



Article

# Optimization of a WGA-Free Molecular Tagging-Based NGS Protocol for CTCs Mutational Profiling

Giuseppa De Luca <sup>1</sup>, Barbara Cardinali <sup>2</sup>, Lucia Del Mastro <sup>2,3</sup>, Sonia Lastraioli <sup>1</sup>, Franca Carli <sup>4</sup>, Manlio Ferrarini <sup>5</sup>, George A. Calin <sup>6</sup>, Anna Garuti <sup>3</sup>, Carlotta Mazzitelli <sup>2</sup>, Simona Zupo <sup>1</sup> and Mariella Dono <sup>1,\*</sup>

<sup>1</sup> Molecular Diagnostic Unit, IRCCS Ospedale Policlinico San Martino, 16132 Genova, Italy; giuseppa.deluca@hsanmartino.it (G.D.L.); sonia.lastraioli@hsanmartino.it (S.L.); simonettazu@hotmail.it (S.Z.)

<sup>2</sup> Breast Unit, IRCCS Ospedale Policlinico San Martino, 16132 Genova, Italy; barbara.cardinali@hsanmartino.it (B.C.); lucia.delmastro@hsanmartino.it (L.D.M.); carlotta.mazzitelli@edu.unige.it (C.M.)

<sup>3</sup> Department of Internal Medicine (Di.M.I.), University of Genova, 16132 Genova, Italy; anna.garuti@unige.it

<sup>4</sup> Department of Pathology, IRCCS Ospedale Policlinico San Martino, 16132 Genova, Italy; franca.carli@hsanmartino.it

<sup>5</sup> DIMES, Anatomy Section, Medical School, University of Genova, 16132 Genova, Italy; ferrarini.manlio@gmail.com

<sup>6</sup> Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, Texas, TX 77030, USA; gcalin@mdanderson.org

\* Correspondence: maria.dono@hsanmartino.it; Tel.: +39-010-555-8940

Received: 24 April 2020; Accepted: 17 June 2020; Published: 19 June 2020



**Abstract:** Molecular characterization of Circulating Tumor Cells (CTCs) is still challenging, despite attempts to minimize the drawbacks of Whole Genome Amplification (WGA). In this paper, we propose a Next-Generation Sequencing (NGS) optimized protocol based on molecular tagging technology, in order to detect CTCs mutations while skipping the WGA step. MDA-MB-231 and MCF-7 cell lines, as well as leukocytes, were sorted into pools (2–5 cells) using a DEPArray™ system and were employed to set up the overall NGS procedure. A substantial reduction of reagent volume for the preparation of libraries was performed, in order to fit the limited DNA templates directly derived from cell lysates. Known variants in *TP53*, *KRAS*, and *PIK3CA* genes were detected in almost all the cell line pools (35/37 pools, 94.6%). No additional alterations, other than those which were expected, were found in all tested pools and no mutations were detected in leukocytes. The translational value of the optimized NGS workflow is confirmed by sequencing CTCs pools isolated from eight breast cancer patients and through the successful detection of variants. In conclusion, this study shows that the proposed NGS molecular tagging approach is technically feasible and, compared to traditional NGS approaches, has the advantage of filtering out the artifacts generated during library amplification, allowing for the reliable detection of mutations and, thus, making it highly promising for clinical use.

**Keywords:** Circulating Tumor Cells; breast cancer; liquid biopsy; Next-Generation Sequencing; molecular tagging; Whole Genome Amplification-free; single-cell genomics; biomarkers; precision medicine

## 1. Introduction

Circulating Tumor Cells (CTCs) in the bloodstream, which originate from primary and/or metastatic tumors, are both phenotypically and molecularly highly heterogeneous [1,2].

An increasing number of data have confirmed the role of CTCs count and their molecular characterization in patient prognosis, relapse prediction, and therapy outcome, indicating that CTCs represent an attractive blood-based biomarker with high potential for understanding the clinical behavior of the associated disease [3–10].

To date, major efforts towards the molecular characterization of CTCs have been focused on the optimization of genetic analysis protocols, mainly including Whole Genome Amplification (WGA) as a first step, which allows for the isolation of an acceptable amount of DNA from few starting molecules [11–16]. Unfortunately, different studies have demonstrated a series of drawbacks arising from WGA procedures, such as allelic dropout, chimeric molecules, low uniformity, and polymerase errors, which lead to bias in the interpretation of molecular data [17–21]. These biases are partly caused by an unavoidable delay in the mutational profile analysis of CTCs in routine laboratory practice.

Meanwhile, different workflows have been developed to provide alternative options to the traditional techniques which do not require this preamplification step, instead of utilizing downstream whole-genome or targeted sequencing [22–24].

This work proposes a modified workflow for a targeted Next-Generation Sequencing (NGS) panel, which already was tested in our laboratory for the molecular characterization of Circulating Tumor DNA (ctDNA) and which is based on molecular tagging technology. This approach is capable of intercepting even single DNA molecules in difficult biological samples, such as ctDNA. The original protocol was adapted for NGS analysis of few pooled cells isolated on a DEPArray™ system, with the overall aim of completely skipping WGA; instead, directly proceeding with library construction while obtaining reliable sequencing results.

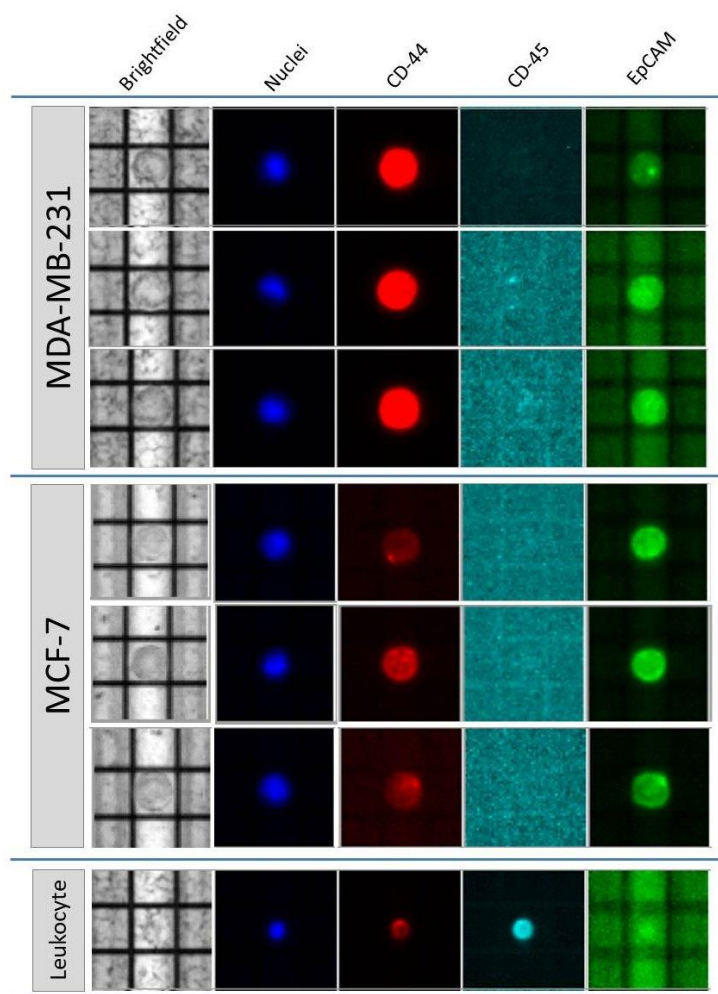
## 2. Results

### 2.1. DEPArray™ Sorting and Validation of Optimized NGS Procedures

#### 2.1.1. Cell Detection and DNA Integrity Assessment

MDA-MB-231 (MDA) and MCF-7 cell lines were employed as surrogate Breast Cancer (BC) CTCs and used in spiking-in experiments by addition to a healthy donor blood sample before DEPArray™ isolation.

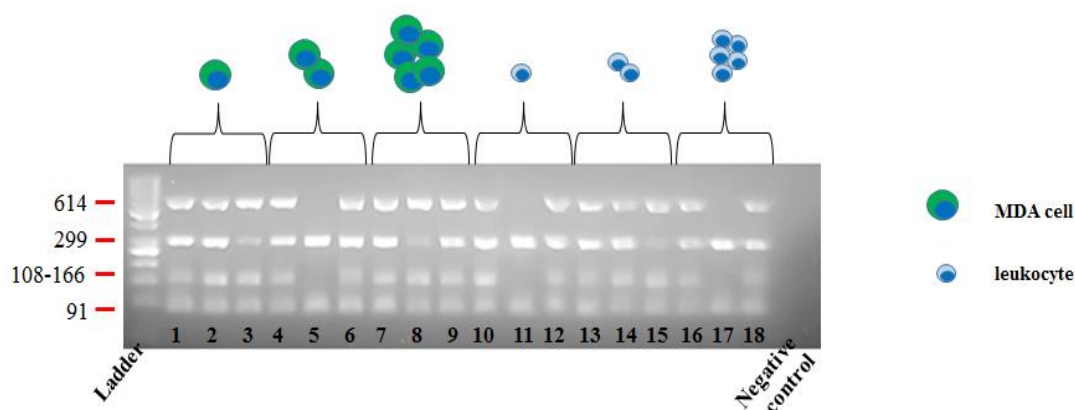
Sorting was realized on the basis of morphological and phenotypic features of MDA/MCF-7 cells: they presented an oval shape and round features and high nuclear–cytoplasmic ratio. Furthermore, they variably expressed Epithelial Cell Adhesion Molecule (EpCAM) and Cluster of Differentiation 44 (CD44) markers, the expression of the latter being almost complete (95–100%) in MDA, but almost negative in MCF-7 (~1–2%); whereas EpCAM was, on average, expressed by 45% of MDA and on the totality of MCF-7 cultured cells, respectively. Neither of the two cell lines showed a Cluster of Differentiation 45 (CD45) expression (by Fluorescent-Activated Cell Sorter, FACS, data not shown). Leukocytes were mostly identified and isolated on the basis of their small size and high CD45 expression. Examples of DEPArray™ sorted and identified cells are reported in Figure 1.



