

## mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients

Anieta M. Sieuwerts<sup>1\*</sup>, Bianca Mostert<sup>2</sup>, Joan Bolt-de Vries<sup>1</sup>, Dieter Peeters<sup>3</sup>, Felix E. de Jongh<sup>4</sup>, Jacqueline M.L. Stouthard<sup>5</sup>, Luc Y. Dirix<sup>3</sup>, Peter A. van Dam<sup>3</sup>, Anne van Galen<sup>1</sup>, Vanja de Weerd<sup>1</sup>, Jaco Kraan<sup>2</sup>, Petra van der Spoel<sup>2</sup>, Raquel Ramírez-Moreno<sup>6</sup>, Carolien H.M. van Deurzen<sup>7</sup>, Marcel Smid<sup>1</sup>, Jack X. Yu<sup>8</sup>, John Jiang<sup>8</sup>, Yixin Wang<sup>8</sup>, Jan W. Gratama<sup>2</sup>, Stefan Sleijfer<sup>2</sup>, John A. Foekens<sup>1</sup>, and John W.M. Martens<sup>1</sup>.

<sup>1</sup> Department of Medical Oncology, Josephine Nefkens Institute and Cancer Genomics Centre, Erasmus Medical Center, Rotterdam, The Netherlands.

<sup>2</sup> Department of Medical Oncology, Daniel den Hoed Cancer Center, Erasmus Medical Center, Rotterdam, The Netherlands.

<sup>3</sup> Translational Cancer Research Group, Laboratory of Pathology, University Hospital Antwerp and General Hospital Sint-Augustinus, Wilrijk, Belgium.

<sup>4</sup> Ikazia Hospital, Department of Internal Medicine, Rotterdam, The Netherlands.

<sup>5</sup> Maastad Hospital, Department of Internal Medicine, Rotterdam, The Netherlands.

<sup>6</sup> Department of Biochemistry and Molecular Biology Physiology, Genetic and Immunology, University of Las Palmas de Gran Canaria, Canary Islands Cancer Research Institute, Canary Islands, Spain.

<sup>7</sup> Department of Pathology, Josephine Nefkens Institute, Erasmus Medical Center, Rotterdam, The Netherlands.

<sup>8</sup> Molecular Diagnostics R&D, Veridex LLC, North Raritan, New Jersey, United States of America.

\* Corresponding author, address reprint requests to:

Anieta M. Sieuwerts  
Erasmus MC, Josephine Nefkens Institute Room BE-400  
Dr. Molewaterplein 50  
3015 GE Rotterdam  
Netherlands  
telephone: +31 10 7044372  
FAX: +31 10 7044377  
Email: [a.siewerts@erasmusmc.nl](mailto:a.siewerts@erasmusmc.nl)

### Funding

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

**Running title:** Gene transcripts in primary tumors and CTCs

**Keywords:** circulating tumor cells; CellSearch profile; mRNA; microRNA; leukocytes; metastatic breast cancer.

## **STATEMENT OF 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. CTCs 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 demonstrate 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 based on primary tissue.

## Abstract

**Purpose:** Molecular characterization of circulating tumor cells (CTCs) 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 EpCAM-based CellSearch™ Profile Kit. Selected genes were measured by real-time RT-PCR in CTCs of 50 metastatic breast cancer patients collected before starting first line systemic therapy, in blood from 53 healthy blood donors (HBDs) 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 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 through 4 had developed both visceral and non-visceral metastases. Comparing transcript levels in CTCs with those measured in corresponding primary tumors showed clinically relevant discrepancies in ER 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.

## Introduction

Molecular characterization of primary tumors has already greatly contributed to the personalized treatment of cancer patients. High-throughput techniques have yielded knowledge of mutations or epigenomic changes in certain genes as well as prognostic and predictive models based on mRNA and 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 under treatment. However, most of these prognostic and predictive models have been developed based on primary tumor tissue while 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.

CTCs are tumor cells circulating in the peripheral blood of patients, shed from either the primary tumor or its metastases. A recently developed technology to quantify the number of CTCs in whole blood is the CellSearch™ CTC Test (Veridex LLC, Raritan, NJ). So far this is the only test which has been approved by the US Food and Drug Administration (FDA)(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-EpCAM antibodies, isolated cells are stained with

fluorescently labeled monoclonal antibodies specific for epithelial cells (CK-8/18/19), for leukocytes (CD45) and their nuclei with a nuclear staining dye (DAPI), and subsequently enumerated by a semi-automated 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, while EpCAM-based enrichment eliminates a large proportion of leukocytes (~4-log depletion), there are still considerable quantities of contaminating leukocytes (DAPI<sup>+</sup>/CD45<sup>+</sup>) 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 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 RT-PCR on as little as five CTCs isolated from 7.5 mL of blood, which is considered the clinically relevant cut off in patients with metastatic breast cancer(24-26), in an environment containing excess quantities of up

---

*AM Sieuwerts et al.*

to 1,000(18) contaminating leukocytes. This robust and novel method allows the determination of miRNA and mRNA expression levels in CTCs and the exploration of their clinical relevance as shown in this study for patients with metastatic breast cancer.

## 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 x 7.5 mL blood samples were prospectively taken for CTC enumeration and isolation (for details see below) 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 below), 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, 10 in the Ikazia Hospital Rotterdam and 10 in the Maasstad Hospital Rotterdam, The Netherlands, and 21 patients in the Oncology Center GZA St-Augustinus, Antwerpen, Belgium). For 8 out 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 clinico-pathological information for these 50 patients and the 8 matching primary tissues is given in **Table 1** and in addition after dichotomization of patients at the for breast cancer clinically relevant level of 5 CTCs(24-26) in **Supplemental Table 1**. Fifty-three HBD blood samples were drawn from laboratory volunteers and blood donors of the Sanquin Blood Bank South-west Region (Rotterdam, The Netherlands).

## **Enumeration of CTCs**

7.5 mL blood of HBDs and of metastatic breast cancer patients prior to the administration of first line systemic therapy was drawn in CellSave™ tubes (Veridex LLC, Raritan, NJ). For CTC enumeration, samples were processed on the CellTracks™ AutoPrep System (Veridex LLC) 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 before(27, 28).

