A Six-Gene Signature Predicting Breast Cancer Lung Metastasis

Thomas Landemaine,² Amanda Jackson,³ Akeila Bellahcene,⁴ Nadia Rucci,⁵ Soraya Sin,¹,² Berta Martin Abad,⁶ Angels Sierra,⁶ Alain Boudinet,⁷ Jean-Marc Guinebretière,¹ Enrico Ricevuto,⁵ Catherine Noguès,²,¹,₂ Marianne Briffod,¹,² Ivan Bièche,² Pascal Chereu,¹ Teresa Garcia,³ Vincent Castronovo, Anna Teti,³ Rosette Lidereau,¹,² and Keltouma Driouch¹,²

¹Centre René Huguenin, Fédération Nationale des Centres de Lutte Contre le Cancer and ²Institut National de la Santé et de la Recherche Medeciale, U735, Saint-Cloud, France; ³Proskelia, Romainville, France; ⁴Metastasis Research Laboratory, Center of Experimental Cancer Research, University of Liège, Liège, Belgium; ⁵Department of Experimental Medicine, University of L'Aquila, L'Aquila, Italy; and ⁶Center of Molecular Oncology, Institut d'Investigació Biomèdica de Bellvitgé, L'Hospitalet de Llobregat, Barcelona, Spain.

Abstract

The lungs are a frequent target of metastatic breast cancer cells, but the underlying molecular mechanisms are unclear. All existing data were obtained either using statistical association between gene expression measurements found in primary tumors and clinical outcome, or using experimentally derived signatures from mouse tumor models. Here, we describe a distinct approach that consists of using tissue surgically resected from lung metastatic lesions and comparing their gene expression profiles with those from non-pulmonary sites, all coming from breast cancer patients. We show that the gene expression profiles of organ-specific metastatic lesions can be used to predict lung metastasis in breast cancer. We identified a set of 21 lung metastasis–associated genes. Using a cohort of 72 lymph node–negative breast cancer patients, we developed a 6-gene prognostic classifier that discriminated breast primary cancers with a significantly higher risk of lung metastasis. We then validated the predictive ability of the 6-gene signature in 3 independent cohorts of breast cancers consisting of a total of 721 patients. Finally, we show that the signature improves risk stratification independently of known standard clinical variables and a previously established lung metastasis signature based on an experimental breast cancer metastasis model. [Cancer Res 2008;68(15):6092–9]

Introduction

Metastasis is the main cause of death from breast cancer, reducing the chances of long-term survival from 90% to around 5% (1). However, breast cancer is a heterogeneous disease, with respect to many characteristics, including those associated with metastasis. Thus, patients differ widely in prognosis and survival.

Metastasis was long thought to result from a cryptic minority of tumor cells with increased metastatic capacity at the primary site (2). Studies with genome-wide microarray techniques recently raised the concept that the clinical outcome of breast cancer patients could be predicted by gene expression signatures present in the primary tumors at the time of diagnosis. Several studies have identified gene expression profiles associated with the risk of distant metastasis (3). Transcriptome analysis of distant metastases further supported this concept. Distant metastases and paired primary breast tumors showed extensive genetic similarities, and a “metastatic signature” identified in adenocarcinoma metastases was also found in a subset of primary tumors (4–6). However, the presence of a site-specific metastatic expression profile has not been tackled in these studies.

Breast cancer, such as other cancers, tends to metastasize more frequently to certain organs. In the case of breast, the sites most frequently colonized are bone and lungs, and in a lower extent, liver and brain (2). The organ specificity has mainly been investigated in animal models (7, 8). In particular, Massague’s group (9–11) developed an experimental system based on the in vivo selection of MDA-MB-231–derived breast cancer cell lines with specific organotropism. The gene expression analysis of these cell lines allowed the identification of genes mediating metastasis to bone or lungs. The genes related to lung metastasis showed very little overlap with the list of genes identified in the bone metastasis signature, suggesting the involvement of different organ-specific pathways. Importantly, in a cohort of human breast cancer primary tumors, those expressing the lung metastasis signature had a significantly poorer lung metastasis–free survival but not a bone metastasis–free survival. The “lung metastasis signature” was found to be predictive of a high risk for developing lung metastases in three independent series of breast tumors (10–12).