**Figure 1.** Analysis of MDA/MCF-7 and leukocyte cells on DEPArray™. The panel shows the morphological and fluorescent characteristics of MDA-MB-231 (MDA) or MCF7 cells spiked into healthy donor blood and isolated on DEPArray™. MDA cell is typically CD45-negative and Hoechst-positive (light blue and blue stains). In the first lane, a low EpCAM-positive MDA cell can be observed. MCF7 cells are all EpCAM-positive and show different expressions of CD44, almost negative (lines 4 and 6), and occasionally positive (line 5). In comparison, below, a CD44-positive/CD45-positive and Hoechst-positive leukocyte is shown.

To check for DNA integrity, MDA and leukocytes were subsequently isolated on DEPArray™ and a total of 18 samples deriving from two independent experiments were analyzed (i.e., nine MDA cells and nine leukocytes).

Then, DNA lysates from cells pools composed of one, two, and five cells (each run tested in triplicate) were checked using Polymerase Chain Reaction (PCR) by amplification of four chromosomal regions of different lengths (91, 108–166, 299, and 614 bp, respectively of 12 p, 5 p, 17 p, and 6 p chromosomes), where PCR products were visualized by electrophoresis. Optimal DNA integrity was demonstrated by the presence of four bands, respectively, in 8/9 MDA (88.9%) and 7/9 leukocytes (77.8%) samples (Figure 2).



**Figure 2.** DNA integrity testing in sorted MDA/leukocytes cells. Multiplex PCR was performed and four PCR products of different lengths from four chromosomal regions were amplified. The number of bands obtained by each PCR product was visualized on a 3% agarose gel through ultraviolet light. Each lane represents a single/cellular pool.

However, as downstream NGS analysis is ensured only for cells with at least three genomic amplicons (as per Ampli1 QC kit manufacturer instructions and [25]), our preparation and cell isolation procedures demonstrated outperformance, as 83% (15/18) of tested samples were considered potentially suitable for successive NGS experiments.

The remaining three samples—composed of one cell (one sample) or two cells (two samples) and derived from MDA cells in one case and from leukocytes in the other two—displayed only two amplified bands, thus showing that loss of isolated cell/s during experimental manipulation did not occur, although some genetic loss/fragmentation could be detected. Consequently, these data confirmed the robustness and reliability of the cell recovery step, as well as that of the whole sorting procedure.

### 2.1.2. From Canonical to Optimized Molecular Tagging NGS Workflow: Experimental Setting and Development

In the preliminary NGS assessment, MDA lysates of one, three, and five cells were added to 5 ng of normal DNA. Then, these DNA mixtures were employed to prepare libraries using the Oncomine Breast cfDNA Research Assay v2 (OBcfRAv2) NGS targeted panel. Two hotspots present in the MDA cell line—the *KRAS* p.Gly13Asp and *TP53* p.Arg280Lys—were followed and detected.

In particular, the library prepared starting from three MDA cells showed 0.61% and 1.06% molecular allele frequencies (MAFs) for *KRAS* p.Gly13Asp and *TP53* p.Arg280Lys, respectively; likewise, libraries prepared from five MDA cells displayed 0.67% and 0.95% MAFs for the two variants, respectively. Interestingly, variants detection was still feasible from one MDA cell lysate, despite the fact that the addition of 5 ng of normal genomic DNA resulted in 1000× dilution of the ~6 pg mutated alleles, where MAFs obtained for *KRAS* p.Gly13Asp and *TP53* p.Arg280Lys were 0.18% and 0.38%, respectively (data not shown).

Collectively, these preliminary experiments were encouraging, as they demonstrated that tagging NGS is a robust and highly sensitive approach for mutation detection at the single-cell level, being able to discriminate one mutated allele within a bulk of wild-type alleles.

Consequently, we optimized the sequencing protocol by directly using isolated MDA cells as a unique starting genetic material, in the complete absence of normal genomic DNA or WGA, for subsequent library preparation.

As this strategy can dramatically decrease the input amount of DNA, an adjustment of the NGS protocol was developed. The entire new procedure presented main technical changes consisting of the following four items: (1) DNA derived only from sorted cells was used as a starting material for the preparation of libraries; (2) a three-times volume reduction of the reaction reagents was applied for the

two PCRs planned by the OBcfRAv2 protocol, as reported in detail in Table S1; (3) libraries were mixed up to 24 samples for Ion 520™ Chip loading (low-coverage sequencing, Table 1).

The modified protocol was tested not only on MDA cell lines, but also on pools of MCF-7 cells sorted and collected as already described, in order to verify the protocol feasibility in a second cell line harboring different mutations.

**Table 1.** Comparison between canonical and optimized molecular tagging Next-Generation Sequencing (NGS) workflows.

|  | Canonical Molecular Tagging NGS Workflow |   |   | Optimized Molecular Tagging NGS Workflow                |   |   |  |     |     |     |                         |   |  |
|--|--|---|---|---|---|---|--|-----|-----|-----|-------------------------|---|--|
|  | MDA cells (single or pools)              |   |   | MDA <sup>a</sup> /MCF-7 <sup>b</sup> cells (only pools) |   |   | MDA/MCF-7: leukocytes (combined pools) |     |     |     | Leukocytes (only pools) |   |  |
| Types of starting isolated cells             | 1  | 3 | 5 | 2   | 4 | 5 | 4:1                                    | 3:2 | 2:3 | 1:4 | 2                       | 5 |  |
| n of cells                                   | 1  | 3 | 5 | 2   | 4 | 5 | 4:1                                    | 3:2 | 2:3 | 1:4 | 2                       | 5 |  |
| 5ng of DNA from healthy donor                | +  | + | + | -   | - | - | -                                      | -   | -   | -   | -                       | - |  |
| leukocytes OBcfRAv2 volume reagents reaction | As recommended by manufacturer           |   |   | Modified: 3× reduction of volume reagents               |   |   |  |     |     |     |                         |   |  |
| Thermal PCR conditions                       | As recommended by manufacturer           |   |   | As recommended by manufacturer                          |   |   |  |     |     |     |                         |   |  |
| Libraries quantification                     | qPCR                                     |   |   | qPCR  |   |   |  |     |     |     |                         |   |  |
| Libraries multiplexing on chip               | Five libraries on Ion 530™ Chip          |   |   | 24 libraries on Ion 520™ Chip (low-coverage sequencing) |   |   |  |     |     |     |                         |   |  |

<sup>a</sup> MDA sorting experiments were all performed in triplicate. <sup>b</sup> MCF-7 sorting experiments were performed in duplicate, except for pools comprising two tumor cells and 1:4 MCF-7/leukocytes combined pools, which were performed in triplicate. + addition of 5 ng of normal genomic DNA. – no addition of normal genomic DNA.

By applying the optimized NGS protocol, a total of 47 libraries were prepared from nine different cellular pool types. In particular, three pools composed of two, four, and five tumor cells only and four combined pools (4:1, 3:2, 2:3, and 1:4 MDA/leukocytes ratios, respectively) were investigated for MDA, together with two pools of leukocytes only as negative controls. For MCF-7 analysis, duplicates were considered for all pools, except for pools consisting of two cells only and the 1:4 MCF-7/leukocytes combination, respectively, which were tested in triplicate. In this context, the cellular combinations at different ratios were intended to test the detection accuracy of mutated alleles in the presence of wild-type ones.

First, sufficient amounts for subsequent massive sequencing (range 11–246 pMol, median 65 pMol) were obtained in 45/47 samples (95.7%) and only two libraries failed to be quantified (from MDA cells/leukocytes combinations at 3:2 and 2:3 ratios), probably indicating a complete loss of genetic material in these samples. Subsequently, we decided to analyze two new MDA cellular preparations to complete the triplicates and, thus, a total of 49 libraries were prepared, with a final performance of 47/49 samples (95.9%).

Second, successful sequencing was realized in 46/47 libraries (97.9%), with a 12,272–294,265 range of total reads (median value 83,723). *KRAS* p.Gly13Asp and *TP53* p.Arg280Lys variants were successfully found in the MDA triplicates, as well as *PIK3CA* Glu545Lys in MCF-7 (except for one combination), regardless of the number of tumor cells present in a pool (Table 2).

**Table 2.** Variant calling results from MDA, MCF-7, and leukocytes libraries.

| Pool Type    | <i>n</i> of Cells | MDA-MB-231        |                        |                                  | MCF-7             |                        |                                  |                   |                        |                                  |
|--------------|-------------------|-------------------|------------------------|----------------------------------|-------------------|------------------------|----------------------------------|-------------------|------------------------|----------------------------------|
|              |                   | KRAS Gly13Asp     |                        |                                  | TP53 Arg280Lys    |                        |                                  | PIK3CA Glu545Lys  |                        |                                  |
|              |                   | Expected MAFs (%) | Mean Observed MAFs (%) | Median Allele Molecular Coverage | Expected MAFs (%) | Mean Observed MAFs (%) | Median Allele Molecular Coverage | Expected MAFs (%) | Mean Observed MAFs (%) | Median Allele Molecular Coverage |
| tumor cells  | 2                 | 100               | 100                    | 7                                | 100               | 100                    | 10                               | 50                | 38                     | 2                                |
|              | 4                 | 100               | 100                    | 8                                | 100               | 100                    | 12                               | 50                | 66.7                   | 2.5                              |
|              | 5                 | 100               | 100                    | 10                               | 100               | 95.8                   | 14                               | 50                | 43.7                   | 6.5                              |
| tumor cells: | 4:1               | 80                | 76.2                   | 5                                | 80                | 87.5                   | 7                                | 40                | 33.3                   | 6                                |
|              | 3:2               | 60                | 83.4 <sup>a</sup>      | 2 <sup>a</sup>                   | 60                | 57.5 <sup>a</sup>      | 6 <sup>a</sup>                   | 30                | 50                     | 3                                |
| leukocytes   | 2:3               | 40                | 37.2                   | 5                                | 40                | 57.6                   | 5                                | 20                | 36.4 <sup>b</sup>      | 4 <sup>b</sup>                   |
|              | 1:4               | 20                | 24.4                   | 2                                | 20                | 45.1                   | 3                                | 10                | 14.8                   | 1                                |
| Leukocytes   | 2                 | 0                 | 0                      | 0                                | 0                 | 0                      | 0                                | 0                 | 0                      | 0                                |
|              | 5                 | 0                 | 0                      | 0                                | 0                 | 0                      | 0                                | 0                 | 0                      | 0                                |

<sup>a</sup> Calculated on 2 replicates, due to a library sequencing failure. <sup>b</sup> calculated on a single replicate.

