Gene Expression Profiling of Circulating Tumor Cells in Breast Cancer

Emanuela Fina, Maurizio Callari, Carolina Reduzzi, Francesca D’Aiuto, Gabriella Mariani, Daniele Generali, Marco A. Pierotti, Maria G. Daidone, and Vera Cappelletti

BACKGROUND: Determining the transcriptional profile of circulating tumor cells (CTCs) may allow the acquisition of clinically relevant information while overcoming tumor heterogeneity-related biases associated with use of tissue samples for biomarker assessment. However, such molecular characterization is challenging because CTCs are rare and outnumbered by blood cells.

METHODS: Here, we describe a technical protocol to measure the expression of >29 000 genes in CTCs captured from whole blood with magnetic beads linked with antibodies against epithelial cell adhesion molecule (EpCAM) and the carcinoma-associated mucin, MUC1, designed to be used for CTC characterization in clinical samples. Low numbers of cells (5–200) from the MCF7 and MDA-MB-468 breast cancer cell lines were spiked in healthy donor blood samples and isolated with the AdnaTest EMT-1/Stem CellSelect kit. Gene expression profiles (GEPs) were obtained with the WG-DASL HT assay and compared with GEPs obtained from RNA isolated from cultured cell lines and unspiked samples.

RESULTS: GEPs from samples containing 25 or more spiked cells correlated (r = 0.95) with cognate 100-ng RNA input samples, clustered separately from blood control samples, and allowed MCF7 and MDA-MB-468 cells to be distinguished. GEPs with comparable technical quality were also obtained in a preliminary series of clinical samples.

CONCLUSIONS: Our approach allows technically reliable GEPs to be obtained from isolated CTCs for the acquisition of biologically useful information. It is reproducible and suitable for application in prospective studies to assess the clinical utility of CTC GEPs, provided that >25 CTCs can be isolated. © 2014 American Association for Clinical Chemistry

Attempting to guide cancer treatment on the basis of the features of the primary tumor (PT) may not represent an optimal approach owing to frequently reported discordance between the primary tumor and metastatic sites and the occurrence of intratumoral clonal heterogeneity. Moreover, in clinical practice, metastatic lesions are seldom biopsied owing to their anatomic inaccessibility or the comorbidity associated with the procedure. Blood-based biomarker monitoring represents a new direction in the development of a precision medicine approach tailored to provide information on the specific progression step of the disease.

Circulating tumor cells (CTCs) are the purported intermediates of metastatic dissemination and are likely to contain cellular clones responsible for disease progression; CTCs therefore represent a preferred source for the identification of drug targets. Beyond the clinical validity of CTC counts in both early and metastatic breast cancer patients, CTCs offer the possibility of obtaining information on the disease in real time without invasive biopsies; a biological characterization of such cells is likely to be more representative of the disease evolution and treatment resistance than the PT. Unfortunately, molecular characterization of CTCs is seriously hampered by their low numbers and by the contamination of blood samples with leukocytes. Primarily for these 2 reasons, most studies have pursued a candidate gene approach focusing on a limited number of genes known to give therapeutic information, and only few studies have performed an unbiased characterization of numerous genes by array approaches. Because CTCs
are likely to represent a very heterogeneous cell population, single-cell profiling has been applied to investigate their transcriptional heterogeneity (19–22) as well as multimarker approaches on the bulk of CTC population (23–27).

With the goal of providing a reliable assay allowing the acquisition of valuable information on CTC features in the clinical setting, we have adapted a commercially available method that captures CTCs by use of beads coated with antibodies against epithelial cell adhesion molecule (EpCAM) and the carcinoma-associated mucin, MUC1. Extensive gene expression profiling is then performed on the captured cells with the cDNA-mediated annealing, selection, extension, and ligation (DASL) platform, which allows measurement of the expression of 29,000 genes in low-quantity RNA samples (28). The method has been tested with cell line cells spiked at different numbers into healthy donor blood (HDB) samples, and has been subsequently validated in metastatic breast cancer patients entering systemic treatment at the Fondazione Istituto di Ricovero e Cura a CarattereScientifico (IRCCS) Istituto Nazionale Tumori of Milan and at the Breast Unit of Cremona Hospital.

