Title

Expanded Genomic Profiling of Circulating Tumor Cells in Metastatic Breast Cancer Patients to Assess Biomarker Status and Biology Over Time (CALGB 40502 and CALGB 40503, Alliance)

Journal

Clinical Cancer Research, 24(6)

ISSN

1078-0432

Authors

Magbanua, Mark Jesus M
Rugo, Hope S
Wolf, Denise M
et al.

Publication Date

2018-03-15

DOI

10.1158/1078-0432.ccr-17-2312

Peer reviewed

HHS Public Access

Author manuscript

Clin Cancer Res. Author manuscript; available in PMC 2019 March 15.

Published in final edited form as:

Clin Cancer Res. 2018 March 15; 24(6): 1486–1499. doi:10.1158/1078-0432.CCR-17-2312.

Expanded genomic profiling of circulating tumor cells in metastatic breast cancer patients to assess biomarker status and biology over time (CALGB 40502 and CALGB 40503, Alliance)

Mark Jesus M. Magbanua[1,*], **Hope S. Rugo**[1], **Denise M. Wolf**[2], **Louai Hauranieh**[1], **Ritu Roy**[3], **Praveen Pendyala**[1], **Eduardo V. Sosa**[1], **Janet H. Scott**[1], **Jin Sun Lee**[1], **Brandelyn Pitcher**[4], **Terry Hyslop**[4], **William T. Barry**[5], **Steven J. Isakoff**[6], **Maura Dickler**[7], **Laura van't Veer**[2], and **John W. Park**[1,*]

[1]Division of Hematology/Oncology, University of California at San Francisco, San Francisco, CA

[2]Department of Laboratory Medicine, University of California at San Francisco, San Francisco, CA

[3]Helen Diller Family Comprehensive Cancer Center and Computational Biology and Informatics, University of California at San Francisco, San Francisco, CA

[4]Alliance Statistics and Data Center, Duke University, Durham, NC

[5]Alliance Statistics and Data Center, Dana–Farber/Partners CancerCare, Boston, MA

[6]Massachusetts General Hospital Cancer Center, Boston, MA

[7]Memorial Sloan Kettering Cancer Center, New York, NY

Abstract

Purpose—We profiled circulating tumor cells (CTCs) to study the biology of blood-borne metastasis and to monitor biomarker status in metastatic breast cancer (MBC).

Methods—CTCs were isolated from 105 MBC patients using EPCAM-based immunomagnetic enrichment and fluorescence-activated cells sorting (IE/FACS), 28 of whom had serial CTC analysis (74 samples, 2–5 time points). CTCs were subjected to microfluidic-based multiplex QPCR array of 64 cancer-related genes (n=151) and genome-wide copy number analysis by array comparative genomic hybridization (n=49).

Results—Combined transcriptional and genomic profiling showed that CTCs were 26% *ESR1−ERBB2−*, 48% *ESR1+ERBB2−*, and 27% *ERBB2+*. Serial testing showed that *ERBB2* status was more stable over time compared to *ESR1* and proliferation (*MKI67*) status. While cell-to-cell heterogeneity was observed at the single cell level, with increasingly stable expression in larger pools, patient-specific CTC expression 'fingerprints' were also observed. CTC copy number profiles clustered into three groups based on the extent of genomic aberrations and the presence of

[*]Correspondence: Mark Jesus M. Magbanua, PhD and John W. Park, MD, Division of Hematology Oncology, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco CA USA 94115, Tel. No. 415 514 3969, mark.magbanua@ucsf.edu; john.park@ucsf.edu.

The authors declare no potential conflicts of interest.

large chromosomal imbalances. Comparative analysis showed discordance in *ESR1*/ER (27%) and *ERBB2*/HER2 (23%) status between CTCs and matched primary tumors. CTCs in 65% of the patients were considered to have low proliferation potential. Patients who harbored CTCs with high proliferation (*MKI67*) status had significantly reduced progression-free survival (p=0.0011) and overall survival (p=0.0095) compared to patients with low proliferative CTCs.

Conclusions—We demonstrate an approach for complete isolation of EPCAM-positive CTCs and downstream comprehensive transcriptional/genomic characterization to examine the biology and assess breast cancer biomarkers in these cells over time.

INTRODUCTION

The biology of metastasis is not well understood, and studies have been hampered by availability of metastatic cells for analysis. Metastatic biopsy is limited by inaccessibility of certain disease sites, procedural risks, and patient discomfort and inconvenience. A non-invasive source of tumor material are metastatic cells detected in the blood, also called circulating tumor cells (CTCs). The accessibility of CTCs, which can be obtained via a blood draw, provides an opportunity for serial analysis to monitor disease and response to treatment (1, 2). Indeed, a major impetus to CTC research is that it can enable “liquid biopsy” to circumvent limitations associated with tissue biopsy. Capturing pure CTCs, however, has proven difficult because they are extremely rare (3). Moreover, nucleic acids derived from these cells are fairly limited, creating yet another obstacle to CTC DNA and RNA characterization (3). Our group has developed an approach for complete isolation and downstream molecular characterization of EPCAM-positive CTCs (4–7). Our isolation method —called IE/FACS (immunomagnetic enrichment/fluorescence-activated cell sorting) — yields CTC samples with minimal leukocyte contamination (4–7). Previous analyses indicated >90% purity of IE/FACS-isolated CTCs (5, 8, 9). Copy number analysis confirmed the malignant nature of these cells and revealed strong clonal relationship between CTCs and corresponding primary tumors, although genetic divergence was also detected (5, 6).

To shed further light on the molecular biology of CTCs, we performed gene expression profiling of CTCs in metastatic breast cancer (MBC) patients. We also conducted, in parallel, genome-wide copy number analysis on duplicate pools of CTCs isolated from the same blood samples. We assessed the status of therapeutic biomarkers (ER and HER2) in CTCs and compared them with matched primary tumors. We analyzed CTCs in serial blood draws to track changes in gene expression, and to evaluate biomarker status over time. We also performed single cell expression analysis to explore tumor heterogeneity in CTCs. Finally, we examined associations between CTC profiles and patient outcomes. To our knowledge, this study represents the largest example of integrated DNA and RNA analysis of CTCs.

PATIENTS AND METHODS

Patient samples and CTC enumeration

Clinical samples were collected between May 2010 to May 2012 from MBC patients enrolled in local (University of California San Francisco, UCSF) and nationwide clinical

trials Cancer and Leukemia Group B (CALGB) 40502 (NCT00785291) and 40503 (NCT00601900), and Translational Breast Cancer Research Consortium (TBCRC) 009 (NCT00483223) (Supplementary Table 1). CALGB is now part of the Alliance for Clinical Trials in Oncology. All patients gave informed consent under a protocol approved by institutional review boards in respective participating centers. Blood was drawn into CellSearch™ CellSave preservative tubes (Veridex) for CTC enumeration (Figure 1a). Additional volumes of blood were drawn into tubes containing EDTA for CTC isolation. Blood samples from other institutions were shipped overnight to the Park Laboratory at UCSF and processed immediately for CTC enumeration. Previous studies have shown that CTC numbers obtained from CellSearch™ and IE/FACS (described below) were highly concordant (2, 10). To determine whether a blood sample contained CTCs for isolation, we first screened blood samples using the CellSearch™ assay. Blood samples with ≥8 CTCs/7.5mL (~ ≥1 CTC/mL) were subjected to IE/FACS. CTC isolation was performed between 24–36 hours after blood draw. Both CTC enumeration (CellSearch) and isolation (IE/FACS) were performed in the Park Laboratory at UCSF. Detailed descriptions of the methods are provided in the Supplementary Information.

Cell isolation by IE/FACS

CTCs were isolated via IE/FACS as previously described (4–6) (Figure 1b). Briefly, magnetic beads coated with EPCAM monoclonal antibody were used to enrich for tumor cells. The tumor-enriched samples were then subjected to FACS analysis using differentially labeled monoclonal antibodies to distinguish tumor cells (nucleated, EPCAM+/CD45−) from leukocytes (nucleated EPCAM−/CD45+) during cell sorting. IE/FACS allowed for the isolation of single or small pools of cells, which could then be subjected to downstream molecular analyses.

Panel of 64 genes for transcript analysis

We used a panel that has been previously validated for gene expression analysis of rare tumor cells (11). The list includes *EPCAM* and *PTPRC* (encodes CD45), which are markers for epithelial cells and leukocytes, respectively. Also included are clinically relevant cancer genes (e.g., *ESR1* and *ERBB2*), stem cell and epithelial-mesenchymal transition (EMT) markers, and candidate reference genes for data normalization (Fig. 1C). The complete list can be found in Supplementary Table 2.

Taqman Low-Density Array Quantitative PCR (aQPCR) analysis

For multiplex aQPCR analysis, we used a custom microfluidic card (384-well format) containing two sets of 64 Taqman probes printed in triplicate. Details of the cell lysis, reverse transcription, specific transcript (cDNA) preamplification, and aQPCR analysis are described in the Supplementary Methods. Results of performance evaluation of the aQPCR assay are presented in the Supplementary Results and Supplementary Figure 1a–e.

Reference genes

To select the reference genes for normalization, we used the geNorm algorithm within RealTime StatMiner® (see below) to calculate the gene stability measure (M) for candidate

genes: *ACTB*, *GAPDH*, *GUSB*, *RPLP0*, *TFRC*, and *RPS18*. Genes *ACTB* and *RPS18* were chosen because they showed the lowest M values indicating least variable expression across 105 CTC and 76 leukocyte samples (Supplementary Figure 1a).

Quality controls

Prior to aQPCR analysis, preamplified cDNA samples were screened for expression levels of *ACTB*, *GAPDH*, and *RPS18* using conventional QPCR. Results of initial experiments revealed that samples with low transcript levels of *RPS18* (Cycle threshold, Ct ≥26) (12) and/or *ACTB* and *GAPDH* (Ct >36) resulted in a failed aQPCR analysis, indicating insufficient quantity and/or poor quality RNA. Samples with <20% detection rate (i.e., detection of ≤12 of the 64 genes by aQPCR analysis) and those with undetectable expression (missing values) of references genes, *RPS18* and *ACTB,* were excluded from the analysis.

Single cell gene expression analysis

Single cells were sorted by IE/FACS and subjected to expression profiling. Transcript levels for a subset of 32 genes were measured via the Fluidigm™ method following manufacturer's instructions (Supplementary Table 2). Details of the protocol are described in the Supplementary Methods.

