

# A Six-Gene Signature Predicting Breast Cancer Lung Metastasis

Thomas Landemaine,<sup>1,2</sup> Amanda Jackson,<sup>3</sup> Akeila Bellahcène,<sup>4</sup> Nadia Rucci,<sup>5</sup> Soraya Sin,<sup>1,2</sup> Berta Martin Abad,<sup>6</sup> Angels Sierra,<sup>6</sup> Alain Boudinet,<sup>1</sup> Jean-Marc Guinebrière,<sup>1</sup> Enrico Ricevuto,<sup>5</sup> Catherine Noguès,<sup>1,2</sup> Marianne Briffod,<sup>1,2</sup> Ivan Bièche,<sup>1,2</sup> Pascal Cherel,<sup>1</sup> Teresa Garcia,<sup>3</sup> Vincent Castronovo,<sup>4</sup> Anna Teti,<sup>5</sup> Rosette Lidereau,<sup>1,2</sup> and Keltouma Driouch<sup>1,2</sup>

<sup>1</sup>Centre René Hugué, Fédération Nationale des Centres de Lutte Contre le Cancer and <sup>2</sup>Institut National de la Santé et de la Recherche Médicale, U735, Saint-Cloud, France; <sup>3</sup>Proskelia, Romainville, France; <sup>4</sup>Metastasis Research Laboratory, Center of Experimental Cancer Research, University of Liège, Liège, Belgium; <sup>5</sup>Department of Experimental Medicine, University of L'Aquila, L'Aquila, Italy; and <sup>6</sup>Center of Molecular Oncology, Institut d'Investigació Biomèdica de Bellvitge, 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 tumorigenesis-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

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Current address for A. Jackson: Division of Cell & Molecular Biology, Imperial College London, London, United Kingdom.

**Requests for reprints:** Keltouma Driouch, Institut National de la Santé et de la Recherche Médicale U735/Oncogénétique, Centre René Hugué, 35 rue Dailly, 92210 Saint-Cloud, France. Phone: 331-47-11-15-66; Fax: 331-47-11-16-96; E-mail: k.driouch@stcloud-huguenin.org.

©2008 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-08-0436

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'Investigació Biomèdica de Bellvitge (IDIBELL), and Centre René Huguénin. 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 ("CRH cohort") was specifically selected from patients with node-negative breast cancer treated at the Centre René Huguénin, 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" ( $n = 344$ ), 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 Huguénin 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.

