possibly originating from an erroneous homologous recombination process between an AluSq2 (Alu family, SINE class; chr17:41206762-41207066) and an AluSz (Alu family, SINE class; chr17:41200521-41200834) motifs. The rearrangement involved a perfectly repeated stretch of 24 bases and resulted in the deletion of 6228 nucleotides encompassing part of IVS20, exons 21-22 and IVS22 (Fig. 5a). The BR963 proband was affected with breast cancer at age 40 and belonged to HBC family. Segregation analysis demonstrated that the mutation came from the maternal lineage (Fig. 6). PCR amplification of genomic DNA from BR1154 patient resulted in an aberrant fragment of approximately 872 bp (also present in her mother, sample BR1148), whose direct sequencing confirmed loss of *BRCA1* exons 20, possibly originating from an erroneous homologous recombination process between an AluY (Alu family, SINE class; chr17:41205398-41205698) and an AluY (Alu family, SINE class; chr17:41205398-41205698) motifs (Fig. 5b). The rearrangement involved a perfectly repeated stretch of 11 bases and resulted in the deletion of 4173 nucleotides encompassing part of IVS19, exon 20 and IVS20. The BR1154 proband was affected with ovarian cancer at age 52 and belonged to a HBOC family (Fig. 6). The segregation analysis demonstrated that the mutation originating from the maternal lineage segregated in three individuals (Fig. 6). Finally, PCR amplification of genomic DNA from BR1379 patient, resulted in an aberrant fragment of approximately 2027 bp, whose direct sequencing confirmed loss of *BRCA1* exons 16 and 17, possibly originated from an erroneous homologous recombination process between an AluSp (Alu family, SINE class; chr17:41224585-41224884)

and an AluSg (Alu family, SINE class; chr17:41218424-41218724) motif (Fig. 5c). The rearrangement involved a perfectly repeated stretch of 16 bases and resulted in the deletion of 6155 nucleotides encompassing part of IVS15, exons 16-17 and IVS17. The BR1379 proband was affected with bilateral breast cancer at age 42 and 58 and belonged to a family with colon cancer and hepatomas cases (Fig. 6).

## Discussion

The simultaneous detection of mutations and copy number alterations is an attractive and useful prospect for clinical settings. In the last years the NGS-based approaches for genetic testing offered a powerful alternative for *BRCA1/2* mutation detection. However, the specificity of this approach is still considered not completely satisfactory for a correct LGRs detection. One of the most relevant aspects concerns the library preparation method, with the amplicon-based approach having a lower specificity compared to target enrichment approaches (Apessos et al., 2018). Here we reported the definition of an operative algorithm to use amplicon-based Ion-PGM/Ampliseq *BRCA1/BRCA2* sequencing data to efficiently predict the occurrence of *BRCA1* LGRs. By comparison of the results obtained with DQ and IR analyses, we demonstrate that DQ had 100% sensitivity and 100% NPV, at variance with IR analysis, which failed in the identification of a *BRCA1* exon 20 deletion. This result is consistent with one known limitation of the IR software, able to detect CNVs only if the region of interest is covered by more than one amplicon (<https://assets.thermofisher.com/TFS-Assets/LSG/brochures/CNV-Detection-by-Ion.pdf>).

Indeed, *BRCA1* exon 20, deleted in BR1154 sample, is covered by only one amplicon in the Ion AmpliSeq™ *BRCA1* and *BRCA2* Panel, making IR incapable of calling this CNV.

Of course, a major caveat deals with the limited specificity and accuracy of our approach, which

could not overcome the limitations also reported by other groups (Feliubadalo et al., 2013; Pilato et al., 2016). Thus, although our operative algorithm cannot fully substitute for MLPA analysis, and if our data will be confirmed in larger data sets, we suggest that combined DQ and IR analyses could be used for selecting samples to be subjected to MLPA analysis following the flow chart depicted in Fig. 4, with significant savings in time and money.

Another important contribution of this paper is the molecular characterization of the three novel *BRCA1* rearrangements up to providing their unique breakpoint coordinates. Deletion of exons 21 and 22 causing damage to the C-terminal BRCT domain of the BRCA1 protein has been reported and characterized in Czech (Vasickova et al., 2007; Ticha et al., 2010) and Malay population (Hasmad et al., 2015), but with different breakpoints. *BRCA1* exon 20 deletion has been described in Italian and Greek population (Montagna et al., 2003; Belogianni et al., 2004; Armaou et al., 2007) but all different from each other and from our own, with respect to their breakpoints. The *BRCA1* exons 16-17 deletion, responsible of BRCA1 loss of function (Carvalho et al., 2009), has been reported in Latin America/Caribbean population, but the breakpoints were not provided by the authors (Judkins et al., 2012). Similar to many other cases (Mazoyer, 2005; Buffone et al., 2007; Ewald et al., 2009), all three novel rearrangements described here, are likely to be due to an erroneous homologous recombination event between perfectly matching Alu repeats.

## Conclusion

In conclusion, here we described a simple approach that require the use of a basic statistical package such as Microsoft Excel, to predict the occurrence of LGRs by the analysis of NGS data designed for Ion AmpliSeq™ BRCA1 and BRCA2 Panel/IT-PGM platform, applicable to all NGS

platforms in use to reduce the number of samples to be subjected to MLPA analysis. We also characterized for the first time the breakpoints of three novel *BRCA1* LGRs.

## Acknowledgements

The authors would like to thank Dr. Alessandro Albiero for advice on data analysis.

