Gene expression profiles in circulating tumor cells to predict prognosis in metastatic breast cancer patients

B. Mostert¹, A. M. Sieuwerts¹,², J. Kraan¹, J. Bolt-de Vries¹, P. van der Spoel¹, A. van Galen¹, D. J. Peeters³, L. Y. Dirix³, C. M. Seynaeve¹, A. Jager¹, F. E. de Jongh⁴, P. Hamberg⁵, J. M. L. Stouthard⁶, J. F. S. Kehrer⁷, M. P. Look¹, M. Smid¹, J. W. Gratama¹, J. A. Foekens¹, J. W. M. Martens¹,² & S. Sleijfer¹,²*

¹Department of Medical Oncology; ²Department of Medical Oncology and Cancer Genomics Netherlands, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, The Netherlands; ³Translational Cancer Research Unit, Center for Oncological Research, GZA Hospitals Sint-Augustinus and University of Antwerp, Antwerp, Belgium; ⁴Ikazia Hospital, Rotterdam; ⁵Sint Franciscus Gasthuis, Rotterdam; ⁶Maasstad Hospital, Rotterdam; ⁷Department of Internal Medicine, IJsselland Hospital, Capelle aan den IJssel, The Netherlands

Background: A circulating tumor cell (CTC) count is an established prognostic factor in metastatic breast cancer (MBC). Besides enumeration, CTC characterization promises to improve outcome prediction and treatment guidance. Having shown the feasibility of quantifying clinically relevant mRNA transcripts in CTCs, we determined the prognostic value of CTC gene expression in MBC.

Patients and methods: CTCs were isolated and enumerated from blood of 197 MBC patients who were about to start first-line systemic therapy. Of these, 180 were assessable for quantification of mRNA expression by RT-qPCR in relation to time-to-treatment failure (TTF). A prognostic CTC gene profile was generated by leave-one-out cross validation in a 103 patient discovery set and validated in 77 patients. Additionally, all 180 patients were randomly divided into two equal sets to discover and validate a second prognostic profile.

Results: CTC count predicted for TTF at baseline (≥5 versus <5 CTCs/7.5 ml blood, hazard ratio (HR) 2.92 [95% confidence interval (CI) 1.71–4.95] P < 0.0001). A 16-gene CTC profile was generated in the first discovery set, which identified patients with death or TTF <9 months versus those with a better outcome. In multivariate analysis, the 16-gene profile was the only factor associated with TTF [HR 3.15 (95% CI 1.35–7.33) P 0.008]. Validation of this profile in the independent patient set pointed into the same direction, but was not statistically significant. A newly generated 8-gene profile showed similarly favorable test characteristics as the 16-gene profile, but did not significantly pass validation either.

Conclusion: A 16-gene CTC profile was identified, which provided prognostic value on top of CTC count in MBC patients. However, validation of this profile in an independent cohort, nor of a second profile, reached statistical significance, underscoring the need to further fine-tune the still promising approach of CTC characterization.

Key words: circulating tumor cells, breast cancer, gene expression, prognosis, prediction, RT-qPCR

introduction

To improve individualization of metastatic breast cancer (MBC) treatment, prognostic and predictive factors have been identified in primary tumors. However, in advanced disease, characteristics of metastases can greatly differ from those of the primary tumor. This heterogeneity has been described for clinically relevant factors such as ER [1], HER2 [2] and KRAS [3]. Consequently, characterization of metastatic tissue would be preferred, but is not always feasible. Circulating tumor cells (CTCs) are thought to represent metastatic tissue and are therefore an attractive alternative [4].

CTC counting is an established prognostic factor in MBC [5–7] and can serve as early response predictor [5]. Additionally, CTC characterization is a promising treatment tailoring tool. However, CTCs are extremely rare cells that, even after sensitive CellSearch™ (Veridex LLC, Raritan, NJ) enrichment, need to be characterized among up to 1000 remaining leukocytes [8]. When measuring CTC-specific gene transcripts by RT-qPCR, this problem of contaminating leukocytes can be overcome by
focusing on genes that are not, or at a much lower level, expressed in leukocytes. Using these stringent selection methods combined with gene-specific preamplification, we were able to reliably quantify a CTC-specific gene panel in MBC patients [9].

In this study, we explored the clinical relevance of CTC characterization by assessing the prognostic value of CTC gene expression profiles from MBC patients.