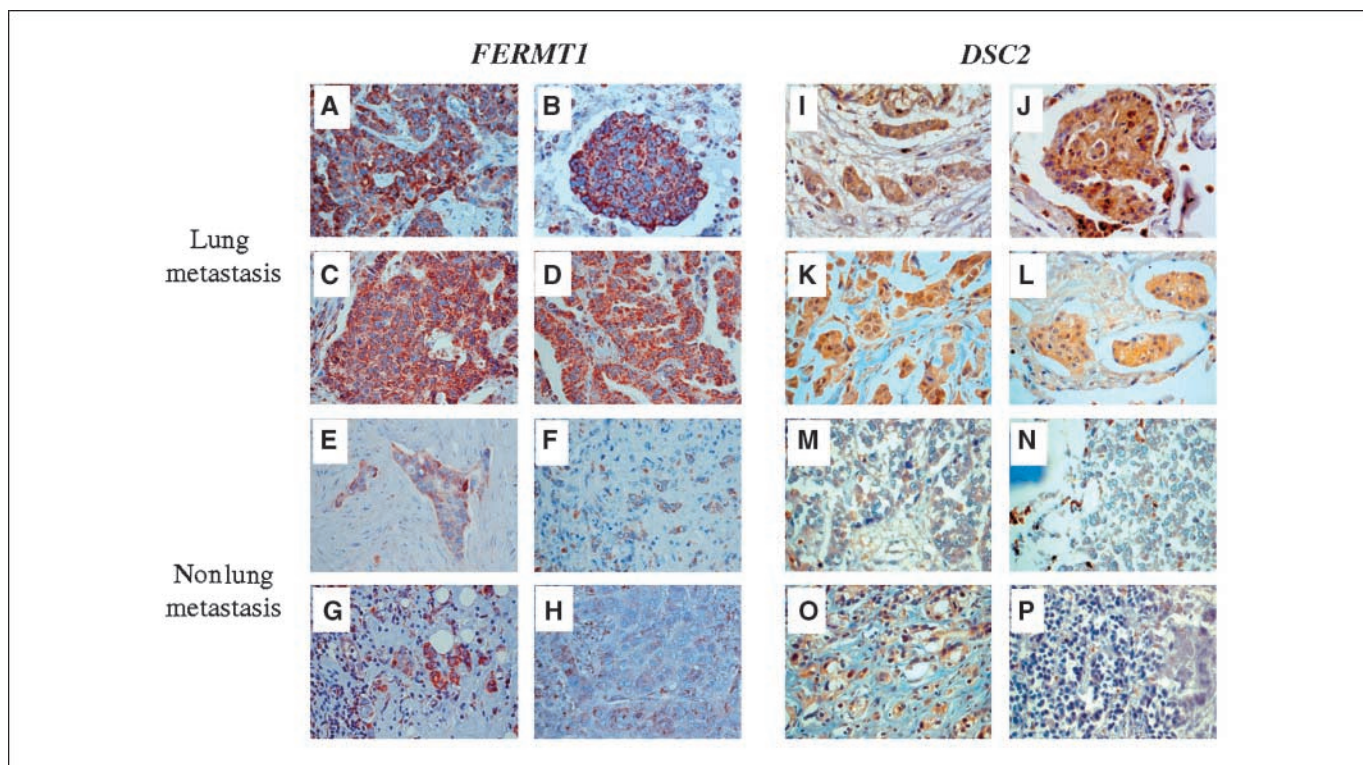
For quantitative RT-PCR analysis, we used cDNA synthesis and PCR conditions described in detail elsewhere (16). All PCR reactions were performed with an ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems) and the SYBR Green PCR Core Reagents kit (Perkin-Elmer Applied Biosystems). 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 ( $n = 18$ )

Probe set	Gene symbol	Gene title	Function/biological process	Fold change	Parametric <i>P</i>
204751_x_at	<i>DSC2*</i>	Desmocollin 2	Cell adhesion	8.2	$P < 0,000001$
223861_at	<i>HORMAD1*</i>	HORMA domain containing 1		21.5	2,00E-06
228171_s_at	<i>PLEKHG4*</i>	Pleckstrin homology domain containing, family G, member 4	Rho protein signal transduction	2.7	2,00E-06
228577_x_at	<i>ODF2L*</i>	Outer dense fiber of sperm tails 2 like		2.9	8,00E-06
220941_s_at	<i>C21orf91*</i>	Chromosome 21 open reading frame 91		3.6	1,00E-05
227642_at	<i>TFCP2L1*</i>	Transcription factor CP2 like 1	Regulation of transcription	5.5	1,00E-05
219867_at	<i>CHODL*</i>	Chondrolectin	Hyaluronic acid binding	2.8	1,40E-05
205428_s_at	<i>CALB2*</i>	Calbindin 2 (calretinin)	Calcium ion binding	9.2	1,60E-05
228956_at	<i>UGT8*</i>	UDP glucosyltransferase 8	Glycosphingolipid biosynthetic process	15.4	1,80E-05
1554246_at	<i>C1orf210*</i>	Chromosome 1 open reading frame 210		2.3	2,10E-05
221705_s_at	<i>SIKE*</i>	Suppressor of IKK $\epsilon$		1.9	2,10E-05
211488_s_at	<i>ITGB8*</i>	Integrin, $\beta$ 8	Cell adhesion/Signal transduction	1.7	2,60E-05
213372_at	<i>PAQR3*</i>	Progesterin and adipoQ receptor family member III	Receptor activity	5.6	2,90E-05
208103_s_at	<i>ANP32E*</i>	Acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	Phosphatase inhibitor activity	6.0	3,20E-05
60474_at	<i>FERMT1*</i>	Fermitin family homologue 1 (drosophila)	Cell adhesion	6.9	3,60E-05
222869_s_at	<i>ELAC1*</i>	elaC homologue 1 (Escherichia coli)	tRNA processing	1.6	3,60E-05
227829_at	<i>GYLTL1B</i>	Glycosyltransferase-like 1B	Glycosphingolipid biosynthetic process	2.6	3,90E-05
226075_at	<i>SPSBI</i>	splA/ryanodine receptor domain and SOCS box containing 1	Intracellular signaling cascade	2.9	5,60E-05
1553705_a_at	<i>CHRM3</i>	Cholinergic receptor, muscarinic 3	G-protein coupled signal transduction	2.0	6,10E-05
225363_at	<i>PTEN</i>	Phosphatase and tensin homologue (mutated in multiple advanced cancers 1)	Phosphatidylinositol signaling	0.4	6,20E-05
203256_at	<i>CDH3</i>	Cadherin 3, type 1, P-cadherin (placental)	Cell adhesion	6.1	9,50E-05