## **miRNA and mRNA isolation from CTCs, FF and FFPE**

For gene expression studies, in parallel with the enumeration studies, 7.5 mL blood of 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). The cells in the enriched CTC fractions were lysed by adding 250  $\mu$ L of Qiagen AllPrep DNA/RNA Micro Kit Lysis Buffer (RLT+ lysis buffer) (Qiagen BV, Venlo, The Netherlands) and stored immediately at  $-80^{\circ}\text{C}$  until RNA isolation was performed with the AllPrep DNA/RNA Micro Kit (Qiagen) according to the manufacturer's instructions. In brief, by using a gradient of ethanol (Absolute Ethyl Alcohol (EtOH) Merck, Darmstadt, Germany) the larger RNAs ( $>200$  nt) were first captured in a RNeasy Mini spin column in the presence of 35% EtOH and eluted separately from the small RNA molecules ( $\leq 200$  nt) present in the flow through. These  $>200$  nt aliquots were treated with DNase I according the manufacturer's instruction. Next, the  $\leq 200$  nt molecules present in the flow through were captured in a new RNeasy Mini spin column in the presence of a final concentration of 60% EtOH and thus eluted separately from the  $>200$  nt molecules. Using this approach consisting of two sequential filtrations with different ethanol concentrations, a 12  $\mu$ L RNA fraction highly enriched in



RNA species  $\leq 200$  nt and a 14  $\mu\text{L}$  RNA fraction enriched in RNA species  $> 200$  nt could be obtained from the same sample.

Total RNA was isolated from FF tissue with RNA-Bee as described before(29) and from FFPE tissue with the column-based High Pure RNA Paraffin Kit (Roche Applied Science, Almere, The Netherlands) according to the manufacturer's instructions.

### **(Stem-loop) cDNA synthesis, pre-amplification and real-time PCR (qRT-PCR)**

The generation of pre-amplified cDNA from total RNA from the FF and FFPE tissues and the  $> 200$  nt RNA fractions and subsequent Taqman-based qRT-PCR analysis, as well as 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 nM each) were pooled, concentrated for 60 min in a speed vacuum centrifuge at  $50^{\circ}\text{C}$  and resuspended in nuclease-free ddH<sub>2</sub>O to a final concentration of 50 nM 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. Twenty-five to 50 ng of total RNA sample aliquots were reverse-transcribed in a final volume of 20  $\mu\text{L}$  with a final concentration of 12.5 nM for each RT primer using the TaqMan MicroRNA for reverse transcription kit (Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands (ABI)) according to the manufacturer's instructions and as described before(21).

For the miRNA quantification in the CTC samples, 3  $\mu\text{L}$   $\leq 200$  nt RNA aliquots were reverse-transcribed in a final volume of 7.5  $\mu\text{L}$  with a final concentration of 12.5 nM for each RT primer (ABI), 0.65 mM of each dNTP (ABI), 3 mM magnesium chloride (Invitrogen, Breda, The Netherlands), 0.3 U/ $\mu\text{L}$  RNase inhibitor (ABI), 15 U/ $\mu\text{L}$  RevertAid™ H Minus enzyme (Fermentas, St. Leon-Rot, Germany) and 1x RT buffer (Fermentas). Cycling conditions

were according to the “Megaplex RT reaction for TaqMan microRNA array” protocol from ABI, *i.e.*, 40 cycles of 16°C for 2 min, 42°C for 1 min and 50°C for 1 sec, followed by a final incubation at 85°C for 5 min and a cool down to 4°C. Prior to PCR, half of the resulting multiplex cDNA was linearly pre-amplified 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 ddH<sub>2</sub>O and analyzed by real-time PCR in a 20 µL reaction volume in a Mx3000P™ Real-Time PCR System (Stratagene, Amsterdam, The Netherlands) using the individual TaqMan MicroRNA 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 pre-amplification run, and the resulting data were used to normalize for in-between experimental variations. 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 reverse transcriptase and samples in which total RNA and cDNA was replaced with ddH<sub>2</sub>O. Quantitative values were obtained from the threshold cycle (Ct) 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 **Supplemental Table 2A** for the miRNAs and **Supplemental Table 2B** for the mRNAs. For the identification phase of potentially CTC-specific miRNA transcripts, the TaqMan Human MicroRNA Assay Set (Sanger miRBase v10; Applied Biosystems (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 below). Next, these levels were compared with those measured in a pool of 6 CellSearch-enriched preparations from HBDs for potentially differentially expressed miRNAs. These pre-screen 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#*, *hsa-miR-379* because of the observed difference between ER-positive and -negative samples in the pre-screen, *RNU6B* as being a potential reference miRNA, and *hsa-miR-210*, which has shown clinically relevance in breast cancer(21, 30) (**Supplemental Table 2A**).

For the mRNA transcripts, clinically relevant and potentially CTC-specific genes were selected *in silico* based on 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 pre-screen 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

---

*AM Sieuwerts et al.*

qRT-PCR, that were potentially higher expressed in breast tumor cells relative to leukocytes (**Supplemental Table 2B**).

### **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*) the control 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 Ct method, which is the difference between the median Ct of the appropriate control genes and the Ct of the target gene. Only samples that were at the median Ct of all miRNAs and the median Ct of *HMBS*, *HPRT1* and *GUSB* able to generate a signal within an arbitrarily chosen cut off set at 26 Ct were considered of sufficient quality and quantity to enter the study. By the use of this threshold, 11 out 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 Ct of the appropriate reference set, after which, for each individual assay, the median Ct measured in CellSearch-enriched HBDs (n=31 for the mRNAs and

n=8 for the miRNAs) was used as a cut off Ct for the CTC samples, with all Ct values exceeding this cut off Ct considered as undetectable.

### **Statistical analysis**

Statistical analysis was performed using 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%), which 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 non-parametric Spearman rank correlation test ( $r_s$ ). Gene expression levels in the various fractions were compared using the non-parametric 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/>)(33), was used to cluster the samples based on the gene expression values and TreeView (<http://rana.lbl.gov/eisen/>)(33) was used to visualize the results. DAVID ([david.abcc.ncifcrf.gov](http://david.abcc.ncifcrf.gov/))(34, 35) was used to functionally annotate genes and to identify 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 two-sided with  $P < 0.05$  considered as statistically significant.