In the present study, we used a strategy based on the use of metastatic human samples rather than human cell clones selected in mice to identify genes involved in breast cancer metastasis to the lungs. We hypothesized that as primary tumors are highly heterogeneous in terms of both their cell populations and their ability to metastasize, the genes responsible of the organotropism might not score using the classic approach that consists of identifying such genes in bulk primary tumor samples. Therefore, we searched for such genes directly by profiling metastatic samples from different secondary sites.

We analyzed the transcriptome of 23 human breast cancer metastases excised from various anatomic locations, including the lungs. We searched for genes differentially expressed between lung metastases and nonlung metastases to eliminate potential common breast tumorogenesis–related genes. We thereby identified a set of 21 differentially expressed genes. From this list, we identified a 6-gene signature that correlated with a significantly increased risk of lung metastasis in a series of 72 lymph node–negative breast
tumors. This signature was then validated in 3 independent series of breast cancer patients \((n = 721)\). Finally, we addressed the issue of what additional predictive information was gained with the six-gene predictor beyond known standard clinical parameters and the previously established lung metastasis signature based on an experimental model \((11, 12)\).

### Materials and Methods

**Patients and samples.** The study was performed according to the local ethical regulations. We first studied the transcriptome of 23 metastases (5 lung, 6 liver, 4 brain, 2 skin, and 6 osteolytic bone metastases) from breast cancer patients that undergone surgery. Ten additional samples were used for reverse transcription-PCR (RT-PCR) validation (three lung, two liver, four skin, and one bone metastases). All metastatic samples were obtained from the University of L'Aquila, Institut d'Investigacio ´B i o m e`dica de Bellvitge (IDIBELL), and Centre René Huguenin. In addition, RNA pools were prepared for normal bone and brain tissues [from six bone samples (L'Aquila)] and four brain samples (IDIBELL), respectively. RNA pools from normal lung and normal liver (at least five samples in each cases) were purchased from Biochain and Invitrogen.

A series of 72 primary breast tumors (“CHRI cohort”) was specifically selected from patients with node-negative breast cancer treated at the Centre René Huguenin, and who did not receive systemic neoadjuvant or adjuvant therapy (median follow-up of 132 mo; range, 22.6–294 mo). During 10 y of follow-up, 38 patients developed distant metastases. Eleven of these patients developed lung metastases as the first site of distant relapse.

We also analyzed 3 independent breast tumor series: the “MSK” \((n = 82)\), “EMC” \((n = 34)\), and “NKI” \((n = 295)\) cohorts described in detail elsewhere \((10, 12–14)\) for which microarray data were freely downloaded from the National Center for Biotechnology Information Web site. Briefly, EMC and NKI cohorts consist of early stage breast cancers, whereas MSK series is consisting of locally advanced tumors.

Finally, paraffin-embedded sections of 13 primary breast tumors and paired lung and/or nonlung metastases were obtained from Liège University and from Centre René Huguenin to perform immunohistologic analyses.

**Gene expression analysis.** For microarray analysis, sample labeling, hybridization, and staining were carried out as previously described \((15)\). Human Genome U133 Plus 2.0 arrays were scanned in an Affymetrix GeneChip scanner 3.000 at 570 nm. The 5’/3’ glyceraldehyde-3-phosphate dehydrogenase ratios averaged 3.5. The average background, noise average, % Present calls, and average signal of Present calls were similar with all the arrays used in this experiment.

For quantitative RT-PCR analysis, we used cDNA synthesis and PCR conditions described in detail elsewhere \((16)\). All PCR reactions were performed with an ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems) and the SYBR Green PCR Core Reagents kit (Perkin-Elmer Applied Biosystems). TATA-box-binding protein \((TBP)\) transcripts were used as an endogenous RNA control, and each sample was normalized on the basis of its \(TBP\) content \((16)\).