Materials and Methods

CELL CULTURE AND CELL SPIKING
We grew breast cancer cell lines MCF7 and MDA-MB-468 (ATCC) in DMEM/F-12 (Lonza) medium supplemented with 10% fetal bovine serum (Lonza) in humidified 5% CO₂ atmosphere. Cell lines were authenticated by short tandem repeats DNA profiling with the StemElite™ ID System kit (Promega). For spiking experiments, cells were harvested with TrypLE™ Select 1X (Invitrogen, Life Technologies), and single cells from highly diluted cell suspensions were micropipetted under an inverted optical microscope directly into 5 mL whole HDB in AdnaCollect EDTA collection tubes (AdnaGen). Samples were processed immediately or stored at 4 °C for ≤2 h after spiking.

CLINICAL SAMPLES
We carried out gene expression profile analysis on CTCs quantified in parallel by different approaches (CellSearch and AdnaTest) from blood of 7 patients with advanced breast cancers entering primary systemic treatment protocols ongoing at the Breast Unit, Istituti Ospitalieri at Cremona (6 cases, CTC counted by CellSearch) and at the Medical Oncology Department at Istituto Nazionale Tumori, Milan (1 case, CTC evaluated by AdnaTest). Blood samples for CTC gene expression (5 mL) were collected before starting a new line of treatment, drawn into BD Vacutainer K₃EDTA tubes, and processed within 1 h after collection. Blood samples (7 mL) for CTC counts with the Cellsearch™ System were collected in parallel into CellSave Preservative sample tubes (Janssen Diagnostik), whereas an additional 5 mL were collected in parallel for CTC determination by AdnaTest. Cremona and Milan institutional review board and ethics committees approved the study protocol, and all patients provided written informed consent.

The patients’ clinicopathologic features and CTC status/counts are summarized in Supplemental Table 1, which accompanies the online version of this article at http://www.clinchem.org/content/vol61/issue1.

TUMOR CELL ENRICHMENT, RNA EXTRACTION, AND SINGLE-GENE EXPRESSION EVALUATION
We isolated tumor cells spiked into blood by the AdnaTest EMT-1/StemCell Select kit (AdnaGen AG). Captured cells were incubated with lysis buffer from the Agencourt RNAadvance Cell v2 kit (Beckman Coulter) for 30 min at room temperature, and bead-free lysates were stored at 80 °C. We performed total RNA extraction according to the manufacturer’s instructions; RNA was eluted in 5 or 10 μL nuclease-free water (Ambion), depending on the experimental setting.

For single-gene expression studies, enriched tumor cells were incubated with the AdnaTest lysis buffer, and mRNA was isolated with the Dynabeads® mRNA Direct™ Micro Kit included in the AdnaTest and reverse-transcribed. We used the cDNA obtained as template for evaluation of PTPRC (protein tyrosine phosphatase, receptor type, C)⁶ (the gene coding for CD45) with the TaqMan assay Hs 00365634_g1 (Applied Biosystems).

WHOLE-GENOME EXPRESSION PROFILING ASSAY
We performed the Illumina Human Whole-Genome DASL HT Assay according to manufacturer’s instructions. The BeadChips were imaged on the BeadArray Reader.

A flow chart summarizing the experimental design is reported in Fig. 1.

DATA PROCESSING AND STATISTICS
Microarray raw data were generated with Illumina BeadStudio 3.8 software and processed with the Lumi package (29) of Bioconductor. For each gene, we selected the probe with the highest interquartile range. Array data were deposited at Gene Expression Omnibus data repository (GSE55470).

