Translational Oncology
www.transonc.com

Volume 8 Number 5 October 2015 pp. 407–416 407

Detection of Clonal and Subclonal Copy-Number Variants in Cell-Free DNA from Patients with Breast Cancer Using a Massively Multiplexed PCR Methodology

Eser Kirkizlar[1], Bernhard Zimmermann[1], Tudor Constantin, Ryan Swenerton, Bin Hoang, Nicholas Wayham, Joshua E. Babiarz, Zachary Demko, Robert J. Pelham, Stephanie Kareht, Alexander L. Simon, Kristine N. Jinnett, Matthew Rabinowitz, Styrmir Sigurjonsson and Matthew Hill

Natera Inc., 201 Industrial Road, Suite 410, San Carlos, CA 94070

Abstract

We demonstrate proof-of-concept for the use of massively multiplexed PCR and next-generation sequencing (mmPCR-NGS) to identify both clonal and subclonal copy-number variants (CNVs) in circulating tumor DNA. This is the first report of a targeted methodology for detection of CNVs in plasma.

Using an in vitro model of cell-free DNA, we show that mmPCR-NGS can accurately detect CNVs with average allelic imbalances as low as 0.5%, an improvement over previously reported whole-genome sequencing approaches. Our method revealed differences in the spectrum of CNVs detected in tumor tissue subsections and matching plasma samples from 11 patients with stage II breast cancer. Moreover, we showed that liquid biopsies are able to detect subclonal mutations that may be missed in tumor tissue biopsies. We anticipate that this mmPCR-NGS methodology will have broad applicability for the characterization, diagnosis, and therapeutic monitoring of CNV-enriched cancers, such as breast, ovarian, and lung cancer.

Translational Oncology (2015) 8, 407–416

Introduction

Cancer is a complex and dynamic disease. Tumors consist of diverse populations of genetically distinct subclones that may be mixed or spatially separated [1,2] and may have developed along different evolutionary paths [3]. Tumor genomes have a broad mutational landscape that includes single-nucleotide variations (SNVs), copy-number variations (CNVs), and/or chromosomal aberrations [4–12]. Genetic heterogeneity has been shown to occur in nearly all cancers, including breast cancer [1,13–15]. From a clinical perspective, it is critical to understand tumor heterogeneity as key clonal and subclonal mutations can indicate a reduction of the sensitivity of tumors to targeted treatments, potentially leading to drug resistance and ultimately metastasis [1,2,16–18].

Assessing the full mutational spectrum of heterogeneous tumors in a clinical setting can be difficult. First, the spatially limited nature of biopsies leads to genomic profiles that may not be representative of the entire tumor, and thus, biopsy data tends to underrepresent the full mutational spectrum of heterogeneous tumors. Second, some tumors are not readily accessible for biopsy. Finally, the invasive nature of biopsies makes them unsuitable for longitudinal analyses of treatment efficacy or to monitor disease progression [19]. Thus, novel methods that can more accurately assess the mutational landscape of tumors are urgently needed.

Liquid biopsies are an attractive alternative to tissue biopsies because they are less invasive and potentially more representative of a patient's overall tumor burden [20–22]. A number of proof-of-concept studies have used next-generation sequencing (NGS)-based methods to detect a wide range of cancer-associated alterations, including aneuploidies, CNVs, SNVs, and epigenetic alterations in plasma cell-free DNA (cfDNA) from both early- and late-stage tumors [23–31]. In addition, plasma DNA analysis has been used to

Address all correspondence to: Matthew Hill, PhD, Natera Inc., 201 Industrial Road, Suite 410, San Carlos, CA 94070, USA.
E-mail: mhill@natera.com
[1]These authors contributed equally to this work.
Received 4 June 2015; Revised 31 July 2015; Accepted 10 August 2015

1936-5233/15
http://dx.doi.org/10.1016/j.tranon.2015.08.004

detect subclonal point mutations in a patient with lung cancer [25] and to elucidate the genetic heterogeneity of both primary and metastatic tumors in a patient with breast cancer [32]. Several whole-genome sequencing (WGS) methods have been reported to detect CNVs in plasma [26,33,34], but these require a circulating tumor DNA (ctDNA) fraction of approximately 5% to achieve good sensitivity and specificity. That may not be low enough to detect subclonal CNVs or early-stage cancers. Furthermore, no tests using a targeted methodology to detect CNVs from liquid biopsies have been reported.

In this proof-of-concept study, we use a novel, targeted method for identifying both clonal and subclonal CNVs, with lower ctDNA fractions than previously reported. This methodology, which uses single-nucleotide polymorphism (SNP)-targeted massively multiplexed PCR (mmPCR) followed by NGS (hereafter referred to as "mmPCR-NGS"), was adapted from previously described approaches for SNP-based CNV detection in noninvasive prenatal screening for fetal aneuploidies and microdeletions [35–39].

We first validated our methodology in tumor samples with a commercially available SNP microarray. We then defined the analytical sensitivity of mmPCR-NGS using artificial plasma DNA mixtures made from matched germline and affected cell lines with known copy number variants. Finally, we applied mmPCR-NGS to plasma samples from eleven patients with stage II breast cancer (BC1–BC11), and analyzed the concordance between CNVs detected in plasma and those detected in tumor samples from each corresponding patient.

Materials and Methods

Samples

Human breast cancer cell lines (HCC1954 and HCC2218) and matched normal cell lines (HCC1954BL and HCC2218BL) were obtained from the American Type Culture Collection. Paired father and child 22q11.2 deletion syndrome cell lines (GM10383 and GM10382) were obtained from the Coriell Cell Repository (Camden, NJ). GM10382 cells are missing the maternal homolog of the 22q11.2 region (hg19 coordinates 17,256,415-19,795,835).

We procured fresh frozen (FF) tumor tissues and matched plasma samples from eleven patients with breast cancer from Geneticist (Glendale, CA) and North Shore-LIJ (Manhasset, NY). All patients signed institutional review board-approved informed-consent forms prior to sample collection. Patient demographics and tumor characteristics are shown in Table 1. Tumor subsections were resected for analysis. Each subsection weighed approximately 25 mg (~27 mm^3). Samples were first taken from the center and then in up to five additional distinct regions, such that the overall spatial pattern was the same for each sample. Prior to tumor resection, blood samples were collected into EDTA tubes. Within six hours, samples were centrifuged at ≥2000*g* for 20 minutes to isolate plasma.

Table 1. Breast Cancer Patient Demographics and Tumor Characteristics

Patient	Sex	Age	Race/Ethnicity	Histology	Grade	Stage
BC1	Female	56	Caucasian	Infiltrating ductal	2	II
BC2	Female	48	Caucasian	Infiltrating ductal	3	II
BC3	Female	74	Caucasian	Invasive ductal	3	II
BC4	Female	31	Asian	Invasive ductal	3	II
BC5	Female	49	Caucasian	Invasive ductal	3	II
BC6	Female	40	African American	Invasive ductal	2	II
BC7	Female	36	Ashkenazi Jew	Invasive ductal	3	II
BC8	Female	42	Caucasian	Invasive ductal	3	II
BC9	Female	43	Caucasian	Infiltrating ductal	3	II
BC10	Female	56	Caucasian	Infiltrating ductal	1	II
BC11	Female	34	Caucasian	Infiltrating ductal	3	II

Blood samples from 30 putative healthy, negative controls were also obtained, along with matched maternal and paternal buccal or blood samples using an institutional review board-approved protocol. Samples were collected into Streck BCT tubes and stored at room temperature. Within four days, plasma was isolated after centrifugation at 2000*g* for 20 minutes, and then at 3220*g* for 30 minutes. cfDNA was extracted from 1 to 4 ml plasma (~6 ng/ml) using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Valencia, CA).

