Imaging, Diagnosis, Prognosis

Clinical Cancer Research

Elucidating Prognosis and Biology of Breast Cancer Arising in Young Women Using Gene Expression Profiling

Hatem A. Azim Jr[1], Stefan Michiels[1], Philippe L. Bedard[3], Sandeep K. Singhal[1], Carmen Criscitiello[1], Michail Ignatiadis[1], Benjamin Haibe-Kains[4], Martine J. Piccart[2], Christos Sotiriou[1], and Sherene Loi[1]

Abstract

Purpose: Breast cancer in young women is associated with poor prognosis. We aimed to define the role of gene expression signatures in predicting prognosis in young women and to understand biological differences according to age.

Experimental Design: Patients were assigned to molecular subtypes [estrogen receptor (ER)$^+$/HER2$^-$; HER2$^+$, ER$^-$/HER2$^-$)] using a three-gene classifier. We evaluated whether previously published proliferation, stroma, and immune-related gene signatures added prognostic information to Adjuvant! online and tested their interaction with age in a Cox model for relapse-free survival (RFS). Furthermore, we evaluated the association between candidate age-related genes or gene sets with age in an adjusted linear regression model.

Results: A total of 3,522 patients (20 data sets) were eligible. Patients aged 40 years or less had a higher proportion of ER$^-$/HER2$^-$ tumors ($P < 0.0001$) and were associated with poorer RFS after adjustment for breast cancer subtype, tumor size, nodal status, and histologic grade and stratification for data set and treatment modality (HR = 1.34, 95% CI = 1.10–1.63, $P = 0.004$). The proliferation gene signatures showed no significant interaction with age in ER$^+$/HER2$^-$ tumors after adjustment for Adjuvant! online. Further analyses suggested that breast cancer in the young is enriched with processes related to immature mammary epithelial cells (luminal progenitors, mammary stem, *c-kit*, *RANKL*) and growth factor signaling in two independent cohorts ($n = 1{,}188$ and 2,334).

Conclusions: Proliferation-related prognostic gene signatures can aid treatment decision-making for young women. However, breast cancer arising at a young age seems to be biologically distinct beyond subtype distribution. Separate therapeutic approaches such as targeting RANKL or mammary stem cells could therefore be needed. *Clin Cancer Res; 18(5); 1341–51. ©2012 AACR.*

Introduction

Around 7% of patients in the developed world and 25% of patients in the developing world are diagnosed with breast cancer below the age of 40 (1, 2). These women have poorer survival and higher risk of relapse than their older counterparts (3, 4). Several factors have been linked to the poor prognosis associated with developing breast cancer at a young age. These include large tumor size at diagnosis, higher tumor grade, mitotic rate, lymphovascular invasion, increased expression of HER2, and lower estrogen and progesterone receptor expression (4, 5). However, even after correction for stage and tumor characteristics, young age at diagnosis remains an independent risk factor for relapse and breast cancer–related death (6–10) and also an indication for aggressive systemic therapy (6).

Using gene expression profiling, at least 3 distinct molecular subtypes have been observed, each associated with different clinical outcome: "luminal", HER2$^+$, and "basal-like" breast cancer (11–14). Recent studies have examined the distribution of breast cancer molecular subtypes and found that breast cancer in young women is enriched with aggressive subtypes (15–17). This has led some investigators to question whether breast cancer diagnosed at a young age has a unique biology or whether it is a just a surrogate for a higher incidence of aggressive molecular subtypes (15).

The use of gene expression profiling in breast cancer has also contributed to the development of biomarkers for improved prognostication (18–23). Several multigene expression signatures have been reported to provide a better indication of clinical outcome than the traditional clinical

Authors' Affiliations: [1]Breast Cancer Translational Research Laboratory (BCTL) J.C. Heuson, Institut Jules Bordet, Université Libre de Bruxelles; [2]Department of Medical Oncology, Institut Jules Bordet, Brussels, Belgium; [3]Division of Medical Oncology and Hematology, Princess Margaret Hospital, University of Toronto, Toronto, Canada; and [4]Computational Biology and Functional Genomic Laboratory, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston

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

Corresponding Author: Sherene Loi, Breast Cancer Translational Research Laboratory (BCTL) J.C. Heuson, Institut Jules Bordet, Boulevard de Waterloo, 125, 1000 Brussels, Belgium. Phone: 32-(0)2-541-34-57; Fax: 32-(0)2-541-33-39; E-mail: sherene.loi@bordet.be

doi: 10.1158/1078-0432.CCR-11-2599

Translational Relevance

Young age (i.e., $\leq$40 years) at breast cancer diagnosis has long associated with a poor prognosis. In this work, we address two key questions: (i) whether prognostic gene expression signatures can also discriminate prognostic subgroups in young patients, as young age alone can be an indicator for adjuvant chemotherapy and (ii) whether breast cancers arising in this age group are biologically distinct. We report that proliferation-related prognostic gene signatures retain their prognostic power independent of age—these results are important because although their clinical utility is currently being evaluated in prospective trials, many young patients are not likely to be recruited in these studies. Next, we report that breast cancer arising at a young age is enriched with unique molecular processes that may explain their poor outcomes. These data, from more than 3,500 patients, provide further rationale for investigating separate therapeutic approaches for breast cancer diagnosed in young women.

and pathologic standards (19, 20, 24). Some of these gene signatures are currently commercially available as diagnostic tests, though prospective clinical trials are ongoing to evaluate their exact clinical utility (25, 26).

In this study, we aimed to conduct a comprehensive analysis of breast cancer with respect to age, taking advantage of the large compendium of publicly available gene expression data sets from more than 3,500 breast cancer patients. We specifically wanted to clarify the relevance of several published prognostic gene signatures in young women ($\leq$40) and to determine whether young age is truly associated with unique disease biology.

Materials and Methods

Patient clinical and gene expression data

We searched 39 publically available data sets and retrieved all clinical and gene expression data. We excluded all patients with missing information on age and cross-checked the different data sets to identify patients who were included in more than 1 data set. Repeated patients were deleted. As stroma and lymphocyte content are known to be altered in different tissue sampling procedures, those neoadjuvant data sets in which fine needle aspirates were known to have been used, were also excluded (27; Fig. 1). Eligible patients were divided into 2 cohorts according to whether they had received systemic adjuvant therapy or not ("untreated," cohort 1 and "treated," cohort 2; Supplementary Table S1). We used normalized microarray 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 with Entrez database version 2007.01.21. When multiple probes were mapped to the same GeneID, the probe with the highest variance in a particular data set was selected to represent the GeneID.

