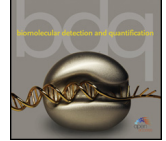




## Biomolecular Detection and Quantification

journal homepage: [www.elsevier.com/locate/bdq](http://www.elsevier.com/locate/bdq)

Research paper

## Feasibility of a workflow for the molecular characterization of single cells by next generation sequencing

Francesca Salvianti<sup>a</sup>, Giada Rotunno<sup>b</sup>, Francesca Galardi<sup>c</sup>, Francesca De Luca<sup>c</sup>, Marta Pestrin<sup>c</sup>, Alessandro Maria Vannucchi<sup>b</sup>, Angelo Di Leo<sup>c</sup>, Mario Pazzagli<sup>a</sup>, Pamela Pinzani<sup>a,\*</sup><sup>a</sup> Department of Clinical, Experimental and Biomedical Sciences, University of Florence, Florence, Italy<sup>b</sup> Department of Clinical and Experimental Medicine, University of Florence, Florence, Italy<sup>c</sup> Sandro Pitigliani Medical Oncology Department, Hospital of Prato, Istituto Toscano Tumori, Prato, Italy

## ARTICLE INFO

## Article history:

Received 29 April 2015

Received in revised form 24 July 2015

Accepted 27 July 2015

Available online 12 September 2015

## Keywords:

Single cell

Circulating tumor cells

NGS

Somatic mutations

Microgenomics

## ABSTRACT

The purpose of the study was to explore the feasibility of a protocol for the isolation and molecular characterization of single circulating tumor cells (CTCs) from cancer patients using a single-cell next generation sequencing (NGS) approach.

To reach this goal we used as a model an artificial sample obtained by spiking a breast cancer cell line (MDA-MB-231) into the blood of a healthy donor.

Tumor cells were enriched and enumerated by CellSearch<sup>®</sup> and subsequently isolated by DEPArray<sup>™</sup> to obtain single or pooled pure samples to be submitted to the analysis of the mutational status of multiple genes involved in cancer.

Upon whole genome amplification, samples were analysed by NGS on the Ion Torrent PGM<sup>™</sup> system (Life Technologies) using the Ion AmpliSeq<sup>™</sup> Cancer Hotspot Panel v2 (Life Technologies), designed to investigate genomic “hot spot” regions of 50 oncogenes and tumor suppressor genes.

We successfully sequenced five single cells, a pool of 5 cells and DNA from a cellular pellet of the same cell line with a mean depth of the sequencing reaction ranging from 1581 to 3479 reads.

We found 27 sequence variants in 18 genes, 15 of which already reported in the COSMIC or dbSNP databases. We confirmed the presence of two somatic mutations, in the *BRAF* and *TP53* gene, which had been already reported for this cells line, but also found new mutations and single nucleotide polymorphisms. Three variants were common to all the analysed samples, while 18 were present only in a single cell suggesting a high heterogeneity within the same cell line.

This paper presents an optimized workflow for the molecular characterization of multiple genes in single cells by NGS. The described pipeline can be easily transferred to the study of single CTCs from oncologic patients.

© 2015 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The molecular analysis of solid tumors is of great impact in the field of oncology research notwithstanding some unsolved challenging aspects in both sample selection and technical approaches. Solid tumors are in fact a heterogeneous mixture of tumor and normal cells differentially contributing to the final results of molecular biology tests on nucleic acids. As a result molecular signatures rep-

resent an average value masking the individual signals deriving from each single cancer cell [1]. The analysis of somatic mutations at the single-cell level can evidence whether one or more variants detected in the bulk tissue are simultaneously present in the same cell or if, on the contrary, derive from different cell clones.

Next-generation sequencing (NGS) has become a major tool for nucleic acid analysis and promises to disclose the basis of diseases, opening perspectives for increased personalized therapies and screening [2]. NGS has been so far applied to nucleic acids deriving from a high number of cells providing global information on the average state of the entire population of cells [1]. By combining whole genome amplification (WGA) with NGS, it is now possible to achieve individual cancer cell sequencing [3].

\* Corresponding author at: Dipartimento di Scienze Biomediche, Sperimentali e Cliniche Università Degli Studi di Firenze Viale Pieraccini, 6 50139 Firenze, Italy.  
E-mail address: [p.pinzani@dfc.unifi.it](mailto:p.pinzani@dfc.unifi.it) (P. Pinzani).

Single-cell sequencing has the potential to address some important questions in oncology such as understanding tumor heterogeneity, reconstructing cell lineages and defining the genome of rare tumor cell populations [3–5].