⁶ Human genes: PTPRC, protein tyrosine phosphatase, receptor type, C; EPCAM, epithelial cell adhesion molecule; MUC1, mucin 1, cell surface associated; ERBB2, v-erb-b2 avian erythroblast leukemia viral oncogene homolog 2, (HER2); AKT2, v-akt murine thymoma viral oncogene homolog 2; PIRK2, phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha; TWIST1, twist family bHLH transcription factor 1; ALDH1A1, aldehyde dehydrogenase 1 family, member A1. See Fig. 6 legend for gene symbols and names mentioned only in that figure.
Unsupervised hierarchical clustering was performed with the Geneplotter package with Euclidean distance and average linkage.

## Results and Discussion

**Suitability of the DASL Assay to Determine Gene Expression Profiles of Epithelial Cells Spiked at Low Frequencies in HDB and Efficacy of AdnaWash Buffer**

To evaluate the reliability of the DASL assay for obtaining gene expression profiles (GEPs) from both low RNA-input samples and CTCs, experiments were run with low predefined numbers of breast cancer cells (MCF7 and MDA-MB-468), spiked into a HDB at final concentrations of 400, 200, and 100 cells/5 mL. For this step, a single healthy donor was chosen to ensure comparable samples when investigating the efficiency of leukocyte removal and the interference by leukocyte-specific genes in GEPs from spiked cells.

Total RNA from captured cells was split in 2 equal aliquots and processed in duplicate on 2 distinct chips of the DASL platform together with total RNA (0.5–}

---

**Fig. 1. Flowchart summarizing the experimental design.**

The experimental steps carried out on isolated RNA extracted from cell lines and from healthy donor blood samples spiked with different amounts of cells are reported highlighting for each step the technical and biological issues addressed by the experiment. BC, breast cancer; HD, healthy donor.
levels in samples with lower RNA input. About one-third of high-RNA-input sample were found to be at background of genes spanning low to intermediate intensities in the line Supplemental Fig. 1A, the expression levels of a subset to r 4 Clinical Chemistry

Experiment was designed to (a) verify the sensitivity of the assay; (b) test the reliability of GEP obtained with 200, 100, and 50 cells; and (c) assess the interassay variability (under conditions in which no interference by capture-associated variability was present). In addition, blood samples from the same healthy donor, not containing any spiked cells, were processed similarly with and without use of the AdnaWash buffer (designed to limit leukocyte contamination in the captured samples) to test its ability to prevent contamination by leukocyte-expressed genes.

We generated diagnostic plots (Fig. 2) to judge the technical reliability of the data computing mean signal intensities and probe detection rates (Fig. 2A). As expected, the lowest mean log2 signal intensities (around 5) and probe detection rates (30%) were observed in healthy controls when captured cells were treated with AdnaWash buffer before lysate preparation. These results suggest that after treatment with the washing buffer, few cells remained attached to the immunomagnetic beads. In fact, both the mean signal values and detection rates increased, respectively, to 6% and 45% compared with the lower values obtained in untreated control samples, demonstrating that AdnaWash buffer effectively reduced leukocyte contamination.

The plots in Fig. 2A indicate that both the spiked samples and the samples with higher RNA input were characterized by mean log2 intensities between 7 and 7.7 and probe detection rates >60%. In the case of isolated RNA samples, only those containing 1 or 0.5 ng of RNA showed lower detection rates and intensities.

Fig. 2B summarizes the correlations between samples. Samples from HDB, either untreated or treated with AdnaWash, were weakly correlated with the remaining samples, with a poor correlation between technical duplicates. A separation on the basis of GEP was obtained between the MCF7 and MDA-MB-468 blood-spiked cells, supporting the biological reliability of the GEPs obtained from cells spiked into HDB and captured with immunomagnetic beads. Of note, MDA-MB-468 cells spiked into blood and samples containing their isolated RNA (in the concentration range 10–100 ng) clustered together. However, low-input RNA samples (1 or 0.5 ng) deriving from MDA-MB-468 cells clustered separately. In fact, when the RNA input was decreased from 10 to 1 or 0.5 ng, the correlation coefficients with profiles obtained from a 100-ng RNA sample (an RNA input within the range suggested by the chip manufacturer) dropped from $r = 0.99$ to $r = 0.89$ and $r = 0.83$, respectively. As reported in online Supplemental Fig. 1A, the expression levels of a subset of genes spanning low to intermediate intensities in the high-RNA-input sample were found to be at background levels in samples with lower RNA input. About one-third of the genes were present only in samples with high RNA content, whereas the remainder were found also in low-RNA-input samples. Importantly, the 6578 probes exclusively detected in high-RNA-input samples represent a group of low-expression genes that were not significantly enriched in any specific biological process (see online Supplemental Fig. 1B), suggesting that a lower RNA input affects low-expression genes belonging to any cellular process, without causing any modification in the biological interpretation of data. Such a result, therefore, allows confidence in the GEPs obtained from samples with lower RNA input, as also suggested by April et al. (28).

