Breast Cancer Research and Treatment (2019) 178:251–261
https://doi.org/10.1007/s10549-019-05374-x

PRECLINICAL STUDY

Detection of breast cancer stem cell gene mutations in circulating free DNA during the evolution of metastases

Zhe-Bin Liu[1,2,13] · Nader E. Ezzedine[3] · Agda K. Eterovic[3] · Joe E. Ensor[4] · Helen J. Huang[5] · Joan Albanell[6,7,8] · Dong S. Choi[2,4] · Ana Lluch[6,7,9] · Yi Liu[2,4] · Federico Rojo[6,7,10] · Helen Wong[2,4] · Eduardo Martínez-Dueñas[6,11] · Ángel Guerrero-Zotano[6,12] · Zhi-Min Shao[1,13] · Jorge G. Darcourt[4] · Gordon B. Mills[3] · Bhuvanesh Dave[2,4] · Jenny C. Chang[2,4]

Received: 10 April 2019 / Accepted: 21 July 2019 / Published online: 6 August 2019

Abstract

Purpose Limited knowledge exists on the detection of breast cancer stem cell (BCSC)-related mutations in circulating free DNA (cfDNA) from patients with advanced cancers. Identification of new cancer biomarkers may allow for earlier detection of disease progression and treatment strategy modifications.

Methods We conducted a prospective study to determine the feasibility and prognostic utility of droplet digital polymerase chain reaction (ddPCR)-based BCSC gene mutation analysis of cfDNA in patients with breast cancer.

Results Detection of quantitative BCSC gene mutation in cfDNA by ddPCR mirrors disease progression and thus may represent a valuable and cost-effective measure of tumor burden. We have previously shown that *hematological and neurological expressed 1-like* (*HN1L*), *ribosomal protein L39* (*RPL39*), and *myeloid leukemia factor 2* (*MLF2*) are novel targets for BCSC self-renewal, and targeting these genetic alterations could be useful for personalized genomic-based therapy.

Conclusion BCSC mutation detection in cfDNA may have important implications for diagnosis, prognosis, and serial monitoring.

Keywords Breast carcinoma · Stem cell · Mutation · Droplet digital polymerase chain reaction · Metastasis

Abbreviations

BCSC	Breast cancer stem cell
cfDNA	Circulating free DNA
ddPCR	Droplet digital polymerase chain reaction
HN1L	Hematological and neurological expressed 1-like

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10549-019-05374-x) contains supplementary material, which is available to authorized users.

Zhe-Bin Liu, Bhuvanesh Dave, and Jenny C. Chang have contributed equally to this work.

✉ Jenny C. Chang
jcchang@houstonmethodist.org

1 Department of Breast Surgery, Fudan University Shanghai Cancer Center, Shanghai 20032, China

2 Houston Methodist Research Institute, Houston, TX 77030, USA

3 Department of Systems Biology and Institute of Personalized Cancer Therapy, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

4 Houston Methodist Cancer Center, 6445 Main St P21-34, Houston, TX 77030, USA

5 Division of Cancer Medicine, Department of Investigational Cancer Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

6 GEICAM (Spanish Breast Cancer Group), San Sebastián de los Reyes, 28703 Madrid, Spain

7 Centro de Investigación Biomédica en Red de Oncología, CIBERONC-ISCIII, 28029 Madrid, Spain

8 Hospital del Mar, 08003 Barcelona, Spain

9 Hospital Clínico Universitario de Valencia, 46010 Valencia, Spain

10 Fundación Jiménez Díaz, 28040 Madrid, Spain

11 Hospital Provincial de Castellón, 12002 Castellón, Spain

12 Instituto Valenciano de Oncología, 46009 Valencia, Spain

13 Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

RPL39	Ribosomal protein L39
MLF2	Myeloid leukemia factor 2
RNA-Seq	RNA deep sequencing
SIFT	Sorting intolerant from tolerant
HER2	Human epidermal growth factor receptor 2
TTM	Time-to-metastasis
NOS	Nitric oxide synthase
ER	Estrogen receptor

Background

Despite improvements in cancer therapies, breast cancer metastasis remains largely incurable and the primary cause of cancer-related deaths. Radiographic measurements, used to monitor and assess tumor burden, often underestimate the true extent of disease due to their relatively low resolution. Similarly, protein biomarkers, such as carcinoembryonic antigen and carcinoma antigen-125, lack sensitivity and specificity for clinical utility. Thus, new biomarkers for measuring tumor burden with high diagnostic sensitivity and specificity are urgently needed [1]. Cancer stem cells (CSCs) are a subpopulation of cells involved in tumor initiation, progression, metastasis, and therapeutic resistance. Therefore, CSCs may serve as diagnostic, prognostic, and predictive biomarkers [2–4]. Therapeutic approaches to target breast CSCs (BCSCs) are being investigated [5]. However, few studies have examined the genetic profile and somatic mutations of BCSCs to identify early genetic alterations in tumor evolution that might improve both early metastatic detection and patient survival. We previously identified a BCSC-related 477-gene signature by comparing CSCs and non-CSCs [4]. Using a shRNA knockdown functional approach, *ribosomal protein L39* (*RPL39*), *myeloid leukemia factor 2* (*MLF2*), and *hematological and neurological expressed 1-like* (*HN1L*) were identified as the top candidate genes that significantly impact BCSC self-renewal [6, 7]. By using RNA deep sequencing (RNA-Seq) and sorting intolerant from tolerant (SIFT) analyses of lung metastases, we described 4 mutations in these BCSC genes including A14V in *RPL39*, R158W in *MFL2*, and P20L and A106V in *HN1L* [6–8].

Most publically available datasets have sequenced primary tumors; however, less comprehensive molecular analysis exists from metastatic sites and paired analysis of primary and metastatic samples are difficult to obtain. Recently, whole-exome and whole-transcriptome sequencing of 500 adult patients with metastatic solid tumors was published. The most prevalent genes somatically altered in metastatic cancer included TP53, CDKN2A, PTEN, PIK3CA, and RB1 [9]. However, specific mutations in CSCs for identifying early genetic alterations in tumor evolution that might improve metastatic detection have not been analyzed.

