Significance, detection and markers of disseminated breast cancer cells

Marc Lacroix¹,²

¹Laboratoire Jean-Claude Heuson de Cancérologie Mammaire, Institut Jules Bordet, Bruxelles, Belgium
²InTextoResearch, 4, chemin de Hoevel, B-4837 Baelen, Wallonia, Belgium

Abstract

The development of distant metastases is the major cause of death from breast cancer. In order to predict and prevent tumour spreading, many attempts are being made to detect small numbers of tumour cells that have shed from the primary lesions and have moved to lymph nodes, blood or bone marrow. This article presents the advantages and the limitations of techniques used for disseminated tumour cells (DTC) detection. DTC markers are listed and the most currently used of them (KRT19, CEACAM5, TACSTD1, MUC1, EGFR, ERBB2, SCGB2A2, SCGB2A1, SCGB1D2, PIP, SBEM, TFF1, TFF3, ANKRD30A, SPDEF, ESR1, SERPINB5 and GABRP) are discussed, notably on the basis of recent data on breast tumour portraits (luminal epithelial-like, basal/myoepithelial-like and ERBB2). The significance of DTC for the prognosis and prediction of response to therapy is examined. DTC viability, the notion of cell dormancy and the concept of breast cancer stem cells are also discussed.

Introduction

Despite the improvements in detection and the treatment of breast cancer, about 40% of patients still succumb to the disease. The development of distant metastases is the major cause of these deaths. Breast cancer is generally no longer curable once metastases are detected by ‘classical’ means: clinical manifestations of the spread, imaging methods (such as tomography) and serum marker assays, such as those based on carcinoma antigen 15.3 (CA15.3) or carcino-embryonic antigen (CEA).

According to a longstanding hypothesis, breast cancer dissemination should involve a succession of clinical and pathological stages starting with carcinoma in situ, progressing into invasive lesion and culminating in metastatic disease. Moreover, it was thought for decades that metastasizing breast cancer cells (BCC) first disseminated to the lymph nodes (LN) before reaching peripheral blood (PB) and distant locations, including bone marrow (BM). Unfortunately, it has now became clear that metastatic spreading occurs in about 50% of cases with apparently localized breast cancer, and that up to 30% of patients with LN-negative disease will develop distant metastases within 5 years (Fisher et al. 2002, Gilbey et al. 2004, Pantel & Brakenhoff 2004, Zieglschmid et al. 2005). Therefore, recurrence is probably due to the establishment of micro-metastases before primary loco-regional treatment. That BCC seem occasionally able to shed from the primary lesion very early in the natural history of tumours, and that a direct haematogenous dissemination route is likely to exist that bypasses the lymphogenous one, strongly supports the search for techniques and tumour markers able to unambiguously identify disseminated tumour cells (DTC). This should allow evaluating the potential of these DTC in predicting the development of metastases and monitoring the response of patients to various adjuvant or neoadjuvant therapies.

Dissemination sites: lymph nodes, peripheral blood, bone marrow

Historically, the detection of DTC is most important in pathological staging of LN specimens. In the last few years, the presence of DTC in BM has also been shown to provide prognostic information. Promising detection strategies for DTC in PB are also being evaluated.
Lymph nodes (LN)

In breast cancer, the risk of metastatic disease is classically estimated by factors, such as tumour size, tumour grade, oestrogen (ESR1) and progesterone (PGR) receptor status, ploidy, ERBB2 (HER2/neu) overexpression and the number of positive axillary lymph nodes (ALN). Numerous studies have shown that the presence of DTC in ALN is the most powerful prognostic factor, being associated with significantly poor disease-free (DFS) and overall survival (OS); for instance, see Valgussa et al. 1978, International (Ludwig) Breast Cancer Study Group 1990, Cote et al. 1999, Braun et al. 2001a, Hawes et al. 2001, Pantel & Brakenhoff 2004).

During the last years, the concept of sentinel lymph node (SLN) has emerged. SLN biopsy implements mapping of the one or two LN that primarily drain the tumour (the ‘sentinel nodes’) and therefore are most likely to harbour the metastatic disease. SLN analysis is now extensively performed in breast cancer, as it can provide prognostic value with minimal-associated morbidity in contrast to complete ALN dissection. The prescreening of SLN with highly sensitive detection methods for micro-metastases thus represents a promising approach.

Considering that significant numbers of LN-negative patients develop metastatic disease, the reliability of current staging procedures to detect DTC in LN has been questioned (see ‘Techniques for DTC detection’).

Peripheral blood (PB)

PB is historically one of the most important diagnostic specimens. For instance, circulating tumour markers have been monitored in serum for years to provide indicative values about metastatic or emerging primary breast cancer. Serum markers may be good indications for tumour load, yet in most cases they fail to provide information about minimal residual disease.

Technically speaking, PB appears as an ideal source for the monitoring of DTC. Indeed, PB sampling is relatively painless and can be done at frequent intervals (for instance, to allow an assessment of the patient’s recovery or potential to develop metastases). Many groups have demonstrated the presence of DTC in PB of patients with early-stage cancer without overt metastases (for instance, see Gaforio et al. 2003, Pierga et al. 2004, Cristofanilli et al. 2005a, Müller et al. 2005, Benoy et al. 2006, Wülling et al. 2006 and the reviews of Gilbey et al. 2004, Pantel & Brakenhoff 2004, Ring et al. 2004, Zieglschmid et al. 2005).

Bone marrow (BM)

In contrast to PB sampling, BM aspiration during surgery (mostly from the medullary space of iliac crest), a site of intensive cellular exchange between blood and the mesenchymal interstitium) appears time consuming and uncomfortable for the patient. However, among the distant organs, BM is a common homing site for DTC derived from breast cancer and other primary carcinoma, even in the absence of LN metastases or clinical signs of overt distant metastases (see notably the review of Pantel & Brakenhoff 2004). In fact, the detection rate of DTC in BM from non-metastatic breast cancer patients has been reported to be in the range from 0% (Fetsch et al. 2000) to 100% (Slade et al. 2005), and this illustrates the variability of results obtained by the use of different techniques or marker genes (see ‘Techniques for DTC detection’).

In a recent, large (more than 3500 cases) study of stages I–III breast cancer patients, the incidence of DTC in BM detected by immunocytochemistry (ICC) ranged from 13 to 43% (Braun & Naume 2005).

The presence of DTC in BM may be useful not only in predicting the development of bone metastases, but also in predicting the development of metastases in other distant organs, such as lung and liver. To date, however, it remains unknown whether BM is a reservoir that allows for DTC to adapt and disseminate later into other organs, or whether the presence of DTC in BM might reflect the general propensity of these cells to disseminate and survive in organs, rather than just in the BM. Until methods are developed to detect the presence of DTC in organs, such as the lung or liver, it will not be possible to distinguish between these two possibilities. That BM could serve as a reservoir in breast cancer is supported by the presence of epithelial (cytokeratin-positive) cells in the PB of patients with overt distant metastases years after the removal of the primary tumour. This suggests that tumour cells could break from bone metastases to recirculate and disseminate to secondary tissues (Pantel & Brakenhoff 2004). This ‘two-step’ metastasis model could explain why the DTC in patients with overt metastases closely resemble each other genetically (Klein et al. 2002; see ‘Genetic alterations in DTC’).

Variability of results in DTC analysis

According to Ring et al. (2004), in studies using antibody-based (cytometric) assays, cells with the characteristics of tumour cells have been shown in the PB of between 0 and 100% of patients with operable (stages I–IIIa) breast cancer and 3–100% of
patients with metastatic disease. Studies with nucleic acid-based techniques have shown cells with the characteristics of tumour cells in the PB of 0–88% of patients with operable (stages I–IIIa) breast cancer and 0–100% of patients with metastatic disease. Along the same line, in a survey on a total of more than 3500 stages I–III breast cancer patients, the incidence of DTC in BM detected by ICC ranged from 13 to 43% (Braun & Naume 2005). In fact, the detection rate of DTC in BM from non-metastatic breast cancer patients has been reported to be in the range from 0% (Fetsch et al. 2000) to 100% (Slade et al. 2005).

The variability of results obtained in DTC detection results from drastic variations in methodology. Factors that may influence the data include:

(1) Heterogeneity of the studied populations according to the:
(a) Stage. The number of positive patients and the absolute numbers of DTC per patient rise as clinical stage rises (see notably Ben Hsieh et al. 2006).
(b) Interval of time separating surgery from the obtaining of DTC. Surgery may increase the number of breast cancer DTC (from 0 to 8000 cells/ml) in the PB, which persist for varying length of times in different patients (Hu et al. 2003).
(c) Metastasis location. The division of populations into those with early and metastatic breast cancer is probably simplistic. Moreover, metastasis sites could be missed when DTC are obtained, leading to a misclassification of the patient in the ‘early breast cancer’ category.

(2) Sample handling and preparation:
(a) Delay between collection and analysis.
(b) Conditions of sample storage.
(c) Contamination with normal epithelial cells. The introduction of skin cells into a PB sample at the time of venopuncture could lead to false-positive results. Many investigators advocate that the first few millilitres of sampled PB are discarded to avoid such contamination. It has also been suggested recently that false positivity of SLN could result from iatrogenic displacement and transport of benign epithelial cells in patients with breast carcinoma (Bleiweiss et al. 2006). Clearly, such epithelial cells do not represent metastasis.

(3) Criteria/threshold of positivity:
Number of cells analysed.
Evaluation or not of the apoptotic status of analysed cells.

(4) Analytical and preanalytical (enrichment) techniques (see ‘Techniques for DTC detection’).
(5) Markers. A number of different markers have been used. They may considerably vary the levels of sensitivity and specificity (see ‘Markers for DTC detection’).

### Techniques for DTC detection

The methods to identify DTC must distinguish between epithelial and other (mainly haematopoietic) cells. Secondarily, it may be desirable, although not necessarily essential, to distinguish between cancer and normal epithelial cells.

The most ‘conventional’ technique has been focussed on LN analysis and involves staining of sectioned tissues, previously embedded in paraffin wax, with two dyes, haematoxylin and eosin (H&E). It is likely that very small amounts of DTC present in the LN cannot be detected by this technique. An increase in sensitivity can be achieved by serial sectioning and histopathologic examination of an extensive number of sections. However, this approach is time consuming which hampers its routine application.

More sensitive approaches have been developed. Also used for LN is immunohistochemistry (IHC), using antibodies that bind to more or less specific breast cancer cell marker(s). IHC is able to detect regions of metastases in LN undetected by H&E staining (Cote et al. 1999). However, IHC has several drawbacks: it is a labour intensive and time-consuming method, particularly because at least 100 000 cells need to be analysed for a reliable assessment of the presence of tumour cells (Silva et al. 2001a). Moreover, IHC requires a trained cytologist to confirm the identity of the stained cells. Most importantly, and although IHC has been previously applied to PB and BM smears, this technique is unable to make an accurate measurement of the frequently low DTC load within PB and BM (Gilbey et al. 2004).

To identify DTC in PB and BM, the two major approaches involve additional antibody- and nucleic acid-based techniques.

### Antibody-based techniques

Approaches by fluorescence microscopy (FM), ICC and flow cytometry (FC) analysis aim to isolate and enumerate individual tumour cells. ICC is still a gold standard for DTC detection, and most of the available clinical data have been gathered by ICC screening, especially in BM (Zieglschmid et al. 2005). An
advantage of this approach is that it may allow further characterization of the cells at a molecular level, in terms of expression of key biological markers, such as \textit{ERBB2} (\textit{ERBB2} gene amplification estimated by FISH analysis) and morphological cell analysis. However, identification of intracellular targets, such as cytokeratins, by antibodies needs cell permeabilization. As a consequence, cell viability is lost, making the important discrimination of dead and viable DTC impossible. Since only viable cells might lead to metastasis, this valuable information cannot be assessed (Zieglschmid \textit{et al.} 2005).

Like IHC, FM and ICC are labour intensive and time consuming, making these techniques too expensive for routine implementation. When compared with ‘conventional’, essentially qualitative FM and ICC, FC offers the advantage of a fully automated technique allowing quantitative measurements with high sensitivity, good resolution, speed, reproducibility and statistical reliability.

For breast tumours, the most used targets for antibody-based techniques are the cytokeratins (see ‘Markers’). ERBB2, MUC1 and TACSTD1, the two latters being known under a variety of names (see Table 1), have also been used as antibody targets to isolate and/or identify DTC.