**Immunohistochemistry.** Human biopsy specimens of primary breast tumors and matched metastases were deparaffinized, treated with 3%

### Table 1. Lung metastasis associated genes obtained from a class comparison of lung \((n = 5)\) and nonlung metastases of breast cancer \((n = 18)\)

<table>
<thead>
<tr>
<th>Probe set</th>
<th>Gene symbol</th>
<th>Gene title</th>
<th>Function/biological process</th>
<th>Fold change</th>
<th>Parametric (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>204731_x_at</td>
<td>DSC2*</td>
<td>Desmocollin 2</td>
<td>Cell adhesion</td>
<td>8.2</td>
<td>(P &lt; 0.000001)</td>
</tr>
<tr>
<td>223861_at</td>
<td>HOMA1*</td>
<td>HORMA domain containing 1</td>
<td>Rho protein signal transduction</td>
<td>21.5</td>
<td>2.0E-06</td>
</tr>
<tr>
<td>228171_s_at</td>
<td>PLEKH4*</td>
<td>Pleckstrin homology domain containing, family G, member 4</td>
<td></td>
<td>2.7</td>
<td>2.0E-06</td>
</tr>
<tr>
<td>228577_x_at</td>
<td>ODF2L*</td>
<td>Outer dense fiber of sperm tails 2 like</td>
<td>Regulation of transcription</td>
<td>2.9</td>
<td>8.0E-06</td>
</tr>
<tr>
<td>220941_s_at</td>
<td>C21orf91*</td>
<td>Chromosome 21 open reading frame 91</td>
<td></td>
<td>3.6</td>
<td>1.0E-05</td>
</tr>
<tr>
<td>227642_at</td>
<td>TFCP2L2</td>
<td>Transcription factor CP2 like 1</td>
<td></td>
<td>5.5</td>
<td>1.0E-05</td>
</tr>
<tr>
<td>219867_at</td>
<td>CHDOL*</td>
<td>Chondrolectin</td>
<td>Hyaluronic acid binding</td>
<td>2.8</td>
<td>1.0E-05</td>
</tr>
<tr>
<td>208428_s_at</td>
<td>CALB2*</td>
<td>Calbindin 2 (calretinin)</td>
<td>Calcium ion binding</td>
<td>9.2</td>
<td>1.6E-05</td>
</tr>
<tr>
<td>228956_at</td>
<td>UGT8*</td>
<td>UDP glycosyltransferase 8</td>
<td>Glycosphingolipid biosynthetic process</td>
<td>15.4</td>
<td>1.8E-05</td>
</tr>
<tr>
<td>155426_at</td>
<td>C1orf210*</td>
<td>Chromosome 1 open reading frame 210</td>
<td></td>
<td>2.3</td>
<td>2.1E-05</td>
</tr>
<tr>
<td>221705_s_at</td>
<td>SIK*</td>
<td>Suppressor of IKK (\epsilon)</td>
<td></td>
<td>1.9</td>
<td>2.1E-05</td>
</tr>
<tr>
<td>211488_s_at</td>
<td>ITGB8*</td>
<td>Integrin, (\beta) 8</td>
<td>Cell adhesion/Signal transduction</td>
<td>1.7</td>
<td>2.6E-05</td>
</tr>
<tr>
<td>213372_at</td>
<td>PAQR3*</td>
<td>Progestin and adipoQ receptor family member III</td>
<td>Receptor activity</td>
<td>5.6</td>
<td>2.9E-05</td>
</tr>
<tr>
<td>208103_s_at</td>
<td>ANP32E*</td>
<td>Acidic (leucine-rich) nuclear phosphoprotein 32 family, member E</td>
<td>Phosphatase inhibitor activity</td>
<td>6.0</td>
<td>3.2E-05</td>
</tr>
<tr>
<td>60474_at</td>
<td>FERMT1*</td>
<td>Fermitin family homologue 1 (drosophila)</td>
<td>Cell adhesion</td>
<td>6.9</td>
<td>3.6E-05</td>
</tr>
<tr>
<td>222869_s_at</td>
<td>ELAC1*</td>
<td>elaC homologue 1 (Escherichia coli)</td>
<td>RNA processing</td>
<td>1.6</td>
<td>3.6E-05</td>
</tr>
<tr>
<td>227829_at</td>
<td>GYLT1L</td>
<td>Glycosyltransferase-like 1B</td>
<td>Glycosphingolipid biosynthetic process</td>
<td>2.6</td>
<td>3.9E-05</td>
</tr>
<tr>
<td>226075_at</td>
<td>SPNS1</td>
<td>spla/ryanodine receptor domain and SOCS box containing 1</td>
<td>Intracellular signaling cascade</td>
<td>2.9</td>
<td>5.6E-05</td>
</tr>
<tr>
<td>155370_at</td>
<td>CHRM3</td>
<td>Cholinergic receptor, muscarinic 3</td>
<td>G-protein coupled signal transduction</td>
<td>2.0</td>
<td>6.1E-05</td>
</tr>
<tr>
<td>225363_at</td>
<td>PTEN</td>
<td>Phosphatase and tensin homologue</td>
<td></td>
<td>0.4</td>
<td>6.2E-05</td>
</tr>
<tr>
<td>203256_at</td>
<td>CDH3</td>
<td>Cadherin 3, type 1, P-cadherin (placental)</td>
<td>Cell adhesion</td>
<td>6.1</td>
<td>9.5E-05</td>
</tr>
</tbody>
</table>