The biopsy procedure of metastases is invasive, and intratumoral heterogeneity can confound genomic analysis. Circulating free DNA (cfDNA) extracted from plasma or serum is a less invasive source for detecting tumor-derived DNA mutation [1, 10]. Plasma cfDNA analysis ("liquid biopsy") can be used to monitor changes in total tumor burden over time in a cost-effective manner and can identify specific alterations that arise during therapy in real time [1, 11]. Its feasibility has been shown in patients with different tumor types [12, 13]. However, mutant DNA accounts for a small fraction of the total circulating DNA [14] and undetectable by next-generation sequencing technology. Technological advances like droplet digital polymerase chain reaction (ddPCR) have enabled the design of high sensitivity cfDNA mutation assays for disease detection. ddPCR can detect and quantify mutant DNA molecules present at $\geq 0.01\%$ of the total DNA [15], as has been reported for somatic mutations such as *BRAF* V600E in cfDNA [16].

While cancer-related gene mutations in cfDNA from patients with advanced cancers have been detected [1], this is the first study to determine the feasibility and prognostic utility of ddPCR-based BCSC gene mutation analysis of cfDNA in patients with early-stage and advanced breast cancer.

Materials and methods

Study design, participants, and sample collection

FFPE tumor tissue samples, blood samples, and clinical data were collected from breast cancer patients at Houston Methodist Hospital. The study was approved by the Institutional Review Board, and written informed consent was obtained from all patients before sample and data collection. FFPE tumor tissue samples were obtained from patients with stage IV breast cancer. Tissue biopsies were performed as part of routine clinical care. Plasma samples were obtained from 124 patients with newly diagnosed stage I–IV breast cancer before treatment, and 86 healthy female volunteers without a history of cancer, who presented for routine blood testing to medical and surgical clinics.

Clinical features of the patients

Clinicopathological characteristics of the patients are shown in Table 1. Of the 124 patients, mean patient age was 54 years (range 30–92 years). 58 (46.7%), 26 (21.0%), and 40 (32.3%) had stage I/II, stage III, and stage IV breast cancer at the time of analysis, respectively. Fifty patients (40.3%) were diagnosed with ER-positive breast cancer, and 66 patients (53.2%) were diagnosed with ER-negative breast cancer. Most patients (n = 104; 83%) had human epidermal

Table 1 Demographic and clinical characteristics of the study participants

	FFPE lung mets	Primary/lung	Primary/other	Plasma
Total subjects (n =)	28	8 pairs	24 pairs	124
Age (years)				
Mean	54	54	61.4	54
Range (min–max)	34–74	34–74	41–81	30–92
Histology				
IDC	27	7	25	102
ILC	0	1	6	7
Unspecified	1	0	5	15
Grade				
1 (Well)	1	1	4	5
2 (Moderate)	3	3	11	32
3 (Poor)	10	3	19	62
Unspecified	14	1	2	25
Stage				
I	NA	NA	NA	15
II	NA	NA	NA	43
III	NA	NA	NA	27
IV	28	8	NA	39
ER				
+	18	4	29	50
−	9	3	7	66
Unspecified	1	1	0	8
PR				
+	12	4	16	37
−	15	3	20	79
Unspecified	1	1	0	8
HER2				
+[a]	4	2	10	10
−	20	5	26	104
Equivocal/unspecified	4	1	0	10
Time-to-metastasis (months)				
Median	67	21	57.6	NA
Range (min–max)	16–199	15–196	0.7–201.1	NA

ER estrogen receptor, *HER2* human epidermal growth factor receptor 2, *PR* progesterone receptor, *IDC* invasive ductal carcinoma, *ILC* invasive lobular carcinoma

[a]IHC +++ or FISH +; Equivocal by immunohistochemical staining 2+, FISH ratio of 1.8–2.2 or HER2 gene copy of 4.0–6.0

growth factor receptor 2 (HER2)-negative disease, while 62 patients (50%) had poorly differentiated tumors. The healthy volunteers were recruited from patients without a cancer diagnosis, who were having routine laboratory testing for other diagnoses.

DNA extraction from FFPE primary and metastatic tumor samples

FFPE primary and metastatic tumor samples were sectioned and stained with hematoxylin and eosin to ensure at least 70% tumor content. Tumor areas were identified before macrodissection. FFPE DNA was purified and isolated from primary and metastatic tumor samples using the RecoverAll™ Total Nucleic Acid Isolation Kit according to the manufacturer's instructions (cat. no. AM1975; Thermo Fisher Scientific, Wilmington, DE, USA). DNA concentrations were determined by measuring optical density at 260 nm (OD260) with a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific), and purity was evaluated by measuring OD260/OD280 ratio.

DNA extraction from blood samples

Briefly, blood (18 ml) was drawn from a central catheter into EDTA tubes, centrifuged at 1000×*g* for 10 min, carefully aliquoted into 1.5 ml tubes, and immediately frozen at − 80 °C. Samples were stored for 1–24 h after collection. Plasma DNA was extracted using the QIAamp Circulating Nucleic Acid Kit according to the manufacturer's instructions (cat. no. 55114; Qiagen, Valencia, CA, USA) and quantified as described for FFPE DNA.

Droplet digital polymerase chain reaction (ddPCR)

ddPCR was performed as described in Sanmamed et al.'s study [16] using the QX200 Droplet Digital PCR System (Bio-Rad Laboratories, Hercules, CA, USA). Custom TaqMan probes and primers for *HN1L* (P20L and A106V) mutations were designed by Applied Biosystems (Life Technologies). Custom *RPL39* (A14V), *MLF2* (R158W), and *PIK3CA* (H1047R and E545K) ddPCR probes and primers were designed by Bio-Rad. cfDNA (2 µl) and tumor DNA samples (1 µl of 10 ng/µl) and 10 µl of 2X Droplet PCR Supermix for Probes (Bio-Rad) were premixed and combined with 1 µl each of primers and probes (20×) to a final 20-µl reaction volume with DNase/RNase-free sterile water. For droplet generation, 20-µl sample was added to 70 µl of Droplet Generation Oil (Bio-Rad). A temperature gradient PCR using both a row/column of restriction-digested 100% WT DNA at a relatively high DNA concentration and a concentration gradient PCR using WT DNA with mutant spike-in DNA were tested to determine the optimal annealing temperature per application guild of ddPCR. Droplets were then thermal cycled as follows: 95 °C for 10 min; 40 cycles of 94 °C for 30 s and 55–58 °C for 1 min; and 98 °C for 10 min. After PCR thermocycling, the emulsions were transferred to the QX200 digital droplet reader for simultaneous fluorescent measurement of FAM and VIC or HEX probes. Positive, negative, and no template controls were included in all pre-amplification and ddPCR steps. Samples with detectable mutation were reanalyzed to confirm mutation status. Data were analyzed using QuantaSoft software (Bio-Rad). Droplets were scored as positive or negative based upon their fluorescence intensity, which was determined by gating a threshold using positive, negative, and no template controls. The fractional abundance of the mutations in each sample was calculated by dividing the number of mutant DNA molecules (copies/µl) to the number of total DNA molecules (mutant plus WT, copies/µl) and expressed as the percentage of mutant DNA molecules in a sample taking into account a Poisson distribution of occupied to unoccupied droplets.