A two-colour ELISOPOT, an immunological assay based on enzyme-linked immunosorbent assay, has been recently used to detect DTC-secreting cathepsin D (CTSD) and mucin-1 (MUC1) (see Table 1; Alix-Panabières \textit{et al.} 2005).

Antibody-based techniques have limitations. Many of the antibodies directed at epithelial and breast cancer cells are known to also stain haematopoietic cells, including cytokeratins (KRT19), TACSTD1, MUC1 (see Table 1). Non-specific staining of plasma cells can also occur due to alkaline phosphatase reaction against the \(\kappa\) and \(\lambda\) light chains on the cell surface (Smerage & Hayes 2006). According to the antibody used, a false-positive detection rate of 1–3\% can be expected (Zieglschmid \textit{et al.} 2005). Since tumour and epithelial-specific cell marker antigens are expressed differentially in DTC, the use of a panel of monoclonal antibodies may help to enrich DTC and facilitate their detection, as notably shown by Hager \textit{et al.} (2005).

\section*{Nucleic acid-based techniques}

PCR, either qualitative or quantitative, has been used to identify and characterize DTC through the detection of genetic (allele-specific expression, micro-satellite instability, loss of heterozygosity) and epigenetic alterations (methylation status) that are specifically associated with cancer cells (Sidransky 1997). This includes the search for tumour-associated point mutations in oncogenes or tumour suppressors. This latter PCR approach, however, is complicated by the substantial degree of genetic variability between tumours. For instance, TP53, the gene coding for p53, is mutated in about 25\% of breast tumours, however, more than 1400 different mutations of this gene have been observed (Lacroix \textit{et al.} 2006).

Of note, PCR has been used to detect free DNA within plasma. For instance, the analysis of DNA methylation status of specific genes (\textit{ESR1}, \textit{APC}, \textit{HSD17B4}, \textit{HIC1}, \textit{RASSF1A}) in serum of breast cancer patients has been shown to be of prognostic value (Müller \textit{et al.} 2003); The PCR-based measurement of \textit{RASSF1A} methylation has been used for monitoring efficacy of adjuvant tamoxifen therapy (Fiegl \textit{et al.} 2005). However, this use of PCR is limited by poor specificity. This is due in part to the high stability of DNA in plasma when compared with mRNA (Silva \textit{et al.} 2002). As a result, it is unclear whether the free DNA that is amplified from plasma is from DTC present in plasma or if the DNA is being shed from primary tumours, metastatic tumours, or from normal tissue (Ring \textit{et al.} 2004).

To identify DNA gains and losses in single DTC, the technique of comparative genomic hybridization (CGH) is increasingly used (see notably Klein \textit{et al.} 1999, Austrup \textit{et al.} 2000, Schmidt-Kittler \textit{et al.} 2003, Schardt \textit{et al.} 2005).

Reverse transcription (RT)-PCR has been used to identify DTC through their expression of epithelial or breast cancer-associated mRNA transcripts. A list of markers that have been evaluated in DTC by RT-PCR is contained in Table 1. RT-PCR is generally more sensitive than antibody-based techniques, but has also been hampered by false positive results in samples from normal volunteers and from patients with haematological malignancies (Ring \textit{et al.} 2004). These false positives stem from multiple sources, including issues with laboratory technique, primer selection, illegitimate expression of the target genes in normal cells, the presence of pseudogenes, or contamination (see KRT19/CK19 for more details).

When using assays based on RT-PCR for detection of DTC, the balance between sensitivity and specificity must be considered. Normally, specificity decreases with the increase in sensitivity, and vice versa. One way to resolve this dilemma is to examine multiple tumour markers in samples. As mentioned below, multiplex RT-PCR assays have revealed a higher efficacy (in both sensitivity and specificity) in comparison with the
Table 1 A list of markers that have been used in mono- or multi-markers assays to detect disseminated tumour cells by antibody (KRT8, KRT18) or nucleic acid-based techniques

<table>
<thead>
<tr>
<th>Marker (gene) name</th>
<th>Gene locus</th>
<th>Standard name</th>
<th>Other frequently-used names</th>
<th>Reference(s) related to DTC detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANKRD30A</td>
<td>10p11.21</td>
<td>Ankyrin repeat domain 30A</td>
<td>Breast cancer antigen NY-BR-1; B726P</td>
<td>Backus et al. (2005), Reinholz et al. (2005), Nissan et al. (2006), Zach &amp; Lutz (2006)</td>
</tr>
<tr>
<td>B305D</td>
<td>21q11.1-q11.2</td>
<td>Antigen B305D</td>
<td>B305D, isoform A (B305D-A) B305D, isoform C (B305D-C)</td>
<td>Backus et al. (2005), Reinholz et al. (2005), Zach &amp; Lutz (2006)</td>
</tr>
<tr>
<td>CD44</td>
<td>11p13-pter</td>
<td>Antigen CD44</td>
<td>Hermes antigen; PGP1 E-cadherin; Uvomorulin</td>
<td>Gilbey et al. (2004)</td>
</tr>
<tr>
<td>CDH1</td>
<td>16q22.1</td>
<td>Cadherin-1 (epithelial)</td>
<td></td>
<td>Harigopal et al. (2005)</td>
</tr>
<tr>
<td>CEACAM5</td>
<td>19q13.2</td>
<td>Carcino-embryonic antigen-related cell adhesion molecule 5</td>
<td>Carcino-embryonic antigen (CEA)</td>
<td>Gilbey et al. (2004), Gillanders et al. (2004), Mikhitarian et al. (2005a,b)</td>
</tr>
<tr>
<td>CTSD</td>
<td>11p15.5</td>
<td>Cathepsin-D</td>
<td></td>
<td>Alix-Panabières et al. (2005), Zach &amp; Lutz (2006)</td>
</tr>
<tr>
<td>CXCR4</td>
<td>2q21</td>
<td>Chemokine, CXC motif, receptor 4</td>
<td>Neuropeptide Y receptor Y3 (NPY3R); Lipopolysaccharide-associated protein 3 (LAP3); Fusin</td>
<td>Alix-Panabières et al. (2005)</td>
</tr>
<tr>
<td>EGFR</td>
<td>7p12.3-p12.1</td>
<td>Epidermal growth factor receptor c-erbB-2</td>
<td>Oncogene ERBB1; c-erbB-1</td>
<td>Gilbey et al. (2004), Ring et al. (2004)</td>
</tr>
<tr>
<td>ERBB2</td>
<td>17q21.1</td>
<td></td>
<td>Her-2 Neu</td>
<td>Gilbey et al. (2004), Meng et al. (2004a), Ring et al. (2004), Backus et al. (2005), Smirnov et al. (2005)</td>
</tr>
<tr>
<td>GABRP</td>
<td>5q32-q33</td>
<td>γ-Aminobutyric acid type A receptor pi (π) subunit</td>
<td>GABA receptor A, pi (π) polypeptide (GABARAP); GABA-A receptor, pi (π) polypeptide (GABA A(π))</td>
<td>Backus et al. (2005), Reinholz et al. (2005), Zach &amp; Lutz (2006)</td>
</tr>
<tr>
<td>GALNT6</td>
<td>12q13</td>
<td>UDP-N-acetyl-β-D-galactosamine:polypeptide N-acetyl-β-D-galactosaminyltransferase 6</td>
<td>β-1,4-N-acetyl-galactosaminyl-transferase 6; GaINAc transferase 6; (GaINAcT6)</td>
<td>Gilbey et al. (2004), Ring et al. (2004), Freire et al. (2006)</td>
</tr>
<tr>
<td>KRT7</td>
<td>12q12-q14</td>
<td>Keratin 7</td>
<td>Cytokeratin 7 (CK7); Sarcolectin (SCL)</td>
<td>Ring et al. (2004), Masuda et al. (2005)</td>
</tr>
<tr>
<td>KRT8</td>
<td>12q13</td>
<td>Keratin 8</td>
<td>Cytokeratin 8 (CK8)</td>
<td>Ouellette et al. (2004), Ring et al. (2004), Mikhitarian et al. (2005a,b)</td>
</tr>
<tr>
<td>KRT18</td>
<td>12q13</td>
<td>Keratin 18</td>
<td>Cytokeratin 18 (CK18)</td>
<td>Ring et al. (2004), Smirnov et al. (2005)</td>
</tr>
<tr>
<td>KRT19</td>
<td>17q21-q22</td>
<td>Keratin 19</td>
<td>Cytokeratin 19 (CK19)</td>
<td>Gilbey et al. (2004), Gillanders et al. (2004), Ring et al. (2004, 2005), Weigel et al. (2004), Backus et al. (2005), Mikhitarian et al. (2005a,b), Smirnov et al. (2005), Brown et al. (2006), Nissan et al. (2006), Zach &amp; Lutz (2006)</td>
</tr>
<tr>
<td>KRT20</td>
<td>17q21.2</td>
<td>Keratin 20</td>
<td>Cytokeratin 20 (CK20) Mage</td>
<td>Gilbey et al. (2004), Kwon et al. (2005)</td>
</tr>
<tr>
<td>MAGE-A subtypes</td>
<td>Xq28</td>
<td>Melanoma antigen family A subtypes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAGEA3</td>
<td>Xq28</td>
<td>Melanoma antigen family A, 3</td>
<td>Mage3</td>
<td>Gilbey et al. (2004), Ring et al. (2004), Abdul-Rasool et al. (2006)</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Marker (gene) name</th>
<th>Gene locus</th>
<th>Standard name</th>
<th>Other frequently-used names</th>
<th>Reference(s) related to DTC detection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MET</strong></td>
<td>7q31</td>
<td>Protooncogene met</td>
<td>Hepatocyte growth factor receptor (HGFR); Renal cell carcinoma, papillary 2 (RCCP2)</td>
<td>Gilbey et al. (2004), Ring et al. (2004)</td>
</tr>
<tr>
<td><strong>MUC1</strong></td>
<td>1q21</td>
<td>Mucin-1, transmembrane</td>
<td>Carcinoma antigen 15.3 (CA15.3); Carcinoma antigen 27.29 (CA27.29); CD227 antigen; Episialin; Epithelial membrane antigen (EMA); Polymorphic epithelial mucin (PEM); Peanut-reactive urinary mucin (PUM); Tumour-associated glycoprotein 12 (TAG12)</td>
<td>Gilbey et al. (2004), Gillanders et al. (2004), Ring et al. (2004), Backus et al. (2005), Schindlbeck et al. (2005), Zach &amp; Lutz (2006)</td>
</tr>
<tr>
<td><strong>PIP</strong></td>
<td>7q32-q36</td>
<td>Prolactin-induced protein</td>
<td>Gross cystic disease fluid protein (GCDFP-15)</td>
<td>Gillanders et al. (2004), Ring et al. (2004), Backus et al. (2005), Mikhitarian et al. (2005a,b), Smirnov et al. (2005), Zach &amp; Lutz (2006)</td>
</tr>
<tr>
<td><strong>PLAUR</strong></td>
<td>19q13</td>
<td>Plasminogen activator receptor, urokinase-type</td>
<td>Urokinase receptor (uPAR); CD87 antigen</td>
<td>Pierga et al. (2005)</td>
</tr>
<tr>
<td><strong>SBEM</strong></td>
<td>12q13.2</td>
<td>Small breast epithelial mucin</td>
<td>BS106, B511S</td>
<td>Weigelt et al. (2004), Backus et al. (2005), Brown et al. (2006)</td>
</tr>
<tr>
<td><strong>SCGB1D2</strong></td>
<td>11q12.2</td>
<td>Secretoglobin family 1D member 2</td>
<td>(Prostatein-like) Lipophilin B (LPB, LPHB); Antigen BU101</td>
<td>Mikhitarian et al. (2005a,b), Brown et al. (2006)</td>
</tr>
<tr>
<td><strong>SCGB2A1</strong></td>
<td>11q12.2</td>
<td>Secretoglobin family 2A, member 1</td>
<td>Mammaglobin 2 (MGB2); Mammaglobin B (MGBB); Lacryoglobin (LGB) Lipophilin C (LPC, LPHC)</td>
<td>Aihara et al. (1999), Gillanders et al. (2004), Ouellette et al. (2004), Mikhitarian et al. (2005a,b), Smirnov et al. (2005), Nissan et al. (2006)</td>
</tr>
<tr>
<td><strong>SCGB2A2</strong></td>
<td>11q12.2</td>
<td>Secretoglobin family 2A, member 2</td>
<td>Mammaglobin (MGB); Mammaglobin 1 (MGB1); Mammaglobin A (MGBA)</td>
<td>Gilbey et al. (2004), Gillanders et al. (2004), Janku et al. (2004), Ouellette et al. (2004), Ring et al. (2004, 2005), Weigelt et al. (2004), Backus et al., (2005), Mikhitarian et al. (2005a,b), Reinholz et al. (2005), Smirnov et al. (2005), Abdul-Rasool et al. (2006), Brown et al. (2006), Zach &amp; Lutz (2006)</td>
</tr>
<tr>
<td><strong>SERPINB5</strong></td>
<td>18q21.3</td>
<td>Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5</td>
<td>Mammary serine protease inhibitor (Maspin)</td>
<td>Gilbey et al. (2004), Ring et al. (2004), Zach &amp; Lutz (2006)</td>
</tr>
<tr>
<td><strong>SPDEF</strong></td>
<td>6p21.3</td>
<td>Sterile alpha motif pointed domain-containing ets transcription factor</td>
<td>Prostate epithelium-specific Ets transcription factor (PDEF)</td>
<td>Gillanders et al. (2004), Backus et al. (2005), Mikhitarian et al. (2005a,b)</td>
</tr>
</tbody>
</table>
To improve the reliability, especially the specificity of RT-PCR assays, quantitative RT-PCR (qRT-PCR) may be used. In addition, qualitative marker information, qRT-PCR uses cut-off values of marker transcript numbers, below which transcripts can be considered as tumour cell derived. Moreover, when compared with ‘conventional’ RT-PCR, qRT-PCR relies not only on primers, but also on internal probes that specifically hybridize to the amplified sequences. In addition, due to the continuous measurement of the amplified signal, false-positive results, which could produce an abnormally shaped, non-linear amplification curve could be easily identified and removed (Zieglschmid et al. 2005).