We calculated the estimated 10-year relapse-free survival (RFS) for the untreated cohort with Adjuvant! Online (AOL, version 8.0). AOL (http://www.adjuvantonline.com) was calculated for patients with available tumor size and nodal status. Patients with unknown histologic grade were considered to have "undefined" histologic grade on the AOL risk assessment model. Estrogen receptor (ER) status was missing in 12 patients (1%), and in these cases the dichotomized ER gene (*ESR1*) mRNA value (positive and negative) was used instead. For patients with node-positive disease, we searched the original publication to retrieve the exact number of positive nodes to accurately estimate AOL risk. When this information was unavailable, we classified them as N1 (i.e., node positive with 1–3 positive nodes).

Microarray analysis

Molecular subtypes. Patients were assigned to molecular subtypes with a 3-gene classifier (*ESR1*, *ERBB2 [HER2]*, and *AURKA*). The cutoff for *ESR1* and *ERBB2* expression from microarray data was derived from fitting 2 normal distributions to the observed distribution of expression values in a single training study of 286 patients (23) which was consequently applied to all other data sets, using a previously described method (28). Three main molecular subtypes were defined as $ER^-/HER2^-$, $HER2^+$, and $ER^+/HER2^-$ (i.e., luminal). The luminal subtype group was also divided into phenotypes "A" and "B" on the basis of median *AURKA* expression—$ER^+/HER2^-$/low proliferation (luminal A) and $ER^+/HER2^-$/high proliferation (luminal B). For this purpose, similarly the *AURKA* cutoff was trained on the same data set and then applied to all the other data sets as previously described (28, 29). To ensure compatibility of expression values across multiple data sets, *ESR1*, *ERBB2*, and *AURKA* gene expression values were rescaled before applying the 3-gene classifier (see below). This method is fully documented and was implemented with the R/Bioconductor package genefu (version 1.3.6; http://www.bioconductor.org/packages/release/bioc/html/genefu.html).

Evaluation of previously published prognostic gene signatures according to age. For this analysis, we used the untreated cohort (cohort 1) as not to have treatment as a confounder on prognostic outcomes. In the 3 breast cancer subtypes ($ER^+/HER2^-$; $HER2^+$, $ER^-/HER2^-$), within the overall patient series and in the different age groups ($\leq$40, 41–52, 53–64, $\geq$65 years), we evaluated (i) 3 proliferation-related prognostic gene signatures (GGI, GENE70, and GENE76; refs. 18, 19, 23); (ii) 3 stroma-related gene signatures (DCN, SDPP, and PLAU; refs. 29–31); and (iii) 3 immune-related gene signatures (IRM, immunomodulatory cluster, STAT1; refs. 29, 32, 33). A summary of the 9 evaluated signatures is provided in Supplementary Table S2.

Evaluation of candidate age-related genes and gene sets identified from the literature. We conducted a MEDLINE search using the terms "breast cancer, young, biology,"

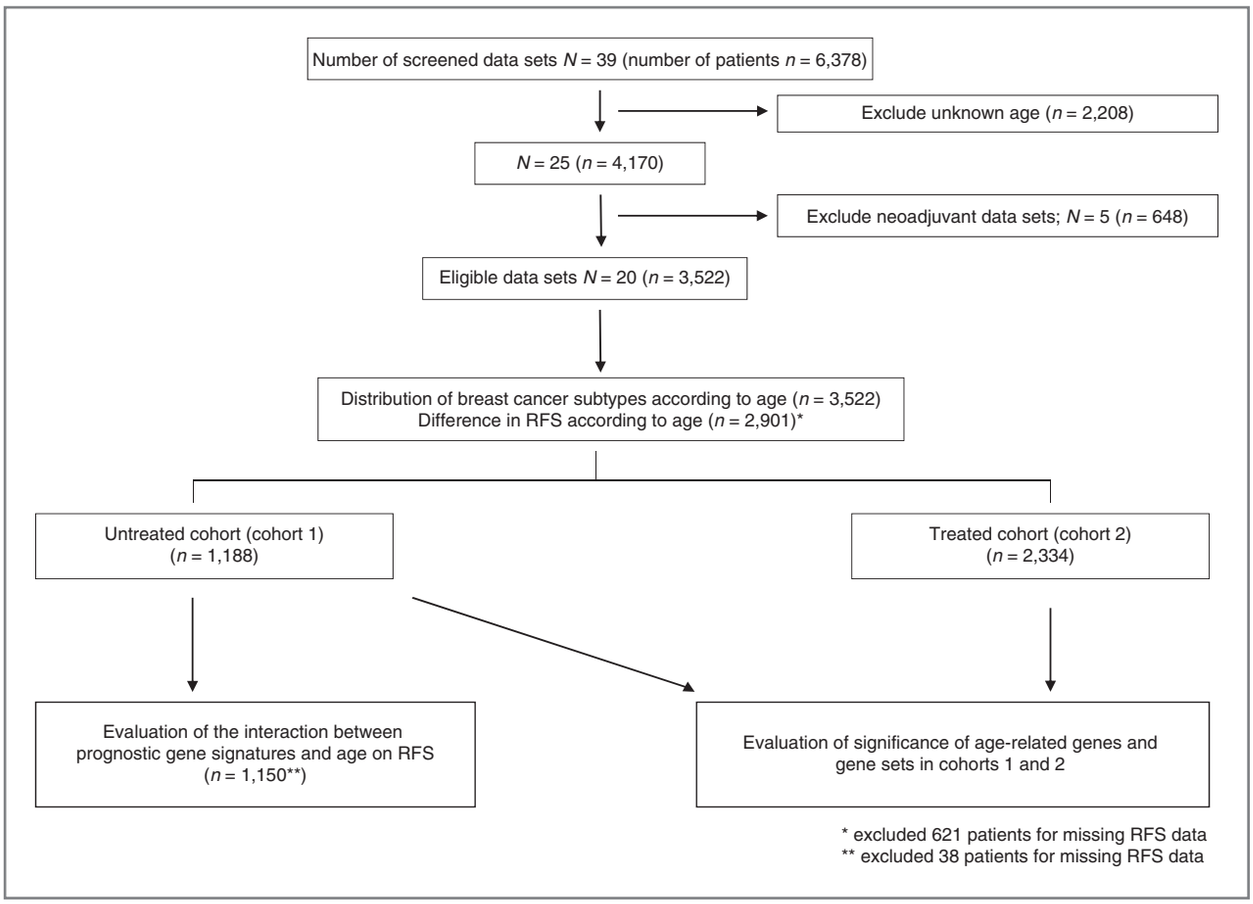

Figure 1. Flow diagram summarizing the gene expression data sets used for the various analyses.