Array comparative genomic hybridization analysis (aCGH)

CTCs collected from a single time point were analyzed by aCGH in 49 of the 102 patients in the study. Genome-wide copy number aberrations were assessed using bacterial artificial chromosome aCGH, as previously described in detail (5, 13). Microarray data was subjected to circular binary segmentation using the DNAcopy package in Bioconductor (14, 15). Details of the computational methodologies for copy number analysis including the use of the Nexus 6.0 software (Biodiscovery) are described in the Supplementary Methods.

Gene expression analysis

We used the RealTime StatMiner® (version 4.2) to analyze gene expression data as previously described (11). The mean Cts for *ACTB* and *RPS18* were used to calculate the ΔCts. Cluster analysis was performed using unsupervised ward linkage hierarchical clustering methods with Pearson correlation distance as a similarity measure. Multidimensional scaling analysis was also performed to visualize similarities in gene expression profiles among samples. Differentially expressed genes between two groups were elucidated using a parametric analysis (Limma) for unpaired samples, and a paired T-test for paired samples. To adjust for multiple comparisons, the Benjamini-Hochberg (BH) correction method was used (16), and an adjusted p value of <0.001 was considered statistically significant. Relative quantification (RQ) was reported in the logarithmic scale ($\log_{10}RQ=\log_{10} 2^{-\Delta\Delta Ct}$). A $\log_{10}RQ=0$ indicated no differential expression, $\log_{10}RQ=1$ or −1 indicated that a gene is expressed 10 times or 1/10 as much in the test sample relative to the calibrator sample, respectively. Percent detection for each of the 64 genes in CTC and leukocyte samples was compared using a two-tailed Z-test. A p value of <0.05 was considered statistically significant.

Serial expression analysis

To determine whether CTC samples from the same patients exhibit a unique gene expression signature, we calculated the mean Pearson and Fisher-transformed Pearson correlation (17, 18) as well as the mean Euclidean distance between samples from the same patient and for all other patients. Patient IDs were then randomly permuted 10,000 times and the mean distance between samples per 'simulated patient' was calculated. A p-value was calculated using the Fisher transformation (17) for actual and simulated random data to test whether serial CTC expression profiles from the same patients were more similar to one another compared to those from other patients.

Assessment of *ESR1*/*ERBB2*/*MKI67* status and intrinsic subtype analysis

We determined *ESR1* status in CTCs using gene expression data. The bimodal distribution of *ESR1* expression allowed for dichotomization into positive and negative groups using the local minimum between modes as a cutoff point. For *ERBB2* status, CTCs were classified as *ERBB2*-positive based on aCGH defined copy number gains or amplification of the *ERBB2* locus. For CTC samples with no aCGH data (n=50), *ERBB2* status was dichotomized based on a simple aQPCR gene expression-based classifier; this was derived by receiver-operating characteristics curve analysis (Youden index) of a subset of samples with both gene expression and aCGH data (n=101) (Supplementary Figure 1f). For proliferation status, we used *MKI67* mRNA expression as a surrogate marker for Ki67 protein expression. We defined the top tertile (threshold=66% quantile) as 'high' and the remainder as 'low'.

Intrinsic subtype was determined using the 16 PAM50 classifier genes present in the assay. We used a previously validated program and algorithm described by Parker and colleagues to make subtype assignments (19). Subtype calls were filtered using confidence measures produced by the program, only samples with >80% confidence were reported (i.e., 69/105 (66%) of patient CTC samples). To validate our approach, we compared the intrinsic subtype calls using all of the PAM50 genes vs. the subset of 16 using the dataset available at https://genome.unc.edu/pubsup/breastGEO/PAM50.zip, and observed 80% concordance (20).

Analysis and visualization of the data was performed using packages in R (21).

Survival analysis

We examined association between with CTC profiles and patient outcomes. The endpoints included progression-free survival (PFS), defined as time from study entry to documented progression or death from any cause, and overall survival (OS), defined as time from study entry to death from any cause. PFS and OS were estimated using the Kaplan-Meier method (22), and the log-rank test (23) was used to compare survival between groups, with $p<0.05$ being considered statistically significant. The survival data used to assess outcome were collected by the Alliance Statistics and Data Center. Statistical analyses were performed on all data available as of July 18, 2016.

RESULTS

Development and performance testing of CTC assays

DNA copy number profiling of CTCs was performed using our previously validated approach of whole genome amplification of CTC genomic DNA followed by aCGH analysis (5, 11).

For gene expression analysis, we developed a protocol involving microfluidic-based multiplex QPCR array (aQPCR) to analyze the transcript levels of 64 cancer-related genes. The method involves reverse transcription of total RNA, preamplification of specific cDNA transcripts and aQPCR analysis. A detailed discussion of the assay optimization and performance analysis is presented in the Supplementary Materials and results are displayed in Supplementary Figure 1b–e.

Proof-of-concept experiments using spike-in models

The performance of IE/FACS and aQPCR assay was further evaluated using spiked cancer cell lines (n=6) into healthy blood or cell admixtures (Supplementary Figure 2). Gene expression analysis of captured cells revealed high purity and specificity of IE/FACS isolation, and the robustness of the preamplification/aQPCR protocol (see Supplementary Materials for details).

Analytical validation of the CTC gene expression assay in clinical samples

We applied our optimized protocols to characterize CTCs from a prospective series of MBC patients who had provided informed consent in different multicenter clinical trials. The workflow of the study is outlined in Figure 1a. Starting from 244 blood samples from 162 consecutive patients, we isolated ~10–20 CTCs by IE/FACS and performed gene expression analysis (Figure 1b–c, Supplementary Table 1). Of the 244 samples, 93 samples were excluded due to low CTC yield (n=46), poor RNA quality (n=37), or failed aQPCR experiment (n=10) (Supplementary Figure 3). The remaining 151 (61%) samples from 105 unique patients yielded high quality expression data and were used for subsequent analysis. The dataset also included additional 46 CTC expression profiles collected at later time points from 28 patients (see below).

Leukocytes were isolated from the same blood samples, and profiled in parallel as non-tumor controls. Matched leukocyte expression data was available for 76 of the 105 patients, 29 patients lacked data because the samples had poor quality RNA.

Cluster analysis showed that CTCs grouped among themselves and away from normal blood (Figure 1d). Multidimensional scaling analysis to visualize expression data from all CTC (n=151) and leukocyte samples further confirmed that CTC expression profiles were clearly distinguishable from those from those of leukocytes (Figure 1e)

Transcripts coding for *EPCAM, CD24*, *KRT19, ERBB2, CCND1, CAPG, BAG1,* as well as reference genes, including *ACTB*, *RPS18*, *RPLP0*, *GAPDH*, and *GUSB,* were frequently detected (>90%) in CTCs (Supplementary Figure 4a).

Expression of EMT, stem cell, and other cancer-related genes in CTCs

To determine differentially expressed genes between CTCs and leukocytes in 76 paired samples (BH corrected $p<0.001$), we compared the relative expression (ΔCt) of genes between the two groups with leukocytes as the calibrator. Consistent with our epithelial cell isolation approach strategy, CTCs exhibited up-regulation of *EPCAM* and down-regulation of *PTPRC/CD45* (Figure 1f). Other breast cancer-related genes including *CCND1*, *KRT7*, *MUC1,* and *TFF3* were up-regulated while *CD68, a* monocyte/macrophage marker, was down-regulated. Expression of EMT (*SNA11:* 51%, *TWIST1:* 22%, *CAV:* 24%, and *VIM:* 78%) and stem cell markers (*CD44:* 88%, *CD24:* 100%, and *ALDH1A1:* 22%) were observed in a subset of CTCs (Supplementary Figure 4b). However, when compared to leukocytes, all four EMT markers were down-regulated in CTCs, as were stem cell markers *CD44* and *ALDH1A1*; while *CD24* was up-regulated (Supplementary Figure 4c). Taken together, our results with isolated EPCAM-positive CTCs do not indicate epithelial/mesenchymal or stem cell-like phenotypes.

Receptor and intrinsic subtype analysis

We classified CTC samples from 105 patients into groups according to receptor and intrinsic subtypes. *ESR1* displayed a bimodal distribution allowing dichotomization into positive and negative groups (Figure 2a). *MKI67* mRNA expression, which also showed a bimodal distribution, was used as a surrogate proliferation marker for Ki67 expression (Figure 2b). For *ERBB2* status, we used both copy number and expression data (see **Methods**).

Approximately 70% of the CTC samples were considered *ESR1*-positive, of which 48% were *ESR1*+*ERBB2*− and 22% *ESR1*+*ERBB2*+ (Figure 2a). *ERBB2*-positive (and either *ESR1*-positive or negative) and *ESR1*−*ERBB2*− CTCs accounted for 27% and 26% of the samples, respectively. 65% of the CTC samples were considered to have low proliferative (*MKI67*) status. Of the 105 samples, 69 were assigned intrinsic subtype calls using the 16 PAM50 genes present in the assay (see **Methods**). These samples were subdivided into 30% luminal A, 6% luminal B, 13% HER2-enriched, 33% basal, and 12% normal-like subtypes. As expected, we found a significant association between intrinsic and receptor subtype calls (Supplementary Figure 5a, chi-squared test, $p=0.00055$).

Clustering analysis using the 16 PAM50 genes revealed three major clusters (Figure 2b). The leftmost cluster (Cluster 1) included CTC samples comprising mostly basal and/or *ESR1*−*ERBB2*− profiles with high proliferative status. This cluster contained CTCs with matched primary tumors that were also ER-negative. Luminal B CTCs were found only in Cluster 2, while luminal A CTCs were found in both Clusters 2 and 3, and so were *ESR1*+*ERBB2*− and *ESR*1+HER+ receptor subtypes. As expected, similar results were obtained when all 64 genes were used for clustering analysis (Supplementary Figure 5b).

Monitoring CTC-based biomarker status over time

To assess the feasibility of serial CTC expression analysis over time, we isolated and profiled CTCs in 74 serial blood samples from 28 patients. Blood was collected from each patient 2 to 5 times at intervals ranging from 17 to 307 days from the first sample collection (T1). Visual inspection of the expression profiles of CTC samples from individual patients

showed fluctuations in expression at the individual gene level. However, intrinsic and clinical biomarker based phenotype assignments were generally consistent across time points (Figure 3a, Supplementary Figure 6a). 100% (28/28) of patients who were assessed for CTCs at multiple time points showed consistent *ERBB2* status (by aCGH/aQPCR) at all time points sampled. *ESR1* and proliferation (*MKI67*) status (by aQPCR) were less consistent (2-sample proportion test: p=0.0036 and p>0.001, respectively), with 32% (9/28) of patients showing a change of *ESR1* status over time and 54% (15/28) exhibiting variable proliferation status.