Variations of the RT-PCR technique, such as nested RT-PCR and competitive nested RT-PCR, have also been used (for instance, see the review of Gilbey et al. 2004).

Fluorescence in situ hybridization (FISH) allows the detection of gene amplifications, for instance ERBB2 amplification in breast cancer. FISH has been used to analyse genetic aberrations in DTC in BM. Considering the importance of ERBB2 as a recent target for successful antibody-based therapy, the use of FISH to detect ERBB2 amplification in DTC appears promising (Meng et al. 2004a).

Preanalytical DTC enrichment techniques

Even in metastatic patients, the number of DTC in PB or in BM is low when compared with the surrounding normal cells. When present in PB, DTC are generally found at a frequency of one cell per $1 \times 10^5$–$10^7$ PB mononuclear cells (PBMC; Ross et al. 1993) or in number <10 DTC/ml. The frequency of DTC in cytological BM preparation from cancer patients has been estimated to be in the range of $10^{-5}$–$10^{-6}$ (Pantel et al. 1999). For most markers used by nucleic acid-based techniques, the sensitivity (one cancer cell detected among $10^7$ PBMC, in most cases) could have been overestimated when it was evaluated by in vitro spiking experiments using chosen cancer cell lines overexpressing the selected markers. Metastatic tumour cells in vivo, however, might not (or at significantly lower level) express the tested markers due to tumour heterogeneity. In addition, sequential sampling might be necessary to improve tumour cell detection since shedding into the circulation could occur intermittently.

These observations led to the development of specific methods to enrich (up to 10 000 times) the DTC population before their differentiation from other PB or BM components. DTC enrichment is usually performed through the use of density gradients.
Markers for DTC detection

During the last years, the number of single markers that have been evaluated for DTC detection, mainly by nucleic acid-based techniques, has considerably increased (see Table 1). For a detailed description of these studies, the reader is invited to consult the recent reviews of Gilbey et al. 2004, Ring et al. 2004, 2005, Zach & Lutz 2006. In the present paper, the same name will be used for the gene and the corresponding protein. For instance, despite the fact that the terms NY-BR-1 and B726P are encountered in the literature, the name of the corresponding gene, ANKRD30A, will also preferentially be used to cite the protein. SCGB2A2 will be used instead of mammaglobin, ESR1 instead of oestrogen receptor-α (ERα), etc.

An ideal marker should be universally, but uniquely expressed on all breast cancer cells. It should be easily detectable, with little variance and bear clinical relevance. Since no single-specific marker that meets these criteria has been identified, attempts are now made to develop assays with multiple tumour markers, of which some are preferably highly specific to breast tissue or breast tumours. The aim is to avoid both false-positive (detection of non-tumour cells, due to the fact that the majority of potential markers have some baseline expression in normal tissues) and false-negative (non-detection of tumour cells, due to the use of high-threshold levels for positivity) cases.

Multi-marker assays have been used by various investigators (see Table 2 and the reviews of Gilbey et al. 2004, Ring et al. 2004, 2005, Zach & Lutz 2006) and have revealed a higher efficacy (sensitivity and specificity) in comparison with the assessment of single markers. A detailed and comparative analysis of these and more recent studies, including studied material (LN, PB or BM), amplification methods, RT-PCR cycling conditions, sensitivity, specificity, single or combined positivity in samples, would deserve a specific article and will thus not be performed here.

Markers with low breast (cancer) specificity

Cytokeratins (KRTs)

Regarding epithelial tumours, the cytoskeleton components KRTs have become the markers of choice for DTC detection. They belong to a large multigene family of more than 30 known members. They are expressed at various levels and compositions in all epithelial tumours, but rarely in other tissues. For antibody-based studies, most researchers use a combination of several monoclonal antibodies that recognize various cytokeratin antigens, or a broad-spectrum anti-cytokeratin monoclonal antibody that recognizes a single epitope that is common to most cytokeratins (for more information, see the review papers of Pantel & Brakenhoff (2004) and Ring et al. (2004)).

For nucleic acid-based studies, cytokeratin 19 (KRT19) and to a lesser extent, cytokeratin 20 (KRT20) have been frequently used as markers.

KRT19, an illustration of the potential sources of false positivity in DTC detection

Due to its high sensitivity, KRT19 is the most used marker for the detection of DTC in breast cancer patients (Gilbey et al. 2004, Ring et al. 2004, Zach & Lutz 2006). Depending on the assays, KRT19 has been shown to be both a specific and a non-specific marker. In fact, KRT19 is an excellent candidate to illustrate the
Potential sources of false positivity in RT-PCR studies: illegitimate transcription, haematological disorders, the presence of pseudogenes, sample contamination.

**Illegitimate transcription.** This term describes the expression in normal tissues of small amounts of mRNA by genes that have no real physiological role in these cells. It can be expected that every promoter could be activated by ubiquitous transcription factors, which leads to an estimated expression level of one tumour marker gene transcript in 500–1000 non-tumour cells (Zieglschmid et al. 2005).

**Haematological disorders.** KRT19 expression can be induced in PB by cytokines and growth factors, which circulate at higher concentrations in inflammatory conditions and neutropenia (Ring et al. 2004). As a consequence, false-positive results are more likely under these circumstances.

**The presence of pseudogenes.** Two KRT19 pseudogenes, KRT19a and KRT19b (Savtchenko et al. 1988, Ruud et al. 1999), have been identified, which have significant sequence homology to KRT19 mRNA. Subsequently, attempts to detect the expression of the authentic KRT19 may result in the detection of either or both of these pseudogenes. To avoid pseudogene amplification, it is recommended to carefully design the primers used for RT-PCR analysis.

**Contamination.** It has been suggested that PB sampling for subsequent analysis could introduce contaminating epithelial cells expressing the KRT19 mRNA into the blood sample. Potential contamination could be minimized or prevented by discarding the first sample of blood taken.

In conclusion, KRT19 appears to be a very sensitive tumour marker, whose use, however, is often hampered by low specificity. It is helpful in detecting disseminated epithelial cells, but is not a true breast cancer marker.

**KRT20**

KRT20 is found in breast cancer cells (Bostick et al. 1998, Corradini et al. 2001, Hu & Chow 2001). However, its expression is less related to breast tissue and more related to gastric and intestinal epithelium, urothelium and Merkel cells (Zieglschmid et al. 2005). Moreover, KRT20 expression has been found in granulocytes (Jung et al. 1999). Due to its lower specificity, when compared with KRT19, the use of KRT20 is not recommended in breast cancer patients.

---

<table>
<thead>
<tr>
<th>Table 2</th>
<th>An overview of markers used in multi-marker studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Markers</strong></td>
<td><strong>Reference of the study</strong></td>
</tr>
<tr>
<td>KRT19, SERPINB5</td>
<td>Luppi et al. (1998)</td>
</tr>
<tr>
<td>CEACAM5, KRT19, KRT20, MUC1, TACSTD1</td>
<td>Bostick et al. (1998)</td>
</tr>
<tr>
<td>CD44, KRT19, MUC1</td>
<td>Eltahir et al. (1998)</td>
</tr>
<tr>
<td>CEACAM5, KRT19, SERPINB5</td>
<td>Lopez-Guerrero et al. (1999)</td>
</tr>
<tr>
<td>CEACAM5, KRT19, MUC1</td>
<td>Goeminne et al. (1999)</td>
</tr>
<tr>
<td>KRT19, EGFR, SCGB2A2</td>
<td>Grunewald et al. (2000)</td>
</tr>
<tr>
<td>CEACAM5, KRT19, MUC1</td>
<td>Berois et al. (2000)</td>
</tr>
<tr>
<td>CEACAM5, EGFR, KRT19, KRT20, MUC1, SCGB2A2, SERPINB5</td>
<td>Corradini et al. (2001)</td>
</tr>
<tr>
<td>ERBB2, SCGB2A2, SERPINB5</td>
<td>Leone et al. (2001)</td>
</tr>
<tr>
<td>KRT19, SCGB2A2</td>
<td>Silva et al. (2001b)</td>
</tr>
<tr>
<td>CGB, KRT19, KRT20</td>
<td>Hu &amp; Chow (2001)</td>
</tr>
<tr>
<td>CEACAM5, ERBB2, KRT19, MUC1, PIP, SCGB2A1, SCGB2A2</td>
<td>Berois et al. (2001)</td>
</tr>
<tr>
<td>CGB, GALNT6, MAGEA3, MET</td>
<td>Taback et al. (2001)</td>
</tr>
<tr>
<td>KRT19, TACSTD1, TFF1, TFF3</td>
<td>Bosma et al. (2002)</td>
</tr>
<tr>
<td>EGFR, KRT19, KRT20</td>
<td>Gradiolene et al. (2003)</td>
</tr>
<tr>
<td>KRT19, MUC1, PIP, SCGB2A1, SCGB2A2, SPDEF</td>
<td>Baker et al. (2003)</td>
</tr>
<tr>
<td>ANKRD30A, B305D, GABRP, SCGB2A2</td>
<td>Zehentner et al. (2004)</td>
</tr>
<tr>
<td>EGFR, KRT19, SBEM, SCGB2A2, TACSTD1, TFF1, TFF3</td>
<td>Weigelt et al. (2004)</td>
</tr>
<tr>
<td>CEACAM5, KRT19, MUC1, PIP, SCGB2A1, SCGB2A2, SPDEF</td>
<td>Gillanders et al. (2004)</td>
</tr>
<tr>
<td>KRT19, SCGB2A2, SERPINB5</td>
<td>Ferrucci et al. (2004)</td>
</tr>
<tr>
<td>CGB, KRT19, SCGB2A2</td>
<td>Fabisiewicz et al. (2004)</td>
</tr>
<tr>
<td>ANKRD30A, B305D, GABRP, KRT19, SCGB2A2</td>
<td>Reinholz et al. (2005)</td>
</tr>
<tr>
<td>KRT19, PIP, SCGB2A2</td>
<td>Ring et al. (2005)</td>
</tr>
<tr>
<td>KRT19, SBEM, SCGB1D2, SCGB2A2</td>
<td>Brown et al. (2006)</td>
</tr>
<tr>
<td>ANKRD30A, KRT19, SCGB2A1</td>
<td>Nissan et al. (2006)</td>
</tr>
<tr>
<td>SCGB2A1, SERPINB5</td>
<td>Mercatali et al. (2006)</td>
</tr>
<tr>
<td>ERBB2, KRT19, TACSTD1, TFF1, TFF3</td>
<td>Quintela-Fandino et al. (2006)</td>
</tr>
</tbody>
</table>
KRT8 and KRT18

KRT8 and KRT18 have been rarely used for DTC detection. In fact, the expression patterns of these epithelial cytokines are very similar to that of KRT19 and they are not expected to provide more specificity than this latter. Of note, KRT8, KRT18 and KRT19 are expressed in the breast epithelium, but at higher levels in the luminal than in the basal component. In view of recent observations that breast tumours may be classified into subtypes, or classes (see ‘recent data on breast cancer classification and progression’), including ‘luminal epithelial-like’ and ‘basal epithelial-like’ classes, one can speculate that these cytokeratins will be less easily detected in DTC originating from basal-like tumours.