Cell Culture

All cell culture reagents were from Life Technologies (Foster City, CA). American Type Culture Collection cells were cultured in 10% fetal bovine serum (FBS) RPMI 1640 (high glucose with pyruvate) with 2 mM L-glutamine at 37°C in a 5% CO_2 atmosphere. GM10382 cells were cultured in 15% FBS DMEM, and GM10383 cells were grown in 15% FBS RPMI 1640. Seed stocks were made after the first passage, and a maximum of five passages was performed.

Genomic DNA Isolation

Genomic DNA (gDNA) from FF tumor tissues, blood, and buccal samples was extracted using the DNeasy Blood and Tissue Kit (Qiagen).

Artificial Plasma DNA Model

Our "PlasmArt" methodology generates mixtures of DNA fragments with the size of mononucleosomal cfDNA fragments naturally found in plasma. First, 9×10^6 cells were lysed with hypotonic lysis buffer (20 mM Tris-Cl (pH 7.5), 10 mM NaCl, and 3 mM $MgCl_2$) for 15 minutes on ice. Then, 10% IGEPAL CA-630 (Sigma, St. Louis, MO) was added to a final concentration of 0.5%. After centrifugation at 3000*g* for 10 minutes at 4°C, pelleted nuclei were resuspended in 1X micrococcal nuclease (MNase) Buffer (New England BioLabs, Ipswich, MA) before adding 1000 U of MNase (New England BioLabs), and then incubating for 5 minutes at 37°C. Reactions were stopped by adding EDTA to a final concentration of 15 mM. Undigested chromatin was removed by centrifugation at 2000*g* for 1 min. Fragmented DNA was purified with the DNA Clean & Concentrator-500 kit (Zymo Research, Irvine, CA). Mononucleosomal DNA produced by MNase digestion was purified and size-selected using AMPure XP magnetic beads (Beckman Coulter, Brea, CA). DNA fragments were sized and quantified with a Bioanalyzer DNA 1000 chip (Agilent, Santa Clara, CA).

To model different ctDNA concentrations, we mixed different fractions of mononucleosomal DNA from HCC1954 and HCC2218 cancer cells with those from the corresponding matched normal cell line (HCC1954BL and HCC2218BL, respectively). Three samples were analyzed at each concentration. Similarly, to model allelic imbalances in plasma DNA in a focal 2.76 Mb region, we generated PlasmArts from DNA mixtures containing different ratios of DNA from a child with a maternal 22q11.2 deletion and from the father. Samples containing only the father's DNA were used as negative controls. Eight samples were analyzed at each concentration.

Massively Multiplexed PCR Primer Design

Massively multiplexed PCR allows simultaneous amplification of many targets in a single reaction. In this study, we targeted 3168

SNPs, which were distributed across five chromosome arms as follows: 646 on 1p, 602 on 1q, 541 on 2p, 707 on 2q, and 672 on the 22q11.2 focal region. These genomic regions were selected for convenience from SNP panels available in our laboratory. Target SNPs had at least 10% population minor allele frequency (1000 Genomes Project data; April 30, 2012 release) to ensure that a sufficient fraction would be heterozygous in any given patient. For each SNP, multiple primers were designed to have a maximum amplicon length of 75 bp and a melting temperature between 54.0 and 60.5°C, using Primer3Plus software [40]. To minimize the likelihood of primer-dimer product formation, primer interaction scores for all possible combinations of primers were calculated, and primers with high scores were eliminated. Candidate PCR assays were ranked and 3168 assays were selected on the basis of target SNP minor-allele frequency, observed heterozygosity rate (from dbSNP), presence in HapMap, and amplicon length.

mmPCR-NGS Workflow

The mmPCR-NGS workflow involved five steps: (1) library preparation, (2) mmPCR, (3) barcoding and pooling, (4) sequencing, and (5) sequence read alignment. First, a library was prepared by blunt-end repairing cfDNA fragments, tailing, and ligating to standard adapters. For mmPCR amplifications, we modified a previously described protocol [38]. Briefly, 3168 SNPs were amplified using one primer pair for each SNP. The mmPCR reaction contained Qiagen Multiplex PCR Master Mix, 70 mM tetramethylammonium chloride (Sigma), 6 μM primers, and 7 μl of library. Thermocycling conditions were 95°C for 15 min, followed by 25 cycles of 96°C for 30 seconds, 65°C for 20 seconds, and 72°C for 30 seconds. A final extension was performed at 72°C for 2 minutes, and then the reaction was cooled at 4°C. Next, sequencing tags and index sequences were added in a barcoding PCR reaction, as described previously [38]. Subsequently, the barcoded PCR products were pooled, purified with the QIAquick PCR Purification Kit (Qiagen), and quantified using the Qubit dsDNA BR Assay Kit (Life Technologies). Amplicons were then single-end sequenced using an Illumina HiSeq 2500 sequencer with 1.5 to 7 M reads/sample for tumor tissue DNA and 18 to 25 M reads/sample for plasma cfDNA. Finally, BWA-MEM [41] (version 0.7.10) and samtools mpileup [42] were used to align sequencing reads with the hg19 human reference genome and determine read base counts, respectively.

Definitions

For dimorphic SNPs that have alleles arbitrarily designated 'A' and 'B', the allele ratio of the A allele is $n_A/(n_A + n_B)$, where n_A and n_B are the number of sequencing reads for alleles A and B, respectively. Allelic imbalance is the difference between the allele ratios of A and B for loci that are heterozygous in the germline. This definition is analogous to that for SNVs, where the proportion of abnormal DNA is typically measured using mutant allele frequency, or $n_m/(n_m + n_r)$, where n_m and n_r are the number of sequencing reads for the mutant allele and the reference allele, respectively.

We report the proportion of abnormal DNA for a CNV (i.e., a variation in the number of copies of a DNA segment) by the average allelic imbalance (AAI) for SNPs in one haplotype of a DNA segment. Because tumor genomes often contain variable copy number in different regions, we prefer the use of the term AAI, which describes the proportion of abnormal genetic material in a sample, to the term "ctDNA fraction," which is usually calculated as the number of genomes from a tumor sample divided by the total number of genomes. ctDNA fraction may be ambiguous in cases with variable copy number; this is especially important in cancers, e.g. breast cancer, which frequently have multiple duplications in key regions [5].