*Genes tested by qRT-PCR on a larger series of metastases.
H₂O₂, and incubated with the primary antibodies against FERMT1 (Abcam) or DSC2 (Progen). The staining signals were revealed with the Ultra-Vision Detection System Anti-Polyvalent horseradish peroxidase/3,3'-diaminobenzidine kit (Lab Vision). The slides were counterstained with Mayer's hematoxylin.

Statistical analysis. Microarray expression profiles were analyzed with BRB Array tools, version 3.3, developed by Dr. R. Simon and Dr. A. Peng Lam. Expression data were collated as CEL files, and normalization was done using the RMA function of Bioconductor. Univariate t tests were used to identify genes differentially expressed between 5 lung metastases and 18 nonlung metastases. Differences were considered statistically significant if the P value was <0.0001. This stringent threshold was used to limit the number of false positives.

Several selection filters were used to refine the lung metastasis–related gene set. First, we filtered potential host-tissue genes characterized by a 1.5-fold higher expression level in normal lung tissue than in each of the other three normal tissue pools (bone, liver, and brain). These genes were considered as potentially expressed by contaminating host tissue. Second, genes with geometric mean intensities below 25 in both lung and nonlung metastases were removed. Finally, probes aligned against human genome not recognizing any transcripts were excluded.

To determine whether gene expression profiles were able to define low- and high-lung metastasis risk populations in the CRH cohort, a risk index was defined as a linear combination of six lung metastasis–related gene expression values. The cutoff point to categorize patients into high- or low-risk groups was evaluated in the training set (CRH cohort). We tested a range of cutoff points (from 65th–85th). The best association of the risk index and the lung metastasis–free survival was observed at the 75th percentile cutoff. The risk index function and the high-/low-risk cutoff point were then applied to the MSK, EMC, and NKI and combined cohorts.

Survival times were estimated by the Kaplan-Meier method, and the significance of differences was determined with the log-rank test. Multivariate analyses were performed using the Cox proportional hazards regression model.

The concordance of predictions obtained from our model and the previously described experimental lung metastasis model (11, 12) was analyzed using ρ statistics (a value of >0.60 indicates a strong relation).

Results

Identification of lung metastasis–associated genes. To identify lung metastasis–associated genes, we first performed a microarray analysis of 23 breast cancer metastases (Affymetrix U133 Plus 2.0 arrays). We compared the gene transcript profiles of 5 lung metastases and 18 metastases from other target organs (bone, liver, brain, and skin), all obtained from breast cancer patients. A class comparison was conducted based on an univariate t test, with a stringent P value of 10⁻⁴, to identify genes differentially expressed by the lung and the nonlung metastatic lesions.

After applying filtering criteria (see Patients and Methods), we identified 21 differentially expressed genes (Table 1; All genes were up-regulated with the exception of the tumor suppressor gene PTEN, which was down-regulated in lung metastases).