CEACAM5

Widely known as CEA, it functions in several biological roles, including cell–cell adhesion. It is one of the most widely expressed markers in breast as well as in various other cancer cells (Gilbey et al. 2004, Ring et al. 2004, Zach & Lutz 2006). Therefore, it suffers from low specificity, as also observed with KRT19, and can similarly be induced in PB by cytokines and growth factors (Goeminne et al. 1999, Ring et al. 2004).

TACSTD1

This epithelial cell–cell adhesion protein is known under a variety of names (Table 1), of which GA733-2 and EpCAM are the most frequently used. Ubiquitously expressed on the surface of epithelial cells, it has been frequently used as target for positive IMS to enrich DTC for RT-PCR analysis (Zieglschmid et al. 2005). Monoclonal antibodies against this antigen have been extensively developed for diagnostic, but also therapeutic, approaches. Although highly sensitive for epithelial malignancies, including breast cancer, its use is, however, hampered by the fact that it is expressed in low amounts in PB cells (de Graaf et al. 1997, Bostick et al. 1998, Zhong et al. 1999).

MUC1

Mucin-1 is a very large, polymorphic and heavily glycosylated mucin. The role of mucus is primarily one of the hydrating and lubricating epithelial linings, but these proteins have also been implicated in modulating both growth factor signalling and cell adhesion. In line with this latter role, it has been suggested that MUC1 expression at the surface of tumour cells could decrease cell adhesion and favour dissemination (Ligtenberg et al. 1992). On the other hand, MUC1 could play a role in the initial attachment of breast tumour cells to tissue at distant sites, facilitating establishment of metastatic sites (Ciborowski & Finn 2002).

Widely expressed in normal epithelial tissues, MUC1 is notably present on the apical surfaces of breast, bronchial, pancreatic, uterine, salivary, intestinal and other glandular tissue cells. Like TACSTD1, MUC1 has been frequently used as target for positive IMS to enrich DTC for RT-PCR analysis (Zieglschmid et al. 2005). Several studies have reported the expression of MUC1 in a significant proportion of healthy blood donors. Indeed, MUC1 expression has been consistently found in PB cells (Zieglschmid et al. 2005). Despite this low specificity, the evaluation of MUC1 expression in DTC is supported by the increasing interest for MUC1-based immunotherapy (Emens et al. 2005).

Although MUC1 is expressed in a majority of breast tumours, its overexpression has been associated with a lower grade and a higher ER-positive phenotype (see notably Rakha et al. 2005).

EGFR

A series of RT-PCR-based mono- or multi-marker studies have evaluated the pertinence of this growth factor receptor for DTC detection (Leitzel et al. 1998, De Luca et al. 2000, Grunewald et al. 2000, Corradini et al. 2001, Gradilone et al. 2003, Weigelt et al. 2004). EGFR appears as more specific, but less sensitive than KRT19. Unfortunately, it has also been found occasionally in the PB of healthy donors (Zieglschmid et al. 2005). Moreover, Weigelt et al. (2004) have found that the median expression of EGFR was higher in normal ALN than in DTC positive ALN! Of note, EGFRvIII, a cancer-specific EGFR variant, has been recently used to detect DTC in breast cancer patients. The mutant was detected in the peripheral blood in 30% of 33 low risk, early-stage patients, 56% of 18 patients selected for neoadjuvant chemotherapy, 63.6% of 11 patients with disseminated disease and, notably, 0 of 40 control women (Silva et al. 2006).

ERBB2

Involved in growth factor signal transduction, ERBB2 plays a major role in breast tumour biology. However, it is not breast-specific (Leone et al. 2001, Mitas et al. 2001) and weak ERBB2 expression has been found in the PB of healthy women in several studies (Zieglschmid et al. 2005). However, it is over-expressed in 20–35% of breast cancer patients, mostly as a consequence of gene amplification, and this predicts for reduced survival. Moreover, in patients
with breast cancer, ERBB2 overexpression by DTC in the BM predicts poor clinical outcome (Braun et al. 2001b). This, as well as the increasing use of ERBB2 as target for immunotherapy (trastuzumab; Emens et al. 2005), supports its evaluation in DTC, at both the mRNA (RT-PCR) and the DNA (FISH) levels.

**Markers with high breast (cancer) specificity**

**The search for new markers**

Using molecular biology techniques, or combinations of techniques, various groups have identified markers specifically expressed in breast and/or breast cancer tissue or cells, when compared with normal PB, BM, or other human tissues.

For instance, genes abundantly expressed in breast cancer tissue, but absent in normal PB and BM have been identified by serial analysis of gene expression (SAGE). By order of decreasing SAGE tag frequency, these genes are SBEM, LACRT, TFF3, COL1A1, MGP, KRT8, MUC1, KRT7, CLEC5F1, IL6ST, APOC1, SCGB2A2, TFF1, TM4SF1, C6, KRT19 (Bosma et al. 2002).

A series of genes coding for secreted proteins overexpressed in breast cancer tissue when compared with corresponding normal tissue and/or other (colon, gastric, kidney, liver, lung, ovary, pancreas, prostate) normal tissues were identified by a combination of annotation/protein sequence analysis, transcript profiling, immunohistochemistry and immunoassay: HAPLN1, GFRA, SCGB1D2, CXCL10, CXCL11, COL1A1, E2F3, TRMT1, CHST2, SERHL2, ZNF324, SCGB2A2, COX6C and SCGB2A1 (Welsh et al. 2003).

Gene expression profiling was used to build a site of origin classifier in order to determine the origin of cancer of unknown primary. From an analysis of 229 primary and metastatic tumours representing 14 tumour types (breast – 34 samples, colorectal, gastric, melanoma, mesothelioma, ovarian, pancreas, prostate, renal, testicular, squamous cell carcinoma, uterine, lung), an ‘optimal’ list of 79 site-specific markers was defined. Genes related to breast-specificity were ACADSB, CCNG2, ESRI, EFHD1, GATA3, SLC39A6, MYB, SCYL3, PIK3R3, PIP, PRLR, RABEP1, TRPS1 and VAV3. Two of them, GATA3 and PIP, were identified that seemed to be strongly and relatively uniformly expressed across the range of breast tumours (Tothill et al. 2005).

Smirnov et al. (2005) obtained PB containing ≥100 DTC from one metastatic colorectal, one metastatic prostate and one metastatic breast cancer patient. In a first step, global gene expression analysis was performed on these samples and a list of cancer-specific DTC genes was obtained. Among genes distinguishing between tumour (colorectal and prostate and breast) and control patients were KRT18, KRT19, TACSTD1, TACSTD2, AGR2, TFF1 and TFF3, all genes known to be associated to the epithelial cell phenotype. Fifty-three genes distinguishing between breast tumour and controls were identified, including ESR1 and ERBB2. In a second step, PB samples immunomagnetically enriched for DTC from 74 metastatic patients (30 colorectal, 31 prostate, 13 metastatic breast cancer patients and 50 normal donors were used to confirm the DTC-specific expression of selected genes by real-time RT-PCR). The genes most restricted to breast cancer patients, when compared with normal donors, colorectal cancer and prostate cancer patients were SCGB2A1, SCGB2A2 and PIP. Two additional genes, S100A14 and S100A16, were restricted to breast and colon cancers. Of note, two genes, KRT19 and AGR2, were expressed in the majority of metastatic samples (colorectal and prostate and breast) and not in the control individuals. This confirms the interest of KRT19 as an epithelial tumour cell marker. To date, AGR2 expression has been less frequently examined (Smirnov et al. 2005).

Mikhitarian et al. (2005a) isolated RNA from a highly metastatic SCGB2A2-overexpressing ALN (only one sample). It was diluted into a pool of normal LN RNA at various ratios. Gene expression (micro-array) analysis was performed and candidate breast cancer-associated genes were then selected based on three criteria: (a) absence of expression in a pool of four normal LN; (b) a high fluorescence signal on microarray and (c) a fluorescence signal also present in the 1:50 dilution. The 34 genes identified by criteria (a), (b) and (c) were sorted by relative intensity of signal in the metastatic ALN. The ‘top15’ genes were SCGB2A2, TFF1, TFF3, KRT19, SCGB1D2, S100P, FOS, SERPINA3, ESR1, TACSTD2, JUN, PGDS, KRT8, AFP. Of note, other genes used for molecular detection of micro-metastatic disease, such as PIP, SPDEF, TACSTD1, CEACAM5 and SCGB2A1, were not present among the top15, although their signal was observed in metastatic ALN. Real-time RT-PCR analysis of pathology-negative ALN (n = 72) showed that of PIP, SCGB2A2, SPDEF, TACSTD1 and TFF1, SCGB2A2 and TFF1 had the highest apparent sensitivity for the detection of micro-metastatic breast cancer (Mikhitarian et al. 2005a).
In a micro-array approach, Backus et al. (2005) analysed RNA from samples covering normal, benign and cancerous tissues from breast, colon, lung, ovarian, prostate and peripheral blood leukocytes from healthy donors. By a combination of this micro-array testing and database/literature searching, a series of candidate breast tissue-specific markers and candidate breast cancer status markers were identified. These potential markers were then submitted to an additional multiuse selection process: some markers were excluded for one of the following reasons: (1) their expression level in white blood cells was too high; (2) their expression in breast cancer was too low and (3) their expression in lung, colon and ovarian cancers was too high. The authors finally obtained 14 markers, of which seven, ANKRD30A, GABRP, KRT19, OR4K11P, PIP, SCGB2A2 and SPDEF, were further selected (the others were CEACAM6, ERBB2, MUC1, S100A7, SCGB2A2 and TNNT1). The utility of these markers for identifying clinically actionable metastases in LN was assessed through RT-PCR analysis of SLN from 254 breast cancer patients. The investigators identified an optimal two gene-expression (KRT19 and SCGB2A2) marker set for detection of the actionable metastasis in breast SLN (Backus et al. 2005).

A series of markers with high breast (cancer) specificity

It is not possible to give here a detailed description of all the markers for which high breast (cancer) specificity has been reported. However, some of these markers emerge, since their specificity has been repeatedly underlined.

SCGB2A2

No breast cancer marker has been shown to be never expressed in healthy volunteers, but some markers are rarely found in controls. SCGB2A2 (Watson & Fleming 1996), widely known as mammaglobin, is one of these markers. It is a member of the secretoglobin superfamily (Klug et al. 2000), a group of small, secretory, rarely glycosylated, dimeric proteins mainly expressed in mucosal tissues, and that could be involved in signalling, the immune response, chemotaxis (Brown et al. 2006) and, possibly, as a carrier for steroid hormones in humans.

SCGB2A2 has become a quasi standard in breast DTC detection by RT-PCR-based methods, being the most widely studied marker after KRT19. It has been used to detect DTC in LN, PB, BM, and even in malignant effusions. SCGB2A2 expression has been detected, rarely and in low levels, in various normal tissues. This could limit its potential use as an immunotherapeutic target (Manna et al. 2003, Jaramillo et al. 2004, Narayanan et al. 2004, Viehl et al. 2005), due to concerns about autoimmune toxicity. Zafrakas et al. (2006a) have recently found an abundant SCGB2A2 expression in malignant and normal tissues of the breast and in the female genital tract, namely the cervix, uterus and ovary, while lower expression levels were rarely found in other tumours and normal tissues (Zafrakas et al. 2006a). These observations might extend the diagnostic potential of SCGB2A2 to the detection of DTC from gynaecologic malignancies.

While SCGB2A2 is considerably more breast cancerspecific than KRT19, it is less ‘universal’ among these tumours. Indeed, SCGB2A2 expression level is highly variable in breast tumours, some of them showing no expression at all. SCGB2A2 expression, evaluated at mRNA or protein level, has been reported in 61–93% of primary and/or metastatic breast cancer biopsies (Min et al. 1998, Watson et al. 1999, Houghton et al. 2001, O’Brien et al. 2002, 2005, Han et al. 2003, Span et al. 2004). By examining SCGB2A2 gene expression levels in 11 BCC lines, BT-474, Evsa-T, Hs578T, IBEP-1, IBEP-2, IBEP-3 (Siwek et al. 1998), KPL-1, MCF-7, MDA-MB-231, MDA-MB-453, T-47D, by micro-array and RT-PCR, we have found elevated SCGB2A2 mRNA level only in Evsa-T BCC, while mild expression was observed in BT-474 BCC (de Longueville et al. 2005). Of note, most of these BCC lines are of metastatic origin (Siwek et al. 1998, Lacroix & Leclercq 2004a, b).