**methods**

**patients**

We conducted a prospective trial at six participating hospitals in the Netherlands and Belgium. Inclusion criteria were MBC and start of first-line endocrine or chemotherapeutic treatment; prior adjuvant therapy was permitted. Before administering the first cycle of treatment, two 7.5 ml blood samples were drawn for CTC enumeration and gene expression profiling (for details see next). After 2–5 weeks of therapy, an additional 7.5 ml blood sample was taken. This study was approved by the Erasmus MC and local Institutional Review Boards (METC 2006-248). All patients gave their written informed consent.

**CTC enumeration**

For CTC enumeration, 7.5 ml blood drawn in CellSave tubes (Veridex) was kept at room temperature and processed within 96 h after collection. Samples were processed on the CellTracks AutoPrep System (Veridex) using the CellSearch Epithelial Cell Kit and CellTracks Analyzer according to the manufacturer’s instructions [10, 11].

**RNA isolation from CTCs, RT-qPCR, and quantification of gene transcripts**

For gene expression studies, in parallel with the enumeration studies, 7.5 ml of blood was drawn in EDTA tubes and enriched for CTCs on the CellTracks AutoPrep System (Veridex) using the CellSearch Epithelial Cell Kit and CellTracks Analyzer according to the manufacturer’s instructions [10, 11].

**statistical analysis**

Primary end point was time-to-treatment failure (TTF), a composite end point defined as the time between start of first-line treatment and start of second-line treatment, due to disease progression or toxicity, or death. Patients who were alive and had not started second-line treatment were censored at last follow-up (FU) date.

Overall survival (OS) was defined as the time between start of first-line treatment and date of death. Hazard ratios (HR) for TTF and OS were calculated by univariate analysis for all established prognostic factors. In multivariate analysis, only variables with $P < 0.05$ in univariate analysis were included.

Patients in the discovery cohort were divided into a poor and a good prognosis group: patients with therapy failure or death <9 months were classified as poor prognosis, all others were classified as good prognosis. Then, a leave-one-out cross validation (LOOCV) was conducted using the Compound Covariate Predictor (CCP) within Biometric Research Branch ArrayTools (http://linus.nci.nih.gov/BRB-ArrayTools.html) starting with the previously established 55 CTC-specific mRNAs [9]. Sixteen genes were identified, of which the combined score was used to select a cut point at which 90% of poor prognosis patients were predicted correctly. This 16-gene profile was then validated in the independent patient set. Based on the profile’s test characteristics, at least 55 patients were needed to validate this profile with a power of 0.80 and a two-sided $\alpha$ of 0.05. All described $P$ values are two-sided. Kaplan–Meier survival curves were compared by log-rank testing.

**results**

**patient characteristics**

Between February 2008 and January 2011, 130 MBC patients were included. CTC counts at baseline and FU were available for 130 and for 82 patients, respectively. Sixteen and 11 patients were excluded because of insufficient mRNA quantity and quality (QQ) and insufficient FU (i.e. <9 months and no treatment failure documented in that time), respectively, leaving 103 patients forming the first discovery set (supplementary Figure S1, available at Annals of Oncology online). Between January 2011 and August 2012, another 78 patients with sufficient FU data were included under the same in- and exclusion criteria, forming the validation set. One patient was excluded because of insufficient QQ mRNA. Among these 77, the 11 patients excluded from the discovery set because of insufficient FU. Characteristics of all 197 included patients are depicted in supplementary Table S1, available at Annals of Oncology online, and were similar between discovery and validation sets. However, when comparing the initial discovery cohort of 130 patients with the 103 QQ mRNA patients, the number of patients that had received adjuvant endocrine treatment was lower in the latter (Fisher’s exact $P = 0.016$).

