The multiplicity of new therapies for breast cancer presents a challenge for treatment selection. We describe a 17-gene digital signature of breast circulating tumor cell (CTC)-derived transcripts enriched from blood, enabling high-sensitivity early monitoring of response. In a prospective cohort of localized breast cancer, an elevated CTC score after three cycles of neoadjuvant therapy is associated with residual disease at surgery ($P = 0.047$). In a second prospective cohort with metastatic breast cancer, baseline CTC score correlates with overall survival ($P = 0.02$), as does persistent CTC signal after 4 weeks of treatment ($P = 0.01$). In the subset with estrogen receptor (ER)-positive disease, failure to suppress ER signaling within CTCs after 3 weeks of endocrine therapy predicts early progression ($P = 0.008$). Drug-refractory ER signaling within CTCs overlaps partially with presence of ESR1 mutations, pointing to diverse mechanisms of acquired endocrine drug resistance. Thus, CTC-derived digital RNA signatures enable noninvasive pharmacodynamic measurements to inform therapy in breast cancer.

**SIGNIFICANCE:** Digital analysis of RNA from CTCs interrogates treatment responses of both localized and metastatic breast cancer. Quantifying CTC-derived ER signaling during treatment identifies patients failing to respond to ER suppression despite having functional ESR1. Thus, noninvasive scoring of CTC-RNA signatures may help guide therapeutic choices in localized and advanced breast cancer.
INTRODUCTION

Recent advances in therapeutics have revolutionized the management of breast cancer, with the approval of more than a dozen drugs in the past few years, including four new agents in 2017. In hormone receptor-positive (HR\(^+\)) breast cancer, new endocrine therapies may be combined with cyclin-dependent kinase 4/6 (CDK4/6) inhibitors (1–3), and drugs inhibiting growth factor receptors (e.g., IGFR and FGFR), oncogenic signaling pathways (e.g., PI3K, AKT, mTOR), and chromatin modifiers (e.g., HDAC) are under active clinical investigation (4). Despite these increasingly effective therapeutic choices, there are few biomarkers to guide initial therapy selection and identify early responses. As a result, treatment choices are frequently empiric, and delayed clinical ascertainment of tumor response limits the ability to rapidly define an effective regimen for an individual patient.

Traditionally, assessing therapy response in metastatic breast cancer involves monitoring serum cancer antigen protein markers, such as CA15-3 and CEA, along with radiographic assessment of tumor volumes (5, 6). However, in many cases, breast cancer involves bone metastases, whose responses to drug treatment are not readily assessed with radiographic imaging (7), and the sensitivity and accuracy of established protein serum markers is limited (4, 8). Recently, longitudinal monitoring of tumor-derived mutations detected in plasma (ctDNA) has been used as a measure of tumor response (9–11). This approach may provide a high degree of genetic information, although it typically requires initial sequencing of the primary tumor to design individualized mutational markers for each patient. Genetic heterogeneity in advanced disease may present an additional challenge in monitoring multiple subclonal mutant alleles with divergent trends following therapy (12).

Although early assessment of tumor response in metastatic breast cancer presents challenges, neoadjuvant treatment of localized breast cancer is empiric and lacks early measurements of drug response, which is only ultimately evident at the time of surgical resection. In women with high-risk localized breast cancer, multiple courses of neoadjuvant preoperative chemotherapy or hormonal therapy may be administered, with the goal of reducing initial tumor burden and
improving the outcome of subsequent surgical resection. ctDNA genotyping has been used to detect early relapse after neoadjuvant treatment and surgical resection, but it typically requires initial tumor sequencing and design of mutation-specific assays to measure the low fraction of mutant alleles in this minimal disease setting (10, 13). Ultrasensitive techniques to detect multiple recurrent somatic mutations in plasma without initial tumor genotyping may provide an early indication of tumor recurrence in colorectal cancer (10, 13, 14), but given the genetic heterogeneity of breast cancer, there is an unmet need for mutation-agnostic biomarkers for noninvasive monitoring.

Circulating tumor cells (CTC) are shed from tumors into the bloodstream, where a small percentage may survive, extravasate, and colonize distant sites (15, 16). As such, CTCs offer a noninvasive source of whole tumor cell-derived material for serial analysis during therapy. Microscopy-based enumeration of CTCs has been established as a biomarker in metastatic breast cancer, with both the CTC number at baseline and treatment-induced changes in the number of CTCs being prognostic of progression-free survival and overall survival (OS) in the context of chemotherapy treatment (17, 18). However, to date, the relatively low sensitivity and technological complexity of CTC imaging, combined with the absence of robust molecular characterization, have limited the clinical application of CTCs to guide therapeutic decision-making (19–21).

Recent advances in microfluidics have enabled the enrichment of viable and intact CTCs through the depletion of normal blood cells (22, 23). Microfluidic antibody-based removal of hematopoietic cells, rather than positive capture of CTCs, enables enrichment of CTCs independent of their variable cell-surface epitopes and also ensures CTC cellular integrity and high RNA quality (24–26). To enhance CTC detection signal following such microfluidic enrichment, we recently developed a quantitative RNA-based digital PCR scoring assay, individualized to cancer type-specific markers (27–29). The transcriptional CTC signature takes advantage of tissue lineage-associated transcripts expressed in cancer cells but absent in the normal blood cells present in the CTC-enriched product. We now describe the development of a breast cancer CTC-specific assay, providing both digital quantitation of CTC burden and intracellular ER signaling measurements, and we test its clinical utility in a prospectively monitored cohort of women receiving neoadjuvant therapy for high-risk localized breast cancer and in a second prospective cohort of women with advanced metastatic breast cancer.

