

mRNA and microRNA Expression Profiles in Circulating Tumor Cells and Primary Tumors of Metastatic Breast Cancer Patients

Anieta M. Sieuwerts¹, Bianca Mostert², Joan Bolt-de Vries¹, Dieter Peeters⁶, Felix E. de Jongh⁴, Jacqueline M.L. Stouthard⁵, Luc Y. Dirix⁶, Peter A. van Dam⁶, Anne Van Galen¹, Vanja de Weerd¹, Jaco Kraan², Petra van der Spoel², Raquel Ramirez-Moreno⁷, Carolien H.M. van Deurzen³, Marcel Smid¹, Jack X. Yu⁸, John Jiang⁸, Yixin Wang⁸, Jan W. Gratama², Stefan Sleijfer², John A. Foekens¹, and John W.M. Martens¹

Abstract

Purpose: Molecular characterization of circulating tumor cells (CTC) holds great promise. Unfortunately, routinely isolated CTC fractions currently still contain contaminating leukocytes, which makes CTC-specific molecular characterization extremely challenging. In this study, we determined mRNA and microRNA (miRNA) expression of potentially CTC-specific genes that are considered to be clinically relevant in breast cancer.

Experimental Design: CTCs were isolated with the epithelial cell adhesion molecule-based CellSearch Profile Kit. Selected genes were measured by real-time reverse transcriptase PCR in CTCs of 50 metastatic breast cancer patients collected before starting first-line systemic therapy in blood from 53 healthy blood donors (HBD) and in primary tumors of 8 of the patients. The molecular profiles were associated with CTC counts and clinical parameters and compared with the profiles generated from the corresponding primary tumors.

Results: We identified 55 mRNAs and 10 miRNAs more abundantly expressed in samples from 32 patients with at least 5 CTCs in 7.5 mL of blood compared with samples from 9 patients without detectable CTCs and HBDs. Clustering analysis resulted in 4 different patient clusters characterized by 5 distinct gene clusters. Twice the number of patients from cluster 2 to 4 had developed both visceral and nonvisceral metastases. Comparing transcript levels in CTCs with those measured in corresponding primary tumors showed clinically relevant discrepancies in estrogen receptor and HER2 levels.

Conclusions: Our study shows that molecular profiling of low numbers of CTCs in a high background of leukocytes is feasible and shows promise for further studies on the clinical relevance of molecular characterization of CTCs. *Clin Cancer Res*; 17(11); 3600–18. ©2011 AACR.

Introduction

Molecular characterization of primary tumors has already greatly contributed to the personalized treatment of cancer

Authors' Affiliations: ¹Department of Medical Oncology, Josephine Nefkens Institute and Cancer Genomics Centre, ²Department of Medical Oncology, Daniel den Hoed Cancer Center, and ³Department of Pathology, Josephine Nefkens Institute, Erasmus Medical Center; ⁴Department of Internal Medicine, Ikazia Hospital; ⁵Department of Internal Medicine, Maastad Hospital, Rotterdam, The Netherlands; ⁶Translational Cancer Research Group, Laboratory of Pathology, University Hospital Antwerp and General Hospital Sint-Augustinus, Wilrijk, Belgium; ⁷Department of Biochemistry and Molecular Biology Physiology, Genetic and Immunology, Canary Islands Cancer Research Institute, University of Las Palmas de Gran Canaria, Canary Islands, Spain; ⁸Molecular Diagnostics R&D, Veridex LLC, North Raritan, New Jersey

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Author: Anieta M. Sieuwerts, Josephine Nefkens Institute, Erasmus Medical Center, Room BE-400, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands. Phone: +31-10-7044372; Fax: +31-10-7044377. E-mail: a.sieuwerts@erasmusmc.nl

doi: 10.1158/1078-0432.CCR-11-0255

©2011 American Association for Cancer Research.

patients. High-throughput techniques have yielded the knowledge of mutations or epigenomic changes in certain genes and prognostic and predictive models on the basis of mRNA and microRNA (miRNA) expression profiles (1–6). Combined with classical tumor characteristics, these models are increasingly used to guide individualized treatment of patients, thereby aiming to avoid over- or undertreatment. However, most of these prognostic and predictive models have been developed based on primary tumor tissue, whereas metastases, rather than the primary tumor, determine the clinical outcome of cancer patients. It has been shown that metastases, which may develop several years after occurrence of the primary tumor and after prior systemic therapy in the adjuvant or neoadjuvant setting, can differ greatly from primary tumor tissue in terms of genetic characteristics (7–13). It is therefore anticipated that molecular characterization of metastases will improve the currently available prognostic and predictive models. Taking biopsies from metastases in patients, however, is an invasive and often painful procedure, and frequently impossible due to the lack of accessible lesions.

Translational Relevance

Metastases, which may develop several years after occurrence of the primary tumor and after prior (neo)adjuvant therapy, can differ greatly from primary tumor tissue in terms of genetic characteristics. Taking biopsies from metastases in patients, however, is an invasive procedure and frequently impossible due to the lack of accessible lesions. Circulating tumor cells (CTC) are tumor cells shed from either the primary tumor or its metastases that circulate in the peripheral blood of patients and can thus be regarded as "liquid biopsies" of metastasizing cells. In this study, we show for the first time the feasibility of extensive molecular characterization of CTCs at both the mRNA and microRNA level in a high background of leukocytes and show its applicability in a cohort of 50 metastatic breast cancer patients. It is anticipated that such an extensive molecular characterization of CTCs will improve the currently available prognostic and predictive models on the basis of primary tissue.

Circulating tumor cells (CTC) are found in the peripheral blood of patients and are shed from either the primary tumor or its metastases. A recently developed technology to quantify the number of CTCs in whole blood (WB) is the CellSearch CTC Test (Veridex LLC). So far, this is the only test that has been approved by the U.S. Food and Drug Administration (FDA; ref. 14) for the detection and enumeration of CTCs in metastatic prostate (15), colorectal (16), and breast (17) cancer as an independent prognostic factor. After enrichment using magnetic beads coated with anti-epithelial cell adhesion molecule (EpCAM) antibodies, isolated cells are stained with fluorescently labeled monoclonal antibodies specific for epithelial cells (CK-8/18/19), leukocytes (CD45), and their nuclei with a nuclear staining dye [4', 6 diamidino 2 phenylindole (DAPI)], and subsequently enumerated by a semiautomated fluorescence microscope.

In addition to enumeration, CTCs can also be isolated for molecular characterization. This may enable insight into the molecular biology of metastasis, the association of their molecular profiles with treatment outcomes, and reveal the presence of potential drugable targets. However, although EpCAM-based enrichment eliminates a large proportion of leukocytes (approximately 4-log depletion), there are still considerable quantities of contaminating leukocytes (DAPI⁺/CD45⁺) present after this enrichment (18). This contamination, together with the low frequency of CTCs, forms a challenge when aiming to characterize CTCs by very sensitive molecular methods such as PCR.

Despite these challenges, we have recently shown the feasibility of determining mRNA expression of epithelial-specific genes in CTC-enriched samples (18). In addition to mRNA, another class of RNAs that increasingly attracts attention is the group of miRNAs. Each miRNA targets, on average, 200 mRNA transcripts by which miRNAs

execute widespread control (19). As might be expected based on these activities, altered expression of specific miRNA genes has already been shown to contribute to the initiation and progression of cancer (20–22). Therefore, miRNA-based cancer gene therapy offers the theoretical appeal of targeting multiple gene networks that are controlled by a single aberrantly expressed miRNA (23), making the profiling of miRNAs in cancer even more appealing, especially in the context of CTCs.

Here, we describe the optimization of a method to perform both miRNA and mRNA expression analysis for multiple genes by real-time reverse transcriptase PCR (RT-PCR) on as little as 5 CTCs isolated from 7.5 mL of blood, which is considered the clinically relevant cutoff in patients with metastatic breast cancer (24–26), in an environment containing excess quantities of up to 1,000 (18) contaminating leukocytes. As shown in this study for patients with metastatic breast cancer, this robust and novel method allows the simultaneous determination of 65 epithelial tumor cell-specific miRNA and mRNA expression levels in CTCs enriched by CellSearch, and the exploration of their clinical relevance on the basis of the identification of 4 different patient clusters with distinct characteristics.

Materials and Methods

Ethics statement

This study was approved by the Erasmus MC and local Institutional Review Boards (METC 2006-248), and all donors and patients gave their written informed consent.

Breast tumor tissues and blood samples

From 61 patients with metastatic breast cancer, 2 × 7.5 mL blood samples were prospectively taken for CTC enumeration and isolation (for details see next) prior to initiation of systemic therapy for metastatic disease. From these 61 samples, 11 (18%) were excluded because of insufficient RNA quality and/or quantity (for details see next), rendering a total number of 50 patients eligible for further analysis. Metastatic breast cancer patients had been included at the start of systemic therapy between February 2008 and December 2009 in 4 hospitals (9 patients in the Erasmus Medical Center, Rotterdam, The Netherlands, 10 in the Ikazia Hospital, Rotterdam, The Netherlands, and 10 in the Maastad Hospital, Rotterdam, The Netherlands, and 21 patients in the Oncology Center GZA St-Augustinus, Antwerpen, Belgium). For 8 of 32 patients with at least 5 CTCs, primary tumor tissue containing at least 50% invasive epithelial tumor cells was available for RNA isolation [5 fresh frozen (FF) and 3 formalin-fixed paraffin-embedded (FFPE)]. These 8 specimens were used for comparison of transcript levels between CTCs and corresponding primary tumors. Detailed clinicopathological information for these 50 patients and the 8 matching primary tissues is given in Table 1 and in Supplementary Table S1 after dichotomization of patients at the breast cancer clinically relevant level of 5 CTCs (24–26). Fifty-three healthy blood donor (HBD) blood samples

Table 1. Clinical, pathological, and biological characteristics of the patients

Characteristic	No. of patients ^a	% ^a	Primary tissue		Molecular profile		<i>P</i> ^b
			Patients from clusters 2–4 ^a	Patients from cluster 1	Patients from clusters 2–4		
All patients	50	100%	8	14	36		
Time between primary surgery and CTC sampling							
≤5 y	23	46%	5	6	17	1.00	
>5 y	15	30%	1	4	11		
Unknown or primary not removed	12	24%	2	4	8		
Age at CTC sampling							
≤50 y	10	20%	0	4	6	0.44	
>50 y	40	80%	8	10	30		
Menopausal status							
Premenopausal	16	32%	0	6	10	0.51	
Postmenopausal	30	60%	7	8	22		
Grade (Bloom–Richardson)							
I, well differentiated	10	20%	3	3	7		
II, moderately differentiated	17	34%	2	5	12		
III, poorly differentiated	16	32%	3	3	13	0.73	
Pathological tumor size							
pT1, ≤2 cm	10	20%	1	3	7	0.38	
pT2–4, >2 cm	31	62%	7	5	26		
Lymph nodes involved							
No	12	24%	2	4	8	1.00	
Yes	29	58%	5	9	20		
ER status ^c							
Negative	12	24%	3	6	6	0.07	
Positive	38	76%	5	8	30		
PGR status ^c							
Negative	24	48%	5	9	15	0.20	
Positive	22	44%	3	4	18		
HER2/neu status ^c							
Negative	29	58%	8	7	22	1.00	
Positive	12	24%	0	3	9		
Histological type							
Lobular	11	22%	3	3	8	1.00	
Ductal	35	70%	5	9	26		
Adjuvant chemotherapy							
No	34	68%	5	11	23	0.50	
Yes	16	32%	3	3	13		
Adjuvant hormonal therapy							
No	31	62%	7	11	20	0.20	
Yes	19	38%	1	3	16		
Any adjuvant therapy							
No	27	54%	5	5	22	0.13	
Yes	23	46%	3	9	14		
Site of metastasis							
Visceral	12	24%	1	6	6	0.05	
Nonvisceral	7	14%	0	3	4		
Both	31	62%	7	5	26		

^aPercentages <100% can be attributed to missing cases.^b*P* for 2-tailed Fisher's exact test.^cAs retrieved from the pathology reports.

were drawn from laboratory volunteers and blood donors of the Sanquin Blood Bank South-west Region.