In detail, considering sequencing data from the pools composed of only MDA/MCF-7 cells, a high concordance was achieved for the MAFs of all hotspots, compared to the expected ones, especially when libraries derived from two cells were studied (Table 2).

When libraries from combined pools of MDA/MCF-7/leukocytes were analyzed, *KRAS*, *TP53*, and *PIK3CA* variants were still detectable in all the replicates (Table 2). These results indicate that, although mutated alleles deriving from MDA/MCF-7 cells were mixed with the wild-type alleles from leukocytes, a high level of sensitivity was maintained, as shown in the 1:4 MDA/MCF-7/leukocytes ratios.

Robustness and sensitivity of the optimized NGS protocol were sustained by the detection of the Glu545Lys variant present in heterozygous mutational status (data not shown), as well as in the combination pools, where the mutated allele was heavily diluted within wild-type ones.

However, at least in some combinations, a loss of library performance may be noticed, which created some unbalance between the expected and observed frequencies (Table 2)—as particularly evident in the 3:2 and 2:3 MDA/leukocytes ratios—and a loss of mutated alleles was evidenced in one of the duplicate pool 2:3 MCF-7 cell line (Table 2). A possible explanation for these findings may be related to the suboptimal pooling of tumor cells and leukocytes into the same tube, or may be due to an alteration of allele frequencies caused by mixture with normal cells.

A sequencing failure occurred only in one out of the 47 libraries prepared (2.1%), confirming again the strength of the proposed modified NGS protocol. Furthermore, the protocol displayed 100% specificity, as no other mutations were found except for those expected and all leukocytes did not show any false positive calls.

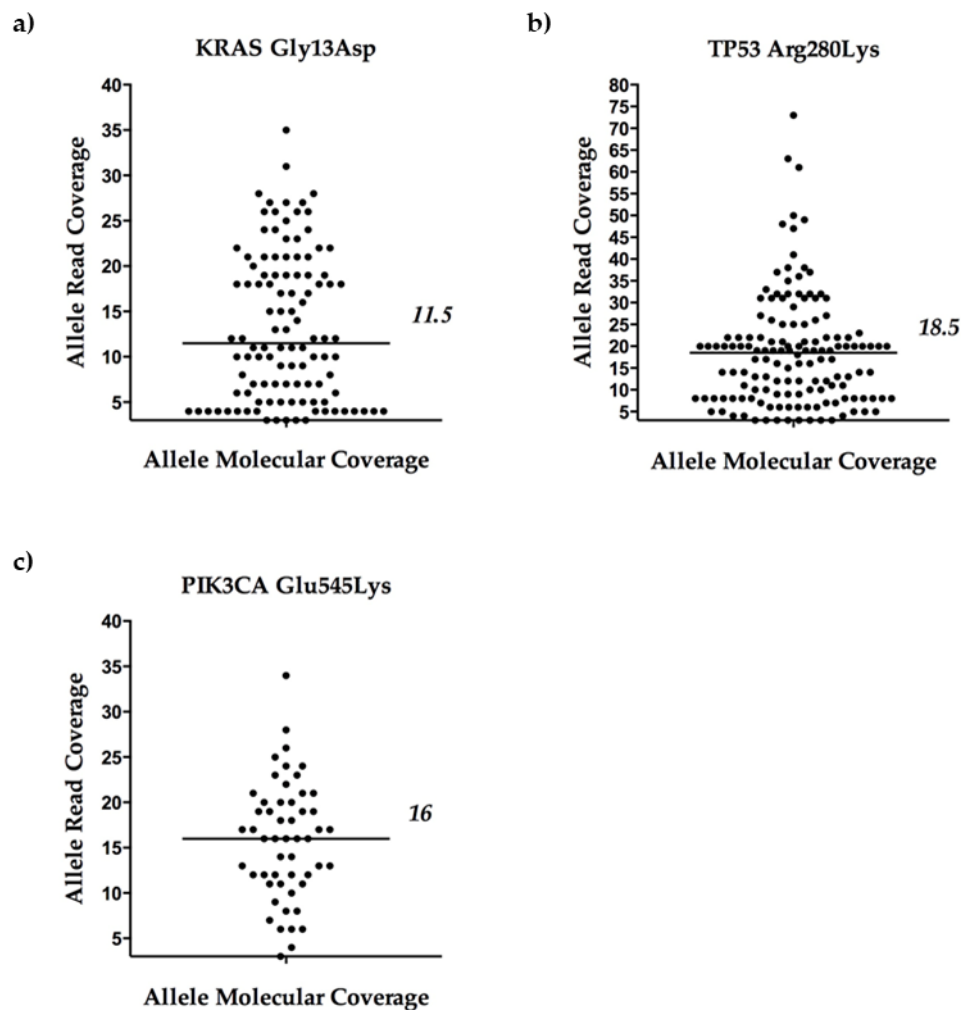
In conclusion, these data demonstrate that the changes introduced in the original OBcRAv2 protocol did not affect its performance.

### 2.1.3. A glance at Variant Calling and Coverage Metrics

In order to better define the coverage metrics of *KRAS* p.Gly13Asp, *TP53* p.Arg280Lys, and *PIK3CA* Glu545Lys variants, an in-depth analysis of the Variant Caller Format (VCF) files was performed. In particular, attention was focused on the Variant Family Size Histogram (VFSH) paired values, a parameter indicating the read count (Allele Read Coverage), and on the number of DNA template molecules (i.e., the tagged original targets), where the count was observed (Allele Molecular Coverage). Overall, this parameter indicates how many reads were attributed to each original mutated DNA molecule and specifies the robustness and reliability of the variant calling. It is appropriate to specify that the VFSH only describes the mutated DNA molecules which passed the quality check of the Torrent Variant Caller (TVC, version 5.10) plugin (i.e., those having at least three reads).

When we analyzed the VFSH values for the mutated libraries (36/47), we found a median value of 11.5 reads obtained for each DNA molecule containing *KRAS* p.Gly13Asp, 18.5 reads for those with *TP53* p.Arg280Lys variants, and 16 reads for *PIK3CA* Glu545Lys (Figure 3). The lower and upper

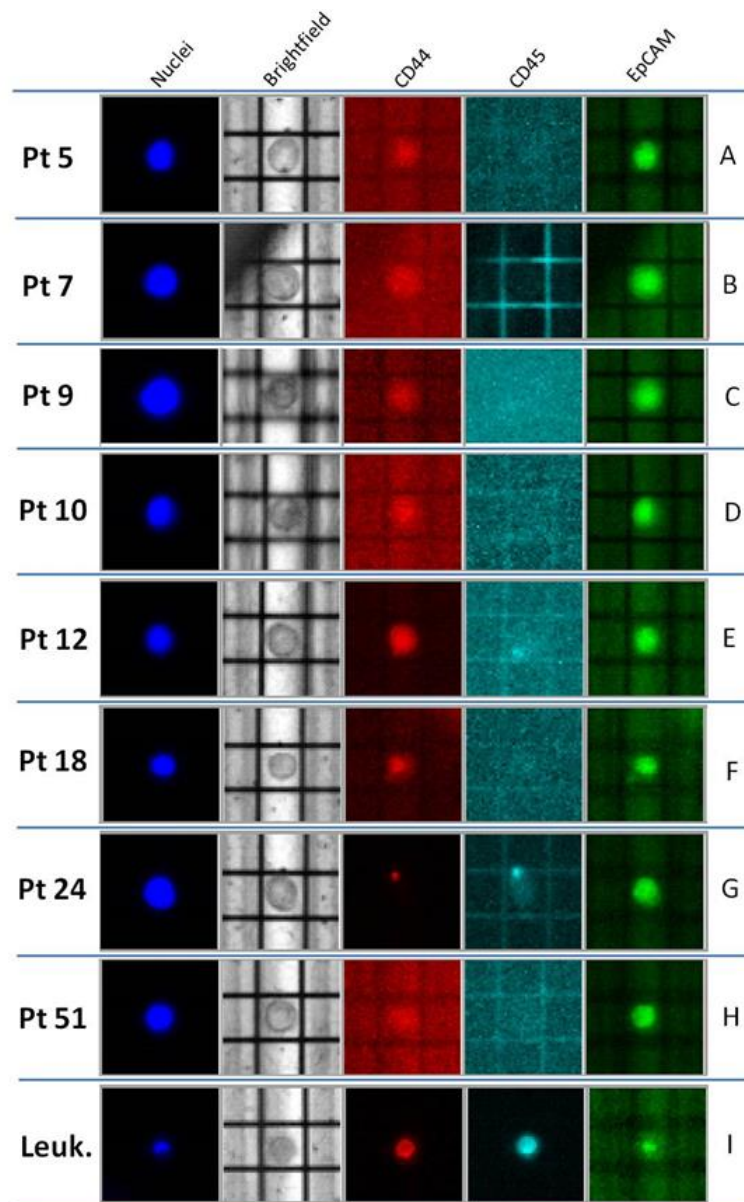
quartiles of the median value of reads were five and 19.25 for *KRAS*, eight and 22.25 for *TP53*, and 11.5 and 20 for *PIK3CA* variant alleles.