RESULTS
Development of a Breast Cancer–Specific RNA Signature for CTC Detection

The microfluidic CTC-iChip achieves approximately 4-log depletion of white blood cells (WBC), red blood cells, and platelets, resulting in an output with 0.1% to 10% CTC purity, depending on initial CTC burden (23, 26). To develop an RNA expression signature capable of detecting breast cancer cells within the background of normal blood cells, we first analyzed RNA-sequencing (RNA-seq) and microarray gene-expression data sets derived from normal breast tissue, breast cancer, and whole blood (Supplementary Fig. S1; Supplementary Table S1). We ultimately selected 17 markers strongly expressed in breast-derived tissues but virtually absent in blood cells (see Methods; Fig. 1A). The markers include breast lineage-specific transcripts (PGR, SCG2B2A1, and PIP) and transcripts highly expressed in breast cancer (MGP and EFHD1), as well as gene products implicated in endocrine signaling (SERPIN2A3 and WFD2), endocrine drug resistance (AGTR2), cancer growth and metastasis (MUC16 and TMRPS4), cellular signaling (FAAT1, FAT2, SFRP1, and SFRP2), epithelial-derived cytokines (CXCL13 and CXCL14), and oncofetal antigens (Prame).

Single-cell RNA-seq confirmed the heterogeneous expression of the 17 markers in 15 individual CTCs isolated from the blood of women with metastatic breast cancer; 5 similarly analyzed single WBCs have negligible expression of these genes (Fig. 1B). For maximal sensitivity and high-throughput capability, we developed a droplet digital PCR assay (ddPCR) for each of the 17 markers, which was applied to the CTC-enriched product after an initial 8-cycle whole transcriptome amplification (WTA; see Methods).

To first test the performance of the assay in reconstitution experiments, we manually micromanipulated 0, 1, 3, 10, or 30 cultured CTCs (cell line BRx-142 derived from a patient with metastatic breast cancer) into 4 mL healthy donor (HD) blood (approximately 20 billion cells), processed these samples through the CTC-iChip, extracted RNA from the CTC-enriched cell population, and digitally quantified the expression of the 17 markers in the product. A robust signal is observed with a single CTC, and signal increases linearly with higher numbers of spiked cells ($R^2 = 0.99$; Fig. 1C; Supplementary Fig. S2). Importantly, the relative expression of detected markers remains consistent from 1 to 30 spiked cells (Fig. 1D). Similar experiments performed using two other breast cancer cell lines [a second CTC-derived line, BRx-68, and the well-characterized triple-negative breast cancer (TNBC) cell line MDA-MB-231] demonstrate differential expression of the 17 markers illustrating the importance of using a diverse panel of cellular transcripts for optimal detection of CTCs (Fig. 1E).

A remarkably similar pattern of expression is observed in the bulk RNA-seq data of these cell lines (Supplementary Fig. S3).

Digital CTC Detection at Different Disease Stages and in Multiple Breast Cancer Subtypes

Having benchmarked the CTC molecular signature assay in vitro, we tested its performance on HDs and in patients with breast cancer. Although we selected breast tissue lineage and breast cancer enriched transcripts for this panel, very low background expression of these markers in WBCs may still occur at levels that are detectable by a highly sensitive technique such as ddPCR. We therefore first applied the assay to an initial cohort of 33 female HDs, establishing normal levels of background for each marker in blood samples without CTCs (see Methods). Consequently, we subtracted this background signal from all further test samples: 20 new female healthy donors (test HD cohort) and a cohort of women presenting with various stages of breast cancer, including untreated (localized presurgical) stage I ($N = 26$), stage II ($N = 42$), or stage III ($N = 12$) breast cancer, and on-treatment samples from women with stage IV (metastatic disease;
Digital RNA Quantitation of Breast Circulating Tumor Cells

81.0

WBC2

WBC3

WBC4

WBC5

their value in a subset of heterogeneous breast cancers. In

expression is completely absent in HD samples, supporting

0.62–0.75) and 4 have lower overall predictive value but their

P

vidual markers (  S4). In patients with metastatic breast cancer, 6 genes show

significant predictive value for the presence of cancer as indi-

S2).

operator characteristic (ROC) analysis (Supplementary Fig.

vidual markers as well as the total CTC score using receiver–

N

A

Figure 1. Development and validation of the breast cancer CTC-ddPCR assay. A, Expression of the 17 selected breast CTC markers in whole blood versus normal breast tissue [GTEx database]. B, Single-cell RNA-seq–derived expression of the 17 breast CTC markers in 5 WBCs and in 15 individually sequenced primary CTCs from women with metastatic breast cancer. C, CTC signal (total transcripts/mL of blood) from 0, 1, 3, 10, and 30 BRx-142 cells introduced into 4 mL HD blood, followed by microfluidic CTC-enrichment and ddPCR analysis ( n = 2; dots represent means; error bars represent SD; best fit line and R-squared statistics of the linear regression model are shown). D, Contribution of individual breast CTC markers to the signal detected from 30 cells each from three different cell lines (BRx-142, BRx-68, and MDA-231) introduced into blood. In all cases, cells were added to 4 mL of whole blood from HD samples and processed through the CTC-iChip for enrichment prior to ddPCR analysis. E, Receiver-operator characteristic (ROC) analysis of total CTC signal in healthy donors ( n = 26) Stage I ( n = 42), stage II ( n = 12), and stage IV ( n = 30) patient samples. AUC values are shown; P values are based on the Wilcoxon rank-sum test.

women with localized breast cancer, no single marker shows statistically significant predictive value, consistent with the generally lower signal in the blood of patients with low tumor burden. Building from these individual markers, the total CTC score achieves a superior, robust, and statistically significant detection capability (AUC = 0.84; P < 0.0001 for metastatic breast cancer; AUC = 0.68, P = 0.0077 for localized breast cancer; Supplementary Fig. S4). As expected, the performance of digital CTC detection improves with increasing