Data and statistical analyses

Statistical analyses were performed using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). Time-to-event endpoints (TTM, PFS) were estimated using the Kaplan–Meier method, and differences were compared using the log-rank test. Individual mutations between paired metastatic and primary samples were compared using a Wilcoxon signed-rank test. To compare fractional abundance between paired samples across all four mutations simultaneously, a mixed-effects generalized linear model was fitted to the fractional abundance difference (i.e., metastasis value—primary value) that contained a random subject effect to account for the intra-patient correlation (i.e., four mutations measured on each patient) with a variance–covariance structure that allowed for heterogeneity among biomarkers. For all analyses, two-tailed p values < 0.05 were considered statistically significant.

Results

The study schema for mutational analysis from tumors and plasma cfDNA is outlined in Fig. 1. Non-matched lung metastases (n = 28), matched primary and lung metastases (n = 8), matched primary and other metastases (n = 24) and plasma cfDNA (n = 210; cancer patients 124 and normal 86) were analyzed.

BCSC mutations in non-matched lung metastatic tumor samples

ddPCR was performed to detect four BCSC mutations (*MLF2* R158W, *RPL39* A14V, *HN1L* P20L, and *HN1L* A106V, Supplementary Table S1) on 28 formalin-fixed, paraffin-embedded (FFPE) non-matched lung metastatic breast cancer samples. Patient demographics and clinical characteristics are shown in Table 1.

Of the 28 non-matched lung metastases, 12 (42.9%) harbored at least 1 of the 4 BCSC gene mutations (Fig. 2a). ddPCR revealed the presence of *HN1L* P20L, *HN1L* A106V, *RPL39* A14V, and *MLF2* R158W mutations in 5, 1, 6, and 3 samples, respectively, yielding mutation frequencies ranging from 3.6 to 21.4% (Fig. 2b). *RPL39* A14V had the highest frequency at 21.4%.

A log-rank test of time-to-metastasis (TTM, which is defined as the interval between surgery and the first metastasis) according to BCSC mutation status was also performed on metastatic patients. All the patients finished the multidisciplinary therapy according to the guideline. Statistically worse outcome was observed in patients with any of the 4 BSCS mutations, with median TTM of 32 months vs. 85 months, $p = 0.0197$ (Fig. 2c).

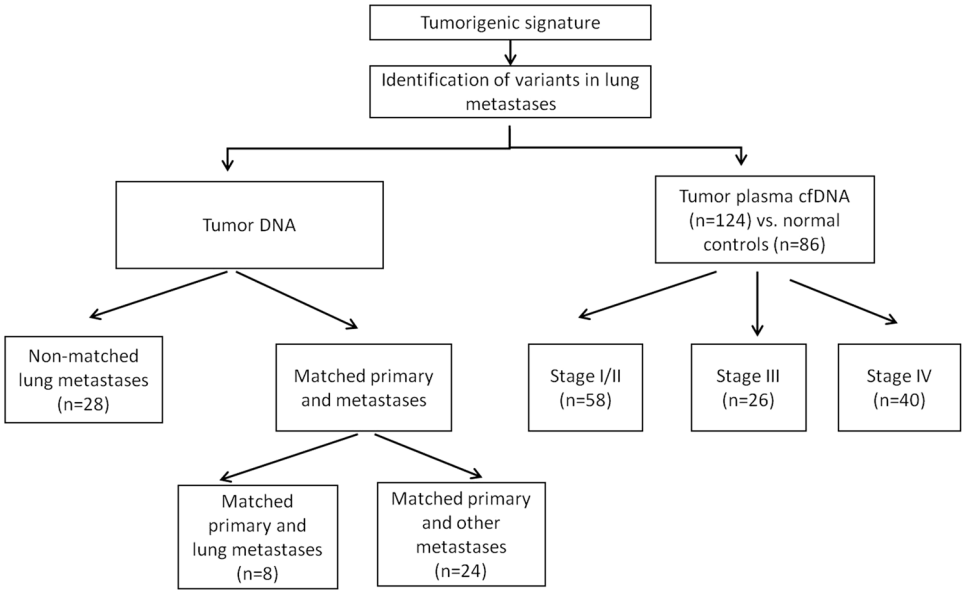


Fig. 1 Flow diagram of mutation analysis of tumor DNA and cfDNA. RNA-Seq was performed to detect the presence of BCSC-specific gene mutations in 8 lung metastatic breast cancer samples. ddPCR analysis of the four BCSC gene mutations was performed on 28 FFPE lung metastatic samples, 8 matched primary breast and lung metastatic pairs, and 24 matched primary breast and other metastases. Plasma cfDNA was analyzed from 210 individuals (86 normal volunteers, and 124 cancer patients with active stage I, II, III, and IV breast cancer). *BCSC* breast cancer stem cell, *cfDNA* circulating free DNA, *ddPCR* droplet digital polymerase chain reaction, *FFPE* formalin-fixed, paraffin-embedded