**Figure 3.** Coverage metrics of *KRAS* Gly13Asp, *TP53* Arg280Lys, and *PIK3CA* Glu545Lys variants. Scatter dot plots in (a,b) show the Allele Read Coverage and the Allele Molecular Coverage for *KRAS* Gly13Asp and *TP53* Arg280Lys, respectively, as found in MDA pools, and in (c) for *PIK3CA* Glu545Lys in MCF-7. Each dot represents a single molecular DNA molecule on the *x*-axis (i.e., Allele Molecular Coverage) and the corresponding number of reads on the *y*-axis (i.e., Allele Read Coverage). The origin of the axes was set to 3, corresponding to the default minimum value of reads to enable a variant calling.

## 2.2. Optimized Molecular Tagging NGS Workflow Revealed to be Reliable for Molecular Analysis of CTCs

Subsequently, we tested the modified protocol in a BC clinical setting. To this purpose, blood samples of eight BC patients were collected and CTCs were isolated using the DEPArray™ system. CTCs were identified as medium-sized cells (diameter > 10 μm) with a high nucleus–cytoplasm ratio (n/c > 0.5), CD45-expression-negative, and nuclei-staining-positive; fluctuation of CD44 marker expression was observed, as well as some variability in EpCAM expression (Figure 4).



**Figure 4.** Representative Circulating Tumor Cells (CTCs) recovered from breast cancer (BC) patients on DEPArray™. CTCs (A–H) are identified as round, middle-sized cells, which are Hoechst and EpCAM-positive and CD45 negative. A low CD44 expression is appreciated only in Pt. 12. In (G), the CD45 fluorescent dot may represent a non-nuclear component of blood (i.e., erythrocyte or platelet). Leukocyte in (I) is CD44- and CD45-positive.

As the analysis of possible CTCs subpopulations (i.e., with epithelial and/or mesenchymal characteristics) was beyond the scope of this study, CTCs recovery was mainly realized on the basis of two parameters: morphology (shape and round features) and the complete absence of CD45 expression. Thirteen pools of different numbers of CTCs (range: 2–6 cells) were recovered and used for subsequent NGS experiments. Libraries were successfully prepared and sequenced from all the 13 CTCs pools (Table 3), thus showing that even if this analysis was limited in number, the optimized molecular tagging NGS workflow was robust when applied to “real” CTCs. Quantification of libraries revealed a yield range of 11–107 pMol, with an optimal performance (in terms of total produced reads) ranging from 20,633 to 294,305.



**Table 3.** Mutational comparison between CTCs and corresponding tumor breast tissue.

| Pt | CTCs                       |                                | Concordance<br>CTC/Tissue | BC Tissue Mutations                                     |   |
|----|----------------------------|--------------------------------|---------------------------|---|---|
|    | (Number<br>Cells/Reaction) | Mutations<br>(OBcfRAv2)        |                           | Detectable by<br>OBcfRAv2 and<br>Custom Panel           | Detectable Only by<br>Custom Panel  |
| 5  | 3                          | <i>TP53</i> c.1100 + 30A > T   | yes                       | <i>TP53</i> c.1100 + 30A > T                            | <i>ERBB2</i> p.Ile654Val<br><i>ERBB2</i> p.Ile655Val<br><i>ERBB2</i> p.Pro1170Ala                               |
| 7  | 3                          | <i>TP53</i> p.Arg213=          | yes                       | <i>TP53</i> p.Arg213=                                   | <i>PIK3CA</i> p.Asn345Lys<br><i>ERBB2</i> p.Ile655Val<br><i>PIK3R1</i> p.Met326Ile<br><i>ERBB2</i> p.Pro1170Ala |
| 9  | 3                          | not found                      | yes                       | not found   | <i>ERBB2</i> p.Pro1170Ala   |
| 10 | 2                          | not found                      | no                        | <i>TP53</i> p.Arg248Gln                                 |   |
|    | 4                          | not found                      |                           |   |   |
| 12 | 5                          | <i>TP53</i> c.1100 + 30A > T   | yes                       | <i>TP53</i> c.1100 + 30A > T                            | <i>ERBB2</i> p.Pro1170Ala   |
| 18 | 3                          | not found                      | no                        | <i>TP53</i> p.Cys275Leufs                               | <i>ERBB2</i> p.Pro1170Ala   |
|    | 3                          | not found                      |                           |   |   |
| 24 | 2                          | <i>TP53</i> c.1100 + 30A > T   | partial                   | <i>TP53</i> p.Arg248Trp<br><i>TP53</i> c.1100 + 30A > T | <i>ERBB2</i> p.Pro1170Ala   |
|    | 5                          | <i>TP53</i> c.1100 + 30A > T   |                           |   |   |
| 51 | 3                          | <i>PIK3CA</i> p.<br>Hys1047Arg | yes                       | <i>PIK3CA</i> p.<br>Hys1047Arg                          |   |
|    | 6                          | <i>PIK3CA</i> p.<br>Hys1047Arg |                           |   |   |

When available, two and/or three different pools of CTCs from the same patient were analyzed. In particular, one sample was tested for Patients 5, 7, 9, and 12; two samples for Patients 18 and 51; and, finally, three independent pools were tested for Patient 10 (Table 3). Analyses performed in more than one replicate showed concordant results.

Table 3 shows the gene variants found in six out of the eight patients studied. In particular, an intronic variant in the *TP53* gene (c.1100 + 30A > T, chr17:7573897) was present in the CTCs of Patients 5, 12, and 24; while patient seven displayed a synonymous variant in the *TP53* gene (p.Arg213=). Interestingly, both the cellular pools containing three and six CTCs from Patient 51 were characterized by a *PIK3CA* variant (p. Hys1047Arg).

We further investigated whether CTCs pools and matched BC tumor tissue shared the same mutational profile (Table 3). To this end, BC tissues were analyzed in targeted NGS by using a custom panel comprised of 20 genes and, among others, all the DNA target regions covered by OBcfRAv2. When the sequencing analysis was restricted to these common regions, six out of eight cases shared mutations between CTCs and matched BC tissues. In detail, 5/8 cases (62.5%; i.e., Patients 5, 7, 9, 12, and 51) displayed a complete mutational concordance between CTCs and tissues; indeed, a partial concordance was achieved in one patient (Patient 24), as her breast tissue showed an additional *TP53* variant (p.Arg248Trp) which was not found in the two CTC pools investigated.

The remaining two BC patients (Patients 10 and 18) were completely discordant, as the results for both showed mutation in the matched tissues but not in the CTCs pools analyzed. In both patients, tissue variants occurred in the *TP53* gene, even if in different hotspots—p.Arg248Gln and p.Cys275Leufs, respectively (Table 3). These results could be explained by the intrinsic nature of CTCs, which may reflect completely the mutational profile of the bulk tumor, but could also diverge.

Finally, the overall analysis of NGS data from BC tissues revealed additional mutations, mainly in *ERBB2* and *PIK3CA* and occurring in 6/8 BC cases, which were not otherwise demonstrable in the matched CTCs as the genomic mutated positions did not overlap with those covered by the OBcfRAv2 panel.

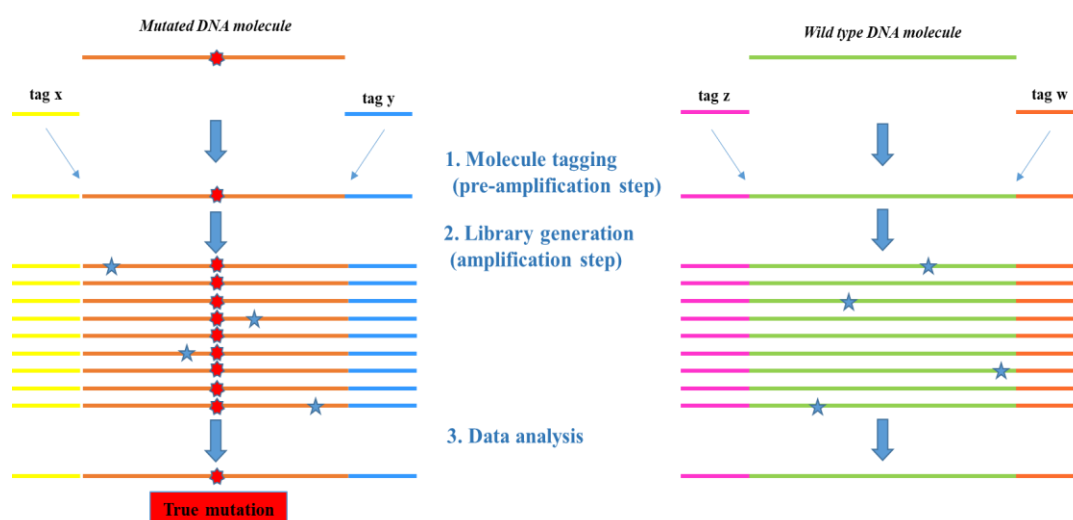
### 3. Discussion

The possible clinical utility of the CTCs obtained by liquid biopsy in routine laboratory practice is still under debate, mostly due to the lack of robust and standardized procedures for their detection and relative molecular characterization. However, especially in BC settings, CTCs have gained a valuable clinical role, as their presence into the bloodstream represents a negative prognostic factor in terms of

malignancy and, moreover, the study of their mutational profile provides precious information about actionable and/or resistance genetic alterations [26–28].