To technically validate the differentially expressed genes, we examined the expression of the highest ranking genes by quantitative RT-PCR using 19 samples analyzed by microarray (2 bone and 2 skin metastases had no more RNA available) and 10

Figure 1. Immunohistochemical analysis of FERMT1 and DSC2 in breast cancer metastases and matched primary tumors. Strong cytoplasmic immunoreactivity was found in both lung metastatic and matched primary tumor (A–B, C–D, I–J, and K–L). Lung parenchyma surrounding metastatic cells stained weakly for DSC2 and negatively for FERMT1. In contrast, nonlung metastases and matched primary tumors showed weak to no signal (E–F, G–H, M–N, and O–P). F and P, liver metastases; H and N, originating from uterus and bone, respectively. When expressed, the 2 proteins are detected in ~ 80% to 90% of the tumor cells. Original magnification, ×400.
additional breast cancer metastases including 3 lung, 2 liver, 4 skin, and 1 bone relapses. This step led to the validation of 7 genes showing a significant variation of expression (Mann-Whitney U test, \( P < 0.05 \)), namely DSC2, HORMAD1, TFCP2L1, UGT8, ITGB8, ANP32E, and FERMT1 (Table 1).

Furthermore, we evaluated the protein expression of two representative genes, DSC2 and FERMT1 (corresponding to the upper and lower \( P \) values; Table 1), using immunohistochemistry on paired samples corresponding to primary tumors and lung or nonlung metastases.

Strong immunoreactivity was detected for both proteins, almost exclusively localized in the tumor cells, and not in the surrounding pulmonary tissue (Fig. 1). High protein expression was observed in the lung metastases and matching primaries (Fig. 1A-D and I-L), whereas the breast tumors not relapsing to lung and paired nonlung metastases were weakly to not stained (Fig. 1E-H and M-P).

The characterized genes were mapped into the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes databases to investigate their functions and biological processes (Table 1). Interestingly, the lung metastasis-related genes showed an overrepresentation of membrane-bound molecules mainly involved in cell adhesion and/or signal transduction (DSC2, UGT8, ITGB8, and FERMT1). Very little is known on the other identified genes (HORMAD1, TFCP2L1, and ANP32E). Several of these molecules have previously been shown to play important roles in the acquisition of proliferative and/or invasive properties of epithelial cells (17–22).

**Development of a predictor of selective breast cancer failure to the lungs.** To determine whether the 7 lung metastasis–associated genes (DSC2, HORMAD1, TFCP2L1, UGT8, ITGB8, ANP32E, and FERMT1) could be expressed in normal and tumor breast tissues, we analyzed by quantitative RT-PCR their expression patterns in a series of 5 normal breast tissue samples, 6 breast cancer cell lines (MCF7, T47D, SKBR3, MDA-MB-231, MDA-MB-361, and MDA-MB-435), and 44 primary breast tumors. All lung metastasis–associated genes were expressed in tissues of mammary origin except HORMAD1 that showed very weak to no expression in all normal breast, all cell lines, and a majority of primary tumors (\( Ct > 35 \)). Therefore, HORMAD1 will not be considered later in our study.

Furthermore, expression levels of the six remaining genes in normal tissues from different organs were analyzed to assure that tumor cells (instead of stromal cells) are the cellular origin of the differential patterns (Supplementary Table S1).

The six genes were then used to develop a gene signature predictive of a higher risk of lung metastasis. We studied a cohort of 72 lymph node–negative patients specifically selected on the basis of the treatment and the metastatic outcome (CRH cohort). All 72 patients had not received neoadjuvant or adjuvant therapy. This meant that the potential prognostic effect of the "lung metastasis classifier" would not be influenced by factors related to systemic treatment. Thirty-eight patients had developed distant metastases (including 11 with lung metastases), and 34 patients remained free of disease within 10 years after their initial diagnosis. The clinical characteristics of this cohort are detailed in Fig. 2A.

The primary tumors were assigned to a high-risk group or a low-risk group, with respect to lung metastasis, according to the risk index calculated on the basis of the six-gene signature. Tumors

---

**Figure 2.** Performance of the 6-gene lung metastasis signature in a series of 72 node negative breast tumors (the CRH cohort). A, detailed clinical and pathologic characteristics of the breast cancer patients and correlation with the six-gene classifier (\( \chi^2 \) test). B, distribution of distant metastases (first and/or only metastatic sites) in the high-risk and low-risk groups identified by the six-gene signature (\( \chi^2 \) test). C, patients with tumors expressing the six-gene signature (high-risk group) had shorter lung metastasis–free survival (log-rank test). D, Kaplan-Meier curves of bone metastasis–free survival showed no significant differences between the high- and low-risk groups identified by the six-gene signature.
expressing high levels of the risk index metastasized significantly more frequently to the lungs than did the other tumors (\( P = 0.04 \), \( \chi^2 \) test; Fig. 2B). In addition, patients with such tumors had significantly shorter lung metastasis–free survival (\( P = 0.008 \); Fig. 2C). The six-gene signature did not correlate with the risk of bone metastasis (Fig. 2D) or liver metastasis (data not shown).

