Detection of Circulating Breast Tumor Cells by Differential Expression of Marker Genes

Astrid J. Bosma, Britta Weigelt, A. Caro Lambrechts, Onno J. H. M. Verhagen, Roelof Pruntel, Augustinus A. M. Hart, Sjoerd Rodenhuis, and Laura J. van 't Veer

Division of Experimental Therapy [A. J. B., B. W., A. C. L., L. J. v. V.] and Departments of Pathology [R. P., L. J. v. V.], Radiotherapy [A. A. M. H.], and Medical Oncology [S. R.], The Netherlands Cancer Institute, 1066 CX Amsterdam, the Netherlands, and Department of Immunotherapy, Central Laboratory of Blood Transfusion and Laboratory for Experimental and Clinical Immunology, Academic Medical Center, University of Amsterdam, 1066 CX Amsterdam, the Netherlands [O. J. H. M. V.]

ABSTRACT

Purpose: We undertook a systematic approach to identify breast cancer (BC) marker genes with molecular assays and evaluated these marker genes for the detection of minimal residual disease in peripheral blood mononuclear cells (PBMCs).

Experimental Design: We used serial analysis of gene expression to identify a range of genes that were expressed in BC but absent in the expression profiles of blood and bone marrow cells. Next, we evaluated a panel of four marker genes (p1B, PS2, CK19, and EGP2) by real-time quantitative PCR in 103 PBMC samples from patients with metastatic BC (stage III/IV) and in 96 PBMC samples from healthy females.

Results: Increased marker gene expression of at least one marker was seen in 33 of 103 patients. Using quadratic discriminant analysis including all four marker genes, we determined a discriminant value with 29% positivity in the BC patient group that did not yield false positive results among the healthy females.

Conclusions: Real-time PCR for the simultaneous expression of multiple cancer-specific genes may ensure the specificity required for the clinical application of mRNA expression-based assays for occult tumor cells.

INTRODUCTION

The accurate detection of micrometastatic disease would constitute a significant advance in the staging of solid tumors. In BC, for example, micrometastatic disease may be cured by systemic therapy, as has been demonstrated convincingly by the success of adjuvant therapy (1, 2). The presence of minimal disease in bone marrow has been shown to be of predictive value (3) and has been proposed as a criterion to select node-negative patients for adjuvant chemotherapy. In addition, many patients with BC receive high-dose therapy at a time when no macroscopic disease can be found with conventional diagnostic tests. In these patients, the monitoring of the presence of minimal disease could provide valuable guidance for physicians.

The detection of minimal disease in blood or bone marrow is usually attempted with immunological methods. Evidence has been presented that these assays are of predictive value (3, 4). However, their execution is laborious and requires a considerable degree of expertise to differentiate between positively stained tumor cells and background staining. In addition, even the most experienced groups report significant false positive rates (e.g., positively stained cells in the bone marrows of healthy volunteers; Ref. 3). Much work has been done to standardize the monoclonal antibodies, antisera, and methods used, but this has not led to uniformity in methodology or to reproducibility of results in different laboratories. As a consequence, there is variation in the findings and conclusions in the literature, and no single technique has been universally adopted as clinically applicable.

We and others have developed assays to detect minimal disease that are not based on the detection of cellular epithelial proteins, but rather on the mRNA expression of genes that are silent in the constituents of peripheral blood and bone marrow (5). These methods usually involve reverse transcription-PCR or a different RNA amplification method, such as nucleic acid sequence based amplification (6). Elaborate methods have been devised to control for nonlinear amplification of the cDNA or RNA, but these have met with little success (7). The main problem of RNA-based assays continues to be the almost universally present background signal (8–15). Two recent technical developments may enable us to overcome these problems. First, a truly quantitative PCR reaction has become available, which is known as “real-time PCR” (TaqMan; Ref. 16). Second, a much higher specificity of RNA-based assays could result from the use of a panel of marker genes rather than a single gene (17). A systematic search for genes that are highly expressed in BC but not in the cellular constituents of blood and bone marrow can be achieved by techniques such as SAGE (18–20). SAGE produces a quantitative representation of all mRNAs and generates a

Received 1/17/02; revised 3/18/02; accepted 3/20/02.

