Title
Expression profiling of circulating tumor cells in metastatic breast cancer

Journal
Breast Cancer Research and Treatment, 149(1)

ISSN
0167-6806

Authors
Lang, Julie E
Scott, Janet H
Wolf, Denise M
et al.

Publication Date
2015

DOI
10.1007/s10549-014-3215-0

Peer reviewed

Breast Cancer Res Treat (2015) 149:121–131
DOI 10.1007/s10549-014-3215-0

CLINICAL TRIAL

Expression profiling of circulating tumor cells in metastatic breast cancer

Julie E. Lang · Janet H. Scott · Denise M. Wolf · Petr Novak · Vasu Punj · Mark Jesus M. Magbanua · Weizhu Zhu · Neal Mineyev · Christopher M. Haqq · Julia R. Crothers · Laura J. Esserman · Debasish Tripathy · Laura van 't Veer · John W. Park

Received: 3 September 2014 / Accepted: 17 November 2014 / Published online: 29 November 2014

Abstract Circulating tumor cells (CTCs) are prognostic in all stages of breast cancer. However, since they are extremely rare, little is known about the molecular nature of these cells. We report a novel strategy for the isolation and expression profiling of pure populations of CTCs derived from peripheral blood. We developed a method to isolate CTCs based on immunomagnetic capture followed by fluorescence-activated cell sorting (IE/FACS). After assay validation using the BT474 cell line spiked into blood samples in vitro, RNA from CTCs isolated from the blood of five metastatic breast cancer (MBC) patients was linearly amplified and subjected to gene expression profiling via cDNA microarrays. We isolated a range of 9-993 captured CTCs from five MBC patients' blood and profiled their RNA in comparison to a diverse panel of primary breast tumors ($n = 55$). Unsupervised hierarchical clustering revealed that CTC profiles clustered with more aggressive subtypes of primary breast tumors and were readily distinguishable from peripheral blood (PB) and normal epithelium. Differential expression analysis revealed CTCs to have downregulated apoptosis, and they were distinguishable from PB by the relative absence of immune-related signals. As expected, CTCs from MBC had significantly higher risk of recurrence scores than primary tumors ($p = 0.0073$). This study demonstrates that it is feasible to isolate CTCs from PB with high purity through IE/FACS and profile them via gene expression analysis. Our approach may inform the discovery of therapeutic predictors and be useful for real-time identification of emerging resistance mechanisms in MBC patients.

Keywords Circulating tumor cells · Micrometastases · Breast cancer · EpCAM · Gene expression

Electronic supplementary material The online version of this article (doi:10.1007/s10549-014-3215-0) contains supplementary material, which is available to authorized users.

J. E. Lang (✉) · W. Zhu · N. Mineyev
Division of Breast and Soft Tissue Surgery, Department of Surgery, Norris Comprehensive Cancer Center (NCCC), University of Southern California (USC), 1510 San Pablo Street Suite 412, Los Angeles, CA 90033-0800, USA
e-mail: julie.lang@med.usc.edu

J. H. Scott · M. J. M. Magbanua · J. R. Crothers · J. W. Park
Division of Hematology and Medical Oncology, Department of Medicine, University of California San Francisco (UCSF), San Francisco, CA, USA

D. M. Wolf · L. van 't Veer
UCSF Department of Laboratory Medicine, San Francisco, CA, USA

P. Novak
Biology Centre ASCR, Institute of Plant Molecular Biology, 37001 Ceske Budejovice, Czech Republic

V. Punj
NCCC Bioinformatics Core and Division of Hematology, Keck School of Medicine, NCCC, University of Southern California (USC), Los Angeles, CA, USA

C. M. Haqq
Atara Biotherapeutics, Westlake Village, Brisbane, CA, USA

L. J. Esserman
Section of Breast Care Surgery, UCSF Department of Surgery, San Francisco, CA, USA

D. Tripathy
Division of Oncology, USC Department of Medicine, NCCC, Los Angeles, CA, USA

Abbreviations

CTC	Circulating tumor cell
IE/FACS	Immunomagnetic enrichment followed by fluorescence-activated cell sorting
RNA	ribonucleic acid
MBC	Metastatic breast cancer
cDNA	Complementary deoxyribonucleic acid
PB	Peripheral blood
QPCR	Quantitative real-time polymerase chain reaction
EpCAM	epithelial cell adhesion marker
mAb	Monoclonal antibody
HER2	Human epidermal growth factor receptor 2
ER	Estrogen receptor
RPMI	Roswell Park Memorial Institute
EDTA	Ethylenediaminetetraacetic acid
CALGB	Cancer and Leukemia Group B
ACRIN	American College of Radiology Imaging Network
StratRef	Stratagene Universal Human Pooled Reference RNA
dIdC	Poly(deoxyinosinic-deoxycytidylic) acid sodium salt
M-MLV	Moloney Murine Leukemia Virus Reverse Transcriptase
GUS	Beta-glucuronidase
ABI	Applied Biosystems
CT	Cycle threshold
HLA	Human leukocyte antigen
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
Cy	Cyanine dye
GSE	Gene Expression Omnibus series format file
AU	Approximately Unbiased
ANOVA	Analysis of variance
BH	Benjamini Hochberg
FDR	False discovery rate
DAVID	the Database for Annotation, Visualization and Integrated Discovery
PAM50	A 50 gene intrinsic subtype classifier
BCL2	B-cell lymphoma 2
CDC6	Cell division cycle 6
NUF2	Kinetochore protein Nuf2
CENPF	Centromere protein F
CEP55	Centrosomal protein 55 kDa
CXXC5	CXXC finger protein 5
EGFR	Epidermal growth factor receptor
ERBB2	Human epidermal growth factor receptor 2
ESR1	Estrogen receptor 1
FGFR4	Fibroblast growth factor receptor 4
FOXC1	Forkhead box C1
GRB7	Growth factor receptor-bound protein 7
NDC80	NDC80 kinetochore complex component
KRT14	Keratin 14
MYBL2	Myeloblastosis oncogene-like 2
PTTG1	Pituitary tumor-transforming 1
RRM2	Ribonucleotide reductase M2
TMEM45B	Transmembrane protein 45B
TYMS	Thymidylate synthetase
UBE2C	Ubiquitin-conjugating enzyme E2C
UBE2T	Ubiquitin-conjugating enzyme E2
ROR ∼ P	Risk of recurrence score
PBS	Phosphate-buffered saline
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene ontology
UTR	Untranslated region
UCSF	University of California, San Francisco
USC	University of Southern California
I-SPY1	Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging And moLecular Analysis 1 clinical trial
SWOG	Southwest Oncology Group
NCI	National Cancer Institute
PE	Phycoerythrin
CI	Confidence interval

Introduction

Circulating tumor cells have been demonstrated to be present in the peripheral blood (PB) from patients with all major types of cancers [1]. CTC status reflects metastatic progression, correlating with imaging to evaluate treatment response or progression of disease [2, 3]. Ideally, longitudinal analysis of CTCs for changes in gene expression and mutation status during treatment may provide insight into the development of resistance mechanisms, information that may be useful in selecting drug combinations that may prolong a patient's life [4].