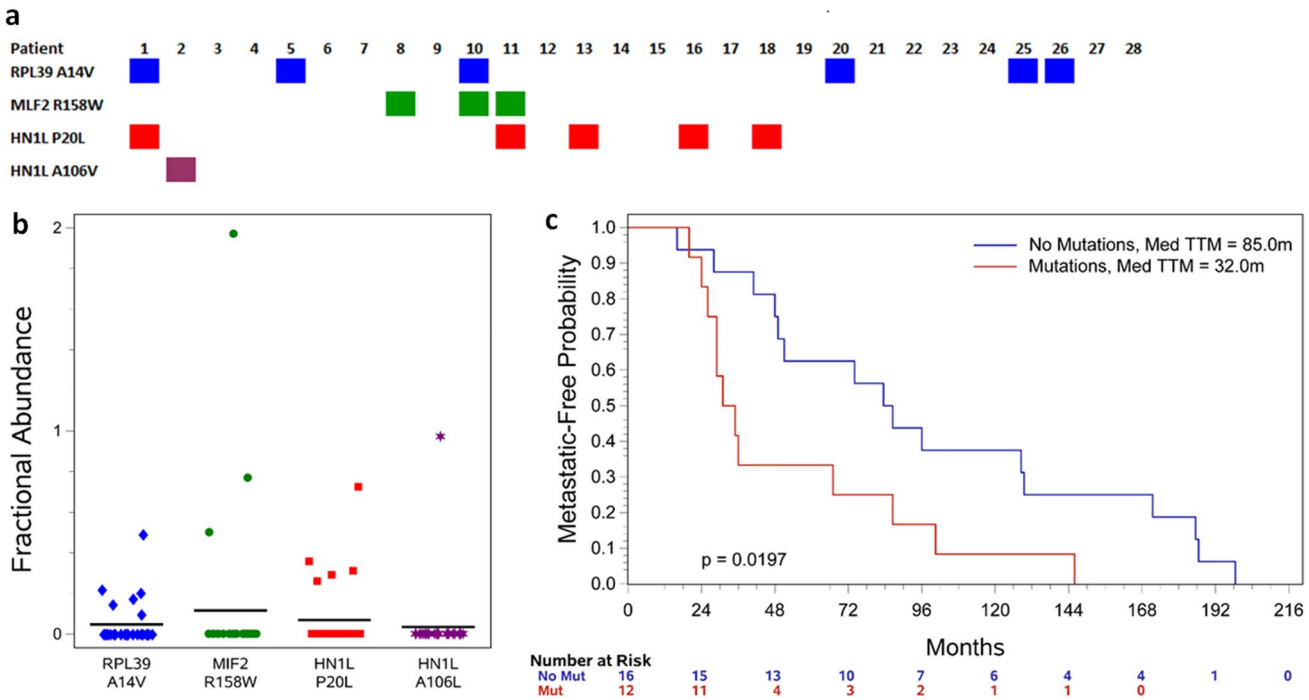


Fig. 2 Mutational spectrum in non-matched lung metastases (n = 28). **a** Mutational frequency in non-matched lung metastatic tumor samples. The frequency of each mutation was calculated as the ratio of mutation-positive samples to total samples. **b** Fractional abundances of BCSC gene alterations in 28 lung metastatic tissue samples. For each sample, the fractional abundance (percentage of mutant DNA molecules) was calculated by dividing the number of mutant DNA molecules (copies/µl) by the number of total DNA molecules (mutant plus WT, copies/µl). **c** Metastasis-free probability by Kaplan–Meier analysis of TTM according to BCSC gene mutation status of FFPE lung metastatic samples. *BCSC* breast cancer stem cel, *FFPE* formalin-fixed paraffin-embedded, *TTM* time-to-metastasis, *WT* wild-type

Mutation analysis in matched primary breast/lung and breast/other metastases

To determine whether *MLF2* R158W, *RPL39* A14V, *HN1L* P20L, and *HN1L* A106V mutations were specific to breast cancer metastases, we assessed the concordance of BCSC gene mutation status in eight matched primary breast and lung metastatic tumor pairs. ER was positive in 4 pairs, negative in 3 pairs, and unspecified in 1 pair. Median time from breast surgery to lung surgery was 41 months (range 15–196 months) (Table 1). Among the 8 metastatic samples, 16 mutations were observed (5 HN1L_P20L, 4 HN1L A106V, 4 RPL39 A14V, 3 MLF2 R158W) and among the 8 primary lesions, 10 mutations were noted (4 HN1L_P20L, 2 HN1L A106V, 1 RPL39 A14V, 3 MLF2 R158W). Of the 32 metastatic/primary pairs (i.e., 8 patients/4 genetic biomarkers), 14 showed an increase in fractional abundance, 5 showed a decrease, and 13 showed no change ($p = 0.0096$, Fig. 3a).

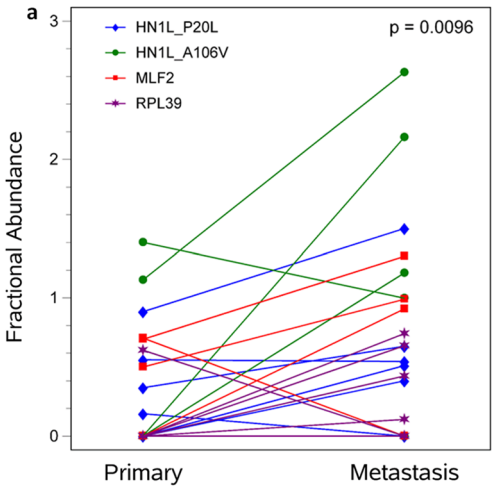
In the second cohort of 24 matched primary and other metastatic sites (ER-positive 29, ER-negative 7), DNA was extracted for ddPCR analysis in 24 pairs. There was a significant increase in mutation frequency in metastatic sites (skin, liver, lung, ovarian, pleura) (Fig. 3b and Table S1).

Plasma cfDNA of BCSC gene mutations

Clinicopathological characteristics of the patients are shown in Table 1. Mutation frequencies in the cfDNA samples were normalized based on the mean fractional abundances of the individual BCSC mutations in control cfDNA samples, to establish a threshold above which a patient sample is considered to be positive for the mutation. At least one of the four BCSC gene mutations was detected in 36 of the 124 (29%) cancer patient cfDNA samples (Fig. 4a, b). Specifically, *HN1L* P20L, *HN1L* A106V, *RPL39* A14V, and *MLF2* R158W mutations were detected in 8 (6.5%), 8 (6.5%), 14 (11.3%), and 15 (12.1%) cancer plasma cfDNA samples, respectively, and 9 patients (7.2%) had more than one mutation. The fractional abundances of the four BCSC gene mutations varied among the samples, ranging from 0.07 to 3.57% (Fig. 4b). Fractional abundances ranged from 0.40 to 2.50% for *HN1L* P20L, 0.70–2.98% for *HN1L* A106V, 0.07–2.77% for *RPL39* A14V, and 0.42–3.57% for *MLF2* R158W.

