Clinical Chemistry 63:2
503–512 (2017)

BRCA Testing by Single-Molecule Molecular Inversion Probes

Kornelia Neveling,[1†] Arjen R. Mensenkamp,[1†] Ronny Derks,[1] Michael Kwint,[1] Hicham Ouchene,[1] Marloes Steehouwer,[1] Bart van Lier,[1] Ermanno Bosgoed,[1] Alwin Rikken,[1] Marloes Tychon,[1] Dimitra Zafeiropoulou,[1] Steven Castelein,[1] Jayne Hehir-Kwa,[1] Djie Tjwan Thung,[1] Tom Hofste,[1] Stefan H. Lelieveld,[1] Stijn M.M. Bertens,[2] Ivo B.J.F. Adan,[2] Astrid Eijkelenboom,[3] Bastiaan B. Tops,[3] Helger Yntema,[1] Tomasz Stokowy,[4,5] Per M. Knappskog,[4] Hildegunn Høberg-Vetti,[4,5] Vidar M. Steen,[4,6] Evan Boyle,[7] Beth Martin,[7] Marjolijn J.L. Ligtenberg,[1,3] Jay Shendure,[7†] Marcel R. Nelen,[1*†] and Alexander Hoischen[1,8†]

BACKGROUND: Despite advances in next generation DNA sequencing (NGS), NGS-based single gene tests for diagnostic purposes require improvements in terms of completeness, quality, speed, and cost. Single-molecule molecular inversion probes (smMIPs) are a technology with unrealized potential in the area of clinical genetic testing. In this proof-of-concept study, we selected 2 frequently requested gene tests, those for the breast cancer genes *BRCA1* and *BRCA2*, and developed an automated work flow based on smMIPs.

METHODS: The *BRCA1* and *BRCA2* smMIPs were validated using 166 human genomic DNA samples with known variant status. A generic automated work flow was built to perform smMIP-based enrichment and sequencing for *BRCA1*, *BRCA2*, and the checkpoint kinase 2 (*CHEK2*) c.1100del variant.

RESULTS: Pathogenic and benign variants were analyzed in a subset of 152 previously BRCA-genotyped samples, yielding an analytical sensitivity and specificity of 100%. Following automation, blind analysis of 65 in-house samples and 267 Norwegian samples correctly identified all true-positive variants (>3000), with no false positives. Consequent to process optimization, turnaround times were reduced by 60% to currently 10–15 days. Copy number variants were detected with an analytical sensitivity of 100% and an analytical specificity of 88%.

CONCLUSIONS: smMIP-based genetic testing enables automated and reliable analysis of the coding sequences of *BRCA1* and *BRCA2*. The use of single-molecule tags, double-tiled targeted enrichment, and capturing and sequencing in duplo, in combination with automated library preparation and data analysis, results in a robust process and reduces routine turnaround times. Furthermore, smMIP-based copy number variation analysis could make independent copy number variation tools like multiplex ligation-dependent probes amplification dispensable.

The growing influence of genetic data on clinical management and therapy demands improvements in both speed and comprehensiveness of genetic testing. Recent examples that reflect this trend include the impact of exome sequencing on therapeutic decisions in neonatal care *(1)* and *BRCA1*[9] and *BRCA2* gene analyses for surgical choices and potential treatment with poly (ADP-ribose) polymerase inhibitors *(2, 3)*. Combined with finite resources to cover growing numbers of diagnostic requests, there is a need to revise existing laboratory practices. Therefore, we set out to develop an automated laboratory work flow that was fast, stable, and flexible with respect to the increasing numbers of requested tests and samples, while also delivering comprehensive high-quality data. We hypothesized that molecular inversion probes (MIPs)[10] would serve these needs for target enrichment.

[1] Department of Human Genetics, Radboud university medical center, Nijmegen, the Netherlands; [2] Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands; [3] Department of Pathology, Radboud university medical center, Nijmegen, the Netherlands; [4] Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; [5] Western Norway Familial Cancer Center, Haukeland University Hospital, Bergen, Norway; [6] Department of Clinical Science, University of Bergen, Bergen, Norway; [7] Department of Genome Sciences, University of Washington, Seattle, WA; [8] Donders Centre for Neuroscience, Radboud University Nijmegen, Nijmegen, the Netherlands.

* Address correspondence to this author at: Department of Human Genetics, Radboud university medical center, PO Box 9101, Nijmegen 6500 HB, the Netherlands. Fax +31-24-36-16658; e-mail marcel.nelen@radboudumc.nl.

† Kornelia Neveling and Arjen R. Mensenkamp contributed equally to this work, and should be considered as first authors. Jay Shendure, Marcel R. Nelen, and Alexander Hoischen contributed equally to this work, and should be considered as last authors.

Received July 26, 2016; accepted September 29, 2016.
Previously published online at DOI: 10.1373/clinchem.2016.263897

[9] Human genes: *BRCA1; BRCA2*; *CHEK2,* checkpoint kinase 2.

[10] Nonstandard abbreviations: MIP, molecular inversion probe; NGS, next-generation sequencing; smMIP, single-molecule MIP; SNV, single nucleotide variant; CNV, copy number variant; SNP, single nucleotide polymorphism; TAT, turnaround time.

First described in 1994 as padlock probes for multiplex sequence detection *(4)* or genotyping *(5)*, MIPs were subsequently adapted to enable multiplex targeted sequence capture in the context of next generation sequencing (NGS) *(6, 7)*. Over the years, protocols for sequence capture with MIPs have been improved, e.g., by adding a single-molecule tag that enables the differentiation between PCR-duplicated and truly independent sequencing reads [single-molecule MIPs (smMIPs)] *(8, 9)*.

There are many potential advantages of the smMIP technology for clinical sequencing, including the fact that it is flexible and easily optimized (e.g., single gene vs panels; adding new genes when needed; adjusting probe concentrations or designs to improve performance), automatable (targeted capture and sample barcoding through a small number of enzymatic reactions; crucial for laboratories dealing with high numbers of diagnostic requests), reproducible [to be used for single nucleotide variant (SNV) as well as copy number variant (CNV) analysis], and inexpensive relative to commercial enrichment kits. However, smMIPs have primarily been used in an academic setting such as targeted sequencing of candidate disease genes in large cohorts.