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

1 Supported by Dutch Cancer Society Grant NKI97-1468.
2 To whom requests for reprints should be addressed, at Department of Pathology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands. Phone: 31-20-512-2754; Fax: 31-20-512-2759; E-mail: lveer@nki.nl.

3 The abbreviations used are: BC, breast cancer; SAGE, serial analysis of gene expression; MC, mononuclear cell; PBMC, peripheral blood mononuclear cell; EST, expressed sequence tag; NKI, Netherlands Cancer Institute; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; QDA, quadratic discriminant analysis; CK19, cytokeratin 19; IHC, immunohistochemistry.

Vol. 8, 1871–1877, June 2002
Clinical Cancer Research
so-called expression profile. Potential marker genes from BC can be selected by comparing the gene expression profile of BC tissue with that from blood or bone marrow of healthy individuals.

We have used both real-time PCR and SAGE to develop a new type of mRNA-based detection system for minimal disease in BC. Our results suggest that this technique may overcome the shortcomings of the earlier ones and is able to identify occult tumor cells in the peripheral blood in the absence of false positive results.

MATERIALS AND METHODS

SAGE. Gene expression profiles were generated from BC tissue, blood of healthy volunteers, and bone marrow of control individuals. Total RNA was isolated using RNAzolB according to the procedure of the supplier (CAMPRO Scientific, Veenendaal, the Netherlands) using 10 × 10⁻⁶ μm slides from 12 different snap-frozen breast carcinomas, from pooled bone marrow samples of 36 patients with hematological malignancies (approximately 5 × 10⁸ MCs in total), and from a pool of three buffycoats of healthy blood donors obtained from the Central Laboratory of Blood Transfusion (9 × 10⁸ PBMCs in total), yielding 690, 670, and 1100 μg of total RNA, respectively.

Approximately 600 μg of total RNA from each of these samples were used to isolate mRNA using Dynabeads Oligo(dT)₂₅ (Dynal, Oslo, Norway), from BC tissue (8.6 μg), normal bone marrow MCs (6.5 μg) and PBMCs (7.0 μg), respectively. Double-stranded cDNA was synthesized from 5 μg of mRNA using superscript II (Life Technologies, Inc., Breda, the Netherlands) and used for SAGE. The construction of the three tag libraries was performed according to the detailed SAGE protocol version 1.0 (kindly provided by Drs. Victor Velculescu and Kenneth Kinzler, Philadelphia, PA (18–20)).

Individual clones, which contain concatenated 11-bp tags (each representative of a specific transcript), were isolated from each library and used for sequence analysis with the Big Dye Terminator kit on an ABI377 automated sequencer (Applied Biosystems, Nieuwerkerk a/d IJssel, the Netherlands). For the identification of tags expressed in BC tissue but not in blood or bone marrow, the data of the BC tissue were compared with the data of blood combined with bone marrow using SAGE software version 1.01 (kindly provided by Drs. Victor Velculescu and Kenneth Kinzler). The sequences of the tags expressed exclusively in BC tissue were submitted to the National Center for Biotechnology Information databases to search for sequence homology with known genes or ESTs.

Blood Samples for Minimal Residual Disease and BC Biopsies. Blood samples were collected from 103 unselected patients with advanced BC (M₁ disease, according to the Union Internationale Contre le Cancer criteria) during a routine follow-up visit in the NKI/Antoni van Leeuwenhoek Hospital between 1997 and 1998 and from 96 healthy female volunteers who work in the NKI. Forty-four invasive BC biopsies were selected from the NKI tissue bank. All patients and volunteers gave informed consent, and the study was approved by the Medical Ethical Committee of the NKI.

Blood (3 × 8 cc) was collected in tubes containing a Ficoll-Hypaque density fluid separated by a polyester gel barrier from a sodium citrate anticoagulant (VACUTAINER CPT; Becton Dickinson, Leiden, the Netherlands). PBMCs were isolated from all these samples, and in patients with metastatic BC, a mean PBMC count of 15.7 × 10⁶ (SD, 5.6 × 10⁵) was found. In healthy individuals, a higher mean PBMC count was found (23.8 × 10⁶ (SD, 5.9 × 10⁵)).