\*Genes tested by qRT-PCR on a larger series of metastases.



**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,  $\times 400$ .

H<sub>2</sub>O<sub>2</sub>, 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 $\beta$ 3, developed by Dr. R. Simon and Dr. A. Peng Lam.<sup>7</sup> Expression data were collated as CEL files, and normalization was done using the RMA function of Bioconductor.<sup>8</sup> 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<sup>9</sup> 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  $\kappa$  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

<sup>7</sup> <http://linus.nci.nih.gov/BRB-ArrayTools.html>

<sup>8</sup> Gene Omnibus Database accession number: GSE11078.

<sup>9</sup> <http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>



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. 1*A-D* and 1*I-L*), whereas the breast tumors not relapsing to lung and paired nonlung metastases were weakly to not stained (Fig. 1*E-H* and 1*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).

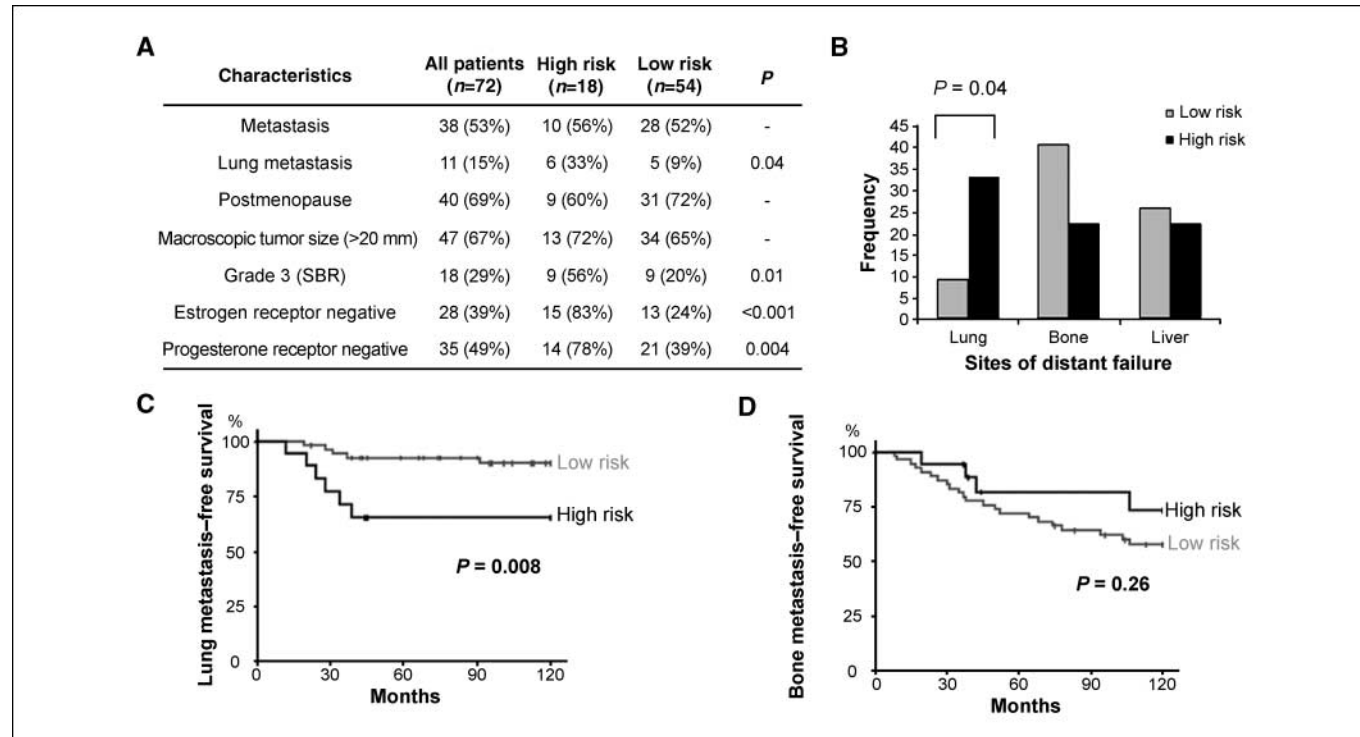
**Development of a predictor of selective breast cancer failure to the lungs.** To determine whether the 7 lung metastasis-