## Results and Discussion

### 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 (~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 non-homogeneously 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 non-homogeneously amplified miRNA and mRNA assays. To achieve this, we combined the already sensitive multiplex stem-loop cDNA approach with the Taqman-based linear pre-amplification method, both from ABI. To validate the sensitivity and linear and homogeneous nature of this combined technique, we performed comparative tests between serially diluted non-amplified and multiplexed pre-amplified cDNA from total RNA of pooled primary breast tumors, as described before(18). The homogeneity of amplification

was set at a cut off of 2 Ct, *i.e.*, for an assay to be considered homogeneously amplified, the number of cycles that were required after pre-amplification should be within a 2 Ct range of the number of cycles that were required for the non-amplified material. After adjusting for the median 15.5 Ct gain due to the pre-amplification procedure, data of 11 miRNA assays were outside this range (**Table 2, lower part**). After testing the 43 miRNAs in a multiplex cDNA PCR reaction in our patient cohort of 50 CTC samples, data of two additional miRNAs (*hsa-miR-10b* and *RNU43*) had to be discarded because they generated very poor amplification curves. Finally, the PCR efficiency of two of the remaining 30 assays were outside our set range of 75% to 125% (*hsa-miR-135b*, 135% and *hsa-miR-452#*, 73%) and were therefore also excluded from our final analyses (**Table 2, column 6**). A resume 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 pre-amplification procedure, are 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 pre-amplification procedure from ABI(18). To certify that the miRNA data generated with these assays and the multiplex pre-amplification procedure were also reproducible in-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 Ct  $\pm$  95% confidence interval) of the 28 miRNAs measured in this control sample in 28 independently performed multiplexed pre-amplified qRT-PCR sessions (**Figure 1**) with a median coefficient of variation (CV) at the absolute Ct level of 6%, ranging from 3% for *hsa-miR-200a#* to 15% for *hsa-miR-184*, illustrates 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 pre-amplification reaction with an anticipated over 10-fold (median 160-fold) higher expression in CTCs relative to blood-derived leukocytes (**Table 2 and Supplemental Table 2A**). Of these 28 small RNAs, only one miRNA (*hsa-miR-183*) was higher expressed in the 32 samples that contained at least 5 CTCs compared with the 9 samples without detectable CTCs after the CellSearch procedure. At a FDR of 10%, 9 additional miRNA transcripts were more abundantly expressed in the preparations that contained at least 5 CTCs relative to whole blood 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 (**Supplemental Table 2B**), 55 were at a FDR of 10% significantly higher expressed in the 32 samples of patients with at least 5 CTCs compared with 31 CellSearch-enriched HBD samples. A gene expression call rate of 55 out 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 whole blood 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 compared with the 32 patient samples with at least 5 CTCs (**Table 3B**).



It should be noted that the CTC counts are derived from one of the 2 aliquots of 7.5 mL blood samples that was processed with the CellSearch Epithelial Kit, while the other aliquot used for the molecular profiling was processed with the CellSearch Profiling Kit. This inevitably introduces stochastic variation between the tumor cell content in the two 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(37). The given cell counts can therefore only be used as a rough estimate for our molecular profile. Nevertheless, with 14 out of 55 mRNAs (25.4%) and only one (*hsa-miR-183*) out of 28 miRNAs (3.6%) higher expressed in the 32 samples that contained at least 5 CTCs compared with the 9 samples without detectable CTCs after the CellSearch enrichment procedure with the Epithelial Kit, it appears 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 were able to measure them might be associated with the remarkable stability of miRNA transcripts in blood(38). Indeed, the detection of an additional 9 out of 28 (32.1%) miRNAs that were higher expressed in breast cancer patients without detectable CTCs than in whole blood preparations of HBDs prior to (n=14) or after (n=8) CellSearch enrichment, compared to an additional 6 out of 55 (10.9%) for mRNAs, further supports this thought.

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

Next, unsupervised two-dimensional average linkage hierarchical cluster analysis(33) was performed to compare the gene expression profiles of our 50 patients. For this we used the

---

*AM Sieuwerts et al.*

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 3B and Figure 2**).

This analysis resulted in a clustering of 4 groups of patients with a clear discrimination in-between **patient cluster 1** and **patient clusters 2 through 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 (**Figure 2**).

Regarding the gene clustering, 5 gene clusters with a correlation >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*(39), *ESR1*, *KRT18*(39) and *MUC1*, of which *MUC1* has previously been used by others for the detection of CTCs in breast cancer(40-44). 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.

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 appears to be inversely related to the “mutagenesis” gene cluster 2 -which includes *ESR1*-, the gene that encodes for the estrogen receptor (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 appears to be dominated by genes associated with cell cycle progression and proliferation such as *DUSP4* (*MKP2(45)*), *KIF11*, *KPNA2* and *MKI67*. Interestingly, a putative stem cell marker (*ITGA6(46)*) is also included in this last cluster. Of note is that half the patients with relatively high CTC-associated *ESR1* levels expressed relatively low levels of *TFF1*. *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).

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-HBDs) from which we had data from both the mRNAs and miRNAs (**Supplemental Figure 1**). These HBDs (marked in green below the cluster) indeed clustered closely together. Also the patients from patient cluster 1 (**Figure 2**, and marked in red below the cluster diagram in **Supplemental Figure 1**), that were characterized by the lack of expression of epithelial marker genes, remain clustered together, next to the HBD cluster. This confirms 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 through 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 expression of epithelial marker genes. Detailed clinico-pathological information of our patient cohort, subdivided in two groups (patient cluster 1 *versus* clusters 2 through 4) based on 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 cluster 2 through 4 displayed a 2-fold higher rate of having both visceral and non-visceral metastases, as opposed to only visceral or non-visceral metastasis for the patients of cluster 1.

Almost identical results were obtained when the associations of primary tumor and patient characteristics were studied based on CTC count subdivided in two groups (patients with less than 5 CTCs *versus* patients with at least 5 CTCs) (**Supplemental Table 1**).

### **Associations of gene transcripts measured in CTCs with current drug targets**

Although we were not able to measure *PGR* transcripts reliably in the CTCs due to the relatively high *PGR* levels present in the contaminating leukocytes, we were able to measure the *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 cluster 2 through 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**. For one of the patients whose primary tumor was assessed to be 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. Vice-versa and perhaps even more disturbing, in 11 out 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, while according to the primary tumor characteristics these patients would have an indication for anti-hormonal treatment, no benefit might be expected from this therapy based on these CTC characteristics. However, due to the limited number of 4 of these 11 patients that were actually treated with anti-hormonal 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. Similarly, the CTCs of at least 4 patients with HER2-negative primary tumors showed to be positive at the time of metastatic disease, while 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 based on primary tumor characteristics, while 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 measured in the primary tumor and metastases and CTCs have been described before at both the mRNA and the protein level(37, 42, 48-51), indicating that the findings with our multi-gene measuring technique may indeed be relevant.

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

We were able to retrieve 8 primary tumor tissues (3x FFPE and 5x 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 (**Figure 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). This indicates that we have indeed succeeded to measure true tumor-specific genes in CTCs with our CTC-specific 65-gene panel, and managed to avoid generation of a predominant leukocyte-derived signal.

The only obvious discrepancy was observed between the CTCs and primary tumor of 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 is reassuring, discrepancies in expression of individual genes, such as for *ESR1* in patients 5, 6 and 8 (**Figure 4**), are 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* compared with the primary tumor (**Figure 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.