Clinical relevance of BCSC gene mutations detected in cfDNA and tumor lung metastases

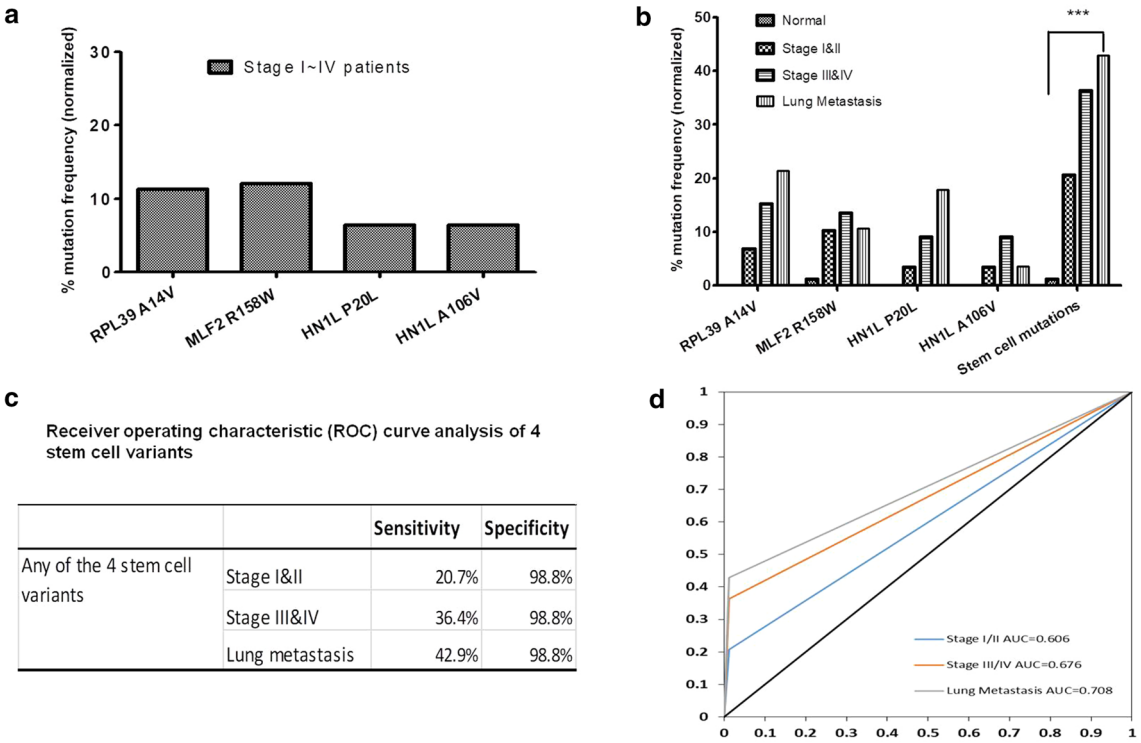
To determine the clinical relevance of BCSC mutations detected in cfDNA, subgroups analyses according to clinical stage were performed. BCSC mutation frequency ranged from 1.7%, 10.3%, 9.1%, and 15.2% in stage I, II, III, and IV subgroups, respectively. The detection of these mutations in tumor lung metastases (n = 28) was also compared. The fractional abundances of the four mutations detected in plasma cfDNA and lung metastases significantly increased from early to late disease stage and in the lung metastases, with high specificity ($p = 0.000$). With increasing stage, the sensitivity of detection of stem cell mutations increased, from stage I and II, III and



b

Biomarker	Median Increase	p-value
RPL39	0.000	0.0017
MLF2	0.205	< 0.0001
HN1L-P20L	0.180	0.0006
HN1L-A106V	0.135	0.0031

Fig. 3 Mutational analysis in matched primary breast/lung metastatic pairs and primary breast/other metastatic pairs. ddPCR analysis of BCSC gene alterations was performed on 8 matched pairs of primary breast cancer and lung metastases, and with 24 pairs of breast and other metastatic sites. **a** Shared mutations, mutations only found in primary breast and lung metastatic samples are shown. Shared mutation indicates mutations expressed in both primary breast cancer and matched lung metastases. Enrichment indicates a higher fractional abundance of mutant gene copies in lung metastases. **b** Analysis of all other metastatic sites was done for 24 matched pairs, and enrichment with higher fractional abundance of mutant gene copies was observed in metastases. *BCSC* breast cancer stem cell, *ddPCR* droplet digital polymerase chain reaction



		Sensitivity	Specificity
Any of the 4 stem cell variants	Stage I&II	20.7%	98.8%
	Stage III&IV	36.4%	98.8%
	Lung metastasis	42.9%	98.8%

Fig. 4 BCSC and gene mutation frequency in plasma cfDNA from normal controls (n = 86) and patients with stage I–IV breast cancer (n = 124). **a** Normalized mutation frequencies of the 4 BCSC in plasma cfDNA samples from 124 patients with stage I–IV breast cancer. Mutation frequencies in patient cfDNA samples were normalized based on the mean fractional abundances of the individual BCSC gene mutations in normal control cfDNA samples. **b** Normalized mutation frequencies of the 4 BCSC in plasma cfDNA samples from 86 normal control cfDNA samples and 124 patients with stage I–IV breast cancer, as well as 28 FFPE samples from lung metastases. Patients with breast cancer were divided into early (stage I or stage II) and late (stage III or stage IV) groups. At least 1 of the 4 BCSC gene mutations defined as stem cell mutations was calculated. ***Represented statistically different among 4 groups (normal, stage I or II, stage III or IV). **c** Sensitivity and specificity of BCSC gene mutation detection in plasma cfDNA and FFPE lung metastatic samples from breast cancer patients compared with plasma cfDNA samples from normal controls. **d** ROC curve analysis of the BCSC gene mutations in cfDNA samples from breast cancer patients according to stage and tumor DNA of lung metastatic samples. *AUC* area under the curve, *BCSC* breast cancer stem cell, *cfDNA* circulating free DNA, *ddPCR* droplet digital polymerase chain reaction, *FFPE* formalin-fixed paraffin-embedded, *ROC* receiver operating characteristic

IV, to lung metastases (20.7%, 36.4%, and 42.9%), respectively (Fig. 4c).