AUG = 0.84

P < 0.0001

women with localized breast cancer, no single marker shows statistically significant predictive value, consistent with the generally lower signal in the blood of patients with low tumor burden. Building from these individual markers, the total CTC score achieves a superior, robust, and statistically significant detection capability (AUC = 0.84; P < 0.0001 for metastatic breast cancer; AUC = 0.68, P = 0.0077 for localized breast cancer; Supplementary Fig. S4). As expected, the performance of digital CTC detection improves with increasing

N = 30) breast cancer. We analyzed the performance of indi-

N = 26) Stage I ( n = 42), stage II ( n = 12), and stage IV ( n = 30) patient samples. AUC values are shown; P values are based on the Wilcoxon rank-sum test.

stage of disease with the AUC values increasing from 0.65 ($P = 0.071$) and 0.68 ($P = 0.015$) in stages I and II, respectively, to 0.82 ($P = 0.0013$) and 0.85 ($P < 0.0001$) in stages III and IV, respectively (Fig. 1F). Sensitivity of the assay at 100% specificity was 19% for stage I, 36% for stage II, 58% for stage III, and 67% for stage IV. Taken together, these results demonstrate that a panel with multiple transcripts surpasses the performance of its individual components and highlight the benefit of multiplex assays for cancer detection in both localized and metastatic breast cancer.

**Persistence of Elevated CTC Score during Neoadjuvant Treatment of Localized Breast Cancer as a Predictor of Residual Disease**

We applied the breast digital CTC score to a cohort of women with localized breast cancer (stages I–III) treated with preoperative (neoadjuvant) therapy (BL-NEO cohort; $N = 54$). Pretreatment baseline and monthly on-treatment blood samples were collected; of the 54 patients, 17% ($n = 9$) had HR+ primary tumors, 46% had TNBC ($n = 25$), and 37% ($n = 20$) had HER2+ subtypes, consistent with the expected distribution among localized disease treated with neoadjuvant therapy (see Table 1 for baseline characteristics of patients). At a set specificity of 100%, baseline CTC scores are positive in 43% of BL-NEO cohort patients (ROC AUC = 0.72, $P = 0.0027$; Supplementary Fig. S5). There was no significant association between baseline CTC score and tumor grade, tumor diameter, and nodal status within this group of patients (Supplementary Fig. S5).

Of the 54 BL-NEO cohort patients, 37 women had their ultimate treatment response determined by both pathology analysis and clinical criteria by a clinician blinded to the CTC score at the time of surgical resection (after approximately 4–6 months of therapy). Patients for whom we did not have clinical or pathologic assessment were not included in further analysis, but all other patients were included. Overall, a trend toward residual disease in patients with a high baseline CTC score (ROC AUC = 0.68, $P = 0.055$) is noted, reflecting the prognostic value of CTCs (Fig. 2).

![Figure 2. Elevated CTC score during presurgical neoadjuvant therapy predicts the probability of residual disease in patients with localized breast cancer at the time of surgical resection. The BL-NEO blood draws were stratified by both treatment cycle (including chemotherapy, endocrine therapy, and/or anti-HER2-targeted therapy) and presence of significant residual disease upon surgery, and their CTC scores were compared. Breast cancer subtypes are noted (HR+, red; HER2+, green; TNBC, blue). High CTC scores in pretreatment and cycle 1 and 2 samples reveal a trend toward presence of significant residual disease, whereas blood draws from ≥3 cycles of therapy predict significant residual disease. ROC curves, AUC values, and $P$ values for each of the conditions are shown. $P$ values were computed by comparing the performance of the CTC score to a random predictor.](image-url)
Serial Monitoring of CTC Score in Women with Metastatic Breast Cancer

We tested the application of the breast digital CTC score for monitoring treatment response in a prospective cohort of women with metastatic breast cancer (TRACK study; \(N=87\)), comparing measurements before initiation of new therapy and then at 3 to 4 weeks after start of treatment. Among the 87 patients, 60 (68%) had HR\(^+\) breast cancer, with 17 (19%) having TNBC and 9 (10%) HER2\(^+\) disease. Baseline clinical characteristics of the women and treatment regimens in this cohort are shown in Table 2.

Table 2. Baseline clinicopathologic characteristics in the prospective study of patients with metastatic breast cancer starting a new line of treatment (TRACK cohort)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Patients (N = 87)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>60</td>
</tr>
<tr>
<td>Range</td>
<td>35–83</td>
</tr>
<tr>
<td>Hormone receptor status at diagnosis, n (%)</td>
<td></td>
</tr>
<tr>
<td>HR(^-)</td>
<td>60 (68%)</td>
</tr>
<tr>
<td>TNBC</td>
<td>17 (19%)</td>
</tr>
<tr>
<td>HER2(^+)</td>
<td>9 (10%)</td>
</tr>
<tr>
<td>Histology at diagnosis, n (%)</td>
<td></td>
</tr>
<tr>
<td>Ductal</td>
<td>70 (79%)</td>
</tr>
<tr>
<td>Lobular</td>
<td>7 (8%)</td>
</tr>
<tr>
<td>Mixed</td>
<td>5 (6%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>6 (7%)</td>
</tr>
<tr>
<td>Current therapy, n (%)</td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>9 (10%)</td>
</tr>
<tr>
<td>Endocrine</td>
<td>38 (43%)</td>
</tr>
<tr>
<td>Anti-HER2</td>
<td>11 (13%)</td>
</tr>
<tr>
<td>Other</td>
<td>29 (34%)</td>
</tr>
<tr>
<td>Types of metastases, n (%)</td>
<td></td>
</tr>
<tr>
<td>Visceral</td>
<td>62 (71%)</td>
</tr>
<tr>
<td>Bone</td>
<td>57 (65%)</td>
</tr>
<tr>
<td>Brain</td>
<td>4 (5%)</td>
</tr>
<tr>
<td>CA15-3 levels, n (%)</td>
<td></td>
</tr>
<tr>
<td>High (&gt;30)</td>
<td>52 (59%)</td>
</tr>
<tr>
<td>Normal (&lt;30)</td>
<td>21 (24%)</td>
</tr>
<tr>
<td>NA</td>
<td>15 (17%)</td>
</tr>
<tr>
<td>Number of prior therapies</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>2</td>
</tr>
<tr>
<td>Range</td>
<td>0–9</td>
</tr>
</tbody>
</table>