"breast cancer, young, gene expression," and "breast cancer, young, prognosis" to retrieve publications related to the biology of breast cancer in young women until February 2011. Any gene or protein alterations that were suggested to be related to the biology of breast cancer in young women were identified. We then evaluated the expression of these genes and gene sets in a linear regression as a function of age and after adjustment for potential confounders to determine whether breast cancer arising in the young is associated with unique cancer biology (see below).

Statistical analysis

The different prognostic gene signatures and age-related gene sets were treated as continuous variables, and they were defined as a weighted average of the included genes using the following formula:

$$\text{Risk score} = \frac{\Sigma_i w_i x_i}{\Sigma_i |w_i|}$$

where x_i is the expression of a gene in the gene set or gene signatures that is present in the data set platform and w_i is either $+1$ or -1 depending on the sign of gene-specific statistic from the original studies. Each risk score was scaled such that quantiles 2.5% and 97.5% equaled -1 and $+1$, respectively, to allow for comparison between data sets using different microarray technologies and normalization procedures.

RFS was the primary survival endpoint, which is defined as the time elapsing between breast cancer diagnosis and date of local or systemic relapse, or death. RFS was evaluated with a Cox regression model stratified by data set and adjuvant treatment modality (hormonal only, chemotherapy, or no therapy), and adjusted for age as a binary ($\leq$40 vs. >40 years) or continuous variable, the 3 breast cancer subtypes, tumor size, nodal status, and histologic grade. Survival plots according to the age groups were drawn using the Kaplan–Meier method, and the differences were evaluated with a log-rank test. Only patients with relapse information available were included in these analyses. When RFS data were not reported, distant metastasis-free survival (DMFS) information was used if available. The median follow-up was calculated with the reversed Kaplan–Meier method (34).

Prognostic gene signatures were evaluated for their ability to provide further prognostic information to AOL. The additional prognostic value of the gene signatures to AOL

was assessed using the change in the likelihood ratio χ^2 value. We also examined whether there was any interaction between age as a continuous variable and the prognostic performance of the different gene signatures across the breast cancer subtypes.

Differences in the incidence of breast cancer subtypes according to age group were assessed by the χ^2 test. To evaluate the association between candidate age-related genes and gene sets, we built a linear regression model for each candidate gene expression score as a function of age as a continuous variable, after controlling for potential confounding factors. The first set of variables entered into the model was age and data set followed by the second set, adding histologic grade, tumor size (<2 cm, 2–5 cm, and >5 cm), nodal status (positive and negative), and the 3 main breast cancer molecular subtypes. The model was applied to the untreated cohort first (i.e., cohort 1), and then we attempted to replicate the significant findings in the treated one (i.e., cohort 2).

To visualize the differences in the prognostic performance of the gene signatures across age, distribution of breast cancer molecular subtypes and breast cancer biology, patients were divided into 4 age groups ($\leq$40, 41–52, 53–64, $\geq$65 years).

As each of the analyzed data sets represented a series of patients from different hospitals and countries, treated heterogeneously, with varying sample collection techniques and profiled on different platforms, we opted to adjust the linear regression analysis for data set and stratify the RFS analysis by data set to avoid potential biases. To control for multiple testing, we used a false discovery rate (FDR) approach as defined by Benjamini and colleagues (35). Reported *P* values are 2-sided. Statistical analyses were conducted with SPSS (version 15.0; SPSS Inc.) and R software (version 2.9.2; http://www.r-project.org).

Results

Patient characteristics

We compiled clinical and gene expression profiling data from 39 published data sets of early breast cancer. After exclusion of patients with missing information on age and those who received neoadjuvant treatment, a total of 3,522 patients from 20 data sets were eligible for this study (Fig. 1; Supplementary Table S1).

When comparing the distribution of breast cancer subtypes across the whole series, as expected, patients aged 40 years or less had nearly doubled the proportion of ER^-/$HER2^-$ tumors (34.3% vs. 17.9%) and half the luminal-A breast cancer than the oldest group (i.e., $\geq$65 years; 17.2% vs. 35.4%, $P < 0.0001$; Supplementary Fig. S3).

We used the untreated cohort (cohort 1) to evaluate the performance of the previously published prognostic gene signatures across age to avoid treatment as a confounder of clinical outcome. Both untreated (cohort 1: $n = 1{,}188$; Table 1) and treated cohorts (cohort 2: $n = 2{,}334$; Table 1) were used to examine the biological differences according to age.

Differences in RFS according to age and breast cancer molecular subtype

Out of 3,522 patients identified for this study, 621 (17.6%) did not have information available on RFS, and thus were not included in this analysis. Of the remaining 2,901 patients, 1,697 patients (52.8%) had DMFS and not RFS available, and hence DMFS values were used in these patients. A total of 952 patients relapsed (33%) at a median follow-up of 5.2 years (interquartile range 1.5–8.6 years). We observed a significantly higher risk of relapse in patients of 40 years or less than in older age groups ($P < 0.0001$, Fig. 2A). As a binary variable, age less than or equal to 40 years was significantly associated with a poor outcome after adjustment compared with ages older than 40 at diagnosis (HR = 1.34, 95% CI = 1.10–1.63, $P = 0.004$). A subgroup analysis per breast cancer molecular subtype suggested an inferior RFS with young age, particularly in the ER^+/$HER2^-$ phenotypes (Fig. 2B–D). Similar results were observed on restricting the analysis to the untreated cohort (Supplementary Fig. S4).

In a multivariate analysis stratified for data set and treatment, and adjusted for tumor size, nodal status, histologic grade, and breast cancer subtype, a 1-year increase in age was associated with a 1% reduction in the risk of relapse (HR = 0.99, 95% CI = 0.98–0.99; $P = 0.043$).

Clinical relevance of previously published prognostic gene signatures according to age and breast cancer subtype

The prognostic value of 3 proliferation related, 3 stroma related, and 3 immune-related gene signatures was evaluated according to age group and breast cancer subtype in the untreated cohort (cohort 1). All analyses were adjusted to the estimated risk of relapse by AOL at 10 years, which was computed for all eligible patients ($n = 1{,}150$). AOL could not be precisely calculated in 163 (13.7%) patients as 116 (9.8%) had an undefined histologic grade (though AOL adjusts its risk assessment for missing histologic grade values) and 47 (3.9%) had an unknown number of positive lymph nodes. One patient had both variables missing.