The function of SCGB2A2 in normal breast and its possible role in breast cancer aetiology are unknown. Attempts have been made to find associations between SCGB2A2 expression and various tumour features. High SCGB2A2 expression has been associated with low-grade, steroid receptors-positive tumours from postmenopausal patients (Miksicek et al. 2002, Guan et al. 2003, Span et al. 2004). In accordance, other investigators have found an association with clinical and biological features defining a less aggressive phenotype (Núñez-Villar et al. 2003). According to Roncella et al. (2006), the lack of SCGB2A2 expression is restricted to the breast tumours with high (G3) grade. O’Brien et al. (2005) have shown that in breast tissue, SCGB2A2 exists in two main forms migrating with approximate molecular mass of 18 and 25 kDa. The high molecular weight form correlates positively with hormone receptors and negatively with tumour grade and proliferation rate (O’Brien et al. 2005).
In conclusion, SCGB2A2 has currently the highest diagnostic accuracy for the detection of metastatic breast cancer. However, although tissue specificity is the most important factor for a marker for circulating cells, sensitivity may fail. Unfortunately, the most aggressive, steroid receptor-negative, high-grade breast tumours and their corresponding DTC are likely to escape detection using SCGB2A2 as marker.

SCGB2A1
SCGB2A1 is a protein far more similar to SCGB2A2 than is to other proteins, including the other members of the secretoglobin superfamily. In breast tumours, SCGB2A1 exhibits a pattern of expression similar to that of SCGB2A2 (Becker et al. 1998). In breast cancer cell lines, SCGB2A1 is highly expressed in MDA-MB-415 BCC, as also observed for SCGB2A2 (Becker et al. 1998).

SCGB2A1 has been detected by RT-PCR in 12 out of 30 (40.0%) SLN from breast cancer patients (Nissan et al. 1998).

In addition to the mammary tissue, SCGB2A1 has been found in lachrymal and ocular glands, in prostate and in the pituitary (Lehrer et al. 1998, Sjodin et al. 2005, Xiao et al. 2005, Stoeckelhuber et al. 2006).

SCGB1D2
Lee et al. (2004) performed a large-scale analysis of mRNA coexpression based on 60 diverse large human datasets containing a total of 62.2 million expression measurements distributed among 3924 micro-arrays. These authors developed a tool (http://benzer.ubic.ca/tmm/websitedoc.html) allowing the finding of genes that are reliably coexpressed (based on the correlation of their expression profiles) in multiple datasets. Using this tool, it appears that SCGB2A1, SCGB2A2 and SCGB1D2 are frequently coexpressed and that their expression cannot be correlated to that of any other gene, including other secretoglobins. This suggests that expression of the three genes, which are localized on the same gene cluster, is probably regulated by common transcriptional mechanisms.

In accordance, a strong correlation between SCGB2A2 and SCGB1D2 levels has been observed in breast cancer. SCGB1D2 may bind to SCGB2A2 in an antiparallel manner forming a covalent tetrameric complex. The significance of this interaction is not known, however, it appears to be the predominant form of both proteins in breast cancer cells (Colpitts et al. 2001, Carter et al. 2002).

As also observed with SCGB2A2, abundant SCGB1D2 expression has been found in malignant and normal tissues of the breast and in the female genital tract, namely the cervix, uterus and ovary (Zafrakas et al. 2006a).

In summary, the secretoglobins SCGB2A1, SCGB2A2 and SCGB1D2 are expressed at variable levels in subsets of breast tumours. Despite their relatively high breast-specificity, they may also be found in several other tissues, notably glands and steroid-rich organs. Of these secretoglobins, SCGB2A2 has been the most used for DTC detection. Since SCGB2A1, SCGB2A2 and SCGB1D2 are frequently coexpressed, it is likely that, in most cases, DTC that do not express SCGB2A2 will also be negative for SCGB2A1 and SCGB1D2 expressions.

PIP
Also known as gross cystic disease fluid protein-15, PIP has been used for years to detect breast cancer and follow breast cancer progression and metastasis. It is a small protein that is considered as a highly specific and sensitive marker of apocrine differentiation (Jones et al. 2001). It has been identified in most breast cancer biopsies (Myal et al. 1998, Clark et al. 1999), in correlation with steroid receptor status. In agreement, androgens, oestrogens and glucocorticoids have been found to regulate PIP expression (Murphy et al. 1987).

However, as observed with SCGB2A1, PIP expression level may considerably vary among breast tumours, some of them showing no expression at all. By examining PIP gene expression levels in 11 BCC lines (see above for SCGB2A2), we found elevated PIP mRNA level only in MDA-MB-453 BCC, supporting the global apocrine phenotype of these cells (de Longueville et al. 2005). Therefore, PIP sensitivity in breast cancer may fail.

Despite being highly breast-specific, PIP has also been detected, although generally at low levels, in various other tissues (Mazoujian et al. 1983, Haagensen et al. 1990, Clark et al. 1999, Liu et al. 2004, Tian et al. 2004).

SBEM
Also known as BS106 (Colpitts et al. 2002). SBEM cDNA was identified based on its preferential representation in libraries prepared from normal breast tissue and breast tumours. SBEM is a small secreted mucin-like protein with strong similarity to many sialomucins (Hubé et al. 2004). In a study of 43 normal human tissues, its presence was largely restricted to the mammary and salivary glands. Regarding cancer tissues, SBEM has
been detected in breast and prostate (Miksicek et al. 2002). Among breast cancer cell lines, SBEM expression has been found in the ER-positive, well-differentiated, ‘luminal epithelial-like’ (Lacroix & Leclercq 2004a, see below ‘recent data on breast cancer classification and progression’) MCF-7, T-47D and ZR-75-1 BCC, but not in the poorly differentiated, ER-negative, ‘basal epithelial-like’ MDA-MB-231 cells (Miksicek et al. 2002).

SBEM expression was detected in > 90% of invasive ductal carcinomas, although with significant differences in expression levels, and correlated with the expression of SCGB2A2. No close correlation was found between SBEM expression and steroid receptor levels or tumour grade (Miksicek et al. 2002).

**ESR1**

Although ESR1 has not been used to detect DTC to date, it represents an essential marker of breast cancer. ESR1 is a transcription factor that allows regulatory functions of female sex steroids, mainly 17β-estradiol, on growth, differentiation and function in several target tissues, including female and male reproductive tract, mammary gland and skeletal and cardiovascular systems. Its key role in the biology and the treatment of breast cancer is well established, as well as the mechanisms underlying its activation and function (for review, see Leclercq et al. 2006). ESR1 is the main mediator of endocrine therapy (tamoxifen, SERMs, aromatase inhibitors), and its detection in tumours and individual cancer cells is thus of considerable clinical importance.

ESR1 is expressed in about two-thirds of all breast cancers. Indeed, ESR1 is the main discriminator in breast tumour classifications. Its presence is characteristic of a specific class (luminal epithelial-like, see ‘recent data on breast cancer classification and progression’) of tumours with a well-differentiated, low-grade phenotype. Significant ESR1 expression has also been found in endometroid and ovarian carcinomas.

**TFF1 and TFF3**

Both are small cysteine-rich acidic-secreted proteins containing one trefoil domain that has several conserved features, including six cysteine residues with conserved spacing.

Trefoil peptides function as ‘luminal epithelium guardians’. They are involved in protection of luminal mucosa and mucosal restitution after damage. Rapid repair of mucous epithelia is essential for preventing inflammation, which is a critical component of cancer progression (Hoffmann 2005).

Abnormal elevated TFF1 and TFF3 levels have been observed in various neoplastic diseases, including breast cancer. TFF3 is widely coexpressed with TFF1 in ER-positive malignant breast cancer cells (Poulsom et al. 1997), and both are upregulated by oestrogens. TFF3 is also induced by growth hormone.

The expression of TFF1 and TFF3 is not found in all breast tumours. Their expression pattern is close to that of ESR1 and the three genes are components of a ‘luminal epithelial’ signature defining a well-differentiated, low-grade subtype that includes about 65% of all breast cancers (see ‘recent data on breast cancer classification and progression’). Therefore, TFF1 and TFF3 may not be viewed as ‘universal’ breast tumour markers. In particular, they are unlikely to be informative in the detection of DTC from most aggressive, ER-negative, high-grade tumours.

**SPDEF**

It is a member of the ‘Ets’ family. These transcription factors regulate a number of biological processes, including cell proliferation, differentiation and invasion and are thought to play an important role in oncogenesis. Unlike the majority of Ets factors, SPDEF is expressed exclusively in tissues with a high epithelial content, such as the prostate and the breast (Oettgen et al. 2000, Ghadersohi & Sood 2001, Mitas et al. 2002). Furthermore, several studies showed SPDEF to be one of the most highly overexpressed mRNAs in human and mouse mammary tumours (Ghadersohi & Sood 2001, Mitas et al. 2002, Galang et al. 2004).

In breast cancer cells, it has been recently shown that SPDEF could cooperate with ERBB2 to promote motility and invasion. These experimental data suggest that the coevaluation of SPDEF and ERBB2 expressions of DTC could be of high prognostic value (Gunawardane et al. 2005).

**ANKRD30A**

ANKRD30A has been previously identified as NY-BR-1 (Nissan et al. 2006) or antigen B726P (Jiang et al. 2002). It was identified based on spontaneous humoral immune responses in breast cancer patients (Jäger et al. 2001, 2002). The protein is regarded as a putative transcription factor, as it contains a bipartite nuclear localization signal motif and a bZIP site (DNA-binding site followed by leucine zipper motif). Additional structural features include five tandem ankyrin repeats, implying a role for ANKRD30A in protein–protein interactions.

In view of its highly restricted expression pattern, ANKRD30A may be considered as a breast protection...
differentiation antigen that could represent a suitable target for immunotherapy (Jäger et al. 2005, Wang et al. 2006). Indeed, it was found in 80% of breast cancer specimens, while tumours of other histological types were ANKRD30A-negative. It was also identified in normal breast, normal testis, was inconsistent in prostate, and not found in other tissues (Jäger et al. 2001, 2002). ANKRD30A expression was found in 40–50% and 60–70% of primary and metastatic breast cancer specimens respectively (Zehentner et al. 2002b), which has been confirmed by other investigators (O’Brien et al. 2003). More recently, ANKRD30A expression was identified by immunohistochemistry in breast (60% of 124 invasive carcinoma lesions), but not in 23 other normal tissues, including prostate and testis, and in breast tumours, but not in lymphoma, seminoma, melanoma, kidney, ovarian, endometrial, prostate and lung cancers (Varga et al. 2006). ANKRD30A has been detected by RT-PCR in 13 out of 30 (43.3%) SLN from breast cancer patients (Nissan et al. 2006).

Therefore, although being a highly sensitive marker, ANKRD30A is not always expressed by breast cancers. Moreover, its expression has been significantly associated with the differentiation grade. For instance, in a study of 124 invasive breast carcinoma lesions, 20 out of 26 grade 1 (77%), 24 out of 38 grade 2 (63%), and 30 out of 60 grade 3 (50%) samples were positive. NYBR-1 expression was also significantly associated with LN negativity, presence of ERBB2 amplification and ER expression (Varga et al. 2006). Therefore, ANKRD30A is more likely to be detected in well-differentiated tumours and related DTC.

SERPINB5

Widely known as maspin, it is an epithelial-specific serine protease inhibitor (serpin) that shares extensive homology to the plasminogen activator inhibitors PAI-1 (SERPINE1) and PAI-2 (SERPINB2).

SERPINB5 expression has been found in the epithelium of several normal organs, including mammary gland (Zhang & Zhang 2002). In breast tissue, the presence of SERPINB5 seems to be restricted to myoepithelial cells (Maass et al. 2001, Bieche et al. 2003), when compared with the luminal epithelial ones and it has been suggested that those myoepithelial cells form a defensive barrier for the progression from ductal carcinoma in situ to more invasive carcinoma (Sternlicht et al. 1997, Polyak & Hu 2005). SERPINB5 has also been identified in tumours of various origins, including breast, although in most cases, its level was reduced when compared with normal counterparts (Pemberton et al. 1997, Zhang & Zhang 2002).