Subgroup analyses for mutation frequency in cfDNA in early and late stage versus lung metastases in breast cancer patients compared to controls showed that *HN1L* P20L, *HN1L* A106V, *RPL39* A14V, and *MLF2* R158W differed significantly among these four subgroups. By calculating the sensitivity against 1 minus the specificity, receiver operating characteristic (ROC) analysis for plasma cfDNA BCSC gene mutation analysis revealed area under the curves (AUCs) of 0.606 and 0.676 for stage I/II versus stage III/IV, and an AUC of 0.708 for tumor DNA obtained from lung metastases (Fig. 4c).

Prognostic value of cfDNA BCSC gene mutations in patients with stage IV breast cancer

We examined the relationship between BCSC gene mutation status detected in cfDNA and clinical outcome in 39 patients with stage IV breast cancer. A log-rank test revealed a shorter progression-free survival (PFS) in patients whose plasma cfDNA samples were positive for at least one of the four mutations compared with those whose plasma cfDNA samples were negative for these mutations (4 months vs. 12 months; $p = 0.035$; Fig. 5). The small number of patients with stage I–III and the low number of

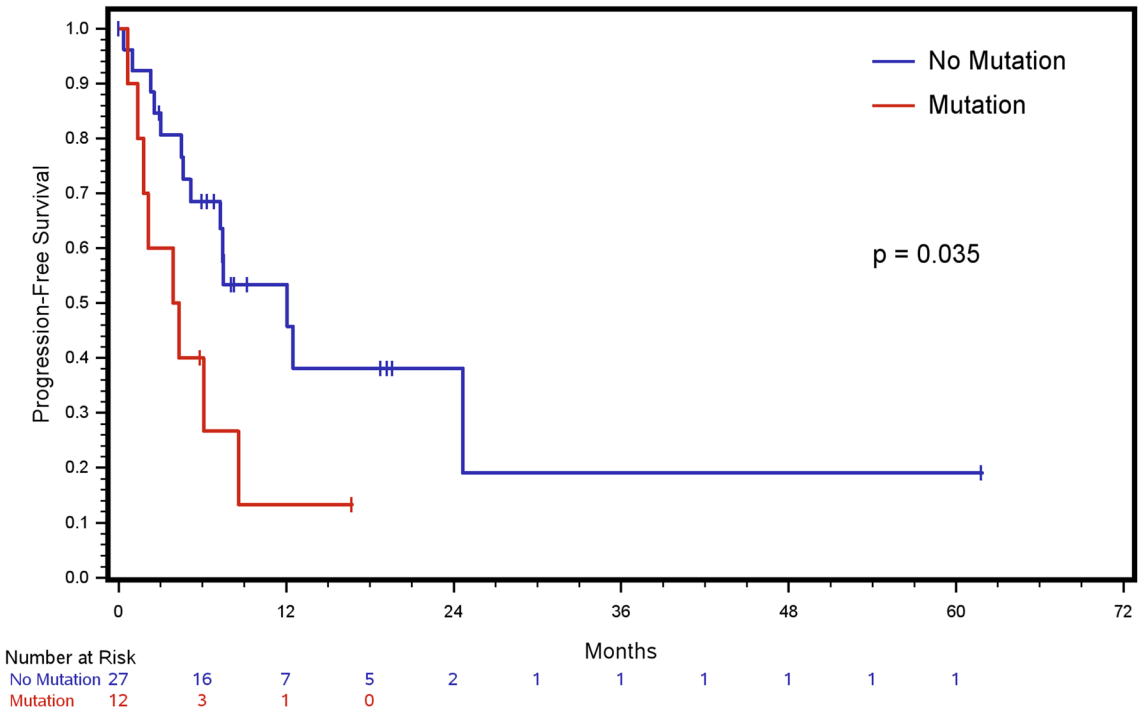


Fig. 5 Kaplan–Meier analysis according to cfDNA BCSC gene mutation status in 40 patients with stage IV breast cancer. *BCSC* breast cancer stem cell, *cfDNA* circulating free DNA, *PFS* progression-free survival

events limited the statistical power and relapse-free survival was not significantly different.

Discussion

We previously identified *RPL39*, *MLF2*, and *HN1L* as novel BCSC targets by screening candidate genes from a tumorigenic signature derived from patient biopsies [6]. Four mutations in these BCSC-specific genes were identified in lung metastases using RNA-Seq [3, 4, 6]. In this study, these BCSC gene mutations were confirmed in non-matched lung metastatic FFPE samples using the ddPCR technique. There was a significant increase in mutation frequency for these genes in paired primary breast versus metastatic sites. The presence of mutations in plasma cfDNA was associated with a significantly worse outcome in patients with metastatic disease. Our results indicate that rare BCSC gene alterations detected in cfDNA have biological implications and may represent promising biomarkers with potential clinical prognostic applicability.

