

## Biological Processes Associated with Breast Cancer Clinical Outcome Depend on the Molecular Subtypes

Christine Desmedt,<sup>1</sup> Benjamin Haibe-Kains,<sup>1,2</sup> Pratyaksha Wirapati,<sup>3,4</sup> Marc Buyse,<sup>5</sup> Denis Larsimont,<sup>1</sup> Gianluca Bontempi,<sup>2</sup> Mauro Delorenzi,<sup>3,4</sup> Martine Piccart,<sup>1</sup> and Christos Sotiriou<sup>1</sup>

**Abstract Purpose:** Recently, several prognostic gene expression signatures have been identified; however, their performance has never been evaluated according to the previously described molecular subtypes based on the estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2), and their biological meaning has remained unclear. Here we aimed to perform a comprehensive meta-analysis integrating both clinicopathologic and gene expression data, focusing on the main molecular subtypes.

**Experimental Design:** We developed gene expression modules related to key biological processes in breast cancer such as tumor invasion, immune response, angiogenesis, apoptosis, proliferation, and ER and HER2 signaling, and then analyzed these modules together with clinical variables and several prognostic signatures on publicly available microarray studies (>2,100 patients).

**Results:** Multivariate analysis showed that in the ER+/HER2- subgroup, only the proliferation module and the histologic grade were significantly associated with clinical outcome. In the ER-/HER2- subgroup, only the immune response module was associated with prognosis, whereas in the HER2+ tumors, the tumor invasion and immune response modules displayed significant association with survival. Proliferation was identified as the most important component of several prognostic signatures, and their performance was limited to the ER+/HER2- subgroup.

**Conclusions:** Although proliferation is the strongest parameter predicting clinical outcome in the ER+/HER2- subtype and the common denominator of most prognostic gene signatures, immune response and tumor invasion seem to be the main molecular processes associated with prognosis in the ER-/HER2- and HER2+ subgroups, respectively. These findings may help to define new clinicogenomic models and to identify new therapeutic strategies in the specific molecular subgroups.

The prognosis and management of breast cancer has always been influenced by the classic variables such as histologic type and grade; tumor size; lymph node involvement; and the status of estrogen receptor (ER; ESR1), progesterone receptor, and human epidermal growth factor receptor 2 (HER2; ERBB2) of the tumor. During the last two decades, the prognosis of breast cancer has extensively been studied using several clinicopathologic parameters and tumor markers reflecting different stages of the disease and breast tumor biology. Recently, different research groups identified gene expression signatures predicting clinical outcome (1–10). A feature common to all these gene expression signatures is that they outperform conventional

clinicopathologic criteria mostly by identifying a higher proportion of low-risk patients not necessarily needing systemic adjuvant treatment, while still correctly identifying the high-risk patients. Although the signatures all address the same clinical question, it might be surprising that there is only little or no overlap between their gene lists, raising questions about their biological meaning. Moreover, although it has repeatedly and consistently been shown that breast cancer, in addition to being a clinically heterogeneous disease, is also molecularly heterogeneous, with subgroups primarily defined by ER and HER2 expression, the different prognostic signatures were never clearly evaluated and compared in these different molecular

**Authors' Affiliations:** <sup>1</sup>Medical Oncology Department, Jules Bordet Institute; <sup>2</sup>Machine Learning Group, Université Libre de Bruxelles, Brussels, Belgium; <sup>3</sup>National Center of Competence in Research Molecular Oncology, Swiss Institute of Experimental Cancer Research, Epalinges, Switzerland; <sup>4</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland; and <sup>5</sup>International Drug Development Institute, Louvain-La-Neuve, Belgium Received 10/29/07; revised 2/28/08; accepted 3/23/08.

**Grant support:** Belgian National Foundation for Cancer Research, FNRS (C. Desmedt, B. Haibe-Kains, and C. Sotiriou); European Commission Framework Programme VI grant FP6-LSHC-CT-2004-503426 (P. Wirapati); the National Center of Competence in Research Molecular Oncology of the Swiss National Science Foundation (M. Delorenzi); The Breast Cancer Research Foundation (C. Sotiriou); and the MEDIC Foundation (C. Sotiriou).

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.

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

C. Desmedt and B. Haibe-Kains contributed equally to this work.

**Requests for reprints:** Christos Sotiriou, Translational Research Unit, Medical Oncology Department, Jules Bordet Institute, 125 Boulevard de Waterloo, 1000 Brussels, Belgium. Phone: 32-2-541-3428; Fax: 11-32-2-538-0858; E-mail: [christos.sotiriou@bordet.be](mailto:christos.sotiriou@bordet.be).