Real-time Quantitative PCR. RNA was isolated from 6 × 10⁶ PBMCs or from 5 × 10⁻⁶ μm tissue sections made from each tumor specimen using RNazolB and resuspended in 30 μl of diethyl pyrocarbonate-treated H₂O (DEPC, Sigma, St. Louis, MO). Two μl of total RNA were used for cDNA synthesis (20 μl), as described previously (6).

The sequences of the real-time quantitative PCR primers (Isogen Bioscience, Maarsen, the Netherlands) and of the fluorescence-labeled probe (Applied Biosystems) for p1B, PS2, CK19, and EGP2 genes were selected using the primer express software (PE Biosystems; Table 1). In addition, commercially available primers and probes for the housekeeping genes, GAPDH and human transcription factor IID/TATA binding factor (HuTBP), were used (Applied Biosystems).

Serially diluted cDNA synthesized from RNA isolated from 6 × 10⁶ MCF7 cells was used to generate standard curves for control and marker gene expression. For all cDNA dilutions, the fluorescence was detected from 0–50 PCR cycles for the control and marker gene and resulted in the threshold cycle (Cₜ) value for each cDNA dilution and each target: the PCR cycle at which a significant increase in fluorescence is detected, due to the exponential accumulation of PCR products, represented in arbitrary units (TaqMan Universal PCR Master Mix Protocol; Applied Biosystems; Ref. 16). The quantities found for the GAPDH control and marker gene were used to calculate the relative quantity of control and marker gene expression in PBMCs of healthy individuals and patients with metastatic BC. The second control gene, HuTBP, was only used for confirmation of GAPDH expression.

Statistics. To optimally use the expression levels of the four marker genes (p1B, PS2, CK19, and EGP2) to separate BC patients from healthy controls, several variants of discriminant analysis were tested, including linear discriminant analysis (Fisher’s linear discriminant analysis), QDA, nonparametric kernel discriminant analysis, and nonparametric k nearest neighbors discriminant analysis (21). The predictive capability was tested using leave-one-out cross-validation (22). The QDA was shown to be the optimal method that gave the maximum number of correctly classified patients in the BC group (specificity was set to 100%).

QDA is based on the following formula capturing the predictive value of the four quantitative marker genes x, w, y, and z:

\[ C_{0000} + C_{1000}w + C_{0100}x + C_{0010}y + C_{0001}z + C_{2000}w^2 + C_{1100}xw + C_{1010}xy + C_{1001}xw + C_{0101}xy + C_{0100}x^2 + C_{0011}xy + C_{0010}xz + C_{0002}z^2 \]

Parameters are as follows: \( w = \ln(CK19/GAPDH + 0.2) \); \( x = \ln(p1B/GAPDH + 50) \); \( y = \ln(PS2/GAPDH + 0.001) \); and \( z = \ln(EGP2/GAPDH) \).
The coefficient values are $C_{0000} = 20.8$, $C_{1000} = 2.58$, $C_{0100} = -6.65$, $C_{0010} = -0.46$, $C_{0001} = -4.3$, $C_{2000} = 0.54$, $C_{1100} = -0.52$, $C_{1010} = -0.01$, $C_{1001} = 0.104$, $C_{0200} = 0.58$, $C_{1100} = 0.52$, $C_{1010} = 0.01$, $C_{1001} = 0.104$, $C_{0200} = 0.58$, and $C_{0002} = 0.45$.

The $C$ constants are derived from the data in such a way that an optimal separation between two groups is attained if the $p$-dimensional ($p$ is the number of markers) distribution of the marker values can be described by a multivariate normal distribution in both groups, with different SDs and/or correlations as well different means in the two groups. To obtain an approximately normal distribution, marker values were logarithmically transformed after adding a gene-specific constant value.

Once the values for the $C$ constants are obtained, the discriminant score can be evaluated for each subject on the basis of her marker values. The higher the score, the more likely it is that the subject is a BC patient. We then put a cutoff value on the score and predict subjects with a score below this cutoff value to be a control and subjects with a score above the cutoff value to be a BC patient. By comparing the predicted and actual status of subjects, the performance of the prediction on the basis of the score function can be evaluated.