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.

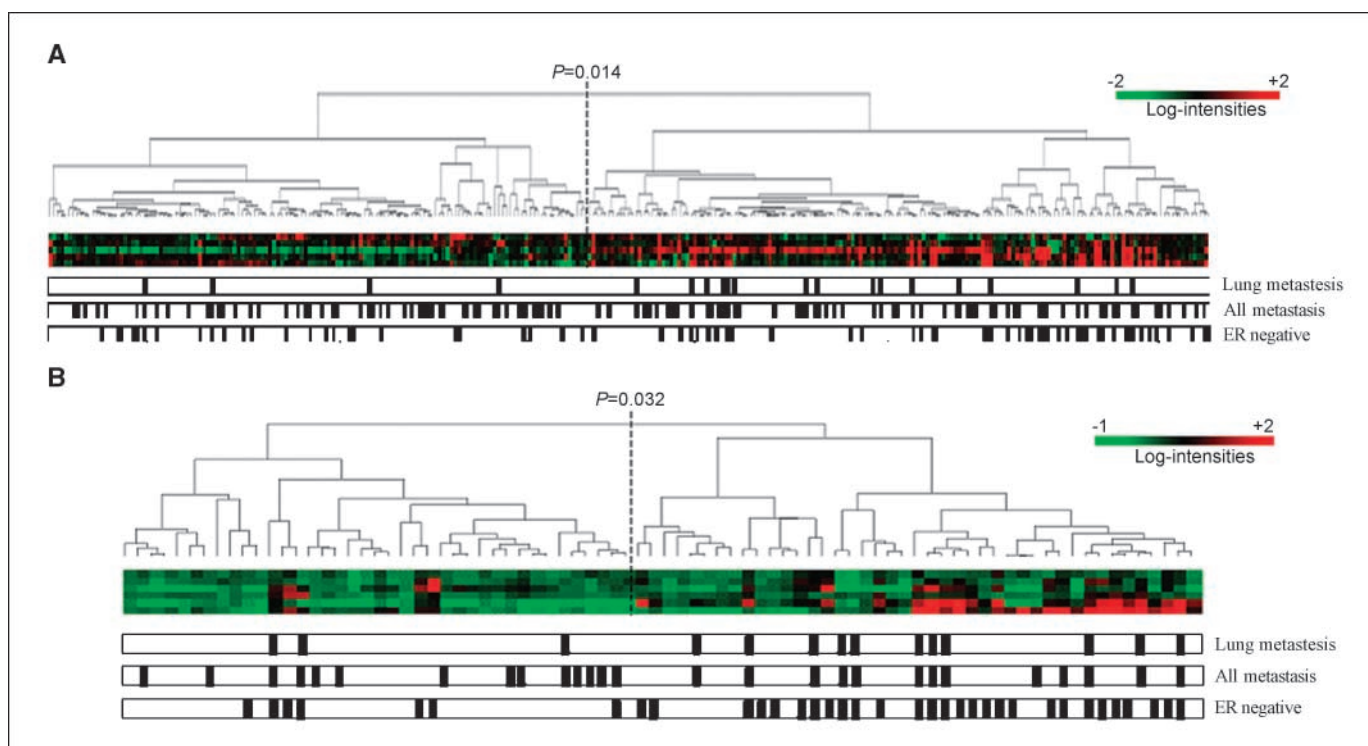
Furthermore, expression levels of the six remaining genes in normal tissues from different organs were analyzed to assure that tumor cells (instead of stromal cells) are the cellular origin of the differential patterns (Supplementary Table S1).

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. 2*A*.

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.