As expected, proliferation-related gene signatures did not add any prognostic information in either ER^-/$HER2^-$ or $HER2^+$ breast cancer (33). In patients with ER^+/$HER2^-$ breast cancer, we did not observe any significant interaction between the prognostic performance of proliferation-related gene signatures and age after adjusting for AOL risk score [GENE70: $P_{\text{interaction}} = 0.56$; GGI: $P = 0.71$; GENE76: $P = 0.56$]. In other words, all proliferation gene signatures added statistically significant prognostic information to AOL irrespective of age with HRs in the overall population: for GENE 70, HR = 3.5, 95% CI = 2.6–4.9; $\Delta\chi^2$: 55.1; $P < 0.0001$; for GGI, HR = 3.2, 95% CI = 2.3–4.2; $\Delta\chi^2$: 57.5; $P < 0.0001$); and for GENE76, HR = 2.7, 95% CI = 2.1–3.4; $\Delta\chi^2$: 61.2; $P < 0.0001$; Fig. 3A–C). Of the stroma and immune-related gene signatures, only one of the former (SDPP) and one of the latter (IRM) added prognostic information to AOL in ER^+/$HER2^-$ breast cancer: for SDPP,

Table 1. Characteristics of patients in the untreated (cohort 1) and treated (cohort 2) cohorts

	Untreated (cohort 1) set ($n = 1,188$)					Treated (cohort 2) set ($n = 2,334$)				
Age group	**≤40 (%)**	**41–52 (%)**	**53–64 (%)**	**≥65 (%)**	***P*[a]**	**≤40 (%)**	**41–52 (%)**	**53–64 (%)**	**≥65 (%)**	
Total number (%)	**191 (16)**	**477 (40)**	**276 (23)**	**244 (21)**		**260**	**683**	**616**	**775**	***P*[a]**
Tumor size										
<2cm	72 (38)	215 (45)	133 (48)	124 (51)		121 (46)	327 (48)	286 (47)	328 (42)	
2–5cm	190 (57)	245 (51)	139 (50)	113 (46)		112 (43)	290 (43)	261 (42)	355 (46)	
> 5cm	2 (1)	4 (1)	2 (1)	2 (1)		5 (2)	9 (1)	11 (2)	9 (1)	
Unknown	8 (4)	13 (3)	2 (1)	5 (2)	0.32	22 (9)	57 (8)	58 (9)	83 (11)	0.517
Nodal status										
Negative	160 (84)	424 (89)	263 (95)	226 (93)		74 (29)	253 (37)	261 (42)	332 (43)	
Positive	23 (12)	41 (9)	11 (4)	8 (3)		170 (65)	376 (55)	299 (49)	357 (46)	
Unknown	8 (4)	12 (2)	2 (1)	10 (4)	<0.0001	16 (6)	54 (8)	56 (9)	86 (11)	<0.0001
Histologic grade										
I	14 (7)	79 (17)	35 (13)	44 (18)		24 (9)	101 (15)	81 (13)	94 (12)	
II	54 (29)	163 (34)	108 (39)	113 (46)		60 (23)	213 (31)	351 (41)	325 (42)	
III	109 (57)	193 (40)	102 (37)	58 (24)		143 (55)	291 (43)	216 (35)	280 (36)	
Unknown	14 (7)	42 (9)	31 (11)	29 (12)	<0.0001	33 (13)	78 (11)	68 (11)	76 (10)	<0.0001
ESR1 gene[b]										
Positive	110 (58)	340 (71.3)	207 (75)	189 (77.5)		141 (54)	376 (55)	320 (52)	433 (56)	
Negative	81 (42)	136 (28.5)	69 (25)	55 (25.5)		117 (45)	268 (39)	194 (31)	213 (27)	
Unknown	0 (0)	1 (<1)	0 (0)	0 (0)	<0.0001	2 (1)	39 (6)	102 (17)	129 (17)	<0.0001
ErbB2 gene[b]										
Positive	42 (22)	107 (22)	44 (16)	27 (11)		76 (29)	159 (23)	132 (21)	142 (18)	
Negative	149 (78)	370 (78)	232 (84)	217 (89)	0.001	184 (71)	524 (77)	484 (79)	633 (82)	0.008
RFS										
Eligible patients, *n*			1,150 (97%)					1,751 (75%)		
Median follow-up			9 y					3.6 y		
Interquartile range			6.6–11.4 y					1–7 y		

[a]*P* calculated using the χ^2 test.

[b]ESR1 and ERBB2 positivity and negativity was defined using gene expression (see Materials and Methods).

HR = 0.5, 95% CI = 0.4–0.6; $\Delta\chi^2$: 24.4; $P < 0.0001$; and for IRM, HR = 0.7, 95% CI = 0.5–0.9; $\Delta\chi^2$: 5.6; $P = 0.019$, again with no age interaction ($P = 0.18$) and ($P = 0.86$), respectively (Supplementary Fig. S5).

In $HER2^+$ breast cancer, all stroma-related gene signatures added significant prognostic information to AOL without interacting with age: PLAU: HR = 2.4, 95% CI = 1.4–4.3; $\Delta\chi^2$: 9.6; $P = 0.0002$, $P_{interaction} = 0.58$; DCN: HR = 2.8, 95% CI = 1.5–5.2; $\Delta\chi^2$: 10.4; $P = 0.001$; $P_{interaction} = 0.85$ and SDPP: HR = 0.3, 95% CI = 0.2–0.6; $\Delta\chi^2$: 12.1; $P = 0.001$; $P_{interaction} = 0.19$ (Supplementary Fig. S6). High expression of STAT1 and IRM were associated with better prognosis, also without an age interaction (Supplementary Fig. S7A–C).

In $ER^-/HER2^-$ tumors, PLAU and DCN showed a significant interaction with age ($P_{interaction} = 0.04$; FDR = 0.18 for both), with high expression associated with poor prognosis only in patients of 40 years or less (HR = 2.4, 95% CI = 1.0–5.4; $\Delta\chi^2$: 4.4; $P = 0.04$) and HR = 2.5, 95% CI = 0.1.1–5.9; $\Delta\chi^2$: 4.6; $P = 0.03$), respectively (Fig. 4A–C). Although previous studies have reported the prognostic potential of immune-related gene signatures in $ER^-/HER2^-$ tumors (32, 33), neither this nor any interaction with age, was observed in the current analysis (Supplementary Fig. S7D–F).