## References

1. Palma M, Ristori E, Ricevuto E, Giannini G, Gulino A. BRCA1 and BRCA2: the genetic testing and the current management options for mutation carriers. *Crit Rev Oncol Hematol*. 2006;57(1):1-23. doi:10.1016/j.critrevonc.2005.05.003.
2. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA*. 2017;317(23):2402-2416. doi:10.1001/jama.2017.7112.
3. Coppa A, Nicolussi A, D'Inzeo S, et al. Optimizing the identification of risk-relevant mutations by multigene panel testing in selected hereditary breast/ovarian cancer families. *Cancer Med*. 2018;7(1):46-55. doi:10.1002/cam4.1251.
4. Rebbeck TR, Mitra N, Wan F, et al. Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. *JAMA*. 2015;313(13):1347-1361. doi:10.1001/jama.2014.5985.



- 315 5. Rebbeck TR, Friebe TM, Friedman E, et al. Mutational spectrum in a worldwide study of  
316 29,700 families with BRCA1 or BRCA2 mutations. *Hum Mutat.* 2018;39(5):593-620.  
317 doi:10.1002/humu.23406.
- 318 6. Giannini G, Capalbo C, Ristori E, et al. Novel BRCA1 and BRCA2 germline mutations and  
319 assessment of mutation spectrum and prevalence in Italian breast and/or ovarian cancer  
320 families. *Breast Cancer Res Treat.* 2006;100(1):83-91. doi:10.1007/s10549-006-9225-9.
- 321 7. Economopoulou P, Dimitriadis G, Psyrri A. Beyond BRCA: new hereditary breast cancer  
322 susceptibility genes. *Cancer Treat Rev.* 2015;41(1):1-8. doi:10.1016/j.ctrv.2014.10.008.
- 323 8. Couch FJ, Antoniou AC, Ramus SJ, et al. Common Variants at the 19p13.1 and ZNF365  
324 Loci Are Associated with ER Subtypes of Breast Cancer and Ovarian Cancer Risk in  
325 BRCA1 and BRCA2 Mutation Carriers. *Cancer Epidemiol Biomarkers Prev.* February  
326 2012. doi:10.1158/1055-9965.EPI-11-0888.
- 327 9. Ottini L, Silvestri V, Saieva C, et al. Association of low-penetrance alleles with male breast  
328 cancer risk and clinicopathological characteristics: results from a multicenter study in Italy.  
329 *Breast Cancer Res Treat.* 2013;138(3):861-868. doi:10.1007/s10549-013-2459-4.
- 330 10. Kuchenbaecker KB, Neuhausen SL, Robson M, et al. Associations of common breast cancer  
331 susceptibility alleles with risk of breast cancer subtypes in BRCA1 and BRCA2 mutation  
332 carriers. *Breast cancer research : BCR.* 2014;16(6):3416. doi:10.1186/s13058-014-0492-9.
- 333 11. Peterlongo P, Chang-Claude J, Moysich KB, et al. Candidate genetic modifiers for breast  
334 and ovarian cancer risk in BRCA1 and BRCA2 mutation carriers. *Cancer Epidemiol*  
335 *Biomarkers Prev.* 2015;24(1):308-316. doi:10.1158/1055-9965.EPI-14-0532.

- 336 12. Moisan A-M, Fortin J, Dumont M, et al. No Evidence of BRCA1/2 genomic rearrangements  
337 in high-risk French-Canadian breast/ovarian cancer families. *Genet Test*. 2006;10(2):104-  
338 115. doi:10.1089/gte.2006.10.104.
- 339 13. Hogervorst FBL, Nederlof PM, Gille JJP, et al. Large genomic deletions and duplications  
340 in the BRCA1 gene identified by a novel quantitative method. *Cancer Res*.  
341 2003;63(7):1449-1453.
- 342 14. Pietschmann A, Mehdipour P, Atri M, et al. Mutation analysis of BRCA1 and BRCA2 genes  
343 in Iranian high risk breast cancer families. *J Cancer Res Clin Oncol*. 2005;131(8):552-558.  
344 doi:10.1007/s00432-005-0678-8.
- 345 15. la Hoya de M, Gutiérrez-Enríquez S, Velasco E, et al. Genomic rearrangements at the  
346 BRCA1 locus in Spanish families with breast/ovarian cancer. *Clin Chem*. 2006;52(8):1480-  
347 1485. doi:10.1373/clinchem.2006.070110.
- 348 16. Hartmann C, John AL, Klaes R, et al. Large BRCA1 gene deletions are found in 3% of  
349 German high-risk breast cancer families. *Hum Mutat*. 2004;24(6):534.  
350 doi:10.1002/humu.9291.
- 351 17. Gad S, Caux-Moncoutier V, Pages-Berhouet S, et al. Significant contribution of large  
352 BRCA1 gene rearrangements in 120 French breast and ovarian cancer families. *Oncogene*.  
353 2002;21(44):6841-6847. doi:10.1038/sj.onc.1205685.
- 354 18. Sluiter MD, van Rensburg EJ. Large genomic rearrangements of the BRCA1 and BRCA2  
355 genes: review of the literature and report of a novel BRCA1 mutation. *Breast Cancer Res*  
356 *Treat*. 2011;125(2):325-349. doi:10.1007/s10549-010-0817-z.