The baseline detection rate using the CTC score in this prospective validation cohort, compared with a new matched set of healthy women with negative breast biopsy findings (see Methods), had an AUC = 0.86 (\(P = 0.0001\); Fig. 3A), with 68% sensitivity at a set specificity of 100%, which is comparable to that observed with the initial test metastatic cohort (AUC = 0.84 \(P < 0.0001\); Fig. 1F). Having validated the assay detection rate in this prospective cohort, we dichotomized the pretreatment CTC scores into high (\(\geq 3,000\) transcripts/mL) and low (\(< 3,000\) transcripts/mL) values (see Methods). Correlating the high and low CTC scores with clinicopathologic variables (age, hormone receptor status, histopathology, type of treatment, location of metastatic sites, number of prior therapies, and CA15-3 tumor marker values) shows none of the variables analyzed to be associated with a higher CTC score (Supplementary Table S2). We then examined the relationships between pretreatment or week 3 to 4 CTC scores and clinical outcomes using multivariable Cox proportional hazards model of OS or time to progression (TTP; see Methods). OS was significantly associated with baseline CTC score, prior CDK4/6 therapy, and presence of ESR1 mutation. Patients with a high baseline CTC score (\(> 3,000\)) had worse OS (HR, 2.70; 95% CI, 1.15–16.7; \(P = 0.02\)), compared with those with a low CTC score (median OS, 11.1 months vs. 17.2 months; Fig. 3B). Better OS was also seen for patients who had received CDK4/6 therapy compared with those who did not (HR, 0.23; 95% CI, 0.07–0.74; \(P = 0.01\), consistent with observed benefits for this class of drugs in combination with endocrine agents (1–3). Lastly, the hazard of death was reduced by two thirds in patients who did not have ESR1 mutations (HR, 0.32; 95% CI, 0.11–0.95, \(P = 0.04\); Supplementary Table S3).

Additional univariate analyses indicated that patients who have a low CTC score at baseline that decreases further (>90%) after 3 to 4 weeks of treatment have significantly better OS compared with those who did not experience such a decrease or who started with a high pretreatment CTC score (\(P = 0.001\); Fig. 3C). Together, these observations suggest that both pretreatment CTC burden as well as the early treatment-induced changes in CTCs are important prognostic factors in predicting patient outcome. Such a correlation has been previously demonstrated using microscopic enumeration of CTCs following chemotherapy, but not endocrine treatment (17, 18).

In our cohort of predominantly HR\(^+\) cancers, neither the pretreatment nor early treatment-induced changes in CTC score are predictive of TTP (Supplementary Fig. S6), a finding that is also consistent with the endocrine therapy–treated subset of patients analyzed by classic microscopic CTC enumeration (17). This difference between OS and TTP is likely related to the known heterogeneity in this patient population, variations in endocrine-based therapeutic regimens, and variable clinical assessment of progression in patients with HR\(^+\) breast cancer whose metastases are not readily evaluable (e.g., bone lesions). Levels of the classic protein tumor marker CA 15-3 are not predictive of either OS or TTP (Fig. 3D; Supplementary Fig. S7), confirming the need for other noninvasive biomarkers in HR\(^+\) breast cancer.
Figure 3. Pretreatment CTC score predicts OS in women with metastatic breast cancer receiving a new line of treatment. **A,** ROC analysis of total CTC signal in pretreatment samples from stage IV patients (n = 87) from the TRACK cohort versus matched negative control (women with negative biopsies despite previously positive mammogram findings; n = 10). AUC values are shown; P values are based on the Wilcoxon rank-sum test. **B,** Kaplan–Meier plot depicting OS in TRACK patients, based on pretreatment CTC score. Patients were divided into two groups at a cutoff of 3,000 transcripts/mL (see Methods). Patients with a high pretreatment CTC score (red) have a longer OS compared with those with a lower pretreatment CTC score (blue). Hazard ratio (HR) and P values are based on the multivariable Cox proportional hazards model are shown. **C,** Kaplan–Meier plot depicting OS in TRACK patients, based on the change in CTC score between pretreatment baseline versus 3 to 4 weeks on-treatment time point. Groups are defined based on low signal at pretreatment (≤3,000 transcripts/mL) with >90% reduction in signal on treatment (green), low signal at pretreatment (≤3,000 transcripts/mL) without >90% reduction in signal on treatment (blue), high signal at pretreatment (>3,000) with >90% reduction in signal at 3 to 4 weeks (orange), or high signal at pretreatment (>3,000) without >90% reduction in signal at 3 to 4 weeks on treatment (red). P values were calculated using the log-rank test. **D,** Kaplan–Meier plot depicting OS in TRACK patients, based on CA15-3 levels at pretreatment. Groups are defined as abnormal CA15-3 levels (>30, blue), normal CA15-3 levels (≤30, green), and missing CA15-3 levels (NA, red). P value based on the log-rank test.
Persistent Estrogen Signaling within CTCs Identifies HR+ Cases with Short TTP on Endocrine Therapy

To identify predictive biomarkers for continued response to endocrine therapy in highly treated HR+ breast cancer (i.e., second or later lines of therapy), we focused on the subset of TRACK cohort patients who had HR+ breast cancer (n = 36 at pretreatment) and were starting a new course of endocrine-based treatment (Supplementary Table S4). At a set specificity of 100%, 67% had detectable CTC signal. We first performed unsupervised clustering of the 17 markers before and after initiation of treatment. Remarkably, at the 3- to 4-week on-treatment time point, a 6-gene subset (PIP, SERPINA3, AGR2, SCGB2A1, EFHD1, and WFD2C) identifies patients with rapid disease progression within 120 days of start of treatment (P = 0.005, Fisher exact test; Fig. 4A). This CTC-derived “Resistance Signature” (RS) emerges at the first on-treatment time point; although the 6 CTC transcripts also cluster together at the pretreatment time point, their expression before initiation of therapy is not associated with clinical outcome (Supplementary Fig. S8), raising the possibility that the persistence of this signature on therapy is biologically significant.