To present, many of the published protocols about CTCs genomic analysis include a preamplification WGA step, prior to “omic” techniques, with the aim of increasing the few DNA starting copies isolated from a single-cell or pools of few cells [11–16]. However, WGA procedures introduce important biases (i.e., allelic dropout, chimeric molecules, low uniformity, and polymerase errors), thus making the interpretation of downstream data both difficult and ambiguous [17–21]. The current major efforts have been focused on the optimization of different commercial WGA protocols coupled with NGS procedures [13,14,29,30].

Recently, molecular tagging NGS technology has found its main application in the liquid biopsy field, particularly for the research and detection of exiguous ctDNA molecules [31,32]. This is possible thanks to the use of unique molecular barcodes which allow us to intercept and “tag” rare DNA alleles, thus providing an effective solution to track and distinguish sequencing reads from the original DNA molecule (Figure 5).



**Figure 5.** Overall principle of sequencing based on molecular tagging NGS approach. Unique molecular tags bound to gene-specific primers are added into every DNA molecule (Step 1) prior to the library amplification step, in order to keep track of possible PCR and/or sequencing artifacts (blue stars). All of the reads (Step 2) are then collapsed with the same genomic coordinates and molecular tag into a single representative read (Step 3), corresponding to an original mutated (on the left) or wild-type (on the right) DNA template.

In this way, the artifacts generated during library amplification are filtered during variant calling and true mutations are correctly detected. This provides a great advantage over traditional amplicon-based NGS approaches, as the latter is more error-prone, especially when low-frequency mutations are studied and more PCR cycles are needed, further increasing the difficulty of discriminating false positives.

With these premises, our study proposes a new NGS workflow based on molecular tagging technology to study the molecular profile of CTCs, with the main goal of avoiding WGA procedures and their relative drawbacks.

Firstly, in the optimization phase of our study, an MDA cell line spiked into healthy donor blood was used as a CTCs surrogate, in order to simulate real-life conditions, and sorted using a DEPArray™ system. To perform the downstream molecular analysis, we used the OBcfRAv2, a molecular tagging-based NGS panel which was originally developed to search for mutations in ctDNA of BC patients. We initially tested the robustness of this assay when applied to single or pooled

MDA cells by adding 5 ng of normal genomic DNA, as this step would intentionally raise the amount of starting nucleic acids to a sufficient quantity for preparation of the NGS libraries.

We obtained optimal sequencing results in each of the MDA pools analyzed, thus demonstrating that high specificity and sensitivity were maintained, even when the maximum dilution of mutated DNA molecules was reached (one cell).

In our opinion, this procedure works well in the case where the mutations followed are known; however, we expect that, in absence of mutations, the distinction between wild-type DNA molecules derived from normal genomic DNA and those from tumor cells is not allowed, thus precluding the translational value of the proposed protocol.

Consequently, a new procedure was developed in order to perform WGA-free library construction, which does not include the addition of any external DNA (other than that derived from cell lysates). The procedure was tested in spike-in experiments using both MDA and MCF-7 cell lines, harboring the different mutations included in the chosen NGS panel.

In this case, massive sequencing data demonstrated that, although the decrease in the starting DNA amount was important (i.e., nanograms versus picograms), the technology was still able to detect *KRAS* p.Gly13Asp, *TP53* p.Arg280Lys, and *PIK3CA* Glu545Lys variants in all the libraries obtained from cellular preparations of two, four, and five MDA/MCF-7 cells (Table 2).

Moreover, the chance to confirm the strength of the optimized NGS workflow was offered by sequencing analysis of MDA/MCF-7/leukocytes combined pools. In fact, although mutated DNA alleles (derived from MDA/MCF-7 cells) were further diluted by wild-type DNA alleles (from leukocytes), the presence of *KRAS* p.Gly13Asp, *TP53* p.Arg280Lys, and *PIK3CA* Glu545Lys variants were still detected in all cellular combinations (even in the 1:4 MDA/MCF-7/leukocytes pools).

We have previously confirmed the presence of these mutations in related cell lines by Sanger sequencing (data not shown) and found that both *KRAS* p.Gly13Asp and *TP53* p.Arg280Lys were homozygous in MDA, whereas *PIK3CA* Glu545Lys was heterozygous in MCF-7. This represented an additional challenge to overcome for the whole procedure, including sorting, library preparation, and a bioinformatic variant calling process. In fact, in such a situation (where only one of the two alleles in a tumor cell contained the mutation), the successful detection in even the 1:4 MCF-7/leukocytes triplicate further demonstrated the robustness of the proposed protocol.

However, some loss of library performance was noticed for the 2:3 and 3:2 combinations, as well as the underestimation of mutated alleles with respect to wild-type ones. This may be related to suboptimal pooling of MDA/MCF-7 cells and leukocytes in the same pool, or the (intentional) contamination from normal cells may have caused unpredictable MAFs.

Other evidence supported the robustness of the optimized workflow. First, DEPAArray<sup>TM</sup> cell recovery was realized in 0.2 mL tubes and the subsequent steps leading to amplification of tagged DNA molecules were conducted in the same reaction tube, thus warranting no or minimal loss of the already exiguous genetic material. Second, we detected no variants different from the expected ones in all of the MDA/MCF-7 pools tested. Although our findings were slightly different compared to the literature (especially regarding the MDA cell line [13,33,34]), the relatively small panel size used in our experiments must be taken into account, as it did not cover large DNA target regions compared to other NGS panels. Third, no false positives were found in the sequencing analysis of the libraries derived from isolated leukocytes, thus confirming that consistent data on the specificity of the NGS process were obtained (Table 2). Fourth, our results showed that the read counts per mutated DNA molecule (median of 11.5 for *KRAS* p.Gly13Asp, 18.5 reads for *TP53* p.Arg280Lys, and 16 for *PIK3CA* Glu545Lys) were consistently higher, compared to the minimum value suggested by the manufacturer to enable a valid variant calling (i.e., equivalent to at least “3” reads).

The proposed workflow, beyond its technical aspects, also has interesting and important cost-efficiency related advantages. In fact, the reagent volume reduction allowed us to triplicate the number of reactions possible for the OBcfRAv2 kit, thus raising the number of libraries to a total of 24 preparations (compared to the eight proposed by the manufacturer). Moreover, the procedure

supports “low-coverage sequencing”, as the scarce genetic material used as input does not require a high level of read depth.

We are aware that the proposed NGS workflow needs further refinements to be definitively applied; for example, a critical point in the proposed procedure regards the allele counts per target region amplified (Total Molecular Coverage). Indeed, the observed read counts were usually higher than those expected (data not shown). We speculated that this could be due to misreading of the molecular tags during the amplification cycles, causing an event known as “barcode resampling” which resulted in an increase of the allele counts [35]. However, this event did not seem to alter the accuracy of the overall variant calling.

In the second phase of our study, the possible translational value of the proposed workflow was proved in a “real” clinical setting. This approach could have great value in clinical settings, where liquid biopsies are now starting to be included in the validated tests panel used to detect mutations for target therapies, such as the use of alpelisib for the treatment of patients with hormone receptor-positive, HER2-negative, and *PIK3CA*-mutated advanced or metastatic BC progressed after previous endocrine therapy as monotherapy [36]. In this context, one of the most intriguing aspects of our protocol is the possibility to study CTCs molecular profiles in women who are progressing many years after the primary tumor and who do not have biopsy from relapsed mammary tissue or from other progression sites. In these cases, we believe that this molecular analysis will allow us to discover emerging mutations, which is important for designing treatment approaches. In this view, comparative studies on ctDNA and CTCs will shed new light on the importance of studying both compartments in the BC setting.

To test the protocol’s applicability in a real clinical setting, it was applied to 13 different pools of CTCs isolated from eight BC patients. We found that 7/13 BC CTCs samples had a mutation occurring; mainly in the *TP53* and *PIK3CA* genes, as these genes are the most frequently mutated in BC [37–39].

Comparison between CTCs and matched primary BC tumor tissues demonstrated complete concordant mutational data in 5/8 BC patients; whereas two patients were discordant, and a partial match was found in one patient. These data were not surprising and are in line with the current literature, showing a certain rate of heterogeneity in BC [30,40–42].

It is necessary to note that the NGS panel utilized for neoplastic tissues sequencing covers a larger number of target regions, compared to the OBcfRAv2 kit; thus, some tissues displayed other mutations not identified in the corresponding CTCs.

To the best of our knowledge, this is the first work where a molecular tagging-based NGS without a preamplification step was applied for the molecular analysis of single CTCs, even if some other studies have adapted amplicon-based NGS technology to the mutational profiling of CTCs [23,24]. Although the idea to skip WGA is not novel, the contributions of more laboratories are still required in order to finally obtain reliable workflows for CTCs purposes. In this regard, we believe that our approach is user-friendly, robust, and easy to implement in a routine setting by molecular clinical laboratories, providing the intriguing possibility to apply an already standardized and accessible (commercial) assay for both ctDNA and CTCs molecular analyses.

Compared to other, highly sensitive approaches which are applied for liquid biopsy purposes, such as BEAMing (Beads, Emulsions, Amplification, and Magnetics) assay [43], this procedure allows the use of minimal input amount of genetic material (picograms) and the examination of many DNA target regions simultaneously. Instead, BEAMing is able to test only few target regions at once and, consequently, are not exactly suitable when dealing with comprehensive mutational investigations of CTCs.

## 4. Materials and Methods

### 4.1. Method Optimization

#### 4.1.1. Cell Line, Spiking Experiments, and DEPArray™ Cell Sorting

BC-derived MDA (triple-negative) and MCF-7 (HR/PgR+, HER2-) cell lines were cultured in complete DMEM medium (Gibco®, Waltham, MA, USA).