19. Montagna M, Dalla Palma M, Menin C, et al. Genomic rearrangements account for more than one-third of the BRCA1 mutations in northern Italian breast/ovarian cancer families. *Hum Mol Genet.* 2003;12(9):1055-1061.
20. Buffone A, Capalbo C, Ricevuto E, et al. Prevalence of BRCA1 and BRCA2 genomic rearrangements in a cohort of consecutive Italian breast and/or ovarian cancer families. *Breast Cancer Res Treat.* 2007;106(2):289-296. doi:10.1007/s10549-007-9499-6.
21. Agata S, Dalla Palma M, Callegaro M, et al. Large genomic deletions inactivate the BRCA2 gene in breast cancer families. *J Med Genet.* 2005;42(10):e64. doi:10.1136/jmg.2005.032789.
22. Woodward AM, Davis TA, Silva AGS, Kirk JA, Leary JA, kConFab Investigators. Large genomic rearrangements of both BRCA2 and BRCA1 are a feature of the inherited breast/ovarian cancer phenotype in selected families. *J Med Genet.* 2005;42(5):e31. doi:10.1136/jmg.2004.027961.
23. Smith TM, Lee MK, Szabo CI, et al. Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. *Genome Res.* 1996;6(11):1029-1049.
24. Feliubadalo L, Lopez-Doriga A, Castellsagué E, et al. Next-generation sequencing meets genetic diagnostics: development of a comprehensive workflow for the analysis of BRCA1 and BRCA2 genes. *Eur J Hum Genet.* 2013;21(8):864-870. doi:10.1038/ejhg.2012.270.
25. Trujillano D, Weiss MER, Schneider J, 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(2):162-170. doi:10.1016/j.jmoldx.2014.11.004.

- 378 26. Tarabeux J, Zeitouni B, Moncoutier V, et al. Streamlined ion torrent PGM-based  
379 diagnostics: BRCA1 and BRCA2 genes as a model. *Eur J Hum Genet.* 2014;22(4):535-541.  
380 doi:10.1038/ejhg.2013.181.
- 381 27. Schenkel LC, Kerkhof J, Stuart A, et al. Clinical Next-Generation Sequencing Pipeline  
382 Outperforms a Combined Approach Using Sanger Sequencing and Multiplex Ligation-  
383 Dependent Probe Amplification in Targeted Gene Panel Analysis. *J Mol Diagn.*  
384 2016;18(5):657-667. doi:10.1016/j.jmoldx.2016.04.002.
- 385 28. Enyedi MZ, Jaksa G, Pintér L, et al. Simultaneous detection of BRCA mutations and large  
386 genomic rearrangements in germline DNA and FFPE tumor samples. *Oncotarget.* August  
387 2016. doi:10.18632/oncotarget.11259.
- 388 29. Schmidt AY, Hansen TVO, Ahlborn LB, Jønson L, Yde CW, Nielsen FC. Next-Generation  
389 Sequencing-Based Detection of Germline Copy Number Variations in BRCA1/BRCA2:  
390 Validation of a One-Step Diagnostic Workflow. *J Mol Diagn.* 2017;19(6):809-816.  
391 doi:10.1016/j.jmoldx.2017.07.003.
- 392 30. Tan R, Wang Y, Kleinstein SE, et al. An evaluation of copy number variation detection  
393 tools from whole-exome sequencing data. *Hum Mutat.* 2014;35(7):899-907.  
394 doi:10.1002/humu.22537.
- 395 31. Wallace AJ. New challenges for BRCA testing: a view from the diagnostic laboratory. *Eur*  
396 *J Hum Genet.* 2016;24 Suppl 1(S1):S10-S18. doi:10.1038/ejhg.2016.94.

32. Nicolussi A, Belardinilli F, Mahdavian Y, et al. Next-generation sequencing of BRCA1 and BRCA2 genes for rapid detection of germline mutations in hereditary breast/ovarian cancer. *PeerJ*. 2019;7:e6661. doi:10.7717/peerj.6661.
33. Capalbo C, Ricevuto E, Vestri A, et al. BRCA1 and BRCA2 genetic testing in Italian breast and/or ovarian cancer families: mutation spectrum and prevalence and analysis of mutation prediction models. *Ann Oncol*. 2006;17 Suppl 7:vii34-vii40. doi:10.1093/annonc/mdl947.
34. Capalbo C, Ricevuto E, Vestri A, et al. Improving the accuracy of BRCA1/2 mutation prediction: validation of the novel country-customized IC software. *Eur J Hum Genet*. 2006;14(1):49-54. doi:10.1038/sj.ejhg.5201511.
35. Coppa A, Buffone A, Capalbo C, et al. Novel and recurrent BRCA2 mutations in Italian breast/ovarian cancer families widen the ovarian cancer cluster region boundaries to exons 13 and 14. *Breast Cancer Res Treat*. 2014;148(3):629-635. doi:10.1007/s10549-014-3196-z.
36. Belardinilli F, Capalbo C, Buffone A, et al. Validation of the Ion Torrent PGM sequencing for the prospective routine molecular diagnostic of colorectal cancer. *Clin Biochem*. 2015;48(13-14):908-910. doi:10.1016/j.clinbiochem.2015.04.003.
37. Schouten JP, McElgunn CJ, Waaijer R, Zwiijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res*. 2002;30(12):e57.