The barriers to such a rational adaptive approach to treating metastatic breast cancer (MBC) are largely technical. CTCs are rare and difficult to isolate. Prior efforts in the molecular profiling of breast cancer CTCs utilized enriched samples with a predominant population of leukocytes even after enrichment procedures [5–7]. Several groups have reported using quantitative real-time polymerase chain reaction (QPCR) for expression analysis of CTCs [8–12]. QPCR requires prior knowledge of genes to be interrogated and may not be suitable to distinguish very low levels of leukocyte contamination. Sieuwerts et al. reported that despite four logarithms of enrichment, large quantities of contaminating leukocytes remained [11]. However, molecular profiling of CTCs in a background of leukocytes does have the potential for detecting malignant

transcripts [6]. Yu et al. reported that it is feasible to sequence CTCs and that epithelial to mesenchymal transition was demonstrated [5]. While these approaches are major advances in rare cell profiling, both required the subtraction of leukocyte signatures for CTC profiling, introducing the potential for biases and false discovery.

We have developed a protocol to collect rare populations of CTCs from PB [13] consisting of an immunomagnetic enrichment step followed by fluorescence-activated cell sorting (IE/FACS). Small populations of pure CTCs can be isolated for downstream molecular analyses. Magbanua et al. reported that IE/FACS-isolated CTCs could be assessed by array comparative genomic hybridization, providing evidence of the clonal relationship between CTCs and primary tumors [14]. We have demonstrated that whole transcriptome profiling is possible at the picogram input level [15].

In this pilot study, we profiled CTCs isolated from the blood of five MBC patients to demonstrate that gene expression profiling of CTCs is feasible using IE/FACS followed by RNA amplification. This approach may serve as a rational basis for optimal treatment selection based on transcriptional profiling.

Methods

Cell lines and CTC model

The BT474 cell line (from the Gray Lab) was used for proof of principle experiments [16]. Cells were grown in Roswell Park Memorial Institute (RPMI) medium supplemented with 10 % FBS and 1 % antibiotics/antimycotics (Invitrogen, Carlsbad, CA). BT474 cells were trypsinized for 5 min and counted using a hemocytometer. Cell count was confirmed with a FACS Calibur using our previously described multimarker FACS assay [13]. 1000–100,000 cells were spiked into 10 mL PB obtained from healthy females; all specimens were processed immediately, and over-exposure to trypsin was avoided by quenching with medium.

Patient samples

Ten to 20 mL of PB was drawn into EDTA tubes from five female MBC patients yielding 9-993 captured CTCs from which expression profiles were obtained; specimens were processed immediately by IE/FACS. Microarray data for primary tumors prior to neoadjuvant chemotherapy were obtained from the I-SPY1 trial (CALGB 150012, ACRIN 6657). Normal breast ($n = 1$) and skin samples ($n = 3$) were collected as core and punch biopsies. Negative control PB samples were obtained from healthy individuals. All patients gave informed consent under a protocol approved by the University of California San Francisco Institutional Review Board. All patients received standard of care therapy for MBC; all patients had been previously treated with several courses of chemotherapy, and samples were obtained at convenient time points in patient care during routine outpatient laboratory PB draws. We selected patients known to be CTC positive for this pilot study based on prior positive CTC test results.

Cell isolation via IE/FACS

IE/FACS was performed as previously described [13] but with an emphasis on the preservation of RNA to be isolated directly from CTCs. As we previously described in Magbanua et al. [13], our IE/FACS assays involve immunomagnetic separation using EpCAM (MJ37) mAb-coated magnetic beads followed by FACS with EpCAM (EBA-1) mAb conjugated to phycoerythrin (PE), thioflavin nucleic acid dye, and CD45 (2D1) mAb conjugated to Cy5-PerCP (all from BD Biosciences, San Jose CA). A threshold of a single cell meeting these criteria was qualified as a positive test result. We prepared the FACS Aria II (BD Biosciences) with RNAse Zap decontamination solution (Ambion, Austin, TX) prior to all sorting. All samples were processed immediately following blood draws, and all lysates were immediately placed on ice. All CTC subjects' specimens were analyzed with a FACS Aria processed using consistent gates.

RNA extraction and amplification

All sorted samples were stored at −80 °C as cell lysates until the time of RNA isolation with PicoPure RNA isolation kits (Life Technologies). For normal PB, RNA was extracted using the Qiagen Blood RNA kit (Qiagen). Total RNA and equivalent amounts of StratRef (Stratagene Universal Human Pooled Reference RNA, Stratagene, La Jolla, CA) were treated with 200 ng of poly dIdC (Sigma-Aldrich, St. Louis, MO). CTC, BT474, PB, and StratRef samples were linearly amplified with 2 rounds using Arcturus RiboAmpHS (Life Technologies). Primary tumor, normal epithelium, and StratRef were amplified using 2 round modified T7 amplification [15, 17]. Concentrations of amplified RNA products were measured using a UV spectrophotometer. The molecular weight and integrity of amplified RNA species were evaluated using the Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA).

Quantitative RT-PCR

BT474 and StratRef total RNA were converted to cDNA using M-MLV reverse transcriptase and random hexamers (Life Technologies). Samples were then incubated at 25 °C

for 10 min then 48 °C for 30 min. Expression levels of 37 genes and a housekeeping gene GUS (beta-glucuronidase) were analyzed using a 5' nuclease assay and TaqMan Gene Expression Assays with an ABI PRISM 7700 instrument (Applied Biosystems (ABI), Foster City, CA). Relative expression levels were calculated relative to GUS. Calculated QPCR expression ratios were derived using the formula relative expression $= 2(^{-\Delta\ \Delta CT})$ [18]. Log-transformed (base 2) values for each gene expression ratio were plotted for both microarray and QPCR methods. We have previously published details regarding the selection strategy and gene list [15]. Subsequent validation of our assay performance was performed by QPCR of spiked, sorted BT474 and PB for EpCAM, ER, HER2, CD45, HLA, and GAPDH using TaqMan assays.

cDNA Microarrays

The 20,862 human cDNAs used in these studies were purchased from Research Genetics (Huntsville, AL), now Invitrogen, and were provided by the Haqq laboratory as cDNA microarrays. On the basis of Unigene build 166, these clones represent 19,740 independent loci. Hybridization, washing, scanning, and primary data analysis were performed as previously described [19–21]. Slides were then washed, scanned, and analyzed with Axon Imager 4000b (Molecular Devices, Sunnyvale, CA), using GENEPIXPRO3.0.

cDNA microarray data included three different array formats, containing 21,600 (set 1), 21,632 (set 2), and 41,664 (set 3) features, respectively, representing in total 85 samples (5 sorted CTC, 7 plated BT474, 6 sorted BT474, 8 PB, 4 normal epithelium, and 55 primary breast tumors). Microarray data were deposited under GSE45965.

cDNA microarray normalization

Two types of normalization were used. First, each set of arrays was normalized separately. Background correction was not performed as this increased noise in low signal features. Instead, a 2D spatial normalization approach using the robust local regression function 'loess' was performed to correct for uneven hybridization across individual arrays using the marray software package [22] in the R [23]. To achieve consistency between arrays, quantile normalization across samples was performed using the limma package [24] in R. The three sets of arrays were merged into a single dataset by including those features consistent to all array sets. This resulted in 21,006 total genes for the combined dataset of 85 samples. To test for residual batch effects after normalization, we calculated the pairwise correlation matrix for all samples to verify that clustering patterns were driven by cell type rather than cDNA array design and processing batch.