Is breast cancer arising in young women associated with unique disease biology?

To understand whether breast cancer in young women is biologically distinct from that diagnosed in older age groups and not just a surrogate for a higher incidence of aggressive breast cancer subtypes, we conducted a MEDLINE search to identify candidate age-related genes and pathways that have been suggested to characterize breast cancer arising at a young age. Out of 280 potentially relevant articles, we identified a total of 41 genes and 13 gene sets related to these aberrations (Supplementary Table S8). We then evaluated the differences in the gene expression values of these candidates using a linear regression model adjusted first for age as a continuous variable and data source and then other potential confounding variables such as breast cancer subtype, tumor size, nodal status, and histologic grade.

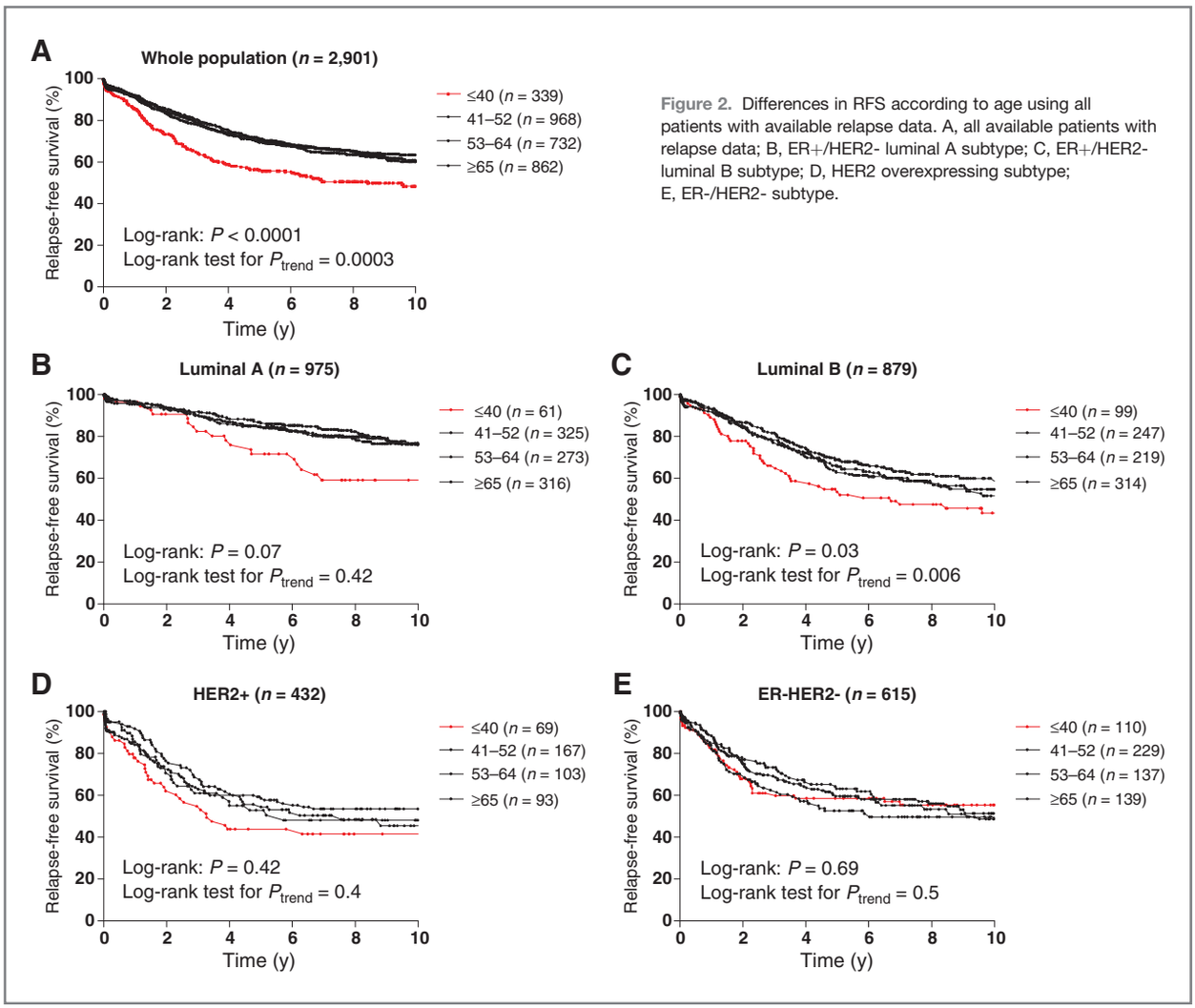

Figure 2. Differences in RFS according to age using all patients with available relapse data. A, all available patients with relapse data; B, ER+/HER2- luminal A subtype; C, ER+/HER2- luminal B subtype; D, HER2 overexpressing subtype; E, ER-/HER2- subtype.

Within the untreated cohort (cohort 1), the expression of 16 genes and gene sets were found to be significantly age dependent after adjustment. We proceeded to replicate these findings in the treated cohort (cohort 2) and found that 12 out of the 16 were still significantly associated with age after adjustment (Table 2). The common themes associated with young age were enrichment of biological processes related to immature mammary cell populations (*RANKL*, *c-kit*, BRCA1-mutated phenotype, mammary stem cells, and luminal progenitors cells), and growth factor signaling [mitogen—activated protein kinase (MAPK), phosphoinositide 3-kinase (PI3K)-related]. There was also downregulation of apoptosis-related genes.

Discussion

To the best of our knowledge, this is the largest work using gene expression data to investigate the biology and prognosis of breast cancer in young women. Notably, we found that breast cancer in the young seems to be associated with a unique biology irrespective of being enriched with more $ER^-/HER2^-$ tumors. We also found that the proliferation-related gene signatures seemed to be clinically relevant in patients aged 40 years or less as well as in the older age groups, imparting prognostic information beyond that provided by AOL (36). This is despite the observation that young breast cancer patients with $ER^+/HER2^-$ disease have an inferior recurrence-free survival compared with older women with $ER^+/HER2^-$ disease. As this is the first report that specifically addresses the relevance of prognostic gene signatures in young patients, our data suggest that they may be helpful in treatment decision-making given that young age alone is considered by many to warrant aggressive systemic therapy.