In addition, single-cell technologies will contribute to elucidate rare cell events such as scarce cancer cells in clinical samples and the dissemination of circulating tumor cells (CTCs) in the blood of cancer patients [1,6]. Since CTCs reflect the heterogeneity of primary or metastatic tumors, their analysis at the single-cell level represent a real time liquid biopsy of the disease giving an insight in cancer progression, helping in the choice of target therapies or in the prediction of the efficacy of chemotherapeutic agents [1] and thus representing a key tool in the development of precision medicine.

Another aspect that can be considered as an application of single-cell sequencing is the analysis of cell lines derived from tumors that are commonly used as models in cancer research. This approach can be successfully adopted to identify those cell lines that more closely resemble the characteristics of the tumor under study [7]. Even if at a much lower extent than tissues, the cell line population usually considered as composed of identical cellular elements can hide a heterogeneity that could be disclosed only by single-cell analysis.

The MDA-MB-231 is an adherent human breast cancer cell line of epithelial origin, which is highly aggressive, invasive and poorly differentiated. MDA-MB-231 cells display a mesenchymal-like phenotype thus representing a good model for metastatic breast cancer [8].

In the present study we investigated the mutational status of multiple genes on an artificial sample obtained by spiking the MDA-MB-231 cell line into the blood of a healthy donor. The tumor cells were immunomagnetically isolated by CellSearch<sup>®</sup> (Veridex, USA), sorted by a dielectrophoretic system (DEPArray<sup>™</sup>, Silicon Biosystems, Italy) and then submitted to WGA and targeted deep sequencing. We used this model with the aim to demonstrate the feasibility of a workflow for the isolation and molecular characterization of single circulating tumor cells from cancer patients using a single-cell NGS approach.

## 2. Methods

### 2.1. Cell culture and spiking experiments

Experiments were conducted on the MDA-MB-231 human breast cancer cell line. MDA-MB-231 cells were obtained from Di.V.A.L. Toscana (Laboratory for Drug Validation and Antibody production, University of Florence) and maintained in DMEM (Lonza, Basel, Switzerland) + 10% FBS (PAA Laboratories, Austria) at 37 °C in 5% CO<sub>2</sub>. Cells were detached from plastic surface by Accutase (PAA Laboratories) to better preserve membrane antigens. MDA-MB-231 cells ( $n=500$ ) were spiked in blood samples from a healthy volunteer collected in CellSave<sup>®</sup> tubes (Veridex LLC, Raritan, NJ) and processed as described below. The remaining cells from the same culture were centrifuged and the resulting cell pellet was submitted to DNA extraction by QIAamp DNA Mini Kit (Qiagen, Germany).

### 2.2. Cell enrichment and enumeration

Cell enrichment and enumeration was performed by the CellSearch<sup>®</sup> System. 7.5 mL of whole blood were processed using the CellSearch<sup>™</sup> Epithelial Cell kit (Veridex LLC), which selects EpCAM positive cells using ferrofluids particles coated with EpCAM antibody. Cells were stained with the nuclear dye 4',6'-diamino-2-phenylindole (DAPI), anti-cytokeratin 8, 18 and 19-phycoerythrin (PE) labelled antibodies, and CD45 antibody labelled with allophy-

cocyanin (APC). After enrichment, isolated and stained cells were resuspended in the MagNest Device (Veridex LLC), labelled cells were analyzed in the CellTracks<sup>®</sup> Analyzer II (Veridex LLC) and cells identified and enumerated according to the criteria specified by the manufacturer's instructions.

### 2.3. Single cell sorting

Following CellSearch<sup>®</sup> enrichment the sample was recovered from the Veridex cartridge and loaded into the DEPArray<sup>™</sup> A300K chip (Silicon Biosystems) according to the manufacturer's instructions. The chip is a single-use, microfluidic cartridge containing an array of individually controllable electrodes, each one with embedded sensors. Chip scanning was performed by an automated fluorescence microscope to generate an image gallery, with cells selected according to their morphology (round shape, round nucleus within the cytoplasm) and staining pattern deriving from that of the CellSearch<sup>®</sup> system: DAPI positive, PE positive (CK8, CK18, CK19 positive cells), APC negative (CD45 negative cells).

After tumor cell identification, single cells or groups of few cells were recovered into 200  $\mu$ l tubes.

### 2.4. Whole genome amplification

Single cells or pools of cells were submitted to WGA using the Ampli1<sup>™</sup> WGA kit (Silicon Biosystems) according to the manufacturer's instructions, in order to obtain a sample suitable for sequencing analysis. The procedure is based on a ligation-mediated PCR following a site specific DNA digestion by *MseI* enzyme. The kit has no needs for precipitation steps avoiding DNA loss and provides a library of fragments of about 0.2–2 kb representing the entire genome. The kit uses a mixture of Taq polymerase with a proofreading enzyme, Pwo polymerase, that has been reported to have error rates more than 10 times lower than the error rate observed with Taq polymerase [9].