Tumor Tissue Genomic DNA Analysis

In tumor tissue samples, CNVs were delineated by transitions between allele frequency distributions. Regions with at least 100 SNPs that had an allele ratio statistically different from 0.50 were considered to be of interest. More specifically, we focused on regions with average allele ratios of ≤0.45 or ≥0.55 for loci that were heterozygous in the germline. We used a segmentation algorithm to exhaustively search DNA sequences in the five aforementioned genomic regions for such regions, and iteratively selected them starting from the longest one until a limit of 100 SNPs was reached. Once a ≥100 SNP region was determined to contain a CNV, it was further segmented by average allelic ratios with a minimum segment size of 50 SNPs as appropriate.

FF tissue samples from three patients with breast cancer were also analyzed using Illumina CytoSNP-12 microarrays as previously described [43].

Circulating Tumor DNA Analysis

In plasma samples, CNVs were identified by a maximum-likelihood algorithm that searched for plasma CNVs in regions where the tumor sample from the same individual also had CNVs using haplotype information deduced from the tumor sample. In the negative control samples, haplotype information was deduced from parental genotypes.

The CNV detection algorithm models expected allelic frequencies across all allelic imbalance ratios at 0.025% intervals for three sets of hypotheses: (1) all cells are normal (no allelic imbalance), (2) some/all cells have a homolog 1 deletion or a homolog 2 amplification, or (3) some/all cells have a homolog 2 deletion or a homolog 1 amplification. The likelihood of each hypothesis was determined at each SNP using a Bayesian classifier based on expected and observed allele frequencies at all heterozygous SNPs, and then the joint likelihood across multiple SNPs was calculated. Finally, the hypothesis with the maximum likelihood was selected.

This algorithm also calculates the confidence of each CNV call by comparing the likelihoods of different hypotheses. A minimum confidence threshold of 99.9% was used in plasma samples from patients with cancer to minimize false-positive results.

Statistical Analyses

A linear regression model was used to compare either expected AAI or tumor input DNA percentage and observed AAI determined by the CNV detection algorithm. $P < .05$ was considered statistically significant. SigmaPlot 12.5 (Systat Software, San Jose, CA) and Matlab 7.12.0 R2011.a (MathWorks, Natick, MA) were used.

Results

Validation of mmPCR-NGS Assay

To assess the ability of mmPCR-NGS to detect CNVs in tumor tissue samples, we first validated our CNV detection algorithm by comparing CNVs detected in three breast cancer samples using mmPCR-NGS with those identified by the Illumina CytoSNP-12

microarray. We observed a strong linear correlation (R^2 = 0.974) between these two different methods (Supplementary Figure 1), which confirmed the accuracy of our methodology.

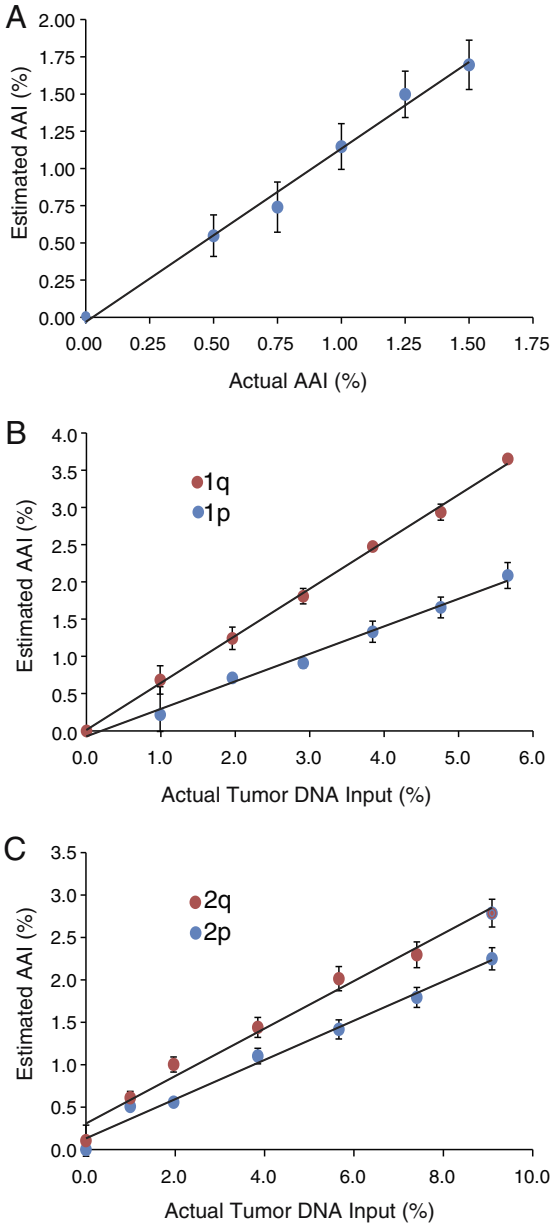

Figure 1. Determination of sensitivity and reproducibility of mmPCR-NGS to detect copy-number variants (CNVs) using artificial cfDNA mixtures ("PlasmArts"). (A) Correlation between actual and estimated average allelic imbalance (AAI) in PlasmArt samples with DNA from GM10382 cells with a 22q11.2 deletion and matched normal cells. (B) Correlation between estimated AAI and actual tumor DNA input in PlasmArt samples with DNA from HCC1954 breast cancer cells with chromosome 1p and 1q CNVs and matched normal HCC1954BL cells, containing 0% to 5.66% tumor DNA fractions. (C) Correlation between estimated AAI and actual tumor DNA input in PlasmArt samples with DNA from HCC2218 breast cancer cells with chromosome 2p and 2q CNVs and matched normal HCC2218BL cells, containing 0% to 9.09% tumor DNA fractions. Data points and error bars indicate the mean and standard deviation (SD), respectively, of 3 to 8 replicates.

Sensitivity of mmPCR-NGS

The sensitivity and reproducibility of mmPCR-NGS to detect CNVs were evaluated using mixtures of abnormal DNA titrated into samples with matched normal DNA. These artificial cfDNA mixtures, which we call "PlasmArts", contain DNA fragments with a size distribution approximating that of natural mononucleosomal cfDNA (see Materials and Methods).

A DNA sample having a deletion of the 22q11.2 region was titrated with 0% to 3.0% input (equivalent to 0% to 1.5% AAI) into a matched normal sample (Figure 1*A*). For CNVs in this region, we observed good agreement between actual and estimated AAIs (Table 2). The algorithm found CNVs of at least 0.45% AAI in all positive control samples (minimum expected AAI of 0.5%), 0% AAI (no CNV) in 7/8 (88%) of the negative control samples, and 0.05% AAI in the remaining negative sample. The estimated AAI values showed high linearity (R^2 = 0.940) and reproducibility (error variance = 0.026%). These results implied that AAIs as low as 0.5% could be detected and accurately quantified.