© 2008 American Association for Cancer Research.  
doi:10.1158/1078-0432.CCR-07-4756

**Table 1.** Characteristics of the publicly available gene expression data sets

	NKI	NKI2	STNO2	NCI	MGH	UPP	STK
	No. patients (%)						
Total	76	295	118	99	60	110	159
Age (y)							
≤50	53 (70)	264 (90)	37 (31)	29 (30)	2 (4)	25 (22)	53 (33)
>50	23 (30)	31 (10)	81 (69)	70 (70)	58 (96)	85 (78)	106 (67)
Size (cm)							
≤2	46 (60)	155 (52)	19 (17)	36 (37)	29 (49)	72 (66)	0
>2	30 (40)	140 (48)	94 (80)	63 (63)	31 (51)	38 (35)	0
Nodal status							
Negative	57 (75)	151 (51)	34 (29)	46 (47)	28 (47)	75 (69)	0
Positive	19 (25)	144 (49)	79 (67)	53 (53)	25 (42)	29 (27)	0
Tumor grade							
1	9 (11)	75 (25)	11 (10)	16 (17)	3 (5)	31 (29)	28 (18)
2	18 (24)	101 (35)	49 (42)	38 (38)	39 (65)	53 (49)	58 (37)
3	49 (65)	119 (40)	53 (45)	45 (45)	18 (30)	25 (22)	61 (39)
Estrogen receptors							
Negative	28 (37)	69 (24)	31 (27)	34 (35)	1 (1)	16 (15)	29 (19)
Positive	48 (63)	226 (76)	82 (70)	65 (65)	59 (99)	92 (84)	130 (81)
Treatment							
Untreated	76 (100)	165 (55)	22 (19)	11 (11)	0	110 (100)	0
Treated	0	130 (45)	96 (81)	88 (89)	60 (100)	0	48 (31)
Platform	Agilent	Agilent	Stanford Microarray	cDNA NCI	Arcturus	Affymetrix	Affymetrix
Reference	(1)	(2)	(17)	(18)	(19)	(9)	(20)

NOTE: Note that some samples are used in several studies. The following study IDs have samples in common: NKI/NKI2 and UPP/STK/UNT/TBAGD/TBVDX/TAM. For all analyses, we removed duplicated patients from small data sets (e.g., NKI) to avoid decreasing the sample size of large data sets (e.g., NKI2).

subgroups. This was probably due to the relatively small sizes of the individual studies, which would have made these findings statistically unstable.

To address these issues, we have undertaken a large comprehensive meta-analysis integrating both the gene expression and clinicopathologic data of more than 2,100 breast cancer patients. We first defined several gene expression modules associated with key biological processes in breast cancer: proliferation, which was the main component of our recently characterized genomic grade index (7); tumor invasion/metastasis; impairment of immune response; sustained angiogenesis; evasion of apoptosis; self-sufficiency in growth signals; and ER and HER2 signaling. We then sought to elucidate the prognostic information of these gene expression modules and the traditional clinicopathologic parameters used in the clinic in the previously defined molecular breast cancer subtypes according to ER and HER2 module status. We further investigated the biological meaning of individual genes included in several gene expression prognostic signatures by correlating their expression with our molecular modules and explored their prognostic performance across the main breast cancer molecular subtypes.

## Materials and Methods

**Gene expression data and probe annotation.** Gene expression data sets were retrieved from public databases or authors' websites. We used normalized data (log 2 intensity in single-channel platforms or log 2 ratio in dual-channel platforms) as published by the original studies. Hybridization probes were mapped to Entrez GeneID (11) through sequence alignment against RefSeq mRNA in the (NM) subset, as in Shi

et al. (12), using RefSeq and Entrez database version 2007.01.21. When multiple probes were mapped to the same GeneID, the one with the highest variance in a particular data set was selected to represent the GeneID.

**Prototype-based coexpression modules.** We considered a set of prototypes (i.e., genes known to be related to specific biological processes in breast cancer) and aimed to identify the genes that are specifically coexpressed with each of them. To this end, we identified a gene module for prototype  $i$  by selecting the genes whose expression is significantly more correlated to the prototype  $i$  than to the other prototypes. We used the orthogonal Gram-Schmidt variable selection (13) and the Friedman's test to compute the correlations and compare them, respectively. The genes in the module  $i$  are referred to as "specific" to prototype  $i$ . The genes being highly correlated to several prototypes are referred to as "related" to these prototypes. This method was applied in a meta-analytical framework, combining results from NKI2 (2) and VDX (3) data sets (see Table 1).

**Module scores.** For a specific data set, the module score was computed for each sample as

$$\text{module score} = \frac{\sum_i w_i x_i}{\sum_i |w_i|}$$

where  $x_i$  is the expression of a gene in the module that is present in the data set platform, and  $w_i$  is either +1 or -1 depending on the sign of the association with the prototype. Robust scaling was done on each module score to have the interquartile range equal to 1 and the median equal to 0 within each data set, allowing for comparison between module scores.

**Gene ontology and functional analysis.** Gene ontology analyses were done using Ingenuity Pathways Analysis tools<sup>6</sup> (Ingenuity Systems), a web-delivered application that enables the discovery, visualization, and exploration of molecular interaction networks in gene expression data.