The efficacy of AdnaWash in removing the contribution of leukocyte-derived genes is further shown in Fig. 2C. Removal of leukocytes is crucial for wide gene expression analysis, since many genes are expressed by both epithelial and mesenchymal cells, and consequently leukocyte contamination can bias GEPs of putative CTCs. The top scatterplot in Fig. 2C correlates GEPs obtained from the same HDB with or without washing steps with AdnaWash buffer. A substantial number of genes with a wide expression range in samples not treated with AdnaWash appeared to be expressed at low levels after leukocyte removal and, overall, samples not treated with AdnaWash were characterized by higher expression levels for many genes. These data suggest high leukocyte contamination in immunomagnetic bead-captured samples not submitted to AdnaWash treatment.

The next 2 scatterplots in Fig. 2C report correlations between GEPs obtained from a healthy donor sample after a washing step with AdnaWash without any spiked cells and samples from the same donor that were spiked with 50 MCF7 or MDA-MB-468 cells. It appears that for the majority of genes, cell-spiked HDB samples showed higher gene expression levels, suggesting an efficient removal of leukocytes without affecting epithelial cells. Furthermore, a distinct set of genes characterized by a wide range of expression in samples spiked with cells showed only a low expression in unspiked HDB samples, suggesting an enrichment in epithelial-specific genes. The specificity of AdnaWash buffer in the removal of leukocytes is finally supported by the last scatterplot in Fig. 2C, which shows that distinguishing cell-type specific genes from genes equally expressed by MCF7 and MDA-MB-468 cells is possible. On the basis of all the reported observations, we concluded that the treatment with AdnaWash buffer represents a useful step to limit leukocyte contamination that does not affect detection of epithelial-specific genes.

Finally, reproducibility of the AdnaWash procedure for the removal of leukocytes was tested by evaluating the levels of PTPRC gene expression with a quantitative PCR approach in 7 samples from the same healthy donor spiked with breast cancer cells. Different
Fig. 2. Intensities and correlations.

(A), Distribution plot of signal mean intensities in gene expression data derived from isolated MDA-MB-468 RNAs (green squares), from 50–100–200 MDA-MB-468 (red dots) or MCF7 (blue dots) cells spiked into healthy donor blood (HDB) and from cell-free HDB (yellow triangles), profiled with the Illumina Human Whole-Genome DASL HT. (B), Heat map reporting the reciprocal correlations (Pearson correlation coefficient) among GEPs obtained from RNA directly isolated from MDA-MB-468 cells or from captured cell RNA corresponding to 200, 100, and 50 MDA-MB-468 or MCF7 cells spiked into HDB. All GEPs were obtained by Illumina Human Whole-Genome DASL HT. Before computation of the correlations, data were separately normalized with the robust spline normalization method. (C), Correlations between samples: AdnaWash buffer-treated vs -untreated HDB samples; AdnaWash buffer-treated HDB containing 50 MCF7 spiked cells vs untreated HDB; AdnaWash buffer-treated HDB containing 50 MDA-MB-468 spiked cells vs untreated HDB; AdnaWash buffer-treated HDB containing 50 MDA-MB-468 spiked cells vs AdnaWash buffer-treated HDB containing 50 MCF7 spiked cells. AW, AdnaWash.
numbers of breast cancer cells (0–50) were spiked into 7 independent 5-mL blood samples obtained from the same healthy donor at a single time point to minimize leukocyte variation. After the standard washing procedure with the AdnaWash buffer, captured cells were lysed, and PTPRC gene expression was assayed by quantitative PCR. The mean cycle threshold (Ct) value was 36.53 (0.46) with a CV of 1.26%. The CV value obtained supports the reproducibility of the leukocyte removal.