Unsupervised hierarchical clustering and principal component analysis

To characterize differential expression between different cell types, we performed principal component analysis (PCA) and hierarchical clustering using the stats package in R. To verify that unsupervised clustering can reliably classify specimens, the p-values for each cluster were calculated via bootstrap resampling using standard Bootstrap Probability (BP) and a multi-scale bootstrapping method that is a better approximation to Approximately Unbiased (AU) (R package pvclust [25]). Only probes of the array which exhibited a hybridization signal 2 fold above background in at least 10 % of specimens and with none of the spots flagged as bad were used for PCA, clustering, and all other analyses (7685 probes).

Differential gene expression and functional enrichment studies

To identify genes with differential expression in sorted CTC samples relative to primary tumors, normal epithelium, and PB samples, we used ANOVA followed by Tukey analysis with Benjamini-Hochberg (BH) multiple testing correction applied at each step to adjust for potential false discovery (FDR). Genes with a FDR adjusted p value < 0.05 were considered significant. Functional and pathway analysis of differentially expressed genes were carried out with the DAVID functional/pathway enrichment analysis tool [26], using multiple testing corrected p-values that control for false discovery with significance threshold $p < 0.05$.

Intrinsic subtype classification

Samples were evaluated for intrinsic subtype using the PAM50 single-sample classification algorithm developed by Parker and Perou [27]. Of the 50 genes in the PAM50 classifier, 21 were identified in the combined, quality filtered dataset (7685 probes), and used for analysis. The 21 genes available for this analysis were BCL2, CDC6, NUF2, CENPF, CEP55, CXXC5, EGFR, ERBB2, ESR1, FGFR4, FOXC1, GRB7, NDC80, KRT14, MYBL2, PTTG1, RRM2, TMEM45B, TYMS, UBE2C, and UBE2T. The remaining 29 genes in the PAM50 were either not present on the cDNA array probe set or did not meet the quality control criteria described above. We used a 90 % confidence interval (CI) in the subtype call to establish that gene expression levels of the 21 genes were correlated to the centroid of the subtype (otherwise the sample was considered unclassified). To assess differential risk of recurrence (ROR $\sim$ P) scores across sample types, we used ANOVA followed by Tukey analysis.

Results

Assay validation using spiked-in breast cancer cells

Our approach to CTC isolation and expression array profiling was first validated using a spiked-in cancer cell line. BT474 cells were spiked into PB or PBS then subjected to IE/FACS; supplemental Fig. 1 shows a representative FACS gating strategy. As a control for the amplification procedure, 500 pg of bulk BT474 RNA was amplified and subjected to cDNA microarray analysis. Expression profiles for all sorted BT474 conditions were similar to the expected BT474 profile (Fig. 1a), which was obtained from non-amplified bulk total RNA analyzed in parallel as a positive control. We previously reported a comparison of amplification techniques at the picogram input level using BT474 as the test cell line [15].

Unsupervised hierarchical clustering confirmed that all BT474 expression profiles were closely related, regardless of whether the cells were spiked into blood, PBS or were bulk cells derived from cell culture; this was independent of the RNA amplification procedure (Fig. 1a). Profiles of BT474 cells isolated from blood were distinct from that of PB (Fig. 1b), demonstrating that IE/FACS-isolated cells appeared devoid of leukocyte contamination. This approach demonstrated gene clusters that were differentially expressed between BT474 cells and blood. A distinct cluster of genes was observed to be upregulated in blood but not in BT474 cells (Fig. 1c), these included fibrinogen-related procoagulant Fgl2, pleckstrin (platelet and leukocyte C kinase substrate), and DCL1, which is frequently expressed in leukocytes but downregulated in cancers, as it is a tumor suppressor gene. Conversely, genes such as HER2 and ESR1 (ER) (Figs. 1d, e), as expected, were upregulated in BT474 but not in blood.

Sorted, spiked BT474 cells via our IE/FACS approach showed no expression of leukocyte markers CD45 or HLA by QPCR, but positive control markers EpCAM, ER, and HER2 were all expressed (Supplemental Fig. 2). As shown in the PCA in Fig. 1f, samples clustered according to tissue of origin rather than batch, confirming successful normalization.

To assess the reproducibility of whole transcriptome amplification, we performed a global pair-wise Pearson correlation of microarrays (Supplemental Fig. 2). Results confirmed high correlation among all BT474 profiles.