Two additional PlasmArt titrations of normal and matched tumor DNA, with CNVs on chromosome 1 or chromosome 2, were also evaluated (Figure 1, *B* and *C*). High linearity (R^2 = 0.952 for HCC1954 1p, 0.993 for HCC1954 1q, 0.977 for HCC2218 2p, and 0.968 for HCC2218 2q) and reproducibility (error variance = 0.027% for HCC1954 1p, 0.012% for HCC1954 1q, 0.014% for HCC2218 2p, and 0.029% for HCC2218 2q) were observed between the known input DNA amounts and those calculated by mmPCR-NGS. A comparison of tumor DNA input vs. estimated AAIs for these CNVs is shown in Table 2. The algorithm found 0%

Table 2. Comparison of Actual and Estimated AAIs of CNVs in PlasmArts Experiments

Genomic Region	Input Fraction† (%)	Actual AAI (%)	Estimated AAI (%)		
			Mean	95% CI	*n*
22q11.2	1.00	0.50	0.55	0.45-0.65	8
	1.50	0.75	0.74	0.62-0.86	8
	2.00	1.00	1.15	1.04-1.26	8
	2.50	1.25	1.50	1.39-1.61	8
	3.00	1.50	1.70	1.58-1.81	8
1p	0.99	Unk	0.22	0-0.65	3
	1.96	Unk	0.71	0.69-0.73	3
	2.91	Unk	0.91	0.84-0.98	3
	3.85	Unk	1.33	1.17-1.49	3
	4.76	Unk	1.66	1.49-1.82	3
	5.66	Unk	2.08	1.88-2.29	3
1q	0.99	Unk	0.68	0.46-0.90	3
	1.96	Unk	1.24	1.07-1.42	3
	2.91	Unk	1.81	1.69-1.93	3
	3.85	Unk	2.48	2.42-2.53	3
	4.76	Unk	2.94	2.81-3.06	3
	5.66	Unk	3.65	3.60-3.70	3
2p	0.99	Unk	0.51	0.46-0.55	3
	1.96	Unk	0.56	0.51-0.60	3
	3.85	Unk	1.10	1.00-1.21	3
	5.66	Unk	1.42	1.29-1.55	3
	7.41	Unk	1.79	1.66-1.93	3
	9.09	Unk	2.25	2.10-2.40	3
2q	0.99	Unk	0.61	0.52-0.70	3
	1.96	Unk	1.00	0.90-1.11	3
	3.85	Unk	1.44	1.31-1.58	3
	5.66	Unk	2.02	1.85-2.18	3
	7.41	Unk	2.30	2.12-2.47	3
	9.09	Unk	2.79	2.60-2.98	3

Abbreviations: CI, confidence interval; Unk, unknown due to tumor heterogeneity.
† Defined as the proportion of nuclei originating from abnormal cells.

Table 3. CNVs Detected in Tumor Subsections and Plasma from Patients with Breast Cancer

BC 1	1p	1q	2p	2q	22q11.2	BC 2	1p	1q	2p	2q	22q11.2
Tumor						Tumor					
Plasma						Plasma					
BC 3	1p	1q	2p	2q	22q11.2	BC 4	1p	1q	2p	2q	22q11.2
Tumor						Tumor					
Plasma						Plasma					
BC 5	1p	1q	2p	2q	22q11.2	BC 6	1p	1q	2p	2q	22q11.2
Tumor						Tumor					
Plasma						Plasma					
BC 7	1p	1q	2p	2q	22q11.2						
Tumor											
Plasma											
BC 8	1p	1q	2p	2q	22q11.2	BC 9	1p	1q	2p	2q	22q11.2
Subsection 1						Subsection 1					
Subsection 2						Subsection 2					
Subsection 3						Subsection 3					
Subsection 4						Subsection 4					
Plasma						Plasma					
BC 10	1p	1q	2p	2q	22q11.2	BC 11	1p	1q	2p	2q	22q11.2
Subsection 1						Subsection 1					
Subsection 2						Subsection 2					
Subsection 3						Subsection 3					
Subsection 4						Subsection 4					
Subsection 5						Subsection 5					
Subsection 6						Subsection 6					
Plasma						Plasma					

Black and gray boxes indicate CNV-positive plasma and tumor samples, respectively. Unshaded boxes indicate CNV-negative samples. BC, breast cancer.

AAI (no CNV) in 11/12 (92%) of the negative control samples, and 0.32% AAI in the remaining negative sample.

Based on these titration experiments, we chose to use a conservative AAI threshold of 0.45% to call CNVs in plasma samples. Together, these findings suggested that mmPCR-NGS would be able to detect CNVs with high sensitivity and reproducibility in actual samples from patients with cancer.

CNV Analyses of Tumor Tissue Samples

We first analyzed a single tissue subsection from a tumor from each of seven patients with breast cancer (BC1 to BC7) using mmPCR-NGS. All seven subsections had a CNV detected in at least one of the five targeted genomic regions (1p, 1q, 2p, 2q, and 22q11.2; Table 3, Supplemental Table 1).

We then analyzed four to six tumor subsections from each of four additional patients with breast cancer (BC8 to BC11). Similar to what was observed with patients BC1 to BC7, all subsections from patients BC8 to BC11 had a CNV detected in at least one of the five targeted genomic regions (Table 3, Supplemental Table 1). A CNV was identified in at least one tumor subsection in 18/20 (90%) genomic regions. Among these 18 CNV-positive regions, 11 (61%) had a CNV detected in that particular region in all subsections.

Different patterns of AAIs across these five chromosomal regions were also observed among different tumor subsections (Figure 2 and Supplementary Figure 2). In patient BC8 (Figure 2), for example, a similar pattern of CNVs was observed for regions 2p, 2q, and 22q11.2 in all four subsections, suggesting that these CNVs are clonal mutations. In contrast, only two of the four subsections for this patient had CNVs observed in the 1p region, and three of the four subsections had CNVs observed in the 1q region, suggesting that those CNVs were subclonal mutations. Similar patterns of possible clonal and subclonal CNVs were observed in patients BC10 and BC11, whereas for patient BC9, CNVs appeared to be more homogenous (Figure 2).

In addition, even when a CNV was detected in all subsections for a particular patient, such as in the 1q region for patient BC10 (Figure 2), the AAI often varied between subsections. Overall AAI patterns also differed between patients. Taken together, these findings suggest that mmPCR-NGS can be used to elucidate both intra- and inter-tumor clonal heterogeneity.