The identification of genetic alterations in cancer has led to the development of new non-invasive approaches for tumor detection [10] and has since been used to detect plasma cfDNA in other tumor types including breast cancer [10–15]. We show that ddPCR-based cfDNA analysis can quantify the number of copies of a given mutant gene fragment and their variation with disease stage, and the fraction of mutant gene sequences in the WT background in circulation. Indeed, ddPCR technology can be used for monitoring cfDNA mutations such as estrogen receptor mutations as an indicator of treatment resistance, for detecting previously unidentified mutational clones [17] or as an indicator of disease recurrence [18]. However, BCSC mutation accounts for a minor fraction of the total plasma cfDNA and not described in bulk primary breast tumors, as these mutations may be present in a small subgroup of CSCs and intratumoral heterogeneity may prevent the identification of specific mutations in distinct subpopulations [19]. To the best of our knowledge, this is the first study to demonstrate that BCSC gene mutation detection in plasma cfDNA samples is feasible and represents a highly sensitive biomarker of tumor burden, with the fractional abundances of these mutations increasing from early to late stage disease and in metastases. We have previously demonstrated that somatic mutations in *RPL39* (A14V) and *MLF2* (R158W) in tumors were associated with shorter median time to relapse in breast cancer patients and upregulation of *HN1L* expression correlates with poor clinical outcome in triple negative breast cancer patients [6, 7]. In this study, BCSC gene mutations were detected in 29% of cfDNA samples from breast cancer patients and in over 40% of lung metastatic samples. Over 40% (14/32) primary/metastatic pairs showed an increase in fractional abundance of these mutations. The increasing

fractional abundance of these mutations with increasing stage is interesting, either as an early marker of disease, or marker of metastatic risk. Potential explanations include either an overall increase in micrometastatic tumor burden, or potentially an accumulation of these mutations during the metastatic process. BCSC gene mutation in tumor and plasma cfDNA correlated with disease stage, where the presence of any of these BCSC gene mutations was associated with significantly worse survival.

Nitric oxide synthase (NOS) signaling is a pathway common to both *RPL39* and *MLF2* [8]. Overexpression and knockdown of *RPL39* and *MLF2* increased and decreased inducible NOS (iNOS) signaling, respectively [8]. *HN1L* has been shown to have a similar effect on CSC self-renewal and promotes BCSCs through LEPR-STAT3 pathway [7]. Targeting NOS or STAT3 signaling may be an effective approach to overcome BCSC-induced chemo-resistance. Understanding mechanistic role of these BCSC gene alterations in self-renewal and monitoring response to targeted therapies by non-invasive "liquid biopsies" will be important in developing potential therapeutic targets for treatment resistance.

Pathologic confirmation through biopsy is the only way to make a definitive diagnosis of metastasis in breast cancer patients. Here, we show that even relatively early-stage cancers give rise to circulating mutant DNA fragments that can be detected with sufficient sensitivity and specificity with ddPCR-based mutation analysis. Yet, the reliability and reproducibility of mutation detection in cfDNA and the management of patients with discordant cfDNA and imaging tests still need to be addressed. Large prospective intervention studies are required to define the optimal cutoff values for the BCSC gene mutations. Additionally, plasma-based assays may be inferior to traditional tumor tissue-based assays for mutant DNA detection due to the shorter half-life of cfDNA. Nonetheless, the ease of execution and repeatability of plasma-based assays compared to invasive tissue biopsy, especially for metastatic sites, makes this an important technological advance, especially if early evolutionary alterations can be detected.

A limitation of this study is that the sample size is small and heterogeneous (biologically and clinically). A well-defined, homogeneous study population in a context of a clinical trial is desirable to test for the effect of covariates while assessing the effect of mutational status on progression-free survival. Our sample of 39 stage IV patients for which follow-up exists for 35 and contains 22 events was not large enough to provide enough power to assess multiple factors simultaneously. We also acknowledge that the analysis of time-to-metastasis included a population of patients who already metastasized and greater number of stage I–III patients are needed to make any conclusion. Nevertheless, the design of this study was based on retrospective samples, and as such, was limited by the accessibility of both matched and unmatched primary-metastatic pairs. Metastatic samples are notoriously difficult to obtain, even in multi-institutional studies like this. These preliminary analysis based on these retrospective samples form the basis to generate data so that future larger randomized prospective studies can be conducted. Nonetheless, the number of plasma cfDNA samples from cancer patients examined was 129, which is a large cohort. As suggested in the REMARK paper, the goals of this paper are consistent with the guidelines to "encourage transparent and complete reporting so that relevant information will be available to others to help them judge the usefulness of the data and understand the context in which the conclusions apply" [20]. We have ongoing clinical trials, which we hope will address some of these issues and shed light on whether the cfDNA could serve as an early marker of disease, or marker of metastatic risk.

Conclusions

In summary, we demonstrated the feasibility of cfDNA for specific BCSC gene alterations in patients with breast cancer. Quantitative BCSC gene mutation detection in plasma cfDNA mirrored disease progression and therefore may represent a valuable marker of tumor burden. *HN1L*, *RPL39*, and *MLF2* are novel targets for BCSC self-renewal and targeting these genetic markers could be useful for personalized genomic-based therapy. Thus, measurement of CSCs in plasma cfDNA may have important implications for diagnosis, prognosis, and serial monitoring.

Acknowledgements Any opinions, findings, and conclusions expressed in this materials are those of the author(s) and do not necessarily reflect those of the American Society of Clinical Oncology® or the Conquer Cancer Foundation, or The Breast Cancer Research Foundation. We would like to thank Dr. Ana María González-Angulo from MD Anderson Hospital who identified the samples from GEICAM investigators.

Author contributions Conceptualization, ZBL, BD, and JCC; Methodology, ZBL, NEE, AKE, BD, HW, and JCC; Investigation, ZBL, DSC, YL, HW, and BD; Formal Analysis, ZBL, NEE, JEE, and BD.; Resources, AKE, HJH, ZMS, JGD, GBM, and JCC; Writing—Original Draft, ZBL; Writing—Review and Editing, ZBL, JGD, DK, BD, and JCC; Supervision, BD and JCC; Funding Acquisition, JCC.

Funding This work was supported by a 2013 Conquer Cancer Foundation of ASCO Long-term International Fellowship (LIFe) in Breast Cancer, supported by The Breast Cancer Research Foundation.

Data availability The datasets supporting the conclusions for the current study are stored in secured shared drive and will be shared by the corresponding author upon reasonable request.

Compliance with ethical standards

Conflict of interests The authors declare no potential conflicts of interest with the material reported in this manuscript and have no financial relationship with the organization that funded this study. Individual conflict of interest unrelated to the data presented here are listed in the conflict of interest form.

Research involving human participants and informed consent All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Institutional Review Board at the Houston Methodist Hospital (IRB protocol numbers 0908-0265, 0811-0147, and 0208-0033) and written informed consent was obtained from all patients before sample and data collection.

Animal subjects This article does not contain any studies with animals performed by any of the authors.