Validation of the predictive ability of the six-gene signature. To validate the predictive value of the six-gene lung metastasis signature, we analyzed expression profiles of three independent cohorts of breast cancer patients, which microarray data are publicly available (11–14). These 3 data sets correspond to 2 large cohorts of early stage breast cancer patients (NKI and EMC series, \( n = 295 \) and 344, respectively) and a cohort of locally advanced breast cancers (the MSK cohort, \( n = 82 \); refs. 11–14).

Hierarchical clustering analysis was performed on all individual series. Within the 3 different cohorts, the 6-gene signature discriminates a subgroup of breast tumors with a higher propensity to metastasize to the lungs (\( P = 0.032, 0.016, \) and 0.014, \( \chi^2 \) test, for MSK, EMC, and NKI, respectively; Fig. 3).

Although significant, the results of hierarchical clustering are only indicative. Thus, using the same procedure as for the CRH cohort, we evaluated the six-gene signature in the three independent cohorts. Patients assigned to the high-risk group had significantly shorter lung metastasis–free survival (\( P = 0.004, 0.001, \) and 0.039 for MSK, EMC, and NKI series, respectively; Fig. 4), whereas there was no difference in bone metastasis–free survival (data not shown). It is noteworthy that the NKI cohort showed a lower discrimination probably due to the evaluation of only five genes of the signature (\( \text{ANP32E} \) was not present on the corresponding chip). When tested on the combined cohort (\( n = 721 \)), the 6-gene signature was highly correlated to the outcome of breast cancer patient with regard to lung metastasis (\( P < 10^{-5} \)).

Correlation to standard clinicopathologic variables and other prognostic signatures. We evaluated whether the six-gene signature provided additional prognostic information that may not be obtained by other signatures and/or standard markers. First, we analyzed the NKI series (for which the complete clinical data were documented) for the main prognostic signatures reported for breast cancers. Consistent with previous reports (9, 12), the primary breast tumors expressing the 6-gene signature were mostly of poor prognosis on the basis of standard pathologic variables [62% estrogen receptor (ER) negative and 70% grade 3] and previously reported poor-prognosis signatures such as the 70-gene signature (3, 13), the wound-healing signature (23, 24), and the basal-like molecular subtype signature (80%, 89%, and 58%, respectively; Fig. 5A; refs. 25, 26).

In addition, when analyzing the clinicopathologic variables available for each of the cohorts of breast cancer patients, we found no difference between the high- and low-risk groups with respect to age, lymph node status, or primary tumor size (data not shown).

To ensure that the 6-gene classifier improved risk stratification independently of the standard clinical variables, we performed a multivariate Cox proportional hazards analysis on the combined cohort (\( n = 721 \); only ER and lymph node status variables were
available for all patients; Fig. 5B). The Cox model showed that the 6-gene signature and ER status were independent predictors of lung metastasis ($P = 0.01$ and $0.04$, respectively).

Finally, we compared the predictive value of our lung metastasis signature to the one derived from a mouse model (named LMS for “Lung Metastasis Signature,” ref. 12). These two signatures are defined by expression patterns of distinct sets of genes with no overlap. When evaluated on the same series of 721 samples, we observe that despite their different derivations, the signatures gave overlapping and concordant predictions outcome. Almost all tumors identified as LMS + were also classified as expressing the six-gene signature. The two models showed 85% agreement in outcome classification of breast cancer patients ($\kappa$ coefficient = 0.57), suggesting that they could track common set of biological characteristics conferring the organ-specific metastatic phenotypes. Indeed, among the molecular pathways underlying the LMS and the six-gene signature, common processes could be highlighted such as the transforming growth factor (TGF)-β pathway and the signaling through focal adhesions (interaction between the extracellular matrix and integrins).