We here describe a fully automated clinical-sequencing work flow based on smMIP enrichment *(9)*, in combination with 2-color NGS (NextSeq500, Illumina). As a proof of principle, we focused on *BRCA1* and *BRCA2*. Since the discovery that pathogenic variants in *BRCA1* and *BRCA2* genes predispose to hereditary breast and ovarian cancer, variant analysis of these genes has been offered to women with a concerning family history, and even population screening is suggested *(10)*. For breast cancer, the availability of results within days is important because the variant status may affect the choice of surgical strategy and chemotherapy *(11–13)*. Furthermore, poly (ADP-ribose) polymerase inhibitors are available for women with ovarian cancer and germline or somatic pathogenic variants in *BRCA1* or *BRCA2* and are evaluated for other cancer types like breast, pancreatic, and prostate cancer *(2, 3, 14)*. Current tests include Sanger sequencing *(15, 16)* and several NGS-based tests *(17–20)*. Here we present a new approach that offers several advantages.

Methods

GENOMIC DNA SAMPLES

All individuals gave written informed consent for *BRCA1* and *BRCA2* testing. Human genomic DNA samples were isolated from EDTA blood. The study cohort consisted of patients who visited the outpatient clinic of the department of Human Genetics of the Radboud University Medical Center in Nijmegen, the Netherlands, or the Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway. All samples were blinded but had a known variant status as determined by either Sanger sequencing or ion semiconductor sequencing after AmpliSeq enrichment.

The MIP design and protocol are based on previous methods *(8)*. Modifications of this protocol and its implementation to an automated work flow are described below.

DESIGN OF smMIPs

smMIPs for *BRCA1, BRCA2*, and checkpoint kinase 2 (*CHEK2*) (GenBank reference sequences *BRCA1*: NM_007294.3, *BRCA2*: NM_00059.3, and *CHEK2*: NM_007194.3) were designed using the MIPgen pipeline *(21)*. For details of the design see Supplemental Methods and Supplemental Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol63/issue2.

PHOSPHORYLATION AND DILUTION OF smMIPs

Phosphorylation was performed after pooling of all smMIPs. The phosphorylation mix included: 50 μL pooled smMIPs after rebalancing, 2 μL T_4 polynucleotide kinase (New England Biolabs), 2 μL H_2O, 6.0 μL 10× T_4 DNA ligase buffer with 10 mmol/L ATP (New England Biolabs) (total volume: 60 μL). The PCR program used the following conditions: 45 min 37 °C, 20 min 65 °C, storage at 4 °C. Phosphorylated smMIP pools were diluted to reach a ratio of 800:1 in the final capture reaction (smMIPs:DNA molecules). This dilution was variable, depending on the number of smMIPs present in the pools.

DNA ISOLATION

Genomic DNA isolation was performed as described previously *(22)*.

AUTOMATED smMIP LIBRARY PREPARATION

A Hamilton Microlab Star Plus robot was used for pipetting the capture mastermix at 4 °C, which contained per reaction: 2.5 μL 10× Ampligase buffer (Epicentre/Illumina), 0.03 μL 0.25 mmol/L dNTPs (deoxynucleotide triphosphates), 0.32 μL HemoKlentaq (New England Biolabs, 10 U/μL), 0.01 μL Ampligase (Epicentre/Illumina, 100 U/μL), 12.44 μL H_2O, patient's genomic DNA (6.7 μL, 15 ng/μL), and 3.0 μL smMIP dilution (2 μL smMIP pool + 88 μL H_2O). smMIP capture was performed for 18 h overnight (10 min 95 °C, 18 h 60 °C, storage at 4 °C), followed by exonuclease treatment [0.5 μL EXO I (New England Biolabs), 0.5 μL EXO III (New England Biolabs), 0.2 μL 10× Ampligase buffer (Epicentre/Illumina), 0.8 μL H_2O] using the following program: 45 min 37 °C, 2 min 95 °C, storage at 4 °C. The exonuclease treatment was performed manually at a distinct location to prevent exonuclease contamination

in pre- and post-PCR environments. Following exonuclease treatment, postcapture PCR including barcodes was performed on a MicroLab Starlet Replicator Robot (Hamilton). The PCR mixture was: 25 μL 2× iProof (Bio-Rad Laboratories), 2.5 μL 5 μmol/L smMIP forward primer (Integrated DNA Technologies), 2.5 μL 5 μmol/L smMIP reverse primer (including barcode, Integrated DNA Technologies), 20 μL Exonuclease-treated smMIP sample (custom PCR barcoded primers were used as described by *(8)*). The PCR program was: 30 s 98 °C, 24× (10 s 98 °C, 30 s 60 °C, 30 s 72 °C), 2 min 72 °C, storage at 4 °C. The mapping from DNA sample to barcode sequence was logged in an associated database; the respective sample sheet was created automatically. After PCR, all samples from one 96-well plate were pooled using the MicroLab Starlet Replicator Robot (Hamilton), and 2 μL PCR-product per sample. The final volume was added up to 200 μL using 0.1× Tris-EDTA buffer. Each pool was independently purified using 0.8× Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), and concentrations were measured using a Qubit 2.0 Fluorometer (Invitrogen). If more than 1 pool was present, pools were combined to 1 megapool.

SEQUENCING

The final pooled smMIP libraries were denaturated and diluted to a sequencing concentration of 1.2 pmol/L. Sequencing was performed on a NextSeq500 (Illumina) according to manufacturer's instructions [300 cycles Mid-Output sequencing kit (Norwegian data set: 300 cycles High-Output kit)]. smMIP libraries required spike-in of custom primers as described previously *(8)*.

VARIANT NOMENCLATURE

The variant nomenclature follows the guidelines indicated by the Human Genome Variation Society v2.0.

DATA ANALYSIS

By using the generated sample sheet, BCL-to-FASTQ conversion and demultiplexing were performed without manual intervention (bcl2fastq v2.14, Illumina). Demultiplexed FASTQ files were uploaded to a server running the analysis software SeqNext (Sequence Pilot™, version 4.2.2 Build 502, JSI medical systems). Within SeqNext, sequencing reads were mapped to region of interest and primer sequences (MIP extension and ligation arms), and variant calling (for SNVs and small indels) was performed (excluding MIP extension and ligation arms). A minimal coverage of 20 unique reads, i.e., unique patient molecules, was required. Since the smMIPs were sequenced in both directions, this corresponded to 40× total reads. This enabled the detection of variants with at least 35% variant reads with >95% confidence. Variants were called at >5% variant reads if present in >3 unique reads. These settings were established for the detection of germline mutations and enabled exclusion of sequencing errors using the random tag of each smMIP. Single molecules, or unique reads, were detected using the filter settings available in the software. Consensus reads were selected on their random 5N tag. The consensus nucleotide was called when present in >50% of the contributing duplicate reads. If >30% of the duplicate reads contributing to a consensus read deviated from the consensus read, the consensus read was ignored.