- 416 38. Wilson EB. 1927. Probable Inference, the Law of Succession, and Statistical Inference.  
417 *Journal of the American Statistical Association* 22:209–212 DOI  
418 10.1080/01621459.1927.10502953.
- 419 39. Apressos A, Agiannitopoulos K, Pepe G, et al. Comprehensive BRCA mutation analysis in  
420 the Greek population. Experience from a single clinical diagnostic center. *Cancer Genet.*  
421 2018;220:1-12. doi:10.1016/j.cancergen.2017.10.002.
- 422 40. Pilato B, Pinto R, De Summa S, et al. BRCA1-2 diagnostic workflow from next-generation  
423 sequencing technologies to variant identification and final report. *Genes Chromosomes*  
424 *Cancer*. 2016;55(10):803-813. doi:10.1002/gcc.22383.
- 425 41. Ticha I, Kleibl Z, Stribrna J, et al. Screening for genomic rearrangements in BRCA1 and  
426 BRCA2 genes in Czech high-risk breast/ovarian cancer patients: high proportion of  
427 population specific alterations in BRCA1 gene. *Breast Cancer Res Treat*. 2010;124(2):337-  
428 347. doi:10.1007/s10549-010-0745-y.
- 429 42. Vasickova P, Machackova E, Lukesova M, et al. High occurrence of BRCA1 intragenic  
430 rearrangements in hereditary breast and ovarian cancer syndrome in the Czech Republic.  
431 *BMC Med Genet*. 2007;8:32. doi:10.1186/1471-2350-8-32.
- 432 43. Hasmad HN, Sivanandan K, Lee V, Yip CH, Mohd Taib NA, Teo SH. Identification of a  
433 recurrent BRCA1 exon 21-22 genomic rearrangement in Malay breast cancer patients. *Clin*  
434 *Genet*. 2015;87(4):392-394. doi:10.1111/cge.12451.

44. Armaou S, Konstantopoulou I, Anagnostopoulos T, et al. Novel genomic rearrangements in the BRCA1 gene detected in Greek breast/ovarian cancer patients. *European Journal of Cancer*. 2007;43(2):443-453. doi:10.1016/j.ejca.2006.09.021.
45. Belogianni I, Apeessos A, Mihalatos M, et al. Characterization of a novel large deletion and single point mutations in the BRCA1 gene in a Greek cohort of families with suspected hereditary breast cancer. *BMC Cancer*. 2004;4:61. doi:10.1186/1471-2407-4-61.
46. Carvalho M, Pino MA, Karchin R, et al. Analysis of a set of missense, frameshift, and in-frame deletion variants of BRCA1. *Mutat Res*. 2009;660(1-2):1-11. doi:10.1016/j.mrfmmm.2008.09.017.
47. Judkins T, Rosenthal E, Arnell C, et al. Clinical significance of large rearrangements in BRCA1 and BRCA2. *Cancer*. 2012;118(21):5210-5216. doi:10.1002/cncr.27556.
48. Mazoyer S. Genomic rearrangements in the BRCA1 and BRCA2 genes. *Hum Mutat*. 2005;25(5):415-422. doi:10.1002/humu.20169.
49. Ewald IP, Ribeiro PLI, Palmero EI, Cossio SL, Giugliani R, Ashton-Prolla P. Genomic rearrangements in BRCA1 and BRCA2: A literature review. *Genet Mol Biol*. 2009;32(3):437-446. doi:10.1590/S1415-47572009005000049.

**Table 1** (on next page)

LGRs in TS and VS



1  
2  
3

	sample Id	genomic variant	exon deletion	ref
TS	BR59	c.5407-?_*(1_?)del	exon 23-24 del	20
	BR328	c.5075-?_5193+?del	exon 18-19 del	20
	BR409	NBR2delEX1_BRCA1delEX1-2	exon 1 NBR2 del exon 1-2 BRCA1 del	3

VS	BR963	NG_005905.2: g.163181_169408del6228	exon 21-22 del	/
	BR1154	NG_005905.2: g.160396_164568del4173	exon 20 del	/
	BR1379	NG_005905.2:g.145185_151339del6155	exon 16-17 del	/

4  
5  
6  
7  
8  
9

## **Table 2**(on next page)

CNVs prediction by IR software algorithm in TS and VS

The confidence score is the probability that the number of copies of the region of interest is different from 2, which is the normal value, while the confident precision indicates how much the algorithm is certain of the accuracy of the number of copies estimated by the analysis.

1  
2

	Sample ID	Locus	Type	Genes	Location	Length	Copy Number	CytoBand	CNV Confidence	CNV Precision
TS	BR59	chr17:41197602	CNV	BRCA1	exon 23-24	2.138kb	1	17q21.31(41197602-41199740)x1	5.66	5.66
	BR328	chr17:41215277	CNV	BRCA1	exon 18-19	749kb	1	17q21.31(41215277-41216026)x1	13.05	13.05
	BR409	chr17:41275973	CNV	BRCA1	exon 2	275kb	1	17q21.31(41275973-41276248)x1	1.14	1.14
VS	BR963	chr17:41201074	CNV	BRCA1	exon 21-22	2.18kb	1	17q21.31(41201074-41203254)x1	9.14	9.14
	BR1379	chr17:41215855	CNV	BRCA1	exon 16-18	7.44kb	1	17q21.31(41215855-41223295)x1	5.11	5.11

3  
4

# **Table 3**(on next page)

## Performance of NGS-dependent LGRs analysis

Validation metrics were defined as: Accuracy =  $(TP + TN)/(TP + FP + TN + FN)$ ; Sensitivity =  $TP/(TP + FN)$ ; Specificity =  $TN/(TN+FP)$ ; FDR =  $FP/(TP + FP)$ , Negative Predictive Value =  $TN/(TN + FN)$ , where TP = true positives, TN = true negatives, FP = false positives, FN = false negatives.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18

			MLPA		Results
		tot	CNV	no CNV	
DQ	CNV	33	3	30	64.7% accuracy [95% CI: 50.6-76.7] 100% sensitivity [95% CI: 22.8-98.4]
	no CNV	52	0	52	63.4% specificity [95% CI: 49-75.8] 100% NPV [95% CI: 87.6-99.9]
IR	CNV	29	2	27	67.1% accuracy [95% CI: 52.9-78.7] 66.7% sensitivity [95% CI: 8.9-98.8]
	no CNV	56	1	55	67.1% specificity [95% CI: 52.7-78.9] 98.2% NPV [95% CI: 85,6-99,9]

# Figure 1

DQ analyses for the training set samples.