Cytospin Preparation and Immunocytochemical Staining. The PBMCs were resuspended at $5 \times 10^6$ cells/15 ml (in 0.9% NaCl), and 21 cytospin slides/sample were made. Cells were attached to amino propyl triethoxy silane (Sigma)-coated slides using a Cytospin 3 centrifuge (Shandon, Runcorn, United Kingdom). Each slide contained two spots of $1 \times 10^5$ cells. Slides were air-dried for 30 min, fixed with acetone, and stored at $70\,^\circ C$. One slide was fixed with methanol and stained with May-Grunwald Giemsa (Merck, Darmstadt, Germany) for morphological analysis. The BC cell line MCF7 mixed with PBMCs was used as a positive and negative control for each immunostaining. Five slides ($1 \times 10^6$ cells) were thawed for staining with a monoclonal antibody specific for CK19 (RCK108; DAKO, Glostrup, Denmark) using an alkaline phosphatase procedure.

In brief, slides were preincubated with 5% goat serum for 15 min, followed by incubation with the primary antibody (1:200) in 1% PBS/BSA for 1 h at room temperature. As a negative control, the slides were incubated with 1% PBS/BSA without the primary antibody.Slides were washed with $1 \times$ PBS, fixed with 1% paraformaldehyde for 5 min, and washed with $2 \times$ PBS. Subsequently, the slides were incubated with the secondary goat antimouse antibody for 30 min (DAKO), washed with $2 \times$ PBS, incubated with the avidin-biotin-alkaline phosphatase complex (StreptABComplex/AP; DAKO) for 30 min, and washed with PBS and 0.2 M Tris-HCl (pH 8.0). The slides were exposed to chromogenic substrate solution containing 0.3 mg/ml naphthol-As-Tris-phosphate, 0.24 mg/ml levamisole, and 0.1% New Fuchsin in HCl mixed with 0.1% NaNO2 (Sigma). Slides were counterstained with hematoxylin. Cells were considered immunocytochemically positive when staining was observed of most of the cytoplasm and the cell membrane. Cells were considered tumor cells if they stained positively and if cell morphology showed features characteristic of malignancy. In addition, all CK19-stained slides were

<table>
<thead>
<tr>
<th>Marker gene</th>
<th>GenBank accession no.</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>TaqMan probe with 5'-FAM label</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1B</td>
<td>L15203</td>
<td>CTGAGAAGGTGCTGGCTGCTG</td>
<td>CTGCAAACCAGTGTGCCGTGCC</td>
<td>124–106: CTGCAAACCAGTGTGCCGTGCC</td>
</tr>
<tr>
<td>PS2</td>
<td>X00474</td>
<td>GAGGCCCAGACAGAGACGTG</td>
<td>CTGCTGTTTCGACGACACCGTTCG</td>
<td>175–198: CTGCTGTTTCGACGACACCGTTCG</td>
</tr>
<tr>
<td>CK19</td>
<td>NM002276</td>
<td>CTACAGCCACTACTACACGAC</td>
<td>CACCATTGAGAACTCCAGGATTGTCCTGC</td>
<td>378–398: CTACAGCCACTACTACACGAC</td>
</tr>
<tr>
<td>EGP2</td>
<td>M32306</td>
<td>CAGTTGGTGCACAAAATACTGTCA</td>
<td>TTGCTCAAAGCTGGCTGCCAAATGTT</td>
<td>149–172: CAGTTGGTGCACAAAATACTGTCA</td>
</tr>
</tbody>
</table>

Numbers correspond to the nucleotide position in the cDNA. All sequences are written 5'-3'.
evaluated using an automated cellular image system (Chromavision; ACIS).

**RESULTS**

**SAGE of BC, Normal Bone Marrow, and Peripheral Blood.** For the identification of genes abundantly expressed in BC tissue but not in blood or bone marrow, SAGE profiles of expressed sequences were generated from BC tissue, control bone marrow MCs, and normal PBMCs.

Sequence analysis was performed on tag libraries of 300–800-bp cloned fragments. We sequenced 560 colonies of the BC library, 700 colonies of the bone marrow MC library, and 1,100 colonies of the PBMC library, yielding DNA sequences of 14,000, 14,000, and 30,000 concatenated 11-bp tags, respectively. Sixty percent of the sequenced tags were identified by SAGE software as known genes or ESTs (8,400 BC, 8,600 MC, and 17,800 PBMC tags, respectively).