## **Concluding remarks**

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 co-captured 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.

## **Acknowledgements**

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.



## References

1. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;6(11):857-66.
2. Fan C, Oh DS, Wessels L, *et al.* Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;355(6):560-69.
3. Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007;128(4):683-92.
4. van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* 2008;452(7187):564-70.
5. Esteller M. Epigenetics in cancer. *N Engl J Med* 2008;358(11):1148-59.
6. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer* 2010;10(6):389-402.
7. Kang Y, Siegel PM, Shu W, *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 2003;3(6):537-49.
8. Mimori K, Kataoka A, Yoshinaga K, *et al.* Identification of molecular markers for metastasis-related genes in primary breast cancer cells. *Clin Exp Metastasis* 2005;22(1):59-67.
9. Suzuki M, Tarin D. Gene expression profiling of human lymph node metastases and matched primary breast carcinomas: clinical implications. *Mol Oncol* 2007;1(2):172-80.
10. Park S, Holmes-Tisch AJ, Cho EY, *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(7):809-15.
11. Cavalli LR. Molecular markers of breast axillary lymph node metastasis. *Expert Rev Mol Diagn* 2009;9(5):441-54.
12. Gevaert O, Daemen A, De Moor B, Libbrecht L. A taxonomy of epithelial human cancer and their metastases. *BMC Med Genomics* 2009;2(69):1-16.
13. Campbell PJ, Yachida S, Mudie LJ, *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 2010;467(7319):1109-13.
14. 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(91):26036-38.
15. Danila DC, Heller G, Gignac GA, *et al.* Circulating tumor cell number and prognosis in progressive castration-resistant prostate cancer. *Clin Cancer Res* 2007;13(23):7053-58.
16. Cohen SJ, Alpaugh RK, Gross S, *et al.* Isolation and characterization of circulating tumor cells in patients with metastatic colorectal cancer. *Clin Colorectal Cancer* 2006;6(2):125-32.
17. Cristofanilli M, Budd GT, Ellis MJ, *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* 2004;351(8):781-91.
18. Sieuwerts AM, Kraan J, Bolt-de Vries J, *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(3):455-68.
19. Krek A, Grun D, Poy MN, *et al.* Combinatorial microRNA target predictions. *Nat Genet* 2005;37(5):495-500.
20. Croce CM, Calin GA. miRNAs, cancer, and stem cell division. *Cell* 2005;122(1):6-7.
21. Foekens JA, Sieuwerts AM, Smid M, *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(35):13021-26.
22. Bartels CL, Tsongalis GJ. MicroRNAs: novel biomarkers for human cancer. *Clin Chem* 2009;55(4):623-31.
23. Tong AW, Nemunaitis J. Modulation of miRNA activity in human cancer: a new paradigm for cancer gene therapy? *Cancer Gene Ther* 2008;15(6):341-55.
24. Budd GT, Cristofanilli M, Ellis MJ, *et al.* Circulating tumor cells versus imaging--predicting overall survival in metastatic breast cancer. *Clin Cancer Res* 2006;12(21):6403-09.
25. Hayes DF, Cristofanilli M, Budd GT, *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(14 Pt 1):4218-24.
26. Cristofanilli M, Broglio KR, Guarneri V, *et al.* Circulating tumor cells in metastatic breast cancer: biologic staging beyond tumor burden. *Clin Breast Cancer* 2007;7(6):471-79.

---

AM Sieuwerts *et al.*



27. Sieuwerts AM, Kraan J, Bolt J, *et al.* Anti-epithelial cell adhesion molecule antibodies and the detection of circulating normal-like breast tumor cells. *J Natl Cancer Inst* 2009;101(1):61-6.
28. Mostert B, Kraan J, Bolt-de Vries J, *et al.* Detection of circulating tumor cells in breast cancer may improve through enrichment with anti-CD146. *Breast Cancer Res Treat* 2010;Epub ahead of print.
29. Sieuwerts AM, Meijer-van Gelder ME, Timmermans 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(20):7311-21.
30. Camps C, Buffa FM, Colella S, *et al.* hsa-miR-210 is induced by hypoxia and is an independent prognostic factor in breast cancer. *Clin Cancer Res* 2008;14(5):1340-48.
31. Mestdagh P, Van Vlierberghe P, De Weer A, *et al.* A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 2009;10(6):R64.
32. Benjamini Y, Y. H. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B* 1995;57(1):289-300.
33. 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(25):14863-68.
34. Dennis G, Jr., Sherman BT, Hosack DA, *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003;4(5):P3.
35. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4(1):44-57.
36. Shi W, de Graaf CA, Kinkel SA, *et al.* Estimating the proportion of microarray probes expressed in an RNA sample. *Nucleic Acids Res* 2010;38(7):2168-76.
37. Flores LM, Kindelberger DW, Ligon AH, *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(10):1495-502.
38. Cho WC. MicroRNAs: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. *Int J Biochem Cell Biol* 2010;42(8):1273-81.
39. Hollestelle A, Nagel JH, Smid M, *et al.* Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res Treat* 2009;121(1):53-64.
40. Mehes G, Witt A, Kubista E, Ambros PF. Circulating breast cancer cells are frequently apoptotic. *Am J Pathol* 2001;159(1):17-20.
41. Alix-Panabieres C, Brouillet JP, Fabbro M, *et al.* Characterization and enumeration of cells secreting tumor markers in the peripheral blood of breast cancer patients. *J Immunol Methods* 2005;299(1-2):177-88.
42. Tewes M, Aktas B, Welt A, *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(3):581-90.
43. Fehm T, Hoffmann O, Aktas B, *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(4):R59.
44. Bolke E, Orth K, Gerber PA, *et al.* Gene expression of circulating tumour cells in breast cancer patients. *European journal of medical research* 2009;14(10):426-32.
45. Wang HY, Cheng Z, Malbon CC. Overexpression of mitogen-activated protein kinase phosphatases MKP1, MKP2 in human breast cancer. *Cancer Lett* 2003;191(2):229-37.
46. Cariati M, Naderi A, Brown JP, *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(2):298-304.
47. Schwartz LH, Koerner FC, Edgerton SM, *et al.* pS2 expression and response to hormonal therapy in patients with advanced breast cancer. *Cancer Res* 1991;51(2):624-28.
48. Gradilone A, Naso G, Raimondi C, *et al.* Circulating tumor cells (CTCs) in metastatic breast cancer (MBC): prognosis, drug resistance and phenotypic characterization. *Ann Oncol* 2010.
49. Pestrin M, Bessi S, Galardi F, *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(3):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, *et al.* Receptor conversion in distant breast cancer metastases. *Breast Cancer Res* 2010;12(5):R75.
52. O'Malley FP. Lobular neoplasia: morphology, biological potential and management in core biopsies. *Mod Pathol* 2010;23 Suppl 2:S14-25.