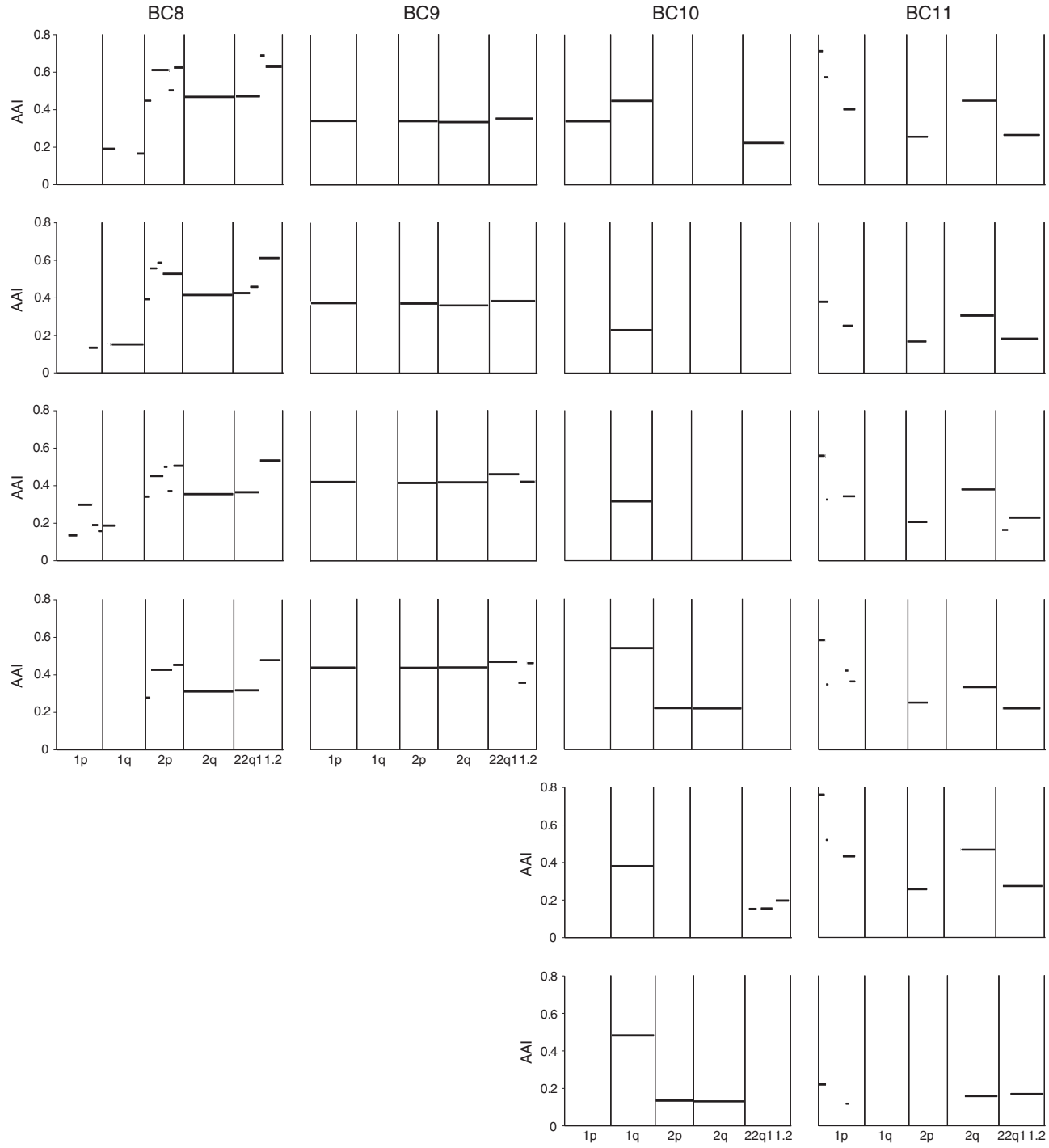

Figure 2. Detection of CNVs in four tumor tissue subsections from breast cancer patients BC8 and BC9, and six subsections from patients BC10 and BC11. Each plot shows the sequential location of the 3168 SNPs in our panel as they map to the genome (*x*-axis) and the mean AAI (*y*-axis) of each detected CNV (horizontal line segment). Vertical lines delineate genomic regions.

Concordance of CNVs in Tumors and Plasma cfDNA

To quantify the amount of overlap between CNVs detected in plasma cfDNA and those detected in tumor tissue gDNA, we used mmPCR-NGS to interrogate the plasma samples corresponding to the eleven tumor tissue samples. In patients BC1 to BC7, 9/21 (43%) CNV-positive genomic regions identified in tumors were also detected in the plasma (0.81% to 5.26% AAI) (Table 3, Supplemental Table 2). No CNVs were detected in any of the five targeted genomic regions in 2/7 (29%) patients (BC6 and BC7; Table 3).

As expected, we found a similar degree of overlap between tumor and plasma samples from patients BC8 to BC11. Specifically, 7/18 (39%) CNV-positive genomic regions identified in tumor subsections were also detected in the plasma (0.77% to 5.80% AAI) (Figure 3, Table 3, Supplemental Table 2). One of these four patients (25%; BC10) did not

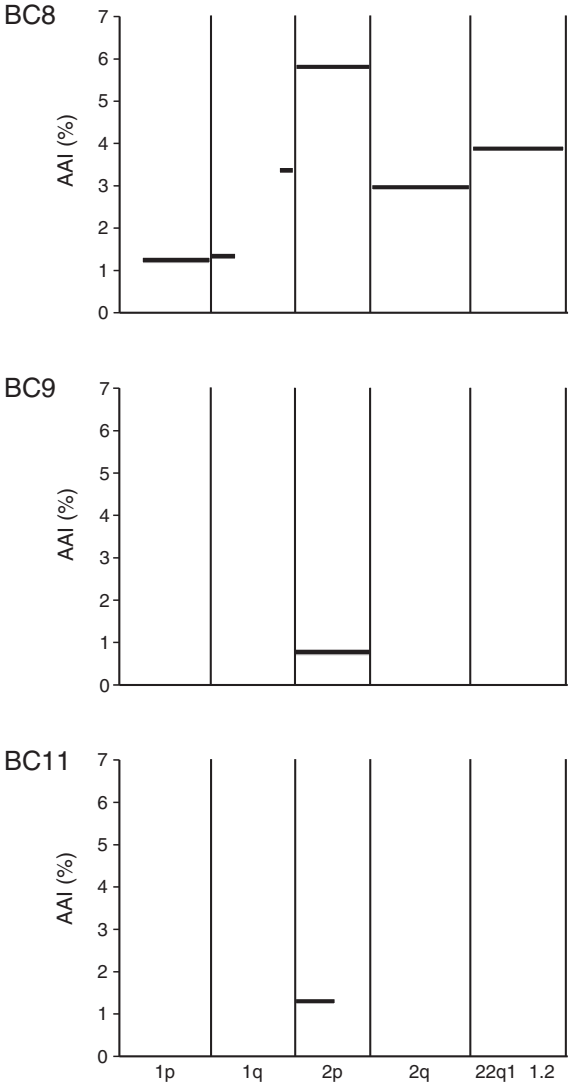

Figure 3. Detection of CNVs in plasma samples from breast cancer patients BC8, BC9, and BC11. Plots show sequential location of the 3168 SNPs in our panel as they map to the genome (*x*-axis) and the mean percentage AAI (*y*-axis) of each detected CNV (horizontal line segment). Vertical lines delineate genomic regions.