### 2.5. WGA quality control

The quality of the output product of the WGA reaction was assessed by the Ampli1<sup>™</sup> QC kit (Silicon Biosystems) according to the manufacturer's instructions. The kit is a PCR-based assay which implies the amplification of two distinct regions of the human genome to produce two amplicons (A and B) of 373 and 167 bp respectively. PCR products A and B were analyzed by capillary electrophoresis on the Agilent 2100 Bionalyzer. The presence of both amplification products indicates a successful WGA and consequently the suitability of the sample for downstream analysis.

### 2.6. Next generation sequencing

Sequencing analysis was performed on the Ion Torrent PGM<sup>™</sup> system (Life Technologies). We amplified the samples using the Ion AmpliSeq<sup>™</sup> Cancer Hotspot Panel v2 (Life Technologies) designed to target 207 amplicons covering mutations from 50 oncogenes and tumor suppressor genes. DNA quantification was assessed using the Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). Ten nanograms of DNA were used to prepare barcoded libraries using the Ion AmpliSeq<sup>™</sup> Library kit 2.0 and Ion Xpress<sup>™</sup> barcode adapters (Life Technologies). The libraries were purified with Agentcourt AMPure XP (Beckman Coulter) and quantified with the Ion Library Quantitation Kit (Life Technologies) on the StepOne Plus system (Applied Biosystem).

Template preparation was performed with the Ion OneTouch<sup>™</sup> 2 System and Ion One Touch ES. Finally sequencing was performed on PGM using Ion PGM<sup>™</sup> Sequencing 200 kit v2 (Life Technologies)

**Table 1**

Classification of genes of the AmpliSeq™ Cancer Hotspot Panel v2 on the basis of the presence of the MseI restriction site in one or more amplicons.

31 Unaffected genes	17 Partially affected genes	2 Totally affected genes
<i>ABL1</i>	<i>APC</i>	<i>NPM1</i>
<i>AKT1</i>	<i>ATM</i>	<i>JAK2</i>
<i>ALK</i>	<i>CDH1</i>	
<i>BRAF</i>	<i>EGFR</i>	
<i>CDKN2A</i>	<i>ERBB4</i>	
<i>CSF1R</i>	<i>FBXW7</i>	
<i>CTNNB1</i>	<i>FGFR2</i>	
<i>ERBB2</i>	<i>KDR</i>	
<i>EZH2</i>	<i>KIT</i>	
<i>FGFR1</i>	<i>KRAS</i>	
<i>FGFR3</i>	<i>MET</i>	
<i>FLT3</i>	<i>PIK3CA</i>	
<i>GNA11</i>	<i>PTEN</i>	
<i>GNAQ</i>	<i>RB1</i>	
<i>GNAS</i>	<i>RET</i>	
<i>HNF1A</i>	<i>SMAD4</i>	
<i>HRAS</i>	<i>VHL</i>	
<i>IDH1</i>		
<i>IDH2</i>		
<i>JAK3</i>		
<i>MLH1</i>		
<i>MPL</i>		
<i>NOTCH1</i>		
<i>NRAS</i>		
<i>PDGFRA</i>		
<i>PTPN11</i>		
<i>SMARCB1</i>		
<i>SMO</i>		
<i>SRC</i>		
<i>STK11</i>		
<i>TP53</i>		

on the Ion 316 chip V1. The run was set in order to achieve a 1000X coverage for each sample.

Due to the WGA method, involving an enzymatic cleavage of DNA by the MseI restriction enzyme, we could not sequence 49/207 amplicons of the panel. Table 1 lists the genes of the AmpliSeq™ Cancer Hotspot Panel v2 which are respectively totally, partially or non-affected by the enzymatic digestion. 31 genes had no amplicon cleaved by the enzyme; 17 genes had some amplicons cut by MseI, and two genes (*NPM1* and *JAK2*) were impossible to analyse because of the enzymatic cleavage.

## 2.7. Data analysis

All samples were processed using the Torrent Suite Software 3.6 and variant calling was performed running the Torrent Variant Caller plugin version 3.6.56708. Moreover, samples were analysed using NextGENe® software 2.3.1 (SoftGenetics, LLC, State College, PA). We considered only variants with balanced numbers of forward and reverse reads.

Each variant was investigated about its potential pathogenetic role using available gene mutations and SNPs databases and prediction algorithms (COSMIC, dbSNP, 1000GENOME, SIFT, Polyphen).

## 3. Results

### 3.1. Design of the workflow for single cell recovery

We defined a workflow for targeted NGS sequencing composed of the following steps covering the procedure from the blood draw to the deep sequencing data analysis. Fig. 1 reports the entire pipeline.