Among the variants reported to occur in MDA and MCF-7 cells (see ATCC website, <https://www.lgcstandards-atcc.org/~{/media/03E3D0D9F0C74FE593CE7D010F982CA2.ashx}>), two specific variants were investigated for MDA (i.e., the c.38G > A, p.Gly13Asp in exon 2 of the *KRAS* gene and the c.839G > A, p.Arg280Lys in exon 8 of *TP53*) and one for MCF-7 (the c.1633G > A, p.Glu545Lys in exon 10 of the *PIK3CA* gene), as their hotspot positions were also covered by the NGS panel utilized for sequencing analyses below.

A total of  $1-10 \times 10^3$  MDA and MCF-7 cells were separately spiked in 7.5 mL healthy donor blood and enrichment was performed using immunologic negative selection (adapted from Bulfoni et al. [44]). Briefly, after red blood cell lysis (Red Blood Cell Lysis Solution, Miltenyi Biotec, Bologna, Italy), cells were incubated with CD45 and CD235 A (Glycophorin A) MicroBeads antibodies (Abs, Miltenyi, Bologna, Italy) and depletion of leukocytes and red blood cells were realized through magnetic LD separation columns. The negative flow-throughs containing MDA/MCF-7 cells were then labeled with anti-EpCAM (clone 9C4; BioLegend, Koblenz, Germany) fluorescein isothiocyanate- (FITC), anti-CD44 phycoerythrin (PE, Clone 515 Becton Dickinson, Milano, Italy), anti-CD45 allophycocyanin (APC) conjugated (Clone 5B1 Miltenyi, Bologna, Italy), and Hoechst 33,342 (Merck Life Science, Milano, Italy). Then, cells were fixed and permeabilized (Inside stain kit, Miltenyi, Bologna, Italy) for complete nuclei staining, reduced to a volume of 13–14  $\mu$ L after washing with SB115, loaded into an A300K V2 DEPArray™ cartridge and, finally, selected and sorted on a DEPArray™ system (Menarini Silicon Biosystems, Bologna, Italy).

MDA/MCF-7 cells were identified on the basis of morphological (as medium- or large-sized cells) and phenotypic features (such as CD44/EpCAM positivity and CD45 negativity), and isolated individually or in pools composed of 2, 3, 4, or 5 cells. Combined pools containing MDA or MCF-7 and leukocytes were arranged according to different cells: leukocytes ratios of 4:1, 3:2, 2:3, and 1:4 were used. Leukocytes (bright CD45 positive small size cells) were also recovered grouped in two and five cells.

Cells were collected into 0.2 mL tubes and stored in a final volume of 1–2  $\mu$ L of PBS at  $-20$  °C until molecular analysis.

#### 4.1.2. Evaluation of DNA Quality

DNA from single or pooled cellular preparations of both MDA cells and leukocytes were assessed by a WGA assay (Ampli1 WGA kit, Silicon Biosystems, Bologna, Italy) and multiplex PCR with the Ampli1 QC kit (Silicon Biosystems, Bologna, Italy), which amplified four different chromosomal regions with different lengths (91, 108–166, 299, and 614 bp, respectively with 12 p, 5 p, 17 p, and 6 p chromosomes), as previously described [25]. The number of bands obtained by each PCR product was visualized on a 3% agarose gel through UV light.

#### 4.1.3. Canonical NGS Workflow Assessment

DNA from single or pools of MDA cells was obtained by treatment with 2  $\mu$ L of Lysis Reaction Mix (Ampli1 WGA kit, Silicon Biosystems, Bologna, Italy).

MDA cell lysates were processed together with 5 ng of normal genomic DNA from healthy donor blood and MDA/normal DNA mixtures were used for library construction with the OBcfRAv2 (ThermoFisher Scientific, Carlsbad, CA, USA), following the manufacturer's instructions ([https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017065\\_Lung\\_Breast\\_](https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017065_Lung_Breast_)

[cfTNA\\_Assay\\_UG.pdf](#). The assay targeted more than 150 hotspot positions relatively to 10 genes: *AKT1*, *EGFR*, *ERBB2*, *ERBB3*, *ESR1*, *FBXW7*, *KRAS*, *PIK3CA*, *SF3B1*, and *TP53* (for further information, see <https://www.thermofisher.com/order/catalog/product/A35865>).

Manually prepared libraries were quantified by qPCR assay with an Ion Library TaqMan™ Quantitation kit (ThermoFisher Scientific, Carlsbad, CA, USA), diluted to 50–60 pMol and then multiplexed up to 5 samples for automated template preparation on an Ion Chef™ System and loaded on an Ion 530™ Chip (ThermoFisher Scientific, Carlsbad, CA, USA). Sequencing runs were performed on an Ion Torrent GeneStudio™ S5 (ThermoFisher Scientific, Carlsbad, CA, USA), according to the manufacturer's user manual.

#### 4.1.4. Optimized Molecular Tagging NGS Workflow

Libraries were prepared using directly the 3 µL cell lysates from isolated cells as a source of DNA. The entire protocol for library preparation with OBcfRAv2 was modified to adjust the amplification of small quantities of DNA input into a small reaction volume.

All reactions used for library construction were set up with a 1:3 dilution reduced volume (see details in Table S1). Thermal conditions and number of amplification cycles were maintained as the canonical original protocol suggested (Table S1).

Manually prepared libraries were quantified, diluted, and used for template preparation, as described in Section 4.1.3. The only relevant modification consisted of multiplexing up to 24 amplified libraries on an Ion 520™ Chip (low-coverage sequencing). Run sequencing was performed on an Ion Torrent GeneStudio™ S5, as described above.

The overall study design and the main changes throughout the entire procedure are summarized in Table 1.

#### 4.1.5. NGS Analysis

Molecular data analysis was performed on the Torrent Suite (version 5.10) software. The Sequencing Binary Alignment Map (BAM) file of each library was generated through an automatic pipeline already set by ThermoFisher, which provides read alignment and mapping on reference hg19, thanks to the Torrent Mapping Alignment Program (TMAP). The TVC plugin was run, in order to get the VCF files and list all the mutations that passed the quality filters. Then, VCFs were visualized and interpreted using the Integrative Genomic Viewer (IGV, Broad Institute, Cambridge, MA, USA).

### 4.2. CTCs Testing

#### 4.2.1. BC Patients

Eight BC patients were randomly chosen from those enrolled in a translational trial conducted at IRCCS Ospedale Policlinico San Martino (Citrucel Protocol approved by Regione Liguria Ethics Committee, Decision No. 0519, 4 May 2016). In particular, six patients were treated in the neoadjuvant settings, one patient in the adjuvant setting, and one patient in the metastatic BC stage. All patients gave their written informed consent before their enrolment.

#### 4.2.2. Enrichment, Fixation, and Sorting of CTCs from BC Patients

Blood draws were collected, in each case, before systemic treatment. BC CTCs were isolated according to the same enrichment, fixation, and sorting procedures used for MDA/MCF-7 cell lines (see Section 4.1.1). Due to the variability of EpCAM and/or CD44 expression in patient cells, DEPArray™ selection of putative CTCs was mainly based on morphology (oval shape and round features), positive Hoechst staining, and the complete absence of CD45 expression; as well as cell diameter higher than 10 µm and a nucleus–cytoplasm ratio > 0.5.

Cell sorting was differentially arranged in various pools through the 8 patients, depending on the number of CTCs identified in their blood (pool range: 2–6 cells). Recovered pools were stored at  $-20^{\circ}\text{C}$  until molecular analysis.

NGS tests on CTCs were performed using the optimized molecular tagging NGS workflow described in Section 4.1.4 and the analysis was conducted as in Section 4.1.5.

### 4.3. NGS of BC Tissues

A custom panel based on the Ion AmpliSeq™ technology was designed to study common alterations in BC, with overlapping target regions present in the OBcfRAv2. Specifically, it is comprehensive of 291 amplicons distributed within 20 genes (i.e., *TP53*, *PIK3CA*, *ERBB2*, *ERBB3*, *ERBB4*, *ESR1*, *MCL1*, *GATA3*, *PTEN*, *CCND1*, *KRAS*, *AKT1*, *CDH1*, *MAP2K4*, *SF3B1*, *FBXW7*, *MAP3K1*, *PIK3R1*, *EGFR*, and *FGFR1*).

Matched Formalin-Fixed Paraffin-Embedded (FFPE) tumor tissue sections (5  $\mu\text{m}$  thickness) were used for automated genomic DNA extraction with a QIASymphony instrument (Qiagen, Milano, Italy). Tumor tissue enrichment was performed by macro-dissection of areas containing at least 50% of neoplastic cells.

A total of 10–15 nanograms of DNA were quantified by a Qubit™ 3.0 fluorometer, used for automated library construction on the Ion Chef™ with an Ion AmpliSeq™ Kit for Chef DL8 (ThermoFisher Scientific, Carlsbad, CA, USA).

Fifty pMol of eight mixed libraries were used for automated template preparation on the Ion Chef™ System and loading on an Ion 520™ Chip. The run sequencing step was performed on the Ion Torrent GeneStudio™ S5.

NGS data analysis was executed in the Ion Torrent Suite and the Coverage Analysis and TVC plugins were run. VCFs were visualized on IGV and variants annotation was performed with the Ion Reporter Software (version 5.10, ThermoFisher Scientific, Carlsbad, CA, USA).

## 5. Conclusions

This study has demonstrated, for the first time, the potential application of a molecular tagging NGS panel, which is already standardized for other liquid biopsy purposes, as a feasible mutational approach for CTCs sequencing. Here, we demonstrated that accurate tumor cells isolation by the DEPArray™ System coupled with a precise and sensitive NGS procedure can represent not only an alternative strategy for CTCs mutational profiling, but also a reliable WGA-free method.