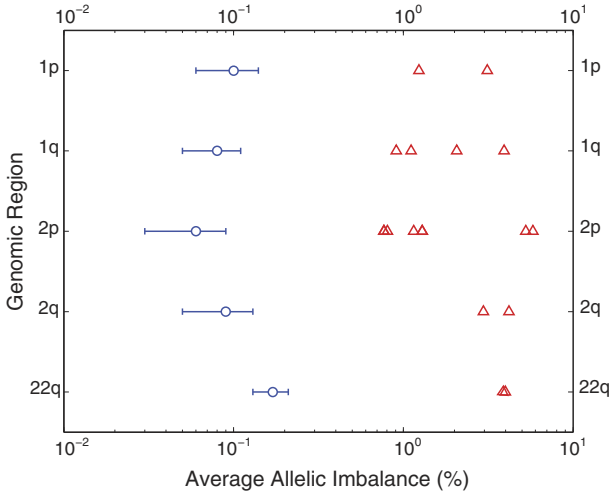

Figure 4. Comparison of estimated AAIs of CNVs detected in plasma samples from negative controls and patients with early-stage breast cancer, using mmPCR-NGS. (*left*) Blue circles and error bars show the mean AAI and two standard errors (*x*-axis; log scale), respectively, of CNVs in five genomic regions (*y*-axis) from 30 putative healthy individuals according to the maximum likelihood hypothesis. (*right*) Red triangles indicate the AAIs of detected CNVs in the same genomic regions from eleven patients with stage II breast cancer.

have any CNVs detected in plasma in any of the targeted chromosomal regions (Table 3).

We further compared the concordance between clonal and subclonal CNVs in tumor tissue subsections and CNVs in matching plasma samples from patients BC8 to BC11. Considering only the 11 clonal CNVs—those that were detected in all tumor subsections—a CNV was detected in four (36%) of the patient-matched plasma samples (estimated AAI: 0.77% to 5.80%). Among the seven subclonal CNVs—those that were not observed in all subsections—we detected a CNV in 3/7 (43%) of the regions (estimated AAI: 1.24% to 3.36%) in the corresponding cfDNA. Of note, in these three regions (BC1, chromosome 1p; BC1, chromosome 1q; and BC4, chromosome 2p), a CNV was detected in 10/14 (71%) of the matched tumor subsections. In contrast, in the other four genomic regions that did not have a CNV detected in the corresponding plasma samples (BC3, chromosomal regions 1p, 2p, 2q, and 22q11.2), we only detected a CNV in 7/24 (29%) of the tissue subsections. These data suggest that the more prevalent a subclonal CNV is within a tumor, the more likely it is to be observed in cfDNA.

In the 150 genomic regions assayed in 30 negative controls, there were no CNVs with AAIs >0.45% and confidence >99.9%. Furthermore, these AAIs were significantly lower than those from regions with CNVs detected in patients with breast cancer (Figure 4). Together, these results suggest that mmPCR-NGS has a low false-positive rate.

Discussion

Despite significant advances in targeted cancer therapies, the majority of patients with cancer still develop resistance due to emergence of new subclonal mutations. Improved control of tumor growth and spread requires methods that are sufficiently sensitive to identify clinically actionable mutations as early as possible, preferably before metastasis occurs. Currently, most clinical methods for cfDNA analysis focus on SNVs [23,24]. However, to maximize detection and characterization of early-stage tumors, methods that interrogate other mutation types, such as CNVs, should be incorporated into cancer tests. To our knowledge, no methods with sufficient analytical sensitivity to enable detection of CNVs in early-stage tumors have been described.

Here we describe a novel methodology to detect CNVs in plasma that has substantially higher sensitivity than any previously reported, allowing us to detect clonal and subclonal mutations in both tissue and plasma from patients with stage II breast cancer. We anticipate that the ability to detect both SNVs and CNVs in plasma samples with low tumor fraction will provide greater genomic coverage and thereby facilitate early detection of cancer and improvement of therapeutic strategies [1,2,16–18].

Unlike traditional tumor biopsies, which require sufficiently large tumors to sample and are often costly and difficult to obtain due to

their invasive nature, liquid biopsies are more amenable to collection at all tumor stages and at multiple time points. Two hurdles, however, have prevented the clinical application of liquid biopsies for CNV detection. First, current methods have not been sufficiently sensitive to detect CNVs in plasma when levels of ctDNA are low, which is typical of patients with early-stage disease or subclonal mutations [20,22,27]. Although the lower limit of CNV detection using WGS has not been rigorously defined, three studies reported that a ctDNA fraction of approximately 3.7% to 5.0% ctDNA fraction (equivalent to 2.2% to 3.7% AAI[2]) was necessary to achieve clinically relevant specificities [26,33,34]. Given that our method can detect CNVs with AAIs as low as 0.5% with excellent specificity, our methodology is at least five times more sensitive than WGS. Furthermore, because CNVs often comprise multiple duplication events, our method may be able to detect tumors from plasma with substantially lower ctDNA fraction. In this study, we were able to detect CNVs in plasma from 8/11 (72.7%) patients with stage II breast cancer, suggesting that this methodology may enable earlier detection of CNVs from liquid biopsies, a point when successful treatment is more likely.

Second, the clinical use of liquid biopsies has been hindered by lack of understanding about the concordance between clonal and subclonal tumor mutations and those observed in ctDNA [44]. Analysis of matched tumor and plasma samples using the mmPCR-NGS methodology could help address this knowledge gap. In this study, we detected CNVs in plasma that were not observed in all corresponding tissue subsections, suggesting that a liquid biopsy may be more representative of tumor heterogeneity than single biopsies. However, in other cases, we did not detect CNVs in the plasma that were observed in some or all of the corresponding tissue subsections, indicating that a liquid biopsy may not be able to detect all tumor CNVs. It is not clear whether this is due to the limit of detection of mmPCR-NGS, or the lack of representation of DNA from tumor-tissue subsections in plasma.

Studies using somatic SNVs as markers have also noted partial concordance between ctDNA and tumor tissue [45,46]. The incomplete overlap is consistent with the idea that plasma ctDNA and tumor biopsies are two different ways the mutational landscape of a tumor can be sampled, and consequently they may capture different mutational profiles. Although both approaches may provide clinically useful information, ctDNA tends to be enriched for mutations that develop early in tumor evolution [44,46] and may therefore be more actionable [2].

Despite its ability to detect subclonal mutations, the mmPCR-NGS methodology described here has two main limitations. First, our methodology requires knowledge of haplotype information, which is necessary to identify CNVs with very low AAI values in ctDNA; in this study, for the patients with cancer, we deduced haplotype information from analysis of tumor tissue, or for the negative controls, from parental genotypic information. However, the ability to detect early-stage disease (i.e., before tumors are large enough to be biopsied) using liquid biopsies necessitates *de novo* haplotype phasing, which could be achieved by other methods [47], such as haplotyping by dilution [48] or long-read sequencing [49].

Second, although mmPCR-NGS can determine the absence or presence of a CNV at low AAI, it is not able to differentiate between copy number gains and losses at low AAI. However, discrimination of duplications from deletions may not be necessary for some clinical applications. Furthermore, because many cancers tend to display either deletions or duplications in specific genomic regions, but not both, knowledge of CNVs in conjunction with single-nucleotide variant (SNV) markers may be sufficient to fingerprint specific cancers [11]. Although the five genomic regions assayed in this study were not optimized for breast cancer diagnosis, they do suggest that recurrent CNV patterns can be detected in limited genomic analyses. For example, in this study, plasma ctDNA analysis detected 2p CNVs in 6/11 (55%) of breast tumors, but identified 1p CNVs in only 2/11 (18%) of breast tumors. Analysis of AAI patterns in ctDNA—analogous to those we performed in tumor-tissue subsections—may also provide more detailed insights into the clonal architecture of tumors to help predict their therapeutic responses and optimize treatment strategies. Therefore, we are currently developing mmPCR-NGS panels to detect clinically actionable CNVs and SNVs. Such panels would be particularly useful for patients with cancers where CNVs represent a substantial proportion of the mutation load, as is common in breast, ovarian, and lung cancer [5].