<sup>6</sup> <http://www.ingenuity.com>

**Table 1.** Characteristics of the publicly available gene expression data sets (Cont'd)

VDX	VDX2	UNT	UNC	TBAGD	TBVDX	TAM
No. patients (%)						
286	180	78	144	109	198	268
129 (45)	50 (28)	29 (38)	57 (40)	62 (56)	142 (71)	21 (8)
157 (55)	129 (72)	49 (62)	79 (55)	47 (44)	56 (29)	247 (92)
278 (97)	95 (53)	49 (63)	104 (73)	66 (60)	102 (51)	110 (42)
8 (3)	85 (47)	29 (37)	30 (21)	43 (40)	96 (49)	158 (58)
286 (100)	180 (100)	78 (100)	62 (44)	109 (100)	198 (100)	116 (44)
0	0	0	75 (53)	0	0	143 (54)
7 (3)	17 (10)	20 (26)	12 (9)	17 (16)	30 (16)	50 (19)
42 (15)	81 (45)	30 (39)	46 (32)	44 (41)	83 (42)	131 (49)
148 (52)	56 (32)	15 (20)	74 (52)	42 (39)	83 (42)	47 (18)
77 (27)	47 (27)	21 (27)	54 (38)	26 (24)	64 (33)	5 (2)
209 (73)	132 (73)	53 (68)	82 (57)	78 (72)	134 (67)	263 (98)
286 (100)	180 (100)	78 (100)	0	109 (100)	198 (100)	0
0	0	0	144 (100)	0	0	268 (100)
Affymetrix (3)	Affymetrix (4)	Affymetrix (7)	Agilent (21)	Agilent (22)	Affymetrix (23)	Affymetrix (24)

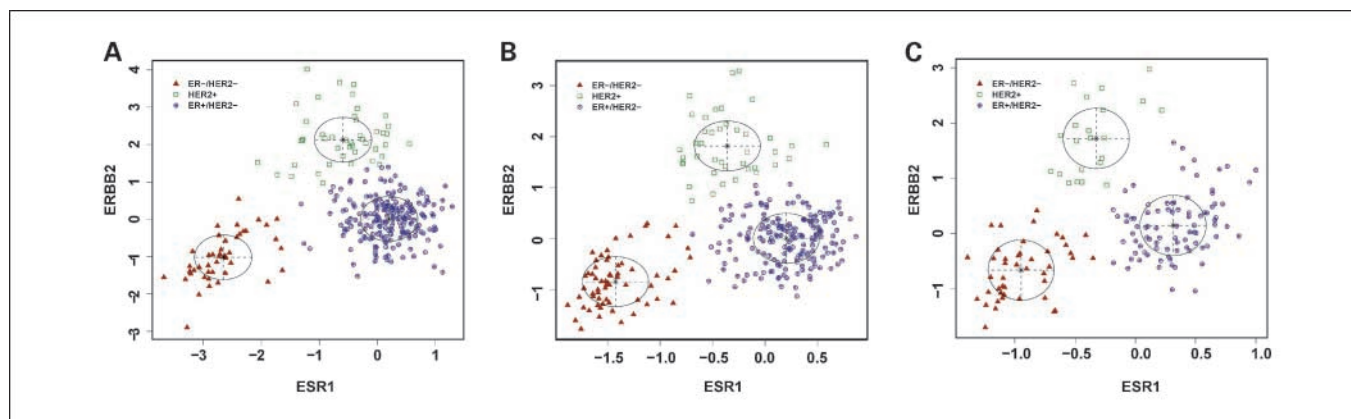
**Clustering.** To consistently identify molecular subgroups across the different data sets, we clustered the tumors using the ESR1 and ERBB2 module scores by fitting Gaussian mixture models (14) with equal and diagonal variance for all clusters. We used the Bayesian Information Criterion (15) to test the number of components. Each tumor was automatically classified to one of the identified molecular subgroups using the maximum posterior probability of membership in the clusters.

**Survival analysis.** We considered relapse-free survival of untreated patients as the survival end point. When relapse-free survival was not available, we used distant metastasis-free survival data. All survival data were censored at 10 years. Survival curves were based on Kaplan-Meier estimates. Hazard ratios (HR) between two or three groups were calculated using Cox regression with the data set as stratum indicator, allowing for different baseline hazard functions between cohorts. For clinical variables and module scores, the HRs were estimated for each data set separately and combined with inverse

variance-weighted method with fixed effect model (16). We used a forward stepwise variable selection in a meta-analytical framework to identify the best multivariate Cox models. The significance thresholds of the combined *P* values (Wald test for HR) for the inclusion of a new variable and for the exclusion of a previously selected variable were set to 0.05.

**Application of the prognostic gene signatures.** When cross-platform mapping was necessary, we only considered genes in the signatures that could be mapped to GeneID. A prediction score was computed for each signature using a linear combination similar to the formula for module score above. Gene-specific weights (coefficients, correlations, or other measures) from the original studies were converted into +1 or -1 depending on the original up-regulation or down-regulation of each gene. Robust scaling was done on each gene signature to allow for comparison between the different gene signatures.

For a more detailed description of the methods, see Supplementary Methods.



**Fig. 1.** Joint distribution between the ER and HER2 module scores for three example data sets: NKI2 (A), UNC (B), and VDX (C). Clusters are identified by Gaussian mixture models with three components. The ellipses shown are the multivariate analogs of the SDs of the Gaussian of each cluster.