CNV ANALYSIS

CNV read-depth analysis was performed to identify deletions and duplications. The FASTQ files were mapped using BWA-MEM (default settings to hg19) *(23)*. BAM files were filtered for Q20 reads and read depth per target calculated. Only unique reads were taken into account. Read depth ratios were calculated using a self-referencing batch and normalizing for the total number of reads in the sample and the mean coverage per exonic target within the sequencing run. Three test runs containing positive controls were analyzed (see online Supplemental Table 2), followed by a blinded set containing 36 samples run in duplo. Segmentation was performed on the test vs reference ratios using 3 states representing a duplication (0.3), no copy number change (0) and a deletion (−0.5), requiring a positive call in both duplo samples.

Results

VALIDATION OF smMIPs

For validation of smMIPs for clinical *BRCA1* and *BRCA2* sequencing, 166 human genomic DNA samples were analyzed, 90 samples carrying a pathogenic variant in *BRCA1*, and 77 samples carrying a pathogenic variant in *BRCA2*. One sample had both a pathogenic *BRCA1* and *BRCA2* variant. The 166 samples contained 110 unique variants, most of them pathogenic variants identified in our laboratory between January 2010 and July 2014. smMIP capture was performed manually and in duplicate. Sequencing was completed in 3 independent sequencing runs (runs 1–3). The mean (SD) coverage per smMIP (unique reads only) was 359× (159×) for *BRCA1* and 289× (137×) for *BRCA2* (Fig. 1, Table 1). On average, 0.4% of the individual *BRCA1* smMIPs and 2.4% of the individual *BRCA2* smMIPs had a coverage <40× (in total 17 smMIPs). Still, the "horizontal" sequence coverage per gene was 100%, with all nucleotides showing >40× coverage due to overlapping smMIPs that redundantly covered the same coordinates (Fig. 1, Table 1). Variant calling resulted in correct identification of all 110 pathogenic variants (Fig. 2, online Supplemental Tables 3–6). Analytical sensitivity and specificity were calculated to be 100%, based on all variants present in a

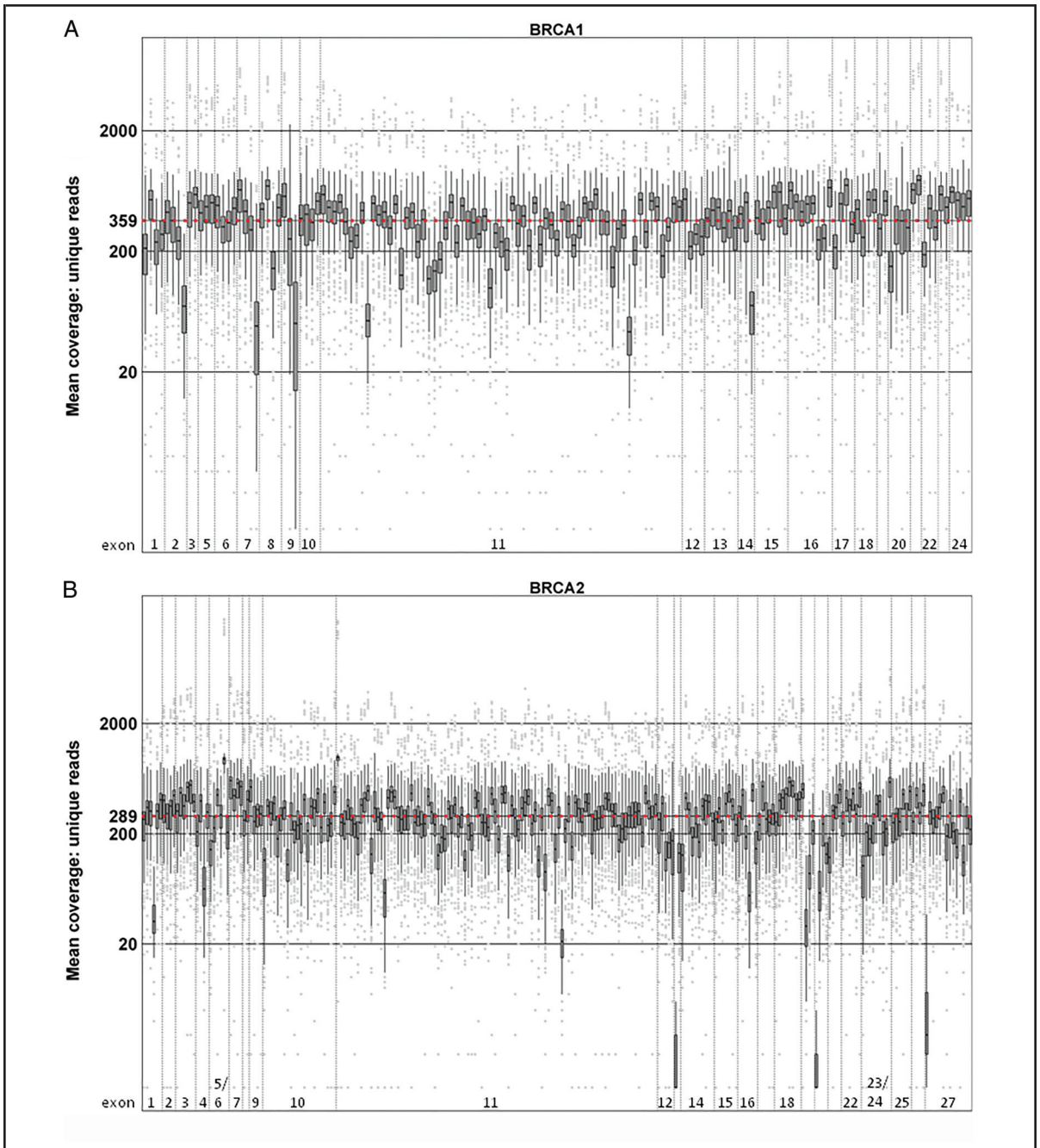

Fig. 1. Coverage plots of *BRCA1* and *BRCA2*. The unique number of reads per smMIP for *BRCA1* and *BRCA2*.