Enumeration of CTCs

Prior to the administration of first-line systemic therapy, 7.5 mL of blood from HBDs and metastatic breast cancer patients was drawn in CellSave tubes (Veridex LLC). For CTC enumeration, samples were processed on the CellTracks AutoPrep System (Veridex LLC) by using the CellSearch Epithelial Cell Kit (Veridex LLC) and CTC counts were determined on the CellTracks Analyzer (Veridex LLC) according to the manufacturer's instructions and as described previously (27, 28).

miRNA and mRNA isolation from CTCs, FF, and FFPE

For gene expression studies, in parallel with the enumeration studies, 7.5 mL of blood from the same healthy donors and patients was drawn in EDTA tubes and enriched for CTCs on the CellTracks AutoPrep System using the CellSearch Profile Kit (Veridex LLC). RNA isolation was performed with the AllPrep DNA/RNA Micro Kit (Qiagen) according to the manufacturer's instructions. A more detailed description is given in the Supplementary Materials and Methods. Total RNA was isolated from FF tissue with RNA-Bee as described previously (29) and from FFPE tissue with the column-based High Pure RNA Paraffin Kit (Roche Applied Science) according to the manufacturer's instructions.

Stem-loop cDNA synthesis, preamplification, and real-time PCR (quantitative RT-PCR)

The generation of preamplified cDNA from total RNA from the FF and FFPE tissues and the >200 nucleotide (nt) RNA fractions and subsequent TaqMan-based quantitative RT-PCR (qRT-PCR) analysis, and the validation procedures to ensure homogeneous amplification, were performed as described before (18).

To analyze miRNAs, a multiplex stem-loop cDNA approach was used essentially as described before (21). In brief, up to 50 different RT primers (250 nmol/L each) were pooled, concentrated for 60 minutes in a speed vacuum centrifuge at 50°C, and resuspended in nuclease-free ddH₂O (double distilled water) to a final concentration of 50 nmol/L each. The use of a specific primer with a hairpin structure during cDNA synthesis and mature miRNA-specific detection probes precluded the detection of precursor miRNAs. A total of 25 to 50 ng of total RNA sample aliquots were reverse-transcribed in a final volume of 20 µL with a final concentration of 12.5 nmol/L for each RT primer using the TaqMan miRNA for reverse transcription kit [Applied Biosystems (ABI)] according to the manufacturer's instructions and as described before (21).

For the miRNA quantification in the CTC samples, 3 µL ≤ 200 nt RNA aliquots were reverse-transcribed in a final volume of 7.5 µL with a final concentration of 12.5 nmol/L for each RT primer (ABI), 0.65 mmol/L of each dNTP (ABI), 3 mmol/L magnesium chloride (Invitrogen), 0.3 U/µL RNase inhibitor (ABI), 15 U/µL RevertAid H Minus enzyme

(Fermentas), and 1x RT buffer (Fermentas). Cycling conditions were according to the "Megaplex RT reaction for TaqMan miRNA array" protocol from ABI, i.e., 40 cycles of 16°C for 2 minutes, 42°C for 1 minute, and 50°C for 1 second, followed by a final incubation at 85°C for 5 minutes and a cooldown to 4°C. Prior to PCR, half of the resulting multiplex cDNA was linearly preamplified in 15 cycles according to the manufacturer's instructions (TaqMan PreAmp from ABI) and as described previously for our multiplex gene expression studies (29). Before performing real-time PCRs for each of the miRNAs separately, RT samples were diluted in nuclease-free ddH₂O and analyzed by real time.

PCR in a 20-µL reaction volume in an Mx3000P Real-Time PCR System (Stratagene) using the individual TaqMan miRNA primer and probe assays in combination with TaqMan Universal PCR Master Mix No AmpErase UNG (ABI) with cycling conditions according to the manufacturer's instructions.

To verify that the multiplex RT approach did not affect the quantification of specific miRNAs, all miRNA data were validated in a uniplex RT reaction. A pool consisting of RNA of different human breast tissues was included in each cDNA synthesis and preamplification run, and the resulting data were used to normalize for variation between experiments. In addition, all cDNA synthesis runs incorporated a minus RT reaction, which proved to be negative for all assays in this study. PCR efficiency, linearity, and the upper and lower detection limits of each of the individual miRNA assays were validated with a standard curve prepared of RNA from a pool of breast tumors. Negative controls included samples without RT and samples in which total RNA and cDNA was replaced with ddH₂O. Quantitative values were obtained from the threshold cycle (C_t) at which the increase in TaqMan probe fluorescent signal associated with an exponential increase of PCR products reached the fixed threshold value of 0.08, which was in all cases at least 10-fold higher than the background signal.

First selection of potentially CTC-specific mRNA and miRNA transcripts

The specifics of the used TaqMan assays are given in Supplementary Table S2A for the miRNAs and Supplementary Table S2B for the mRNAs. For the identification phase of potentially CTC-specific miRNA transcripts, the TaqMan Human MicroRNA Assay Set (Sanger miRBase v10; ABI), consisting of 446 unique assays to quantify 436 miRNAs and 10 controls (small nucleolar RNAs; SNORs/RNUs), was used to screen a pool of 150 primary breast cancer RNAs. Of these 446 miRNAs, 253 were expressed in these breast cancer samples and approximately 200 had an expression level of more than 10% of the expression of the reference miRNA set (see next). Next, these levels were compared with those measured in a pool of 6 CellSearch-enriched preparations from HBDs for potentially differentially expressed miRNAs. These prescreen analyses selected 39 miRNAs with both notable expression in breast tumors and at least a 10-fold higher expression in breast tumors

relative to CellSearch enriched HBDs. Four additional miRNAs were included for other reasons, i.e., *hsa-miR-452* to compare with *hsa-miR-452#* and *hsa-miR-379* because of the observed difference between estrogen receptor (ER)-positive and ER-negative samples in the prescreen, *RNU6B* as being a potential reference miRNA, and *hsa-miR-210*, which has shown clinical relevance in breast cancer (refs. 21, 30; Supplementary Table S2A).

For the mRNA transcripts, clinically relevant and potentially CTC-specific genes were selected *in silico* on the basis of literature data and their reported low expression in white blood cells and higher expression in breast tumor tissues, according to the SAGE Genie Database of the Cancer Genome Anatomy Project (<http://cgap.nci.nih.gov/SAGE/AnatomicViewer>). These prescreen analyses were performed as described in detail before (18) and resulted in 90 mRNA transcripts, including 3 reference genes and 2 reference leukocyte markers that could be measured reliably by qRT-PCR and which were potentially higher expressed in breast tumor cells relative to leukocytes (Supplementary Table S2B).

Reference genes, data normalization, and quality control

Unless stated otherwise, levels of *HMBS*, *HPRT1*, and *GUSB* were used to control sample loading and >200 nt RNA quality, as described previously (29). Bone marrow stromal cell antigen 1 (*BST1*) and protein tyrosine phosphatase receptor type C (*PTPRC* coding for CD45) were the control genes for leukocyte background and keratin 19 (*KRT19*) was the control gene for CTC quantification (29).

Although appropriate reference molecules for miRNAs are still unknown for clinical breast cancer cells with a background of leukocytes, previous studies have shown that normalization on mean or median expression of all miRNAs measured in a sample can adequately reduce technical variation (31). Therefore, miRNA data of each individual sample were normalized on the median level of all miRNAs measured in that particular sample.

After verification of equal PCR efficiency for all assays, the relative expression levels were quantified by using the delta C_t method, which is the difference between the median C_t of the appropriate control genes and the C_t of the target gene. Only samples that were at the median C_t of all miRNAs and the median C_t of *HMBS*, *HPRT1*, and *GUSB* able to generate a signal within an arbitrarily chosen cutoff set at 26 C_t were considered of sufficient quality and quantity to enter the study. By the use of this threshold, 11 of our initial 61 patient CTC samples (18%) were excluded from further analysis.

Finally, all transcript data of the 50 CTC samples, 53 HBD controls, and 8 primary tumors were normalized to the C_t of the appropriate reference set, after which, for each individual assay, the median C_t measured in CellSearch-enriched HBDs ($n = 31$ for the mRNAs and $n = 8$ for the miRNAs) was used as a cutoff C_t for the CTC samples. All genes with C_t values exceeding this cutoff C_t were considered to be undetectable.

Statistical analysis

Statistical analysis was done by SPSS 15.0 and Datan Framework GenEx Pro package version 5.2.5.20 software for real-time PCR expression profiling. Grubbs' test was used to define outlier data points (1.1%) that were replaced with the median value of all samples for the gene in question. The strengths of the associations between continuous variables were tested with the nonparametric Spearman rank correlation test (r_s). Gene expression levels in the various fractions were compared with the nonparametric Wilcoxon's test to test the null hypothesis and the Mann-Whitney U test to identify genes with significantly different expression levels between groups. A false discovery rate (FDR) control of 10% was applied to correct for multiple testing (32). Cluster analysis (<http://rana.lbl.gov/eisen/>; ref. 33) was used to cluster the samples on the basis of the gene expression values and TreeView (<http://rana.lbl.gov/eisen/>; ref. 33) was used to visualize the results. DAVID (Database for Annotation, Visualization, and Integrated Discovery, david.abcc.ncifcrf.gov; refs. 34, 35) was used to functionally annotate genes and identify the over-represented functions, with P values corrected for multiple testing via the Benjamini-Hochberg's procedure. All human genes were used to compare frequencies of functions. Unless stated otherwise, all statistical tests were 2-sided with $P < 0.05$ considered as statistically significant.

Results

Quality control measures taken to ensure reliable measurement of CTC-specific gene transcripts

The first purpose of this study was to establish a sensitive method to perform both mRNA and miRNA expression analysis of transcripts specific for CTCs, in samples often containing only a few CTCs in an environment of excess quantities of contaminating leukocytes.

To select the gene transcripts, we used the approach described in detail in the Materials and Methods section, resulting in 43 putative breast CTC-specific miRNAs, 85 putative breast CTC-specific mRNAs, and 5 control mRNAs. Our first challenge was to find a method that would enable us to measure both mRNAs and miRNAs in RNA isolated from as little as 5 CTCs (approximately 50 pg total RNA), which is considered the clinically relevant cut point in patients with metastatic breast cancer (24–26) in a reliable and quantitative manner. In this respect, as already described and tested for the mRNA assays (18), any individual miRNA expression assay showing as a nonhomogeneously amplified outlier in our tests should be treated with caution because the data may not be truly representative for the original sample. Therefore, our assay had to have a high sensitivity combined with a minimum number of nonhomogeneously amplified miRNA and mRNA assays. To achieve this, we combined the already sensitive multiplex stem-loop cDNA approach with the TaqMan-based linear preamplification method, both from ABI. To validate the sensitivity and linear and homogeneous nature of this combined technique, we performed comparative

tests between serially diluted nonamplified and multiplexed preamplified cDNA from total RNA of pooled primary breast tumors, as described before (18). The homogeneity of amplification was set at a cutoff of 2 C_t , i.e., for an assay to be considered homogeneously amplified, the number of cycles that were required after preamplification should be within a 2 C_t range of the number of cycles that were required for the nonamplified material. After adjusting for the median 15.5 C_t gain due to the preamplification procedure, data of 11 miRNA assays were outside this range (Table 2, lower). After testing the 43 miRNAs in a multiplex cDNA PCR reaction in our patient cohort of 50 CTC samples, data of 2 additional miRNAs (*hsa-miR-10b* and *RNU43*) had to be discarded because they generated very poor amplification curves. Finally, the PCR efficiency of 2 of the remaining 30 assays was outside our set range of 75% to 125% (*hsa-miR-135b*, 135%, and *hsa-miR-452#*, 73%) and these miRNAs were therefore also excluded from our final analyses (Table 2, column 6).