BIOLOGICAL RELIABILITY OF GEPs OBTAINED FROM CELLS SPIKED INTO BLOOD
We assessed the biological reliability of GEPs by comparing genes shared by captured cells and their corresponding isolated RNA and genes selectively expressed by samples derived from isolated RNA (100 ng) or from captured cells (200 cells). Results for MDA-MB-468 cells are shown in Fig. 3A. Reliability of profiles was supported by the high number of shared genes (90.4%) and by the fact that genes exclusively expressed by captured cells (5.5%) were enriched in gene ontology terms referring to leukocytes (e.g., immune response, immune system process, lymphocyte activation), whereas genes exclusively expressed in isolated RNA samples did not show a specific type of gene ontology enrichment. A list of gene ontology terms exclusively enriched in captured cells and in the corresponding isolated RNA is reported in online Supplemental File 1.

The actual enrichment in specific gene categories in cells captured from spiked blood samples compared with control HDB was evaluated considering both types of controls, either washed or not with AdnaWash buffer (Fig. 3B). A proportion of genes (41.5%) expressed by all 3 samples was enriched in terms referring to biological functions common to all types of cells (e.g., RNA processing, cellular macromolecule catabolic process, RNA binding), whereas a smaller set of genes (21.9%) was shared between samples spiked with cells and the control samples that were not treated to remove leukocytes. Such genes that are enriched in cellular biological functions common to all types of cells likely represent common genes with lower expression levels, and this could explain their absence in the AdnaWash-treated samples. The efficacy of leukocyte removal was supported by the gene ontology of the 430 genes exclusively expressed in controls untreated with AdnaWash that were enriched in 151 gene ontology terms that all referred to leukocytes and included immune response, inflammation, leukocyte activation, macrophage activation, and lymphocyte differentiation (see online Supplemental File 2). The same analysis was repeated for MCF7 spiked cells, and a similar interpretation could be made (see online Supplemental Fig. 2 and Supplemental File 3). Similar results were also obtained by repeating the analysis with the replicated samples (data not shown).

SENSITIVITY AND EXPERIMENTAL VARIABILITY OF THE DEVELOPED CTC GENE PROFILING METHOD
The sensitivity and experimental variability of the entire CTC capture and profiling assay was tested by spiking independent triplicates of 50, 25, 10, and 5 MCF7 and MDA-MB-468 cells into HDB. RNA (100 ng) extracted from the 2 cell lines was also hybridized in triplicate as a gold-standard sample, and in the case of the MCF7 cell line, triplicate spikes with 50 cells were also performed in parallel in a different donor.

To understand the quality of the data, we carried out Pearson correlation analysis on all samples. The resulting correlation plot is shown in Fig. 4. One sample (10 spiked MDA-MB-468 cells) found to be completely unrelated to the other samples was rated as a technical failure and was not taken into consideration in the following analyses (see also online Supplemental Fig. 3).

The remaining samples segregated into 2 clusters. A larger cluster (cluster A in Fig. 4) contained all samples derived from isolated RNA and 69% (18/26) of spiked cell samples. In both the RNA samples cluster and the spiked cells cluster, a clear separation between MCF7 and MDA-MB-468 cells was observed. For both MCF7 and MDA-MB-468 cell lines, 73% and 63.6% of spiked samples, respectively, correlated as expected.

The specificity of capture of CTCs and the efficacy of leukocyte removal is also supported by the strong correlation between GEPs derived from 50 MCF7 cells spiked into different HDB.