**BC-specific Tags.** To determine which genes are abundantly expressed in BC tissue and, at the same time, absent in blood and bone marrow, the tag banks from blood and bone marrow were combined to allow comparison with the tag bank of BC tissue using the SAGE software. Thus, of the 8400 BC tags identified by SAGE, 3027 were shown to be BC specific, representing 2490 unique BC tags. The abundance of each of these tags was assessed; 2215 tags (89%) have a frequency of 1, 255 tags (10%) have a frequency between 2 and 5, and 20 tags (<1%) have a frequency of ≥6. By their magnitude of difference in the SAGE, these latter 20 tags were considered good candidates for application as potential marker genes for BC cells. When these tag sequences were compared with the information in the National Center for Biotechnology Information gene and EST databases, 15 of these were associated with single known genes, such as episialin, the cytokeratins 8 and 18, two members of the collagen gene family, and three members of the apolipoprotein gene family or with a single EST (Table 2).

These 15 genes, which were exclusively and highly expressed in the BC tag bank, have been evaluated by Northern blotting for differential expression in a pool of BC RNA versus pools of either peripheral blood or bone marrow RNAs of normal volunteers (data not shown). Two of these genes, p1B and PS2 (T4 and T18, respectively), were confirmed in this independent assay to have a low background in the control mRNA pools and a strong signal in the BC mRNA pool.

**Real-time PCR of Marker Genes to Detect Circulating Tumor Cells.** For the detection of tumor cells in peripheral blood, a panel of four marker genes was selected. Two genes from our SAGE, p1B (T4) and PS2 (T18), were chosen, and a third marker gene, CK19, was chosen because this gene has been used in numerous other RNA-based assays (SAGE tag T34; Table 2; Refs. 6, 10, and 11). In addition, EGP2, a pan-carcinoma antigen, was selected, based on previously reported results (23). The expression of the panel of marker genes was determined by quantitative real-time PCR in 103 peripheral blood samples from patients with advanced BC and in peripheral blood samples from 96 healthy females. The real-time amplification plots of CK19 in a representative series of PBMCs from BC patients and healthy females are shown in Fig. 1B. For all PBMC samples, the expression level of the marker genes

#### Table 2 BC unique tags identified by SAGE

<table>
<thead>
<tr>
<th>Tag</th>
<th>Frequency of detection in BC tagbank</th>
<th>Potential marker genes identified by SAGEa</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>32</td>
<td>EST, cDNA clone</td>
<td>W72837</td>
</tr>
<tr>
<td>T2</td>
<td>30</td>
<td>Glycoprotein lacritin</td>
<td>AY005150</td>
</tr>
<tr>
<td>T3</td>
<td>21</td>
<td>2 matchesb</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>16</td>
<td>Secretory protein p1B</td>
<td>L15203</td>
</tr>
<tr>
<td>T5</td>
<td>16</td>
<td>2 matchesb</td>
<td></td>
</tr>
<tr>
<td>T6</td>
<td>10</td>
<td>Collagen α1 type 1</td>
<td>AF017178</td>
</tr>
<tr>
<td>T7</td>
<td>10</td>
<td>Matrix GLA protein</td>
<td>M58549 J05572</td>
</tr>
<tr>
<td>T8</td>
<td>9</td>
<td>Cytokeratin 8</td>
<td>M77025</td>
</tr>
<tr>
<td>T9</td>
<td>8</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>T10</td>
<td>7</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>T11</td>
<td>7</td>
<td>Episialin</td>
<td>M34088</td>
</tr>
<tr>
<td>T12</td>
<td>7</td>
<td>Keratin 7</td>
<td>BC002700</td>
</tr>
<tr>
<td>T13</td>
<td>7</td>
<td>Lectin</td>
<td>AF0077345</td>
</tr>
<tr>
<td>T14</td>
<td>7</td>
<td>Many different GenBank matches</td>
<td></td>
</tr>
<tr>
<td>T15</td>
<td>7</td>
<td>cDNA clone DKFZp564F053</td>
<td>AL049265</td>
</tr>
<tr>
<td>T16</td>
<td>7</td>
<td>Apolipoprotein C-1</td>
<td>M20902</td>
</tr>
<tr>
<td>T17</td>
<td>7</td>
<td>Mammaglobin</td>
<td>AF015224</td>
</tr>
<tr>
<td>T18</td>
<td>6</td>
<td>PS2</td>
<td>X52003</td>
</tr>
<tr>
<td>T19</td>
<td>6</td>
<td>mRNA of PC3 cell line</td>
<td>X75684</td>
</tr>
<tr>
<td>T20</td>
<td>6</td>
<td>Complement C6</td>
<td>X72188</td>
</tr>
<tr>
<td>T34</td>
<td>4</td>
<td>CK19</td>
<td>NM002276</td>
</tr>
</tbody>
</table>