A summary of the results of these quality control experiments, which left us with 28 potentially breast CTC-specific miRNAs that could be measured reliably after our multiplexed cDNA followed by the preamplification procedure, is listed in Table 2.

Finally, when implementing an assay into clinical diagnostics, it is important that data can be compared in-between qRT-PCR sessions. For our mRNA measurements, we have previously shown that the data are reproducible using the preamplification procedure from ABI (18). To certify that the miRNA data generated with these assays and the multiplex preamplification procedure were also reproducible between different qRT-PCR sessions, a control RNA sample consisting of 300 pg total RNA of a pool of breast tumors was included in each session. The relative expressions (average delta $C_t \pm 95\%$ CI) of the 28 miRNAs measured in this control sample in 28 independently performed multiplexed preamplified qRT-PCR sessions (Fig. 1) with a median coefficient of variation (CV) at the absolute C_t level of 6%, ranging from 3% for *hsa-miR-200a#* to 15% for *hsa-miR-184*, illustrate the robustness of our method.

mRNAs and miRNAs differentially expressed between CTC preparations and leukocytes

The miRNA analyses showed that of the 446 miRNAs investigated, 28 miRNA transcripts could be measured reliably and linearly in a multiplex preamplification reaction with an anticipated more than 10-fold (median 160-fold) higher expression in CTCs relative to blood-derived leukocytes (Table 2 and Supplementary Table S2A). Of these 28 small RNAs, only 1 miRNA (*hsa-miR-183*) was higher expressed in the 32 samples that contained at least 5 CTCs than the 9 samples without detectable CTCs after the CellSearch procedure. At an FDR of 10%, 9 additional miRNA transcripts were more abundantly expressed in the preparations that contained at least 5 CTCs relative to WB preparations of HBDs prior to ($n = 14$) or after ($n = 8$) CellSearch enrichment (Table 3A).

For the mRNA transcripts, we used the approach described in detail before (18). Of the thus *in silico* selected 85 putatively CTC-specific and/or for breast cancer clinically relevant genes (Supplementary Table S2B), 55 were at an FDR of 10% significantly higher expressed in the 32 samples of patients with at least 5 CTCs than 31 CellSearch-enriched HBD samples. A gene expression call rate of 55 of 85 (65%) is within the limits of what can be expected for a profiling study (36). In addition to these 55 mRNA transcripts, another 6 mRNA transcripts were more abundantly present in the 32 samples that contained at least 5 CTCs relative to the 14 WB samples from HBDs prior to CellSearch enrichment. Of the 55 mRNA transcripts, 14 were also more abundantly expressed in the 32 samples with at least 5 CTCs relative to the 9 enriched metastatic breast cancer blood samples without detectable CTCs. Finally, only 6 genes, including the 2 leukocyte control genes *PTPRC* (CD45) and *BST1*, were found to be significantly higher expressed in the 31 CellSearch-enriched HBD samples than the 32 patient samples with at least 5 CTCs (Table 3B).

Unsupervised hierarchical clustering to identify clusters of patients according to gene expression patterns

Next, unsupervised 2-dimensional average linkage hierarchical cluster analysis (33) was done to compare the gene expression profiles of our 50 patients. For this, we used the 65 genes (55 mRNA and 10 miRNA transcripts) that were at a 10% FDR more abundantly expressed in CellSearch-enriched fractions of the 32 patients with at least 5 CTCs (Table 3 and Fig. 2).

This analysis resulted in a clustering of 4 groups of patients with a clear discrimination between **patient cluster 1** and **patient clusters 2 to 4**. The median number of counted CTCs for cluster 1 was 1 (range: 0–173) CTC; for cluster 2, 14 (0–138) CTCs; for cluster 3, 41 (0–2, 262) CTCs; and for cluster 4, 74 (0–886) CTCs (Fig. 2).

About the gene clustering, 5 gene clusters with a correlation more than 0.2 could be identified. In the largest 18-gene cluster (**gene cluster 1**), "signaling" was the most significant common category for 12 genes (*MUCL1*, *FGFR4*, *FGFR3*, *ERBB4*, *CXCL14*, *PLOD2*, *PIP*, *TFF3*, *FKBP10*, *IGFBP2*, *TIMP3*, and *PLAU*) as identified by DAVID (34, 35) analysis (3.9-fold enriched, $P = 0.0014$). In addition to these signaling genes, this gene cluster contains some potentially interesting drug targets such as *ERBB4*, *FGFR3*, and *FGFR4*.

The second-gene cluster (**gene cluster 2**, correlation 0.40) is characterized by luminal genes, such as *CCND1* (37), *ESR1*, *KRT18* (37), and *MUC1*, of which *MUC1* has previously been used by others for the detection of CTCs in breast cancer (38–42). At an enrichment of 8.0-fold, Benjamini $P = 0.008$, "mutagenesis site," i.e., genes with mutational hot spots, was the most significant category identified by DAVID for 6 genes (*MUC1*, *CCND1*, *KRT18*, *ESR1*, *CEP55*, and *FEN1*) in this 7-gene cluster.

Table 2. Quality control qRT-PCR miRNA before and after multiplexed preamplification

miRNA assay	Serially diluted breast tumor pool 1 (Median C_t in uniplex cDNA reaction without preamplification)	Serially diluted breast tumor pool 2 (Median C_t in multiplex cDNA reaction with preamplification)	C_t gain due to multiplexed preamplification (Relative to unamplified uniplex cDNA reaction)	PCR efficiency in uniplex cDNA reaction without preamplification	PCR efficiency in multiplex cDNA reaction with preamplification	Reason for exclusion
(Specific for epithelial breast tumor cells)						
hsa-miR-100	28.67	11.71	16.96	102%	98%	
hsa-miR-10a	31.06	13.86	17.20	98%	108%	
hsa-miR-125b	28.33	12.65	15.68	95%	100%	
hsa-miR-135a	34.89	17.60	17.29	115%	77%	
hsa-miR-141	29.07	12.36	16.71	97%	105%	
hsa-miR-183	32.99	17.57	15.42	92%	120%	
hsa-miR-184	35.66	20.21	15.45	88%	123%	
hsa-miR-187	31.77	17.00	14.77	105%	93%	
hsa-miR-193b	34.05	19.14	14.91	99%	107%	
hsa-miR-200a#	32.73	18.09	14.64	96%	109%	
hsa-miR-200b	27.96	12.65	15.31	97%	89%	
hsa-miR-200c	24.32	10.32	14.00	96%	75%	
hsa-miR-205	29.98	14.15	15.83	94%	93%	
hsa-miR-210	30.15	13.18	16.97	86%	95%	
hsa-miR-214	28.73	12.93	15.80	101%	79%	
hsa-miR-31	29.87	12.40	17.47	101%	76%	
hsa-miR-34a	33.45	17.96	15.49	104%	103%	
hsa-miR-375	30.47	13.24	17.23	90%	75%	
hsa-miR-379	35.42	19.97	15.45	98%	103%	
hsa-miR-424	35.76	19.52	16.24	112%	73%	
hsa-miR-452	35.85	20.67	15.18	108%	94%	
hsa-miR-455	31.00	16.35	14.65	100%	97%	
hsa-miR-497	33.98	17.20	16.78	98%	79%	
hsa-miR-565	30.54	13.98	16.56	98%	123%	
hsa-miR-615	32.47	17.04	15.43	100%	110%	
hsa-miR-9#	33.00	18.59	14.41	104%	117%	
hsa-miR-95	34.47	19.41	15.06	82%	116%	
RNU19	31.78	15.81	15.97	108%	88%	

(Continued on the following page)

Table 2. Quality control qRT-PCR miRNA before and after multiplexed preamplification (Cont'd)

miRNA assay	Serially diluted breast tumor pool 1	Serially diluted breast tumor pool 2	C _t gain due to multiplexed preamplification	PCR efficiency in uniplex cDNA reaction without preamplification	PCR efficiency in multiplex cDNA reaction with preamplification	Reason for exclusion
RNU43	31.02	13.63	17.39	83%	117%	Poor amplification curves
hsa-miR-10b	31.71	16.79	14.92	98%	103%	Poor amplification curves
hsa-miR-135b	28.86	14.16	14.70	99%	132%	Poor PCR efficiency
hsa-miR-452#	31.92	17.98	13.94	105%	73%	Poor PCR efficiency
hsa-miR-149	29.82	10.35	19.47	111%	72%	Not homogeneously amplified
hsa-let-7e	33.61	14.30	19.31	99%	84%	Not homogeneously amplified
hsa-let-7i	33.14	14.77	18.37	88%	116%	Not homogeneously amplified
hsa-miR-196a	30.33	17.06	13.27	97%	81%	Not homogeneously amplified
hsa-miR-199a#	28.18	19.21	8.97	99%	123%	Not homogeneously amplified
hsa-miR-200a	29.62	17.18	12.44	103%	88%	Not homogeneously amplified
hsa-miR-203	30.13	11.79	18.34	97%	91%	Not homogeneously amplified
hsa-miR-429	34.65	25.65	9.00	98%	111%	Not homogeneously amplified
hsa-miR-511	32.01	18.77	13.24	76%	94%	Not homogeneously amplified
hsa-miR-99a	34.94	13.13	21.81	89%	104%	Not homogeneously amplified
snoR-12 (RNU6B)	33.91	16.20	17.71	93%	94%	Not homogeneously amplified

NOTE: RNA of a pool of primary tumors of patients who had not received neoadjuvant systemic therapy was diluted to 100, 25, and 6.25 ng (pool 1) and to 300, 75, and 18 pg (pool 2), respectively. Forty-three serially diluted aliquots of the 3 dilutions of pool 1 were used in 3 × 43 individual RT reactions and after a 20-fold dilution, measured at 25, 6.25, and 1.56 ng in 43 individual miRNA-specific PCRs. For pool 2, only 1 sample of each dilution was used to measure all 43 miRNAs in a multiplex cDNA reaction, resulting, after an additional preamplification and dilution step, in a final input of 6, 1.25, and 0.38 pg for each individual miRNA-specific PCR reaction. For all miRNAs, a minus RT sample at the second dilution was included that proved to be negative for all assays. The median reduction in the number of PCR cycles that were required to generate a signal after preamplification was 15.5. The homogeneity of amplification was therefore set at 15.5 ± 2 C_t. All assays that gained less than 13.5 or more than 17.5 C_t after this multiplexed preamplification procedure were considered as not homogeneously amplified. PCR efficiencies were calculated from the data, thus generated with the 3 serially diluted samples with a PCR efficiency between 75% and 125% considered acceptable.