Conclusions

We have developed a CNV detection methodology, mmPCR-NGS, that can be used to detect CNVs in plasma with better sensitivity than previously described methods. This methodology permits detection of subclonal CNVs in plasma samples from patients with cancer that a single tumor biopsy might miss. We anticipate that future optimization of mmPCR-NGS will result in a rapid and cost-effective diagnostic for CNV-enriched cancers, and that is likely to facilitate personalized cancer medicine.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.tranon.2015.08.004.

Conflict of Interest Statement

All authors are employees of Natera, Inc. and hold stock or options to hold stock in the company.

Statement of Author Contributions

EK, BZ, SS, and MH conceived and designed the study. RS, BH, NW, JB, and TC designed, performed, and analyzed experiments. MR helped develop the CNV detection algorithm. BZ, EK, and KNJ analyzed the data, and EK did all computational experiments. KNJ, ZD, SK, and ALS wrote the manuscript. BZ, SS, RJP, ZD, and MH supervised the project.

Acknowledgements

This study was funded by Natera, Inc. We thank Dr. Jan van Turnout (Natera, Inc.) for reviewing the manuscript and providing helpful comments.

References

[1] Meacham CE and Morrison SJ (2013). Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337. http://dx.doi.org/10.1038/nature12624.

[2] Leary et al. (2012) estimated that ~5% ctDNA fraction is necessary to achieve 99% specificity. AAI is approximately one-half the ctDNA fraction for tumors that contain a deletion or a single duplication, and less than half the ctDNA fraction for tumors where there are multiple duplications of a target region. Therefore, 2.5% AAI represents a lower bound on the sensitivity of this method for clinically relevant specificities. Heidary et al. (2014) estimated 3.7% ctDNA fraction by calculating the mean mutant allele frequency of eight SNVs in three plasma samples, which they also used for CNV analyses. Since mutant allele frequency maps to AAI in a 1:1 manner, their result is also an indirect estimate of the lower bound of the AAIs for the CNVs they detected. Chan et al. (2013) determined the fractional concentration of tumor-derived DNA defined as (A-B)/A, where A and B are the number of sequencing reads from each of the two alleles at SNPs from a genomic region showing loss of heterozygosity in the tumor, and B is the number of reads from the allele on the homolog that was deleted. The AAI corresponds to (A-B)/(A+B), which gives a lower bound of 2.2% AAI for the sample with the lowest ctDNA fraction (4.3%) for which a deletion was detected in the plasma.

[2] Swanton C (2012). Intratumor heterogeneity: evolution through space and time. *Cancer Res* **72**, 4875–4882. http://dx.doi.org/10.1158/0008-5472.can-12-2217.

[3] Gerlinger M, McGranahan N, Dewhurst SM, Burrell RA, Tomlinson I, and Swanton C (2014). Cancer: evolution within a lifetime. *Annu Rev Genet* **48**, 215–236. http://dx.doi.org/10.1146/annurev-genet-120213-092314.

[4] Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz Jr LA, and Kinzler KW (2013). Cancer genome landscapes. *Science* **339**, 1546–1558. http://dx.doi.org/10.1126/science.1235122.

[5] Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, and Sander C (2013). Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* **45**, 1127–1133. http://dx.doi.org/10.1038/ng.2762.

[6] Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang CZ, Wala J, and Mermel CH, et al (2013). Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45**, 1134–1140. http://dx.doi.org/10.1038/ng.2760.

[7] Kim TM, Xi R, Luquette LJ, Park RW, Johnson MD, and Park PJ (2013). Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Res* **23**, 217–227. http://dx.doi.org/10.1101/gr.140301.112.

[8] The Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337. http://dx.doi.org/10.1038/nature11252.

[9] The Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70. http://dx.doi.org/10.1038/nature11412.

[10] Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, and Bignell GR, et al (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404. http://dx.doi.org/10.1038/nature11017.

[11] Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, and Urashima M, et al (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905. http://dx.doi.org/10.1038/nature08822.

[12] Stratton MR, Campbell PJ, and Futreal PA (2009). The cancer genome. *Nature* **458**, 719–724. http://dx.doi.org/10.1038/nature07943.

[13] Russnes HG, Navin N, Hicks J, and Borresen-Dale AL (2011). Insight into the heterogeneity of breast cancer through next-generation sequencing. *J Clin Invest* **121**, 3810–3818. http://dx.doi.org/10.1172/JCI57088.

[14] Fisher R, Pusztai L, and Swanton C (2013). Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer* **108**, 479–485. http://dx.doi.org/10.1038/bjc.2012.581.

[15] Jamal-Hanjani M, Quezada SA, Larkin J, and Swanton C (2015). Translational implications of tumor heterogeneity. *Clin Cancer Res* **21**, 1258–1266. http://dx.doi.org/10.1158/1078-0432.CCR-14-1429.

[16] Seoane J and De Mattos-Arruda L (2014). The challenge of intratumour heterogeneity in precision medicine. *J Intern Med* **276**, 41–51. http://dx.doi.org/10.1111/joim.12240.

[17] Burrell RA, McGranahan N, Bartek J, and Swanton C (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345. http://dx.doi.org/10.1038/nature12625.

[18] Barber LJ, Davies MN, and Gerlinger M (2014). Dissecting cancer evolution at the macro-heterogeneity and micro-heterogeneity scale. *Curr Opin Genet Dev* **30C**, 1–6. http://dx.doi.org/10.1016/j.gde.2014.12.001.

[19] Bedard PL, Hansen AR, Ratain MJ, and Siu LL (2013). Tumour heterogeneity in the clinic. *Nature* **501**, 355–364. http://dx.doi.org/10.1038/nature12627.

[20] Diaz Jr LA and Bardelli A (2014). Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol* **32**, 579–586. http://dx.doi.org/10.1200/JCO.2012.45.2011.

[21] De Mattos-Arruda L, Cortes J, Santarpia L, Vivancos A, Tabernero J, Reis-Filho JS, and Seoane J (2013). Circulating tumour cells and cell-free DNA as tools for managing breast cancer. *Nat Rev Clin Oncol* **10**, 377–389. http://dx.doi.org/10.1038/nrclinonc.2013.80.

[22] Bidard FC, Weigelt B, and Reis-Filho JS (2013). Going with the flow: from circulating tumor cells to DNA. *Sci Transl Med* **5**, 207ps214. http://dx.doi.org/10.1126/scitranslmed.3006305.

[23] Lebofsky R, Decraene C, Bernard V, Kamal M, Blin A, Leroy Q, Rio Frio T, Pierron G, Callens C, and Bieche I, et al (2015). Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. *Mol Oncol* **9**, 783–790. http://dx.doi.org/10.1016/j.molonc.2014.12.003.