a GenBank search.

b Apolipoprotein D (XM003067), ApoD-precursor (AI912925).

Bold indicates genes used for the marker panel.

d Human Bac clone GS1-542D18 from 7q31-q32 (AC002528), collagen type I α2 (COL1α2) (NM000089).

---

relative to the MCF7 standard curve (for CK19, see Fig. 1A) was calculated and corrected for the input of cDNA based on the GAPDH control (in arbitrary units; see “Materials and Methods”).

Thirty-three of the 103 BC patients revealed a positive signal for at least one marker gene that is above any signal seen for each of these marker genes separately in the healthy females. Fig. 2 shows the relative quantities for each of the four marker genes for all individual samples from the healthy females as well as the BC patients, with median expression levels for each group indicated by a horizontal line. These median expression levels were significantly higher for three marker genes in PBMCs of patients with BC than in PBMCs of healthy females as determined by the Mann-Whitney test (see Table 3). Furthermore, evaluating the range of expression levels reveals that, for all four markers, there are many patients who have a higher expression value compared with the healthy females (Fig. 2; Table 3).

Optimal Separation between BC Patients and Healthy Controls. A clinical test to determine the presence of circulating tumor cells should be sensitive and should surely avoid false positive results for each individual BC patient. Based on this consideration, we evaluated the potential of several forms of discriminant analysis (see “Materials and Methods”) to determine in what way the expression levels of each of our marker genes should be weighted to lead to zero false positives in the peripheral blood of the healthy volunteers subjects and as many true positives as possible in the data set of BC patients. The QDA gives the optimal separation, with 30 of 103 patients positive on the basis of marker gene expression in the BC group, and no false positives in the control group. However, on cross-validation, 3 of the 96 (3%) healthy females and 29 of the 103 (28%) BC patients tested positive.

Marker Gene Expression in Primary BCs. Clearly, occult tumor cells can only be detected using the marker gene panel if these tumors express one or more of these genes. We have not been able to test this directly in the BCs of our 103 patients. Instead, we tested a series of 44 primary BCs from other patients for the expression of each of the genes. The selected marker genes were expressed in all or in the large majority of the cancer specimens (Table 4).

Comparison with Immunocytochemical Staining. Cytospin preparations were made from 38 peripheral blood samples from patients with advanced BC and peripheral blood samples from 49 healthy females that were all part of our real-time PCR study. The cytospin preparations (each sample contained one million cells) were stained with monoclonal antibodies directed against CK19.

Fig. 1 Real-time PCR of the CK19 gene. A, standard curves of a real-time PCR experiment for the CK19 gene. PCR amplification curves of undiluted and 10-, 100-, 1,000-, and 10,000-fold dilutions of MCF7 cDNA synthesized from the RNA of 10^6 MCF7 cells are shown. On the Y axis, the absolute emission intensity is indicated, and on the X axis, the number of PCR cycles is indicated. The C_T values for the four MCF7 cDNA dilutions are 18.0, 21.4, 25.0, and 28.7, respectively. B, amplification plot of a real-time PCR experiment for the CK19 gene. Black curves, cDNA of PBMCs from BC patients (n = 13); gray curves, PBMCs from healthy females (n = 13).