#### 3.1.1. Tumor cell enrichment and recovery

The preanalytical variables are taken into account by the use of a stabilizing blood collection tube which also provides cell fixation. Cells spiked into the blood of a healthy donor are enriched, fluorescently labelled and counted by the CellSearch® system. The platform provides the immunomagnetic capture of cells and their selection on the basis of positive staining for cytokeratins and negative signal for CD45. The left-over cartridge is handled in order to achieve the right volume for single cell sorting by the DEPArray™, which combines imaging technologies with the ability to manipulate and recover individual cells from a heterogeneous sample. The system performs a scanning of the sample by an automated fluorescence microscope generating an image gallery. Cells are selected according to their morphology (round shape, round nucleus within the cytoplasm) and fluorescence staining pattern deriving from that of the CellSearch® system. Selected cells are trapped in dielectrophoretic “cages” and moved singularly by manipulating the applied electric field [10].

The workflow has several critical points related to the complete CTC recovery from the Cellsearch® cartridge, to the subsequent volume reduction to fit the input volume of the DEPArray™ chip (14 µl) and to the volume reduction of single cell samples after DEPArray™ recovery.

In particular, after enrichment by the CellSearch® system we identified 130 cells (percentage of recovery = 26%). The cell suspension recovered from the CellSearch® was injected in the DEPArray chip: we identified 86 cells with a recovery ratio (defined as the ratio between the number of single cells identified by the DEPArray™ after CellSearch® enrichment and the total number of cells counted by the CellSearch®) of 66% as already reported in a previous study [11].

We sorted 5 single MDA-MB-231 cells and a pool of 5 cells to be submitted to deep sequencing. Fig. 2A shows the images of 2 single cells.

#### 3.1.2. WGA and quality control

We amplified the whole genome of our samples in a 3-day procedure required for the Ampli1 WGA™ kit. Handling time was reduced, but two overnight incubations were required. After the WGA reaction we verified the quality of the output product before proceeding to sequencing by a PCR based assay using the Ampli1™ QC kit. The five single cells and the pool of 5 cells showed both the expected PCR products A and B by capillary electrophoresis. Fig. 2B shows the capillary electrophoresis of the two products of the Ampli1™ QC kit in two single cells.

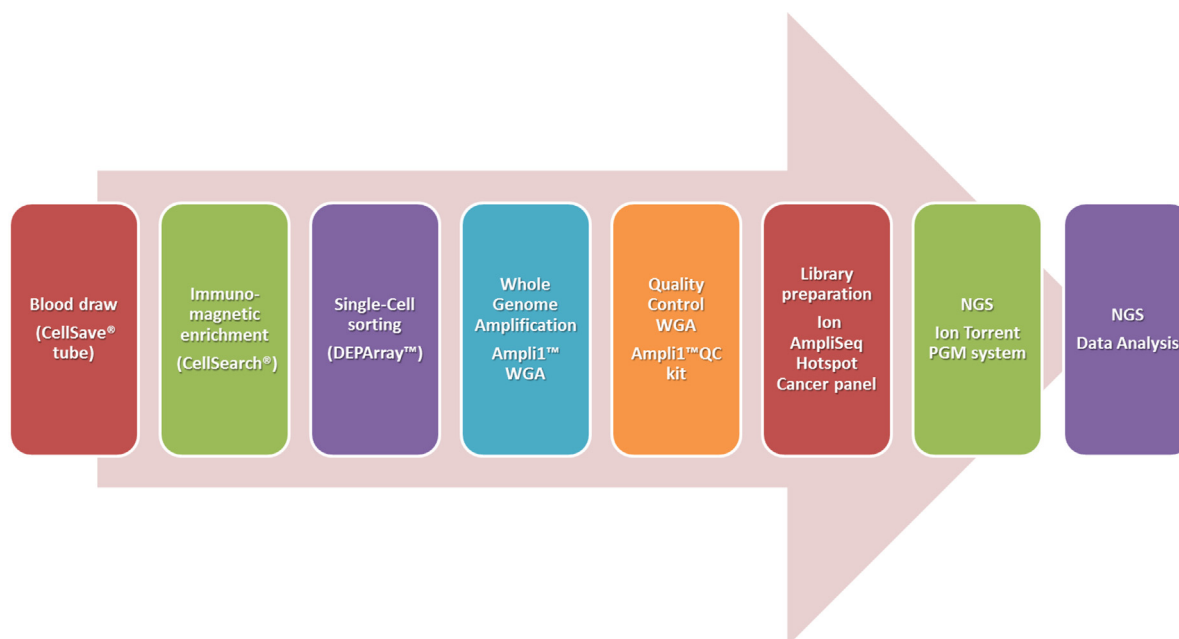
#### 3.1.3. Next generation sequencing on single and pooled cells

We successfully sequenced 5 single cells, a pool of 5 cells and DNA from a pellet of MDA-MB-231 cells with a mean depth of the sequencing reaction of 2966, 2862, 1581, 1884, 1854, 3479 and 2100 reads respectively.