[24] Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, and Alani RM, et al (2014). Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224. http://dx.doi.org/10.1126/scitranslmed.3007094.

[25] Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, Liu CL, Neal JW, Wakelee HA, and Merritt RE, et al (2014). An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* **20**, 548–554. http://dx.doi.org/10.1038/nm.3519.

[26] Chan KC, Jiang P, Zheng YW, Liao GJ, Sun H, Wong J, Siu SS, Chan WC, Chan SL, and Chan AT, et al (2013). Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin Chem* **59**, 211–224. http://dx.doi.org/10.1373/clinchem.2012.196014.

[27] Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, and Mahler-Araujo B, et al (2013). Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **368**, 1199–1209. http://dx.doi.org/10.1056/NEJMoa1213261.

[28] Heitzer E, Ulz P, Belic J, Gutschi S, Quehenberger F, Fischereder K, Benezeder T, Auer M, Pischler C, and Mannweiler S, et al (2013). Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med* **5**, 30. http://dx.doi.org/10.1186/gm434.

[29] Kukita Y, Uchida J, Oba S, Nishino K, Kumagai T, Taniguchi K, Okuyama T, Imamura F, and Kato K, et al (2013). Quantitative identification of mutant alleles derived from lung cancer in plasma cell-free DNA via anomaly detection using deep sequencing data. *PLoS One* **8**e81468. http://dx.doi.org/10.1371/journal.pone.0081468.

[30] Shaw JA, Page K, Blighe K, Hava N, Guttery D, Ward B, Brown J, Ruangpratheep C, Stebbing J, and Payne R, et al (2012). Genomic analysis of circulating cell-free DNA infers breast cancer dormancy. *Genome Res* **22**, 220–231. http://dx.doi.org/10.1101/gr.123497.111.

[31] Narayan A, Carriero NJ, Gettinger SN, Kluytenaar J, Kozak KR, Yock TI, Muscato NE, Ugarelli P, Decker RH, and Patel AA, et al (2012). Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing. *Cancer Res* **72**, 3492–3498. http://dx.doi.org/10.1158/0008-5472.can-11-4037.

[32] De Mattos-Arruda L, Weigelt B, Cortes J, Won HH, Ng CK, Nuciforo P, Bidard FC, Aura C, Saura C, and Peg V, et al (2014). Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann Oncol* **25**, 1729–1735. http://dx.doi.org/10.1093/annonc/mdu239.

[33] Heidary M, Auer M, Ulz P, Heitzer E, Petru E, Gasch C, Riethdorf S, Mauermann O, Lafer I, and Pristauz G, et al (2014). The dynamic range of circulating tumor DNA in metastatic breast cancer. *Breast Cancer Res* **16**, 421. http://dx.doi.org/10.1186/s13058-014-0421-y.

[34] Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, O'Shaughnessy J, Kinzler KW, Parmigiani G, and Vogelstein B, et al (2012). Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* **4**, 162ra154. http://dx.doi.org/10.1126/scitranslmed.3004742.

[35] Pergament E, Cuckle H, Zimmermann B, Banjevic M, Sigurjonsson S, Ryan A, Hall MP, Dodd M, Lacroute P, and Stosic M, et al (2014). Single-nucleotide polymorphism-based noninvasive prenatal screening in a high-risk and low-risk cohort. *Obstet Gynecol* **124**, 210–218. http://dx.doi.org/10.1097/AOG.0000000000000363.

[36] Samango-Sprouse C, Banjevic M, Ryan A, Sigurjonsson S, Zimmermann B, Hill M, Hall MP, Westemeyer M, Saucier J, and Demko Z, et al (2013). SNP-based non-invasive prenatal testing detects sex chromosome aneuploidies with high accuracy. *Prenat Diagn* **33**, 643–649. http://dx.doi.org/10.1002/pd.4159.

[37] Wapner RJ, Babiarz JE, Levy B, Stosic M, Zimmermann B, Sigurjonsson S, Wayham N, Ryan A, Banjevic M, and Lacroute P, et al (2015). Expanding the scope of noninvasive prenatal testing: detection of fetal microdeletion syndromes. *Am J Obstet Gynecol* **212**(332), e331–e339. http://dx.doi.org/10.1016/j.ajog.2014.11.041.

[38] Zimmermann B, Hill M, Gemelos G, Demko Z, Banjevic M, Baner J, Ryan A, Sigurjonsson S, Chopra N, and Dodd M, et al (2012). Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. *Prenat Diagn* **32**, 1233–1241. http://dx.doi.org/10.1002/pd.3993.

[39] Nicolaides KH, Syngelaki A, Gil M, Atanasova V, and Markova D (2013). Validation of targeted sequencing of single-nucleotide polymorphisms for non-invasive prenatal detection of aneuploidy of chromosomes 13, 18, 21, X, and Y. *Prenat Diagn* **33**, 575–579. http://dx.doi.org/10.1002/pd.4103.

[40] Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, and Leunissen JA (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**, W71–W74. http://dx.doi.org/10.1093/nar/gkm306.

[41] Li H and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.

[42] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.

[43] Levy B, Sigurjonsson S, Pettersen B, Maisenbacher MK, Hall MP, Demko Z, Lathi RB, Tao R, Aggarwal V, and Rabinowitz M (2014). Genomic imbalance in products of conception: single-nucleotide polymorphism chromosomal microarray analysis. *Obstet Gynecol* **124**, 202–209. http://dx.doi.org/10.1097/AOG.0000000000000325.

[44] Martinez P, McGranahan N, Birkbak NJ, Gerlinger M, and Swanton C (2013). Computational optimisation of targeted DNA sequencing for cancer detection. *Sci Rep* **3**. http://dx.doi.org/10.1038/srep03309.

[45] Perkins G, Yap TA, Pope L, Cassidy AM, Dukes JP, Riisnaes R, Massard C, Cassier PA, Miranda S, and Clark J, et al (2012). Multi-purpose utility of circulating plasma DNA testing in patients with advanced cancers. *PLoS One* **7**e47020. http://dx.doi.org/10.1371/journal.pone.0047020.

[46] Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, Rosner J, Shumansky K, Kalloger S, and Senz J, et al (2013). Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J Pathol* **231**, 21–34. http://dx.doi.org/10.1002/path.4230.

[47] Snyder M, Adey A, Kitzman JO, and Shendure J (2015). Haplotype-resolved genome sequencing: experimental methods and applications. *Nat Rev Genet* **16**, 344–358. http://dx.doi.org/10.1038/nrg3903.

[48] Kaper F, Swamy S, Klotzle B, Munchel S, Cottrell J, Bibikova M, Chuang HY, Kruglyak S, Ronaghi M, and Eberle MA, et al (2013). Whole-genome haplotyping by dilution, amplification, and sequencing. *Proc Natl Acad Sci U S A* **110**, 5552–5557. http://dx.doi.org/10.1073/pnas.1218696110.

[49] Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, Kertesz M, and Synder M, et al (2014). Whole-genome haplotyping using long reads and statistical methods. *Nat Biotechnol* **32**, 261–266. http://dx.doi.org/10.1038/nbt.2833.