**CTC count predicts TTF and OS**

TTF was chosen as the primary end point to reflect the heterogeneity in clinical decision making in MBC patients. TTF reliably reflects the benefit a patient derives from a treatment, as it captures the time gained by administering that treatment. Of note, apart from two patients who switched therapies because of toxicity, all patients started second-line treatment after disease progression. However, in contrast to PFS [5, 6], TTF had not previously been correlated with CTC count. Therefore, we first verified whether CTC counts were associated with TTF in our cohort. Indeed, the 63 patients with $\geq$5 CTCs at baseline had a shorter TTF than the 67 patients with $<$5 CTCs [median TTF 10.2 versus 20 months, HR 2.92 (95% CI 1.71–4.95) $P < 0.0001$]. After 2–5 weeks of therapy, CTCs were again enumerated in 82 patients. Patients with $\geq$5 CTCs at FU had shorter TTF [9.2 versus 16.1 months, HR 2.83 (95% CI 1.39–5.76) $P = 0.004$] than patients with $<$5 CTCs. Looking at change in CTC count during therapy, all patients with $\geq$5 CTCs at FU, regardless of CTC count at baseline, had a shorter TTF than patients with persistently low CTC counts [HR 3.83 (95% CI 1.71–8.86) $P = 0.001$]. No statistically significant difference was seen between the TTF of patients with persistently low CTC counts versus those with a decline to $<$5 CTCs after a high baseline count ($P = 0.066$, for Kaplan–Meier curves, see supplementary Figure S2, available at Annals of Oncology online). In univariate analysis, besides CTC count, the presence of visceral metastases and a higher number of metastases were associated with shorter TTF. In multivariate analysis, only baseline CTC count was independently associated with TTF [HR 2.54 (95% CI 1.45–4.46) $P = 0.001$ supplementary Table S2, available at Annals of Oncology online]. CTC count at
Figure 1. Receiver operating characteristic (ROC) curves (A and B), test performance (C) and test characteristics (D) of CTC count and 16-gene CTC profile in 103 patients. AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.
baseline [HR 2.44 (95% CI 1.27–4.69) \(P = 0.007\)], at FU [HR 2.77 (95% CI 1.18–6.52) \(P = 0.019\)), and CTC count change during therapy [HR 3.26 (95% CI 1.23–8.61) \(P = 0.017\)] were also associated with OS.

### CTC gene expression

To establish the prognostic value of CTC gene expression, we chose to base our analysis on the 55 mRNA genes that were previously established to be CTC-specific [9]. While this 55-gene panel is based on its expression in patients with \(\geq 5\) CTCs, cell line spiking experiments showed the ability of this panel to detect epithelial signal in as little as one tumor cell spiked into 7.5 ml blood [9]. We were therefore confident that this CTC-specific panel would be able to pick up CTC gene expression, if indeed present, in patients with \(< 5\) counted CTCs, and included all 103 patients in our discovery set, regardless of CTC count, for the subsequent analyses.

### 16-gene CTC profile predicts for TTF

Of the 103 patients, 42 patients were classified as poor prognosis and 61 patients as good prognosis based on a cutoff at 9 months TTF. This cutoff, chosen based on the median PFS in first-line MBC patients [12, 13], was deemed valid as the median TTF in our cohort was 8.9 months (95% CI 7.3–10.2).

![Figure 2](image-url)

**Figure 2.** Kaplan–Meier plots for patient subgroups as defined by CTC count (A), the 16-gene CTC profile (B), and the combination of CTC count and 16-gene CTC profile (C). Only one patient had more than five CTCs and a favorable 16-gene CTC profile; therefore, no curve is depicted for this subgroup. Panel D combines panels A and C and shows that the 16-gene CTC panel is able to distinguish a truly good (dashed green line) and an intermediate prognosis group (dotted green line) among patients with \(< 5\) CTCs (solid green line), while no added value is seen in patients with \(\geq 5\) CTCs (blue lines; no line depicted for patients with \(\geq 5\) CTCs and a favorable profile as there was only one such patient). u.p., unfavorable profile; f.p., favorable profile.
A predictor was built based on the 55 CTC-specific genes. In univariate analysis, nine genes were at a $P < 0.05$ (supplementary Table S3, available at *Annals of Oncology* online) and 16 genes at a $P < 0.1$ differentially expressed between the good and poor prognosis group. LOOCV was carried out with these latter 16 genes, and a CCP was calculated for each sample, for which the receiver operating characteristic (ROC) curve is depicted in Figure 1A. At an area under the curve (AUC) of 0.69 (95% CI 0.59–0.80, $P = 0.0001$), the 16-gene CTC profile carried out similarly to the CTC count [AUC 0.62 (95% CI 0.51–0.73) $P = 0.0145$] (Figure 1B). Because our primary interest was correctly predicting patients with early therapy failure, we aimed for our CTC profile to identify poor prognosis patients with 90% sensitivity. At this cutoff, 76 patients had an unfavorable profile, half of whom failed treatment before 9 months (positive predictive value, PPV 50%). Twenty-seven patients had a favorable profile, of whom 23 indeed experienced no treatment failure (negative predictive value, NPV 85%). Test characteristics of both the 16-gene CTC profile and count are depicted in Figure 1C and D.