Accumulated evidence shows that SERPINB5 may act as a tumour suppressor. Its extracellular form is sufficient to inhibit tumour cell motility, extracellular matrix degradation and invasion in vitro, and inhibits tumour growth and metastasis in vivo (Zou et al. 1994, Shi et al. 2001). It also inhibits tumour-induced angiogenesis (Zhang et al. 2000). Intracellular SERPINB5 is responsible for an increased cellular sensitivity to apoptosis (Latha et al. 2005, Lockett et al. 2006).

It has been previously suggested that SERPINB5 expression in breast tumours declined with progression and that high SERPINB5 levels were associated to low aggressiveness. For instance, a significant stepwise decrease in maspin expression was shown to occur in the sequence ductal cancer in situ – invasive cancer – lymph-node metastasis (Maass et al. 2001).

According to various studies, however, SERPINB5 overexpression has been observed only in a subset (10–35%) of breast tumours (Maass et al. 2001, Umekita et al. 2002, Kim et al. 2003, Mohsin et al. 2003). In these studies, SERPINB5 levels in breast carcinomas have been directly correlated to tumour size, high grade, high S-phase fraction, aneuploidy, positive p53 status, the presence of comedo-necrosis and of lymphocyte-rich stroma, inversely correlated to the presence of steroid receptors, and identified as a strong indicator of poor prognosis, with shorter relapse-free survival (RFS) and OS (Martin et al. 2000, Umekita et al. 2002, Bieche et al. 2003, Kim et al. 2003, Mohsin et al. 2003, Umekita & Yoshida 2003). Therefore, despite its tumour suppressor function, SERPINB5 expression seems to be a characteristic of aggressive tumours, supporting its use for DTC detection.

GABRP

The γ-aminobutyric acid (GABA) receptor is a multimeric transmembrane chloride ion channel. Sixteen subtypes of GABA-receptor subunits have been categorized within five structural classes (α1–6, β1–3, γ1–3, ε, θ, π). These subunits are thought to assemble in different pentameric complexes (Hedblom & Kirkness 1997, Zafrakas et al. 2006b).

GABRP was previously identified by in silico analysis of four million ESTs as a candidate gene differentially expressed in breast cancer. It codes for the π-subunit of the GABA receptor. In a study of 23 normal human tissues, the GABRP expression level was most abundant in the breast. In breast tissue, GABRP is mainly expressed in myoepithelial/basal cells and it is hypothesized that its function could be related to tissue contractility.

GABRP expression was found to be lower in a majority of primary breast tumours when compared
with corresponding normal tissues. Along the same line, strong GABRP expression was observed in normal epithelial and benign papilloma breast cells, but no signal could be detected in invasive ductal carcinoma, suggesting that GABRP is progressively downregulated with tumour progression, and that it may be useful as a prognostic marker in breast cancer (Zafrakas et al. 2006b). In contrast, in a study of 203 invasive breast cancers, GABRP expression was found high in a subset (16%) of ER-negative, ERBB2-negative, high-grade tumours with basal-like (undifferentiated) phenotype (Symmans et al. 2005). How to explain these discrepancies?

Most in situ breast tumours are of luminal epithelial origin. They express no, or low levels of, SERPINB5 and GABRP, but are located close to the SERPINB5- and GABRP-producing normal intact basal/myoepithelial cell layers. When these tumours progress, they invade and destroy the normal basal/myoepithelial cell layers. Since tumour samples most often include some normal surrounding tissue, we suggest that this might explain why several authors, such as Zafrakas et al. (2006b) have concluded that invasive lesions expressed less SERPINB5 and GABRP than in situ tumours. On the other hand, a minority of breast tumours have a ‘basal/myoepithelial-like’ phenotype (see below) and likely originate from the transformation of normal SERPINB5- and GABRP-expressing basal/myoepithelial cells. These tumours are most often steroid-receptor negative, ERBB2 negative, have a high grade and are aggressive lesions, supporting the observations of Symmans et al. (2005).

Viability of DTC

Are most DTC precursors of clinically relevant metastases or just transiently shed cells with limited lifespan?

Clearly, tumour cells are very inefficient in causing metastasis. It has been estimated that only one in 10,000 DTC is able to establish metastatic lesions (Liotta & Stetler-Stevenson 1991). One reason is that the lifespan of many DTC circulating in PB is short. Indeed, the examination of DTC has revealed a high frequency of apoptosis (Mehes et al. 2001, Chambers et al. 2002). It may be speculated that DTC hardly survive their vigorous passage in PB.

The fraction of DTC in PB and BM that express the proliferation marker Ki-67 (absent in the G0 and early-G1 phases of the cell cycle) is small and most DTC do not proliferate at the time of primary surgery (Pantel et al. 1993, Braun & Pantel 1999, Müller et al. 2005). Therefore, many DTC escaping apoptosis are likely in a latent stage (dormant cell-cycle arrest). However, a proportion of DTC isolated from the BM are capable of clonogenic growth in vitro (Ross et al. 1993). Moreover, DTC have been obtained in up to 90% of breast cancer cases by culturing BM in standard in vitro culture medium, a percentage that was higher than the percentage of DTC directly detected in BM aspirates (Solakoglu et al. 2002, Loo et al. 2005). It is likely that DTC reaching BM are prevented to proliferate by their specific environment. In fact, most DTC appear to remain in the state of dormant cell-cycle arrest for many years; however, their persistence is associated with an unfavourable clinical outcome, suggesting that at least some of these DTC can eventually escape ‘dormancy control’ and start to expand towards an overt metastasis (Janni et al. 2001, 2005, Wiedswang et al. 2004). Studies in animal models have shown that the last step (resumed proliferation) seems to be particularly rate limiting in the formation of overt metastases (Luzzi et al. 1998). It has been suggested that when reaching BM, DTC are ‘immature’ and need alterations, possibly subtended by genetic changes to form overt metastases driven by the specific selective pressures of the bone-marrow environment (Gray 2003). At the present time, little is known about what is required for DTC to survive the vigorous passage in blood and the subsequent invasion of organs in patients.

The low proliferative activity or dormancy of individual DTC in BM of patients with non-metastatic cancer at the time of primary surgery (Pantel et al. 1993), might explain the relative resistance of these cells to conventional chemotherapy (Braun et al. 2000a, Naumov et al. 2003). This observation should lead to the development of new therapies that work equally well on proliferating and quiescent cells (e.g. immunotherapy). It must be noted that a series of recent articles discuss the mechanisms underlying cancer cell dormancy (Aguirre-Ghiso 2006, Felsher 2006, Indraccolo et al. 2006, Klein & Holzel 2006, Marches et al. 2006, Naumov et al. 2006, Ranganathan et al. 2006, Townsend & Chambers 2006, White et al. 2006).

Genetics and phenotype of DTC, when compared with primary tumours

Recent data on breast cancer classification and progression

Breast tumour classification

Recent technological advances have allowed the simultaneous evaluation of multiple RNAs (DNA micro-arrays) or proteins (tissue arrays) in tumour
samples. These studies have revealed that the breast tumours could be sorted into a very few classes characterized by the high level of expression of specific groups of genes/proteins (signatures, for instance, see Bertucci et al. 2000, Ross & Perou 2001, Lacroix et al. 2002, 2004, Callagy et al. 2003, Sorlie et al. 2003, Abd El-Rehim et al. 2005, Brenton et al. 2005, Hu et al. 2006). According to these studies, about two-thirds of tumours express features reminiscent of the luminal epithelial component of the breast. These lesions are often well differentiated, have a low grade and demonstrate relatively high levels of steroid receptors, cytokeratins KRT8, KRT18 and KRT19, BCL2, CDH1, MUC1, the transcription factors GATA3, FOXA1, XBPI (Lacroix & Leclercq 2004c), TFF1, TFF3, SLC39A6, CDKN1A, CDKN1B and CCND1. In contrast to the ‘luminal epithelial-like’ lesions, about 15% of tumours have a low level of the above-cited markers, whereas they express relatively high levels of cytokeratins KRT5 and KRT17, CDH3, EGFR, FOXC1, KIT, SERPINB5, TRIM29, GABRP, MMP7, SLPI and various proliferation markers. Most of these ‘basal/myoepithelial-like’ tumours are poorly differentiated and have a high grade (Nielsen et al. 2004). Part of them is associated with the rare medullary carcinomas (Bertucci et al. 2006) and mutations in the familial cancer susceptibility BRCA1 gene (Foulkes et al. 2003, Lacroix & Leclercq 2005, 2006). Tumours overexpressing ERBB2 as a consequence of gene amplification may be grouped into a separate class (ERBB2 subtype), more closely related to the ‘basal/myoepithelial-like’ than to the ‘luminal epithelial-like’ lesions. Of interest, the ‘luminal epithelial-like’, ‘basal/myoepithelial-like’ and ‘ERBB2’ classes are also found in breast cancer cell lines (Lacroix & Leclercq 2004a, Charafe-Jauffret et al. 2006), most of which are derived from DTC (obtained in most cases from pleural effusions).

It must be noted that among the markers listed above, many are more or less associated to a specific class. EGFR, SERPINB5 and GABRP are mostly expressed by ‘basal/myoepithelial-like’ tumours, while high ERBB2 levels are obviously expressed in lesions of the ‘ERBB2’ class. ESR1, TFF1 and TFF3, the expression of which is closely correlated, are found at high levels only in ‘luminal epithelial-like’ tumours. Other markers related to this well-differentiated, low-grade class are the secreted proteins PIP, SCGB2A1, SCGB2A2 and SCGB2D1, as well as the mucins MUC1 and SBEM, the transcription factor SPDEF and ANKRD30A.

**Stable portrait of breast cancer during progression, despite increasing genetic complexity**

The existence of breast tumour classes defined by gene/protein signatures suggests that any tumour biology reflects to a large extent the biology of the cell of origin at the time of initiation. Tumours originating from more undifferentiated epithelial cells have a rapid growth pattern and more aggressive behaviour and outcome compared with those originating in a more differentiated epithelial cells. Therefore, the ‘portrait’ of tumours seems to be stable during progression.

We previously compiled and analysed a number of data regarding breast cancer biology, pathology and genetics. We concluded that during progression to metastasis, although undergoing increasing genetic alterations, most breast tumours largely maintain their portrait (luminal epithelial-like, basal/myoepithelial-like, ERBB2). Indeed, the grade (I–III) and the expression of markers, such as ESR1, PGR, TFF1, EGFR, ERBB2, P53 and various proliferation markers, etc. are generally concordant between primaries and metastases (Lacroix et al. 2004). In fact, gene signatures underlying these portraits are preserved throughout the metastatic process of breast cancer (Weigelt et al. 2005). This opposes to the ‘classical’ view, according to which tumour progression is frequently associated with some degree of dedifferentiation (i.e. loss of ER) and supposes a deep change in the biological status of cancer cells. One consequence is that DTC are expected to express the same markers and, likely, the same properties (for instance, sensitivity or resistance to chemotherapeutic agents) than tumour cells in the corresponding primaries.

While the portrait of tumours appears stable, their progression from in situ to metastasis is accompanied by an increasing genetic complexity. This probably results from the accumulation of various minor (low-frequency) genetic or epigenetic events at many different sites of the genome, giving rise to a number of different patterns, each restricted to a small cell subpopulation. This genetic micro-heterogeneity has minor effects on the global portrait, but it eventually alter the molecular balances controlling cell adhesion, migratory ability, proteolysis, angiogenesis and, possibly, allow DTC to colonize distant organs and produce secondary tumours (Lacroix et al. 2004).

Although genetic complexity is a hallmark of breast cancer, recent studies have, however, allowed subclassifying tumours into a few categories, based on array-CGH analysis. Among breast tumours, DNA gains in chromosome 1q and loss in 16q appear to be the most frequent alterations. Some ER-positive, low-grade
tumours have very few copy number alterations in addition to gain of 1q and loss of 16q and are associated with the best patient outcome. At the other extreme of genome instability are tumours with many low level copy number aberrations. Copy number losses involving chromosomes 3p, 4, 5q, 11p, 14q, 15q, 17q and 18q are more prevalent in this group composed mainly of ER-negative, high-grade lesions from patients experiencing significantly worse outcome (Loo et al. 2004, Fridlyand et al. 2006). Fridlyand et al. (2006) have identified an additional subgroup comprised of both ER-positive and ER-negative tumours and characterized by the presence of low-level gains and losses and recurrent amplifications. The more frequently seen amplifications in this group, which occurred predominantly in the ER-positive tumours, involved 8p, including FGFR1, 11q13, including CCND1 and regions of 20q, including ZNF217.