References

1. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N (2013) Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 368(13):1199–1209. https://doi.org/10.1056/NEJMoa1213261
2. Smalley M, Ashworth A (2003) Stem cells and breast cancer: a field in transit. Nat Rev Cancer 3(11):832–844. https://doi.org/10.1038/nrc1212
3. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, Hilsenbeck SG, Pavlick A, Zhang X, Chamness GC, Wong H, Rosen J, Chang JC (2008) Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst 100(9):672–679
4. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, Rimm DL, Wong H, Rodriguez A, Herschkowitz JI, Fan C, Zhang X, He X, Pavlick A, Gutierrez MC, Renshaw L, Larionov AA, Faratian D, Hilsenbeck SG, Perou CM, Lewis MT, Rosen JM, Chang JC (2009) Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc Natl Acad Sci 106(33):13820–13825. https://doi.org/10.1073/pnas.0905718106
5. Luo M, Clouthier SG, Deol Y, Liu S, Nagrath S, Azizi E, Wicha MS (2015) Breast cancer stem cells: current advances and clinical implications. Methods Mol Biol 1293:1–49. https://doi.org/10.1007/978-1-4939-2519-3_1
6. Dave B, Granados-Principal S, Zhu R, Benz S, Rabizadeh S, Soon-Shiong P, Yu KD, Shao Z, Li X, Gilcrease M, Lai Z, Chen Y, Huang TH, Shen H, Liu X, Ferrari M, Zhan M, Wong ST, Kumaraswami M, Mittal V, Chen X, Gross SS, Chang JC (2014) Targeting RPL39 and MLF2 reduces tumor initiation and metastasis in breast cancer by inhibiting nitric oxide synthase signaling. Proc Natl Acad Sci USA 111(24):8838–8843. https://doi.org/10.1073/pnas.1320769111
7. Liu Y, Choi DS, Sheng J, Ensor JE, Liang DH, Rodriguez-Aguayo C, Polley A, Benz S, Elemento O, Verma A, Cong Y, Wong H, Qian W, Li Z, Granados-Principal S, Lopez-Berestein G, Landis MD, Rosato RR, Dave B, Wong S, Marchetti D, Sood AK, Chang JC (2018) HN1L promotes triple-negative breast cancer stem cells through LEPR-STAT3 pathway. Stem cell Rep 10(1):212–227. https://doi.org/10.1016/j.stemcr.2017.11.010
8. Dave B, Gonzalez DD, Liu ZB, Li X, Wong H, Granados S, Ezzedine NE, Sieglaff DH, Ensor JE, Miller KD, Radovich M, KarinaEtrovic A, Gross SS, Elemento O, Mills GB, Gilcrease MZ, Chang JC (2017) Role of RPL39 in Metaplastic Breast Cancer. J Natl Cancer Inst. https://doi.org/10.1093/jnci/djw292
9. Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, Cao X, Rabban E, Kumar-Sinha C, Raymond V, Schuetze S, Alva A, Siddiqui J, Chugh R, Worden F, Zalupski MM, Innis J, Mody RJ, Tomlins SA, Lucas D, Baker LH, Ramnath N, Schott AF, Hayes DF, Vijai J, Offit K, Stoffel EM, Roberts JS, Smith DC, Kunju LP, Talpaz M, Cieslik M, Chinnaiyan AM (2017) Integrative clinical genomics of metastatic cancer. Nature 548(7667):297–303. https://doi.org/10.1038/nature23306
10. Schwarzenbach H, Hoon DS, Pantel K (2011) Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer 11(6):426–437. https://doi.org/10.1038/nrc3066
11. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, Parkinson C, Chin SF, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, Brenton JD, Caldas C, Rosenfeld N (2013) Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 497(7447):108–112. https://doi.org/10.1038/nature12065
12. Kimura H, Kasahara K, Kawaishi M, Kunitoh H, Tamura T, Holloway B, Nishio K (2006) Detection of epidermal growth factor receptor mutations in serum as a predictor of the response to gefitinib in patients with non-small-cell lung cancer. Clin Cancer Res 12(13):3915–3921. https://doi.org/10.1158/1078-0432.CCR-05-2324
13. Sorenson GD (2000) Detection of mutated KRAS2 sequences as tumor markers in plasma/serum of patients with gastrointestinal cancer. Clin Cancer Res 6(6):2129–2137
14. Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, Diaz LA Jr, Goodman SN, David KA, Juhl H, Kinzler KW, Vogelstein B (2005) Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci USA 102(45):16368–16373. https://doi.org/10.1073/pnas.0507904102
15. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, Vessella RL, Tewari M (2013) Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat Methods 10(10):1003–1005. https://doi.org/10.1038/nmeth.2633
16. Sanmamed MF, Fernandez-Landazuri S, Rodriguez C, Zarate R, Lozano MD, Zubiri L, Perez-Gracia JL, Martin-Algarra S, Gonzalez A (2015) Quantitative cell-free circulating BRAFV600E mutation analysis by use of droplet digital PCR in the follow-up of patients with melanoma being treated with BRAF inhibitors. Clin Chem 61(1):297–304. https://doi.org/10.1373/clinchem.2014.230235
17. Chu D, Paoletti C, Gersch C, VanDenBerg DA, Zabransky DJ, Cochran RL, Wong HY, Toro PV, Cidado J, Croessmann S, Erlanger B, Cravero K, Kyker-Snowman K, Button B, Parsons HA, Dalton WB, Gillani R, Medford A, Aung K, Tokudome N, Chinnaiyan AM, Schott A, Robinson D, Jacks KS, Lauring J, Hurley PJ, Hayes DF, Rae JM, Park BH (2016) ESR1 mutations in circulating plasma tumor DNA from metastatic breast cancer patients. Clin Cancer Res 22(4):993–999. https://doi.org/10.1158/1078-0432.CCR-15-0943
18. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, Cheang M, Osin P, Nerurkar A, Kozarewa I, Garrido JA, Dowsett M, Reis-Filho JS, Smith IE, Turner NC (2015) Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Science translational medicine 7(302):302ra133. https://doi.org/10.1126/scitranslmed.aab0021
19. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela

I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366(10):883–892. https://doi.org/10.1056/NEJMoa1113205

20. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, Statistics Subcommittee of NCIEWGoCD (2006) REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat 100(2):229–235. https://doi.org/10.1007/s10549-006-9242-8