To quantify changes in gene expression in serial CTC samples, we calculated the correlation between the first two samples (T1 vs. T2) from each patient (n=28), while noting the time interval between the two blood draws. We observed that CTCs collected and analyzed further apart in time—i.e., greater number of days elapsed between T1 and T2—showed lower Pearson correlation than those with shorter intervals between sample collections ($r=-0.434$, p=0.021, Figure 3b, Supplementary Figure 6b). To extend this analysis, we examined the correlation of T1 samples with all samples from individual patients, including those collected at later time points. We observed that, in general, the median Pearson correlation of patients' samples decreased across time points, although this pattern was not statistically significant (Figure 3c, Supplementary Figure 6c). Using the information above, we estimated an average Pearson correlation change of −0.31/year.

Visual examination of the expression profiles showed similarities in gene expression in serial CTC samples from individual patients. To test whether serial CTCs from a particular patient (ranging from 2 to 5 samples) were more similar to one another than to CTCs from other patients, we compared the correlation between metachronous samples from individual patients vs. correlations between randomly paired samples from a large simulated dataset (generated by permuting sample labels 10,000 times). This comparative analysis revealed that correlation among samples from the patient data (Fisher transformed mean $\rho=0.66$) were significantly higher compared to those from the simulated dataset (Fisher transformed mean $\rho=0.80$), indicating that CTC samples from individual patients were more related to each other than to randomly selected samples (p<0.0001; Figure 3d, Supplementary Figure 6e). Moreover, visualization by multidimensional scaling analysis in four patients with the highest number of serial samples showed that CTCs from the same patient over time are more similar to one another than to samples from other patients (Figure 3e). Taken together, these observations indicate the reproducibility of our assay, and suggest that CTC expression profiles can be obtained from individual patients.

Single cell gene expression analysis reveals heterogeneity among individual CTCs

To assess tumor heterogeneity at the single cell level, we performed expression profiling on individual CTCs by multiplex aQPCR analysis using a microfluidic-based dynamic array (Fluidigm™). Pooled and single CTCs were isolated using IE/FACS from two MBC patients: 4042 (Figure 4a) and 4043 (Figure 4b). Expression analysis revealed higher levels of heterogeneity observed among single cells compared to pooled cells (Figure 4c). Despite the heterogeneity, some genes (e.g., *ESR1*) did exhibit more consistent expression across all samples. Multidimensional scaling (Figure 4d) and clustering analysis (Figure 4e) showed

that CTCs from each patient displayed a unique gene expression signature, consistent with results from serial analysis of CTC samples. In addition, the expression signature seemed to be conserved down to individual cells, as single cell samples from each patient clustered together (Figure 4d).

Copy number analysis of CTCs reveals aberrations frequently seen in primary breast cancers

Duplicate pools of CTCs were isolated from the same tumor-enriched blood samples in 49 of the 105 patients for aCGH profiling. Genome-wide copy number analysis detected genomic aberrations in all of our CTC samples (Figure 5a; Supplementary Figure 7a). Frequent aberrations included gains in 1q and 8q, as well as losses in 8p and 16q (Supplementary Table 3). Gene amplifications were also detected for oncogenes, such as *CCND1* (31%), *ERBB2* (12%), and *MYC* (29%). Similar recurrent aberrations were observed when compared to a publicly-available dataset for primary breast tumors, consistent with our previous findings (Supplementary Figure 7b) (24).

Cluster analysis of copy number data revealed three distinct groups (Figure 5b). Examples of CTC copy number profiles for each cluster are shown in Supplementary Figure 7c. Cluster 1 contained CTCs that exhibited significantly less genomic aberrations compared to CTCs in Clusters 2 and 3 (Supplementary Figure 7d). Cluster 2 included CTCs with 8q gain, while Cluster 3 included those with 1q gain/11q loss. Interestingly, CTCs with ESR1−*ERBB2*−/basal phenotype were mostly observed in Cluster 1 containing CTC samples that display low genomic instability. Associations between genomic aberrations detected in CTCs vs. patient outcomes were not analyzed due to the limited sample size.

ERBB2 copy number and expression in CTCs

Our previous studies have shown that CTC aCGH analysis can accurately detect copy number changes involving *ERBB2* (5, 11). We therefore performed aCGH analysis of the CTCs in the present study, and observed *ERBB2*-positive and *ERBB2*-negative CTCs (Figure 5c).

Other studies have shown that *ERBB2* mRNA expression and copy number in primary breast cancer are highly concordant (25, 26). To evaluate concordance in CTCs, we compared *ERBB2* expression and copy number from 49 patients in this study and an additional 88 patients from our previous work (5). Comparative analysis showed that the mean expression levels of *ERBB2* in *ERBB2*-positive CTCs (by copy number) was numerically higher compare to *ERBB2*-negative CTCs; however, the agreement between copy number and expression data was not strong (Figure 5D). Also, broad ranges of *ERBB2* expression levels was observed in both *ERBB2*-positive and *ERBB2*-negative CTCs (by copy number), but more so in the latter group.

Biomarker status in CTCs and matched primary tumors

To investigate whether biomarkers relevant to breast cancer can change during disease progression, we compared the *ESR1* (by aQPCR) and *ERBB2* (by aCGH) status in CTCs with clinical ER and HER2 status of matched primary tumor. Comparative analyses between

CTCs and corresponding metastases were not performed because metastatic tissue was not available. Also, we did not analyze gene expression in matched primary tumors.

Of the 105 patients in this study, 70% harbored *ESR1*-positive CTCs. ER status for 102 corresponding primary tumors was available for comparison, of these 85% were ER-positive. ER status was in agreement in 73% (74 of 102) matched samples (Figure 5e). Conversely, 25% (22 of 87) of patients with ER-positive primary tumors showed *ESR1*-negative CTCs, while 40% (6 of 15) of patients with ER-negative primary tumors had *ESR1*-positive CTCs.

Assessment of *ERBB2* status using available aCGH data revealed that CTCs from 26% (35 of 137) of the patients showed copy number gains, and therefore were considered *ERBB2*-positive. Comparison of *ERBB2*/HER2 status between 130 matched CTCs and primary tumors revealed a concordance of 77% (100 of 130) (Figure 5f). Of note, *ERBB2*-positive CTCs were detected in 24% (30 of 127) of HER2-negative primary tumors.

CTC profiles vs. patient outcome

We examined the correlation between CTC biomarker status and patient outcome. Follow-up period was approximately 40 months. There were no significant differences in OS between patients with *ESR1*-negative CTCs vs. those with *ESR1*-positive CTCs (Supplementary Table 4, Supplementary Figure 8a), and patients with *ERBB2*-negative CTCs vs. those with *ERBB2*-positive CTCs (Supplementary Figure 8b). We did observe that patients harboring CTCs with a more aggressive subtype (i.e., *ESR1*-negative or *ERBB2*-positive) had numerically shorter survival compared to those with less aggressive phenotype.

Survival analysis based on receptor subtype (combined *ESR1* and *ERBB2* status) revealed that patients with *ESR1*−*ERBB2*+ CTCs have significantly shorter PFS (log-rank p=0.015) compared to other subtypes (Supplementary Figure 9a). No significant differences were observed for OS (log-rank p=0.68). Patients with basal-like CTCs showed a trend towards shorter PFS (log-rank p=0.1053) and OS (log-rank p=0.1322), which was not statistically significant (Supplementary Figure 9b). Analysis based on proliferation status revealed that patients with CTCs showing high expression of *MKI67* (encodes the proliferative marker Ki67) have significantly shorter PFS (log-rank p=0.0011, Figure 6a) and OS (log-rank p=0.01, Figure 6b) compared to patients with CTCs that express the gene at low levels.

ESR1 status in CTCs and ER status in primary tumors vs. patient outcome

We also investigated whether combining biomarker status in primary tumors and CTCs was correlated with patient outcome. Stratification based on ER status revealed that patients whose CTCs and primary tumor were both *ESR1*−/ER− had significantly shorter PFS (log-rank p=0.01, Figure 6c) and OS (log-rank p=0.0035, Figure 6d) compared to patients who were *ESR1*+/ER+ in both compartments. In patients where *ESR1*/ER status in primary tumor and CTCs changed from positive to negative or vice versa, the PFS and OS were similar to survival probabilities based on the original ER status of the primary tumor. Survival analysis based on *ERBB2*/HER2 status was not performed due to the limited number of evaluable HER2-positive patients in the study.

DISCUSSION

Genomic analysis of CTCs has been limited by technical hurdles including isolation of CTCs away from contaminating leukocytes and limited amounts of recoverable tumor DNA/RNA (3). Nevertheless, initial studies of CTC-enriched samples have been performed (27–35). Attempts to isolate individual tumor cells by micromanipulation or laser microdissection, although feasible, can be technically challenging (36–41). To address these issues, we have used an immunomagnetic enrichment/FACS-based (IE/FACS) approach for direct isolation of CTCs, which are then amenable to detailed molecular analysis (4–7), including transcriptional profiling in the present study.

Differential expression analysis confirmed that CTCs isolated by IE/FACS were clearly distinct from normal blood cells. Consistent with our EPCAM-based purification approach, CTCs displayed high expression of *EPCAM* (relative to normal blood), while hematopoietic markers (e.g., *PTPRC/CD45* and *CD68*) were down-regulated. Detection of *MUC1*, *ESR1* and *ERBB2* expression in CTCs was consistent with the epithelial origin of these cells, while the up-regulation of known oncogenes, e.g., *CCND1* and *CCNB1*, provided evidence of malignant transformation.

The EPCAM-positive CTCs captured by IE/FACS did not exhibit dual epithelial-mesenchymal properties, nor the $CD44^{+}/CD24^{-/low}$ ALDH1+ phenotype attributed to breast cancer stem cells (42). These results are in contrast with previous reports showing EMT- and stem cell-like characteristics in EPCAM-positive CTCs (43–45). Those studies, however, were performed on CTC-enriched samples and not on isolated CTCs, and expression levels were not normalized to those of blood cells. Other studies involving direct isolation of EPCAM-negative CTCs suggest that these cells can display markers consistent with EMT (46) or cancer stem cells (47).

Clinical tests to evaluate biomarker status in breast cancer typically consist of immunohistochemical (IHC, e.g., ER, PR, HER2, Ki67) and fluorescence in situ hybridization (FISH, e.g., HER2) assays. While other studies have described similar immunocytochemical and FISH based assays of CTCs (for review, see (48)), we used transcriptomic panels and genome wide copy number analysis in this study in order to demonstrate the feasibility of expanded multigene profiling of CTCs. Therefore, we analyzed standard biomarkers in CTCs using alternative aQPCR (*ESR1*, *PGR*, *ERBB2*, *MKI67*) and aCGH (*ERBB2*) analyses.