It is well known that specific gene amplification occurs frequently in breast cancer. For instance, ERBB2, EGFR, MYC, CCND1, MDM2, NCOA3/AIB1, FGFR1, TOP2A, CTTN/EMS1, FG3, AKT2 and ZNF217 are genes for which amplification has been described in previous breast cancer studies (Albertson et al. 2003, Al-Kuraya et al. 2004). For some of them, their amplification has been linked more or less clearly to the degree of tumour aggressiveness. For instance, ERBB2 and MYC amplifications have been linked to shortened survival, while ERBB2/MYC-coamplified cancers have a worse prognosis than tumours with only one of these amplifications (Al-Kuraya et al. 2004).

Therefore, a decrease of survival is observed with increasing genome instability in primary tumours, but specific DNA gains/losses combinations as well as genes amplifications appear to have more weight in this regard.

**Genetic alterations in DTC**

There are indications that DTC may exhibit a considerable genetic diversity, reflecting the instability and micro-heterogeneity observed in primary tumours.

Using a procedure involving whole-genome amplification and subsequent CGH of single immunostained cells, it has been shown that cytokeratin-positive DTC in the BM of breast cancer patients without clinical signs of overt metastases (stage M0) were genetically heterogeneous (Klein et al. 2002). This heterogeneity was reduced with the emergence of clinically evident metastasis (stage M1). The fact that DTC in M1 patients closely resemble each other genetically suggests that cells could detach from lesions at secondary sites (e.g. BM) and recirculate, and may be cause the appearance of other metastatic sites. As mentioned earlier, it has been hypothesized that BM could serve as a ‘reservoir’ allowing for DTC to adapt and disseminate later into other organs.

Investigators using a combination of ICC and FISH found that the pattern of genetic aberrations in BM-derived DTC varied considerably among different breast cancer patients (Solakoglu et al. 2002). This is consistent with the CGH-based data of Klein et al. (2002) supporting a plethora of different random changes in M0 cells.

Schmidt-Kittler et al. (2003) also reported a high genetic heterogeneity in M0 cells, although these DTC displayed fewer chromosomal aberrations than primary tumours or cells from M1-stage patients. Numerous M0 DTC without detectable aberration (CGH analysis) were also found by these authors. In M0 cells, genetic aberrations appeared to be randomly generated, while characteristic chromosomal imbalances were observed in M1 cells. This suggests that in breast cancer, tumour cells may disseminate in a far less progressed genomic state than previously thought, and that they acquire aberrations typical of metastatic cells thereafter.

Along the same line, Gangnus et al. (2004) analysed tumour cells in BM of early-stage breast tumour patients for genomic changes by single-cell CGH. The viable disseminated cancer cells had a plethora of copy number changes in their genome. All examined cells showed chromosomal copy number changes with a substantial intercellular heterogeneity and differences to the matching primary tumours.

The further development of M0 cells into metastasis, and hence M1 cells, apparently is a matter of mutation and selection, leading to a plausible explanation for tumour dormancy. In this interpretation, dormancy reflects the time needed for M0 cells to acquire the full capacity of unrestrained growth. This selection model is in agreement with the fact that DTC in patients with overt metastases closely resemble each other genetically (Klein et al. 2002).

It must be noted that the genetic changes as observed in DTC from BM (Klein et al. 2002, Solakoglu et al. 2002, Schmidt-Kittler et al. 2003, Gangnus et al. 2004) and PB (Fehm et al. 2002) confirm the tumoural nature of these DTC.

Since specific DNA gains/losses combinations and genes amplifications in primary tumours are associated to prognosis, it would be useful to assess whether such changes are also found in DTC, as well as the possible relationships between their presence in these cells and various parameters (survival of DTC, time before clinical appearance of metastases, metastasis target organs). For instance, the prognostic value of genomic alterations in
breast DTC has been examined by Austrup et al. (2000). These authors found significant correlations between genomic alterations of the DCC and ERBB2 genes in DTC and relapse-free survival. Furthermore, increasing numbers of genomic imbalances measured in DTC were significantly associated with worse prognosis of recurrent disease.

Some of the genes that are frequently amplified in breast tumours encode proteins that are or could be targeted by specific therapies. For instance, Her-2/neu, the product of ERBB2, is targeted by the antibody trastuzumab, while attempts are made to design molecules preventing the interaction between the ubiquitin ligase MDM2 and the p53 oncogene (Lacroix et al. 2006). At term, the identification of specific gene amplifications in DTC, notably by a combination of array-CGH and FISH could allow to apply specific therapies (Bussey et al. 2006).

Phenotype of DTC

Individual phenotype comparisons between primary tumours and their corresponding DTC have not been performed as yet, to the best of our knowledge. What is clear, however, is that DTC have been observed in patients independent of the grade, the differentiation status and the ESR1 content of their tumours (for instance, see Gaforio et al. 2003, Cristofanilli et al. 2005a,b, Müller et al. 2005, Benoy et al. 2006). Therefore, the presence of DTC is not restricted to patients with aggressive ‘basal/myoepithelial’ or ‘ERBB2’ tumours, which are known to be associated with less favourable outcome. Indeed, DTC seem to distribute along a wide range of phenotypes, as judged by their variable expression of specific markers.

Since the portrait of primary tumours is generally maintained in their corresponding metastases, it can be hypothesized that DTC should also closely resemble the tumour from which they have shed. According to this, DTC from ‘luminal epithelial-like’ tumours are not expected to express high SERPINB5 or GABRP levels, while markers, such as ESR1, TFF1, TFF3 or the three secretoglobins, should not be found in DTC in patients with a ‘basal epithelial-like’ primary. As this latter class of tumours is also characterized by lower levels of KRT19, MUC1 and CEACAM5, it is suggested that preanalytical enrichment techniques based on the detection of these markers could occasionally miss the DTC originating from ‘basal epithelial-like’ lesions. In the near future, the biological characteristics of breast tumours should play a mandatory role in the choice of marker(s) for DTC detection.

DTC are expected to closely resemble their corresponding primaries. However, it has been repeatedly shown that a fraction of ERBB2-negative tumours could be associated with ERBB2-positive DTC (Braun et al. 2001b, Hayes et al. 2002, Zidan et al. 2005, Solomayer et al. 2006). These observations are in line with the fact that a significant percentage of patients with HER2-negative primary tumours develop high concentrations of serum HER2 during tumour progression (Meng et al. 2004a). Although very early ERBB2 amplification in breast tumours has been demonstrated (Schardt et al. 2005), it is possible that this amplification could, in some cases, be acquired relatively late during breast cancer progression, so that the entire primary tumour would be considered as ERBB2 negative, while part of the DTC resulting from this lesion could be ERBB2 positive. Moreover, there could be a specific preselection of ERBB2-positive cells during tumour cell dissemination. The presence of ERBB2-overexpressing cells in a tumour is associated to a higher density of micro-vessels (Sopel et al. 2005), which might favour the haematogenous dissemination of these cells. Moreover, ERBB2-overexpression was found to provide an advantage in in vitro extravasation experiments using disaggregated cells and cell clusters from primary breast cancer tissue (Roetger et al. 1998).

Significance of DTC in LN, PB and BM

Prognosis and correlations

Numerous studies have concluded that the presence of DTC in BM, evaluated by ICC or RT-PCR, correlates strongly with an early relapse of breast cancer and decreased patient survival (Landys et al. 1998, Mansi et al. 1999, Braun et al. 2000b, Gebauer et al. 2001, Gerber et al. 2001, Wiedswang et al. 2003, 2004, 2006, Schindlbeck et al. 2004, Weinschenker et al. 2004, Müller & Pantel 2005, Pantel & Woelfle 2005, Benoy et al. 2006). As demonstrated by clinical follow-up data on more than 4000 breast cancer patients studied in prospective trials by several international groups, the presence of DTC in BM (identified by ICC at primary diagnosis) predicts the postoperative occurrence of overt metastases in bone and other organs (Braun et al. 2005). Of note, strong correlations between the presence of BM micro-metastases and poor survival have been reported in breast cancer independent from LN metastases (Cote et al. 1991, Braun et al. 2000b, 2001a). That the presence of DTC in the BM of breast cancer patients does not correlate with the outspread of tumour cells into LN has been reported by other authors (for instance, see Fehm et al. 2004, Benoy et al. 2005,
Troccoli et al. (2005). For instance, in a study of 68 breast cancer patients, Benoy et al. (2005) showed that the presence or the absence of DTC in LN did not predict the DTC status of BM. Concordance between BM status and LN status was present in only 49% of all patients.

PB samples are easier to collect than BM samples, yet the prognostic significance of DTC in PB is still uncertain, as it has been less investigated than for DTC in BM. However, a series of recent works indicate that the presence of DTC in PB may be associated with bad prognosis. Using a KRT19-based RT-PCR on 128 patients with stages I and II after removal of the primary breast tumour and before adjuvant chemotherapy, Stathopoulou et al. (2002) found that patients with DTC in PB had a reduced DFS and a significant lower OS. A similar association between the presence of KRT19-positive DTC in PB and reduced DFS was noted in a study of 100 operated patients before the initiation of adjuvant chemotherapy and local radiotherapy (Giatromanolaki et al. 2004). In another series of 100 patients, the presence of CEACAM5-positive DTC in PB either before or after surgery was also indicative of a reduced DFS; patients who were DTC positive at both times had the worst prognosis (Jotsuka et al. 2004). Using an immunoctrometric method, it was found that patients with elevated DTC in PB prior to therapy had worse DFS and OS (Gaforio et al. 2003). Studies involving up to 177 patients with metastatic breast cancer demonstrated that an elevated number of DTC in PB predicted extremely short median DFS and OS, irrespective of the line of treatment, when compared with patients with low or negative DTC (Criscilofani et al. 2004, 2005a,b, Hayes et al. 2006). More recently, in 167 LN-negative patients, the presence of KRT19-positive DTC in PB before the initiation of adjuvant chemotherapy was associated with reduced DFS and OS (Xenidis et al. 2006).

Several authors have shown a significant positive correlation between the presence of DTC in PB and BM (as detected by ICC or RT-PCR), although in their studies, only the presence of DTC in BM could be clearly correlated with metastatic relapse or OS (Pierga et al. 2004, Müller et al. 2005, Benoy et al. 2006). For instance, in a recent RT-PCR study using KRT19 and SCGB2A2 as markers, Benoy et al. (2006) found that in contrast to the DTC status in BM, the presence of DTC in the PB had no impact on the OS of the patients. This suggests that the DTC that are able to find their way to the BM and survive there seem to have an increased ability to develop into overt metastases (Pantel & Brakenhoff 2004). That DTC in BM could be more informative than DTC in PB is notably suggested by a recent study, in which PB and BM were prospectively collected from 341 breast cancer patients median 40 months after operation. DTC were present in PB of 10% of the patients and in BM of 14%. DTC status in PB and BM were both significantly associated with DFS and breast cancer-specific survival. The presence of DTC in both PB and BM (8 out of 341 patients) resulted in an especially poor prognosis. However, in LN-negative patients, DTC status in BM, but not in PB, predicted differences in DFS (Wiedswang et al. 2006).

It has been suggested that the presence of DTC in PB was an indication of micro-metastasis in BM, but not in LN. For instance, in a study of 47 patients without overt metastasis, all the patients with DTC in PB had micrometastasis in BM, whereas 36% of patients with micrometastasis in BM had no circulating cells in PB; there was no correlation between DTC in PB or BM and micrometastasis in SLN (Zhu et al. 2005). The absence of correlation between DTC in LN and DTC in PB has also been shown by other investigators (see the recent papers of Stathopoulos et al. 2005, Wülffing et al. 2006).

It is important to note that, while many studies, including those that are discussed in the present article, have evaluated the clinical significance of DTC in axillary LN, the significance of micro-metastasis in SLN, especially when detected by highly sensitive techniques, such as mono- or multi-marker RT-PCR (for instance, see Mikhitarian et al. 2005b, Dell’orto et al. 2006, Hughes et al. 2006, Nissan et al. 2006) is still under investigation. To date, even quantitative RT-PCR cannot be regarded as a substitute for an extensive histopathological scrutiny of the SLN in the clinical practice.

In summary, the data currently available from DTC detection in LN, PB and BM suggest that these sites may provide partially non-redundant prognostic information. Whether these information might be complementary remains unclear (Müller and Pantel 2005).