x Axis: unique smMIPs used to enrich *BRCA1* and *BRCA2*, respectively, sorted per exon. *BRCA1*: exon 1–24 (exon 4 is not existing), *BRCA2*: exon 1–25 (exons 5 and 6, and 23 and 24 are combined, respectively). *y* Axis: mean coverage (unique reads) per smMIP.

subset of 152 samples that were previously sequenced for *BRCA1* and *BRCA2* either by Sanger or semiconductor sequencing (Table 2). Three variants were identified that had been missed previously either by Sanger (2 variants) or by semiconductor sequencing (1 variant). Two of them were benign variants not detected earlier due to adjacent frameshift-causing variant; the other one was a benign variant not present in the data due to allelic dropout, introduced by the pathogenic variant located underneath the overlapping PCR primer (see online Supplemental Table 7).

Table 1. Validation of *BRCA* smMIPs.

	Run 1	Run 2	Run 3
Cluster density per run	78 000 K/mm^2	110 000 K/mm^2	60 000 K/mm^2
Mean (SD) coverage (unique reads) for smMIPs *BRCA1*	416 (129)	360 (195)	302 (153)
Mean (SD) coverage (unique reads) for smMIPs *BRCA2*	297 (99)	296 (170)	276 (141)
% *BRCA1* smMIPs with >40 unique reads	100%	100%	98.7%
% *BRCA2* smMIPs with >40 unique reads	97.6%	97.6%	97.6%
% *BRCA1* smMIPs with >100 unique reads	96.0%	100%	95.3%
% *BRCA2* smMIPs with >100 unique reads	95.2%	94.8%	92.3%
Horizontal coverage *BRCA1* (−20..+20) >40 unique reads	100%	100%	100%
Horizontal coverage *BRCA2* (−20..+20) >40 unique reads	100%	100%	100%
% *BRCA1* nucleotides covered ≥2 smMIPs	100%	100%	98.0%
% *BRCA2* nucleotides covered ≥2 smMIPs	97.6%	97.6%	97.6%

IMPLEMENTATION AND VALIDATION OF AN AUTOMATED smMIP SEQUENCING WORK FLOW

Despite 100% horizontal coverage of both genes, 35 smMIPs were redesigned, covering the regions of the 17 poorly performing smMIPs for *BRCA1* and *BRCA2,* and adding smMIPs for *CHEK2*_c.1100del. We also included additional smMIPs for which single nucleotide polymorphisms (SNPs) in one of the hybridization arms were identified. These SNP-smMIPs only differ at the position of the SNP, thus allowing both alleles to be captured and minimizing allelic dropout. All smMIPs for *BRCA1*, *BRCA2*, and *CHEK2* were pooled in a single pool. The new pool was validated using 22 known samples, carrying 10 pathogenic *BRCA1* variants, 10 pathogenic *BRCA2* variants, and 2 pathogenic *CHEK2* variants (see online Supplemental Table 8). For these experiments, smMIP capture and library construction was performed using a newly developed generic automated smMIP work flow (Fig. 3). Three independent test runs (library preparation plus sequencing) were performed. FASTQ files were automatically imported into the SeqNext module (JSI medical systems). All variants were identified correctly (see online Supplemental Table 8).

PROCESS OPTIMIZATION

Several scenarios regarding implementation of the new work flow (from available isolated DNA until availability of data) were simulated to achieve the best possible outcome concerning fast and stable processing with low flow times and costs. On the basis of these simulations, we elected to execute the new work flow twice per week, with every sample being tested in duplo.

Running the process twice per week was a consequence of simulations of turnaround times (TATs) when sequencing daily, or once, twice, or 3 times a week. With sequencing once per week, the waiting time for a sample until a processing was started had an upper limit of 6 days (see online Supplemental Fig. 2A). For sequencing twice per week, the waiting time was at most 3 days. Sequencing more often would only reduce TATs when weekends were included as working days (see online Supplemental Fig. 2A). We now initiate smMIP capture twice a week, on Mondays and Thursdays, and sequencing data are available on Thursdays and Mondays, respectively. Samples arriving just after a process has started need to wait, with a maximum of 3 days. The practical work flow is therefore approximately 4 days for up to 400 samples a week, with 2 runs per week (and no weekend shift) (see online Supplemental Fig. 2B).

Capturing and sequencing samples in duplo was beneficial to obviate potential inefficiencies in the work flow (see online Supplemental Fig. 3). In case of potential (technical) failure of a sample, duplicate samples likely produce useful data. In case data from both duplicates are available but coverage is critical (usually due to poor DNA quality or quantity), data from duplicates can be combined to reach sufficient sequencing depth. In the more likely scenario where both duplicates perform well and a variant is identified, the duplicate can be used as confirmation, thereby excluding sample swaps and obviating the need for a serial run to provide validation and thereby extending the TAT.

DOUBLE-BLIND SEQUENCING OF ROUTINE BRCA SAMPLES

Diagnostic testing was performed using the work flow as described above, in parallel to the routine method (ion semiconductor sequencing after AmpliSeq enrichment, in combination with Sanger sequencing) for 65 different samples (see online Supplemental Table 9). All tests were completed independently by different employees, and results were compared afterward. Data analyses showed