Interestingly, in the $ER^-/HER2^-$ subtype, we observed a prognostic value of stroma-related gene signatures, DCN and PLAU only in patients aged 40 years or less. However, although there was consistency in the prognostic

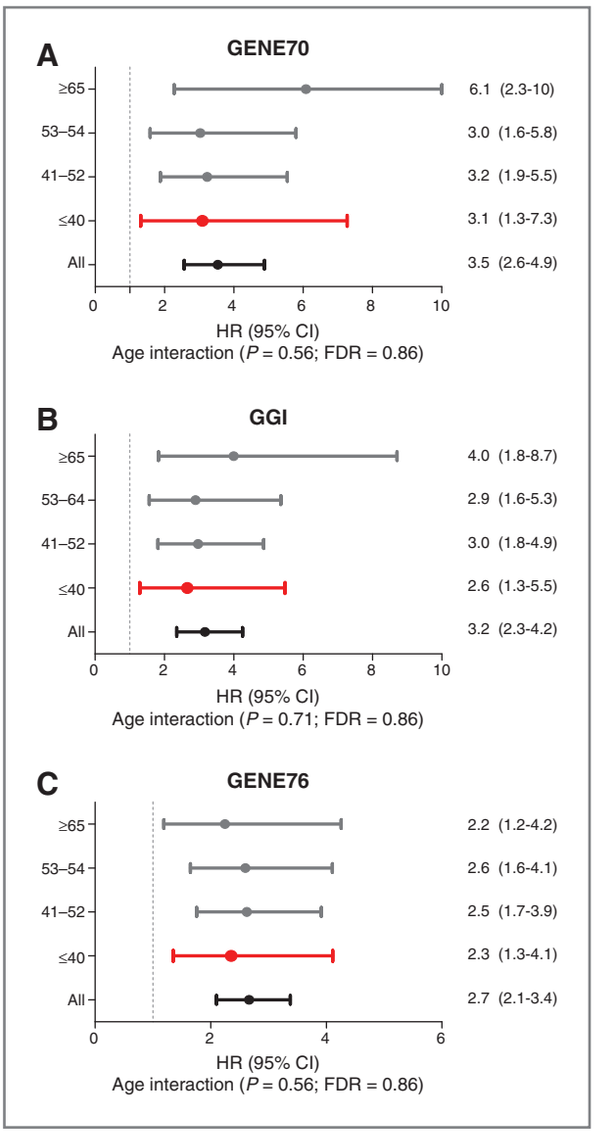

Figure 3. Prognostic evaluation of proliferation gene signatures in the $ER^+/HER2^-$ subtype according to age groups. Dotted line represents a HR of 1.0 and error bars represent 95% CIs. All HR shown have been adjusted for AOL. A, GENE70; B, GGI; C, GENE76. The $P_{interaction}$ between age as a continuous variable and the gene signature in a Cox-model and corresponding FDR value is shown for each gene signature.

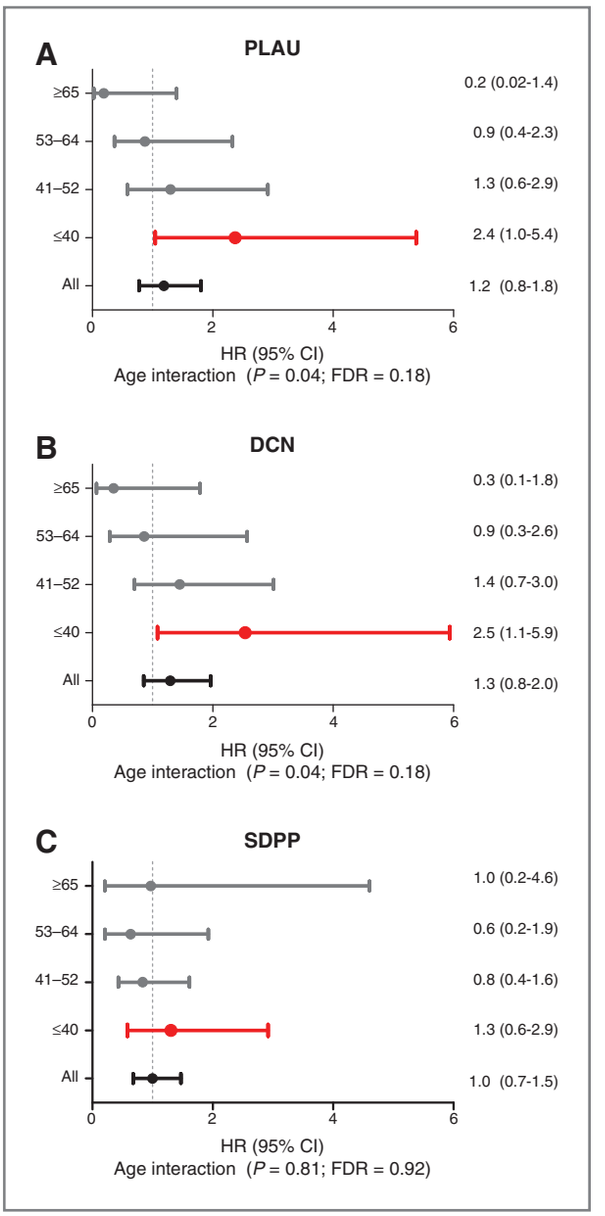

Figure 4. Prognostic evaluation of stroma-related gene signatures in the $ER^-/HER2^-$ subtype according to age groups. Dotted line represents a hazard ratio (HR) of 1.0, and error bars represent 95% CIs. All HR shown have been adjusted for AOL. A, PLAU; B, DCN; C, SDPP. The p-value of an interaction between age as a continuous variable and the gene signature in a Cox-model and corresponding FDR value is shown for each gene signature.

performance of the proliferation-related gene signatures, this was not observed with the 3 stroma-related ones. Of note, SDPP was developed by microdissection of tumor-associated stroma, which was not the case for DCN and PLAU, which were both developed *in silico* and are highly correlated ($R = 0.88$). Regardless, these results highlight a potential role of the microenvironment in mediating breast cancer growth in young women, particularly for those with $ER^-/HER2^-$ breast cancer. As breast stroma in young women is highly responsive to growth factor stimulation potentially to accommodate pregnancy and lactation during child-bearing years, it is not inconceivable that this microenvironment could also be advantageous for aggressive tumor growth (37–40). Therefore, it may be worth developing therapeutic approaches to target the microenvironment and stroma for $ER^-/HER2^-$ subgroup in young women.