Subtyping of receptor status in CTCs revealed frequencies (*ESR1*-positive: 70% positive and *ERBB2*-positive: 27% positive) similar to those observed in primary breast cancers (49). 35% of the patient CTC samples were considered to have high proliferation (*MKI67*) status. Interestingly, 16 (70%) of the 23 basal-like CTCs (*ESR1*−*ERBB2*−) showed high *MKI67* expression, indicating high proliferative status in these cells.

Serial expression analysis showed that *ERBB2* status in CTCs remained unchanged across time points, while *ESR1* and proliferation (*MKI67*) status were less stable. Moreover, the good correlation observed among serial CTCs samples from individual patients indicate the reproducibility of our assay.

Copy number analysis corroborated our previous findings that IE/FACS captures CTCs that are unequivocally malignant cells [5]. Genomic aberrations frequently found in primary breast tumors, including 1q/8q gain, 8p/16q loss, *MYC*, *CCND1* and *ERBB2* amplification (24), were also observed in CTCs. Cluster analysis of CTC copy number profiles revealed three major groups. One cluster displayed less genomic alterations compared to the other clusters. Large chromosome imbalances, like 8q gain and 1q gain/11q loss, appeared to underlie the separation of the latter two clusters. Serial copy number analysis was not performed in this study. We previously reported that copy number profiles of CTCs obtained at different time points from individual MBC patients display unequivocal relatedness, although genetic divergence was also observed (5).

Quantitative studies of single cell expression analysis have shown that stochastic noise can arise from interrogation of low copy number transcripts (50, 51), which represents a potential limitation of the present study as well. To minimize this noise in our single cell profiling, we chose a subset of the 64 genes that were highly expressed in CTCs based on our initial aQPCR experiments in CTC pools from 102 patients. Results from our single cell expression analysis revealed increased heterogeneity in single CTCs compared to pooled CTCs. Despite this variability, we observed that expression profiles of single CTCs analyzed from two patients clustered away from each other (Figure 4d–e), suggesting the overall consistency of patients' CTC gene expression profiles at the single cell level, as well as the reproducibility of our assay.

Exploratory survival analysis revealed that proliferation status in CTCs was correlated with clinical outcome. Patients with CTCs that express *MKI67* at high levels displayed inferior survival probabilities compared to those with CTCs with low level *MKI67* expression. It has also been reported that castration-resistant prostate cancer patients harboring Ki67-positive CTCs have significantly shorter survival compared to patients with Ki67-negative CTCs (52). We observed that patients with *ESR1*−*ERBB2*+ CTCs had inferior PFS compared to patients harboring CTCs of other receptor subtypes, although the very small sample size of this subgroup precludes any conclusions.

Studies comparing ER and HER2 status between primary breast cancer and corresponding metastatic tumors have found that biomarker status can shift during the course of the disease (53, 54). In this study, we observed that *ESR1*/ER and *ERBB2*/HER2 status between CTCs vs. matched primary tumors changed in about 25% of the patients. This discordance may be due to several factors, such as tumor heterogeneity, subclonal selection and expansion, as well as variability in assays for biomarker testing (55).

Existing guidelines by the National Comprehensive Cancer Network recommend re-biopsy and retesting of receptor status in recurrent tumors (55). The clinical utility of CTC-based assessment of receptor status is currently being investigated (56, 57). An advantage of CTCs over conventional tissue-based biopsy is that the status of clinically important biomarkers can be monitored serially to help guide treatment decisions in real-time. Information regarding changes in biomarker status, especially those involving a shift from negative to positive, can potentially aid in therapeutic decisions in the clinic.

Clinical trials are currently evaluating whether HER2-targeted therapy provides benefit in patients with conventionally HER2 negative status derived from primary tumor testing but with HER2-positive CTCs (56, 58). We observed some CTCs with *ERBB2* overexpression but no *ERBB2* copy number gain or amplification. It is possible that HER2-targeted therapy could also provide benefit in these patients.

In our study, we observed that the change in ER status between primary tumor and CTCs did not appear to affect patient outcome, i.e., patients with ER-negative primary tumors had significantly shorter survival compared to ER-positives, regardless of the *ESR1* status in CTCs. The impact of CTC biomarker information on prognosis cannot be fully appreciated in this study as treatment regimens were not changed based on biomarker status in CTCs. Moreover, this study consisted of correlative science performed in conjunction with clinical trials incorporating various therapeutic regimens for MBC (59–61). The study was not designed nor powered to evaluate effects of specific therapies on CTC gene expression.

A limitation of our approach is that it utilizes EPCAM-based methods for enrichment and purification of CTCs, and therefore may fail to detect EPCAM-negative CTCs such as those undergoing EMT. While results from numerous studies have now shown that EPCAM-positive CTCs are strongly and unequivocally associated with poor prognosis—including a recent meta-analysis involving 1,944 patients from 20 studies (62)—the clinical significance of EPCAM-negative CTCs has yet to be established.

Assessment of CTC-based biomarkers can potentially better inform treatment decisions and improve cancer care, as these cells may be more relevant to the current disease compared to the primary tumor. Furthermore, comparative gene expression profiling of CTCs and matched primary tumors/metastases can provide important information regarding clonal-relatedness and tumor heterogeneity/evolution during disease progression. In principle, molecular profiling of CTCs via copy number and expression analyses as described here could be complemented by next generation sequencing to identify potentially actionable mutations in candidate genes. We have recently reported the feasibility of whole genome sequencing of CTCs for this purpose (8).

CONCLUSIONS

We demonstrate the feasibility of direct isolation and expanded molecular profiling of EPCAM-positive CTCs. Our approach involving combined copy number and gene expression profiling of CTCs enabled the assessment of biomarkers relevant to breast cancer. We also demonstrated the feasibility of monitoring CTC gene expression and biomarker status over time. Assessment of breast cancer biomarkers in CTCs has the potential to help guide therapeutic decisions in the face of evolving or progressing metastatic breast cancer. However, for CTCs to be used in the context of personalized medicine, the clinical utility of CTC-based biomarkers has yet to be demonstrated (63).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Alliance for Clinical Trials in Oncology (formerly Cancer and Leukemia Group B) for funding and providing most of the patient samples (CALGB 40502 and 40503). We also thank the Alliance Statistics and Data Center and the Breast Correlative Science Committee for critical review of the manuscript. We thank Sarah Elmes (UCSF Laboratory for Cell Analysis), Kathryn Thompson, Jennifer Dang and Kirsten Copren (UCSF Genome Core) for the technical assistance. We thank Caroline Dando, Florence Ng and Amy Hamilton (Fluidgim Corp.) for processing the samples for single cell analysis. We thank Nicholas Eng and Hannah Fuchshuber for assisting in the review of CellSearch™ data. We thank the HDFCCC Computational Biology Core for assisting in bioinformatic analysis. Many thanks to Prithi Polavarapu for critical review of the manuscript.

Funding: Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers U10CA180821 and U10CA180882 (to the Alliance for Clinical Trials in Oncology), U10CA138561, U10CA180791, and U10CA180867. Also supported in part by funds from the National Cancer Institute Cancer Center Support Grant (5P30CA082103). Additional funding was provided by the Breast Cancer Research Foundation (BCRF-17-140 awarded to HSR and MJM), The Avon Foundation for Women and the Susan G. Komen for the Cure for support of Translational Breast Cancer Research Consortium (TBCRC) Trial 009. We also acknowledge support from MSKCC Core grant P30 CA 008748. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute.

References

1. Liu MC, Shields PG, Warren RD, Cohen P, Wilkinson M, Ottaviano YL, et al. Circulating tumor cells: a useful predictor of treatment efficacy in metastatic breast cancer. J Clin Oncol. 2009; 27:5153–9. [PubMed: 19752342]
2. Magbanua MJ, Carey LA, DeLuca A, Hwang J, Scott JH, Rimawi MF, et al. Circulating tumor cell analysis in metastatic triple-negative breast cancers. Clin Cancer Res. 2015; 21:1098–105. [PubMed: 25524311]
3. Magbanua MJ, Park JW. Advances in genomic characterization of circulating tumor cells. Cancer Metastasis Rev. 2014; 33:757–69. [PubMed: 24867683]
4. Magbanua MJ, Park JW. Isolation of circulating tumor cells by immunomagnetic enrichment and fluorescence-activated cell sorting (IE/FACS) for molecular profiling. Methods. 2013; 64:114–8. [PubMed: 23896286]
5. Magbanua MJ, Sosa EV, Roy R, Eisenbud LE, Scott JH, Olshen A, et al. Genomic profiling of isolated circulating tumor cells from metastatic breast cancer patients. Cancer research. 2013; 73:30–40. [PubMed: 23135909]
6. Magbanua MJ, Sosa EV, Scott JH, Simko J, Collins C, Pinkel D, et al. Isolation and genomic analysis of circulating tumor cells from castration resistant metastatic prostate cancer. BMC Cancer. 2012; 12:78. [PubMed: 22373240]
7. Paszek MJ, DuFort CC, Rossier O, Bainer R, Mouw JK, Godula K, et al. The cancer glycocalyx mechanically primes integrin-mediated growth and survival. Nature. 2014; 511:319–25. [PubMed: 25030168]
8. Gulbahce N, Magbanua MJM, Chin R, Agarwal MR, Luo X, Liu J, et al. Quantitative Whole Genome Sequencing of Circulating Tumor Cells Enables Personalized Combination Therapy of Metastatic Cancer. Cancer Res. 2017; 77:4530–41. [PubMed: 28811315]
9. Magbanua MJ, Sosa EV, Scott JH, Simko J, Collins C, Pinkel D, et al. Isolation and genomic analysis of circulating tumor cells from castration resistant metastatic prostate cancer. BMC Cancer. 2012; 12:78. [PubMed: 22373240]
10. Lee JS, Melisko ME, Magbanua MJ, Kablanian AT, Scott JH, Rugo HS, et al. Detection of cerebrospinal fluid tumor cells and its clinical relevance in leptomeningeal metastasis of breast cancer. Breast Cancer Res Treat. 2015; 154:339–49. [PubMed: 26520840]
11. Magbanua MJ, Melisko M, Roy R, Sosa EV, Hauranieh L, Kablanian A, et al. Molecular profiling of tumor cells in cerebrospinal fluid and matched primary tumors from metastatic breast cancer patients with leptomeningeal carcinomatosis. Cancer Res. 2013; 73:7134–43. [PubMed: 24142343]