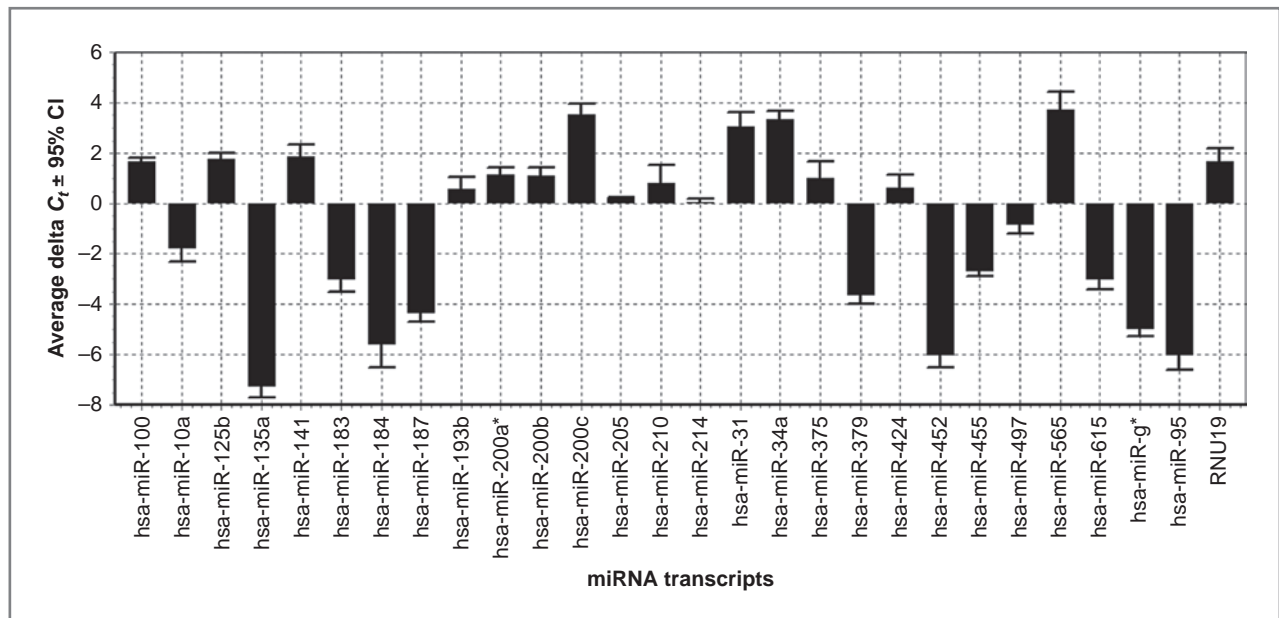


Figure 1. Reproducibility of the measurement of 28 miRNAs in a multiplex cDNA preamp reaction. Bars depict miRNA expression of 28 individual miRNA assays relative to the median level of all miRNA assays (average delta $C_t \pm 95\%$ CI) as measured in a control sample at 300 pg total RNA input in 28 independently performed multiplexed preamplified qRT-PCR sessions.

One distinct gene cluster (**gene cluster 3**, correlation 0.35) was responsible for the association with the absence of CTCs, i.e., patient cluster 1. This 14-gene cluster holds in addition to the previously identified CTC-specific genes *KRT19*, *AGR2*, *S100A16*, and *KRT7*, and as could be expected *TACSTD1*, the gene encoding EpCAM, the antigen that was used to enrich for CTCs, also the miRNAs *hsa-miR-452* and *hsa-miR-34a*.

Notably, the miRNA-cluster (**gene cluster 4**, correlation 0.20) containing *hsa-miR-183*, *hsa-miR-184*, *hsa-miR-379*, and *hsa-miR-424* shows an expression pattern that seems to be inversely related to the "mutagenesis" gene cluster 2—which includes *ESR1*-, the gene that encodes for the ER. This suggests that these miRNAs might be negatively regulated by ER or, *vice versa*, that these miRNAs negatively regulate ER.

Although no specific category was identified by DAVID as significantly enriched in the last cluster (**gene cluster 5**, correlation 0.20), this cluster seems to be dominated by genes associated with cell-cycle progression and proliferation such as *DUSP4* (*MKP2*; ref. 43), *KIF11*, *KPNA2*, and *MKI67*. Interestingly, a putative stem cell marker (*ITGA6*; ref. 44) is also included in this last cluster.

To ascertain that the signals we generated were indeed tumor CTC specific, we also performed a clustering analysis with inclusion of the 14 full blood HBDs (FB-HBD) from which we had data from both the mRNAs and miRNAs (Supplementary Fig. S1). These HBDs (marked in green below the cluster) indeed clustered closely together. Also, the patients from patient cluster 1 (Fig. 2, and marked in red below the cluster diagram in Supplementary Fig. S1), which were characterized by the lack of expression of

epithelial marker genes, remain clustered together, next to the HBD cluster.

To further validate that our identified 65-gene expression profile is able to clearly discriminate between signals derived from leukocytes that remain after CellSearch enrichment and signals derived from epithelial cells, we performed a proof-of-principle spiking experiment. For this, gene expression profiles of cells from 4 different breast cancer cell lines were compared with those of HBD samples of 5 different healthy volunteers and an HBD sample of a healthy volunteer in which the RNAs of the 4 different tumor cell lines were spiked in a final quantity equivalent to approximately 1 CTC (approximately 10 pg) per 1.5 mL blood. As can be appreciated from Supplementary Figure S2, a clear distinction can be seen between mixed and unmixed HBD and cell line samples. More importantly, no clear distinction can be seen between the final expression data by using RNA of the cell lines and the cell lines mixed with RNA from HBD.

These data point to a lack of contribution of the leukocytes to the overall gene expression results and confirm that our molecular CTC profile is indeed able to discriminate between signals from leukocytes and epithelial-specific signals from CTCs.

Associations of the CTC molecular profile with primary tumor characteristics

For the association of the molecular profile with primary tumor characteristics, we continued with the 36 patients in patient clusters 2 to 4. These patients displayed a molecular CTC profile with very distinct patterns from the 14 patients in patient cluster 1, which were characterized by the lack of

Table 3A. Evaluation of miRNA transcript levels in HBDs and CTC samples

Gene symbol	≥5 CTCs (n = 32)				HBD_CS (n = 8)				HBD_WB (n = 14)				0 CTCs (n = 9)			
	Normalized on the median level of all miRNAs	Average	SEM	Normalized on the median level of all miRNAs	Ratio ≥5 CTCs vs. HBD_CS	P value	10% FDR	Normalized on the median level of all miRNAs	Ratio ≥5 CTCs vs. HBD_WB	P value	10% FDR	Normalized on the median level of all miRNAs	Ratio ≥5 CTCs vs. 0 CTCs	P value	10% FDR	
																Average
hsa-miR-379	0.25	0.17	0.05	0.00	4.77	9.4E-05	*	0.05	0.00	4.77	1.3E-06	*	0.30	0.28	0.84	7.3E-01
hsa-miR-210	10.57	4.69	1.30	2.78	8.12	1.4E-04	*	1.32	0.73	7.99	4.3E-06	*	11.10	7.77	0.95	8.1E-01
hsa-miR-424	1.87	2.32	0.62	0.00	3.05	1.6E-04	*	0.62	0.00	3.05	3.1E-06	*	2.96	2.69	0.63	1.5E-01
hsa-miR-452	0.04	0.02	0.01	0.00	4.72	2.8E-03	*	0.01	0.00	3.06	3.0E-03	*	0.02	0.01	1.80	2.8E-01
hsa-miR-184	0.04	0.02	0.01	0.00	6.16	3.3E-03	*	0.01	0.00	5.76	2.6E-04	*	0.03	0.01	1.59	4.1E-01
hsa-miR-497	1.43	3.76	0.50	0.19	2.86	4.8E-03	*	0.45	0.03	3.15	9.0E-05	*	1.13	9.21	1.26	4.9E-01
hsa-miR-183	0.33	0.34	0.09	0.02	3.49	6.5E-03	*	0.10	0.00	3.30	2.7E-03	*	0.13	0.06	2.55	3.3E-02
hsa-miR-565	13.74	5.66	5.61	2.08	2.45	2.2E-02	*	4.41	0.12	3.12	1.8E-04	*	23.23	10.81	0.59	4.4E-01
RNU19	2.11	2.02	1.30	0.28	1.62	2.6E-02	*	1.27	0.00	1.66	1.6E-03	*	4.24	2.86	0.50	5.3E-02
hsa-miR-34a	8.85	2.14	6.52	0.00	1.36	8.2E-02	*	6.52	0.00	1.36	3.1E-02	*	10.85	3.48	0.82	3.4E-01
hsa-miR-125b	2.96	0.46	3.12	0.44	0.95	2.2E-01	*	3.12	0.15	0.95	2.9E-02	*	2.83	0.00	1.04	7.9E-01
hsa-miR-214	1.00	149.28	0.96	0.26	1.04	3.7E-01	*	1.14	0.82	0.88	1.6E-02	*	0.96	0.00	1.04	9.0E-01
hsa-miR-95	0.01	0.00	0.01	0.00	0.96	4.3E-01	*	0.01	0.00	0.58	3.0E-02	*	0.01	0.00	0.79	9.4E-01
hsa-miR-31	3.12	0.22	4.34	0.87	0.72	2.0E-03	*	3.69	0.20	0.85	1.3E-02	*	3.07	0.00	1.01	9.0E-01
hsa-miR-205	1.11	0.00	1.17	0.36	0.95	7.2E-03	*	1.20	0.26	0.93	1.6E-02	*	1.11	0.00	1.00	9.0E-01
hsa-miR-200c	7.79	1.83	10.60	2.58	0.74	2.2E-02	*	12.27	1.20	0.64	2.4E-02	*	8.24	1.06	0.95	2.6E-01
hsa-miR-200b	1.78	1.32	1.81	0.60	0.98	8.2E-02	*	2.17	0.33	0.82	3.1E-02	*	1.81	1.14	0.98	8.6E-01
hsa-miR-455	0.15	0.02	0.17	0.02	0.85	9.4E-02	*	0.18	0.02	0.80	3.6E-02	*	0.14	0.00	1.07	6.8E-01
hsa-miR-187	0.05	0.01	0.05	0.00	1.13	1.6E-01	*	0.06	0.00	0.88	2.8E-01	*	0.05	0.01	0.93	8.4E-01
hsa-miR-193b	1.24	3.99	1.03	3.50	1.21	1.8E-01	*	1.04	1.60	1.19	9.0E-01	*	1.41	3.20	0.88	5.5E-01
hsa-miR-141	2.62	1.22	2.57	0.30	1.02	2.2E-01	*	2.78	0.16	0.94	5.6E-02	*	2.47	0.32	1.06	9.9E-01
hsa-miR-100	3.11	0.84	3.30	0.59	0.94	2.3E-01	*	2.97	0.06	1.04	1.7E-01	*	2.81	0.00	1.10	3.3E-01
hsa-miR-200a*	2.05	1.01	1.73	0.00	1.19	2.9E-01	*	1.98	0.27	1.04	3.0E-01	*	1.76	0.17	1.17	2.3E-01
hsa-miR-375	1.66	3.52	1.32	1.62	1.25	2.9E-01	*	1.25	0.46	1.33	4.2E-01	*	1.31	2.40	1.26	6.0E-01
hsa-miR-615	0.12	0.05	0.09	0.01	1.29	4.9E-01	*	0.09	0.00	1.31	6.5E-01	*	0.12	0.07	0.99	4.4E-01
hsa-miR-10a	0.28	0.21	0.16	0.01	1.73	5.3E-01	*	0.20	0.02	1.41	4.8E-01	*	0.44	0.39	0.65	3.9E-02
hsa-miR-135a	0.01	0.00	0.01	0.00	0.92	7.2E-01	*	0.01	0.00	0.84	8.5E-01	*	0.01	0.00	1.00	5.2E-01
hsa-miR-9*	0.03	0.01	0.03	0.00	1.14	8.5E-01	*	0.04	0.00	0.93	8.3E-01	*	0.04	0.01	0.93	7.6E-01