Consistent with recent publications using both gene expression-defined molecular subtypes and immunohistochemistry, we found that breast cancer in young women

Table 2. Genes and gene sets significantly associated with age (young age) in cohorts 1 ($n = 1,188$) and 2 ($n = 2,334$)

Function	Untreated cohort (cohort 1, $n = 1,188$)					Treated cohort (cohort 2, $n = 2,334$)			
	Genes	Gene sets	Effect of age adjusted for data set	Effect of age adjusted for all covariates	FDR of 2nd adjustment	Effect of age adjusted for data set	Effect of age adjusted for all covariates	FDR of 2nd adjustment	Up- or downregulated
Apoptosis related	FAS		1.7E-04	6.6E-03	0.03	7.5E-04	3.9E-03	0.008	down
	CASP3		2.2E-03	2.2E-02	0.08	3.3E-03	2.5E-02	0.04	
	BAD		3.8E-03	3.2E-02	0.11	4.0E-03	1.7E-02	0.03	
MAP kinase related		MAPK	1.2E-13	5.8E-07	<0.0001	1.6E-08	5.9E-05	0.0002	up
mTOR/PI3K related	PDPK1		3.3E-03	2.1E-02	0.08	2.3E-05	7.6E-04	0.002	up
		PIK3CA-GS	1.4E-12	6.7E-09	<0.0001	1.7E-11	5.0E-11	<0.0001	
BRCA related	BRCA1		3.3E-04	3.8E-04	0.003	2.4E-02	4.4E-02	0.06	down
		BRCA1 mutant	5.4E-09	4.5E-03	0.02	5.6E-06	2.5E-03	0.006	up
Stem cell related	RANKL		5.8E-08	1.8E-10	<0.0001	1.3E-06	1.6E-06	<0.0001	up
		MaSC	8.0E-11	1.5E-09	<0.0001	3.5E-18	3.2E-15	<0.0001	up
Luminal progenitor	c-kit		5.8E-12	3.3E-13	<0.0001	7.9E-08	1.3E-07	<0.0001	up
		Luminal progenitor	1.7E-09	1.1E-03	0.007	2.4E-05	1.9E-02	0.04	up

NOTE: For complete reference list of genes and gene sets, please refer to Supplementary 8.

is enriched with $ER^-/HER2^-$ tumors, with a lower incidence of luminal-A type tumors (15, 16). However, one of the most controversial questions is whether young age is associated with unique cancer biology. Recently, Anders and colleagues concluded that age alone did not seem to induce biological influence beyond that of breast cancer subtype and grade (15). These results were in direct contrast with the poor outcome of young breast cancer patients after adjusting for ER, grade, and HER2 status documented in several studies, including this one (6–10).

There are several differences between our study and that of Anders and colleagues (15). We studied candidate genes and gene sets based on a literature search, thereby reducing the potential bias associated with multiple testing. In addition, we included 873 patients aged 45 years or less compared with 130 patients in their analysis. We also investigated trends across age as a continuous variable, rather than dichotomizing age at its extremes. This allowed us potentially to detect subtle biological and clinical differences in gene expression across age.

The results of our biological analyses propose several interesting hypotheses to be further validated. We confirm a previous finding that suggested that breast cancer in young women is enriched with genes involved in extracellular signal—regulated kinase and PI3K signaling (5). Similarly, we also found that the single gene *BRCA1* and a gene set developed from BRCA1 germ line mutant breast tumors was significantly associated with breast cancer arising at a young age, suggesting similar biological processes (18). Of note, in all the analyzed series, there were only 15 known germ line mutant BRCA1 carriers documented in 4 data sets. In addition, we observed significant enrichment of gene sets representing luminal progenitor cells, mammary stem cells, and high levels of *RANKL* and *c-kit*. These gene sets were strongly correlated with young age, independent of breast cancer subtype, in 2 large independent and heterogeneous cohorts of patients.

Although these age-specific findings could be difficult to functionally validate in an experimental model, the RANKL results are particularly interesting. The normal breast in young women is enriched with immature mammary cell subpopulations (stem cells and progenitors), which have been shown to increase significantly with pregnancy, menstrual cycles, and lactation. *RANKL* has been shown to be a key mediator of this effect and RANKL inhibition could be antiproliferative as well as mediate reductions in the mammary stem cell compartment which is thought to predispose to cancer (41–43). Similarly, Lim and colleagues recently proposed that the cell of origin of *BRCA1*-associated and "basal-like" tumors were probably luminal progenitor cells rather than the stem-cell enriched population with *c-kit* identified as a key marker (44). Our results may also help to explain the higher incidence of $ER^-/HER2^-$ tumors but also suggest a common shared biology in the young. This could account for the worst clinical outcome long associated with young age and imply that approaches such as suppression of mammary stem cell function or *RANKL* signaling may need to be explored in the young population. This and other data now provide significant scientific rationale to take these concepts forward into the clinical setting—inhibition of these pathways with specific drugs such as a RANKL inhibitor and their effects on mammary epithelial populations and tumor growth could be initially examined in preoperative "window of opportunity" studies in young women with newly diagnosed tumors to show "proof-of-concept" (for example, EudraCT number 2011-006224-21).

Our study has potential limitations that should be considered while interpreting its results. We chose to use a 3-gene classifier for breast cancer subtyping rather than the intrinsic gene list (12, 13, 45–47) as we felt that this would more closely approximate clinically used subgrouping and results would be therefore comparable with previous publications using IHC. Currently, no gene expression method for the assignment of molecular subtypes is considered as the "gold standard" as hierarchical clustering allocation is subjective and interobserver reproducibility remains modest (48, 49). Nevertheless, the 3-gene classifier used in the current study has been shown to provide a robust classification of the major molecular breast cancer subtypes using gene expression of the ER (*ESR1*), *ERBB2*, and a proliferation gene (*AURKA*) and provide similar prognostic information to that of PAM50 (46). We also acknowledge that the AOL calculation was not completely accurate in 100% of patients due to missing clinical variables. However we do not believe that the missing information could have significantly overestimated the value of prognostic gene signatures after adjustment. One should also note that for many of the prognostic gene signatures, while the exact published algorithms were not used, the approximated versions still produced a strong prognostic signal. Another limitation was the minimal information available on survival and specific adjuvant treatment modalities given to these women as it would have been interesting to look at regimen-specific treatment effects according to age and subtype.

In summary, we conclude that proliferation-related prognostic gene signatures could aid in treatment decision-making independent of age. This may be particularly clinically relevant for the young given the potential long-term side effects of adjuvant systemic chemotherapy. Furthermore, we find that young age adds extra biological complexity, which is independent of differences in breast cancer subtype distribution. Although these results require further validation, either experimentally or in other clinical data sets, we suggest that separate therapeutic approaches may need to be specifically designed to improve outcomes for breast cancer arising in young women.