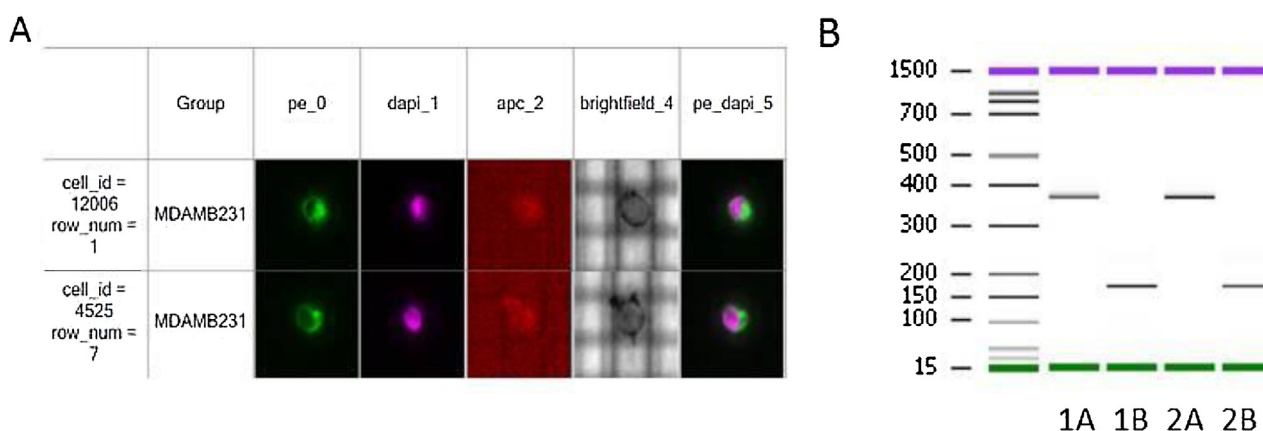
On the whole 27 sequence variants were detected in 18 genes (Table 2). Fifteen variants had been already reported in the COSMIC or dbSNP databases, while 12 were described for the first time.

Only two somatic mutations, p.G464V in exon 11 of the *BRAF* gene and p.R280K in exon 7 of the *TP53* gene, had been already described in this cells line, while the other mutations and single nucleotide polymorphisms (SNPs) have never been reported before.

Two somatic mutations (p.G464V in exon 11 of the *BRAF* gene and p.R280K in exon 7 of the *TP53* gene) and a SNP (p.P72R in exon 3 of the *TP53* gene), were detected in single and pooled cells as well as in the reference cellular pellet; three synonymous variants (p.Q787Q in *EGFR*, p.P567P in *PDGFRA* and p.L769L in *RET*) and the deleterious mutation p.N1161fs in *IDH1* were detected in at least two samples (single or pooled cells or in the reference pellet), while



**Fig. 1.** Flow chart representing the single steps of the proposed procedure for single cell analysis by NGS.



**Fig. 2.** (A) Cell staining pattern visualized by the DEPArray™: the cell show positive fluorescent signals for nuclear DAPI and cytokeratins (PE) and no signal for CD45 (APC). (B) Analysis of Ampli1™ QC PCR products for MDA-MB-231 single cell 1 and 2 by capillary electrophoresis. “A” = PCR product A; “B” = PCR product B.

18 variants were present only in a single cell sample (Table 2). Two variants (p.S614I in *ATM* and p.R927L in *ERBB4*) were found only in the pool (Table 2). All the variants detected in the cellular pellet were present in at least one single cell sample (Table 2).

In addition to the identified variants, Table 2 reports for all samples the number of forward and reverse reads of the mutated base and the homozygous or heterozygous status of the variant detected in each single cell.

Sequence variants common to the 6 samples were detected with a high number of reads, while a few variants found just in a single sample had a low number of reads of the mutated nucleotide.

Table 3 reports the mean coverage of each amplicon per sample. In addition, for samples with a detected sequence variant, we reported the coverage of the specific position.

#### 4. Discussion

We described a workflow for the molecular characterization of multiple genes in single cells by NGS. The study represents a proof-of-principle demonstrating the possibility to investigate the molecular status of about 50 genes related to cancer development

in a single tumor cell. The described pipeline can be easily transferred to the study of single circulating tumor cells from oncologic patients. In fact, we demonstrated the feasibility of deep sequencing on single and pooled pure MDA-MB-231 cells which represent a model of metastatic breast cancer cells.

One of the purposes of the study was to bypass some of the limitations of the CellSearch®, one of the most widely used systems for CTC enrichment, in the molecular analysis of patients samples. In fact, while the instrument was extensively evaluated for its ability to accurately count CTCs in blood and for the consequent prognostic performances [12–14], it is unable to provide pure samples to be submitted to molecular characterization: contaminating leucocytes are still present in the enriched samples and the low number of isolated CTCs substantially limits subsequent molecular analyses [15,16].