Table 3B. Evaluation of mRNA transcript levels in HBDs and CTC samples

Gene symbol	≥5 CTCs (n = 32)			HBD_CS (n = 31)			HBD_WB (n = 14)			0 CTCs (n = 9)			
	Reference gene set normalized (×10 ³)	Ratio ≥5 CTCs vs. HBD_CS	P value	Reference gene set normalized (×10 ³)	Ratio ≥5 CTCs vs. HBD_CS	P value	Reference gene set normalized (×10 ³)	Ratio ≥5 CTCs vs. HBD_WB	P value	Reference gene set normalized (×10 ³)	Ratio ≥5 CTCs vs. 0 CTCs	P value	
	Average SEM	Average SEM	10% FDR	Average SEM	Average SEM	10% FDR	Average SEM	Average SEM	10% FDR	Average SEM	Average SEM	10% FDR	
KRT19	12.01	10.03	0.01	0.00	1113.54	1.0E-08	0.03	0.00	428.90	9.9E-07	318.06	1.9E-04	*
S100A16	3.90	2.70	0.01	0.00	293.30	1.0E-08	0.03	0.00	122.38	3.4E-05	125.09	3.5E-04	*
AGR2	0.52	0.31	0.00	0.00	135.71	3.3E-07	0.01	0.00	98.37	3.0E-04	91.98	1.9E-03	*
IGFBP5	0.26	0.11	0.00	0.00	66.98	1.0E-07	0.01	0.00	25.89	1.3E-03	42.10	2.4E-03	*
KRT7	12.17	8.39	0.51	0.00	24.08	1.0E-08	0.84	0.02	14.49	4.9E-04	15.19	2.5E-03	*
CLDN3	4.67	2.88	0.18	0.01	25.44	1.2E-07	0.24	0.00	19.23	7.3E-04	16.12	2.9E-03	*
FOXA1	0.58	0.31	0.00	0.00	152.45	1.0E-08	0.01	0.00	67.05	2.0E-04	63.80	2.9E-03	*
LAD1	0.13	0.05	0.00	0.00	33.58	1.0E-07	0.01	0.00	12.13	8.3E-03	15.18	5.3E-03	*
PKP3	0.11	0.04	0.00	0.00	29.90	9.9E-07	0.00	0.00	29.90	1.3E-04	18.29	5.3E-03	*
TFP3	9.36	7.67	0.27	0.00	35.26	2.9E-05	0.37	0.01	24.97	6.5E-03	24.74	5.9E-03	*
SCGB2A2	0.33	0.21	0.00	0.00	86.02	5.3E-05	0.00	0.00	86.02	1.6E-03	86.02	7.4E-03	*
SPDEF	1.14	0.57	0.05	0.00	22.88	2.9E-06	0.09	0.00	13.04	1.7E-02	13.13	1.6E-02	*
IGFBP2	13.75	8.70	1.31	0.00	10.52	1.0E-08	2.42	0.09	5.68	5.8E-02	7.33	9.4E-03	*
FKBP10	0.28	0.10	0.02	0.00	13.40	9.9E-07	0.04	0.00	7.11	2.3E-02	7.23	1.2E-02	*
SEPP1	3.69	1.76	0.27	0.00	13.52	1.0E-08	0.46	0.01	8.06	9.9E-04	5.23	1.8E-02	*
SCGB1D2	0.23	0.14	0.00	0.00	59.47	6.7E-04	0.00	0.00	59.47	7.8E-03	59.47	2.4E-02	*
CRABP2	7.29	4.83	0.38	0.00	19.16	1.0E-08	0.58	0.01	12.58	8.9E-03	6.64	5.3E-02	*
TNRC9	0.02	0.01	0.00	0.00	6.43	1.4E-03	0.00	0.00	6.43	1.3E-02	4.13	1.1E-01	*
SBEM/MUCL1	0.52	0.30	0.00	0.00	119.71	6.7E-04	0.01	0.00	35.30	1.0E-02	4.38	2.3E-01	*
PIP	0.07	0.03	0.00	0.00	17.68	3.0E-04	0.00	0.00	15.61	9.9E-03	2.15	6.1E-01	*
IL17BR3	0.13	0.02	0.08	0.00	1.69	1.0E-08	0.15	0.00	0.89	1.5E-01	1.68	3.3E-01	*
DTX3	3.77	0.71	2.65	0.00	1.42	1.0E-08	4.19	0.13	0.90	1.0E-01	0.82	7.5E-01	*
PLAU	0.45	0.09	0.30	0.01	1.52	1.5E-08	0.46	0.01	0.98	2.0E-01	1.08	9.4E-01	*
FGFR4	0.20	0.04	0.10	0.00	2.09	1.0E-08	0.17	0.00	1.17	4.7E-01	0.69	8.6E-01	*
ITGA6	54.40	25.94	31.74	1.07	1.71	2.1E-05	45.25	2.17	1.20	2.0E-01	1.49	7.2E-01	*
DTL	1.29	0.31	0.55	0.00	2.34	1.0E-08	1.00	0.04	1.29	5.9E-01	1.60	3.9E-01	*
MKI67	0.66	0.17	0.32	0.00	2.09	1.0E-08	0.43	0.01	1.54	2.1E-01	1.84	6.9E-01	*
PTRF	1.94	0.49	0.83	0.00	2.35	1.0E-08	1.23	0.03	1.58	6.1E-01	1.91	3.4E-01	*
CEP55	0.79	0.20	0.28	0.00	2.76	1.0E-08	0.48	0.01	1.64	9.4E-01	1.21	4.9E-01	*
MELK	0.67	0.16	0.21	0.00	3.23	1.3E-04	0.40	0.01	1.66	5.6E-01	1.30	2.5E-01	*
SMA/ACTA1	0.01	0.00	0.00	0.00	3.91	1.9E-02	0.01	0.00	1.68	9.0E-01	3.38	4.5E-01	*
FEN1	4.91	1.48	2.12	0.10	2.31	1.0E-08	2.88	0.07	1.70	9.2E-01	1.31	4.7E-01	*

(Continued on the following page)

Table 3B. Evaluation of mRNA transcript levels in HBDs and CTC samples (Cont'd)

Gene symbol	≥5 CTCs (n = 32)			HBD_CS (n = 31)			HBD_WB (n = 14)			0 CTCs (n = 9)				
	Reference gene set normalized (×10 ³)	Ratio >5 CTCs vs. HBD_CS	P value	Reference gene set normalized (×10 ³)	Ratio >5 CTCs vs. HBD_CS	P value	Reference gene set normalized (×10 ³)	Ratio >5 CTCs vs. HBD_WB	P value	Reference gene set normalized (×10 ³)	Ratio >5 CTCs vs. 0 CTCs	P value		
	Average SEM	Average SEM		Average SEM	Average SEM		Average SEM	Average SEM		Average SEM	Average SEM			
KIF11	3.67	1.41	1.13	0.00	3.25	6.7E-04	1.77	0.05	2.07	6.9E-01	2.70	0.85	1.36	3.2E-01
KPNA2	18.19	8.71	4.80	0.03	3.79	3.0E-05	7.57	0.28	2.40	2.4E-01	9.39	2.81	1.94	1.3E-01
CD24	3.60	1.52	1.07	0.03	3.36	3.8E-02	1.47	0.04	2.44	9.5E-01	2.09	0.51	1.72	4.5E-01
TM4SF13	13.88	8.34	2.29	0.00	6.06	5.3E-05	5.00	0.28	2.78	3.8E-01	8.01	4.20	1.73	1.9E-01
DUSP4	4.41	2.01	1.11	0.00	3.98	2.8E-02	1.56	0.03	2.84	8.8E-01	1.11	0.00	3.99	1.2E-01
ESR1	3.62	1.78	0.71	0.00	5.06	1.4E-03	1.15	0.03	3.15	6.7E-01	1.84	0.58	1.97	3.3E-01
IGFBP4	37.09	28.32	7.55	0.26	4.92	1.7E-03	11.41	0.46	3.25	4.1E-01	20.40	11.52	1.82	3.1E-01
MUC1 (EMA)	6.66	3.49	1.25	0.00	5.31	6.7E-04	1.92	0.05	3.46	4.7E-01	2.46	0.55	2.71	3.9E-01
CITN1/EM51	16.48	10.46	3.49	0.21	4.73	4.7E-04	4.48	0.13	3.68	2.9E-01	5.61	1.24	2.94	1.2E-01
KRT17	0.02	0.00	0.00	0.00	4.66	1.1E-02	0.00	0.00	3.71	2.0E-01	0.01	0.00	3.20	3.4E-01
PLOD2	4.57	2.32	0.62	0.00	7.32	1.2E-02	1.19	0.05	3.85	1.3E-01	1.24	0.27	3.68	6.5E-02
FGFR3	0.05	0.01	0.01	0.00	8.63	1.0E-08	0.01	0.00	4.02	2.7E-01	0.01	0.00	5.69	1.0E-01
CXCL14	0.02	0.01	0.00	0.00	6.16	5.7E-03	0.00	0.00	4.97	1.6E-01	0.00	0.00	5.00	1.7E-01
TIMP3	5.01	2.58	0.61	0.00	8.21	1.0E-08	1.00	0.03	5.03	1.6E-01	0.71	0.05	7.02	3.3E-02
TFPI (pS2)	0.02	0.01	0.00	0.00	5.81	1.9E-02	0.00	0.00	5.81	6.8E-02	0.00	0.00	5.71	2.4E-01
TACSTD1	2.23	1.04	0.17	0.00	12.98	7.9E-06	0.38	0.02	5.83	7.7E-02	0.66	0.20	3.37	1.9E-01
EEF1A2	1.32	0.57	0.11	0.00	12.44	2.9E-03	0.22	0.01	5.98	4.2E-02	0.46	0.18	2.86	8.9E-02
ERBB3	0.40	0.15	0.04	0.00	11.17	7.9E-06	0.06	0.00	6.30	7.0E-02	0.09	0.02	4.38	6.3E-02
KRT18	31.07	25.42	3.73	0.45	8.33	4.7E-05	4.85	0.14	6.40	7.7E-02	8.56	3.61	3.63	1.0E-01
ERBB4	0.03	0.01	0.00	0.00	7.87	2.9E-03	0.00	0.00	7.26	3.9E-02	0.00	0.00	7.87	4.9E-02
S100A7	0.03	0.01	0.00	0.00	7.68	3.4E-02	0.00	0.00	7.68	9.7E-02	0.00	0.00	6.74	3.1E-01
CCND1	2.35	1.23	0.18	0.00	13.35	2.1E-05	0.29	0.01	8.09	5.2E-02	0.48	0.13	4.86	1.0E-01
CEA(CAM-5)	0.10	0.05	0.00	0.00	26.83	3.0E-04	0.01	0.00	14.65	5.0E-02	0.01	0.00	12.35	1.2E-01
BST1	464.20	927.07	878.90	4353.26	0.53	3.5E-04	291.33	34.20	1.59	9.8E-01	211.85	20.19	2.19	1.6E-01
CD29	869.70	13784.47	2268.13	1953.69	0.38	8.7E-05	360.54	76.79	2.41	3.9E-01	415.78	676.83	2.09	1.5E-01
CD44	1552.60	5355.03	7748.54	1165.71	0.20	3.5E-06	808.44	496.02	1.92	4.6E-01	864.87	5661.39	1.80	1.9E-01
NOTCH3	1.48	0.55	1.70	0.21	0.87	9.9E-07	0.30	0.00	4.91	9.0E-02	0.29	0.02	5.13	8.3E-02
PSMD10	4.02	1.08	6.86	0.84	0.59	2.5E-04	2.59	0.07	1.55	7.7E-01	2.97	0.64	1.35	3.2E-01
PTPRC (CD45)	1968.20	4821.11	21728.33	4262.51	0.09	2.1E-07	1438.22	803.53	1.37	7.1E-01	1182.81	6086.11	1.66	7.9E-01
IGFBP3	1.34	0.09	1.24	0.00	1.08	8.4E-01	2.26	0.09	0.59	1.5E-02	1.24	0.00	1.08	9.0E-01

(Continued on the following page)

Table 3B. Evaluation of mRNA transcript levels in HBDs and CTC samples (Cont'd)