All six RS transcripts are significantly enriched in ER+ tumors in The Cancer Genome Atlas (TCGA) database (Supplementary Fig. S9A), suggesting that their expression may be related to estrogen signaling. Indeed, a metascore based on their mean expression shows a significant correlation with the Hallmark estrogen signaling pathway before and after hormone therapy (P = 0.008, consistent with its providing pharmacodynamic evidence of inadequate suppression of ER signaling by the administered endocrine therapy (Fig. 4F and G). Remarkably, there is limited overlap between patients with a high RS score and those with ESR1 mutations: Of the 5 patients with ESR1 mutations for whom we had an on-treatment blood sample, only three have high RS scores (Supplementary Table S6). Conversely, 3 of 13 patients with a high on-treatment RS score harbor an ESR1 mutation. Failure of endocrine therapy to suppress ER signaling in advanced breast cancers is only partially attributable to the acquisition of ESR1-activating mutations, suggesting distinct mechanisms that contribute to refractory disease.

DISCUSSION

By combining initial microfluidic depletion of hematopoietic cells to enrich for intact CTCs, together with quantitative digital PCR of multiple breast-specific transcripts, we have developed a platform for high-throughput, noninvasive characterization of cancer cells in the circulation of women with breast cancer. Compared with imaging-based CTC quantification (15, 16) or analysis of individual RNA markers (36–38), the CTC score has the sensitivity and complexity to interrogate different cancer-related pathways, including ER signaling, during the course of therapy. Indeed, failure of endocrine therapy to suppress intracellular ER signaling within cancer cells sampled in the bloodstream is associated with a very poor clinical outcome and, to our knowledge, it constitutes the first application of transcriptomic-based treatment to predict clinical outcomes in breast cancer.
Figure 4. Markers associated with persistent ER signaling (RS signature) identify HR+ patients at high risk of progression on endocrine treatment. A, Unsupervised clustering of breast CTC marker expression in HR+ patients receiving endocrine therapy for 3 to 4 weeks. A set of markers (boxed) identifies a group of patients (colored blue) significantly enriched for progression within 120 days. P = 0.0051 was calculated using Fisher exact test. ESR1 mutation status for each patient is shown. B, Correlations between a metascore based on the expression of the six high-risk RS genes and gene set enrichment analysis signatures associated with estrogen signaling (top) and endocrine resistance (bottom) across multiple publicly available data sets. Correlations with endocrine resistance signatures based on 100 random sets of six genes are shown in blue circles. C, RS expression based on bulk RNA-seq in tamoxifen (Tam) sensitive or resistant MCF7 cells, left untreated or treated with estrogen (E2) alone or together with tamoxifen (4-OHT). *, P < 0.05; P values based on two-sided t test. D, Kaplan-Meier plots depicting OS in HR+ patients receiving endocrine therapy based on the presence of ESR1 mutations at pretreatment. Cases with ESR1 mutations (red) are compared with those with wild-type ESR1 (blue). P values were calculated using the log-rank test. E, Kaplan-Meier plots depicting TTP in HR+ patients receiving endocrine therapy based on the presence of ESR1 mutations at pretreatment. Cases with ESR1 mutations (red) are compared with those with wild-type ESR1 (blue). P values were calculated using the log-rank test. F, Kaplan-Meier plots of OS of HR+ patients receiving endocrine therapy based on 3 to 4 weeks on-treatment RS score. Groups were divided at 25 transcripts/mL. Patients with high RS CTC score (red) are compared with those having a low RS CTC score (blue). P values were calculated using the log-rank test. G, Kaplan-Meier plots of OS of HR+ patients receiving endocrine therapy based on 3 to 4 weeks on-treatment RS score. Groups were divided at 25 transcripts/mL. Patients with high RS CTC score (red) are compared with those having a low RS CTC score (blue). P values were calculated using the log-rank test.
noninvasive pharmacodynamic monitoring in the therapy of breast cancer. The clinical utility of CTC enumeration was first demonstrated by the observation that baseline measurements of $>5$ CTCs/7.5 mL of blood, assayed by microscopic visualization and scoring, predict adverse OS in women with metastatic breast cancer treated primarily with chemotherapy (17). Women whose CTC counts failed to decline following chemotherapy also had a worse prognosis, although switching to an alternative standard chemotherapy regimen did not lead to a better outcome (18). However, because chemotherapy resistance is broadly displayed against multiple agents, and there is no predictive marker to indicate sensitivity or resistance to specific regimens, this negative result for treatment selection based on CTC monitoring is not surprising. Our digital RNA-based readout supports the prognostic value of CTC measurements at baseline and following initiation of therapy and extends their clinical relevance to women treated primarily with endocrine-based therapies. This clinical population is particularly relevant for blood-based monitoring, because there are currently multiple therapeutic regimens available, without reliable biomarkers to direct treatment selection. It is in this context that interrogation of the ER pathways through CTC gene-expression profiling holds promise for guiding treatment selection.