Disclosure of Potential Conflicts of Interest

C. Sotiriou is a named inventor of the Genomic Grade Index (GGI). S. Loi and C. Sotiriou are named co-inventors of a PI3K prognostic gene signature (PIK3CA-GS). No potential conflicts of interest were disclosed by the other authors.

Acknowledgments

The authors thank Carolyn Straehle for editorial assistance and all of the patients who have generously donated their tumor tissue for research.

Grant Support

H.A. Azim Jr is supported by an ESMO translational research grant.

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.

Received October 10, 2011; revised December 14, 2011; accepted December 29, 2011; published OnlineFirst January 18, 2012.

References

1. Brinton LA, Sherman ME, Carreon JD, Anderson WF. Recent trends in breast cancer among younger women in the United States. J Natl Cancer Inst 2008;100:1643–8.
2. El Saghir NS, Khalil MK, Eid T, El Kinge AR, Charafeddine M, Geara F, et al. Trends in epidemiology and management of breast cancer in developing Arab countries: a literature and registry analysis. Int J Surg 2007;5:225–33.
3. El Saghir NS, Seoud M, Khalil MK, Charafeddine M, Salem ZK, Geara FB, et al. Effects of young age at presentation on survival in breast cancer. BMC Cancer 2006;6:194.
4. Bharat A, Aft RL, Gao F, Margenthaler JA. Patient and tumor characteristics associated with increased mortality in young women (< or = 40 years) with breast cancer. J Surg Oncol 2009;100:248–51.
5. Anders CK, Hsu DS, Broadwater G, Acharya CR, Foekens JA, Zhang Y, et al. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. J Clin Oncol 2008;26:3324–30.
6. Fredholm H, Eaker S, Frisell J, Holmberg L, Fredriksson I, Lindman H. Breast cancer in young women: poor survival despite intensive treatment. PLoS One 2009;4:e7695.
7. Adami HO, Malker B, Holmberg L, Persson I, Stone B. The relation between survival and age at diagnosis in breast cancer. N Engl J Med 1986;315:559–63.
8. Chung M, Chang HR, Bland KI, Wanebo HJ. Younger women with breast carcinoma have a poorer prognosis than older women. Cancer 1996;77:97–103.
9. Kollias J, Elston CW, Ellis IO, Robertson JF, Blamey RW. Early-onset breast cancer–histopathological and prognostic considerations. Br J Cancer 1997;75:1318–23.
10. Kim EK, Noh WC, Han W, Noh DY. Prognostic significance of young age (<35 years) by subtype based on ER, PR, and HER2 status in breast cancer: a nationwide registry-based study. World J Surg 2011;35:1244–53.
11. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
12. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, 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.
13. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418–23.
14. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, 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.
15. Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-Demore N, et al. Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes?J Clin Oncol 2011;29:e18–20.
16. Cancello G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, Pruneri G, et al. Prognosis and adjuvant treatment effects in selected breast cancer subtypes of very young women (<35 years) with operable breast cancer. Ann Oncol 2010;21:1974–81.
17. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer 2007;109:1721–8.
18. van `t Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6.
19. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, 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.
20. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.
21. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007;356:217–26.
22. Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 2004;5:607–16.
23. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671–9.
24. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
25. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. J Clin Oncol 2008;26:721–8.
26. Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the MINDACT trial. J Clin Oncol 2008;26:729–35.
27. Symmans WF, Ayers M, Clark EA, Stec J, Hess KR, Sneige N, et al. Total RNA yield and microarray gene expression profiles from fine-needle aspiration biopsy and core-needle biopsy samples of breast carcinoma. Cancer 2003;97:2960–71.
28. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res 2008;10:R65.
29. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res 2008;14:5158–65.
30. Farmer P, Bonnefoi H, Anderle P, Cameron D, Wirapati P, Becette V, et al. A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. Nat Med 2009;15:68–74.
31. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med 2008;14:518–27.
32. 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.
33. Sadun RE, Sachsman SM, Chen X, Christenson KW, Morris WZ, Hu P, et al. Immune signatures of murine and human cancers reveal unique mechanisms of tumor escape and new targets for cancer immunotherapy. Clin Cancer Res 2007;13:4016–25.
34. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Controlled clinical trials 1996;17:343–6.
35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289–300.

36. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, 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.
37. Bemis LT, Schedin P. Reproductive state of rat mammary gland stroma modulates human breast cancer cell migration and invasion. Cancer Res 2000;60:3414–8.
38. Bhowmick NA, Moses HL. Tumor-stroma interactions. Curr Opin Genet Dev 2005;15:97–101.
39. Kim JB, Stein R, O'Hare MJ. Tumour-stromal interactions in breast cancer: the role of stroma in tumourigenesis. Tumour Biol 2005;26: 173–85.
40. McDaniel SM, Rumer KK, Biroc SL, Metz RP, Singh M, Porter W, et al. Remodeling of the mammary microenvironment after lactation promotes breast tumor cell metastasis. Am J Pathol 2006;168: 608–20.
41. Asselin-Labat ML, Vaillant F, Sheridan JM, Pal B, Wu D, Simpson ER, et al. Control of mammary stem cell function by steroid hormone signalling. Nature 2010;465:798–802.
42. Gonzalez-Suarez E, Jacob AP, Jones J, Miller R, Roudier-Meyer MP, Erwert R, et al. RANK ligand mediates progestin-induced mammary epithelial proliferation and carcinogenesis. Nature 2010; 468:103–7.
43. Tiede B, Kang Y. From milk to malignancy: the role of mammary stem cells in development, pregnancy and breast cancer. Cell Res 2011; 21:245–57.
44. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med 2009;15:907–13.
45. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006;7:96.
46. Haibe-Kains B, Desmedt C, Loi S, Culhane AC, Bontempi G, Quackenbush J, et al. A three-gene model to robustly identify breast cancer molecular subtypes. J Natl Cancer Inst. 2012 Jan 18 [Epub ahead of print].
47. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160–7.
48. Mackay A, Weigelt B, Grigoriadis A, Kreike B, Natrajan R, A'Hern R, et al. Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement. J Natl Cancer Inst 2011;103:662–73.
49. Weigelt B, Mackay A, A'Hern R, Natrajan R, Tan DS, Dowsett M, et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. Lancet Oncol 2011;11:339–49.