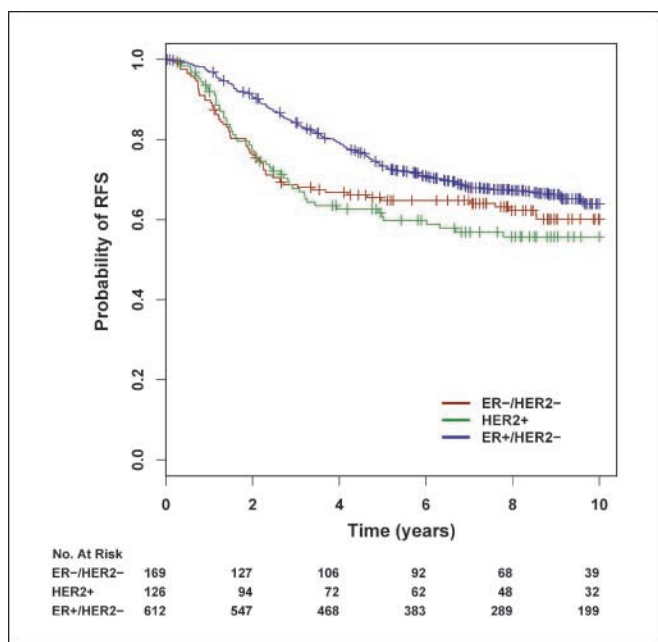


Fig. 2. Survival curves for untreated patients stratified by molecular subtypes ER-/HER2-, HER2+, and ER+/HER2-.

### Results

**Definition of the gene expression modules of breast cancer.** To develop the gene expression modules, we first selected genes to act as “prototypes” for each biological process, based on the literature. We then applied a comparison of linear models to generate modules of genes specifically associated with each of the prototype genes underlying different biological processes in breast cancer. The selected prototype genes were *AURKA* (also known as *STK6*, *STK7*, or *STK15*), *PLAU* (also known as *uPA*), *STAT1*, *VEGF*, *CASP3*, *ESR1*, and *ERBB2*, representing the proliferation, tumor invasion/metastasis, immune response,

angiogenesis, apoptosis phenotypes, and the ER and HER2 signaling, respectively.

These lists of genes are available as Supplementary Table S1. The main characteristics of these molecular modules are that they are identified as genes that are coexpressed consistently with the chosen prototypes in data sets using Agilent and Affymetrix microarray platforms and that they are identified without looking at clinical variables and gene annotations.

The seven gene lists were uploaded into the Ingenuity Pathway Knowledge Database for the analysis of functional annotations. The detailed results can be found in Supplementary Table S2. In brief, as expected, many genes included in these modules were known to be associated with the chosen biological process.

A module score was then defined by the difference of the sums of the positively and negatively correlated genes for the chosen prototype only. We then mapped and computed each of these module scores on several published microarray data sets (1–4, 7, 9, 17–24), totaling to more than 2,100 tumor samples (Table 1).

**Identification of the ER-/HER2-, ER+/HER2-, and HER2+ subgroups.** In our analysis, we reexamined the molecular classification of breast cancer based on ER and HER2 module scores. Interestingly, when the ER and HER2 module scores were combined, the tumor samples grouped into three main clusters instead of the four that would have been observed had the two scores been completely independent. These groups represented the ER-/HER2-, HER2+, and ER+/HER2- tumors, corresponding roughly to the intrinsic basal-like, HER2, and combined luminal A/B subtypes, respectively, as defined by the Stanford group (17, 25, 26). Notably, the HER2+ cluster showed intermediate levels of ER module scores, highlighting the intrinsic biological differences that characterize this molecular subtype with respect to ER signaling (Fig. 1; Supplementary Fig. S3).

The clinicopathologic characteristics per molecular subgroup are illustrated in Supplementary Table S5. As one would expect, the vast majority of the tumors in the ER-/HER2- and ER+/HER2- subgroups were negative and positive, respectively, with respect to ER protein status, in contrast to the HER2+ subgroup that comprised a mixture of tumors.