The Kaplan–Meier curves for the 16-gene CTC profile, CTC count and for the combination are shown in Figure 2. Panel A shows that a CTC count $\geq 5$ identifies poor prognosis patients among the 103 in whom the CTC profile was generated (log-rank $P < 0.001$). In Figure 2B, an early and clear distinction into a poor and good prognosis group is seen when separating patients according to the CTC profile (log-rank $P < 0.001$). The added value of the profile appears to lie mainly in its ability to further classify patients with <5 CTCs (Figure 2C and D, log-rank for trend $P < 0.001$), while this CTC profile does not identify prognostic groups among patients with $\geq 5$ CTCs.

In univariate analysis, the 16-gene CTC profile was significantly associated with TTF [HR 4.57 (95% CI 2.20–9.50) $P < 0.0001$, Table 1], as were the number of metastases [HR 1.39 (95% CI 1.13–1.72) $P = 0.002$], presence of visceral metastases [HR 1.84 (95% CI 1.05–3.23) $P = 0.035$] and CTC count at baseline [HR 3.00 (95% CI 1.73–5.19) $P < 0.001$]. Other known prognostic factors such as triple-negative status were not associated with TTF. Among patients with <5 CTCs, those with an unfavorable CTC profile had a shorter TTF [HR 4.23 (95% CI 1.57–11.42) $P = 0.004$] than those with a favorable profile. Among patients with $\geq 5$ CTCs, only one patient had an unfavorable CTC profile, so the CTC profile did not have added value. In multivariate analysis including all these prognostic factors, only the 16-gene CTC profile was an independent predictor of TTF [HR 3.15 (95% CI 1.35–7.33) $P = 0.008$].

**validation of the 16-gene profile**

Having identified a highly prognostic 16-gene CTC profile, its external validity was assessed in 77 patients who were included under the same study protocol, but whose FU completed at a later time point. All known prognostic factors were similar between the original discovery set and the validation set (supplementary Table S1, available at *Annals of Oncology* online).

In the validation set, the 16-gene profile correctly identified 18 of 22 patients with early treatment failure (82% sensitivity). Specificity was lower at 20% (Figure 3). The 16-gene profile predicted 62 patients to have a poor prognosis, of whom 18 experienced early treatment failure (PPV 29%). Eleven of 15 patients were correctly predicted to have a good prognosis (NPV 73%).

To explain the lack of statistically significant validation of our 16-gene CTC profile in the independent patient set, we hypothesized that—while the discovery and validation set were balanced for known prognostic factors—additional unknown prognostic factors could have interfered with the validation. To overcome this potential problem, we randomly divided the total group of 180 patients into two groups of 90. These two groups too were well balanced for known prognostic factors (supplementary Table S1, available at *Annals of Oncology* online). Using LOOCV, a new 8-gene CTC profile was generated (supplementary Table S3, available at *Annals of Oncology* online). At an AUC of 0.77 [(95% CI 0.67–0.87) $P < 0.0001$], the 8-gene CTC profile carried out similarly to the 16-gene profile. This 8-gene profile was then validated in the other 90 patients. Sixty-seven patients were predicted to have a poor prognosis, 28 of whom failed treatment before 9 months (PPV 42%). Twenty-three patients had a predicted favorable prognosis, of whom 18 experienced no or late treatment failure (NPV 78% $P = 0.0851$).

**discussion**

CTCs provide a unique opportunity to assess prognostic and predictive markers repeatedly during the course of disease [14–17].

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TTS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>&lt;5 versus $\geq 5$ CTCs at baseline</td>
<td>3.00</td>
<td>1.73–5.19</td>
</tr>
<tr>
<td>Presence of visceral metastases</td>
<td>1.84</td>
<td>1.05–3.23</td>
</tr>
<tr>
<td>Number of metastases</td>
<td>1.39</td>
<td>1.13–1.72</td>
</tr>
<tr>
<td>CTC 16-gene profile</td>
<td>4.57</td>
<td>2.20–9.50</td>
</tr>
</tbody>
</table>