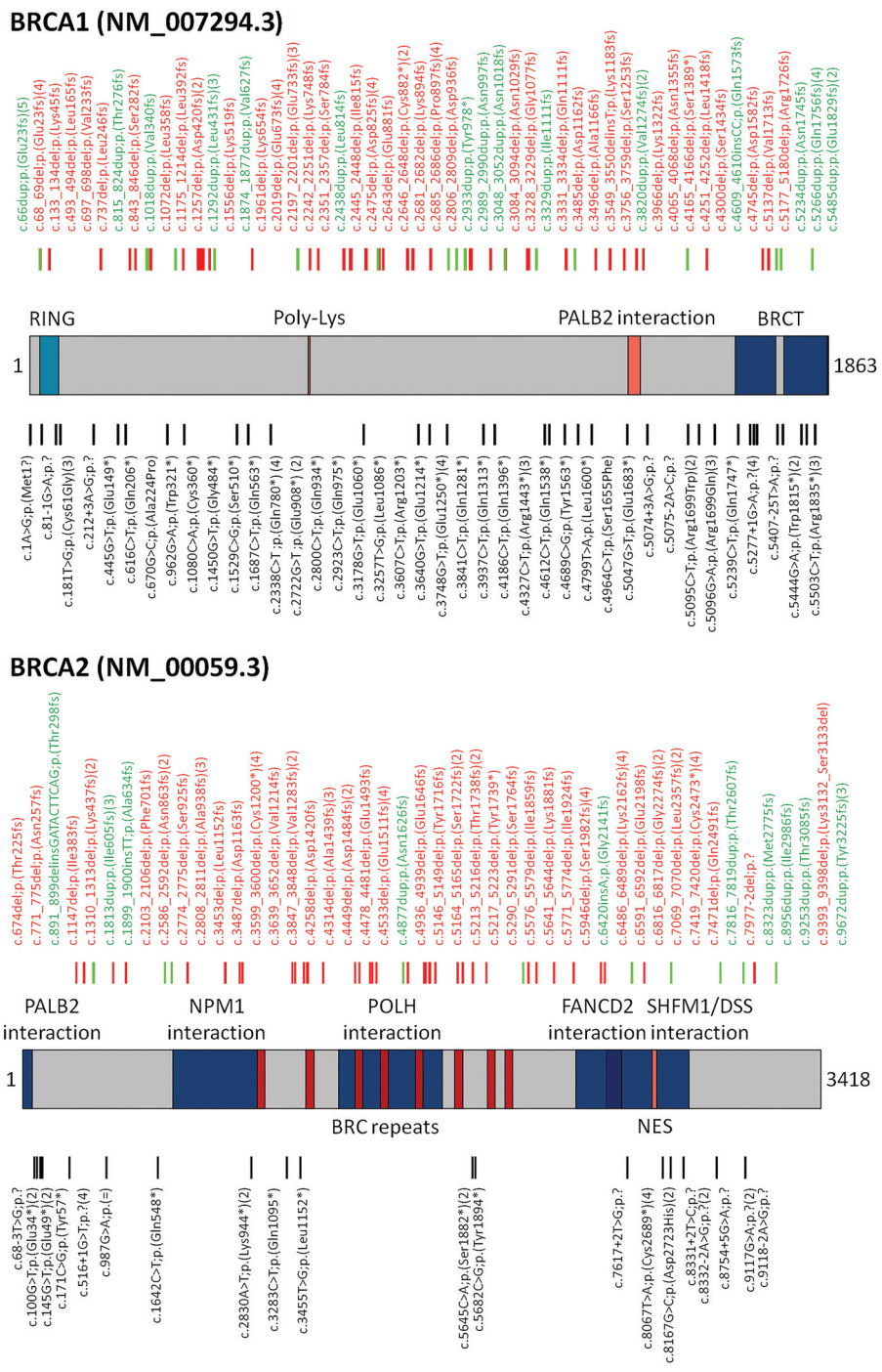

Fig. 2. Schematic overview of the BRCA1 and BRCA2 proteins, with all variants identified.

Vertical bars indicate *BRCA1* and *BRCA2* variants. Red: deletions, green: insertions, black: substitutions, numbers in brackets: number of cases with identical variants. In total 110 unique variants from 166 samples are shown, most of them being pathogenic. Protein domains and regions are based on UniProt (www.uniprot.org). RING, RING (really interesting new gene) domain; PALB2, partner and localizer of *BRCA2*; BRCT, *BRCA1* c-terminus domain; NPM1, nucleophosmin 1; POLH, DNA polymerase eta; FANCD2, Fanconi anemia complementation group D2; SHFM1/DSS, 26S proteasome complex subunit protein DSS (deleted in split hand/split foot), encoded by the gene *SHFM1*; BRC repeats, breast cancer repeats; NES, nuclear export signal.

Table 2. Analytical sensitivity and specificity of manual runs.

Number of samples	152	
Total of sequenced bases	2 688 728	
TP[a]	1 821	Variant calls: pathogenic + benign variants
FP	0	False-positive calls
TN	2 686 907	Bases identical to reference
FN	0	Missed variants
Sensitivity		
TP rate	100%	= TP/(TP + FN)
FP rate	0%	= FP/(FP + TN)
Accuracy	100%	= (TP + TN)/(TP + TN + FP + FN)
Precision	100%	= TP/(TP + FP)
Specificity		
TN rate	100%	= TN/(FP + TN)

[a] TP, true positives; FP, false positives; TN, true negatives; FN, false negatives. TP/(TP + FN) = 1 821/(1 821 + 0) = 1, FP/(FP + TN) = 0/(0 + 2 686 907) = 0, (TP + TN)/(TP + TN + FP + FN) = (1 821 + 2 686 907)/(1 821 + 2 686 907 + 0 + 0) = 1, TP/(TP + FP) = 1 821/(1 821 + 0) = 1, TN/(FP + TN) = 2 686 907/(0 + 2 686 907) = 1.

that all variants identified by semiconductor sequencing or Sanger sequencing were also detected in the smMIP-based data, including 13 pathogenic variants (see online Supplemental Table 9).

INTERLABORATORY VALIDATION

The BRCA smMIP work flow was independently validated using 267 Sanger-sequenced BRCA-samples of Norwegian origin. Six samples (2.2%) did not contain any sequencing data. These samples were excluded from further analysis. For the residual samples, the variant tables were analyzed blinded. A total of 3692 variants, 61 pathogenic and 3631 benign variants, were identified by the smMIP approach, all previously seen by Sanger sequencing, resulting in a 100% analytical sensitivity and 100% analytical specificity. The overall false-positive rate was 0% (see online Supplemental Table 10).

ANALYSIS OF BRCA1 CNVs

We next assessed the ability to detect *BRCA1* CNVs in the smMIP data by including 5 positive control samples, an exon 1–8 deletion, an exon 3–20 deletion, an exon 11–12 deletion, an exon 13 deletion, and an exon 22 deletion, in duplo in 3 different sequencing runs which also included diagnostic samples (see online Supplemental Table 2 and online Supplemental Fig. 4). This resulted in an analytical sensitivity of 100% and an analytical specificity of 88%. Subsequent application of the CNV analysis identified an exon 22 deletion in a patient sample in a blinded fashion (see online Supplemental Fig. 4).