A second cluster of less well-correlated samples (cluster B) contained the 3 negative samples (HDB with no spiked cells) and 8 spiked samples (both with MFC7 and with MDA-MB-468 cells) with low cells numbers (4 samples spiked with 5 cells, 3 with 10 cells, and 1 with 25 cells). Although the overall number of spiked cells samples processed was not high, the data suggest that in samples containing 50 CTCs, reliable GEPs can be obtained, whereas for samples with 10 or fewer CTCs, the failure rate exceeds the success rate. In fact, the majority of low-input samples were more strongly correlated with donors rather than with other samples containing the same cell type. Finally, profiling of 25 captured cells appears to be feasible, although with a success rate of around 80% (5/6). Considering the low numbers of CTCs usually recovered in clinical samples, we therefore suggest that to ensure the successful profiling of CTCs in clinical samples, a larger volume of blood (10–15 mL) should be used.

In GEPs of CTCs from patients, the main aim is to detect biologically significant differences among individuals. To understand whether the developed method was suitable for this purpose, fold changes in gene expression
Fig. 3. Gene ontology.

(A), Eulero-Venn diagrams highlighting the numbers of common and exclusive genes detected in samples derived from 200 spiked MDA-MB-468 cells and 100 ng isolated RNA derived from MDA-MB-468 cells. Gene ontologies analysis results for exclusive genes in each set are reported with character sizes and color intensities directly proportional to the $P$ values for the gene set enrichment.

(B), Eulero-Venn diagrams highlighting the numbers of common and exclusive genes detected in samples derived from 200 MDA-MB-468 cells spiked into HDB, HDB not treated with AdnaWash buffer, and HDB treated with AdnaWash buffer. Gene ontologies analysis results for exclusive genes in each set are reported with character sizes directly proportional to the gene set enrichment.
Fig. 4. Heat map reporting the reciprocal correlations (Pearson correlation coefficient) among GEPs obtained from 100 ng RNA directly isolated from MCF7 and MDA-MB-468 cells, or from RNA extracted from 50, 25, 10, and 5 MCF7 or MDA-MB-468 cells spiked in independent triplicates into HDB.

All GEPs were obtained by Illumina Human Whole-Genome DASL HT. Before computation of the correlation values, data were separately normalized with the robust spline normalization method. On the basis of correlations, samples separated in 2 clusters: cluster A containing samples derived from isolated RNA and samples with higher number of spiked cells, and cluster B containing controls and samples spiked with low cell numbers. AW, AdnaWash.
between the luminal MCF7 cell line and the triple-negative MDA-MB-468 cell line were computed. The quality of the data obtained from spiked cells was assessed by correlating fold changes obtained from 50, 25, and 10 spiked cells with those obtained with gene expression data derived from RNA (100 ng) directly isolated from cultured cells. Data are reported in Fig. 5. Correlations among log2 fold changes dropped significantly when the number of spiked cells was reduced (fold-change correlations by Pearson coefficient: $r = 0.66$ for 50 cells, $r = 0.54$ for 25 cells, and $r = 0.26$ for 5 cells).

**APPLICATION TO CLINICAL SAMPLES**

The approach developed for spiked samples was applied to 7 clinical samples. A CellSearch-based CTC number estimation was available for 6 samples, whereas in 1 sample CTC status was evaluated by AdnaTest EMT-1/Stem CellDetect kit without any enumeration. CTC numbers ranged from 0 to 200 cells (in 5 mL whole blood). It is important to underline that this is only an estimation of the true number of CTCs profiled for gene expression, as the profiled CTCs were captured with a different kit. Although the CellSearch approach and the AdnaTest are both based on EpCAM expression, the AdnaTest also contains antibodies directed against cell-surface MUC1 antigen. In the literature, an agreement of 69% and 73% is reported between the 2 methods when using the 5 or 2 CTC/7.5 mL cutoffs (30).