Gene symbol	≥5 CTCs (n = 32)				HBD_CS (n = 31)				HBD_WB (n = 14)				0 CTCs (n = 9)			
	Reference gene set normalized (×10 ³)	Average SEM	Ratio ≥5 CTCs vs. HBD_CS	P value	10% FDR	Reference gene set normalized (×10 ³)	Average SEM	Ratio ≥5 CTCs vs. HBD_WB	P value	10% FDR	Reference gene set normalized (×10 ³)	Average SEM	Ratio ≥5 CTCs vs. CTCs	P value	10% FDR	
<i>Nm23-H1/NMIE1</i>	45.63	25.49	10.10	0.00	6.4E-02	18.06	0.95	2.53	1.4E-01	24.59	12.02	1.86	3.1E-02			
<i>CCNB1</i>	1.69	0.60	0.40	0.00	2.0E-01	0.71	0.02	2.36	4.0E-01	0.93	0.23	1.81	1.8E-01			
<i>CCNE1</i>	1.82	0.41	1.17	0.00	9.0E-02	1.96	0.06	0.93	1.9E-01	2.26	0.49	0.80	9.9E-01			
<i>CCNE2</i>	2.41	0.74	0.87	0.00	9.0E-02	1.65	0.06	1.46	8.1E-01	1.15	0.14	2.11	2.5E-01			
<i>CD133</i>	0.00	0.00	0.00	0.00	6.7E-01	0.01	0.00	0.72	4.2E-01	0.00	0.00	1.13	7.9E-01			
<i>CDH1</i>	7.47	4.17	1.27	0.01	9.9E-01	1.96	0.05	3.80	3.7E-01	3.32	1.27	2.25	2.5E-01			
<i>COL1A1</i>	0.01	0.00	0.00	0.00	4.0E-01	0.00	0.00	1.50	5.1E-01	0.00	0.00	1.48	9.1E-01			
<i>COL2A1</i>	0.00	0.00	0.00	0.00	8.4E-01	0.00	0.00	1.03	8.5E-01	0.00	0.00	1.03	9.0E-01			
<i>EGFR</i>	0.01	0.00	0.00	0.00	2.9E-01	0.00	0.00	1.76	4.1E-01	0.01	0.00	1.08	8.6E-01			
<i>ERBB2</i>	5.79	3.05	1.24	0.00	2.0E-01	2.62	0.11	2.21	9.0E-01	2.09	0.38	2.77	3.8E-01			
<i>FGFR2</i>	0.14	0.02	0.09	0.00	2.0E-01	0.16	0.00	0.93	2.0E-01	0.09	0.00	1.67	4.0E-01			
<i>GALGT</i>	0.01	0.00	0.00	0.00	9.0E-02	0.00	0.00	2.23	1.9E-01	0.01	0.00	1.23	5.8E-01			
<i>GATA3</i>	166.72	136.63	155.43	72.22	2.7E-01	139.97	14.56	1.19	3.2E-01	119.49	52.73	1.40	9.5E-01			
<i>KRTHB1/KRT81</i>	0.01	0.00	0.00	0.00	2.9E-01	0.01	0.00	1.07	1.2E-01	0.01	0.00	0.98	8.9E-01			
<i>MAGEA3</i>	0.00	0.00	0.00	0.00	5.3E-01	0.00	0.00	1.26	6.2E-01	0.00	0.00	1.26	6.8E-01			
<i>MET</i>	0.04	0.00	0.03	0.00	2.9E-01	0.05	0.00	0.81	1.0E-01	0.05	0.01	0.88	9.1E-01			
<i>MSMB</i>	0.05	0.01	0.02	0.00	9.0E-02	0.04	0.00	1.23	3.1E-01	0.03	0.00	1.50	8.6E-01			
<i>MYL3</i>	0.00	0.00	0.00	0.00	9.9E-01	0.00	0.00	1.00	9.9E-01	0.00	0.00	1.00	9.9E-01			
<i>PLK1</i>	1.37	0.47	0.39	0.00	2.0E-01	0.64	0.02	2.13	9.5E-01	0.84	0.19	1.62	4.9E-01			
<i>PTPRK</i>	0.60	0.21	0.10	0.00	9.9E-01	0.19	0.01	3.15	4.2E-01	0.13	0.01	4.64	7.0E-02			
<i>SELE</i>	0.01	0.00	0.00	0.00	2.9E-01	0.00	0.00	1.92	4.1E-01	0.00	0.00	1.92	4.9E-01			
<i>STK6/AURKA</i>	0.01	0.00	0.00	0.00	1.4E-01	0.01	0.00	0.90	3.5E-01	0.01	0.00	1.04	6.7E-01			
<i>TOP2a</i>	5.17	1.99	1.80	0.00	6.7E-01	2.93	0.10	1.76	9.0E-01	3.36	0.74	1.54	4.1E-01			
<i>TWIST</i>	0.02	0.00	0.01	0.00	2.9E-01	0.02	0.00	0.79	1.0E-01	0.01	0.00	1.30	8.1E-01			
<i>VEGFR2</i>	0.01	0.00	0.00	0.00	1.4E-01	0.01	0.00	1.13	6.3E-01	0.01	0.00	1.31	6.5E-01			

NOTE: The expression levels of 28 miRNAs (Table 3A) and 87 mRNAs (Table 3B) were measured in RNA isolated with the CellSearch Profiling Kit, as described in the Materials and Methods section. To identify putative CTC-specific genes, transcript levels were compared between various groups with and without CTCs as established by the CellSearch Epithelial Kit in a separate 7.5-mL blood sample. The average expression ±SEM is given by the number of samples indicated at the top in the table. The 2-tailed Mann-Whitney U test was used to identify genes with a significantly different expression level between groups. To compensate for multiple testing, an FDR of 10% was applied on these statistics. *, significant at a 10% FDR (2-tailed $P < 0.05$). CS, enriched by EpCAM-based CellSearch.

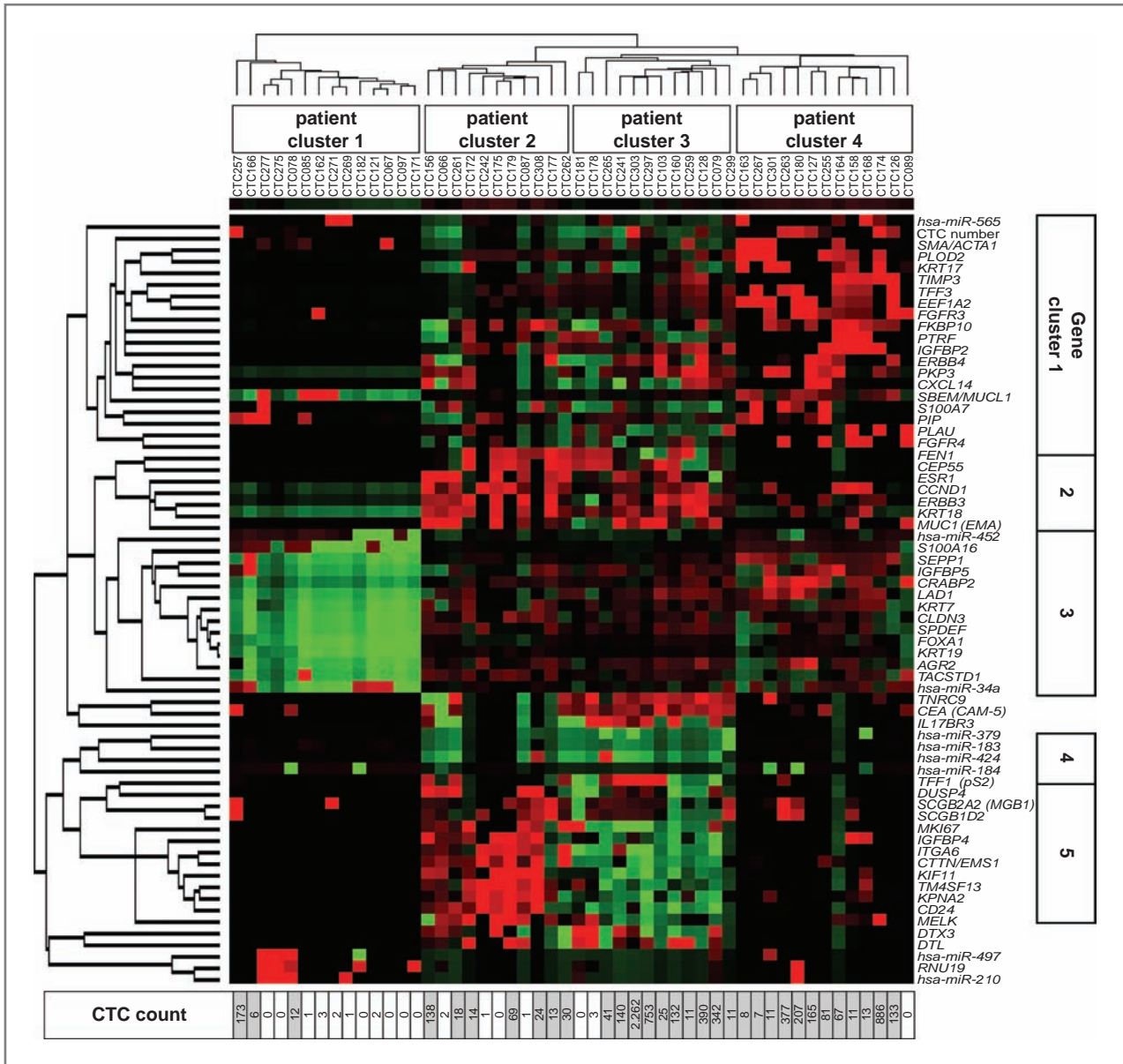


Figure 2. Unsupervised hierarchical cluster analysis comparing gene expression profiles in CTC-enriched blood samples of metastatic breast cancer patients. Expression levels were analyzed with real-time RT-PCR with 65 TaqMan Gene Expression Assays in cDNA generated from RNA isolated from the CTC-enriched fractions of 50 metastatic breast cancer patients. Sample loading and RNA integrity were controlled with 3 additional universal reference genes (*GUSB*, *HPRT1*, and *HMBS*). Prior to real-time PCR, cDNA was preamplified in 15 cycles with the PreAmp method from ABI, as described in the Materials and Methods section, by using the same TaqMan Gene Expression Assays that were used for the real-time PCR. Data shown have been subjected to median normalization of each individual gene across all samples followed by median normalization of each individual sample across all genes. Each horizontal row represents a gene, and each vertical column corresponds to a sample. Red color indicates a transcript level above the median level, black color indicates a median transcript level, and green color indicates a transcript level below the median level of the particular assay as measured in all samples. Depicted gene clusters were identified at an average linkage correlation greater than 0.2. The number of CTCs as established by the CellSearch Epithelial Kit is given below the graph, with the samples with at least 5 CTCs according these cell counts marked gray for easy identification.

expression of epithelial marker genes. Detailed clinicopathological information of our patient cohort, subdivided in 2 groups (patient cluster 1 versus clusters 2 to 4) on the basis of our molecular CTC-specific profile, is given in Table 1. There were no differences between both groups in terms of nodal status, tumor size, histological tumor type,

grade, ER, PR, and HER2 status. The only significant association with clinical information was that the patients of clusters 2 to 4 displayed a 2-fold higher rate of having both visceral and nonvisceral metastases, as opposed to only visceral or nonvisceral metastasis for the patients of cluster 1.