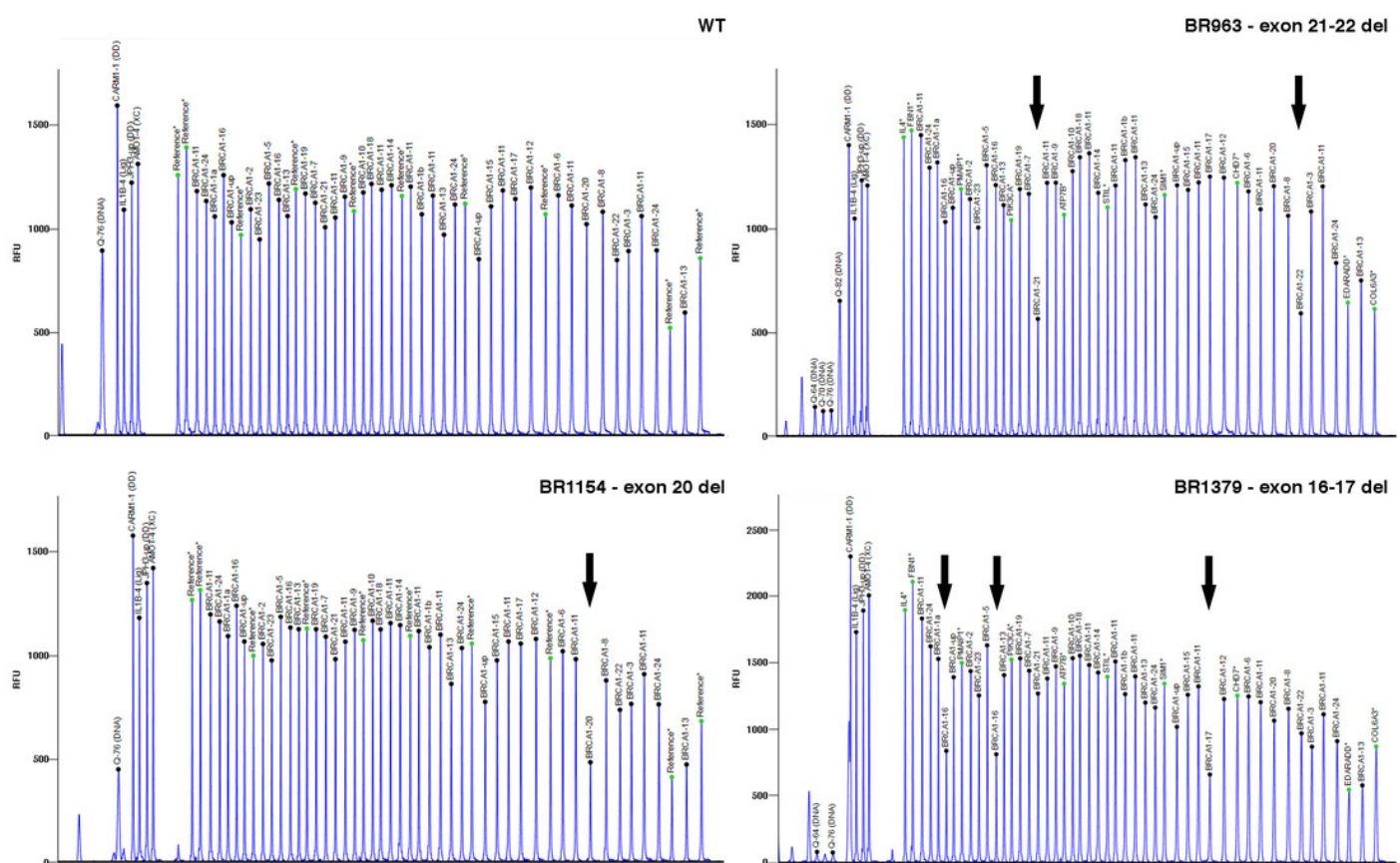
**a:** for each sample, every peak represents the ratio of the amplicon read count normalized on *BRCA1/BRCA2* total reads and the average of normalized amplicon read counts from all samples on a single chip (DQA). **b:** for each sample, every peak represents the ratio of the amplicon read count normalized on *BRCA1/BRCA2* total reads and the average of the coverage data of a baseline built from 127 LGRs negative samples (DQB1). The threshold =  $\text{mean} \pm 2 \text{ SD}$ . Value  $> \text{mean} \pm 2 \text{ SD}$  is indicative of a duplication; Value  $< \text{mean} \pm 2 \text{ SD}$  is indicative of a deletion. \* indicated the amplicons included in the region involved in the rearrangement as confirmed by MLPA analysis.



# Figure 2

BRCA1 MLPA electropherogram showing aberrant profiles in BR963, BR1154, and BR1379 patients.

Black arrows indicate the deletion of *BRCA1* exons 21-22 (BR963), 20 (BR1154), and 16-17 (BR1379), respectively. Wild-type sample (WT) has been showed in the first panel.