To determine whether the use of the two signatures together would result in a better model than the use of any one alone, we derived a single model based on the common findings of the two models separately. The performance of this model according to the Kaplan-Meier analysis was noticeably better than each of the 2 models (Fig. 5C and D) demonstrating that gene signatures derived from 2 distinct approaches can be complementary, as recently reported for the 70-gene and WR signatures (23).

**Discussion**

Molecular and cellular mechanisms mediating breast cancer metastasis to the lungs are poorly understood. All existing data were obtained using experimentally derived mouse tumor models, in particular, an experimental model based on MDA-MB-231 breast cancer cell lines with specific organotropism selected in vivo (9, 10). From this model, Massague and colleagues (9, 10, 27) generated gene signatures mediating specific breast cancer metastasis to the bone and the lungs. Importantly, the MDA-derived LMS was found to be predictive of a high risk for developing lung metastases in human breast cancer patients (10–12).

In the present study, we aimed to investigate the mechanisms that drive the selective breast cancer relapse to the lungs directly from human clinical samples. Thus, to functionally filter expression patterns mediating breast cancer lung metastasis, we applied the concept of in vivo selection of metastatic properties as designed in the MDA-based model in the context of human cancer progression. Therefore, we analyzed a large panel of breast cancer metastases. Such samples are rarely available, as most metastatic patients receive systemic treatment and do not have surgery.

The comparison of lung and nonlung metastases identified 21 differentially expressed genes. A new set of cell adhesion molecules that might contribute to lung metastasis were highlighted. Such adhesion molecules belong to various protein families including integrins (ITGB8), cadherins (CDH3), desmosomal proteins (DSC2), and focal adhesion molecules (FERMT1). These findings are in agreement with previous works emphasizing

**Figure 4.** Validation of the six-gene lung metastasis signature in three independent series of breast cancer patients. Lung metastasis–free survival was analyzed for MSK (A), EMC (B), and NKI (C) cohorts (82, 344 and 295 patients, respectively) and the combined cohort of 721 breast cancer patients (D). Kaplan-Meier analysis distinguished patients who expressed (high-risk group) and did not express (low-risk group) the six-gene signature. Patients with a high-predicted risk of lung metastasis had shorter lung metastasis–free survival.
that diverse alterations in adhesive properties of cancer cells play key roles in the enhancement of their metastatic spread (7). Furthermore, it has been shown that adhesion molecules might facilitate specific interaction of tumor cells with the pulmonary tissue. Such molecules have previously been reported to be involved in breast cancer metastasis to the lungs (28, 29). In the present study, none of the identified genes have been previously shown to be involved in metastasis, several of the encoded molecules have been shown to play important roles in the acquisition of proliferative and/or invasive properties of epithelial cells. In particular, one signaling pathway that might be involved is the regulation of the epithelial-mesenchymal transition (EMT). Indeed, integrin β (8) mediates activation of TGF-β signaling, thereby regulating the EMT (30), whereas FERMT1 may be an effector of the TGF-β–mediated EMT (21). Interestingly, the tumor suppressor gene PTEN was the only down-regulated lung metastasis–associated gene in our panel. PTEN is a major regulator of proliferation, growth, cell survival, and migration signaling pathways, and is mutated or deleted in many tumor types. PTEN also regulates EMT, cell motility, and CXCR4 chemotaxis, which could contribute to its involvement in lung metastasis (31–33).

Taken together, these findings highlight the molecular pathways that might drive the organ preference of breast tumor cells for lung. A better understanding of their functions might stimulate the development of targeted therapeutics to prevent such specific metastases.

From this set of candidate lung metastasis–associated genes, we developed a six-gene predictor of selective breast cancer relapse to the lungs. This six-gene signature (DSC2, TFCP2L1, UGT8, ITGB8, ANP32E, and FERMT1) was generated from a series of lymph node–negative breast cancer patients who did not receive any neoadjuvant or adjuvant therapy to allow the analysis of the signature prognostic effect along with the natural history of the disease. The predictor would thus not be influenced by factors related to systemic treatment. The six-gene classifier was then validated in two independent data sets of patients of early stage (as the training set) generated from two distinct microarray platforms and a third series of locally advanced breast cancer patients. In all tested individual series and in the combined cohort (including >700 cases), the 6-gene signature had a strong predictive ability for breast cancer lung metastasis. When compared with the standard clinical variables, the six-gene signature was found to be an independent predictor of lung metastasis.