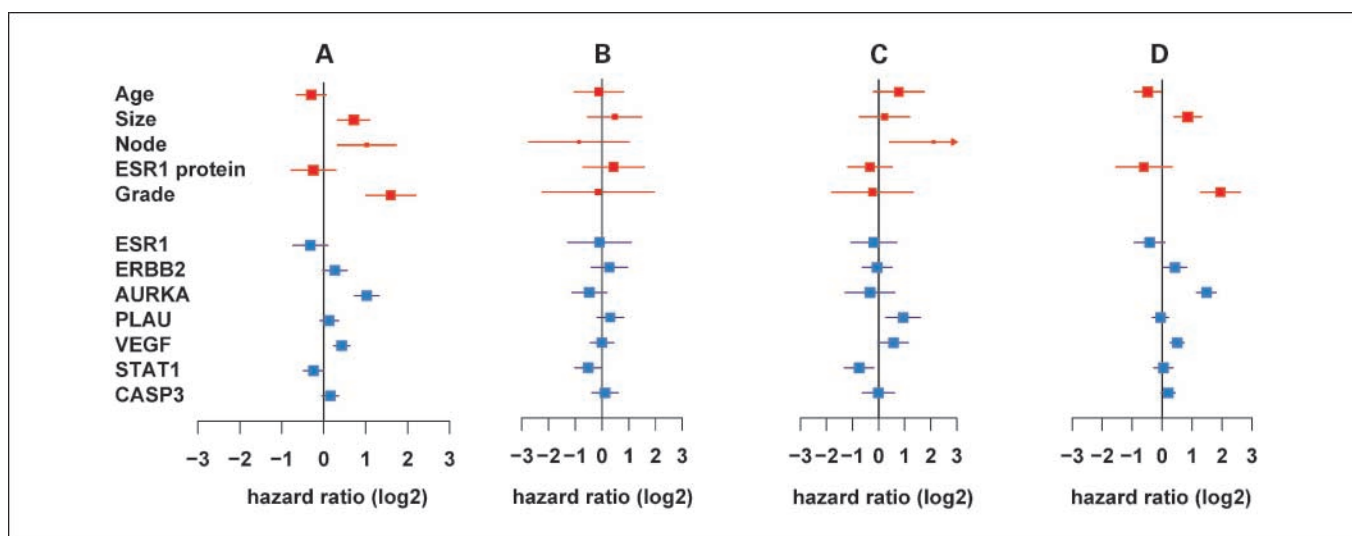
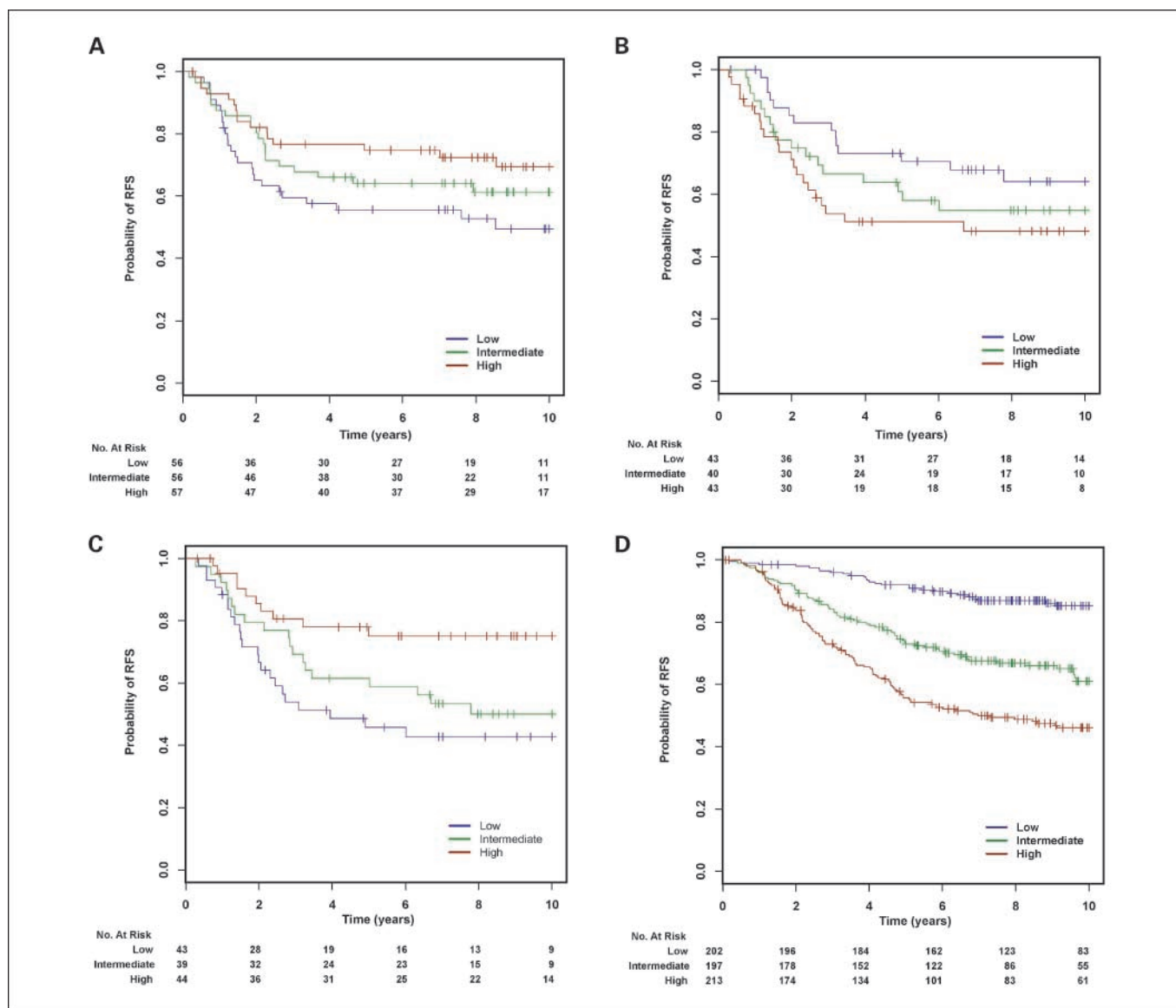


Fig. 3. Forest plots showing the log<sub>2</sub> HRs (and 95% CI) of the univariate survival analyses in the global population (A) and in the ER-/HER2- (B), HER2+ (C), and ER+/HER2- (D) subgroups of untreated breast cancer patients. All clinical indicators are considered as discrete variables (see Table 1) and all modules are considered as continuous variables.