We already demonstrated the possibility to sequence by Sanger method pure CTC samples enriched by CellSearch® and sorted by DEPArray™ [11] and now we would like to extend the procedure to a higher throughput sequencing system like targeted NGS.

In all our samples we verified the presence of two expected somatic mutations detectable by the Ion AmpliSeq Cancer Hotspot

**Table 2**

List of the sequence variants detected in single and pooled MDA-MB-231 cells. White boxes represent wild type sequences, while grey boxes indicate the presence of the mutation. In single cell samples, “homo” and “het” indicate the presence of the mutation in homozygosis or heterozygosis respectively. For all samples, the number of forward and reverse reads of the mutated base has been reported for each sequence variant. Sequence variants with possible deleterious effect on the protein phenotype are written in bold, while benign variants are reported in plain text. Somatic mutations specifically investigated by the Ion AmpliSeq™ Cancer Hotspot Panel are highlighted in light grey.

Gene	exon	seq variant	COSMIC/ dbSNP	already reported	cell 1	cell 2	cell 3	cell 4	cell 5	pool 5 cells	pellet
ABL1	6	p.M337IM		no			het 176;123				
APC	15	p.P1458A	rs143796828	no			het 156;155				
ATM	11	p.S614I		no						40;31	
ATM	11	p.T616A	rs587780615	no		het 6;3					
BRAF	11	p.G464V	COSM450	yes	homo 1715;1288	homo 1081;916	homo 431;373	homo 550;387	homo 944;846	2091;1585	742;756
BRAF	15	p.S607P	COSM21544	no	het 19;20						
EGFR	20	p.Q787Q	COSM1451600	no			homo 378;540	homo 1008;1162	homo 549;615		634;836
ERBB4	23	p.R927L		no						35;26	
ERBB4	23	p.P942H		no		het 37;56					
FBXW7	9	p.H495N		no				het 33;26			
FLT3	14	p.E588Vfs		no				het 129;109			
IDH1	2	p.C114F		no	het 663;982						
IDH1	2	p.N116Ifs		no					het 378;385		99;101
KIT	10	p.L526M		no			het 42;43				
KRAS	3	p.R135G		no				het 26;22			
MLH1	12	p.E414E		no				het 142;186			
PDGFRA	11	p.P567P	COSM1430082	no			homo 2080;1155	homo 1444;1062	homo 1807;1346		964;747
PTEN	1	p.K6K	COSM1349460	no				het 103;151			
PTEN	1	p.D19Y	rs121909233	no		het 107;104					
RET	13	p.L769L	rs1800861	no				homo 80;65			976;822
SMARCB1	2	p.A55D		no			het 154;228				
SMARCB1	5	p.R201Efs	COSM1133055	no			het 228;221				
SMO	11	p.T640A	COSM13150	no				het 161;154			
TP53	3	p.P72R	rs1042522	no	homo 754;1563	homo 1304;2427	homo 858;1096	homo 1681;2182	homo 637;817	459;1050	1135;1460
TP53	4	p.S166L	COSM44289	no					homo 140;133		
TP53	7	p.R280K	COSM10728	yes	homo 1668;1808	homo 1533;1778	homo 1281;1240	homo 749;728	homo 882;791	2130;2487	865;989
TP53	9	p.G356W	COSM44830	no		het 184;108					

panel on the basis of what is reported for MDA-MB-231 in the COSMIC database, demonstrating that NGS fulfils the analytical features of specificity and accuracy required for single cell analysis.

With the exception of the above mentioned mutations and the p.P72R SNP in the *TP53* gene, we evidenced a high heterogeneity among different cells of the same cancer cell line: nineteen variants were in fact present in only one of the five analyzed single

cells. This finding modifies the idea of cell lines as a pool of isogenic and homogeneous cells, in fact cells in a culture may be subjected to divergent clonal evolution resulting in a heterogeneous mixture of different genomes. Moreover, all the variants detected in the cellular pellet were confirmed in at least one single cell indicating that single cells represent the different clones of the original population.

**Table 3**

List of the sequence variants detected in single and pooled MDA-MB-231 cells with details about the coverage of the next generation sequencing reaction. For single wild type cells we reported the mean coverage of the sequenced amplicon; while for mutated single or pooled cells we reported the mean coverage of the amplicon as well as the coverage of the specific mutated nucleotide position. When more than a cell is considered for calculation, the number of cells is reported in brackets.