**Figure 3.** Hierarchical clustering of 2 cohorts of breast tumors (NKI early-stage tumors and MSK locally advanced tumors) based on the expression of six-gene lung metastasis signature. *A*, hierarchical clustering of the NKI breast cancer patients ( $n = 295$ ) with the lung metastasis gene signature. The tumors are separated into two main clusters. Corresponding clinical variables and outcomes are indicated. *Black vertical bars*, patients who developed lung metastasis or overall metastasis and patients whose tumors were ER negative. The cluster of patients expressing high levels of the six-gene signature has a higher propensity to metastasize to the lungs ( $P = 0.014$ ). *B*, hierarchical clustering of MSK primary breast carcinomas ( $n = 82$ ) based on the expression of the 6 lung metastasis-associated genes. *Black vertical bars*, tumors from patients who developed lung metastasis. Patients with tumors expressing the 6-gene signature had a higher propensity to metastasize to the lungs ( $P = 0.032$ ).

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. 2*B*). In addition, patients with such tumors had significantly shorter lung metastasis-free survival ( $P = 0.008$ ; Fig. 2*C*). The six-gene signature did not correlate with the risk of bone metastasis (Fig. 2*D*) 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 (*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. 5*A*; 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<sup>+</sup> 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)- $\beta$  pathway and the signaling through focal adhesions (interaction between the extracellular matrix and integrins).

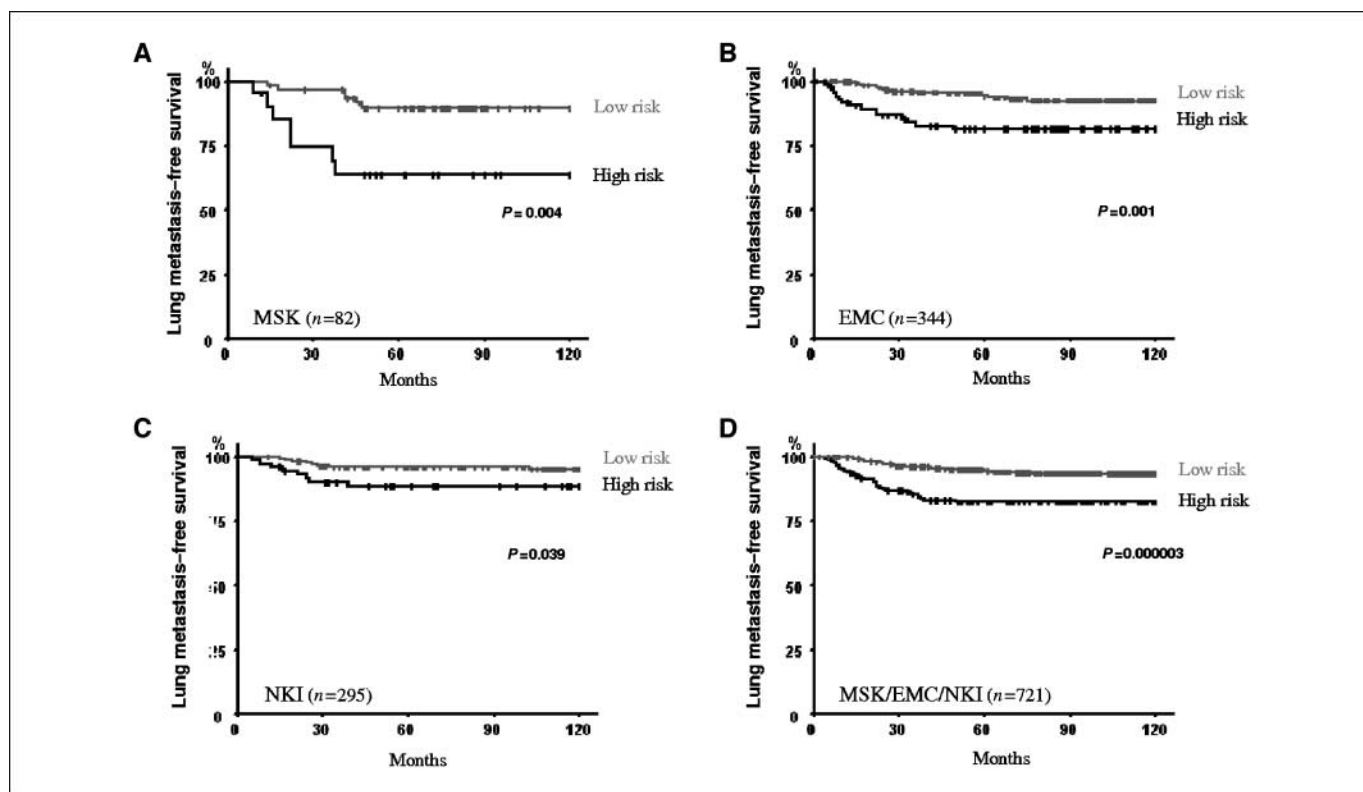
To determine whether the use of the two signatures together would result in a better model than the use of any one alone, we derived a single model based on the common findings of the two models separately. The performance of this model according to the Kaplan-Meier analysis was noticeably better than 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.