**Fig. 4.** Kaplan-Meier curves of the module scores that were significant in the univariate analysis in the molecular subgroup analysis. The module scores were split according to their 33% and 66% quantiles: STAT1 module in the ER-/HER2- subgroup (A), PLAU module in the HER2+ subgroup (B), STAT1 module in the HER2+ module (C), and AURKA module in the ER+/HER2- subgroup (D).

Interestingly, these molecular subtypes showed differences in clinical outcomes. Indeed, as shown in Fig. 2, the survival curve for relapse-free survival of the ER+/HER2- subgroup was significantly different from the two others ( $P = 0.03$  for ER-/HER2- and  $P = 0.003$  for HER2+). However, no difference in survival was noticed between the ER-/HER2- and HER2+ subgroups ( $P = 0.56$ ).

**Prognostic value of the gene expression module scores according to breast cancer subgroups based on the ER and HER2 module scores.** We also analyzed the prognostic value of the gene expression modules and currently used clinicopathologic parameters according to the breast cancer subgroups defined above.

Interestingly, in the high-risk ER-/HER2- subpopulation ( $n = 189$ ), only the immune response module showed a significant association with clinical outcome in both univariate

and multivariate analyses {HR, 0.70 [95% confidence interval (95% CI), 0.50-0.98];  $P = 0.04$ ; Figs. 3 and 4; Supplementary Table S3}.

In the ER+/HER2- subpopulation ( $n = 628$ ), age, tumor size, and histologic grade were associated with relapse-free survival, together with the HER2, proliferation, and angiogenesis modules. In multivariate analysis, only the proliferation module [HR, 2.68 (95% CI, 2.02-3.55);  $P = 9 \times 10^{-12}$ ] and histologic grade [HR, 2.00 (95% CI, 1.18-3.37);  $P = 0.01$ ] remained significant, with the proliferation module having the highest HR and the most significant  $P$  value.

In the HER2+ tumors ( $n = 129$ ), nodal status, tumor invasion, angiogenesis, and immune response module scores were significantly associated with relapse-free survival in the univariate model, whereas only tumor invasion [HR, 2.07 (95% CI, 1.32-3.25);  $P = 0.001$ ] and immune response [HR,



**Table 2.** Univariate analysis of different gene classifiers per molecular subgroup of untreated breast cancer patients

	ER-/HER2-			HER2+			ER+/HER2-		
	HR (95% CI)	P	No. patients	HR (95% CI)	P	No. patients	HR (95% CI)	P	No. patients
GENE70	1.12 (0.73-1.72)	0.60	154	1.29 (0.75-2.20)	0.36	120	2.11 (1.67-2.66)	$3 \times 10^{-10}$	566
GENE76	1.30 (0.78-2.15)	0.32	99	0.81 (0.49-1.34)	0.42	85	1.52 (1.24-1.88)	$2 \times 10^{-5}$	422
P53	1.01 (0.42-2.42)	0.98	163	1.04 (0.51-2.11)	0.92	126	2.23 (1.64-3.03)	$4 \times 10^{-7}$	605
WOUND	0.90 (0.65-1.26)	0.54	160	1.24 (0.79-1.93)	0.35	126	1.48 (1.25-1.75)	$5 \times 10^{-6}$	598
GGI	0.78 (0.44-1.36)	0.38	165	0.79 (0.40-1.53)	0.48	126	3.16 (2.46-4.06)	$2 \times 10^{-19}$	598
ONCOTYPE	0.86 (0.36-2.08)	0.74	156	1.00 (0.50-2.02)	1.00	126	4.79 (3.43-6.68)	$3 \times 10^{-20}$	605
IGS	1.08 (0.73-1.61)	0.70	169	0.96 (0.63-1.46)	0.85	126	2.12 (1.73-2.60)	$6 \times 10^{-13}$	605

NOTE: All signatures are considered here as continuous variables. GENE70, 70-gene signature (1, 2); GENE76, 76-gene signature (3, 4); P53, p53 signature (9); WOUND, Wound response signature (5, 6); GGI, Genomic Grade Index (7); ONCOTYPE, 21-gene Recurrence Score (8); IGS, 186-gene "invasiveness" gene signature (10).

0.56 (95% CI, 0.36-0.86);  $P = 0.009$ ] modules remained significantly associated with relapse-free survival in the multivariate model.