**Figure 1: Reproducibility of the measurement of 28 miRNAs in a multiplex cDNA pre-amped reaction.**

Bars depict miRNA expression of 28 individual miRNA assays relative to the median level of all miRNA assays (average delta Ct  $\pm$  95% confidence interval) as measured in a control sample at 300 pg total RNA input in 28 independently performed multiplexed pre-amplified qRT-PCR sessions.

**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 was controlled with 3 additional universal reference genes (*GUSB*, *HPRT1* and *HMBS*). Prior to real-time PCR, cDNA was pre-amplified in 15 cycles with the PreAmp method from ABI as described in the Materials and Methods section, 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  $>0.2$ . 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 grey for easy identification.

**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 non-parametric Mann-Whitney U test to identify genes with significantly different expression levels in-between groups.

**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 was controlled with 3 additional universal reference genes (*GUSB*, *HPRT1* and *HMBS*). Prior to real-time PCR, cDNA was pre-amplified in 15 cycles with the PreAmp method from ABI as described in the Materials and Methods section, 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 grey 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.

**Supplemental Figure 1, on line only:**

**Unsupervised hierarchical cluster analysis comparing gene expression profiles in healthy blood donors and 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 blood of 14 healthy blood donors and CTC-enriched fractions of 50 metastatic breast cancer patients. Sample loading and RNA integrity was controlled with 3 additional universal reference genes (*GUSB*, *HPRT1* and *HMBS*). Prior to real-time PCR, cDNA was pre-amplified in 15 cycles with the PreAmp method from ABI as described in the Materials and Methods section, 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. The number of CTCs as established by the CellSearch Epithelial Kit is given right of the graph. Here, HBD samples are marked green and the in Figure 2 identified patients of cluster 1 are marked red. The grey marked cells indicate that the clustering of these CTC samples agrees with the expected absence of a epithelial-derived CTC signal.

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

Characteristic	No of patients*	%*	Primary tissue patients from cluster 2-4*	Molecular profile		P**
				patients cluster 1	patients cluster 2-4	
<b>All patients</b>	<b>50</b>	<b>100%</b>	<b>8</b>	<b>14</b>	<b>36</b>	
<b>Time between primary surgery and CTC sampling</b>						
≤5 years	23	46%	5	6	17	1.00
>5 years	15	30%	1	4	11	
unknown or primary not removed	12	24%	2	4	8	
<b>Age at CTC sampling</b>						
≤50 year	10	20%	0	4	6	0.44
>50 year	40	80%	8	10	30	
<b>Menopausal status</b>						
premenopausal	16	32%	0	6	10	0.51
postmenopausal	30	60%	7	8	22	
<b>Grade (Bloom-Richardson)</b>						
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
<b>Pathological tumor size</b>						
pT1, ≤2 cm	10	20%	1	3	7	0.38
pT2-4, >2 cm	31	62%	7	5	26	
<b>Lymph nodes involved</b>						
no	12	24%	2	4	8	1.00
yes	29	58%	5	9	20	
<b>ER status<sup>†</sup></b>						
negative	12	24%	3	6	6	0.07
positive	38	76%	5	8	30	
<b>PgR status<sup>†</sup></b>						
negative	24	48%	5	9	15	0.20
positive	22	44%	3	4	18	
<b>HER2/neu status<sup>†</sup></b>						
negative	29	58%	8	7	22	1.00
positive	12	24%	0	3	9	
<b>Histological type</b>						
lobular	11	22%	3	3	8	1.00
ductal	35	70%	5	9	26	
<b>Adjuvant chemotherapy</b>						
no	34	68%	5	11	23	0.50
yes	16	32%	3	3	13	
<b>Adjuvant hormonal therapy</b>						
no	31	62%	7	11	20	0.20
yes	19	38%	1	3	16	
<b>Any adjuvant therapy</b>						
no	27	54%	5	5	22	0.13
yes	23	46%	3	9	14	
<b>Site of metastasis</b>						
visceral	12	24%	1	6	6	0.05
non-visceral	7	14%	0	3	4	
both	31	62%	7	5	26	

\* Due to missing cases numbers to not always add up to 100%.

\*\* P for two-tailed Fisher's exact test.

† As retrieved from the pathology reports.