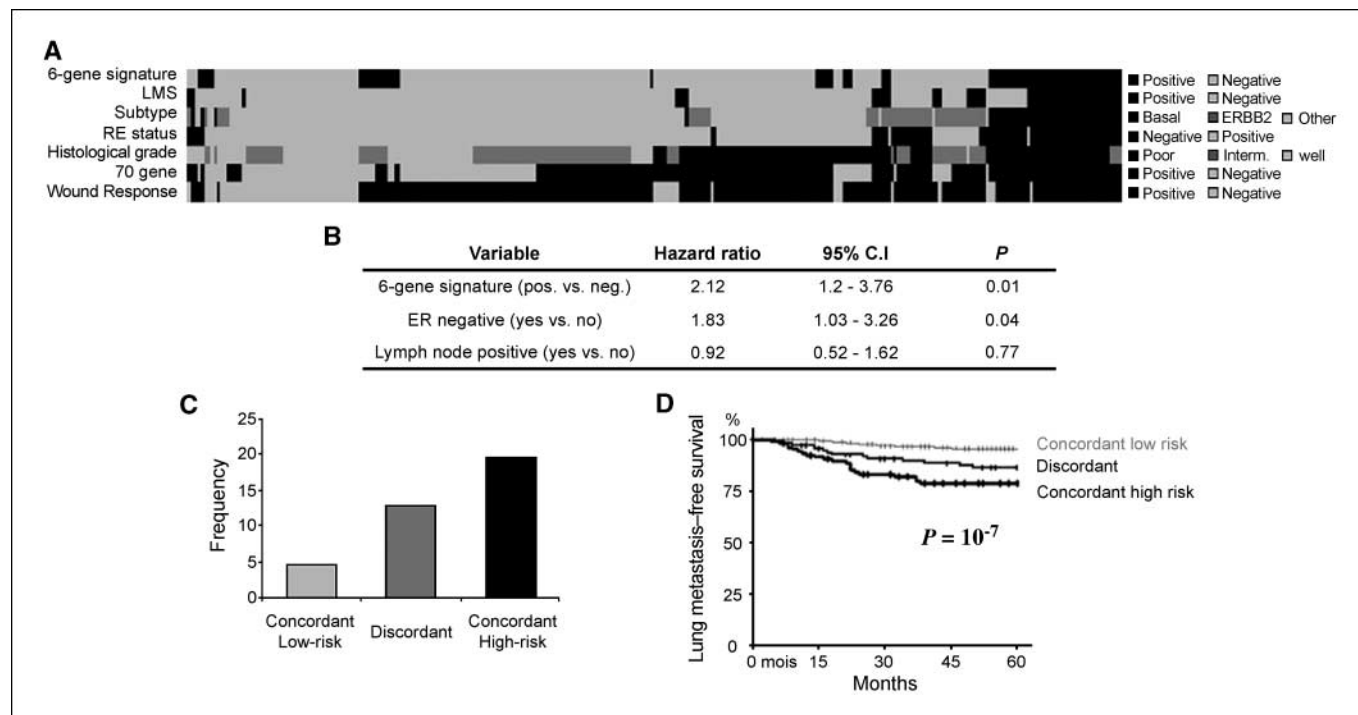
that diverse alterations in adhesive properties of cancer cells play key roles in the enhancement of their metastatic spread (7). Furthermore, it has been shown that adhesion molecules might facilitate specific interaction of tumor cells with the pulmonary tissue. Such molecules have previously been reported to be involved in breast cancer metastasis to the lungs (28, 29). In the present study, although 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  $\beta$  (8) mediates activation of TGF- $\beta$  signaling, thereby regulating the EMT (30), whereas FERMT1 may be an effector of the TGF- $\beta$ -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 knowledge 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