**Dissecting prognostic gene expression signatures using gene expression modules.** To gain further insight into the biological significance of individual genes included in several published gene prognostic signatures (1–10), we investigated their correlation with our gene expression module prototypes. Supplementary Table S4 illustrates the percentage of genes of each signature related to or specifically associated (value in brackets) with a chosen prototype. Our analysis showed that more than half of the genes in each signature investigated in this study were related to the proliferation prototype. Moreover, the highest percentages of specific association (i.e., association with one prototype but not with the others) were also reported for AURKA, highlighting the importance of proliferation genes in several prognostic signatures.

**Evaluating the effect of the prognostic signatures according to different breast cancer subgroups.** We also examined the prognostic performance of several previously published gene prognostic signatures (1–10) according to different breast cancer subgroups based on the ER and HER2 gene expression module scores. Because the exact algorithms for generating the different gene signatures cannot be applied on different microarray platforms, we computed these signatures using the direction of the association reported in the respective original publications. Concerned that a signed average might be less efficient than the original algorithm used for the development of these prognostic signatures, we performed comparison studies and found that the original and modified scores were highly correlated with similar performances (see Supplementary data for details).

Intriguingly, as shown in Table 2, all seven prognostic signatures tested in this study were highly informative with regard to discriminating good versus poor clinical outcome patients in the ER+/HER2- subgroup only, whereas they were much less informative for the ER-/HER2- and HER2+ subgroups.

## Discussion

To reveal the thread connecting molecular subtyping, gene expression prognostic signatures, and conventional clinicopathologic prognostic factors, we introduced the concept of

gene expression modules (comprehensive lists of genes with highly correlated expression) associated with key biological processes in breast cancer tumorigenesis. Wishing to extend our previous results on the genomic grade signature (7), capturing mainly proliferation, we added into the model several other expression modules representing key biological processes in breast cancer such as proliferation, tumor invasion, immune response, angiogenesis, apoptosis, and estrogen and HER2 signaling.

In this study, we showed that these modules contain distinct prognostic information according to different breast cancer subtypes based on ER and HER2 module scores, and we highlighted the importance of proliferation-related genes in predicting clinical outcome in breast cancer. Furthermore, we showed that several previously published gene expression prognostic signatures discriminate poor versus good clinical outcome patients in the ER+/HER2- subtype only.

In the ER+/HER2- subgroup, proliferation module and histologic grade were the two variables that remained associated with survival in the multivariate analysis, with the proliferation module having the most significant  $P$  value. This is consistent with our finding that two clinically distinct ER-positive molecular subgroups can be defined by genomic grade, which captures mainly proliferation (24).

In the HER2+ subgroup, tumor invasion and immune response seemed to be the main processes associated with poor clinical outcome. This finding supports the recent report by Urban et al. (27), which highlighted that mRNA expression of PLAU was a powerful prognostic indicator in HER2-positive tumors.

In the ER-/HER2- subgroup, only immune response seemed to predict prognosis. It has been reported that tumors that do not express the hormone receptors and ERBB2, commonly called the "triple-negative" or "basal-like" tumors, are more aggressive. Given their triple negative status, these patients cannot be treated with the conventional targeted therapies currently available for breast cancer, such as endocrine or ERBB2 therapies, leaving chemotherapy as the only option.

In this context, several authors have suggested that chemotherapy might be the most efficient in this disease subtype (28, 29). However, defining the optimal chemotherapy regimen remains controversial. Because BRCA1 pathway activity seems to be impaired in many of these tumors and

because BRCA1 functions in DNA repair and cell cycle checkpoints, some authors have suggested that these tumors might be associated with sensitivity to DNA-damaging chemotherapy and with resistance to spindle poisons (30).

In this study, we showed that in this triple-negative subgroup of patients, impaired immune response might be linked with the development of distant metastases. Indeed, high expression levels of the immune module were associated with a significantly better outcome, both at the univariate and multivariate levels. Interestingly, Teschendorff et al. (31) recently published similar findings.

It has been shown that signal transducer and activator of transcription 1 (STAT1) is particularly important in activating IFN- $\gamma$  and its antitumor effects. In addition to inhibiting proliferation and survival, IFN- $\gamma$  enhances the immunogenicity of tumor cells, in part, by enhancing the STAT1-dependent expression of MHC proteins (32). Based on this observation and the fact that an attenuated STAT1 signaling in tumors might be correlated with their malignant behavior, Lynch et al. (33) recently postulated that enhancing gene transcription mediated by STAT1 may be an effective approach to cancer therapy. They screened 5,120 compounds and identified one molecule, 2-(1,8-naphthyridin-2-yl)phenol, which enhanced STAT1-mediated gene activation more so than seen with a maximally efficacious concentration of IFN. Because STAT1 activation seems to play an important role in the killing of tumor cells in response to cytotoxic agents through repression of pro-survival genes and activation of apoptosis genes, it may be particularly important for patients receiving chemotherapy and particularly for these ER-/HER2- patients for whom most therapeutic approaches rely on cytotoxic agents that induce cell death in a nonspecific manner.