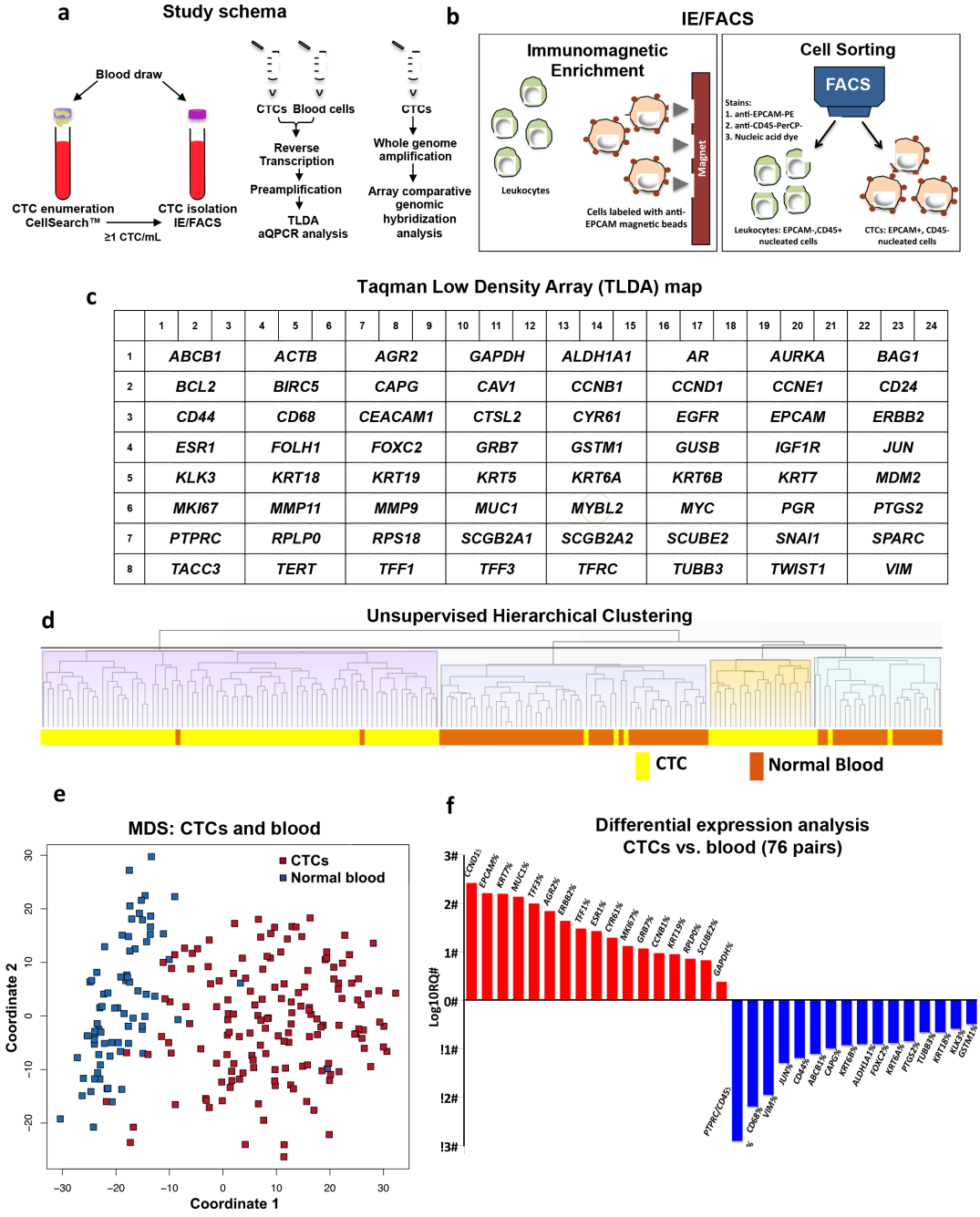
12. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006; 7:3. [PubMed: 16448564]
13. Blaveri E, Brewer JL, Roydasgupta R, Fridlyand J, DeVries S, Koppie T, et al. Bladder cancer stage and outcome by array-based comparative genomic hybridization. Clin Cancer Res. 2005; 11:7012–22. [PubMed: 16203795]
14. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004; 5:557–72. [PubMed: 15475419]
15. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5:R80. [PubMed: 15461798]
16. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995; 57:289–300.
17. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. Biometrika. 1915; 10:507–21.
18. Pearson K. Notes on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London. 1895; 58:240–2.
19. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009; 27:1160–7. [PubMed: 19204204]
20. Gendoo DM, Ratanasirigulchai N, Schroder MS, Pare L, Parker JS, Prat A, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. Bioinformatics. 2016; 32:1097–9. [PubMed: 26607490]
21. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
22. Kaplan EL, Meier P. Nonparametric-Estimation from Incomplete Observations. J Am Stat Assoc. 1958; 53:457–81.
23. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep. 1966; 50:163–70. [PubMed: 5910392]
24. Fridlyand J, Snijders AM, Ylstra B, Li H, Olshen A, Segraves R, et al. Breast tumor copy number aberration phenotypes and genomic instability. BMC Cancer. 2006; 6:96. [PubMed: 16620391]
25. Baehner FL, Achacoso N, Maddala T, Shak S, Quesenberry CP Jr, Goldstein LC, et al. Human epidermal growth factor receptor 2 assessment in a case-control study: comparison of fluorescence in situ hybridization and quantitative reverse transcription polymerase chain reaction performed by central laboratories. J Clin Oncol. 2010; 28:4300–6. [PubMed: 20697093]
26. Perez EA, Baehner FL, Butler SM, Thompson EA, Dueck AC, Jamshidian F, et al. The relationship between quantitative human epidermal growth factor receptor 2 gene expression by the 21-gene reverse transcriptase polymerase chain reaction assay and adjuvant trastuzumab benefit in Alliance N9831. Breast Cancer Res. 2015; 17:133. [PubMed: 26429296]
27. Barbazan J, Alonso-Alconada L, Muinelo-Romay L, Vieito M, Abalo A, Alonso-Nocelo M, et al. Molecular characterization of circulating tumor cells in human metastatic colorectal cancer. PloS one. 2012; 7:e40476. [PubMed: 22811761]
28. Lu J, Fan T, Zhao Q, Zeng W, Zaslavsky E, Chen JJ, et al. Isolation of circulating epithelial and tumor progenitor cells with an invasive phenotype from breast cancer patients. International journal of cancer Journal international du cancer. 2010; 126:669–83. [PubMed: 19662651]
29. O'Hara SM, Moreno JG, Zweitzig DR, Gross S, Gomella LG, Terstappen LW. Multigene reverse transcription-PCR profiling of circulating tumor cells in hormone-refractory prostate cancer. Clin Chem. 2004; 50:826–35. [PubMed: 14988224]
30. Sieuwerts AM, Kraan J, Bolt-de Vries J, van der Spoel P, Mostert B, Martens JW, et al. Molecular characterization of circulating tumor cells in large quantities of contaminating leukocytes by a multiplex real-time PCR. Breast Cancer Res Treat. 2009; 118:455–68. [PubMed: 19115104]
31. Sieuwerts AM, Mostert B, Bolt-de Vries J, Peeters D, de Jongh FE, Stouthard JM, et al. mRNA and microRNA Expression Profiles in Circulating Tumor Cells and Primary Tumors of Metastatic Breast Cancer Patients. Clin Cancer Res. 2011; 17:3600–18. [PubMed: 21505063]

32. Smirnov DA, Zweitzig DR, Foulk BW, Miller MC, Doyle GV, Pienta KJ, et al. Global gene expression profiling of circulating tumor cells. Cancer Res. 2005; 65:4993–7. [PubMed: 15958538]
33. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science. 2013; 339:580–4. [PubMed: 23372014]
34. Mostert B, Sieuwerts AM, Kraan J, Bolt-de Vries J, van der Spoel P, van Galen A, et al. Gene expression profiles in circulating tumor cells to predict prognosis in metastatic breast cancer patients. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2014
35. Yu M, Ting DT, Stott SL, Wittner BS, Ozsolak F, Paul S, et al. RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. Nature. 2012; 487:510–3. [PubMed: 22763454]
36. Chen CL, Mahalingam D, Osmulski P, Jadhav RR, Wang CM, Leach RJ, et al. Single-cell analysis of circulating tumor cells identifies cumulative expression patterns of EMT-related genes in metastatic prostate cancer. The Prostate. 2013; 73:813–26. [PubMed: 23280481]
37. Ozkumur E, Shah AM, Ciciliano JC, Emmink BL, Miyamoto DT, Brachtel E, et al. Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. Science translational medicine. 2013; 5:179ra47.
38. Powell AA, Talasaz AH, Zhang H, Coram MA, Reddy A, Deng G, et al. Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. PLoS One. 2012; 7:e33788. [PubMed: 22586443]
39. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nature biotechnology. 2012; 30:777–82.
40. Cann GM, Gulzar ZG, Cooper S, Li R, Luo S, Tat M, et al. mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer. PLoS One. 2012; 7:e49144. [PubMed: 23145101]
41. Ting DT, Wittner BS, Ligorio M, Vincent Jordan N, Shah AM, Miyamoto DT, et al. Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. Cell reports. 2014; 8:1905–18. [PubMed: 25242334]
42. de Beca FF, Caetano P, Gerhard R, Alvarenga CA, Gomes M, Paredes J, et al. Cancer stem cells markers CD44, CD24 and ALDH1 in breast cancer special histological types. J Clin Pathol. 2013; 66:187–91. [PubMed: 23112116]
43. Kasimir-Bauer S, Hoffmann O, Wallwiener D, Kimmig R, Fehm T. Expression of stem cell and epithelial-mesenchymal transition markers in primary breast cancer patients with circulating tumor cells. Breast Cancer Res. 2012; 14:R15. [PubMed: 22264265]
44. Barriere G, Riouallon A, Renaudie J, Tartary M, Rigaud M. Mesenchymal and stemness circulating tumor cells in early breast cancer diagnosis. BMC Cancer. 2012; 12:114. [PubMed: 22443102]
45. Aktas B, Tewes M, Fehm T, Hauch S, Kimmig R, Kasimir-Bauer S. Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells of metastatic breast cancer patients. Breast Cancer Res. 2009; 11:R46. [PubMed: 19589136]
46. Satelli A, Mitra A, Brownlee Z, Xia X, Bellister S, Overman MJ, et al. Epithelial-mesenchymal transitioned circulating tumor cells capture for detecting tumor progression. Clin Cancer Res. 2015; 21:899–906. [PubMed: 25516888]
47. Vishnoi M, Peddibhotla S, Yin W, ATS, George GC, Hong DS, et al. The isolation and characterization of CTC subsets related to breast cancer dormancy. Sci Rep. 2015; 5:17533. [PubMed: 26631983]
48. Krebs MG, Metcalf RL, Carter L, Brady G, Blackhall FH, Dive C. Molecular analysis of circulating tumour cells-biology and biomarkers. Nat Rev Clin Oncol. 2014; 11:129–44. [PubMed: 24445517]
49. Rack B, Schindlbeck C, Juckstock J, Andergassen U, Hepp P, Zwingers T, et al. Circulating tumor cells predict survival in early average-to-high risk breast cancer patients. J Natl Cancer Inst. 2014:106.