TURNAROUND TIME

TATs for *BRCA1* and *BRCA2* testing in our laboratory in Nijmegen have varied over the years, depending on the respective test performed (denaturing gradient gel electrophoresis in combination with protein truncation test; Sanger sequencing; AmpliSeq/IonTorrent PGM, smMIP/NextSeq500). Looking at monthly intervals over the last 3 years (2013–2015), the mean TAT at our center

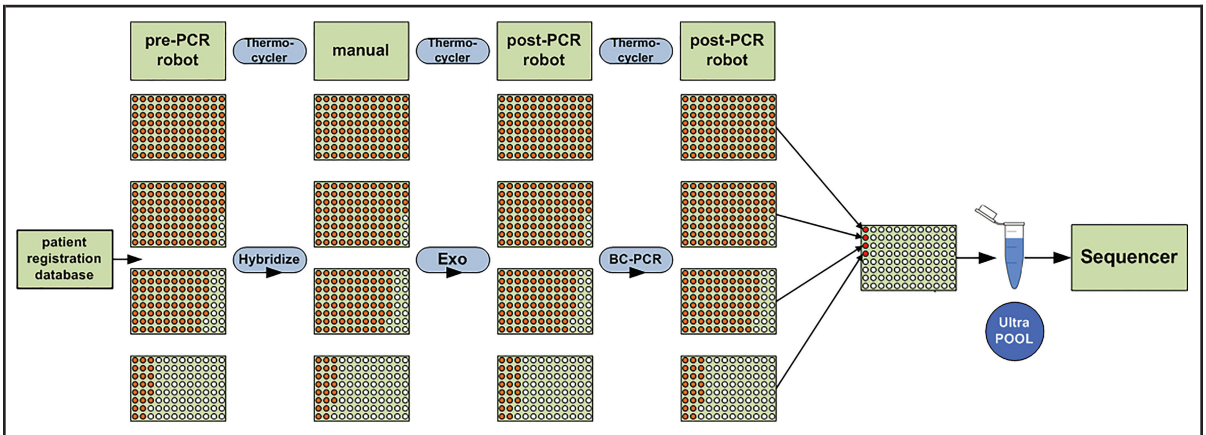

Fig. 3. Automated work flow for smMIP analysis.

Prehybridization reactions were performed on a pre-PCR robot. Hybridization, extension, and ligation occurred overnight. Exonuclease treatment was performed manually the next day. The post-PCR robot pipetted the posthybridization reaction, including a PCR to incorporate barcodes. Following PCR, pooling was performed per plate. Purification occurred automatically per pool. Pools were subsequently combined, diluted, and denaturated for sequencing. Exo, exonuclease treatment; BC-PCR, barcode PCR.

was 18–32 days, measured from acceptance of sample to reporting (see online Supplemental Fig. 5A). smMIP BRCA testing was started in July 2015, resulting in a steep decrease of TATs to a current mean of 11–18 days. Interestingly, numbers of requested BRCA tests rose simultaneously (currently approximately 150 samples/month; see online Supplemental Fig. 5B).

Discussion

Results of clinical genetic tests, such as *BRCA1* and *BRCA2*, increasingly influence clinical management and therapeutic decision-making. Therefore, completeness, accuracy, and reproducibility of clinical sequencing data are becoming more important than ever. Here we describe a newly developed diagnostic sequencing work flow, based on smMIP enrichment technology, that reached 100% analytical sensitivity and specificity for the breast-cancer genes *BRCA1* and *BRCA2*.

BRCA1 and *BRCA2* genetic testing is performed in many laboratories and often is still Sanger-based sequencing *(14, 15, 24)*. Recently, studies have shown that Sanger sequencing can be readily transferred to NGS *(22, 25)*, concluding that NGS is as good or even better *(24, 26)*. Accordingly, new technologies to perform *BRCA1* and *BRCA2* testing have been described, including in-solution capture with complementary RNA probes and sequencing on an Illumina Genome Analyzer *(17)*, long-range PCR in combination with Genome Analyzer *(27)* or GS Junior 454 *(28)* and multiplex PCR strategies for the 454 GS-FLX *(19, 29)* and the IonTorrent PGM *(20)*. Reported analytical sensitivities were close to 100%, whereas analytical specificities ranged from 94.4–97.5% *(30–32)*, still leaving room for improvement. Besides completeness, accuracy, and reproducibility, increasing numbers of *BRCA1* and *BRCA2* tests require a high throughput, and fast and predictable TATs at acceptable costs *(33, 34)*.

In an attempt to satisfy these needs, we implemented a new NGS-based work flow for analyses of single genes and small gene panels. The TAT of this new wet laboratory process is 4 days (excluding DNA isolation and reporting). The resulting data are complete and accurate. The new work flow is based on smMIP technology *(6–9, 21)*. smMIPs have been shown to reach 99% analytical sensitivity and 98% positive predictive value for single nucleotide variants at well-covered positions, i.e., 92%–98% of targeted bases, at relatively low reagent costs *(8)*. Recently, another study reported both analytical sensitivity and specificity >99% for smMIP targeted enrichment *(35)*.

A major advantage of the smMIP technology is that smMIPs, if well designed, do not interfere with each other, enabling smMIPs to be combined to test for complete genes or even several genes in a single pool (in a single well). Solely for this reason, the number of reactions is reduced substantially compared to most conventional approaches, decreasing handling time and costs. Further, the availability of column-synthesized individual smMIPs enables rebalancing of individual probes, giving the user opportunities to optimize uniformity of sequencing depth independent of any supplying company. High reproducibility and even coverage then allows both SNV and CNV calling in a single experiment. smMIPs for newly discovered variants or genes could easily be added to an existing pool. The quantities of each smMIP oligonucleotide obtained are sufficient to support resequencing from millions of human DNA samples. Furthermore, smMIP-based enrichment is platform-independent, and has been described for Illumina *(36, 37)* and IonTorrent sequencing *(38)*. Using the smMIP approach per se is therefore already beneficial for genetic analyses.