**Table 2: Quality control qRT-PCR miRNA before and after multiplexed pre-amplification.**

miRNA assay	Serially diluted breast tumor pool 1	Serially diluted breast tumor pool 2	Ct gain due to multiplexed pre-amplification	PCR efficiency in uniplex cDNA reaction without pre-amplification	PCR efficiency in multiplex cDNA reaction with pre-amplification	Reason for exclusion
(specific for epithelial breast tumor cells)	(median Ct in uniplex cDNA reaction without pre-amplification)	(median Ct in multiplex cDNA reaction with pre-amplification)	(relative to unamplified uniplex cDNA reaction)			
<i>hsa-miR-100</i>	28.67	11.71	16.96	102%	98%	
<i>hsa-miR-10a</i>	31.06	13.86	17.20	98%	108%	
<i>hsa-miR-125b</i>	28.33	12.65	15.68	95%	100%	
<i>hsa-miR-135a</i>	34.89	17.60	17.29	115%	77%	
<i>hsa-miR-141</i>	29.07	12.36	16.71	97%	105%	
<i>hsa-miR-183</i>	32.99	17.57	15.42	92%	120%	
<i>hsa-miR-184</i>	35.66	20.21	15.45	88%	123%	
<i>hsa-miR-187</i>	31.77	17.00	14.77	105%	93%	
<i>hsa-miR-193b</i>	34.05	19.14	14.91	99%	107%	
<i>hsa-miR-200a#</i>	32.73	18.09	14.64	96%	109%	
<i>hsa-miR-200b</i>	27.96	12.65	15.31	97%	89%	
<i>hsa-miR-200c</i>	24.32	10.32	14.00	96%	75%	
<i>hsa-miR-205</i>	29.98	14.15	15.83	94%	93%	
<i>hsa-miR-210</i>	30.15	13.18	16.97	86%	95%	
<i>hsa-miR-214</i>	28.73	12.93	15.80	101%	79%	
<i>hsa-miR-31</i>	29.87	12.40	17.47	101%	76%	
<i>hsa-miR-34a</i>	33.45	17.96	15.49	104%	103%	
<i>hsa-miR-375</i>	30.47	13.24	17.23	90%	75%	
<i>hsa-miR-379</i>	35.42	19.97	15.45	98%	103%	
<i>hsa-miR-424</i>	35.76	19.52	16.24	112%	73%	
<i>hsa-miR-452</i>	35.85	20.67	15.18	108%	94%	
<i>hsa-miR-455</i>	31.00	16.35	14.65	100%	97%	
<i>hsa-miR-497</i>	33.98	17.20	16.78	98%	79%	
<i>hsa-miR-565</i>	30.54	13.98	16.56	98%	123%	
<i>hsa-miR-615</i>	32.47	17.04	15.43	100%	110%	
<i>hsa-miR-9#</i>	33.00	18.59	14.41	104%	117%	
<i>hsa-miR-95</i>	34.47	19.41	15.06	82%	116%	
<i>RNU19</i>	31.78	15.81	15.97	108%	88%	
<i>RNU43</i>	31.02	13.63	17.39	83%	117%	poor amplification curves
<i>hsa-miR-10b</i>	31.71	16.79	14.92	98%	103%	poor amplification curves
<i>hsa-miR-135b</i>	28.86	14.16	14.70	99%	132%	poor PCR efficiency
<i>hsa-miR-452#</i>	31.92	17.98	13.94	105%	73%	poor PCR efficiency
<i>hsa-miR-149</i>	29.82	10.35	19.47	111%	72%	not homogeneously amplified
<i>hsa-let-7e</i>	33.61	14.30	19.31	99%	84%	not homogeneously amplified
<i>hsa-let-7i</i>	33.14	14.77	18.37	88%	116%	not homogeneously amplified
<i>hsa-miR-196a</i>	30.33	17.06	13.27	97%	81%	not homogeneously amplified
<i>hsa-miR-199a#</i>	28.18	19.21	8.97	99%	123%	not homogeneously amplified
<i>hsa-miR-200a</i>	29.62	17.18	12.44	103%	88%	not homogeneously amplified
<i>hsa-miR-203</i>	30.13	11.79	18.34	97%	91%	not homogeneously amplified
<i>hsa-miR-429</i>	34.65	25.65	9.00	98%	111%	not homogeneously amplified
<i>hsa-miR-511</i>	32.01	18.77	13.24	76%	94%	not homogeneously amplified
<i>hsa-miR-99a</i>	34.94	13.13	21.81	89%	104%	not homogeneously amplified
<i>snoR-12 (RNU6B)</i>	33.91	16.20	17.71	93%	94%	not homogeneously amplified

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 x 43 individual reverse transcription (RT) reactions and,

AM Sieuwerts et al.

after a 20-fold dilution, measured at 25, 6.25 and 1.56 ng in 43 individual miRNA-specific PCRs. For pool 2, only one sample of each dilution was used to measure all 43 miRNAs in a multiplex cDNA reaction, resulting, after an additional pre-amplification 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 was required to generate a signal after pre-amplification was 15.5. The homogeneity of amplification was therefore set at  $15.5 \pm 2$  Ct. All assays that gained less than 13.5 or over 17.5 Ct after this multiplexed pre-amplification 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 in-between 75% and 125% considered acceptable.



Gene transcripts in primary tumors and CTCs  
**Table 3A: Evaluation of miRNA transcript levels in healthy blood donors 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		Normalized on the median level of all miRNAs		Normalized on the median level of all miRNAs		Normalized on the median level of all miRNAs		Normalized on the median level of all miRNAs		Normalized on the median level of all miRNAs		Normalized on the median level of all miRNAs				
	Average	SEM	Average	SEM	Ratio ≥5 CTCs vs HBD_CS	P-value vs HBD_CS	10% FDR	Average	SEM	Ratio ≥5 CTCs vs HBD_WB	P-value vs HBD_WB	10% FDR	Average	SEM	Ratio ≥5 CTCs vs 0 CTCs	P-value	10% FDR
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	2.3E-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	





Gene transcripts in primary tumors and CTCs

MELK	0.67	0.16	0.21	0.00	3.23	1.3E-04	*	0.40	0.01	1.66	5.6E-01	0.51	0.13	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	0.00	0.00	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	3.75	0.98	1.31	4.7E-01
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
TMS6F13	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
CTTN/EMS1	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
SI00A7	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
Nm23-H1/NME1	45.63	25.49	10.10	0.00	4.52	6.4E-02	*	18.06	0.95	2.53	1.4E-01	24.59	12.02	1.86	3.1E-02
CCNB1	1.69	0.60	0.40	0.00	4.22	2.0E-01	*	0.71	0.02	2.36	4.0E-01	0.93	0.23	1.81	1.8E-01
CCNE1	1.82	0.41	1.17	0.00	1.56	9.0E-02	*	1.96	0.06	0.93	1.9E-01	2.26	0.49	0.80	9.9E-01
CCNE2	2.41	0.74	0.87	0.00	2.76	9.0E-02	*	1.65	0.06	1.46	8.1E-01	1.15	0.14	2.11	2.5E-01
CD133	0.00	0.00	0.00	0.00	1.13	6.7E-01	*	0.01	0.00	0.72	4.2E-01	0.00	0.00	1.13	7.9E-01
CDH1	7.47	4.17	1.27	0.01	5.87	9.9E-01	*	1.96	0.05	3.80	3.7E-01	3.32	1.27	2.25	2.5E-01