The 17 genes that constitute the breast CTC signature were selected to include multiple tissue-derived and cancer-related transcripts that are not expressed in contaminating blood cells. As such, the six genes included in the RS subsignature do not represent canonical ER targets but their expression is, nonetheless, highly correlated with both ER signaling and resistance to endocrine therapy. Their persistent expression within CTCs after treatment initiation in patients with metastatic breast cancer identifies women with greatly reduced response to endocrine therapy and worse outcomes. The fact that this CTC signature emerges as a strong predictive factor 3 to 4 weeks after the start of novel ER-targeting therapy suggests that it may reflect the differential effectiveness of drug-mediated ER suppression within tumor cells. In women with highly pretreated advanced HR+ breast cancer, initiation of a new course of endocrine therapy presumably suppresses ER signaling in susceptible cancers, whereas persistent pathway activity remains evident in those where the drug fails to suppress its intended target. Given the small number of cases analyzed in this way, larger studies will be required to confirm the clinical relevance of this observation. However, it raises the possibility that ER activity continues to be an important driver of proliferation in a subset of HR+ treatment-refractory cases and points to a need for a better understanding of underlying mechanisms that could be targeted through alternative agents. A similar rationale underlies the development of novel androgen receptor targeting agents in castrate-resistant prostate cancer (39, 40). Activating mutations in the ligand-binding domain of the ER gene ESRI have been recently identified in advanced HR+ breast cancers, particularly following prolonged treatment with AIs, with reported frequencies as high as 37% (13, 25, 31–33, 41). In addition, rare ESRI translocations, ESRI amplifications, as well as mutations in pathways with substantial cross-talk with ER and in the regulatory regions of ER cofactors have also been reported (42–44). In our study, in the subset of patients initiating a new course of endocrine therapy, 21% had ESRI mutations as determined by CTC-ddPCR and tissue genotyping. Thus, presence of ESRI mutations does not account for all patients who had rapid progression (<120 days) on endocrine agents. Persistent ER signaling at 3 to 4 weeks of treatment with endocrine therapy, as measured by the CTC-derived RS expression signature, emerged as an independent prognostic factor in these patients. Although again these observations need to be confirmed in larger clinical trials, they suggest that persistent drug-refractory ER signaling may not be solely attributable to acquired ESRI mutations, and further support the need to fully define the range of mechanisms driving ER activation and potential strategies to suppress this critical signaling pathway in HR+ breast cancer.

The fact that some ESRI-mutant breast cancers had low RS scores raises the possibility that these mutations confer constitutive ER signaling, yet at relatively lower levels of activity than cases with high on-treatment RS scores. In addition, ESRI mutations are frequently subclonal (11) and multiple endocrine drug-resistance mechanisms may coexist within a single patient. Although our data set is too small to allow detailed analysis of predictive power for both of these markers, it is noteworthy that patients who had a high RS score with ESRI mutation (3 cases) or without ESRI mutation (10 cases) had a median TTP of 56 and 57 days, respectively; in contrast, women with an ESRI mutation and a low RS score (2 cases) had a median TTP of 139 days, and those with neither ESRI mutation nor RS score (10 cases) had a median TTP of 251 days (Supplementary Table S6). Although based on small patient numbers, these data suggest that persistent expression of ER signaling in patients treated with endocrine therapy is an independent risk factor in assessing a patient’s likelihood to benefit from such treatments. Therapeutic targeting of ESRI-mutant protein through novel ER degraders is currently a major focus for drug development (45), and additional strategies may be required to target breast cancers with high RS in the absence of ESRI mutations.

Finally, we explored the application of digital CTC scoring in the neoadjuvant treatment of localized high-risk breast cancer. Setting the test specificity to a stringent level of 100%, positive CTC signal was detectable at baseline in 43% of women whose early-stage breast cancer was considered sufficiently high risk to warrant preoperative therapy (Supplementary Fig. S5). A major challenge in the administration of neoadjuvant chemotherapy and/or hormonal therapy is the absence of early markers of response, such that up to 6 months of treatment may be administered before surgical resection of the primary tumor may reveal either the desired tumor shrinkage or persistent disease. Our finding that after 3 months of neoadjuvant therapy, women who have higher CTC signal will have substantial residual tumor at the time of surgical resection, compared with those with lower CTC signal, suggests that CTC monitoring may help guide the presurgical evaluation of drug response and supports previous evidence for the utility of CTC as a prognostic marker in
the neoadjuvant setting (46). Larger trials will be required to confirm the clinical validity of this promising blood-based predictor, and whether its clinical value varies among different histologic subtypes of breast cancer subjected to different neoadjuvant treatment modalities.

The application of “liquid biopsies” to breast cancer therapeutics is rapidly evolving, in parallel with the advent of novel therapeutic agent and drug combinations. Advances in ctDNA technology now allow for the detection of multiple somatic mutations in plasma, while pushing the limit of sensitivity to earlier-stage disease (11, 14). Here, we have leveraged the high specificity and signal amplification inherent in RNA-based biomarkers to provide an orthogonal assay to plasma genotyping—one that allows for the interrogation of intracellular pathways critical to understanding drug effects. ctDNA and CTCs derive from different processes within the tumor: ctDNA originates from tumor cells undergoing apoptosis or necrosis and thus likely enriches for tumor subpopulations sensitive to treatment, whereas CTCs are live cells that have intravasated into the bloodstream and are likely derived from invasive and potentially drug-resistant subclones. Thus, high-throughput microfluidic enrichment of CTCs followed by multiplex digital RNA quantification may provide a novel and complementary strategy to monitor and guide therapy in both localized and advanced breast cancer.