50. Reiter M, Kirchner B, Muller H, Holzhauer C, Mann W, Pfaffl MW. Quantification noise in single cell experiments. Nucleic Acids Res. 2011; 39:e124. [PubMed: 21745823]
51. Bengtsson M, Hemberg M, Rorsman P, Stahlberg A. Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. BMC Mol Biol. 2008; 9:63. [PubMed: 18631407]
52. Lindsay CR, Le Moulec S, Billiot F, Loriot Y, Ngo-Camus M, Vielh P, et al. Vimentin and Ki67 expression in circulating tumour cells derived from castrate-resistant prostate cancer. BMC Cancer. 2016; 16:168. [PubMed: 26923772]
53. Lindstrom LS, Karlsson E, Wilking UM, Johansson U, Hartman J, Lidbrink EK, et al. Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. J Clin Oncol. 2012; 30:2601–8. [PubMed: 22711854]
54. de Duenas EM, Hernandez AL, Zotano AG, Carrion RM, Lopez-Muniz JI, Novoa SA, et al. Prospective evaluation of the conversion rate in the receptor status between primary breast cancer and metastasis: results from the GEICAM 2009-03 ConvertHER study. Breast Cancer Res Treat. 2014; 143:507–15. [PubMed: 24414130]
55. Carlson RW, Allred DC, Anderson BO, Burstein HJ, Edge SB, Farrar WB, et al. Metastatic breast cancer, version 1.2012: featured updates to the NCCN guidelines. J Natl Compr Canc Netw. 2012; 10:821–9. [PubMed: 22773798]
56. Schramm A, Friedl TW, Schochter F, Scholz C, de Gregorio N, Huober J, et al. Therapeutic intervention based on circulating tumor cell phenotype in metastatic breast cancer: concept of the DETECT study program. Arch Gynecol Obstet. 2016; 293:271–81. [PubMed: 26354331]
57. Paoletti C, Regan M, Liu M, Marcom P, Hart L, Smith J II, et al. Circulating tumor cell number and CTC-endocrine therapy index predict clinical outcomes in ER positive metastatic breast cancer patients: Results of the COMETI Phase 2 trial [abstract]. Cancer Res. 2017; 77(4 Suppl) Abstract nr P1-01-01.
58. Lee JS, Magbanua MJM, Park JW. Circulating tumor cells in breast cancer: applications in personalized medicine. Breast Cancer Res Treat. 2016; 160:411–24. [PubMed: 27761678]
59. Dickler MN, Barry WT, Cirrincione CT, Ellis MJ, Moynahan ME, Innocenti F, et al. Phase III Trial Evaluating Letrozole As First-Line Endocrine Therapy With or Without Bevacizumab for the Treatment of Postmenopausal Women With Hormone Receptor-Positive Advanced-Stage Breast Cancer: CALGB 40503 (Alliance). J Clin Oncol. 2016; 34:2602–9. [PubMed: 27138575]
60. Isakoff SJ, Mayer EL, He L, Traina TA, Carey LA, Krag KJ, et al. TBCRC009: A Multicenter Phase II Clinical Trial of Platinum Monotherapy With Biomarker Assessment in Metastatic Triple-Negative Breast Cancer. J Clin Oncol. 2015; 33:1902–9. [PubMed: 25847936]
61. Rugo HS, Barry WT, Moreno-Aspitia A, Lyss AP, Cirrincione C, Leung E, et al. Randomized Phase III Trial of Paclitaxel Once Per Week Compared With Nanoparticle Albumin-Bound Nab-Paclitaxel Once Per Week or Ixabepilone With Bevacizumab As First-Line Chemotherapy for Locally Recurrent or Metastatic Breast Cancer: CALGB 40502/NCCTG N063H (Alliance). J Clin Oncol. 2015; 33:2361–9. [PubMed: 26056183]
62. Bidard FC, Peeters DJ, Fehm T, Nole F, Gisbert-Criado R, Mavroudis D, et al. Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. Lancet Oncol. 2014; 15:406–14. [PubMed: 24636208]
63. Lee JS, Magbanua MJ, Park JW. Circulating tumor cells in breast cancer: applications in personalized medicine. Breast Cancer Res Treat. 2016; 160:411–24. [PubMed: 27761678]

TRANSLATIONAL RELEVANCE

Molecular characterization of circulating tumor cells (CTCs) offers a unique approach to examine biologic mechanisms and biomarkers that are associated with cancer progression and treatment resistance. Indeed, liquid biopsy via CTC profiling provides obvious advantages over tissue biopsy, including non-invasive access and ease of serial monitoring. We isolated circulating tumor cells for copy number and gene expression profiling to evaluate biomarkers relevant to breast cancer. Molecular characterization of CTCs offers an approach to evaluate biomarkers potentially associated with cancer progression and treatment resistance.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	*ABCB1*			*ACTB*			*AGR2*			*GAPDH*			*ALDH1A1*			*AR*			*AURKA*			*BAG1*		
2	*BCL2*			*BIRC5*			*CAPG*			*CAV1*			*CCNB1*			*CCND1*			*CCNE1*			*CD24*		
3	*CD44*			*CD68*			*CEACAM1*			*CTSL2*			*CYR61*			*EGFR*			*EPCAM*			*ERBB2*		
4	*ESR1*			*FOLH1*			*FOXC2*			*GRB7*			*GSTM1*			*GUSB*			*IGF1R*			*JUN*		
5	*KLK3*			*KRT18*			*KRT19*			*KRT5*			*KRT6A*			*KRT6B*			*KRT7*			*MDM2*		
6	*MKI67*			*MMP11*			*MMP9*			*MUC1*			*MYBL2*			*MYC*			*PGR*			*PTGS2*		
7	*PTPRC*			*RPLP0*			*RPS18*			*SCGB2A1*			*SCGB2A2*			*SCUBE2*			*SNAI1*			*SPARC*		
8	*TACC3*			*TERT*			*TFF1*			*TFF3*			*TFRC*			*TUBB3*			*TWIST1*			*VIM*		

Figure 1. Study schema, IE/FACS isolation method and gene expression profiling of CTCs and matched leukocytes

a) Blood sample was drawn into CellSearch™ CellSave Preservative Tubes (Veridex) for CTC enumeration. Additional volume of blood was drawn into tubes containing EDTA for CTC isolation. To increase the likelihood of isolating CTCs for molecular analysis, we first performed CTC enumeration in 7.5mL of blood using the CellSearch™ assay to identify patients with ≥1 CTC/mL. Gene expression and copy number analyses were performed by Taqman Low Density Array (TLDA) aQPCR and array comparative genomic hybridization, respectively; **b**) IE/FACS is a two-step process for isolation of CTCs. It composed of

immumomagnetic enrichment (IE) followed by fluorescence-activated cell sorting. CTCs are defined as nucleated cells that stain positive for EPCAM (epithelial marker) and negative for CD45 (leukocyte marker). Blood cells (leukocytes) can also be collected to serve as non-tumor controls; **c**) Map of the TLDA microfluidic card containing 64 genes printed in triplicate; **d**) Unsupervised hierarchical clustering analysis showing that CTCs (n=105) clustered away from leukocytes (n=76) samples; **e**) Metric multidimensional scaling analysis (MDSA) was performed using gene expression data from 64 genes profiled by aQPCR analysis in 151 CTC and 76 leukocyte samples; **f**) Genes differentially expressed between matched CTCs and leukocytes (n=76 pairs) isolated from the same tumor-enriched blood sample by IE/FACS. A paired t-test was performed using leukocytes as calibrator samples. Genes with an adjusted p value <0.001 were considered statistically significant. Relative quantification (RQ) is reported in the logarithmic scale ($\log_{10}RQ=\log_{10} 2^{-\Delta\Delta Ct}$). A $Log_{10}RQ=1$ or −1 indicates that a gene is expressed 10 times or 1/10 as much, respectively, in the CTCs relative to leukocyte samples.

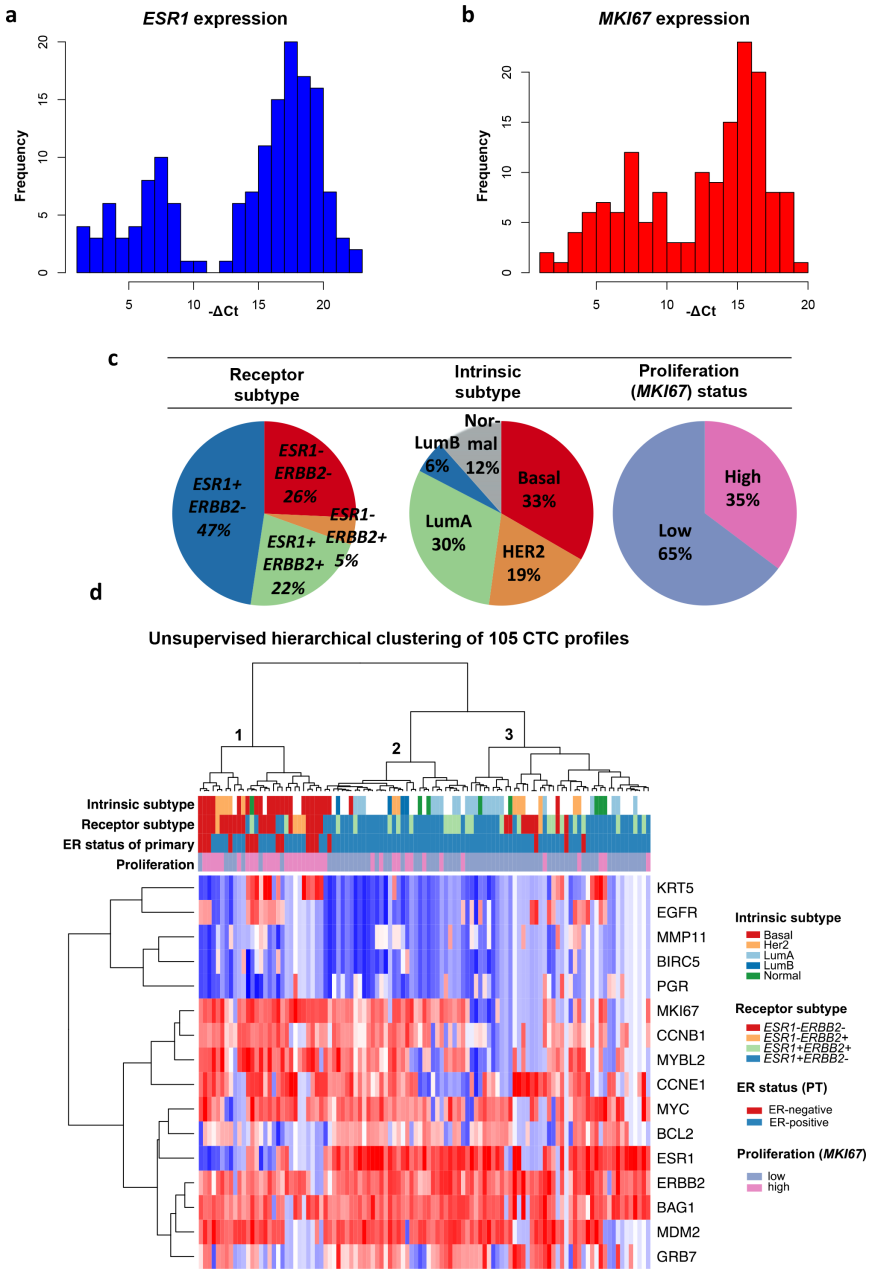


Figure 2. Intrinsic subtype, *ESR1/ERBB2*, and proliferation (*MKI67*) status in CTCs