**Monitoring the response to therapy – predictive potential of DTC**

An important potential application of DTC detection is the monitoring of therapeutic efficacy in the adjuvant setting. Indeed, the efficacy of adjuvant systemic therapy can be assessed currently only retrospectively in large-scale clinical trials following an observation period of at least 5 years. Consequently, the progress in this form of therapy is slow and is not possible to tailor therapy to an individual patient. The potential of a surrogate marker assay that would permit immediate assessment of therapy-induced effects on DTC is
Therefore evident. It could be possible to identify patients who need additional adjuvant therapy, and even to define this additional therapy, based on the DTC characteristics (e.g. the administration of trastuzumab when DTC overexpress ERBB2) or site of dissemination (e.g. biphosphonate treatment, which might eliminate tumour cells in BM persisting after adjuvant treatment).


In the study of Braun et al. (2000a), adjuvant chemotherapy was devoid of effect on the elimination of single dormant tumour cells in BM of high-risk breast cancer patients. Indeed, as most chemotherapeutic agents act by interfering with cell proliferation, they are not expected to eliminate dormant cells. This underlines the need to develop therapeutic agents active on non-proliferating cells. An example of successful immunotherapy is provided by Bozionellou et al. (2004), who were able to remove chemotherapy-resistant KRT19-positive DTC in PB and BM of breast cancer patients by administration of the anti-ERBB2 antibody trastuzumab. Pachmann et al. (2005) observed that the response of DTC in PB to chemotherapy was patient-specific. Moreover, this response faithfully reflected the response of the whole tumour to adjuvant therapy. It is well known that tumours may exhibit considerable differences in their response to specific agents. The data of Pachmann et al. (2005) are in agreement with the fact that the global characteristics of BCC, including resistances, do not change during progression from in situ to metastatic tumour (see notably Lacroix et al. 2004). That the DTC number in PB may reflect the outcome of systemic breast cancer treatment was also shown by Smith et al. (2000). Pantel & Woelfle (2005) concluded that the persistence of DTC in the BM after chemotherapy (paclitaxel–epirubicin) was an independent predictor of reduced OS. Hennessy et al. (2005) observed that in patients receiving primary chemotherapy, those achieving pathologic complete disappearance of DTC in ALN were associated with an excellent prognosis. More recently, a study of 177 patients with metastatic breast cancer has shown that the detection of elevated DTC numbers in PB at any time during therapy was an accurate indication of subsequent rapid disease progression and mortality (Hayes et al. 2006). In these patients, a change in therapy would be needed. In the study of Quintela-Fandino et al. (2006), PB micro-metastases presence after adjuvant chemotherapy was found to predict both relapse and death more powerful than classical factors (oestrogen receptor and progesterone receptor status, tumour size, age, tumour grade, number of nodes affected) in high-risk breast cancer patients undergoing high-dose chemotherapy. Micro-metastases search using a five-gene panel appeared to be a more accurate procedure than classical approaches involving only one or two genes.

Regarding adjuvant chemotherapy, it has been suggested that it could have profound and long-lasting negative effects on the BM immune system (Solo-Mayer et al. 2003). Since an increased incidence of DTC has been observed in patients with immune dysfunction (Campbell et al. 2005), it is possible that in some cases, chemotherapy could contribute to both eradicate DTC and favour their persistence in BM.

Large prospective clinical studies are now required to evaluate whether eradication of DTC in PB and BM after systemic therapy translates into a longer DFS and OS.

**A new concept: breast cancer stem cells**

In recent years, studies using animal models have suggested that only a small proportion of breast tumour cells, the so-called ‘breast cancer stem cells’ (BCSC), have the capacity for extensive proliferation and regrowth of the tumour. Indeed, these cells, which comprise between 1 and 10% of the total cell population, display the defining stem-cell properties of self-renewal and differentiation. Self-renewal drives tumourigenesis, whereas differentiation contributes to tumour phenotypic heterogeneity.

Based on these observations, the stem-cell model of carcinogenesis proposes that breast cancers originate in tissue stem or progenitor cells probably through deregulation of self-renewal pathways (including the Wnt, Notch and Hedgehog pathways, see Liu et al. 2005). This leads to expansion of this cell population which then may undergo further genetic or epigenetic changes to become fully transformed (for a recent review on tumour stem cells, see Wicha et al. 2006). Although the concept of cancer stem cells is very appealing, it must be mentioned that there remain many uncertainties, both theoretical and technical, about the interpretation of the results (see notably Hill 2006).
Regarding the dissemination process, it is possible that although either BCSC or their more differentiated progeny may be capable of forming micro-metastases, only BCSC have the self-renewal capacity to create clinically relevant macroscopic metastases. Moreover, by analogy with normal stem cells, BCSC might remain dormant at metastatic sites until they are activated by the appropriate signals from the micro-environment. Therefore, the most interesting breast DTC could well be the BCSC. This implies that techniques aiming to identify disseminated breast cancer cells through their expression of differentiation markers will be unable to recognize the most ‘dangerous’ tumour cells. BCSC seem to express CD44 and ESA/TACSTD1 (Al–Hajj et al. 2003). However, as pointed out by Hill (2006), these markers are not obviously related to stemness, but rather are features of a differentiating phenotype. Additional BCSC-specific markers remain to be defined. Recently, Glinsky et al. (2005) developed an 11-gene (ANK3, BUB1, CCNB1, CES2, FGFR2, GBX2, HCFC1, KNTC2, MKI67, RNF2, USP22) signature whose expression was regulated by the stem-cell self-renewal polycomb gene BMI1 (Liu et al. 2006). Remarkably, expression of this ‘stem-cell gene’ signature was associated with a poor prognosis for ten different types of human malignancies, including breast cancer (Glinsky et al. 2005). Future works should examine potential correlations between the expression of this ‘death signature’ by disseminated breast tumour cells and the development of clinically detectable metastases.

As discussed earlier, DTC escaping apoptosis are likely in a latent stage (dormant cell-cycle arrest). This concept of tumour cell dormancy may directly relate to stem-cell biology. Stem cells usually exist in a quiescent G0 state and self-renew only when they receive appropriate signals from their niche environment.

Recent technical advances have allowed the detection of single or small groups of breast cancer cells disseminated in LN, PB and BM, thus making visible these intermediates between primary tumours and metastases. This has notably made clear that two distinct routes may lead to tumour cell dissemination. Some cells may transit by LN before accessing the PB and BM (lymphogenous route), while other DTC appear able to directly enter the blood stream (haematogenous route). The mechanism leading to direct haematogenous tumour cell dissemination is not clearly established as yet, but it is likely favoured by a high micro-vessel density (MVD) in the primary lesion, as this latter feature has been correlated to the presence of DTC in PB or BM (Fox et al. 1997, Gerber et al. 2001, Giatromanolaki et al. 2004, Benoy et al. 2005). Antiangiogenic therapies are, therefore, expected to decrease the number of DTC in patients. Of interest, tumours overexpressing ERBB2 seem to be associated with higher MVD (Sopel et al. 2005), suggesting that cells from these tumours and ERBB2-positive cells incidentally emerging in an ERBB2-negative tumour could have a specific advantage for dissemination, at least via the haematogenous route.

The ability to detect DTC with high sensitivity and specificity, when compared with classical serum tumour markers, opens interesting clinical perspectives. Potential applications include:

- Screening of women at risk for breast cancer.
- Since tumour cells may in some cases disseminate very early in the natural history of breast cancer, one can envisage the detection of DTC in women apparently without cancer, but who are regularly screened because they are considered at high risk.
- Estimating prognosis after tumour resection.
- Evaluating the need for a therapy. At the present time, the selection of patients is based on their statistical risk of developing tumour recurrence, without knowing whether they actually harbour any DTC. This uncertainty may lead to overtreatment of patients with cancer with toxic agents that exert severe side effects. For example, only 20–25% of LN-negative breast cancer patients undergo metastatic relapse within 10 years postsurgery, but more than 90% of these patients are supposed to receive chemotherapy according to recommendations (Goldhirsch et al. 2003).

Conclusion

The age-adjusted death rate for cancer (all types) has not significantly declined over the last 50 years (Leaf 2004). Moreover, the incidence of many cancers, including breast cancer, is increasing. At the present time, the probability for a woman to develop breast cancer in Western countries is higher than 0.13 (Jemal et al. 2005). In Europe, about 130 000 women die from breast cancer each year. In most cases, death results from the dissemination of cancer cells and their proliferation at secondary sites, underlining the importance of controlling and preventing these events.
Monitoring the efficacy of a therapy. This might contribute to predict which patients with early stage or metastatic disease will recur. This may also possibly support the shift to another treatment. Monitoring for recurrence after apparently successful adjuvant therapy in patients with early stage or metastatic disease.

Destroying DTC before they develop into metastases. One can speculate that the observed moderate rate of response in advanced cancer patients might be caused by the fact that solid metastases form physiological barriers that prevent the access of macromolecules such as antibodies from the circulation in the metastatic lesion (Jain 1990). From this point of view, DTC are expected to be more easily accessible for intravenously applied immunoglobulins.

Contrasting with LN, which are often removed at surgery, PB and BM might theoretically allow repeated detection of DTC in patients. However, as BM aspiration is rather uncomfortable, PB seems to be the most suitable DTC source. One might consider consecutive evaluations of DTC in PB as a succession of regular and easily practicable real-time biopsies, which could not only be included in the future in the normal follow up of breast cancer patients, but also in the design of clinical trials using biological therapies directed against specific targets.

Before introduction of DTC detection into clinical use, much job remains to be done. To clearly establish the prognostic and predictive value of DTC, a major requirement is the standardization of detection systems and the obtainment of an agreement on threshold values. Another crucial step is the definition of optimal multi-marker assays, as no single ideal marker exists for DTC detection. The choice of markers should be based on various considerations. The biology of the primary tumour is important, as different breast cancer subtypes, associated to different gene expression patterns, have been identified. This may explain why most markers used, to date, are found at variable levels among DTC. For several of these markers, such as the secretoglobins (SCGB2A1, SCGB2A2, SCGB1D2), the amplitude of variation in gene expression level may be enormous, from the absence of expression in some tumour cells to very high mRNA levels in other DTC. In some cases, marker overexpression may be the consequence of gene amplification. Since amplification events are characteristic of tumour cells, these markers, such as ERBB2, have a high potential interest. Another important criterion for marker choice is the possibility to use the selected marker as a target for therapy. ESR1 (for endocrine therapy), MUC1, ERBB2, ANKRD30A (for immunotherapy), are examples of such markers. In the future, a big challenge will be the identification and the establishment of clinical usefulness of additional immunotherapy targets, such as NY-ESO-1 (Sugita et al. 2004, Jäger and Knuth 2005), as standard cytotoxic chemotherapy is inefficient in non-proliferating DTC.

Acknowledgements

There is no conflict of interest that would prejudice the impartiality of this article. The author was partly supported by ‘Fondation Fornarina’ and ‘SciMedWeb’. This article is dedicated to the memory of Albert Lacroix (1935–2006). Merci pour toujours, Papa!

References


International (Ludwig) Breast Cancer Study Group 1990


Leaf C 2004 Why we’re losing the war on cancer and how to win it. Avastin, Erbitux, Gleevec. The new wonder drugs might make you think we’re finally beating this dreaded scourge. We’re not. Here’s how to turn the fight around. *Fortune* **149** 76–97.


Müller V & Pantel K 2005 BM micrometastases and circulating tumor cells in breast cancer patients: where have we been, where are we now and where does the future lie? Cytotherapy 7 478–482.


Narayanan K, Jaramillo A, Benhoff ND, Campbell LG, Fleming TP, Dietz JR & Mohanakumar T 2004 Response of established human breast tumors to vaccination with mammaglobin-A cDNA. Journal of the National Cancer Institute 96 1388–1396.

Namaov GN, Townsend JL, MacDonald IC, Wilson SM, Bramwell V, Groom AC & Chambers AF 2003


Poulson R, Hanby AM, Lalani EN, Hauser F, Hoffmann W & Stamp GW 1997 Intestinal trefoil factor (TFF 3) and pS2 (TFF 1), but not spasmolytic polypeptide (TFF 2) mRNAs are co-expressed in normal, hyperplastic, and neoplastic human breast epithelium. *Journal of Pathology* **183** 30–38.


Zach O & Lutz D 2006 Tumor cell detection in peripheral blood and bone marrow. *Current Opinion in Oncology* **18** 48–56.