**Figure 5.** Integration of diverse clinicopathologic markers and gene expression signatures for lung metastasis risk prediction. **A**, hierarchical clustering of the NCI cohort of breast cancer patients ( $n = 295$ ) was performed with the indicated pathologic and genomic markers consisting of the MDA-MB-231-based LMS (11, 12), molecular subtype (25, 26), ER status, histologic grade, 70-gene prognosis signature (3, 13), and wound-response signature (23, 24). The legend for the codes for each variable is shown. **B**, multivariate analysis for lung metastasis in the combined cohort of 721 breast cancer patients (MSK/EMC/NCI) using a Cox proportional hazard ratio model. **C**, distribution of lung metastases in the combined cohort ( $n = 721$ ) according to the clinically and experimentally derived signatures. Patients that are negative for both signatures, those with divergent assignment, and patients found positive for both signatures are shown. The 6-gene and LMS signatures showed 85% agreement in outcome classification of breast cancer patients with respect to lung metastasis ( $\kappa$  coefficient = 0.57). **D**, Kaplan-Meier analysis of breast cancer patients ( $n = 721$ ) according to the 6-gene and MDA-derived lung metastasis signatures. Prediction of breast cancer lung metastasis is improved by using a combination of predictors derived from two distinct models.



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 models, platforms, or mathematical methods has the potential of adding prognostic information. Here, we show that the experimentally and clinically derived lung metastasis signatures classified tumors into coherent and internally consistent groups. Distinct analytic approaches can be complementary even in the field of organ-specific 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 (CEE 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 Breast Cancer Research Foundation (USA; S. Sin). A. Bellahçène 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 Liège), Dr. M. Brell and Dr. S. Boluda (Neurosurgery and Pathology Services, Bellvitge Hospital, IDIBELL), Dr. M. Gil (Oncology Service, Catalan Institute of Oncology, IDIBELL), Dr. F. Martella (Department of Experimental Medicine, University of L'Aquila), Dr. O. Moreschini (Orthopedics Department, University “La 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. Buccione (Consorzio Mario Negri Sud, Italy), P. Clément-Lacroix (Proskelia, France), P. Clezardin (Institut National de la Santé et de la Recherche Médicale, France), S. Eccles (Institute of Cancer Research, UK), M. Ugorski (Wrocław University of Environmental and Life Sciences, Poland), and G. van der Pluijm (Leiden University Medical Center, the Netherlands).