a) Frequency distribution of *ESR1*, and **b**) *MKI67* gene expression levels across all samples. Expression levels are shown as −ΔCt; **c**) Pie charts showing distribution of receptor and intrinsic subtype, proliferation (*MKI67*) status in CTCs; **d**) Heat map showing results of unsupervised hierarchical clustering analysis of 105 CTCs using 16 of the PAM50 intrinsic subtype classifier. Intrinsic and receptor subtypes along with proliferation (*MKI67)* status, and clinical ER status of matched primary tumors are shown above the heat map.

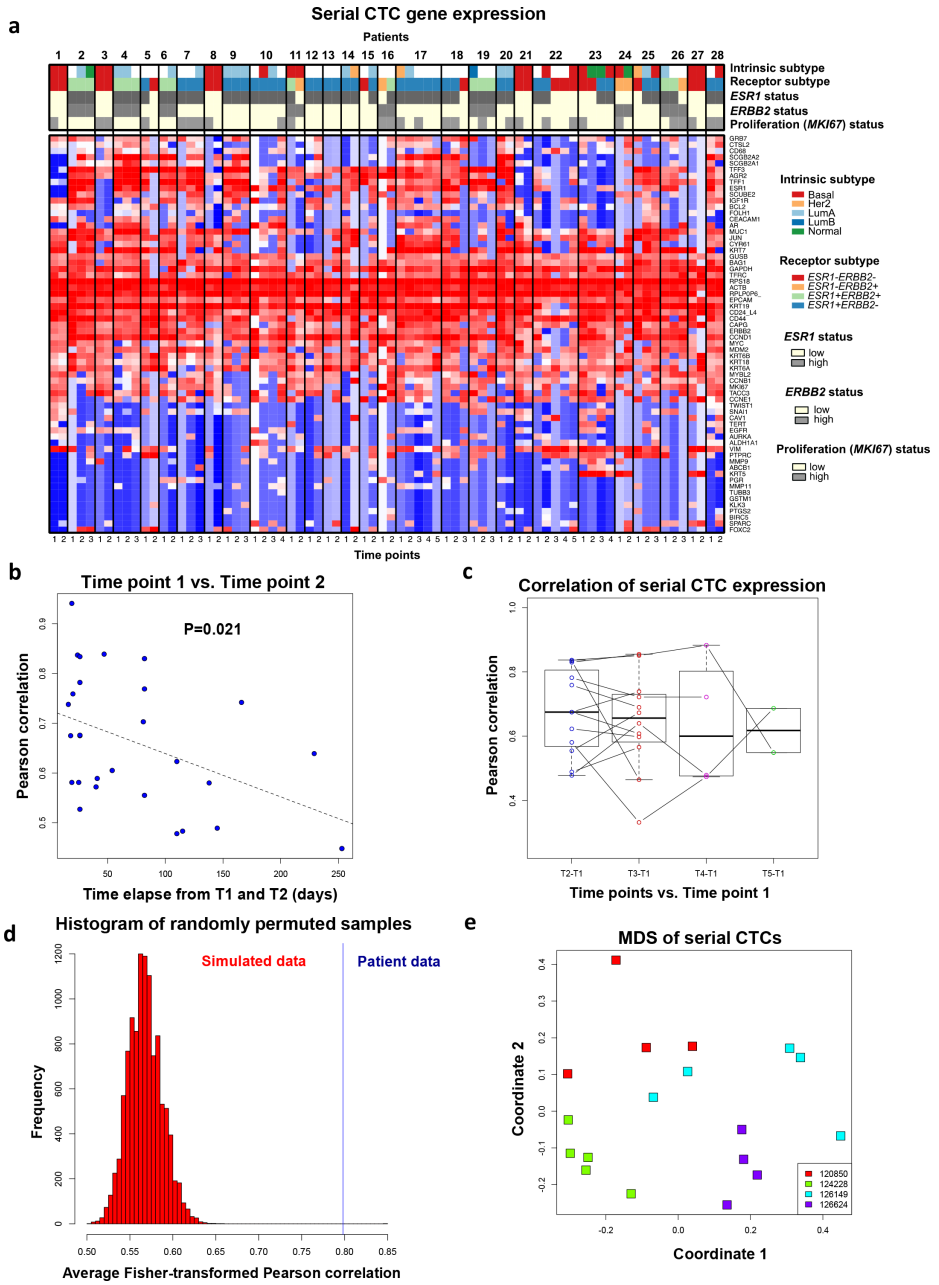


Figure 3. Serial expression analysis reveals patient-specific CTC expression signatures

a) Heat map showing results of supervised clustering analysis of CTC expression profiles from 28 patients (74 samples). Intrinsic and receptor subtypes along with *ESR1*, *ERBB2*, and proliferation (*MKI67*) status are shown above the heat map and the time points are indicated below; **b**) Pearson correlation of serial CTC samples from the same patient collected at time points 1 and 2 (y-axis) and the days elapsed between two time points (x-axis); **c**) Pearson correlation of samples between time point 1 (T1) and other time points. Each line indicates an individual patient. The dark line inside the box plot shows median Pearson correlation across all patients; **d**) Distribution of Pearson correlations of simulated gene expression data generated by 10,000 random permutations (histogram) and median

Pearson correlation among samples from the same patient (blue line); **e**) Metric multidimensional scaling analysis (MDSA) was performed using gene expression data from 64 genes profiled by aQPCR analysis in serial CTC samples from four patients (pts).

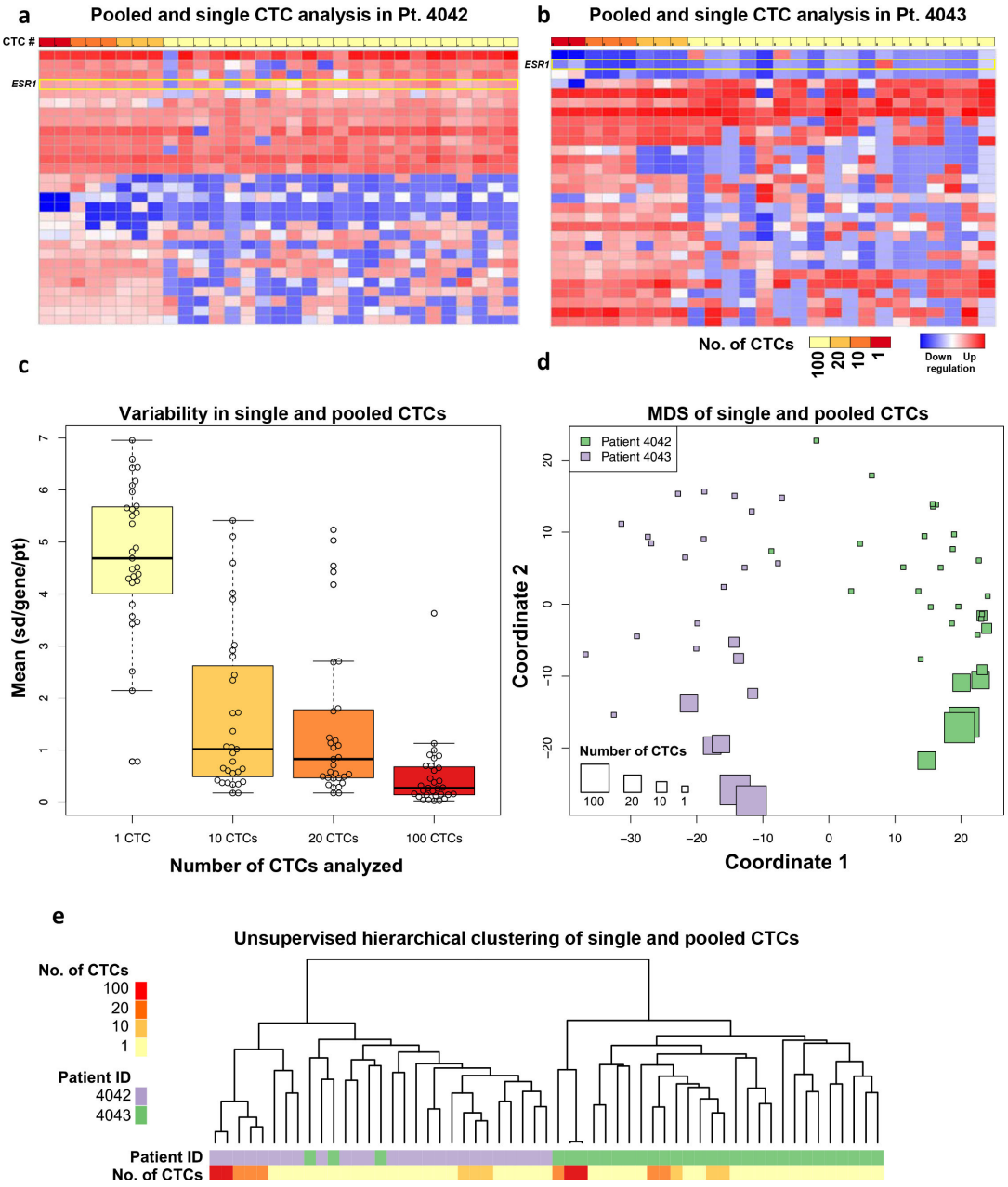
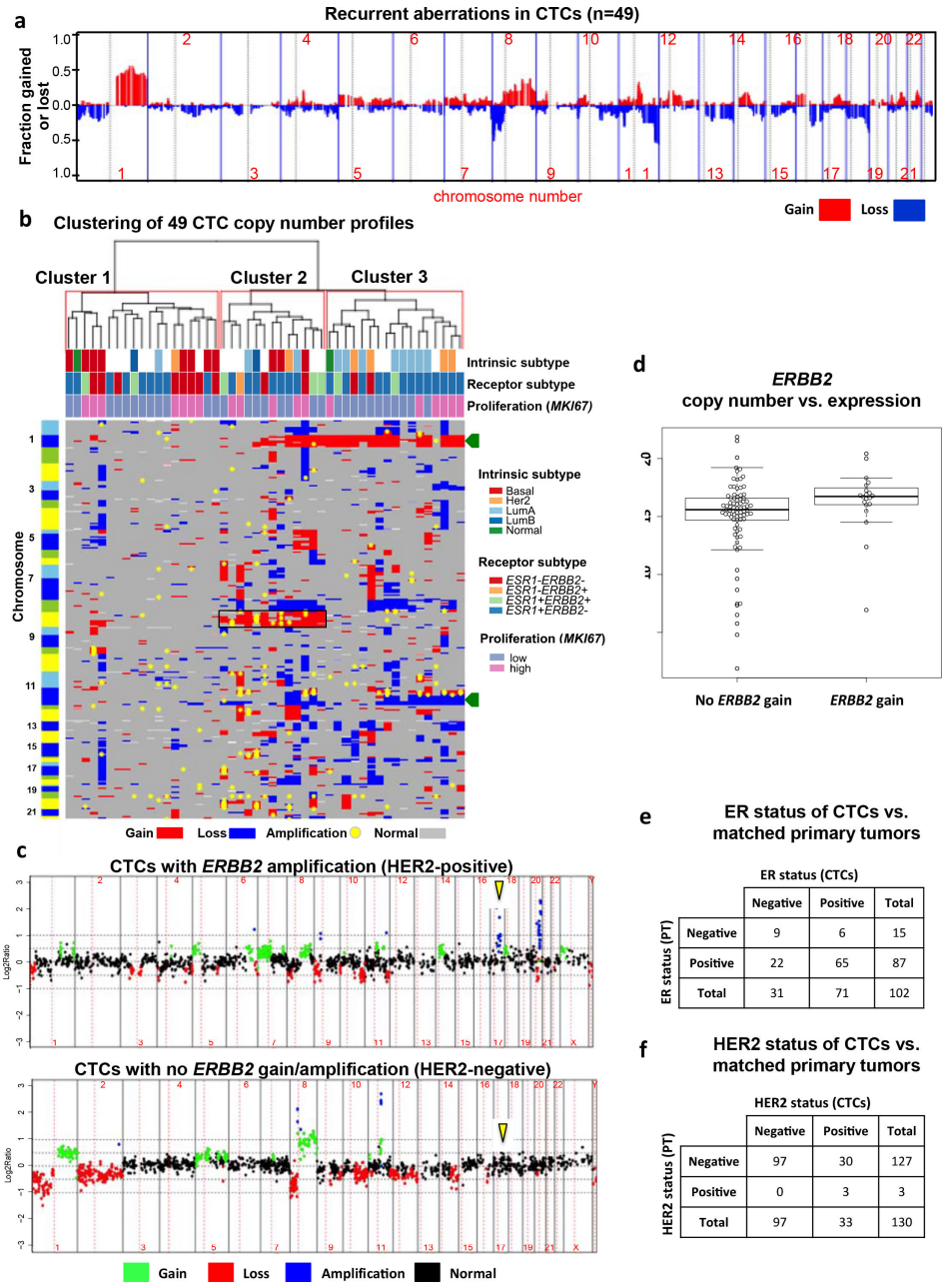