Fig. 2 Relative quantity of expression of four marker genes in the blood of healthy females and patients. A, p1B; B, PS2; C, CK19; D, EGP2. ○, healthy females; ●, BC patients. The median expression level for each marker gene within a group is indicated by a horizontal line.
Marker Genes for Circulating Breast Tumor Cells

Unequivocal tumor cells were identified in none of the healthy females, and six samples showed some degree of staining, although this was interpreted as background signal. Two samples from the 38 BC patients contained a morphologically identifiable tumor cell without excessive numbers of positively stained cells. An additional four samples stained positive with a much higher number of stained nontumor cells compared with the background of healthy females.

The 38 samples from the BC patients were also tested by quantitative real-time PCR for expression of the four marker genes. Sixteen samples were positive based on the QDA of four marker genes. The six positive samples obtained with QDA and two of the four samples with an SD of 38, 17%). The two BC patients with morphologically certified tumor cells, as well as two of the four samples stained positive with a much higher number of stained nontumor cells compared with the background of healthy females.

Using QDA, it was possible to determine a discriminant value that retrospectively separated samples from 30 BC patients from all other samples. The increased marker gene expression was found in 30 of 103 patients with advanced BC, whereas elevated expression levels were absent in all 96 samples from healthy females. However, cross-validation indicated that this might still be a somewhat overoptimistic picture, and hence these results need to be validated in a prospective study. At this point, we cannot yet determine the true level of sensitivity of the assay. It can be anticipated that only a proportion of patients with advanced BC had circulating tumor cells at the time of sample procurement. The blood samples were drawn at convenient times when patients visited the outpatient clinic of the hospital, and many were in satisfactory remission of their BC with endocrine therapy or after chemotherapy.

We believe that high sensitivity is less important than a high degree of specificity for a clinical assay of this kind. Should circulating tumor cells prove to be of prognostic or predictive value, a positive test may eventually lead to additional or more intensive treatments, which are, unavoidably, associated with toxicity. Thus, every effort must be made to avoid false positive results, even if this compromises the sensitivity of the test. Based on experiences from our own group and from others (5, 6), it is unlikely that false positive results can be avoided if only a single marker gene is used in a reverse-transcription-PCR-based test system. Elevated expression levels of genes such as CK19 are found from time to time in the peripheral blood cells of healthy volunteers, which may be a result of a phenomenon called “illegitimate expression” (24). Using more than a single marker gene, as applied in our experiments, is a potential way to overcome this problem, assuming that there is a little chance of encountering significant illegitimate expression of more than one gene at a time.

Our experiments have focused mainly on samples from peripheral blood. Future studies will focus on bone marrow and possibly on lymph nodes, both of which could in theory have background expression levels of one or more of the evaluated genes. The observation that the background problem can be significant was shown previously for the expression of carcinoembryonic antigen, which is absent in peripheral blood but present in bone marrow (5).

A drawback of mRNA-based tests continues to be that it is difficult to accurately quantify the number of tumor cells corresponding to the mRNA levels. This can be done in assays in previous experiments by our group and others (CK19 and EGP2).

### Table 3

<table>
<thead>
<tr>
<th>Marker gene</th>
<th>Healthy females ($n = 96$)</th>
<th>Patients ($n = 103$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Range</td>
</tr>
<tr>
<td>$p1B$</td>
<td>0.0</td>
<td>–</td>
</tr>
<tr>
<td>$PS2$</td>
<td>0.0</td>
<td>–</td>
</tr>
<tr>
<td>$CK19$</td>
<td>0.7</td>
<td>0.0–9.1</td>
</tr>
<tr>
<td>$EGP2$</td>
<td>57.7</td>
<td>13.1–416.7</td>
</tr>
</tbody>
</table>

*Median and range of the expression levels were calculated relative to the MCF7 standard curve and corrected for the input cDNA based on the GAPDH housekeeping control gene.

*Man-Whitney-test. Bold indicates significant values.

### Table 4

<table>
<thead>
<tr>
<th>Marker gene</th>
<th>Positive expression in primary breast cancers ($n = 44$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p1B$</td>
<td>41 (93%)</td>
</tr>
<tr>
<td>$PS2$</td>
<td>37 (84%)</td>
</tr>
<tr>
<td>$CK19$</td>
<td>44 (100%)</td>
</tr>
<tr>
<td>$EGP2$</td>
<td>44 (100%)</td>
</tr>
</tbody>
</table>

*Marker gene expression was evaluated by real-time PCR. Positivity was defined as expression levels higher than the mean ± 2 SDs of the expression level in peripheral blood samples from healthy females.