*Gene transcripts in primary tumors and CTCs*

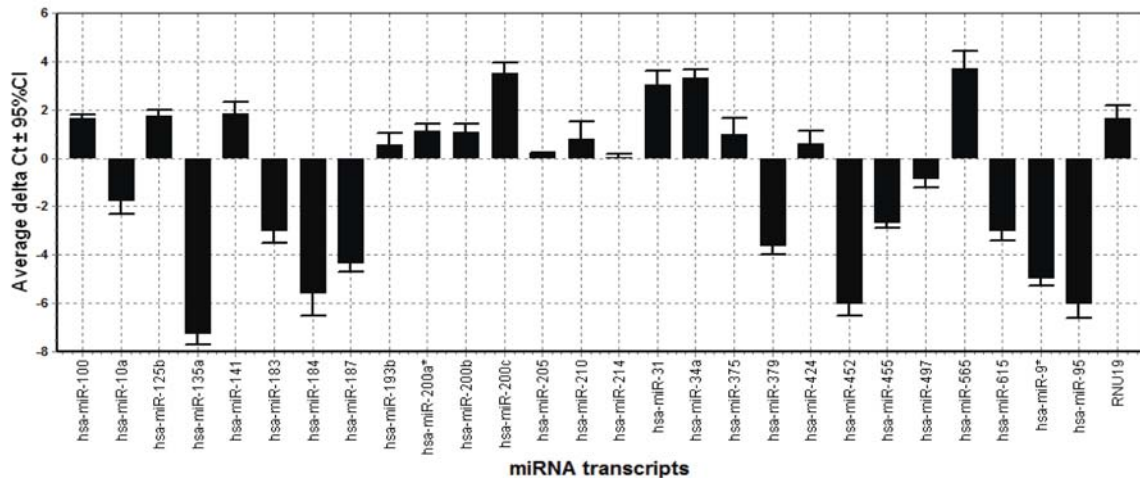
COL1A1	0.01	0.00	0.00	1.50	4.0E-01	0.00	0.00	1.50	5.1E-01	0.00	0.00	1.48	9.1E-01
COL2A1	0.00	0.00	0.00	1.03	8.4E-01	0.00	0.00	1.03	8.5E-01	0.00	0.00	1.03	9.0E-01
EGFR	0.01	0.00	0.00	1.76	2.9E-01	0.00	0.00	1.76	4.1E-01	0.01	0.00	1.08	8.6E-01
ERBB2	5.79	3.05	1.24	4.68	2.0E-01	2.62	0.11	2.21	9.0E-01	2.09	0.38	2.77	3.8E-01
FGFR2	0.14	0.02	0.09	1.67	2.0E-01	0.16	0.00	0.93	2.0E-01	0.09	0.00	1.67	4.0E-01
GALGT	0.01	0.00	0.00	2.23	9.0E-02	0.00	0.00	2.23	1.9E-01	0.01	0.00	1.23	5.8E-01
GATA3	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
KRTHB1/KRT81	0.01	0.00	0.00	1.72	2.9E-01	0.01	0.00	1.07	1.2E-01	0.01	0.00	0.98	8.9E-01
MAGEA3	0.00	0.00	0.00	1.26	5.3E-01	0.00	0.00	1.26	6.2E-01	0.00	0.00	1.26	6.8E-01
MET	0.04	0.00	0.03	1.32	2.9E-01	0.05	0.00	0.81	1.0E-01	0.05	0.01	0.88	9.1E-01
MSMB	0.05	0.01	0.02	2.31	9.0E-02	0.04	0.00	1.23	3.1E-01	0.03	0.00	1.50	8.6E-01
MYL3	0.00	0.00	0.00	1.00	9.9E-01	0.00	0.00	1.00	9.9E-01	0.00	0.00	1.00	9.9E-01
PLK1	1.37	0.47	0.39	3.54	2.0E-01	0.64	0.02	2.13	9.5E-01	0.84	0.19	1.62	4.9E-01
PTPRK	0.60	0.21	0.10	5.78	9.9E-01	0.19	0.01	3.15	4.2E-01	0.13	0.01	4.64	7.0E-02
SELE	0.01	0.00	0.00	1.92	2.9E-01	0.00	0.00	1.92	4.1E-01	0.00	0.00	1.92	4.9E-01
STK6/AURKA	0.01	0.00	0.00	1.40	1.4E-01	0.01	0.00	0.90	3.5E-01	0.01	0.00	1.04	6.7E-01
TOP2a	5.17	1.99	1.80	2.88	6.7E-01	2.93	0.10	1.76	9.0E-01	3.36	0.74	1.54	4.1E-01
TWIST	0.02	0.00	0.01	1.35	2.9E-01	0.02	0.00	0.79	1.0E-01	0.01	0.00	1.30	8.1E-01
VEGFR2	0.01	0.00	0.00	2.11	1.4E-01	0.01	0.00	1.13	6.3E-01	0.01	0.00	1.31	6.5E-01

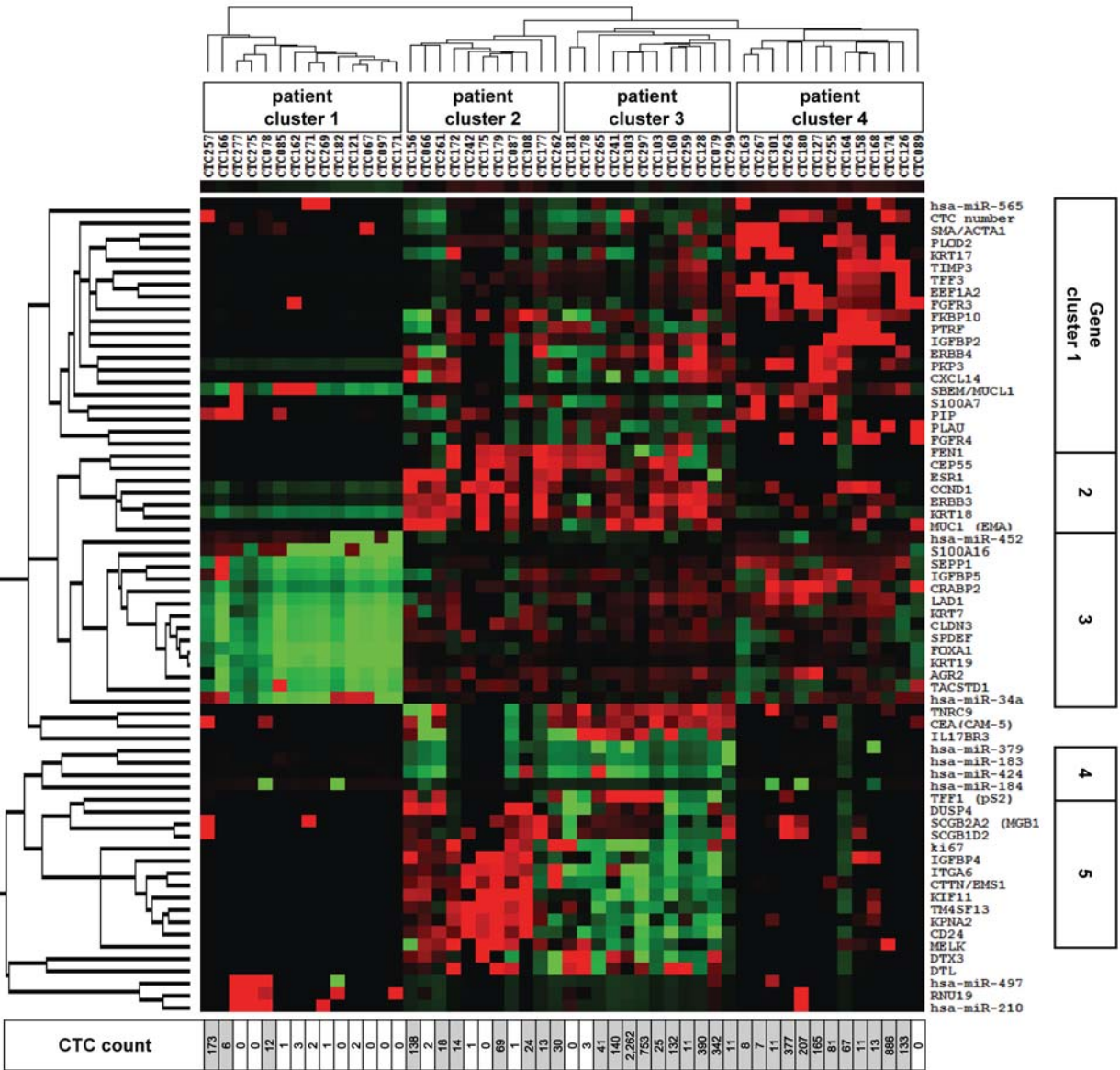
The expression levels of 28 miRNAs (Table 3A) and 87 mRNAs (Table 3B) were measured in RNA isolated with the CellSearch Profiling Kit as described in the Materials and Methods section. To identify putative CTC-specific genes, transcript levels were compared in-between various groups with and without CTCs as established by the CellSearch Epithelial Kit in a separate 7.5 mL blood sample. The average expression  $\pm$  standard error of the mean (SEM) is given for the number of samples indicated at the top in the Table.