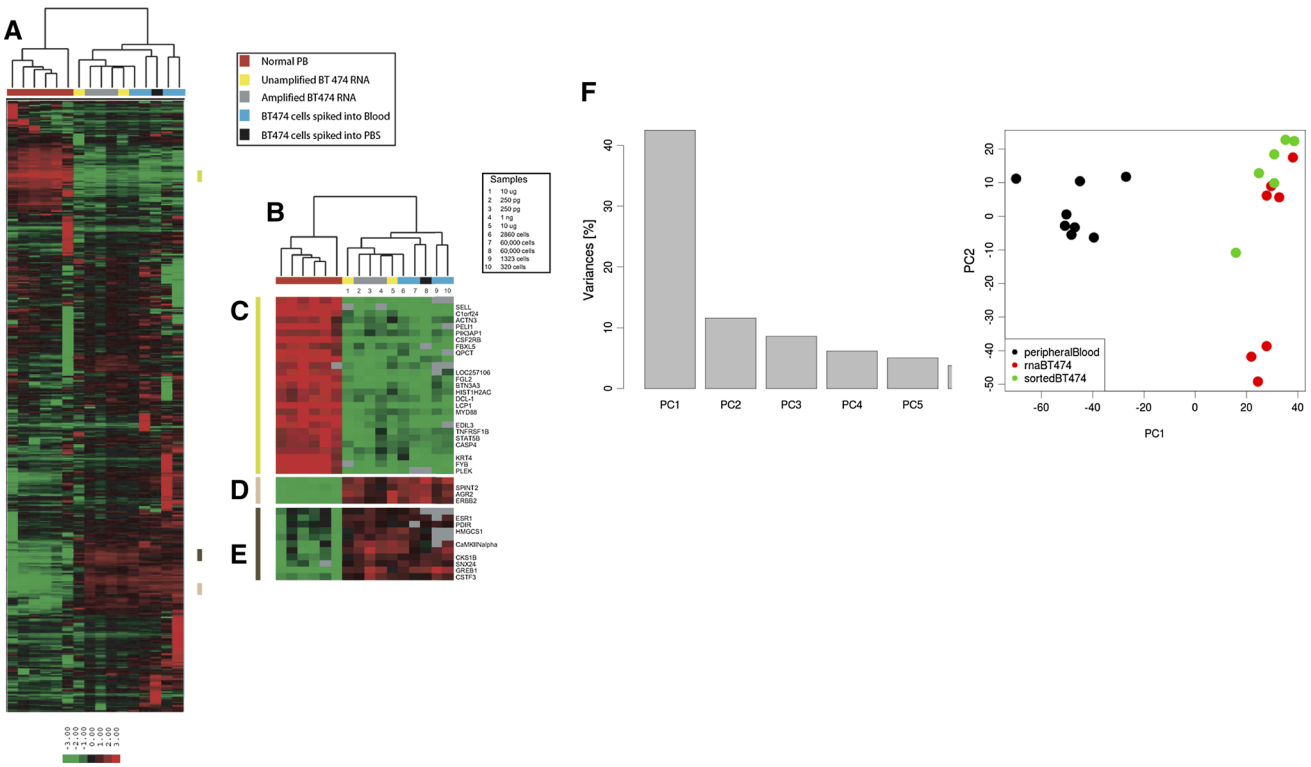

Fig. 1 Microarray analysis of spiked, sorted cells and controls. Unsupervised hierarchical clustering of six samples of normal PB (*red*), three samples of amplified (*gray*), and two samples of unamplified (*yellow*) BT474 RNA, and four replicates of BT474 cells spiked into blood (*blue*) or PBS (*black*) (one sample); **a** A scaled-down representation of the entire cluster of 8041 genes based on similar gene expression; **b** An enlarged dendrogram showing the samples; **c** Gene cluster showing genes that are highly expressed in normal PB; **d** Gene cluster containing HER2 (Erbb-2); **e** Gene cluster containing ESR1. *Brown*, *tan* and *green vertical bars* correspond to location of the gene clusters. The *color bar* scale denotes fold upregulation (*red*) and fold downregulation (*green*). **f** Principal component analysis demonstrated that spiked, sorted BT474 cells are more similar to bulk BT474 RNA than peripheral blood

Furthermore, no correlation was observed between any BT474 profile and PB, including the profile of BT474 cells isolated from blood.

QPCR independently validated the microarray analysis. Thirty-seven target genes were chosen for comparison of QPCR and microarray data. Expression results showed 86.4 % agreement (32/37 genes) between microarray and QPCR (Supplemental Fig. 3).

Isolation and gene expression profiling of CTCs from MBC patients

We next tested the feasibility of this approach in five MBC patients. RNA from $n = 9$–993 CTCs was captured and isolated using IE/FACS, linearly amplified, and subjected to gene expression profiling via cDNA microarrays. We also profiled gene expression in a panel of primary tumors, PB, and normal epithelium using the same platform.

Unsupervised hierarchical clustering of the 500 most variable genes from the CTC samples showed that CTCs clustered separately from PB and normal epithelium (Fig. 2a), a partitioning also evident in PCA analysis of all 7685 probes (Fig. 2b). Moreover, a bootstrapping analysis of the clustering pattern in the dendrogram produced an unbiased estimate of $p = 1$ that PB clustered separately from CTCs, tumors, and epithelium. These analyses provide strong evidence that IE/FACS-isolated CTCs from the PB of patients allowing profiling their gene expression on a transcriptome scale.

Intrinsic subtype and ROR analysis

Intrinsic subtype analysis of the CTCs classified 3/5 samples with a 90 % CI, however, 2/5 could not be classified using the 21 available genes of the PAM50 on the cDNA array. Similarly, 35/55 (63.6 %) primary tumors had subtype calls. The CTC specimens that could be assigned subtype were HER2-like (2/5) and Luminal B (1/5), a distribution consistent with the expected HER2 status based on the primary tumors for these MBC patients (4/5 were HER2 positive) (Table 1). As expected given the classifications, in an unsupervised hierarchical clustering of the available PAM50 genes (21/50), the CTCs were

Table 1 Numbers of CTCs analyzed per MBC patient and primary tumor characteristics

Patient	No. of CTCs analyzed	Volume of blood	CTCs/ml	Her2	ER	PR
1	993	16	62.1	+	+	+
2	456	10	45.6	−	+	+
3	9	20	0.5	+	+	+
4	279	10	27.9	+	+	+
5	195	20	9.8	+	−	−

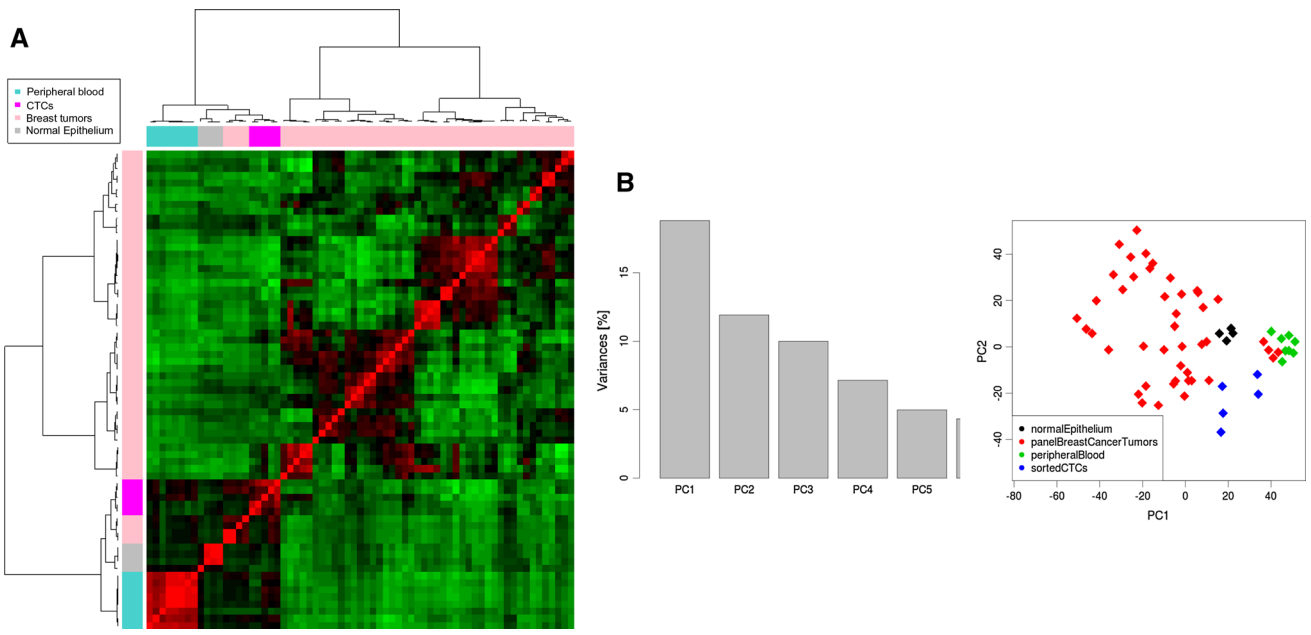

Fig. 2 Gene expression profiling of Stage IV patient CTCs ($n = 5$) versus peripheral blood ($n = 6$), breast tumors ($n = 55$), and normal epithelium ($n = 4$). **a** Unsupervised hierarchical clustering of the top 500 most variable genes (ward linkage) is shown. CTCs cluster most closely with breast tumors. **b** A principal component analysis demonstrated that sorted CTCs form a group distinct from peripheral blood