Figure 4. Single cell expression analysis reveals inter- and intra-patient heterogeneity

Microfluidic-based dynamic array expression analysis (Fluidigm™) of single and pooled CTCs isolated by IE/FACS from patients **a**) 4042 and **b**) 4043. The samples (columns) are arranged in a supervised fashion according to the number of cell input for expression analysis indicated above the heat map; **c**) Mean levels of heterogeneity (standard deviation/gene/patient) in single and pooled CTCs from patients 4042 and 4043. Gene expression levels of 32 genes listed in Supplementary Table 2 were analyzed via Fluidigm™; **d**) Metric multidimensional scaling analysis (MDSA) was performed using 32 genes analyzed in single and pooled CTCs from patients 4042 and 4043; **e**) Unsupervised hierarchical clustering showing that CTC samples from the same patients clustered together.



ER status of CTCs vs. matched primary tumors

ER status (PT)	ER status (CTCs) Negative	Positive	Total
Negative	9	6	15
Positive	22	65	87
Total	31	71	102

HER2 status of CTCs vs. matched primary tumors

HER2 status (PT)	HER2 status (CTCs) Negative	Positive	Total
Negative	97	30	127
Positive	0	3	3
Total	97	33	130

Figure 5. Copy number and *ERBB2*/*ESR1* status in CTCs

a) plot showing recurrent aberrations in CTCs from 49 metastatic breast cancer patients; **b**) Unsupervised clustering analysis of copy number profiles of CTCs with matched expression data (n=49). Each column on the heat map represents a sample. The colors, red (gain), blue (loss), grey (normal) and yellow (high-level amplification), indicate copy number status. The green arrowheads indicate 1q gain and 11q loss and the black box indicates 8q gain. The bar to the left indicates chromosome locations with chromosome 1pter to the top and 22qter to the bottom; note that only the odd-numbered chromosomes are indicated. The short/p-arms are colored aqua or green while the long/q-arms are colored blue or yellow. The annotation strip shows assignment to groups based on gene expression results. **c**) Representative array

comparative genomic hybridization (aCGH) profiles showing HER2-positive (top panel) and HER2-negative CTCs (bottom panel). The arrows point to the chromosome region containing the *ERBB2/HER2* gene. The log2 ratio value for each BAC clone is plotted on the y-axis. The x-axis represents the genomic position of each BAC clone on the array, with chromosome numbers indicated. Vertical solid lines indicate chromosome boundaries, and vertical red dashed line represents the centromeric region dividing each chromosome into the p- or short arm (to the left of centromere) and the q- or long arm (to the right of the centromere). Color represents copy number status: red=loss, green=gain, blue=amplification, and black=normal; **d**) *ERBB2* expression levels (−ΔCt) and HER2 status by aCGH analysis; Concordance between **e**) *ESR1* status (by QPCR) and **f**) *ERBB2* status (by aCGH) in CTCs vs. clinical ER/HER2 status of corresponding primary tumors.

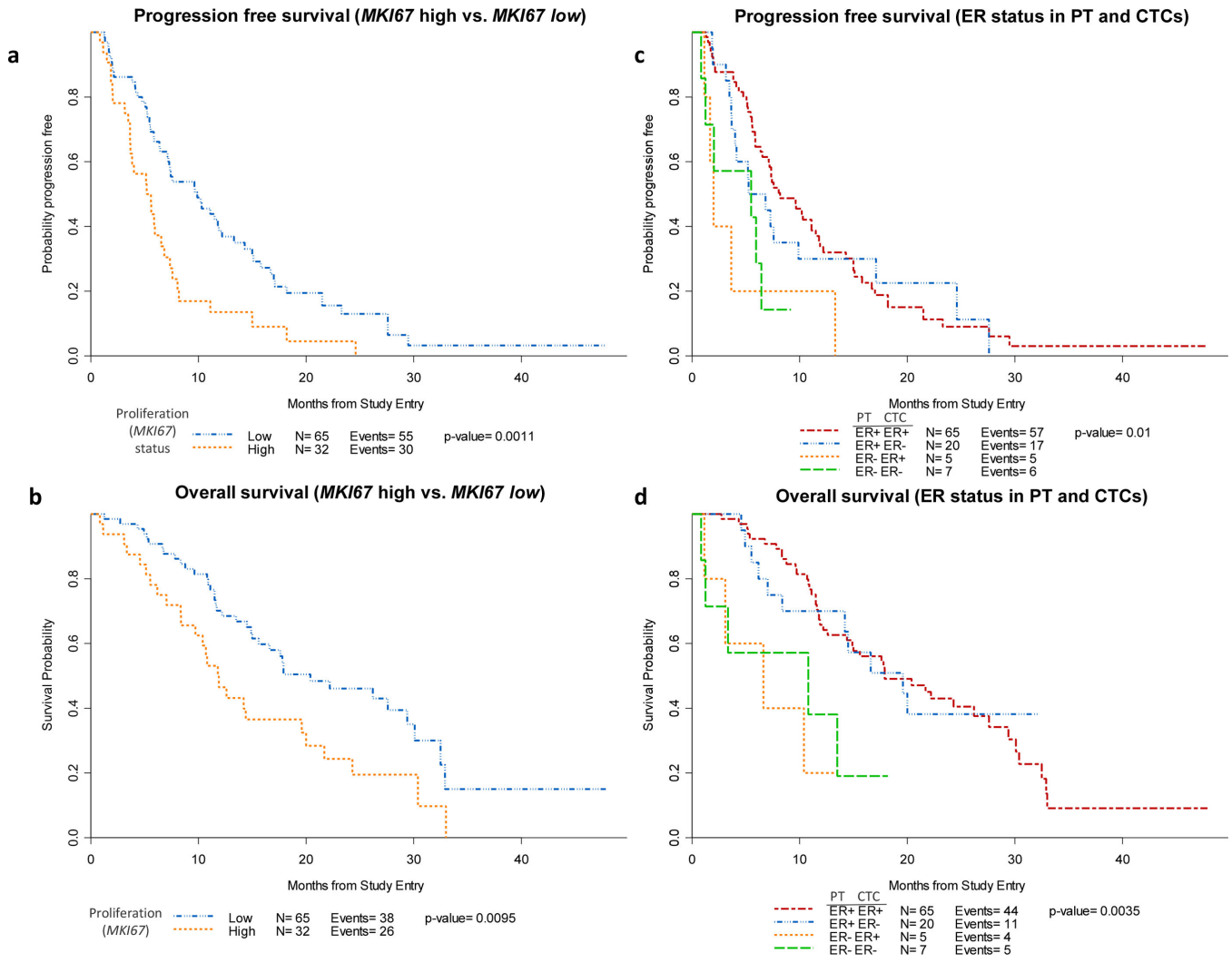


Figure 6. Circulating tumor cell phenotype and patient outcome

Kaplan-Meier plots showing estimates for **a**) progression free survival and **b**) overall survival of patients harboring CTCs with low or high proliferation status based on the expression of *MKI67* (encodes Ki67); **c**) Progression free survival and **d**) overall survival of patients based on *ESR1*/ER status in CTCs and primary tumor (PT).