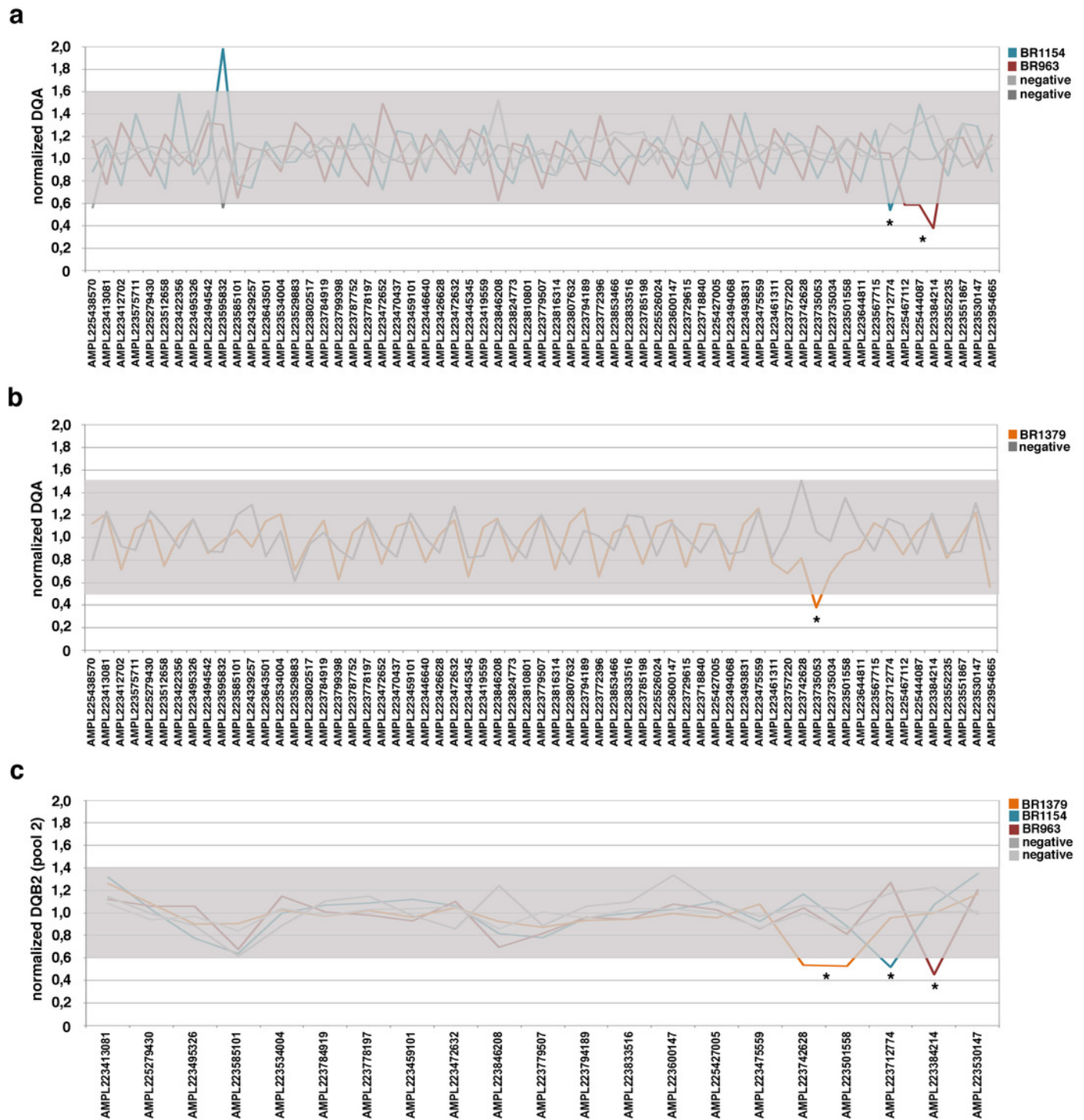




# Figure 3

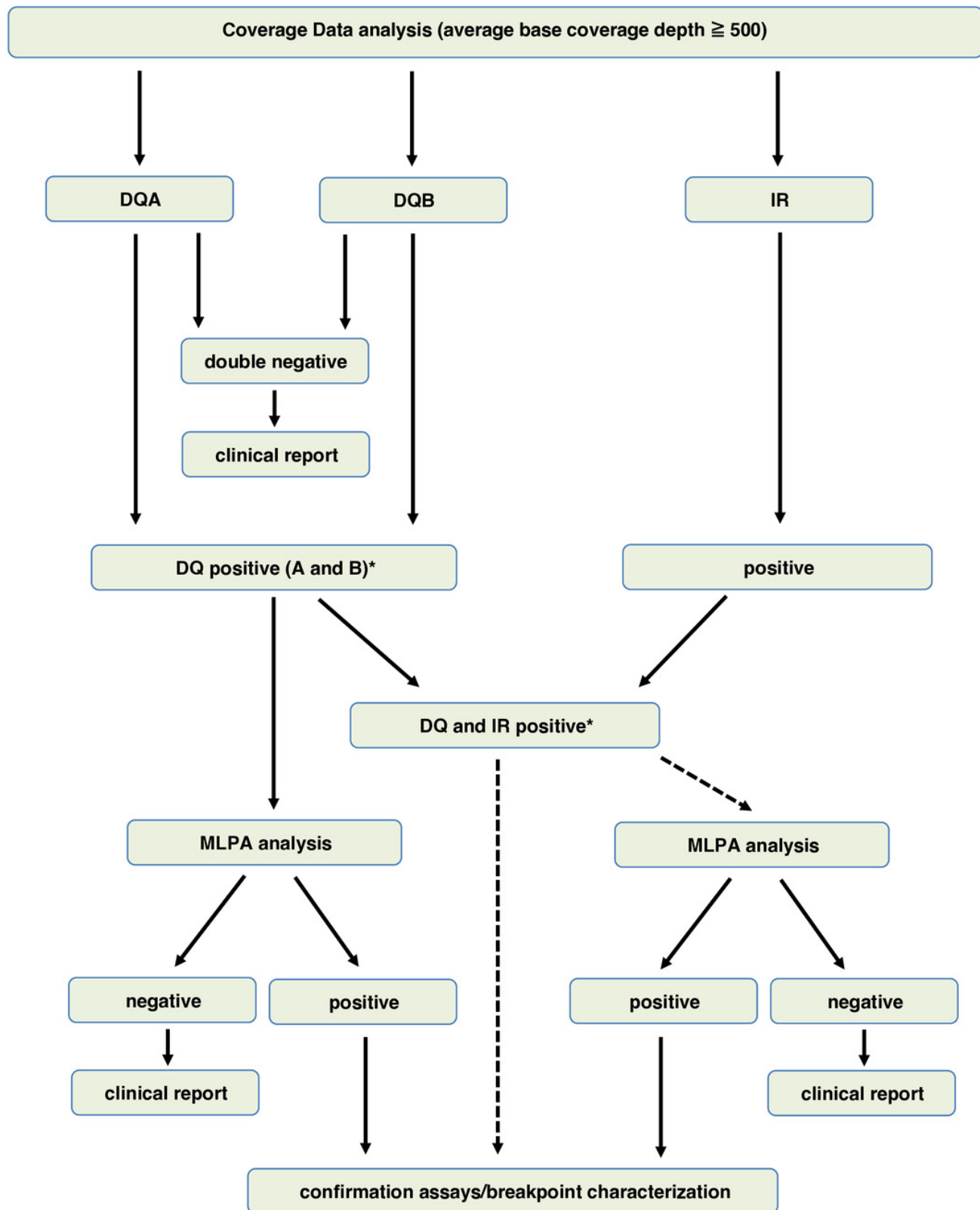
DQ analyses for the representative samples for VS