Univariate and multivariate analysis are depicted of established prognostic factors and CTC profile. In the validation set, the 16-gene profile correctly identified 18 of 22 patients with early treatment failure (82% sensitivity). Specificity was lower at 20% (Figure 3). The 16-gene profile predicted 62 patients to have a poor prognosis, of whom 18 experienced early treatment failure (PPV 29%). Eleven of 15 patients were correctly predicted to have a good prognosis (NPV 73%). To explain the lack of statistically significant validation of our 16-gene CTC profile in the independent patient set, we hypothesized that—while the discovery and validation set were balanced for known prognostic factors—additional unknown prognostic factors could have interfered with the validation. To overcome this potential problem, we randomly divided the total group of 180 patients into two groups of 90. These two groups too were well balanced for known prognostic factors (supplementary Table S1, available at *Annals of Oncology* online). Using LOOCV, a new 8-gene CTC profile was generated (supplementary Table S3, available at *Annals of Oncology* online). At an AUC of 0.77 [(95% CI 0.67–0.87) $P < 0.0001$], the 8-gene CTC profile carried out similarly to the 16-gene profile. This 8-gene profile was then validated in the other 90 patients. Sixty-seven patients were predicted to have a poor prognosis, 28 of whom failed treatment before 9 months (PPV 42%). Twenty-three patients had a predicted favorable prognosis, of whom 18 experienced no or late treatment failure (NPV 78% $P = 0.0851$).
We have previously shown that measurement of a CTC-specific panel of 55 mRNAs in CTCs is feasible despite their low numbers and presence in a leukocyte background [9].

In the current study, a 16-gene CTC profile distinguished patients with poor prognosis (defined as treatment failure <9 months after start of treatment) from patients with good prognosis. The profile was designed based on a cutoff with 90% sensitivity, resulting in an 85% NPV. In clinical practice, it is particularly important to identify poor prognosis patients to balance the chance of benefit from treatment against its potential toxic effects and costs. Half of the patients with an unfavorable CTC profile indeed experienced early treatment failure and would have benefitted from more intense response evaluation to minimize prolonged administration of ineffective and toxic therapy. In patients with <5 CTCs, who would be classified as good prognosis according to their CTC count the 16-gene CTC profile distinguished a good from an intermediate prognosis group, providing additional information on top of a CTC count. Because the 16-gene profile is heavily influenced by the presence of epithelial markers such as cytokeratins and EpCAM, it probably identifies patients with CTCs or CTC fragments that do not meet the CellSearch criteria for morphology or marker expression [11], and could thus identify patients with false-negative CTC counts. Therefore, we propose that in these patients with <5 CTCs as determined by CellSearch, our 16-gene CTC profile better reflects the circulating tumor load.

However, validation of the 16-gene CTC profile did not result in clinically useful test characteristics. While test sensitivity was sufficient at almost 82%, a PPV of 29% means that 71% of patients predicted to have a poor prognosis will experience an average or above benefit from the chosen therapy. Despite the fact that the discovery and validation sets had equal distribution of known prognostic factors, we were concerned that unknown prognostic factors hampered our external validation. Therefore, we randomly constructed new discovery and validation sets and generated a new CTC profile. The resulting 8-gene profile had similar favorable test characteristics in the discovery set and showed higher PPV and NPV in the validation set, but with only a trend toward statistical significance.

There are several hypothetical reasons why validation was not successful. Breast cancer heterogeneity, reflected by its five intrinsic subtypes [18], is likely to lead to differing clinical value of CTCs among subtypes. Recently, it was shown that CTCs were not prognostic in patients with HER2-positive primary tumors treated with anti-HER2 therapy, in strong contrast to its prognostic value in the overall study population [5, 19]. A similar phenomenon may apply to CTC gene expression profiles. Ideally, the clinical value of CTC profiles would be assessed per tumor subtype, requiring larger patient cohorts. Also, CellSearch enriches CTCs but the resulting fractions still contain a surplus of leukocytes. For this reason, we were limited to determining only those genes that are not, or only at a very low level, expressed by these contaminating leukocytes. The incorporation of other genes might have enabled us to find a validated profile. Lastly, a prognostic profile is optimally investigated in an untreated population. As it is, however, not ethical to withhold treatment from metastatic cancer patients, we chose to investigate patients receiving a broad spectrum of first-line treatments. This heterogeneity in treatment can have impacted our prognostic profiles.

In conclusion, we show that gene expression signatures of CTCs correlate with prognosis, additional to—but also irrespective of—CTC count. While these CTC profiles could not be validated, their promising test characteristics justify further research and optimization of CTC detection assays. The ability to measure clinically relevant genes in CTCs underlines the potential of CTC characterization as a tool to improve individualized cancer treatment.

**acknowledgements**

We thank the patients for their willingness to participate and the pathologists, research nurses, and medical oncologists for...
their assistance in collecting samples and patient’s clinical FU data.

funding

This work was supported by the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (NWO); the Stichting Coolsingel (Coolsingel Foundation); and by Veridex LLC, Raritan, NJ, USA (no grant numbers).

disclosure

The authors have declared no conflicts of interest.

references