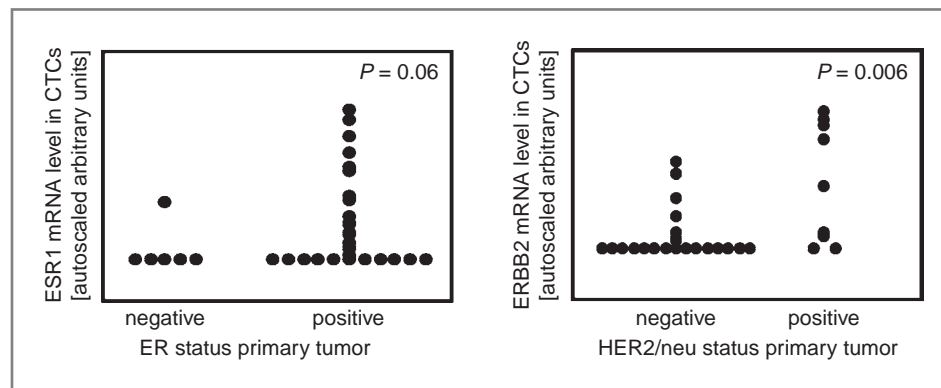


Figure 3. Association between gene expression in CTCs of metastatic breast cancer patients and expression of their corresponding protein in the primary tumor. Gene transcript levels were analyzed with real-time RT-PCR and normalized as described in the Materials and Methods section from CellSearch-enriched fractions of 36 breast cancer patients with molecularly identifiable CTCs. Gene expression levels were compared with expression of their corresponding protein in the primary tumor using the nonparametric Mann–Whitney U test to identify genes with significantly different expression levels between groups.

Almost identical results were obtained when the associations of primary tumor and patient characteristics were studied on the basis of CTC count subdivided in 2 groups (patients with less than 5 CTCs versus patients with at least 5 CTCs; Supplementary Table S1).

Associations of gene transcripts measured in CTCs with current drug targets

Although we could not measure *PGR* transcripts reliably in the CTCs due to the relatively high *PGR* levels present in the contaminating leukocytes, we could measure *ESR1* and *ERBB2* mRNA transcript levels, the genes for ER and HER2, respectively, in the CellSearch-enriched CTCs. *ESR1* and *ERBB2* expression levels measured in the 36 patients from clusters 2 to 4 with expression of epithelial marker genes and compared with ER and HER2 status of the primary tumor as assessed by routine pathological immunohistochemical procedures (with additional FISH for the HER2++ cases), respectively, are shown in Figure 3.

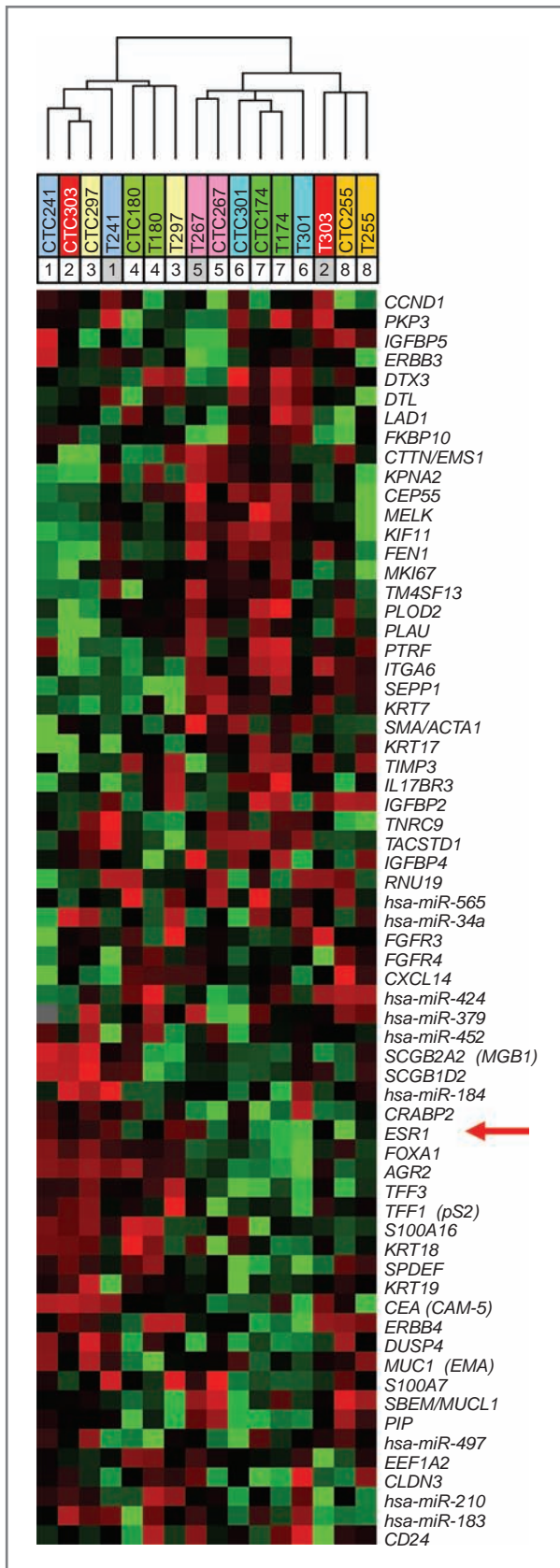
Comparison of gene profiles measured in the CTCs and corresponding primary tumors of metastatic breast cancer patients

We could retrieve 8 primary tumor tissues (3× FFPE and 5× FF) of our cohort of patients with at least 5 CTCs at the time of metastatic disease (median:174, range 7–2, 262 CTCs). We measured the 65 genes of our mRNA and miRNA panel in these tissues after adjusting levels measured in FFPE to those measured in FF.

From the unsupervised average linkage correlation clustering (Fig. 4), it became clear that most CTC samples clustered well with the corresponding primary tumor tissue (T) and that the clustering was not dependent on the origin of the primary tissue (FF or FFPE).

Discussion

In this study, we describe a robust method to simultaneously determine the expression of 65 epithelial tumor cell-specific miRNA and mRNA expression levels in CTCs enriched by CellSearch. The rationale of our study using the CellSearch technique as a starting point was to develop a simple PCR-based molecular characterization that can be performed on material obtained in a clinical setting. Because the CellSearch method is currently the only FDA-approved semi-automated method to capture CTCs, taking CellSearch-enriched CTCs as a starting point for our method will enable its implementation in clinical studies and broadens its application possibilities. However, although the EpCAM-based enrichment employed by the CellSearch technique eliminates a large proportion of leukocytes (approximately 4-log depletion), there are still considerable quantities of leukocytes present after this enrichment (18). This remaining leukocyte contamination, together with the low frequency of CTCs, forms a challenge when aiming to characterize CTCs by the expression of multiple genes. Despite these challenges, our data indicate that we have succeeded to measure true epithelial tumor cell-specific genes in CTCs with our CTC-specific 65-gene panel, and managed to avoid generation of a predominant leukocyte-derived signal. First, by only selecting genes highly expressed in breast cancer samples and not, or at a much lower level, in blood from HBDs. Second, by validating the true epithelial-specific expression with clustering analyses, which showed that based on the expression of the 65 genes of our molecular profile, the HBDs and breast cancer patients without detectable CTCs clustered closely together and could be clearly separated from the breast cancer patients with detectable CTC numbers (Fig. 2 and Supplementary Fig. S4). In addition, after using our 65-gene profile, most CTC samples clustered well with the corresponding primary tumor tissue (Fig. 4). Finally, as a



proof of principle, we showed that profiling with our 65-gene panel before and after spiking RNA of HBDs with RNA from 4 different cell lines in a final quantity equivalent to approximately 1 CTC per 1.5 mL blood clearly separated the mixed and unmixed cells (Supplementary Fig. S2). These data confirmed that our molecular 65-gene profile is indeed able to discriminate between signals from leukocytes and epithelial-specific signals from CTCs.

On the basis of the expression levels of this 65-gene profile, we could identify 4 different patient clusters characterized by 5 distinct gene clusters (Fig. 2). One distinct 14-gene cluster (gene cluster 3) was responsible for the association with the absence of CTCs. To further appreciate the strength of our 65-gene profile in relation to CTC count, it should be noted that the CTC counts were derived from 1 of the 2 aliquots of 7.5 mL blood samples that were processed with the CellSearch Epithelial Kit, whereas the other aliquot used for the molecular profiling was processed with the CellSearch Profiling Kit. This inevitably introduced stochastic variation between the tumor cell content in the 2 aliquots, which is more profound in the lower range of CTC counts. Discussion has also started about the actual number of isolated CTCs differing between the enumeration and profiling kit (45). The given cell counts could therefore only be used as a rough estimate for our molecular profile.

Nevertheless, with 14 of 55 mRNAs (25.4%) and only 1 (*hsa-miR-183*) of 28 miRNAs (3.6%) higher expressed in the 32 samples that contained at least 5 counted CTCs compared with the 9 samples without detectable CTCs after the CellSearch enrichment procedure with the Epithelial Kit, it seems to be easier to discriminate between CTC-specific and leukocyte-derived mRNAs than between CTC-specific and leukocyte-derived miRNAs. Possibly, the detected miRNA transcripts were derived from cell fragments present in the blood of cancer patients without detectable intact CTCs. The fact that we could measure them might be associated with the remarkable stability of miRNA transcripts in blood (46). Indeed, the detection of an additional 9 of 28 (32.1%) miRNAs that were higher expressed in breast cancer patients

Figure 4. Unsupervised hierarchical cluster analysis comparing gene expression profiles in CTC-enriched blood samples of metastatic breast cancer patients and their corresponding primary tumor. Expression levels were analyzed with real-time RT-PCR with 65 TaqMan Gene Expression Assays in cDNA generated from RNA isolated from the CellSearch-enriched fractions of 8 breast cancer patients with metastatic disease and their corresponding primary tumors. Sample loading and RNA integrity were controlled with 3 additional universal reference genes (*GUSB*, *HPRT1*, and *HMBS*). Prior to real-time PCR, cDNA was preamplified in 15 cycles with the PreAmp method from ABI, as described in the Materials and Methods section, by using the same TaqMan Gene Expression Assays that were used for the real-time PCR. Data shown have been subjected to median normalization of each individual gene across all samples followed by median normalization of each individual sample across all genes. Each horizontal row represents a gene, and each vertical column corresponds to a sample (T = tissue). The 3 FFPE primary tumor tissues from patients 1, 2, and 5 are gray colored on top in the graph for easy identification. Red color indicates a transcript level above the median level, black color indicates a median transcript level, and green color indicates a transcript level below the median level of the particular assay as measured in all samples.

without detectable CTCs than in WB preparations of HBDs prior to ($n = 14$) or after ($n = 8$) CellSearch enrichment, compared with an additional 6 of 55 (10.9%) for mRNAs, further supports this thought. For these reasons, we felt confident to continue our analyses with those samples that did contain CTCs according to our molecular profile (patient clusters 2 to 4 in Figure 2), irrespective of the CTC count in the blood sample that was processed in parallel with the CellSearch Epithelial Kit.

To show the potential clinical utility of measuring these 65 marker genes in CTCs, we had a further look in the data we generated with our molecular profiling on the levels of 2 well-known genes in breast cancer, ER and HER2 (Fig. 3). For 1 of the patients whose primary tumor was assessed to be ER negative, a clearly positive *ESR1* signal was detected in the CTCs (CTC087 in Figure 2) obtained at the time of metastatic disease 7 years after surgical removal of the primary tumor. However, and perhaps even more disturbing, in 11 of 30 patients (37%) whose primary tumor was ER-positive, no detectable *ESR1* transcript levels were measured in the CTCs obtained 1 to 149 months after primary surgery. Thus, although according to the primary tumor characteristics, these patients would have an indication for antihormonal treatment, no benefit might be expected from this therapy on the basis of these CTC characteristics. However, due to the limited number of 4 of these 11 patients that were actually treated with antihormonal treatment, no conclusion can be drawn yet on the efficacy of hormonal treatment in these patients with *ESR1*-negative CTCs and ER-positive primary tumors. Of note in this respect is that half the patients with relatively high CTC-associated *ESR1* levels expressed relatively low levels of *TFF1* (Fig. 2). *TFF1* is a gene under the control of ER. Perhaps, assessment of simultaneous *TFF1* expression in CTCs might be able to identify a subset of patients with ER-positive CTCs with functionally active ER, which is more likely to respond to hormonal treatment (47).