Gene	Sequence variant	Single wild type cells	Single mutated cells		Pool 5 cells	
		Mean coverage	Mean coverage	Single base coverage	Mean coverage	Single base coverage
ABL1	p.M337IM	9965 (n=4)	5493	4718	8377	
APC	p.P1458A	7483 (n=4)	5721	4784	11545	
ATM	p.S614I	336 (n=5)			1101	985
ATM	p.T616A	372 (n=4)	191	176	1101	
BRAF	p.G464V		1937 (n=5)	1712 (n=5)	4363	3688
BRAF	p.S607P	549 (n=4)	808	739		
EGFR	p.Q787Q	6050 (n=2)	1664 (n=3)	1423 (n=3)		
ERBB4	p.R927L	558 (n=5)			1356	1200
ERBB4	p.P942H	538 (n=4)	640	559	1356	
FBXW7	p.H495N	2788 (n=4)	600	545	4475	
FLT3	p.E588Vfs	1932 (n=4)	1095	975	3502	
IDH1	p.C114F	9337 (n=4)	4856	3306	15573	
IDH1	p.N116Ifs	8012 (n=4)	10155	9003	15573	
KIT	p.L526M	1227 (n=4)	929	818	2405	
KRAS	p.R135G	597 (n=4)	416	350	1980	
MLH1	p.E414E	7435 (n=4)	4831	4222	14523	
PDGFRA	p.P567P	2586 (n=2)	3513 (n=3)	2973 (n=3)	3933	
PTEN	p.K6K	3823 (n=4)	3066	2837	7575	
PTEN	p.D19Y	3562 (n=4)	4109	3380	7575	
RET	p.L769L	52 (n=4)	167	150	45	
SMARCB1	p.A55DA	9550 (n=4)	6130	5279	10136	
SMARCB1	p.R201Efs	7089 (n=4)	4686	4055	6902	
SMO	p.T640A	3114 (n=4)	9737	8450	4748	
TP53	p.P72R		3321 (n=5)	2668 (n=5)	1989	1514
TP53	p.S166L	14711 (n=4)	6567	5564	17535	
TP53	p.R280K		2886 (n=5)	2508 (n=5)	5417	4670
TP53	p.G356W	3243 (n=4)	3998	3408	2897	

Previous studies showed that amplification and sequencing errors are a concern for single cell mutation analysis [17], but the punctual and mean high coverage reached for our samples make us confident on the reliability of even minor variants found only in single samples. In fact achieving high physical coverage of the targeted sequences is crucial for calling mutations at the same regions across multiple single cells Navine, [3].

As a matter of fact, we cannot exclude technical errors deriving from the WGA procedure. As already pointed out by other authors, WGA could affect subsequent sequencing results by introducing a number of technical variables such as allelic dropout, inadequate coverage, false positive and negative results [3]. For this reason we chose to adopt a WGA method which has been shown to reliably amplify the entire cellular genome homogeneously [18].

It is generally accepted that, due to high error rates in NGS, mutations must be detected in multiple single cells (at least 3 cells) to make a mutant call [3]. Nonetheless with the aim to report the entire results of the proposed workflow and to highlight cell heterogeneity, we decided to report all the variants detected in each single cell provided that the number of forward and reverse reads for the mutated base was balanced.

The feasibility of high-resolution single cell analysis is very important when applied to the study of CTCs which are extremely rare in the bloodstream and, due to the challenge in the development of isolation procedures, can be often recovered in low numbers from a single blood draw.

Single cell analysis may assume even higher relevance when dealing with samples from cancer patients in whom such methods could help elucidating tumor heterogeneity in a variety of specimens, such as primary tissues, metastases and circulating tumor cells. CTCs in particular are rare in the circulation and require highly sensitive and specific methods for their isolation and high throughput procedures for nucleic acid characterization in order to highlight their metastatic potential and their usefulness as dynamic markers of tumor development.

Our study is an example of optimization of single-cell analysis procedures which are still in development and hold promise to establish robust single-cell diagnostics which will be crucial for the realization of precision medicine implying novel therapeutic approaches [6].

### Conflict of interest

All the authors declare to have no conflict of interest.

### Acknowledgement

This work has received financial support from Ente Cassa di Risparmio di Firenze (Firenze, Italy) (2011.1035) and Regione Toscana (project "CYTOPEM" POR CRO FSE 2007-2013).