METHODS

Patients and Healthy Donors

All studies were conducted in accordance with Belmont Report ethical guidelines. Patient samples were collected after written informed consent through an Institutional Review Board–approved protocol for CTC collection (DFHCC 05-300). Ten to 20 mL of peripheral blood was collected. Thirty on-treatment samples from 23 unique stage IV patients were collected for the initial clinical benchmarking of the assay along with 26 presurgical samples from patients with newly diagnosed localized breast cancer (25 stage I and 2 stage II).

Pretreatment samples and samples prior to each subsequent round of neoadjuvant treatment were prospectively collected from women with newly diagnosed localized breast cancer (1 stage I, 41 stage II, and 12 stage III unique patients, BL-NEO cohort; Table 1). The pretreatment samples were used for establishing the performance of the assay; both pretreatment and on-treatment samples from the BL-NEO cohort were used to determine if CTC monitoring of patients receiving neoadjuvant treatment is predictive of surgical outcomes. For quantitation of residual tumor burden after neoadjuvant therapy, we performed a blinded review of the surgical pathology report together with observations of the treating clinician as to whether the neoadjuvant treatment had achieved the desired goal of treatment shrinkage before surgery.

In order to determine if CTC monitoring through the breast CTC-ddPCR assay is predictive of treatment outcome and OS in metastatic patients, we prospectively collected pretreatment and 3 to 4 weeks on-treatment draws from patients with metastatic breast cancer initiating a new therapy (TRACK cohort). At least one sample was collected from each of the 87 patients (baseline characteristics in Table 2). Disease progression was determined by treating physician (blinded to the CTC result) based on standard clinical and/or radiologic criteria.

Thirty-three samples from female healthy donors were obtained from the blood bank (9 mL average) to establish the normal expression of each marker (initial HD cohort). An additional 20 HD samples were obtained from the blood bank to establish the performance of the assay (test HD donors). To validate the detection characteristics established in the initial phase of assay development on the TRACK cohort, we also collected samples from 10 healthy women with negative breast biopsies after suspicious mammogram findings.

Marker Selection

To build the breast CTC assay, we first analyzed publicly available databases, including GTEx, Oncomine, TCGA, and others, and breast CTC sequencing data to identify transcripts with abundant expression in normal breast tissue (lineage markers) or breast cancer (breast cancer markers; Supplementary Fig. S1). We cross-referenced potential markers to WBC gene-expression data that were publicly available or generated by our lab, eliminating transcripts significantly expressed in blood. This approach identified 45 potential markers (Supplementary Table S1), which we tested, first by RT-qPCR, to confirm expression in CTC cell lines and lack of expression within HD-derived WBCs; and then by ddPCR, using WTA-amplified CTC-cChip blood sample products from 10 HDs and 10 patients with metastatic breast cancer. The 17 markers described in this article (specific probes listed in Supplementary Table S7) were chosen based on two criteria: (i) significantly higher expression in the 10 patients versus HDs; or (ii) no expression in the 10 HDs with some expression in patients, and the assay was locked in its current format and further validated in vitro, and in patient and HD samples as described below.

Cell Spike-In Experiments

BRx-142 and BRx-68 CTC cell lines were derived in our lab and have been described previously (25, 47). MDA-231 cells were obtained from the ATCC and have been authenticated using short tandem repeat profiling. All cell lines tested negative for Mycoplasma. Cells were used within 5 to 20 passages from thawing. For the initial in vitro testing of the panel and to determine the linearity of the signal, we micromanipulated increasing numbers of cultured cells into 4 mL of HD blood, ran the samples through the CTC-cChip, and performed RNA extraction WTA and ddPCR as described above.

CTC Score Calculation

To normalize for differences in blood volumes among samples, all raw data were corrected for the blood-volume equivalent used in each ddPCR reaction. To further normalize the signal against HD background, the mean and twice the standard deviation of the expression of each marker within the initial cohort of 33 healthy donors were established. The product of the two values was then subtracted from every patient and HD sample analyzed in this study. If the result was less than 0, it was replaced with 0. The total CTC score was calculated by summing the normalized expression of all markers in a sample without additional weighting and reported as transcripts/mL of blood-volume equivalent used.

ESR1 Mutation Detection

Probes specific for the L536R, Y537C, Y537N, Y537S, and D538G ESR1 mutations have been previously published (13). Their amplification efficiency, as well as that of their respective wild-type probes, was tested on synthetic sequences (data not shown). We established the ability of Y537S to detect mutations present in cDNA from CTC-enriched Inertial Focusing Device product by micromanipulating increasing numbers of BRx-68 cells into HD blood and then processing it as described above. Eighteen-cycle WTA was performed using one third of the extracted RNA with the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech) following the manufacturer’s protocols; 1 µL of undiluted WTA product was used per reaction. Patient samples
were treated in an identical manner; probe specificity was established at 100% after testing at least 5 HD samples per probe. The cutoff for the presence of ESRI mutation was established as >3 positive droplets.

**Statistical Analysis**

ROC analysis was performed to establish the specificity and sensitivity of each marker and the total CTC score for different cancer stages in our initial test cohort consisting of 30 on-treatment stage IV samples, 26 pretreatment stage I samples, and the pretreatment samples from the BL-NEO cohort (42 stage II and 12 stage III) compared with the 20 test HDs. The analysis was performed in R using the ROC package. The specific script is available upon request. Wilcoxon tests were performed to establish significance of the AUC. The specificity and sensitivity in stage IV cancer were validated using a new set of HDs (women with negative findings after a breast mammogram) and the pretreatment samples from the TRACK cohort. All patients, regardless of whether they were defined as CTC positive or negative, were included in downstream analysis, as absence of CTCs is considered meaningful biological information that has prognostic value for clinical outcomes.

For the longitudinal analysis within the BL-NEO cohort, samples for which minimal residual disease or RECIST criteria data were not available were removed from the analysis, resulting in 37 patients analyzed. The CTC score at various clinical intervals (pretreatment and the end of cycle 1, cycle 2, or cycle 3 onward) was compared to the presence of substantial residual disease burden present at the time of surgery, as determined by a clinician blinded to CTC score. The R script used to perform the analysis is available upon request.