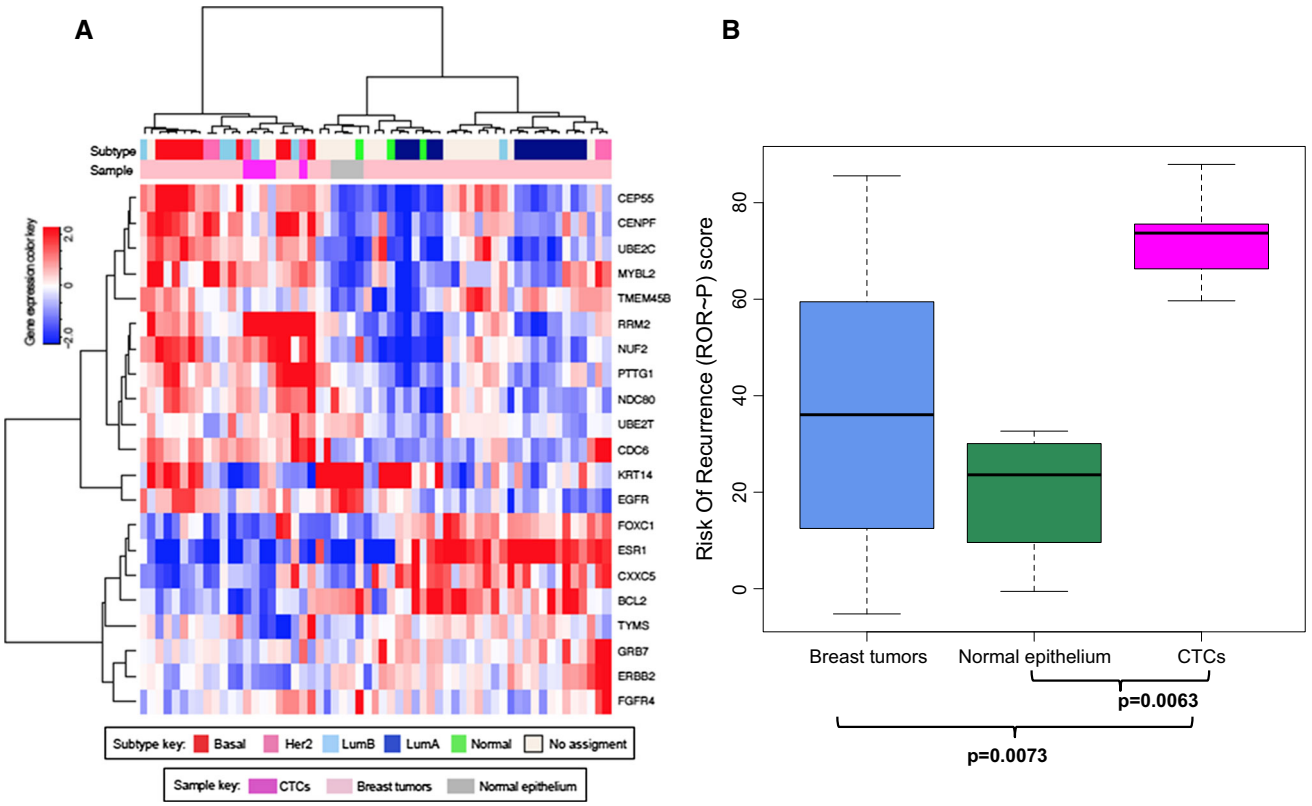

Fig. 3 **a** Intrinsic subtype analysis. Exploratory analysis based on the subset of 21 genes from the PAM50 classifier available in our dataset for analysis, CTCs clustered with the more aggressive breast cancer subtypes (HER2 positive and luminal B) based on a 90 % confidence interval ($p < 0.01$) for assigning PAM50 subtype based on the available probes in the cDNA microarray. **b** As expected, CTCs derived from Stage IV patients had a higher risk of recurrence score in comparison to I-SPY1 primary tumors (most of which were high risk by a 70-gene classifier)

clustered with HER2-enriched, Luminal B, and Basal-like primary breast tumors (see Fig. 3b). Although the CTCs had low ERBB2 (HER2) expression, high RRM, CDC6, GRB7, UBE2T, and FGFR4 and low ESR1, FOXC1, KRT14, and BCL2 resulted in classification of two samples with the HER2 centroid.

We calculated the ROR ∼ P [27] for CTCs, breast tumors, and normal epithelium as a continuous score and present these results in Fig. 3b. This box plot shows that CTCs had a significantly higher ROR ∼ P score on average than do breast tumors, as expected given that CTCs were isolated from MBC patients. This is connected to more aggressive subtype (HER2 and Luminal B). ROR ∼ P scores were different across the different types of samples (F-test $p = 0.0038$), with significantly higher levels in CTCs relative to breast tumors ($p = 0.0073$), and normal epithelial ($p = 0.0063$).

Pathway analysis

We performed a supervised analysis using ANOVA and Tukey post hoc statistical tests to compare CTCs to PB cells, breast tumors, and normal epithelium. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg (BH) algorithm. Of the 7,685 evaluable probes, 1,116 genes were differentially expressed between CTCs and PB, and 1,214 between CTCs and primary breast tumors (see Fig. 4 and Supplemental Table 2).

Pathway and functional enrichment analysis applied to differentially expressed genes, performed using the software tool DAVID with BH adjusted p-values to correct for multiple comparisons (BH $p < 0.05$), identified a number of immune-related pathways and categories significantly upregulated in PB relative to CTCs (see Supplemental Table 3). In addition to a low level of immune signaling, CTCs differed from PB and also from primary tumors in that they have significantly lower levels of apoptosis signaling (Fig. 5). Although no specific pathways were found to be upregulated in CTCs, Supplemental Table 2 provides a complete list of genes up or downregulated in CTCs. Genes downregulated in CTCs relative to PB were enriched for a plethora of ribosome-related pathways and terms. Together, these results suggest that the CTCs had downregulated apoptotic pathways, were distinguishable from PB cells by