### References

- [1] N. Navin, J. Hicks, Future medical applications of single-cell sequencing in cancer, *Genome Med.* 3 (2011) 31.
- [2] C. Richards, How next-generation sequencing came to be, *Drug Target Rev.* 2 (2015) 36–37.
- [3] N.E. Navin, Cancer genomics: one cell at a time, *Genome Biol.* 15 (2014) 452.
- [4] E. Shapiro, T. Biezuner, S. Linnarsson, Single-cell sequencing-based technologies will revolutionize whole-organism science, *Nat. Rev. Genet.* 14 (2013) 618–630.
- [5] Y. Wang, J. Waters, M.L. Leung, A. Unruh, W. Roh, X. Shi, K. Chen, P. Scheet, S. Vattathil, H. Liang, A. Multani, H. Zhang, R. Zhao, F. Michor, F. Meric-Bernstam, N.E. Navin, Clonal evolution in breast cancer revealed by single nucleus genome sequencing, *Nature* 512 (7513) (2014) 155–160.
- [6] M.R. Speicher, Single-cell analysis: toward the clinic, *Genome Med.* 5 (2013) 74.
- [7] S. Domcke, R. Sinha, D.A. Levine, C. Sander, N. Schultz, Evaluating cell lines as tumour models by comparison of genomic profiles, *Nat. Commun.* 4 (2013) 2126.
- [8] K.B. Yin, The Mesenchymal-Like Phenotype of the MDA-MB-231 Cell Line, in: M. Gunduz Prof. (Ed.), *Breast Cancer—Focusing Tumor Microenvironment, Stem cells and Metastasis*, InTech., 2011, pp. 385–402.
- [9] P. McInerney, P. Adams, M.Z. Hadi, Error rate comparison during polymerase chain reaction by DNA polymerase, *Mol. Biol. Int.* 2014 (2014) 287430.

- [10] A.B. Fuchs, A. Romani, D. Freida, G. Medoro, M. Abonnenc, L. Altomare, I. Chartier, D. Guergour, C. Villiers, P.N. Marche, M. Tartagni, R. Guerrieri, F. Chatelain, N. Manaresi, Electronic sorting and recovery of single live cells from microlitre sized samples, *Lab. Chip* 6 (2006) 121–126.
- [11] M. Pestrin, F. Salvianti, F. Galardi, F. De Luca, N. Turner, L. Malorni, M. Pazzagli, A. Di Leo, P. Pinzani, Heterogeneity of PIK3CA mutational status at the single cell level in circulating tumor cells from metastatic breast cancer patients, *Mol. Oncol.* 9 (2015) 749–757.
- [12] M. Cristofanilli, D.F. Hayes, G.T. Budd, M.J. Ellis, A. Stopeck, J.M. Reuben, G.V. Doyle, J. Matera, W.J. Allard, M.C. Miller, H.A. Fritsche, G.N. Hortobagyi, L.W. Terstappen, Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer, *J. Clin. Oncol.* 23 (2005) 1420–1430.
- [13] S.J. Cohen, C.J. Punt, N. Iannotti, B.H. Savidman, K.D. Sabbath, N.Y. Gabrail, J. Picus, M. Morse, E. Mitchell, M.C. Miller, G.V. Doyle, H. Tissing, L.W. Terstappen, N.J. Meropol, Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer, *J. Clin. Oncol.* 26 (2008) 3213–3221.
- [14] J.S. de Bono, H.I. Scher, R.B. Montgomery, C. Parker, M.C. Miller, H. Tissing, G.V. Doyle, L.W. Terstappen, K.J. Pienta, D. Raghavan, Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer, *Clin. Cancer Res.* 14 (2008) 6302–6309.
- [15] A.M. Sieuwerts, J. Kraan, J. Bolt-de Vries, P. van der Spoel, B. Mostert, J.W. Martens, J.W. Gratama, S. Sleijfer, J.A. Foekens, Molecular characterization of circulating tumor cells in large quantities of contaminating leukocytes by a multiplex real-time PCR Breast, *Cancer Res. Treat.* 118 (2009) 455–468.
- [16] A.M. Sieuwerts, B. Mostert, J. Bolt-de Vries, D. Peeters, F.E. de Jongh, J.M. Stouthard, L.Y. Dirix, P.A. van Dam, A. Van Galen, V. de Weerd, J. Kraan, P. van der Spoel, R. Ramírez-Moreno, C.H. van Deurzen, M. Smid, J.X. Yu, J. Jiang, Y. Wang, J.W. Gratama, S. Sleijfer, J.A. Foekens, J.W. Martens, mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients, *Clin. Cancer Res.* 17 (2011) 3600–3618.
- [17] E. Heitzer, M. Auer, C. Gasch, M. Pichler, P. Ulz, E.M. Hoffmann, S. Lax, J. Waldspuehl-Geigl, O. Mauermann, C. Lackner, G. Höfler, F. Eisner, H. Sill, H. Samonigg, K. Pantel, S. Riethdorf, T. Bauernhofer, J.B. Geigl, M.R. Speicher, Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing, *Cancer Res.* 73 (2013) 2965–2975.
- [18] C.A. Klein, O. Schmidt-Kittler, J.A. Schardt, K. Pantel, M.R. Speicher, G. Riethmüller, Comparative genomic hybridization, loss of heterozygosity and DNA sequence analysis of single cells, *Proc. Natl. Acad. Sci. U. S. A.* 96 (8) (1999) 4494–4499.