A heat map for samples clustered on the basis of the expression of the PAM50 gene panel (31) is reported in Fig. 6. Except for a sample with a very high number of CTCs (sample CTC001; 200 CTCs in 5 mL blood), no correlation was observed between number of CTCs estimated by CellSearch and the expression levels of PAM50 genes. However, it should be noted that CTC counts were quite low (median CTC count 13). Moreover, the CTCs captured and profiled in this still-preliminary series of clinical samples were obtained from patients characterized by primary tumors of similar subtypes, which does not allow us to observe molecular subtype–related differences. An interesting observation was noted for sample CTC003, in which CTC status was defined as negative by AdnaGen assay on the basis of the low expression of 7 selected genes \( \text{EPICAM} \) (epithelial cell adhesion molecule), \( \text{MUC1} \) (mucin 1, cell surface associated), \( \text{ERBB2} \) (v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2, [HER2]), \( \text{AKT2} \) (v-akt murine thymoma viral oncogene homolog 2), \( \text{PIK3CA} \) (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha), \( \text{TWIST1} \) (twist family bHLH transcription factor 1), and \( \text{ALDH1A1} \) (aldehyde dehydrogenase 1 family, member A1), but still showed an expression for PAM50 panel genes not dissimilar from that of true CTC-positive samples.
Fig. 6. Unsupervised hierarchical cluster analysis of gene expression profiles from 7 metastatic patient blood samples enriched for CTC with the AdnaTest EMT-1/Stem CellSelect kit.

Robust spline normalization was applied, and probes with a detection $P < 0.01$ in at least 1 sample were selected. Horizontal rows represent genes from the PAM50 panel, and vertical columns correspond to clinical samples. CDH3, cadherin 3, type 1, P-cadherin (placental); KRT5/17/14, keratin 1/17/14; MAPT, microtubule-associated protein tau; BIRC5, baculoviral IAP repeat containing 5; KIF2C, kinesin family member 2C; EXO1, exonuclease 1; CENPF, centromere protein F, 350/400kDa; FGFR4, fibroblast growth factor receptor 4; NUF2, NDC80 kinetochore complex component (formerly CDCA1); CDC6/20, cell division cycle 6/20; SFRP1, secreted frizzled-related protein 1; EGFR, epidermal growth factor receptor; MYBL2, v-myb avian myeloblastosis viral oncogene homolog-like 2; PGR, progesterone receptor; FOXC1/A1, forkhead box C1/A1; MELK, maternal embryonic leucine zipper kinase; ANLN, anillin, actin binding protein; TYMS, thymidylate synthetase; CEP55, centrosomal protein 55kDa; MKI67, marker of proliferation Ki-67; UBEZC, ubiquitin-conjugating enzyme E2C; UBE2T, ubiquitin-conjugating enzyme E2T (putative); ESR1, estrogen receptor 1; MLPH, melanophilin; GRB7, growth factor receptor-bound protein 7; TMEM45B, transmembrane protein 45B; CCNB1/E1, cyclin B1/E1; RRM2, ribonucleotide reductase M2; ACTR3B, ARP3 actin-related protein 3 homolog B (yeast); MIA, melanoma inhibitory activity; MMP11, matrix metalloproteinase 11 (stromelysin 3); SLC39A6, solute carrier family 39 (zinc transporter), member 6; MYC, v-myc avian myelocytomatosis viral oncogene homolog; MDM2, MDM2 proto-oncogene, E3 ubiquitin protein ligase; BAG1, BCL2-associated athanogene; CXXC5, CXXC finger protein 5; BCL2, B-cell CLL/lymphoma 2; ORC6L, origin recognition complex, subunit 6; PTTG1, pituitary tumor-transforming 1; GRP160, G protein-coupled receptor 160; PHGDH, phosphoglycerate dehydrogenase; NAT1, N-acetyltransferase 1 (arylamine N-acetyltransferase); BLVRA, biliverdin reductase A.
Conclusion

CTCs are rapidly moving from a simple tumor burden marker to a potential biomarker providing information on tumor heterogeneity and tumor biology (32). A recent pooled analysis of patient data has confirmed that CTC enumeration is an independent prognostic marker of overall survival and progression-free survival that provides an early assessment of treatment response in breast cancer (9, 33). It is, however, worth reflecting on the advantages offered by CTC molecular characterization in the clinical setting where, so far, the most common approach has been dealing with determination of HER2 status on CTCs.