Our study also highlighted that proliferation-related genes are the main and common denominator for predicting clinical outcome of several previously published gene expression signatures. Because defects in cell cycle deregulation are a fundamental characteristic of breast cancer, it is not surprising that these genes are involved in breast cancer prognosis. Several studies have indeed shown that increased expression of cell-cycle- and proliferation-associated genes was correlated with poor outcome (reviewed in ref. 34). There are of course differences in the exact proliferation-associated genes due to the difference in population analyzed or platform used. Although the use of proliferation-associated cell markers is not new (the protein expression levels of Ki67 and proliferating cell nuclear antigen have already been used

as prognostic markers for decades), gene expression profiling studies suggest that measuring proliferation with a more objective, automated, and quantitative assay may be more robust than less quantitative assays such as immunohistochemistry.

Interestingly, we have also shown that the prognostic abilities of several prognostic gene signatures differ according to the breast cancer subtypes. Indeed, we showed that their prognostic discriminative power was limited essentially to the ER+/HER2-molecular subgroup composed of ER-positive patients only. This concurs with our findings that (a) proliferation seems to be the main contributor of these signatures and (b) the ER+/HER2- subgroup is the only molecular subgroup that displays a wide range of proliferation values. However, Ma et al. (35) recently showed that their two-gene ratio HOXB13/IL17BR provided prognostic information in ER+ breast cancer patients in addition to and independent of the tumor grade, suggesting that a biological process other than proliferation might be associated with prognosis in that particular subgroup of patients.

These findings thus emphasize the need for additional prognostic markers for the other two molecular subgroups, and more specifically for the ER-/HER2- subgroup, which is associated with poor prognosis and limited therapeutic options. Therefore, we strongly believe that studying the immune response mechanisms in this particular subgroup of patients might help to better understand these tumors and to develop efficient novel targeted therapies.

To conclude, by identifying gene expression modules representing key biological mechanisms involved in breast cancer, we were able to better characterize the biological foundation of various prognostic signatures and to understand the mechanisms that trigger the tumors of different subgroups to progress. These findings may help to define new clinicogenomic models and identify new targets in the molecular subgroups to further advance toward truly personalized medicine.

### Disclosure of Potential Conflicts of Interest

C. Sotiriou, M. Delorenzi, and M. Piccart are named inventors on a patent application for the Genomic Grade signature used in this study.

### Acknowledgments

We thank C. Straehle for technical assistance.

### References

- 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.
- 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.
- Foekens JA, Atkins D, Zhang Y, et al. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. *J Clin Oncol* 2006;24:1665–71.
- 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.
- 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.
- Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006;98:262–72.
- Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–26.
- Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* 2005;102:13550–5.
- Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 2007;356:217–26.
- Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 2007;35(Database issue):D26–31.
- MAQC Consortium, Shi L, Reid LH, et al. The Micro-Array Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;9:1151–61.
- Chen S, Billings SA, Luo W. Orthogonal least

- squares methods and their application to non-linear system identification. *Proc Natl Acad Sci U S A* 1989; 30:1873–96.
14. McLachlan G, Peel D. *Finite mixture models*. New York: John Wiley & Sons; 2000. p. 419.
  15. Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978;6:461–4.
  16. Cochrane WG. Problems arising in the analysis of a series of similar experiments. *J Roy Stat Soc* 1937;4: 102–18.
  17. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003; 00:8418–23.
  18. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 2003;100:10393–8.
  19. Ma XJ, Wang Z, Ryan PD, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 2004;6: 607–16.
  20. Pawitan Y, Bjohle J, Amler L, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 2005;6:R953–64.
  21. Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol* 2006;24:1656–64.
  22. Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006;98:1183–92.
  23. Desmedt C, Piette F, Loi S, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 2007;13:3207–14.
  24. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 2007;25:1239–46.
  25. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumors. *Nature* 2000;406: 747–52.
  26. 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.
  27. Urban P, Vuaroqueaux V, Labuhn M, et al. Increased expression of urokinase-type plasminogen activator mRNA determines adverse prognosis in ErbB2-positive primary breast cancer. *J Clin Oncol* 2006;24: 4245–53.
  28. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* 2005;11: 5678–85.
  29. Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res* 2007;13: 2329–34.
  30. Kennedy RD, Quinn JE, Mullan PB, Johnston PG, Harkin DP. The role of BRCA1 in the cellular response to chemotherapy. *J Natl Cancer Inst* 2004;96: 1659–68.
  31. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol* 2007; 8:R157.
  32. Muhlethaler-Mottet A, Di Berardino W, Otten LA, Mach B. Activation of the MHC class II transactivator CIITA by interferon- $\gamma$  requires cooperative interaction between Stat1 and USF-1. *Immunity* 1998;8: 157–66.
  33. Lynch RA, Etchin J, Battle TE, Frank DA. A small-molecule enhancer of signal transducer and activator of transcription 1 transcriptional activity accentuates the antiproliferative effects of IFN- $\gamma$  in human cancer cells. *Cancer Res* 2007;67:1254–61.
  34. Colozza M, Azambuja E, Cardoso F, Sotiriou C, Larsimont D, Piccart MJ. Proliferative markers as prognostic and predictive tools in early breast cancer: where are we now? *Ann Oncol* 2005;11: 1723–39.
  35. Ma XJ, Salunga R, Dahiya S, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. *Clin Cancer Res* 2008;14:2601–8.