In this direction, these findings open the way to future translational applications of the proposed workflow in precision medicine laboratories which deal with the liquid biopsy field, as well as further supporting the importance of CTCs as a useful tool for prognostic, diagnostic, and therapeutic applications.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1422-0067/21/12/4364/s1>, Table S1. Details of single NGS library preparation steps in both canonical and optimized workflows.

**Author Contributions:** Conceptualization, G.D.L. and M.D.; methodology, G.D.L., B.C., and S.L.; software, G.D.L. and B.C.; validation, G.D.L. and M.D.; formal analysis, G.D.L. and B.C.; investigation, G.D.L. and M.D.; resources, L.D.M., F.C., A.G., and C.M.; data curation, G.D.L., B.C., and L.D.M.; writing—original draft preparation, G.D.L. and B.C.; writing—review and editing, G.D.L., B.C., L.D.M., M.F., G.A.C., S.Z., and M.D.; visualization, G.D.L. and B.C.; supervision, S.Z.; project administration, S.Z. and M.D.; funding acquisition, L.D.M. and S.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by AIRC IG2017, grant number 20760 and Italian Ministry of Health  $5 \times 1000$ , funds 2015–2016.

**Acknowledgments:** We acknowledge Alessandro Albiero (ThermoFisher Scientific) for contribution to interpretation of the bioinformatics data, Andrea Sciuotto for technical assistance, Daniele Reverberi for help in FACS analysis and Francesca Poggio for patients selection

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

|          |  |
|----------|--|
| CTCs     | Circulating Tumor Cells  |
| WGA      | Whole Genome Amplification   |
| NGS      | Next-Generation Sequencing   |
| ctDNA    | Circulating tumor DNA  |
| BC       | Breast Cancer  |
| MDA      | MDA-MB-231   |
| EpCAM    | Epithelial Cell Adhesion Molecule                                      |
| FITC     | fluorescein isothiocyanate   |
| PE       | phycoerythrin  |
| APC      | allophycocyanin  |
| FACS     | Fluorescent-Activated Cell Sorter                                      |
| PCR      | Polymerase Chain Reaction  |
| OBcfRAv2 | Oncomine Breast cfDNA Research Assay v2                                |
| KRAS     | Kirsten rat sarcoma 2 viral oncogene homolog                           |
| TP53     | Tumor Protein 53   |
| MAFs     | Molecular Allele Frequencies   |
| VFSH     | Variant Family Size Histogram  |
| TVC      | Torrent Variant Caller   |
| PIK3CA   | Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha |
| ERBB2    | Erb-B2 Receptor Tyrosine Kinase 2                                      |
| ERBB3    | Erb-B2 Receptor Tyrosine Kinase 3                                      |
| ERBB4    | Erb-B2 Receptor Tyrosine Kinase 4                                      |
| ESR1     | Estrogen Receptor 1  |
| MCL1     | MCL1 Apoptosis Regulator   |
| GATA3    | GATA Binding Protein 3   |
| PTEN     | Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase                 |
| CCND1    | Cyclin D1  |
| AKT1     | v-akt murine thymoma viral oncogene homolog 1                          |
| CDH1     | Cadherin 1   |
| MAP2K4   | Mitogen-Activated Protein Kinase Kinase 4                              |
| SF3B1    | Splicing factor 3b subunit 1   |
| FBXW7    | F-box/WD repeat-containing protein 7                                   |
| MAP3K1   | Mitogen-Activated Protein Kinase Kinase Kinase 1                       |
| PIK3R1   | Phosphoinositide-3-Kinase Regulatory Subunit 1                         |
| EGFR     | Epidermal Growth Factor Receptor                                       |
| FGFR1    | Fibroblast Growth Factor Receptor 1                                    |
| FFPE     | Formalin-Fixed Paraffin-Embedded                                       |
| BAM      | Binary Alignment Map   |
| VCF      | Variant Caller Format  |
| IGV      | Integrative Genomic Viewer   |
| TMAP     | Torrent Mapping Alignment Program                                      |
| DMEM     | Dulbecco's Modified Eagle Medium                                       |
| PBS      | Phosphate Buffered Saline  |
| BEAMing  | Beads, Emulsions, Amplification, and Magnetics                         |

## References

1. Yu, M.; Stott, S.; Toner, M.; Maheswaran, S.; Haber, D.A. Circulating tumor cells: Approaches to isolation and characterization. *J. Cell Biol.* **2011**, *192*, 373–382. [[CrossRef](#)] [[PubMed](#)]
2. Alix-Panabieres, C.; Pantel, K. Circulating tumor cells: Liquid biopsy of cancer. *Clin. Chem.* **2013**, *59*, 110–118. [[CrossRef](#)] [[PubMed](#)]



3. Bidard, F.C.; Peeters, D.J.; Fehm, T.; Nolé, F.; Criado, R.G.; Mavroudis, D.; Grisanti, S.; Generali, D.; Garcia-Saenz, J.A.; Stebbing, J.; et al. Clinical validity of circulating tumour cells in patients with metastatic breast cancer: A pooled analysis of individual patient data. *Lancet Oncol.* **2014**, *15*, 406–414. [[CrossRef](#)]
4. Huang, X.; Gao, P.; Song, Y.; Sun, J.; Chen, X.; Zhao, J.; Xu, H.; Wang, Z. Meta-analysis of the prognostic value of circulating tumor cells detected with the CellSearch System in colorectal cancer. *BMC Cancer* **2015**, *15*, 202. [[CrossRef](#)]
5. Tol, J.; Koopman, M.; Miller, M.C.; Tibbe, A.; Cats, A.; Creemers, G.J.M.; Vos, A.H.; Nagtegaal, I.D.; Terstappen, L.W.M.M.; Punt, C.J.A. Circulating tumour cells early predict progression-free and overall survival in advanced colorectal cancer patients treated with chemotherapy and targeted agents. *Ann. Oncol.* **2010**, *21*, 1006–1012. [[CrossRef](#)]
6. Krebs, M.G.; Sloane, R.; Priest, L.; Lancashire, L.; Hou, J.M.; Greystoke, A.; Ward, T.H.; Ferraldeschi, R.; Hughes, A.; Clack, G.; et al. Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer. *J. Clin. Oncol.* **2011**, *29*, 1556–1563. [[CrossRef](#)]
7. Sparano, J.; O'Neill, A.; Alpaugh, K.; Wolff, A.C.; Northfelt, D.W.; Dang, C.T.; Sledge, G.W.; Miller, K.D. Association of Circulating Tumor Cells with Late Recurrence of Estrogen Receptor-Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Oncol.* **2018**, *4*, 1700–1706. [[CrossRef](#)]
8. Alix-Panabières, C.; Pantel, K. Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. *Cancer Discov.* **2016**, *6*, 479–491. [[CrossRef](#)]
9. Moon, D.H.; Lindsay, D.P.; Hong, S.; Wang, A.Z. Clinical indications for, and the future of, circulating tumor cells. *Adv. Drug Deliv. Rev.* **2018**, *125*, 143–150. [[CrossRef](#)]
10. Lianidou, E.; Pantel, K. Liquid Biopsies. *Genes Chromosomes Cancer* **2019**, *58*, 219–232. [[CrossRef](#)]
11. Zhang, H.; Gong, S.; Liu, Y.; Liang, L.; He, S.; Zhang, Q.; Si, M.; Yu, Z. Enumeration and molecular characterization of circulating tumor cell using an in vivo capture system in squamous cell carcinoma of head and neck. *Chin. J. Cancer Res.* **2017**, *29*, 196–203. [[CrossRef](#)]
12. They, L.; Meddis, A.; Cabel, L.; Proudhon, C.; Latouche, A.; Pierga, J.Y.; Bidard, F.C. Circulating Tumor Cells in Early Breast Cancer. *JNCI Cancer Spectrum.* **2019**, *3*, pkz026. [[CrossRef](#)] [[PubMed](#)]
13. Salvianti, F.; Rotunno, G.; Galardi, F.; De Luca, F.; Pestrin, M.; Vannucchi, A.M.; Di Leo, A.; Pazzagli, M.; Pinzani, P. Feasibility of a workflow for the molecular characterization of single cells by next generation sequencing. *Biomol. Detect. Quantif.* **2015**, *5*, 23–29. [[CrossRef](#)]
14. De Luca, F.; Rotunno, G.; Salvianti, F.; Galardi, F.; Pestrin, M.; Gabellini, S.; Simi, L.; Mancini, I.; Vannucchi, A.M.; Pazzagli, M.; et al. Mutational analysis of single circulating tumor cells by next generation sequencing in metastatic breast cancer. *Oncotarget* **2016**, *7*, 26107–26119. [[CrossRef](#)] [[PubMed](#)]
15. Yee, S.S.; Lieberman, D.B.; Blanchard, T.; Rader, J.; Zhao, J.; Troxel, A.B.; DeSloover, D.; Fox, A.J.; Daber, R.D.; Kakrecha, B. A novel approach for next-generation sequencing of circulating tumor cells. *Mol. Genet. Genom. Med.* **2016**, *4*, 395–406. [[CrossRef](#)] [[PubMed](#)]
16. Sho, S.; Court, C.M.; Winograd, P.; Lee, S.; Hou, S.; Graeber, T.G.; Tseng, H.R.; Tomlinson, J.S. Precision oncology using a limited number of cells: Optimization of whole genome amplification products for sequencing applications. *BMC Cancer* **2017**, *17*, 457. [[CrossRef](#)] [[PubMed](#)]
17. Pinard, R.; de Winter, A.; Sarkis, G.J.; Gerstein, M.B.; Tartaro, K.R.; Plant, R.N.; Egholm, M.; Rothberg, J.M.; Leamon, J.H. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genom.* **2006**, *7*, 216. [[CrossRef](#)] [[PubMed](#)]
18. Lasken, R.S.; Egholm, M. Whole genome amplification: Abundant supplies of DNA from precious samples or clinical specimens. *Trends Biotechnol.* **2003**, *21*, 531–535. [[CrossRef](#)]
19. Hughes, S.; Arneson, N.; Done, S.; Squire, J. The use of whole genome amplification in the study of human disease. *Prog. Biophys. Mol. Biol.* **2005**, *88*, 173–189. [[CrossRef](#)]
20. Huang, L.; Ma, F.; Chapman, A.; Lu, S.; Xie, X.S. Single-cell whole-genome amplification and sequencing: Methodology and applications. *Annu. Rev. Genom. Hum. Genet.* **2015**, *16*, 79–102. [[CrossRef](#)]
21. Zhu, Z.; Qiu, S.; Shao, K.; Hou, Y. Progress and challenges of sequencing and analyzing circulating tumor cells. *Cell Biol. Toxicol.* **2018**, *34*, 405–415. [[CrossRef](#)] [[PubMed](#)]
22. Zahn, H.; Steif, A.; Laks, E.; Eirew, P.; VanInsberghe, M.; Shah, S.P.; Aparicio, S.; Hansen, C.L. Scalable whole-genome single-cell library preparation without preamplification. *Nat. Methods* **2017**, *14*, 167–173. [[CrossRef](#)] [[PubMed](#)]