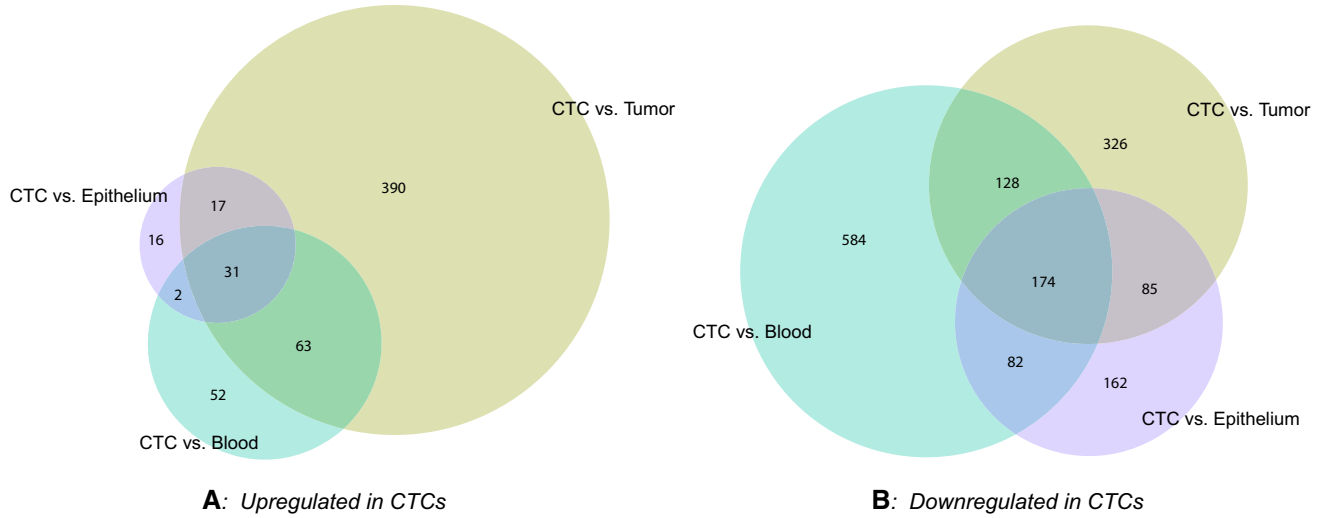

Fig. 4 Venn diagram comparing CTCs, breast tumors, peripheral blood, and normal epithelium. This analysis is based on ANOVA followed by Tukey analysis with a BH p value < 0.05. **a** Genes upregulated in CTCs; **b** Genes downregulated in CTCs

the absence of immune signals, and were characterized by downregulation of ribosomes despite high expression of multiple proliferation genes in the PAM50 gene set.

Negative controls

As a negative control, we performed IE/FACS on healthy females at UCSF ($n = 23$) and USC ($n = 10$). IE/FACS yielded 0–0.5 CTCs per 20 mL of PB.

Discussion

This study demonstrates the feasibility of expression profiling of pure CTCs. Using IE/FACS, we isolated small populations of CTCs and amplified their RNA for microarray analysis. Five MBC patients' CTCs were profiled without evidence of contaminating leukocyte RNA. Our proof of principle data with spiked, sorted cells demonstrated that IE/FACS could isolate CTCs and characterize them without altering their gene expression by either assay manipulation or contamination by background leukocytes. Additionally, negative controls showed essentially no evidence of CTCs. Unsupervised hierarchical clustering analysis differentiated CTC profiles from PB. CTCs clustered most closely with a panel of primary breast tumors but not with normal epithelial samples, in keeping with their malignant nature. Our observation that CTCs clustered with the HER2-enriched and basal-like groups in an exploratory analysis are consistent with the predominance of HER2 positivity of these patients' tumors, and should be validated in larger trials taking into consideration tumor biology and treatment variables [28].

CTCs had a higher ROR score than a panel of primary tumors from the I-SPY1 trial (pre-neoadjuvant chemotherapy). The vast majority of the primary tumors for patients participating in I-SPY1 [29] were found to be high ROR based on the 70 gene classifier [30, 31]. We demonstrated that CTCs gene expression also appeared to be aggressive by a genomic classifier. We recognize that CTCs were from a cohort of five MBC patients, and therefore it is anticipated that their gene expression would show a high ROR. Our intent with this experiment is to provide further evidence of the malignant nature of CTCs and of our ability to profile pure populations of such cells as proof of principle rather than to definitively classify them.

Genomic classifiers in breast cancer are typically based on early stage disease rather than MBC [27, 32]. However, given the unavailability of fresh frozen biopsies of MBC to be interrogated on the same cDNA platform, we established several lines of evidence that we have a methodology for gene expression profiling of pure populations of EpCAM-positive CTCs. Our data provide evidence that profiles of CTCs captured by IE/FACS are consistent with breast cancer. This approach could be used as a "liquid biopsy" since PB is much more readily available than tumor tissue and could be re-assessed longitudinally to assess tumor biology in real time as tumors evolve under the selection pressure of cytotoxic therapies.

Using the same IE/FACS platform, Magbanua et al. successfully profiled 102 MBC patients' CTCs by array comparative genomic hybridization [14]. In the present study, we demonstrated the ability to isolate high quality RNA from highly purified CTCs suitable for microarray profiling. Taken together, RNA and DNA profiling could