**a** and **b**: for each sample, every peak represents the ratio of the amplicon read count normalized on *BRCA1/BRCA2* total reads and the average of normalized amplicon read counts from all samples on a single chip (DQA). **c**: for each sample, every peak represents the ratio of the amplicon read count normalized on *BRCA1/BRCA2* total reads and the average of the coverage data of a baseline built from 127 LGRs negative samples considering separately the amplicon pools (DQB2, pool 2). The threshold= mean  $\pm$  2 SD. Value  $>$  mean  $\pm$  2 SD is indicative of a duplication; Value  $<$  mean  $\pm$  2 SD is indicative of a deletion. \* indicated the amplicons included in the region involved in the rearrangement as confirmed by MLPA analysis.



# Figure 4

Operative algorithm to select samples for MLPA analysis.



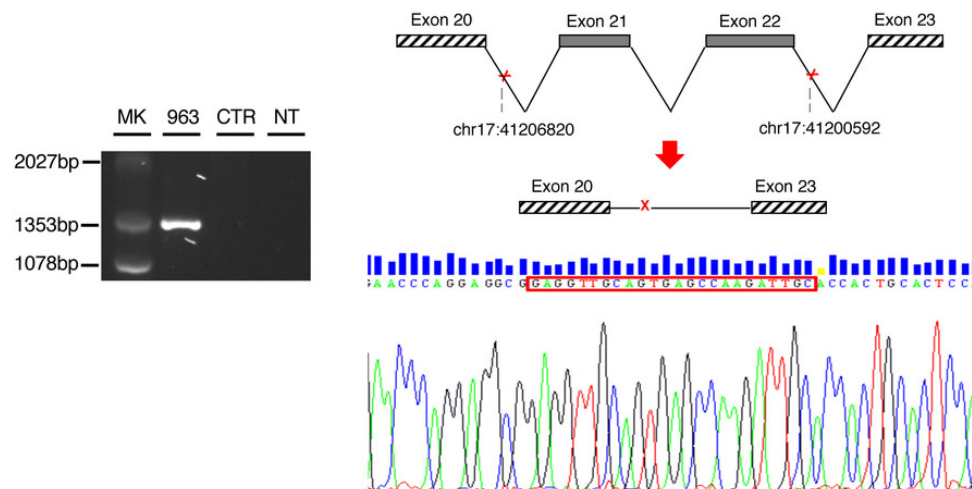
# Figure 5

## Characterization of BRCA1 LGRs

**a:** On the left, gel image of PCR products. PCR amplification of the genomic region spanning the *BRCA1* rearrangement resulted in a fragment of approximately 1353 bp present only in the proband BR963. On the right, schematic representation and electropherogram showing the *BRCA1* exons 21 and 22 deletion. **b:** On the left, gel image of PCR products. PCR amplification of the genomic region spanning the *BRCA1* rearrangement resulted in a fragment of approximately 872 bp present in the proband BR1154 and in her mother BR1148. On the right, schematic representation and electropherogram showing the *BRCA1* exon 20 deletion. **c:** On the left, gel image of PCR products. PCR amplification of the genomic region spanning the *BRCA1* rearrangement resulted in a fragment of approximately 2027 bp present only in the proband BR1379. On the right, schematic representation and electropherogram showing the *BRCA1* exons 16 and 17 deletion. MK, marker; NT, no template; CTR healthy individual DNA.