A direct involvement of CTCs in the metastatic cascade is unquestionable, but probably not all CTC subpopulations have the same metastatic potential. Although not accepted by all (34), epithelial-mesenchymal transition has been suggested to be crucial for dissemination of tumor cells to distant organs (35). Yu et al. (36) demonstrated the occurrence of dynamic changes in epithelial and mesenchymal composition of CTCs in patients with metastatic breast cancer and found an association between treatment resistance and presence of CTCs with mesenchymal features in patients who were serially monitored through disease progression. Such a study provides proof of concept that variation of specific CTC subpopulations can serve as an indicator of the rate of cancer adaptation during treatment and represents, therefore, an opportunity to identify new therapeutic targets and resistance markers possibly anticipating clinical progression.

This concept is further reinforced in the light of the well-known intratumor spatial/temporal heterogeneity of both primaries and metastases, which can be overcome by exploiting liquid biopsies (37) containing materials (e.g., circulating tumor DNA, CTCs, circulating microRNAs) shed in the bloodstream by primary tumors, but also by the complete set of metastases.

To fully exploit the opportunity offered by CTC molecular characterization in the clinical setting, it therefore is mandatory to have an approach allowing an extensive molecular characterization of CTCs, enabling collection of data that offer timely information on tumor progression and/or development of treatment resistance and are suitable for clinical studies in a multicenter setting.

Despite the many challenges, primarily having to do with low sample input and leukocyte contamination, the experiments run on HDB definitely support the possibility of obtaining technically reliable gene expression profiles from as few as 25 cells in 5 mL blood. Our data also show that the obtained profiles convey biologically useful information that may allow differences among samples to be distinguished.

Several published studies have reported gene expression from CTCs, but these have been limited to tens or hundreds of genes primarily detected by PCR. Whereas some of the studies (23) used an EpCAM-based enrichment approach, other studies, mostly dealing with single cells in prostate cancer, relied on complex techniques that are not suitable for translation to the clinic (38).

Our study has the advantage of providing a pipeline for obtaining biologically reliable profiles in a way that is easily applicable to the clinical context; this contrasts with single-cell profiling methods that are not yet ready to be transferred into the daily routine. A limitation in the developed method is represented by the need of somewhat high numbers of cells compared with the median values of CTCs identified in peripheral blood from breast cancer patients. However, such a problem can be solved by drawing higher volumes of blood and improving the CTC-enrichment methods to avoid losses of not fully epithelial CTCs. In fact, on the basis of preliminary data obtained in our laboratory on a limited set of breast cancer patients (21 with early and 9 with metastatic disease), the use of improved antibody cocktails (AdnaTest EMT-1/Stem vs Adna Test EMT-2/Stem CellSelect kits) for CTC enrichment raised the positivity percentage 4- and 3-fold in early and metastatic breast cancer, respectively. Although this does not represent a direct cell count like those obtainable with the CellSearch approach, the increased positivity percentage is likely to be associated with an increase in the absolute number of CTCs. So far, 5 CTCs/7.5 mL is considered the clinically significant CTC threshold for breast cancer patients; even at the lowest end of the CTC number distribution in breast cancer patients, a blood draw of 38 mL would be enough to guarantee the critical number of 25 CTCs necessary for our gene expression approach. Therefore, when using the conventional CTC-enrichment methods, our GEP protocol appears to be more suitable for application in the metastatic context where the number of CTCs is higher than it is in early breast cancer. The AdnaTest-based CTC-capture approach, with its improved antibody cocktail, offers an effective means to increase the number of detectable CTCs.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: None declared.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honorary: None declared.

Research Funding: This work was supported by the European Commission under the 7th Framework Programme under grant agreement no. 260791, EurocanPlatform, and Italian Health Ministry. M.A. Pierotti, AIRC grant 5/1000 #12162; M.G. Daidone, AIRC grant IG #10611.

Expert Testimony: None declared.

Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References