The excellent performance and predictability of our work flow is achieved by small but important adjustments to the originally described implementations of MIPs for sequencing capture, namely single-molecule tags, double tiling/capturing, and sequencing, combined with complete automation of both wet laboratory and subsequent data transfer and analyses.

Single-molecule tags are random tags of (in our case) 5 nucleotides in length that are incorporated adjacent to 1 of the hybridization arms *(9, 21)*. The complexity of 1024 ($5N = 5^4$) tagged single molecules enables tracing of individual molecules throughout capture and sequencing. If needed, the 5N tag can be adapted to a larger N tag, e.g., 10N, if larger numbers of unique reads are required *(9)*. Following sequencing, this random tag can be used to distinguish whether reads are coming from the same molecule, or represent unique reads from independent molecules. Reads showing the same tag are combined to a single consensus read, whereby random errors incorporated during PCR or sequencing can be removed *(9, 39)*.

Double-tiling enrichment indicates that each single nucleotide of the gene/panel of interest is targeted by at least 2 independent smMIPs. In our approach, we aimed to cover the complete region of interest with overlapping smMIPs, for both DNA strands independent of each other. Double tiling creates an intrinsic validation, since every variant is detected with at least 2 independent probes. An additional advantage of maintaining strand information throughout the experiments is that this can be used to exclude strand-specific errors or biases, e.g., formalin-fixed paraffin-embedded deamination. Finally, differences in variant percentages for the same nucleotide position hint to potential allelic dropout, meaning that allelic dropouts are likely to be uncovered when using double tiling.

In our work flow, each sample is captured and sequenced twice. Experience and simulation has shown that duration of wet laboratory analysis often depends on the number of failed experiments that need to be repeated, e.g., a single amplicon failure that needs to be repeated to have a complete gene sequenced. Sequencing each sample in duplo reduces the rework and leads to shorter TATs.

The newly implemented work flow has been extensively tested for altogether more than 500 samples from 2 different laboratories. Concerning SNV analysis, the 3 features—single molecule tags, double tiling, and double enrichment/sequencing—enabled an analytical sensitivity and specificity of 100%, a value not reached before by any other technology used in our laboratory. Preliminary tests concerning CNV analysis of 5 well-defined *BRCA1* deletions gave 100% analytical sensitivity, with an analytical specificity of approximately 88%. Since the numbers were low, there is still room for improvement. Due to the small insert size, smMIP-based sequencing also enables the analysis of *BRCA1, BRCA2*, and other genes on DNA isolated from formalin-fixed, paraffin-embedded material *(14, 39)*, using the same automated work flow. Absolute coverage correlates with input amounts of DNA. In another study variants were still reliably detected with an input amount as low as 19 ng of DNA *(39)*. This will not only enable genetic analyses on deceased patients, but also the analysis of tumor-specific variants for therapeutic interventions. An assay as suggested here might enable population-based screening of *BRCA1* and *BRCA2 (10, 40, 41)*.

Altogether, the smMIP-enrichment technology turns out to be highly competitive compared to previous technologies. The combination of smMIP enrichment and subsequent NGS delivers sequencing data that are of outstanding quality. With smMIP enrichment, we are able to achieve 100% analytical sensitivity and specificity, and substantially reduced TATs. Therefore, we believe that smMIPs are not only beneficial for research purposes, but also provide remarkable value for clinics.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: M.J.L. Ligtenberg, AstraZeneca.
Stock Ownership: None declared.
Honoraria: A.R. Mensenkamp and M.J.L. Ligtenberg, AstraZeneca.
Research Funding: None declared.
Expert Testimony: None declared.
Patents: E. Boyle, patent no. 20160055293; J. Shendure, patent no. WO2014160736 A1.
Other Remuneration: M.J.L. Ligtenberg, AstraZeneca.

Role of Sponsor: No sponsor was declared.

Acknowledgments: We are grateful: to the Eichler and Shendure labs (University of Washington), in particular to Evan Eichler and Bradley Coe, for assistance with the initial MIP design and protocol; to the Genome Technology Center, Laboratory of Tumor Genetics, Department of Pathology and Service-Team Genetics (Radboudumc Nijmegen); to Paal Borge for assistance and JSI medical systems for collaboration on smMIP analysis.

References

1. Saunders CJ, Miller NA, Soden SE, Dinwiddie DL, Noll A, Alnadi NA, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Sci Transl Med 2012;4:154ra35.
2. Hennessy BT, Timms KM, Carey MS, Gutin A, Meyer LA, Flake DD 2nd, et al. Somatic mutations in BRCA1 and BRCA2 could expand the number of patients that benefit from poly (ADP ribose) polymerase inhibitors in ovarian cancer. J Clin Oncol 2010;28:3570–6.
3. Maxwell KN, Domchek SM. Cancer treatment according to BRCA1 and BRCA2 mutations. Nat Rev Clin Oncol 2012;9:520–8.
4. Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. Padlock probes: circularizing oligonucleotides for localized DNA detection. Science 1994;265:2085–8.
5. Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, et al. Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat Biotechnol 2003;21:673–8.
6. Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, et al. Multiplex amplification of large sets of human exons. Nat Methods 2007;4:931–6.
7. Turner EH, Lee C, Ng SB, Nickerson DA, Shendure J. Massively parallel exon capture and library-free resequencing across 16 genomes. Nat Methods 2009;6:315–6.
8. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 2012;338:1619–22.
9. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res 2013;23:843–54.
10. King MC, Levy-Lahad E, Lahad A. Population-based screening for BRCA1 and BRCA2: 2014 Lasker Award. JAMA 2014;312:1091–2.
11. Domchek SM, Friebel TM, Singer CF, Evans DG, Lynch HT, Isaacs C, et al. Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. JAMA 2010;304:967–75.
12. Trainer AH, Lewis CR, Tucker K, Meiser B, Friedlander M, Ward RL. The role of BRCA mutation testing in determining breast cancer therapy. Nat Rev Clin Oncol 2010;7:708–17.
13. Wevers MR, Hahn DE, Verhoef S, Bolhaar MD, Ausems MG, Aaronson NK, Bleiker EM. Breast cancer genetic counseling after diagnosis but before treatment: A pilot study on treatment consequences and psychological impact. Patient Educ Couns 2012;89:89–95.
14. Weren RD, Mensenkamp AR, Simons M, Eijkelenboom A, Sie AS, Ouchene H, et al. Novel BRCA1 and BRCA2 tumor test as basis for treatment decisions and referral for genetic counselling of patients with ovarian carcinomas. Hum Mutat [Epub ahead of print 2016 Oct 21].
15. Muendlein A, Rohde BH, Gasser K, Haid A, Rauch S, Kinz E, et al. Evaluation of BRCA1/2 mutational status among German and Austrian women with triple-negative breast cancer. J Canc Res Clin Oncol 2015;141:2005–12.
16. Bu R, Siraj AK, Al-Obaisi KA, Beg S, Al Hazmi M, Ajarim D, et al. Identification of novel BRCA founder mutations in Middle Eastern breast cancer patients using capture and Sanger sequencing analysis. Int J Cancer 2016;139:1091–7.
17. Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proc Natl Acad Sci U S A 2010;107:12629–33.
18. De Leeneer K, Hellemans J, De Schrijver J, Baetens M,