## References

- Greenberg PA, Hortobagyi GN, Smith TL, et al. Long-term follow-up of patients with complete remission following combination chemotherapy for metastatic breast cancer. *J Clin Oncol* 1996;14:2197–205.
- Fidler IJ. The pathogenesis of cancer metastasis: the seed and soil hypothesis revisited. *Nat Rev Cancer* 2003;3:453–8.
- van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
- Weigelt B, Glas AM, Wessels LF, et al. Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci U S A* 2003;100:15901–5.
- Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003;33:49–54.
- Drionch K, Landemaine T, Sin S, Wang S, Lidereau R. Gene arrays for diagnosis, prognosis and treatment of breast cancer metastasis. *Clin Exp Metastasis* 2007;24:575–85.
- Gupta GP, Massague J. Cancer metastasis: building a framework. *Cell* 2006;127:679–95.
- Steege PS. Tumor metastasis: mechanistic insights and clinical challenges. *Nat Med* 2006;12:895–904.
- Kang Y, Siegel PM, Shu W, et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 2003;3:537–49.
- Minn AJ, Kang Y, Serganova I, et al. Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J Clin Invest* 2005;115:44–55.
- Minn AJ, Gupta GP, Siegel PM, et al. Genes that mediate breast cancer metastasis to lung. *Nature* 2005;436:518–24.
- Minn AJ, Gupta GP, Padua D, et al. Lung metastasis genes couple breast tumor size and metastatic spread. *Proc Natl Acad Sci U S A* 2007;104:6740–5.
- van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009.
- Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365:671–9.
- Jackson A, Vayssières B, Garcia T, et al. Gene array analysis of Wnt-regulated genes in C3H10T1/2 cells. *Bone* 2005;36:585–98.
- Bieche I, Parfait B, Le Doussal V, et al. Identification of CGA as a novel estrogen receptor-responsive gene in breast cancer: an outstanding candidate marker to predict the response to endocrine therapy. *Cancer Res* 2001;61:1652–8.
- Hardman MJ, Liu K, Avilion AA, et al. Desmosomal cadherin misexpression alters  $\beta$ -catenin stability and epidermal differentiation. *Mol Cell Biol* 2005;25:969–78.
- Windoffer R, Borchert-Stuhltrager M, Leube RE. Desmosomes: interconnected calcium-dependent structures of remarkable stability with significant integral membrane protein turnover. *J Cell Sci* 2002;115:1717–32.
- Mu D, Cambier S, Fjellbirkeland L, et al. The integrin  $\alpha(v)\beta 8$  mediates epithelial homeostasis through MT1-MMP-dependent activation of TGF- $\beta 1$ . *J Cell Biol* 2002;157:493–507.
- Lakhe-Reddy S, Khan S, Konieczkowski M, et al.  $\beta 8$  integrin binds Rho GDP dissociation inhibitor-1 and activates Rac1 to inhibit mesangial cell myofibroblast differentiation. *J Biol Chem* 2006;281:19688–99.
- Kloeker S, Major MB, Calderwood DA, et al. The Kindler syndrome protein is regulated by transforming growth factor- $\beta$  and involved in integrin-mediated adhesion. *J Biol Chem* 2004;279:6824–33.
- Herz C, Aumailley M, Schulte C, et al. Kindlin-1 is a phosphoprotein involved in regulation of polarity, proliferation, and motility of epidermal keratinocytes. *J Biol Chem* 2006;281:36082–90.
- Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* 2005;102:3738–43.
- Chang HY, Sneddon JB, Alizadeh AA, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2004;2:E7.
- Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52.
- Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
- Nguyen DX, Massague J. Genetic determinants of cancer metastasis. *Nat Rev Genet* 2007;8:341–52.
- Abdel-Ghany M, Cheng HC, Elble RC, Pauli BU. The breast cancer  $\beta 4$  integrin and endothelial human CLCA2 mediate lung metastasis. *J Biol Chem* 2001;276:25438–46.
- Brown DM, Ruoslahti E. Metadherin, a cell surface protein in breast tumors that mediates lung metastasis. *Cancer Cell* 2004;5:365–74.
- Araya J, Cambier S, Morris A, Finkbeiner W, Nishimura SL. Integrin-mediated transforming growth factor- $\beta$  activation regulates homeostasis of the pulmonary epithelial-mesenchymal trophic unit. *Am J Pathol* 2006;169:405–15.
- Gao P, Wange RL, Zhang N, Oppenheim JJ, Howard OM. Negative regulation of CXCR4-mediated chemotaxis by the lipid phosphatase activity of tumor suppressor PTEN. *Blood* 2005;106:2619–26.
- Phillips RJ, Mestas J, Gharrae-Kermani M, et al. Epidermal growth factor and hypoxia-induced expression of CXCR4 chemokine receptor 4 on non-small cell lung cancer cells is regulated by the phosphatidylinositol 3-kinase/PTEN/AKT/mammalian target of rapamycin signaling pathway and activation of hypoxia inducible factor-1 $\alpha$ . *J Biol Chem* 2005;280:22473–81.
- Leslie NR, Yang X, Downes CP, Weijer CJ. PtdIns(3,4,5)P(3)-dependent and -independent roles for PTEN in the control of cell migration. *Curr Biol* 2007;17:115–25.
- Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;355:560–9.



# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## A Six-Gene Signature Predicting Breast Cancer Lung Metastasis

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

*Cancer Res* 2008;68:6092-6099.

**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#ref-list-1>

**Citing articles** This article has been cited by 12 HighWire-hosted articles. Access the articles at:  
<http://cancerres.aacrjournals.org/content/68/15/6092.full#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](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cancerres.aacrjournals.org/content/68/15/6092>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.