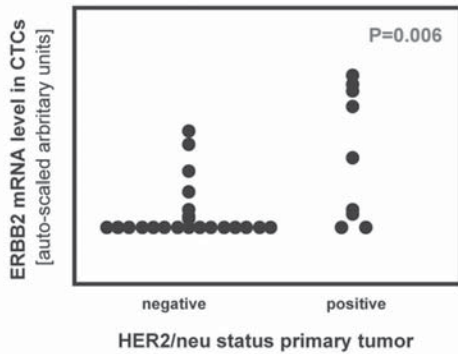
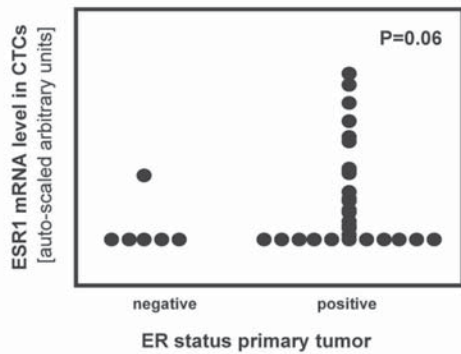
The 2-tailed Mann-Whitney U test was used to identify genes with a significantly different expression level in-between groups. To compensate for multiple testing, a FDR of 10% was applied on these statistics.

\* Significant at a 10% FDR (2-tailed  $P < 0.05$ ).

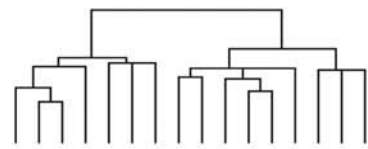
CS, enriched by EpCAM-based CellSearch; CTC, circulating tumor cell; HBD, healthy blood donor; WB, whole blood.



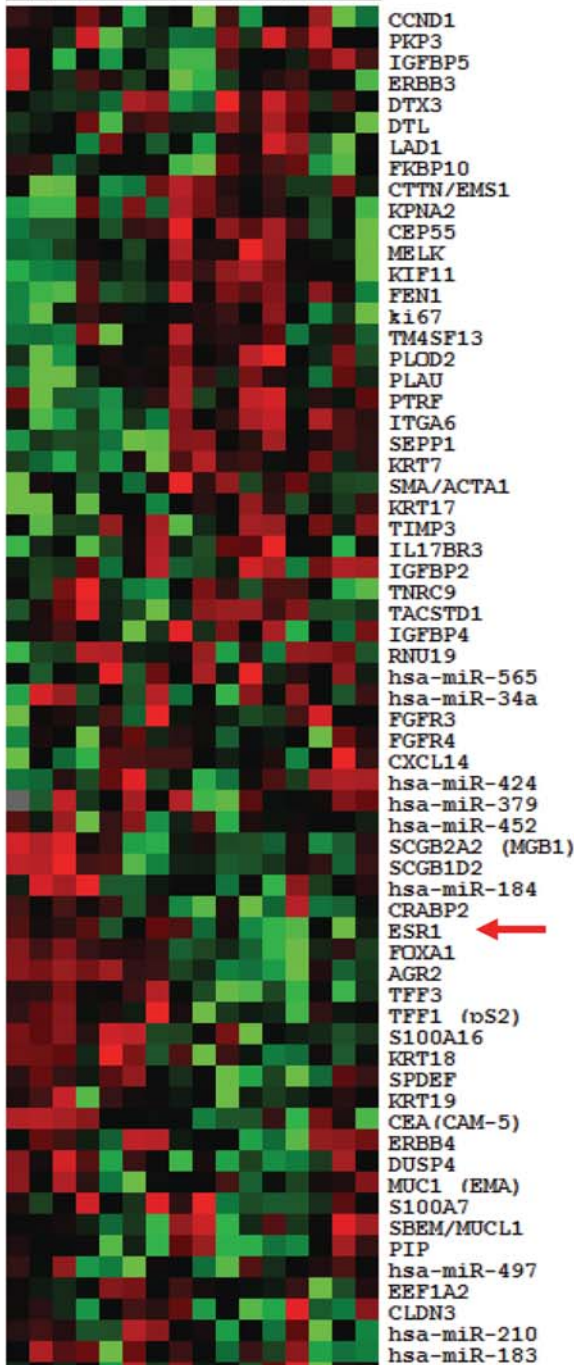








CTC241	CTC303	CTC297	T241	CTC180	T180	T297	T267	CTC267	CTC301	CTC174	T174	T301	T303	CTC255	T255
1	2	3	1	4	4	3	5	5	6	7	7	6	2	8	8



# Clinical Cancer Research

## mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients

Anieta M Sieuwerts, Bianca Mostert, Joan Bolt-de Vries, et al.

*Clin Cancer Res* Published OnlineFirst April 19, 2011.

<b>Updated version</b>	Access the most recent version of this article at: doi: <a href="https://doi.org/10.1158/1078-0432.CCR-11-0255">10.1158/1078-0432.CCR-11-0255</a>
<b>Supplementary Material</b>	Access the most recent supplemental material at: <a href="http://clincancerres.aacrjournals.org/content/suppl/2011/04/21/1078-0432.CCR-11-0255.DC1">http://clincancerres.aacrjournals.org/content/suppl/2011/04/21/1078-0432.CCR-11-0255.DC1</a>
<b>Author Manuscript</b>	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, contact the AACR Publications Department at [permissions@aacr.org](mailto:permissions@aacr.org).