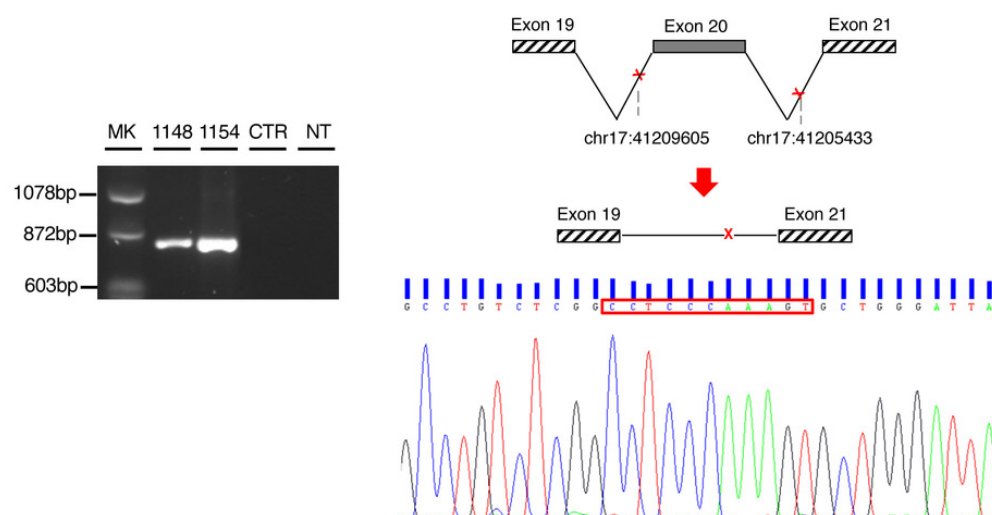
**a**

**BR963 NG\_005905.2:g.163181\_169408del6228**



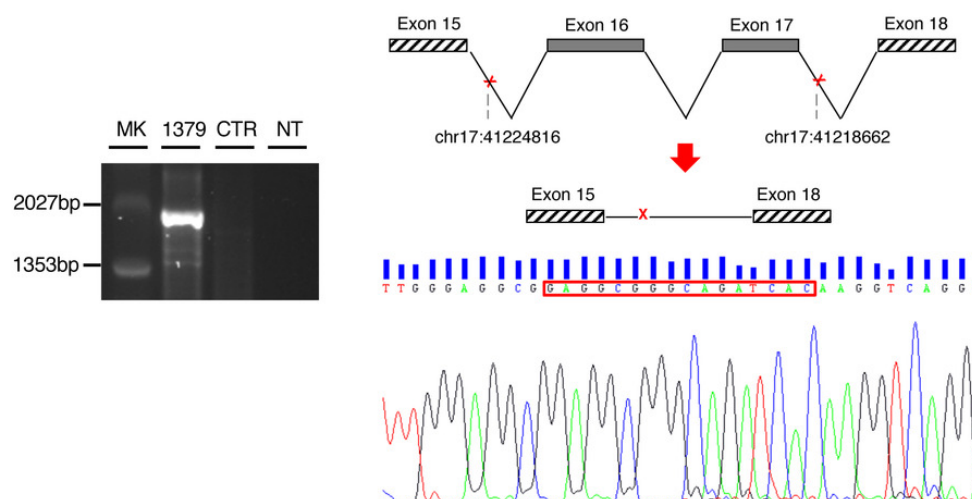
**b**

**BR1154 e BR1148 NG\_005905.2:g.160396\_164568del4173**



**c**

**BR1379 NG\_005905.2:g.145185\_151339del6155**

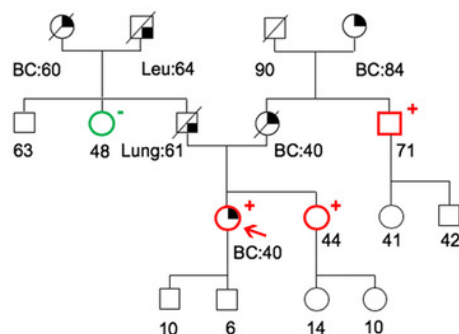


# Figure 6

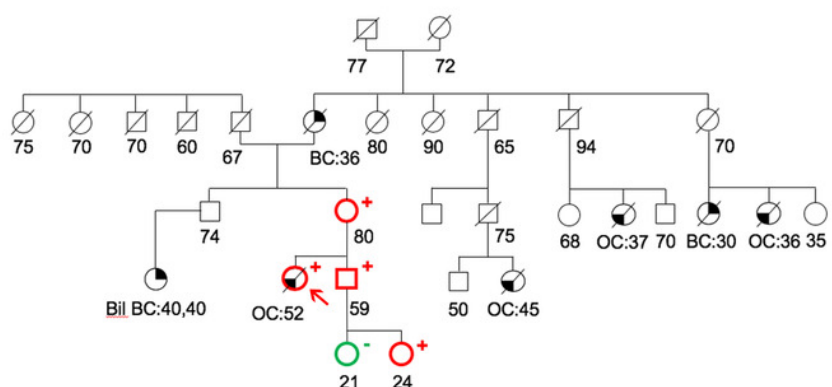
Pedigree of the HBC or HBOC family carriers of *BRCA1* LGRs

Proband, indicated with an arrow, present novel LGRs involving the exons 21-22 (BR963), the exon 20 (BR1154) and the exon 16-17 (BR1379) of *BRCA1*. Cancer type and age at diagnosis are reported and described as: BC, breast cancer; Pan, pancreas; Leu, leukemia; Lung; bil BC, bilateral breast cancer; OC, ovarian cancer; Hep, hepatoma; CC, colon cancer

**BR963**  
BRCA1:NG\_005905.2:g.163181\_169408del6228



**BR1154**  
BRCA1: NG\_005905.2:g.160396\_164568del4173



**BR1379**  
BRCA1: NG\_005905.2:g.145185\_151339del6155