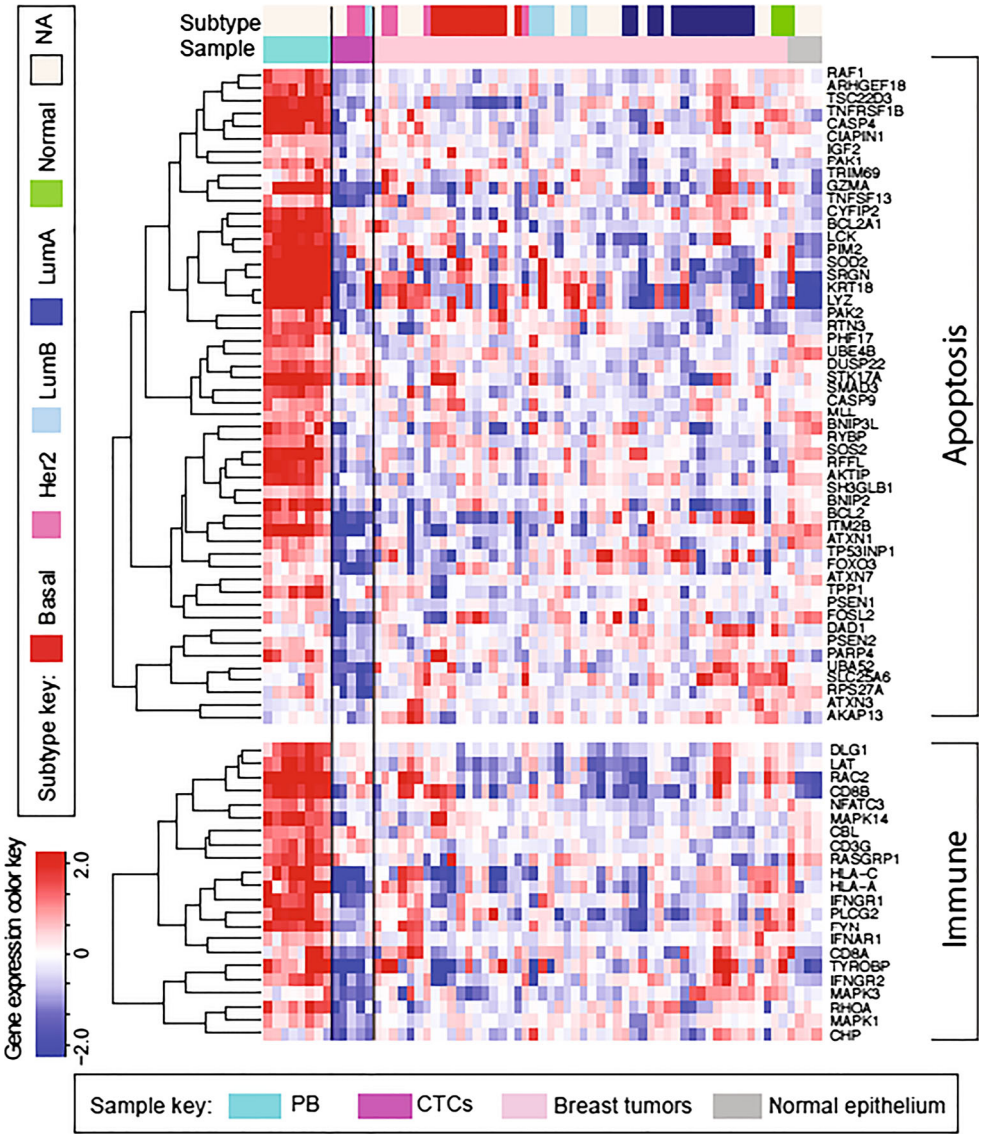

Fig. 5 Pathway analysis. A DAVID functional enrichment analysis demonstrated that CTCs were downregulated for the pathways of apoptosis and immune function relative to peripheral blood

be utilized to further characterize the biology of CTCs and identify opportunities for targeted therapy and potentially novel therapies.

Our exploratory analysis suggested that the CTCs had downregulated apoptotic pathways, were distinguishable from PB cells by an absence of immune signals, and were characterized by a curious downregulation of ribosomes relative to all other cell types, perhaps suggesting a relatively quiescent state while in transit through the blood stream as suggested by Meng et al. [33]. However, ribosomal RNA is neither specifically amplified nor depleted using RNA amplification. This calls into question whether ribosomal RNA is truly downregulated in CTCs or if this represents a problem related to transcript abundance in the starting material causing differential degrees of ribosomal RNA detection. Another explanation could be amplification artifact since although both the CTCs and tumors were subjected to two rounds of linear amplification, different amplification strategies were used since the tumors were profiled as part of the I-SPY1 trial rather than for the purpose of comparison to CTCs. Previously, we found inter-method Pearson correlations of 0.85-0.92 between these amplification methods after taking into consideration data processing [15]. Proliferation markers such as MYBL2 and RRM2 were strongly upregulated in CTCs, which is at odds with the finding of ribosomal inactivity. Further ongoing studies of expression profiling of CTCs will be required to evaluate this apparent incongruent finding of downregulated ribosomal gene expression in CTCs. Additionally, future studies analyzing individual CTCs will shed light on the heterogeneity of CTC biology.

SWOG S0500 found that changing to an alternative chemotherapy in MBC based on CTCs did not improve survival [34]. However, the authors acknowledged that molecular

profiling of CTCs might provide predictive information that simple enumeration does not. In contrast to several previous reports, IE/FACS yields pure CTC populations without requiring background subtraction of residual PB.

We acknowledge that a major limitation of this study is the small sample size of CTCs derived from MBC. This study was intended to demonstrate feasibility of global gene expression profiling of rare CTCs. Future studies including larger numbers of samples are warranted to provide a better understanding of the biology of CTCs. Comparison of gene expression profiles of CTCs with primary tumors and metastatic lesions may shed light on tumor evolution and progression. Unfortunately, primary tumor samples were unavailable for the MBC patients in our study. As gene expression platforms have evolved, additional CTC patients cannot be accrued and co-analyzed with this proof of concept data, necessitating further prospective studies involving contemporary gene expression assays, such as RNA Seq; these studies are currently in progress.

In summary, we demonstrate the feasibility of gene expression profiling of rare CTCs. Molecular characterization of CTCs may yield insights into their potential as biomarkers to allow for specific targeting of these cells in patients who respond poorly to current therapies.

Acknowledgments We would like to thank Daisuke Nonaka, Daniel Khodabakhsh, Kavitha Krishnan, Hema Parmar, Alfred Au, Scot Federman, Pamela Derish, David Ginzinger, Joe Gray, and Eduardo Sosa for their contributions to this work. We acknowledge William Hyun and the UCSF Comprehensive Cancer Center Molecular Cytometry Core. We acknowledge Emily Park and Tom Frey of BD Biosciences for contributing reagents used for these experiments.

Financial support Dr. Park's laboratory was supported by the National Cancer Institute (NCI) Interdisciplinary Research Teams for Molecular Target Assessment (U54 CA90788). Dr. Lang's laboratory was supported by a Ginny Clements Research Award, a Better than Ever Award (University of Arizona), a Society of Surgical Oncology Clinical Investigator Award, a California Breast Cancer Research Program Idea Award, and a STOP Cancer Award. The project described was supported in part by award number P30CA014089 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. Southwest Environmental Health Sciences Center Grant ES006694 supported Petr Novak. We acknowledge the I-SPY Program for additional statistical support.

Conflict of interest Christopher M. Haqq owns stock in Johnson & Johnson. The other authors have no relevant financial disclosures.