**DISCUSSION**

To develop a sensitive detection assay for circulating BC cells, we designed a mRNA-based assay system that uses quantitative real-time PCR and four different marker genes. The four marker genes share the following characteristics: all are expressed at high levels in the large majority of primary BCs but are not expressed or are expressed only at very low levels in the cellular elements of peripheral blood or bone marrow. Two of the genes were derived from a SAGE designed to define differentially expressed genes in BC, and two had already been used in previous experiments by our group and others (CK19 and EGP2).

Using QDA, it was possible to determine a discriminant value that retrospectively separated samples from 30 BC patients from all other samples. The increased marker gene expression was found in 30 of 103 patients with advanced BC, whereas elevated expression levels were absent in all 96 samples from healthy females. However, cross-validation indicated that this might still be a somewhat overoptimistic picture, and hence these results need to be validated in a prospective study. At this point, we cannot yet determine the true level of sensitivity of the assay. It can be anticipated that only a proportion of patients with advanced BC had circulating tumor cells at the time of sample procurement. The blood samples were drawn at convenient times when patients visited the outpatient clinic of the hospital, and many were in satisfactory remission of their BC with endocrine therapy or after chemotherapy.

We believe that high sensitivity is less important than a high degree of specificity for a clinical assay of this kind. Should circulating tumor cells prove to be of prognostic or predictive value, a positive test may eventually lead to additional or more intensive treatments, which are, unavoidably, associated with toxicity. Thus, every effort must be made to avoid false positive results, even if this compromises the sensitivity of the test. Based on experiences from our own group and from others (5, 6), it is unlikely that false positive results can be avoided if only a single marker gene is used in a reverse-transcription-PCR-based test system. Elevated expression levels of genes such as CK19 are found from time to time in the peripheral blood cells of healthy volunteers, which may be a result of a phenomenon called “illegitimate expression” (24). Using more than a single marker gene, as applied in our experiments, is a potential way to overcome this problem, assuming that there is a little chance of encountering significant illegitimate expression of more than one gene at a time.

Our experiments have focused mainly on samples from peripheral blood. Future studies will focus on bone marrow and possibly on lymph nodes, both of which could in theory have background expression levels of one or more of the evaluated genes. The observation that the background problem can be significant was shown previously for the expression of carcinoembryonic antigen, which is absent in peripheral blood but present in bone marrow (5).

A drawback of mRNA-based tests continues to be that it is difficult to accurately quantify the number of tumor cells corresponding to the mRNA levels. This can be done in assays in...
which cultured breast tumor cells, such as MCF7 cells, are added to PBMC pools, but because the marker gene expression levels in MCF7 may not correspond to those in BCs occurring in patients, such an analysis is relatively unreliable. We have attempted to compare the sensitivity of the assay with that of a standard immunocytochemical assay on cytospin preparations. On the immunocytochemical assay two of the 38 BC samples were detected as positive in form of a stained single cell, whereas in the RNA-based assay 16 samples were positive. We conclude that the real-time PCR-based assay is as sensitive or more sensitive than the immunocytochemical one.

Our findings indicate that the combination of real-time PCR and a panel of suitable marker genes is able to reliably detect circulating tumor cells in patients with BC. The panel, rather than the single marker gene, decreases the likelihood of false positive results in blood from healthy volunteers. It is possible or even likely that the panel of marker genes could be chosen even more favorably: at least 56 other genes from our SAGE remain to be tested. It is hoped that this novel approach will eventually result in a standardized assay system that can be used in a routine clinical setting.

ACKNOWLEDGMENTS

We thank C. P. Schröder for supplying the EGP2 DNA sequence.

REFERENCES

Detection of Circulating Breast Tumor Cells by Differential Expression of Marker Genes

Astrid J. Bosma, Britta Weigelt, A. Caro Lambrechts, et al.


Updated version
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/8/6/1871

Cited articles
This article cites 23 articles, 9 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/8/6/1871.full.html#ref-list-1

Citing articles
This article has been cited by 15 HighWire-hosted articles. Access the articles at:
/content/8/6/1871.full.html#related-urls

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.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.