Similarly, the CTCs of at least 4 patients with HER2-negative primary tumors showed to be positive at the time of metastatic disease, whereas in 2 patients with an HER2-positive primary tumor, no detectable *ERBB2* mRNA could be measured in their CTCs. For those 4 patients with *ERBB2*-positive CTCs, anti-HER2 therapy is not indicated on the basis of primary tumor characteristics, whereas this treatment could nonetheless be beneficial based on their CTC characteristics.

No clinically relevant cut point has yet been established for ER and HER2 measured by qRT-PCR in CTCs. Nevertheless, such discrepancies between the levels of ER and HER2 measured in the primary tumor and metastases and CTCs have been described before at both the mRNA and the protein level (40, 45, 48–51), indicating that the findings with our multigene measuring technique may indeed be relevant, not only for ER and HER2 but also for the other markers included in our panel.

After clustering CTCs and primary tumors based on the expression of all 65 genes, the only obvious discrepancy we observed between the CTCs and the corresponding primary tumors of 8 different patients concerned patient 2. With

2,262 CTCs, this was the patient with the highest number of CTCs, and thus with an expected negligible effect of the presence of contaminating leukocytes in the expression analysis. The primary tumor of this patient was originally assessed as lobular, low-grade, pT2, ER-positive, PR-positive, and HER2-negative. Such a lobular tumor, with scattered epithelial cell clusters, and associated contaminating RNA from many stromal cells (52), may have contributed to this poor correlation with the expression profile of the high number of CTCs.

Although the high degree of homology in the gene expression profiles of CTCs and corresponding primary tumors was reassuring, discrepancies in expression of individual genes, such as for *ESR1* in patients 5, 6, and 8 (Fig. 4), were detected. Another example in this respect is patient 8, from whom the CTCs expressed much higher levels of markers associated with cell-cycle progression and proliferation such as *DTL*, *KIF11*, *KPNA2*, *KIF11*, and *MKI67* than the primary tumor (Fig. 4). Such differences between the primary tumor and CTCs isolated at the time of metastatic disease might prove clinically relevant and thus deserve further research.

In summary, by excluding genes with a relatively higher expression in leukocytes, our CTC-specific 65-gene set, consisting of 55 mRNAs and 10 miRNAs, is able to generate a huge amount of highly relevant CTC-specific data, even in the presence of a leukocyte background signal derived of leukocytes cocaptured with CTCs when using the CellSearch procedure.

Although assessed in a relatively small series, we found discrepancies in several important factors such as ER, HER2, and other genes between primary tumor tissue and CTCs. This is not surprising given the time elapsing between primary tumor resection and CTC collection, which occurred at the diagnosis of metastatic disease, and the fact that several patients received prior adjuvant systemic therapy. The discrepancies in molecular characteristics between primary tumor tissue and CTCs clearly stress the importance of further studies on molecular characterization of CTCs.

Disclosure of Potential Conflicts of Interest

The salaries and bench fee for this collaborative work was covered in part by Veridex. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We especially thank the patients for their willingness to participate and the surgeons, pathologists, and medical oncologists for their assistance in collecting samples and patient's clinical follow-up data.

Grant Support

This study was in part financially supported by the Netherlands Genomic Initiative (NGI)/Netherlands Organization for Scientific Research (NWO).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 1, 2011; revised March 13, 2011; accepted April 11, 2011; published OnlineFirst April 19, 2011.

References

- Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;6:857–66.
- Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;355:560–9.
- Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007;128:683–92.
- van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* 2008;452:564–70.
- Esteller M. Epigenetics in cancer. *N Engl J Med* 2008;358:1148–59.
- Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer* 2010;10:389–402.
- Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 2003;3:537–49.
- Mimori K, Kataoka A, Yoshinaga K, Ohta M, Sagara Y, Yoshikawa Y, et al. Identification of molecular markers for metastasis-related genes in primary breast cancer cells. *Clin Exp Metastasis* 2005;22:59–67.
- Suzuki M, Tarin D. Gene expression profiling of human lymph node metastases and matched primary breast carcinomas: clinical implications. *Mol Oncol* 2007;1:172–80.
- Park S, Holmes-Tisch AJ, Cho EY, Shim YM, Kim J, Kim HS, et al. Discordance of molecular biomarkers associated with epidermal growth factor receptor pathway between primary tumors and lymph node metastasis in non-small cell lung cancer. *J Thorac Oncol* 2009;4:809–15.
- Cavalli LR. Molecular markers of breast axillary lymph node metastasis. *Expert Rev Mol Diagn* 2009;9:441–54.
- Gevaert O, Daemen A, De Moor B, Libbrecht L. A taxonomy of epithelial human cancer and their metastases. *BMC Med Genomics* 2009;2:1–16.
- Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 2010;467:1109–13.
- Food and Drug Administration HHS. Medical devices; immunology and microbiology devices; classification of the immunomagnetic circulating cancer cell selection and enumeration system. Final rule. *Fed Regist* 2004;69:26036–8.
- Danila DC, Heller G, Gignac GA, Gonzalez-Espinoza R, Anand A, Tanaka E, et al. Circulating tumor cell number and prognosis in progressive castration-resistant prostate cancer. *Clin Cancer Res* 2007;13:7053–8.
- Cohen SJ, Alpaugh RK, Gross S, O'Hara SM, Smirnov DA, Terstappen LW, et al. Isolation and characterization of circulating tumor cells in patients with metastatic colorectal cancer. *Clin Colorectal Cancer* 2006;6:125–32.
- Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* 2004;351:781–91.
- 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.
- Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. *Nat Genet* 2005;37:495–500.
- Croce CM, Calin GA. miRNAs, cancer, and stem cell division. *Cell* 2005;122:6–7.
- Foekens JA, Sieuwerts AM, Smid M, Look MP, de Weerd V, Boersma AW, et al. Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. *Proc Natl Acad Sci U S A* 2008;105:13021–6.
- Bartels CL, Tsongalis GJ. MicroRNAs: novel biomarkers for human cancer. *Clin Chem* 2009;55:623–31.
- Tong AW, Nemunaitis J. Modulation of miRNA activity in human cancer: a new paradigm for cancer gene therapy? *Cancer Gene Ther* 2008;15:341–55.
- Budd GT, Cristofanilli M, Ellis MJ, Stopeck A, Borden E, Miller MC, et al. Circulating tumor cells versus imaging—predicting overall survival in metastatic breast cancer. *Clin Cancer Res* 2006;12:6403–9.
- Hayes DF, Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Miller MC, et al. Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival. *Clin Cancer Res* 2006;12:4218–24.
- Cristofanilli M, Broglio KR, Guarneri V, Jackson S, Fritsche HA, Islam R, et al. Circulating tumor cells in metastatic breast cancer:biologic staging beyond tumor burden. *Clin Breast Cancer* 2007;7:471–9.
- Sieuwerts AM, Kraan J, Bolt J, van der Spoel P, Elstrodt F, Schutte M, et al. Anti-epithelial cell adhesion molecule antibodies and the detection of circulating normal-like breast tumor cells. *J Natl Cancer Inst* 2009;101:61–6.
- Mostert B, Kraan J, Bolt-de Vries J, van der Spoel P, Sieuwerts AM, Schutte M, et al. Detection of circulating tumor cells in breast cancer may improve through enrichment with anti-CD146. *Breast Cancer Res Treat* 2011;127:33–41.
- Sieuwerts AM, Meijer-van Gelder ME, Timmermans M, Trapman AM, Garcia RR, Arnold M, et al. How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. *Clin Cancer Res* 2005;11:7311–21.
- Camps C, Buffa FM, Colella S, Moore J, Sotiriou C, Sheldon H, et al. hsa-miR-210 is induced by hypoxia and is an independent prognostic factor in breast cancer. *Clin Cancer Res* 2008;14:1340–8.
- Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, et al. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 2009;10:R64.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;57:289–300.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–8.
- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003;4:P3.
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.
- Shi W, de Graaf CA, Kinkel SA, Achtman AH, Baldwin T, Schofield L, et al. Estimating the proportion of microarray probes expressed in an RNA sample. *Nucleic Acids Res* 2010;38:2168–76.
- Hollestelle A, Nagel JH, Smid M, Lam S, Elstrodt F, Wasielewski M, et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res Treat* 2009;121:53–64.
- Mehes G, Witt A, Kubista E, Ambros PF. Circulating breast cancer cells are frequently apoptotic. *Am J Pathol* 2001;159:17–20.
- Alix-Panabières C, Brouillet JP, Fabbro M, Yssel H, Rousset T, Maudelonde T, et al. Characterization and enumeration of cells secreting tumor markers in the peripheral blood of breast cancer patients. *J Immunol Methods* 2005;299:177–88.
- Tewes M, Aktas B, Welt A, Mueller S, Hauch S, Kimmig R, et al. Molecular profiling and predictive value of circulating tumor cells in patients with metastatic breast cancer: an option for monitoring response to breast cancer related therapies. *Breast Cancer Res Treat* 2009;115:581–90.
- Fehm T, Hoffmann O, Aktas B, Becker S, Solomayer EF, Wallwiener D, et al. Detection and characterization of circulating tumor cells in blood of primary breast cancer patients by RT-PCR and comparison to status of bone marrow disseminated cells. *Breast Cancer Res* 2009;11:R59.
- Bölke E, Orth K, Gerber PA, Lammering G, Mota R, Peiper M, et al. Gene expression of circulating tumour cells in breast cancer patients. *Eur J Med Res* 2009;14:426–32.
- Wang HY, Cheng Z, Malbon CC. Overexpression of mitogen-activated protein kinase phosphatases MKP1, MKP2 in human breast cancer. *Cancer Lett* 2003;191:229–37.

44. Cariatì M, Naderi A, Brown JP, Smalley MJ, Pinder SE, Caldas C, et al. Alpha-6 integrin is necessary for the tumorigenicity of a stem cell-like subpopulation within the MCF7 breast cancer cell line. *Int J Cancer* 2008;122:298–304.
45. Flores LM, Kindelberger DW, Ligon AH, Capelletti M, Fiorentino M, Loda M, et al. Improving the yield of circulating tumour cells facilitates molecular characterisation and recognition of discordant HER2 amplification in breast cancer. *Br J Cancer* 2010;102:1495–502.
46. Cho WC. MicroRNAs: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. *Int J Biochem Cell Biol* 2010;42:1273–81.
47. Schwartz LH, Koerner FC, Edgerton SM, Sawicka JM, Rio MC, Bellocq JP, et al. pS2 expression and response to hormonal therapy in patients with advanced breast cancer. *Cancer Res* 1991;51:624–8.
48. Gradilone A, Naso G, Raimondi C, Cortesi E, Gandini O, Vincenzi B, et al. Circulating tumor cells (CTCs) in metastatic breast cancer (MBC): prognosis, drug resistance and phenotypic characterization. *Ann Oncol* 2011;22:86–92.
49. Pestrin M, Bessi S, Galardi F, Truglia M, Biggeri A, Biagioni C, et al. Correlation of HER2 status between primary tumors and corresponding circulating tumor cells in advanced breast cancer patients. *Breast Cancer Res Treat* 2009;118:523–30.
50. Cristofanilli M. The biological information obtainable from circulating tumor cells. *Breast* 2009;18 Suppl 3:S38–40.
51. Hoefnagel LD, van de Vijver MJ, van Slooten HJ, Wesseling P, Wesseling J, Westenend PJ, et al. Receptor conversion in distant breast cancer metastases. *Breast Cancer Res* 2010;12:R75.
52. O'Malley FP. Lobular neoplasia: morphology, biological potential and management in core biopsies. *Mod Pathol* 2010;23 Suppl 2:S14–25.