Although there is no targeted therapy for lung metastasis, such as bisphosphonates for bone metastasis, the knowledge of organ-specific metastasis has been emphasized the last few years and might lead to targeted therapeutics in the near future. By delineating the risk for lung metastasis based on gene signatures, it might be possible that these high-risk breast cancer patients may benefit from these therapies targeting specific secondary failures.

Our results also support previous studies showing that a subset of primary tumors express genes predictive of metastasis to...
specific organs (9, 10). Interestingly, although none of the six genes of our signature were identified in the MDA-MB-231 model (9–12), the predictive value of the two signatures on the same series of samples were similar, suggesting that organ-specific outcomes can readily be predicted by a limited number of genes sufficiently representative of a given cellular phenotype.

The six-gene signature provides additional prognostic information beyond the previously reported LMS signature. As suggested by Fan and colleagues (34), each new signature generated from distinct organspecific metastasis. The combined information obtained from both signatures improved risk stratification compared with individual signatures as previously described for breast cancer prognosis using the Wound-Response and 70-gene signatures (3, 23).

The possible role of the six lung metastasis–associated genes in specific steps of the lung metastatic process needs now to be functionally validated. Dissection of lung metastasis molecular pathways might lead to the development of new prognostic markers that could help identifying patients who are most likely to develop lung metastases and who might benefit from lung-specific antimetastatic therapies. However, the translation of these findings to the clinic should include a preliminary screening step to prospectively identify those cancer patients that might develop lung metastases.

In this study, we show, for the first time, that the gene expression profiles of human metastatic samples could be used to predict lung metastasis in breast cancer patients. Using a similar strategy to other secondary sites could greatly improve the prognosis of breast cancer patients.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
Received 2/5/2008; revised 4/25/2008; accepted 5/5/2008.

Grant support: European Commission Framework Programme VI MetaBre (CCE LSHC-CT-2004-503049) and "La Ligue contre le cancer des Hauts de Seine," the European Community within the MetaBre consortium (T. Landemaine and N. Rucci), and the Dutch Breast Cancer Research Foundation (L. van der Kooi, S. Vos). A. Bellachv de is a Research Associate at the National Fund for Scientific Research (Belgium).

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

We thank Prof. J. Boniver and Dr. L. de Leval (Department of Pathology, University of Lille), Dr. M. Bredl and Dr. S. Bohlud (Neurosurgery and Pathology Services, Bellvitge Hospital, IDIBELL), Dr. M. Gil (Onocology Service, Catalan Institute of Oncology, IDIBELL), Dr. F. Martella (Department of Experimental Medicine, University of L’Aquila), Dr. O. Morcenschi (Orthopedics Department, University 'a Sapienza,' Rome) for the selection of breast cancer metastatic samples, P. Heneaux (Metastasis Research Laboratory, University of Liège), and S. Vacher (Laboratoire d’oncogénétique, Centre René Huguenin) for expert technical assistance, and all the partners of the METABRE consortium for fruitful discussions: M. Bracke (Ghent University Hospital, Belgium), R. Bucciene (Consorzio Mario Negri Sud, Italy), P. Clément-Lacroix (Proskeïlia, France), P. Clezardin (Institut National de la Sante et de la Recherche Medicale, France), S. Eccles (Institute of Cancer Research, UK), M. Ugorski (Wroclaw University of Environmental and Life Sciences, Poland), and G. van der Pluijm (Leiden University Medical Center, the Netherlands).

www.aacrjournals.org

Cancer Res 2008; 68: (15), August 1, 2008

6099

References
A Six-Gene Signature Predicting Breast Cancer Lung Metastasis

Thomas Landemaine, Amanda Jackson, Akeila Bellahcène, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/68/15/6092

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2008/07/28/68.15.6092.DC1

Cited articles
This article cites 34 articles, 15 of which you can access for free at:
http://cancerres.aacrjournals.org/content/68/15/6092.full.html#ref-list-1

Citing articles
This article has been cited by 14 HighWire-hosted articles. Access the articles at:
/content/68/15/6092.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

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