23. Palmirotta, R.; Lovero, D.; Silvestris, E.; Felici, C.; Quaresmini, D.; Cafforio, P.; & Silvestris, F. Next-generation Sequencing (NGS) Analysis on Single Circulating Tumor Cells (CTCs) with No Need of Whole-genome Amplification (WGA). *Cancer Genom. Proteom.* **2017**, *14*, 173–179. [[CrossRef](#)] [[PubMed](#)]
24. D'Oronzo, S.; Lovero, D.; Palmirotta, R.; Stucci, L.S.; Tucci, M.; Felici, C.; Cascardi, E.; Giardina, C.; Caforio, P.; Silvestris, F. Dissection of major cancer gene variants in subsets of circulating tumor cells in advanced breast cancer. *Sci. Rep.* **2019**, *9*, 17276. [[CrossRef](#)] [[PubMed](#)]
25. Polzer, B.; Medoro, G.; Pasch, S.; Fontana, F.; Zorzino, L.; Pestka, A.; Andergassen, U.; Meier-Stiegen, F.; Czyz, Z.T.; Alberter, B. Molecular profiling of single circulating tumor cells with diagnostic intention. *EMBO Mol. Med.* **2014**, *6*, 1371–1386. [[CrossRef](#)]
26. Cristofanilli, M.; Budd, G.T.; Ellis, M.J.; Stopeck, A.; Matera, J.; Miller, M.C.; Reuben, J.M.; Doyle, G.V.; Allard, W.J.; Terstappen, L.W.M.M.; et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N. Engl. J. Med.* **2004**, *351*, 781–791. [[CrossRef](#)]
27. Hayes, D.F.; Cristofanilli, M.; Budd, G.T.; Ellis, M.J.; Stopeck, A.; Miller, M.C.; Matera, J.; Allard, W.J.; Doyle, G.V.; Terstappen, L.W.W.M. Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival. *Clin. Cancer Res.* **2006**, *12*, 4218–4224. [[CrossRef](#)]
28. Kanwar, N.; Hu, P.; Bedard, P.; Clemons, M.; McCready, D.; Done, S.J. Identification of genomic signatures in circulating tumor cells from breast cancer. *Int. J. Cancer* **2015**, *137*, 332–344. [[CrossRef](#)]
29. Liu, H.E.; Triboulet, M.; Zia, A.; Vuppapalaty, M.; Kidess-Sigal, E.; Coller, J.; Natsu, V.S.; Shokoohi, V.; Che, J.; Renier, C.; et al. Workflow optimization of whole genome amplification and targeted panel sequencing for CTC mutation detection. *NPJ Genom. Med.* **2017**, *2*, 34. [[CrossRef](#)]
30. Bingham, C.; Fernandez, S.V.; Fittipaldi, P.; Dempsey, P.W.; Ruth, K.J.; Cristofanilli, M.; Katherine Alpaugh, R. Mutational studies on single circulating tumor cells isolated from the blood of inflammatory breast cancer patients. *Breast Cancer Res. Treat.* **2017**, *163*, 219–230. [[CrossRef](#)]
31. Dono, M.; De Luca, G.; Lastraioli, S.; Anselmi, G.; Dal Bello, M.G.; Coco, S.; Vanni, I.; Grossi, F.; Vigani, A.; Genova, C.; et al. Tag-based next generation sequencing: A feasible and reliable assay for EGFR T790M mutation detection in circulating tumor DNA of non small cell lung cancer patients. *Mol. Med.* **2019**, *25*, 15. [[CrossRef](#)] [[PubMed](#)]
32. De Luca, G.; Lastraioli, S.; Conte, R.; Mora, M.; Genova, C.; Rossi, G.; Tagliamento, M.; Coco, S.; Dal Bello, M.G.; Zupo, S.; et al. Performance of the OncoPrint™ Lung cfDNA Assay for Liquid Biopsy by NGS of NSCLC Patients in Routine Laboratory Practice. *Appl. Sci.* **2020**, *10*, 2895. [[CrossRef](#)]
33. Abaan, O.D.; Polley, E.C.; Davis, S.R.; Zhu, Y.J.; Bilke, S.; Walker, R.L.; Pineda, M.; Gindin, Y.; Jiang, Y.; Reinhold, W.C. The exomes of the NCI-60 panel: A genomic resource for cancer biology and systems pharmacology. *Cancer Res.* **2013**, *73*, 4372–4382. [[CrossRef](#)]
34. Hollestelle, A.; Nagel, J.; Smid, M.; Lam, S.; Elstrodt, F.; Wasielewski, M.; Ng, S.; French, P.; Peeters, J.; Rozendaal, M. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res. Treat.* **2010**, *121*, 53–64. [[CrossRef](#)] [[PubMed](#)]
35. Peng, Q.; Vijaya Satya, R.; Lewis, M.; Randad, P.; Wang, Y. Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC Genom.* **2015**, *7*, 589. [[CrossRef](#)] [[PubMed](#)]
36. André, F.; Ciruelos, E.; Rubovszky, G.; Campone, M.; Loibl, S.; Rugo, H.S.; Iwata, H.; Conte, P.; Mayer, I.A.; Kaufman, B.; et al. SOLAR-1 Study Group. Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. *N. Engl. J. Med.* **2019**, *380*, 1929–1940. [[CrossRef](#)]
37. Forbes, S.A.; Beare, D.; Boutselakis, H.; Bamford, S.; Bindal, N.; Tate, J.; Cole, C.G.; Ward, S.; Dawson, E.; Ponting, L.; et al. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **2017**, *45*, D777–D783. [[CrossRef](#)] [[PubMed](#)]
38. Nik-Zainal, S.; Davies, H.; Staaf, J.; Ramakrishna, M.; Glodzik, D.; Zou, X.; Martincorena, I.; Alexandrov, L.B.; Martin, S.; Wedge, D.C. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **2016**, *534*, 47–54. [[CrossRef](#)]
39. Pereira, B.; Chin, S.F.; Rueda, O.M.; Volland, H.K.; Provenzano, E.; Bardwell, H.A.; Pugh, M.; Jones, L.; Russell, R.; Sammut, S.J. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun.* **2016**, *7*, 11479. [[CrossRef](#)]

40. Pestrin, M.; Salvianti, F.; Galardi, F.; De Luca, F.; Turner, N.; Malorni, L.; Pazzagli, M.; Di Leo, A.; Pinzani, P. Heterogeneity of PIK3CA mutational status at the single cell level in circulating tumor cells from metastatic breast cancer patients. *Mol. Oncol.* **2015**, *9*, 749–757. [[CrossRef](#)]
41. Fernandez, S.V.; Bingham, C.; Fittipaldi, P.; Austin, L.; Palazzo, J.; Palmer, G.; Alpaugh, K.; Cristofanilli, M. TP53 mutations detected in circulating tumor cells present in the blood of metastatic triple negative breast cancer patients. *Breast Cancer Res.* **2014**, *16*, 445. [[CrossRef](#)] [[PubMed](#)]
42. Paolillo, C.; Mu, Z.; Rossi, G.; Schiewer, M.J.; Nguyen, T.; Austin, L.; Capoluongo, E.; Knudsen, K.; Cristofanilli, M.; Fortina, P. Detection of Activating Estrogen Receptor Gene (ESR1) Mutations in Single Circulating Tumor Cells. *Clin. Cancer Res.* **2017**, *23*, 6086–6093. [[CrossRef](#)] [[PubMed](#)]
43. García-Foncillas, J.; Alba, E.; Aranda, E.; Díaz-Rubio, E.; López-López, R.; Tabernero, J.; Vivancos, A. Incorporating BEAMing technology as a liquid biopsy into clinical practice for the management of colorectal cancer patients: An expert taskforce review. *Ann. Oncol.* **2017**, *28*, 2943–2949. [[CrossRef](#)] [[PubMed](#)]
44. Bulfoni, M.; Gerratana, L.; Del Ben, F.; Marzinotto, S.; Sorrentino, M.; Turetta, M.; Scoles, G.; Toffoletto, B.; Isola, M.; Beltrami, C.A.; et al. In patients with metastatic breast cancer the identification of circulating tumor cells in epithelial-to-mesenchymal transition is associated with a poor prognosis. *Breast Cancer Res.* **2016**, *18*, 30. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).