Dichotomous cutoff points for baseline CTC score, change in CTC score, and baseline and on-treatment RS score were determined to maximize the associations between high versus low scores and clinical outcomes. All patient samples were included. Cutoff points for the time-to-event endpoints were estimated using leave-one-out jackknife resampling of the algorithm of Cantal-O’Quigley. For each score, the selected cutoff point was the median of the distribution of possible cutoff points. Comparisons of clinical variables between resulting groups are based on Fisher exact tests for categorical characteristics and exact Wilcoxon rank-sum tests for continuous characteristics.

To examine the relationship between pretreatment and on-treatment clinical factors, CTC scores, and outcome, multivariable Cox proportional hazards models were fit for OS and TTP. OS was defined as the interval between the date of initiation of new therapy and first death of any cause. The follow-up of patients who did not die was censored at the date of last assessment of vital status. TTP was defined as the interval between the date of initiation of new therapy and first death from any cause. The follow-up of patients who did not die was censored at the date of last assessment of vital status. TTP was defined as the interval between the date of initiation of new therapy and first documentation of progressive disease. In the absence of documented progressive disease, follow-up was censored at date of last disease assessment. Candidate predictors in the models were factors associated with outcome based on univariate log-rank P values of 0.2 or less: CTC score (either at pretreatment or at 3 to 4 weeks on treatment divided at 3,000 transcripts/mL), breast cancer type at diagnosis (HR+, HER2+, TNBC), prior endocrine therapy (yes/no), prior chemotherapy (yes/no), treatment with CDK4/6 inhibitor (yes/no), presence of visceral metastases (yes/no), presence of bone metastases (yes/no), presence of brain metastases (yes/no), CA15-3 tumor marker levels (normal defined as <30, abnormal defined as >30, missing), number of prior therapies (divided at 2), and age (divided at the median of 60 years). The model of OS was stratified by age to allow for differences in the underlying baseline hazard of death for the two different age categories. Age was a covariate in the TTP model. For on-treatment unsupervised clustering and time-to-event analyses, only patients with samples available at both pretreatment and 3 to 4 weeks on treatment were included. P values for the Kaplan-Meier analyses are based on log-rank tests. Hazard ratios from the Cox models are presented with 95% confidence intervals estimated using log(-log) methods and Wald P values. Unsupervised clustering of pretreatment and 3 to 4-week on-treatment samples was performed using single linkage.

**Disclosure of Potential Conflicts of Interest**

M. Toner is a consultant/advisory board member for Torpedo Diagnostics. D.A. Haber’s institution (MGH) has filed for patent protection for the CTC-iChip and digital signatures of CTCs. S. Maheswaran’s institution (MGH) has filed for patent protection for the CTC-iChip and digital signatures of CTCs. No potential conflicts of interest were disclosed by the other authors.

**Authors’ Contributions**

**Conception and design:** T.T. Kwan, A. Bardia, L.M. Spring, M. Kalinich, M. Toner, D.A. Haber, S. Maheswaran


**Acquisition of data** (provided animals, acquired and managed patients, provided facilities, etc.): A. Bardia, L.M. Spring, T. Dubash, T. Sundaresan, J.A. LiCausi, U. Ho, E.J. Silva, L.V. Sequist

**Analysis and interpretation of data** (e.g., statistical analysis, biostatistics, computational analysis): T.T. Kwan, A. Bardia, L.M. Spring, A. Giobbie-Hurder, M. Kalinich, T. Sundaresan, B.S. Wittner, S. Maheswaran

**Writing, review, and/or revision of the manuscript:** T.T. Kwan, A. Bardia, L.M. Spring, A. Giobbie-Hurder, M. Kalinich, T. Dubash, T. Sundaresan, X. Hong, L.V. Sequist, D.T. Miyamoto, D.A. Haber, S. Maheswaran

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** L.M. Spring, L.V. Sequist

**Study supervision:** M. Toner, D.A. Haber, S. Maheswaran

**Acknowledgments**

We thank all the patients and healthy donors who participated in this study. We acknowledge the clinical research coordinators J.B. Grinnell, P.Y. Chan, and G. Mahvorsa for their assistance with sample and data collection. We are grateful to the Massachusetts General Hospital nurses for their help with the study and to L. Libby for invaluable technical support. This work was supported by NIH grant 2R01CA129933, the Breast Cancer Research Foundation, the Howard Hughes Medical Institute, and the National Foundation for Cancer Research (to D.A. Haber), NIH Quantum Grant 2U01EB012493 (to M. Toner and D.A. Haber), NIH grant U01CA124927 (to M. Toner, D.A. Haber, and S. Maheswaran), MD/PhD Training Grant T32GM007753 (to M. Kalinich), 1-F30CA224588-01 grant (to M. Kalinich), ESSCO Breast Cancer Research (to S. Maheswaran), K12 5K12CA087723 (to A. Bardia), KL2 TR001100 (to L.M. Spring), the ASCO Young Investigator Award (to L.M. Spring), and the Prostate Cancer Foundation (to D.T. Miyamoto). The CTC-iChip and digital signatures of CTCs. No potential conflicts of interest were disclosed by the other authors.

**Received April 23, 2018; revised July 7, 2018; accepted August 8, 2018; published first August 13, 2018.**

**REFERENCES**


A Digital RNA Signature of Circulating Tumor Cells Predicting Early Therapeutic Response in Localized and Metastatic Breast Cancer

Tanya T. Kwan, Aditya Bardia, Laura M. Spring, et al.

Cancer Discov 2018;8:1286-1299. Published OnlineFirst August 13, 2018.