Poppe B, Van Criekinge W, et al. Massive parallel amplicon sequencing of the breast cancer genes BRCA1 and BRCA2: opportunities, challenges, and limitations. Hum Mutat 2011;32:335–44.
19. Michils G, Hollants S, Dehaspe L, Van Houdt J, Bidet Y, Uhrhammer N, et al. Molecular analysis of the breast cancer genes BRCA1 and BRCA2 using amplicon-based massive parallel pyrosequencing. J Mol Diagn 2012; 14:623–30.
20. Costa JL, Sousa S, Justino A, Kay T, Fernandes S, Cirnes L, et al. Nonoptical massive parallel DNA sequencing of BRCA1 and BRCA2 genes in a diagnostic setting. Hum Mutat 2013;34:629–35.
21. Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics 2014;30:2670–2.
22. Diekstra A, Bosgoed E, Rikken A, van Lier B, Kamsteeg EJ, Tychon M, et al. Translating sanger-based routine DNA diagnostics into generic massive parallel ion semiconductor sequencing. Clin Chem 2015;61:154–62.
23. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:13033997 2013.
24. D'Argenio V, Esposito MV, Telese A, Precone V, Starnone F, Nunziato M, et al. The molecular analysis of BRCA1 and BRCA2: next-generation sequencing supersedes conventional approaches. Clin Chim Acta 2015; 446:221–5.
25. De Leeneer K, Hellemans J, Steyaert W, Lefever S, Vereecke I, Debals E, et al. Flexible, scalable, and efficient targeted resequencing on a benchtop sequencer for variant detection in clinical practice. Hum Mutat 2015; 36:379–87.
26. Beck TF, Mullikin JC, Program NCS, Biesecker LG. Systematic evaluation of Sanger validation of next-generation sequencing variants. Clin Chem 2016;62: 647–54.
27. Ozcelik H, Shi X, Chang MC, Tram E, Vlasschaert M, Di Nicola N, et al. Long-range PCR and next-generation sequencing of BRCA1 and BRCA2 in breast cancer. J Mol Diagn 2012;14:467–75.
28. Hernan I, Borras E, de Sousa Dias M, Gamundi MJ, Mane B, Llort G, et al. Detection of genomic variations in BRCA1 and BRCA2 genes by long-range PCR and next-generation sequencing. J Mol Diagn 2012;14: 286–93.
29. De Leeneer K, De Schrijver J, Clement L, Baetens M, Lefever S, De Keulenaer S, et al. Practical tools to implement massive parallel pyrosequencing of PCR products in next generation molecular diagnostics. PLoS One 2011;6:e25531.
30. Feliubadalo L, Lopez-Doriga A, Castellsague E, del Valle J, Menendez M, Tornero 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;21:864–70.
31. Ruiz A, Llort G, Yague C, Baena N, Vinas M, Torra M, et al. Genetic testing in hereditary breast and ovarian cancer using massive parallel sequencing. Biomed Res Int 2014;2014:542541.
32. Ermolenko NA, Boyarskikh UA, Kechin AA, Mazitova AM, Khrapov EA, Petrova VD, et al. Massive parallel sequencing for diagnostic genetic testing of BRCA genes–a single center experience. Asian Pac J Cancer Prev 2015;16:7935–41.
33. Freedman R, Mountain H, Karina D, Schofield L. A retrospective exploration of the impact of the 'Angelina Jolie Effect' on the Single State-Wide Familial Cancer Program in Perth, Western Australia. J Genet Couns [Epub ahead of print 2016 Jun 6].
34. Pilato B, Pinto R, De Summa S, Petriella D, Lacalamita R, Danza K, et al. BRCA1–2 diagnostic workflow from NGS to variant identification and final report. Genes Chromosomes Cancer 2016;55:803–13.
35. Umbarger MA, Kennedy CJ, Saunders P, Breton B, Chennagiri N, Emhoff J, et al. Next-generation carrier screening. Genet Med 2014;16:132–40.
36. Kumar A, Boyle EA, Tokita M, Mikheev AM, Sanger MC, Girard E, et al. Deep sequencing of multiple regions of glial tumors reveals spatial heterogeneity for mutations in clinically relevant genes. Genome Biol 2014; 15:530.
37. Ockeloen CW, Khandelwal KD, Dreesen K, Ludwig KU, Sullivan R, van Rooij IA, et al. Novel mutations in LRP6 highlight the role of WNT signaling in tooth agenesis. Genet Med 2016;18:1158–62.
38. de Voer RM, Hahn MM, Mensenkamp AR, Hoischen A, Gilissen C, Henkes A, et al. Deleterious germline BLM mutations and the risk for early-onset colorectal cancer. Sci Rep 2015;5:14060.
39. Eijkelenboom A, Kamping EJ, Kastner-van Raaij AW, Hendriks-Cornelissen SJ, Neveling K, Kuiper RP, et al. Reliable next-generation sequencing of formalin-fixed, paraffin-embedded tissue using single molecule tags. J Mol Diagn 2016;18:851–63.
40. Gabai-Kapara E, Lahad A, Kaufman B, Friedman E, Segev S, Renbaum P, et al. Population-based screening for breast and ovarian cancer risk due to BRCA1 and BRCA2. Proc Natl Acad Sci U S A 2014;111: 14205–10.
41. Levy-Lahad E, Lahad A, King MC. Precision medicine meets public health: population screening for BRCA1 and BRCA2. J Natl Cancer Inst 2015;107:420.