References

1. Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, Tibbe AG, Uhr JW, Terstappen LW (2004) Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clin Cancer Res 10(20):6897–6904. doi:10.1158/1078-0432.CCR-04-0378
2. Budd GT, Cristofanilli M, Ellis MJ, Stopeck A, Borden E, Miller MC, Matera J, Repollet M, Doyle GV, Terstappen LW, Hayes DF (2006) Circulating tumor cells versus imaging–predicting overall survival in metastatic breast cancer. Clin Cancer Res 12(21):6403–6409
3. Liu MC, Shields PG, Warren RD, Cohen P, Wilkinson M, Ottaviano YL, Rao SB, Eng-Wong J, Seillier-Moiseiwitsch F, Noone AM, Isaacs C (2009) Circulating tumor cells: a useful predictor of treatment efficacy in metastatic breast cancer. J Clin Oncol 27(31):5153–5159. doi:10.1200/JCO.2008.20.6664
4. Maheswaran S, Sequist LV, Nagrath S, Ulkus L, Brannigan B, Collura CV, Inserra E, Diederichs S, Iafrate AJ, Bell DW, Digumarthy S, Muzikansky A, Irimia D, Settleman J, Tompkins RG, Lynch TJ, Toner M, Haber DA (2008) Detection of mutations in EGFR in circulating lung-cancer cells. N Engl J Med 359(4):366–377. doi:10.1056/NEJMoa0800668
5. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM, Concannon KF, Donaldson MC, Sequist LV, Brachtel E, Sgroi D, Baselga J, Ramaswamy S, Toner M, Haber DA, Maheswaran S (2013) Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science 339(6119):580–584. doi:10.1126/science.1228522
6. Sieuwerts AM, Mostert B, Bolt-de Vries J, Peeters D, de Jongh FE, Stouthard JM, Dirix LY, van Dam PA, Van Galen A, de Weerd V, Kraan J, van der Spoel P, Ramirez-Moreno R, van Deurzen CH, Smid M, Yu JX, Jiang J, Wang Y, Gratama JW, Sleijfer S, Foekens JA, Martens JW (2011) mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients. Clin. Cancer Res 17(11):3600–3618. doi:10.1158/1078-0432.CCR-11-0255
7. Smirnov DA, Zweitzig DR, Foulk BW, Miller MC, Doyle GV, Pienta KJ, Meropol NJ, Weiner LM, Cohen SJ, Moreno JG, Connelly MC, Terstappen LW, O'Hara SM (2005) Global gene expression profiling of circulating tumor cells. Cancer Res 65(12):4993–4997. doi:10.1158/0008-5472.CAN-04-4330
8. Ring A, Smith IE, Dowsett M (2004) Circulating tumour cells in breast cancer. Lancet Oncol 5(2):79–88
9. Xenidis N, Perraki M, Kafousi M, Apostolaki S, Bolonaki I, Stathopoulou A, Kalbakis K, Androulakis N, Kouroussis C, Pallis T, Christophylakis C, Argyraki K, Lianidou ES, Stathopoulos S, Georgoulias V, Mavroudis D (2006) Predictive and prognostic value of peripheral blood cytokeratin-19 mRNA-positive cells detected by real-time polymerase chain reaction in node-negative breast cancer patients. J Clin Oncol 24(23):3756–3762
10. Apostolaki S, Perraki M, Kallergi G, Kafousi M, Papadopoulos S, Kotsakis A, Pallis A, Xenidis N, Kalmanti L, Kalbakis K, Agelaki S, Kalykaki A, Stournaras C, Stathopoulos E, Georgoulias V, Mavroudis D (2009) Detection of occult HER2 mRNA-positive tumor cells in the peripheral blood of patients with operable breast cancer: evaluation of their prognostic relevance. Breast Cancer Res Treat 117(3):525–534. doi:10.1007/s10549-008-0239-3
11. Sieuwerts AM, Kraan J, Bolt-de Vries J, van der Spoel P, Mostert B, Martens JW, Gratama JW, Sleijfer S, Foekens JA (2009) Molecular characterization of circulating tumor cells in large quantities of contaminating leukocytes by a multiplex real-time PCR. Breast Cancer Res Treat 118(3):455–468. doi:10.1007/s10549-008-0290-0
12. Powell AA, Talasaz AH, Zhang H, Coram MA, Reddy A, Deng G, Telli ML, Advani RH, Carlson RW, Mollick JA, Sheth S, Kurian AW, Ford JM, Stockdale FE, Quake SR, Pease RF, Mindrinos MN, Bhanot G, Dairkee SH, Davis RW, Jeffrey SS (2012) Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. PLoS One 7(5):e33788. doi:10.1371/journal.pone.0033788

13. Magbanua MJ, Park JW (2013) Isolation of circulating tumor cells by immunomagnetic enrichment and fluorescence-activated cell sorting (IE/FACS) for molecular profiling. Methods. doi:10.1016/j.ymeth.2013.07.029
14. Magbanua MJ, Sosa EV, Roy R, Eisenbud LE, Scott JH, Olshen A, Pinkel D, Rugo HS, Park JW (2013) Genomic profiling of isolated circulating tumor cells from metastatic breast cancer patients. Cancer Res 73(1):30–40. doi:10.1158/0008-5472.CAN-11-3017
15. Lang JE, Magbanua MJ, Scott JH, Makrigiorgos GM, Wang G, Federman S, Esserman LJ, Park JW, Haqq CM (2009) A comparison of RNA amplification techniques at sub-nanogram input concentration. BMC Genom 10:326. doi:10.1186/1471-2164-10-326
16. Lasfargues EY, Coutinho WG, Redfield ES (1978) Isolation of two human tumor epithelial cell lines from solid breast carcinomas. J Natl Cancer Inst 61(4):967–978
17. Baugh LR, Hill AA, Brown EL, Hunter CP (2001) Quantitative analysis of mRNA amplification by in vitro transcription. Nucleic Acids Res 29(5):E29
18. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25(4):402–408. doi:10.1006/meth.2001.1262
19. Haqq C, Nosrati M, Sudilovsky D, Crothers J, Khodabakhsh D, Pulliam BL, Federman S, Miller JR 3rd, Allen RE, Singer MI, Leong SP, Ljung BM, Sagebiel RW, Kashani-Sabet M (2005) The gene expression signatures of melanoma progression. Proc Natl Acad Sci USA 102(17):6092–6097
20. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephaniants SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19(4):342–347
21. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 14(4):457–460
22. Yang YH, Paquet A, Dudoit S (2009) marray: Exploratory analysis for two-color spotted microarray data.. R package version 1.36.0. edn.,
23. R_Core_Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
24. Smyth GK (2005) Limma: linear models for microarray data. Bioinformatics and Computational Biology Solutions using R and Bioconductor. Springer, New York
25. Suzuki R, Hidetoshi S (2011) pvclust: Hierarchical clustering with P-values via multiscale bootstrap resampling. R package version 1.2–2. edn.,
26. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 4(5):P3
27. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27(8):1160–1167. doi:10.1200/JCO.2008.18.1370
28. McShane LM, Hayes DF (2012) Publication of tumor marker research results: the necessity for complete and transparent reporting. J Clin Oncol 30(34):4223–4232. doi:10.1200/JCO.2012.42.6858
29. Esserman LJ, Berry DA, DeMichele A, Carey L, Davis SE, Buxton M, Hudis C, Gray JW, Perou C, Yau C, Livasy C, Krontiras H, Montgomery L, Tripathy D, Lehman C, Liu MC, Olopade OI, Rugo HS, Carpenter JT, Dressler L, Chhieng D, Singh B, Mies C, Rabban J, Chen YY, Giri D, van 't Veer L, Hylton N (2012) Pathologic complete response predicts recurrence-free survival more effectively by cancer subset: results from the I-SPY 1 TRIAL–CALGB 150007/150012, ACRIN 6657. J Clin Oncol 30(26):3242–3249. doi:10.1200/JCO.2011.39.2779
30. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871):530–536. doi:10.1038/415530a
31. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347(25):1999–2009. doi:10.1056/NEJMoa021967347/25/1999
32. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100(14):8418–8423
33. Meng S, Tripathy D, Frenkel EP, Shete S, Naftalis EZ, Huth JF, Beitsch PD, Leitch M, Hoover S, Euhus D, Haley B, Morrison L, Fleming TP, Herlyn D, Terstappen LW, Fehm T, Tucker TF, Lane N, Wang J, Uhr JW (2004) Circulating tumor cells in patients with breast cancer dormancy. Clin. Cancer Res 10(24):8152–8162
34. Smerage JB, Barlow WE, Hortobagyi GN, Winer EP, Leyland-Jones B, Srkalovic G, Tejwani S, Schott AF, O'Rourke MA, Lew DL, Doyle GV, Gralow JR, Livingston RB, Hayes DF (2014) Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. J Clin Oncol. doi:10.1200/JCO.2014.56.2561