

# Concordance among gene expression-based predictors for ER-positive breast cancer treated with adjuvant tamoxifen

A. Prat<sup>1,2</sup>, J. S. Parker<sup>1</sup>, C. Fan<sup>1</sup>, M. C. U. Cheang<sup>1</sup>, L. D. Miller<sup>3</sup>, J. Bergh<sup>4,5</sup>, S. K. L. Chia<sup>6</sup>, P. S. Bernard<sup>7</sup>, T. O. Nielsen<sup>6,8</sup>, M. J. Ellis<sup>9</sup>, L. A. Carey<sup>1,10</sup> & C. M. Perou<sup>1,11,12\*</sup>

<sup>1</sup>Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, USA; <sup>2</sup>Department of Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain; <sup>3</sup>Department of Cancer Biology, Comprehensive Cancer Center, Wake Forest School of Medicine, Winston Salem, USA; <sup>4</sup>Department of Oncology-Pathology, Karolinska Institutet & Cancer Center Karolinska, Stockholm, Sweden; <sup>5</sup>Department of Medical Oncology, Paterson Institute, Christie Hospital and Manchester University, Manchester, UK; <sup>6</sup>British Columbia Cancer Agency, Vancouver, Canada; <sup>7</sup>Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, USA; <sup>8</sup>Department of Pathology, University of British Columbia, Vancouver, Canada; <sup>9</sup>Department of Medicine, Division of Oncology, Siteman Cancer Center at Washington University, St. Louis; <sup>10</sup>Department of Medicine, Division of Hematology and Oncology, University of North Carolina, Chapel Hill; <sup>11</sup>Departments of Genetics; <sup>12</sup>Pathology & Laboratory Medicine, University of North Carolina, Chapel Hill, USA

Received 16 December 2011; revised 9 February 2012; accepted 10 February 2012

**Background:** ER-positive (ER+) breast cancer includes all of the intrinsic molecular subtypes, although the luminal A and B subtypes predominate. In this study, we evaluated the ability of six clinically relevant genomic signatures to predict relapse in patients with ER+ tumors treated with adjuvant tamoxifen only.

**Methods:** Four microarray datasets were combined and research-based versions of PAM50 intrinsic subtyping and risk of relapse (PAM50-ROR) score, 21-gene recurrence score (OncotypeDX), Mammprint, Rotterdam 76 gene, index of sensitivity to endocrine therapy (SET) and an estrogen-induced gene set were evaluated. Distant relapse-free survival (DRFS) was estimated by Kaplan–Meier and log-rank tests, and multivariable analyses were done using Cox regression analysis. Harrell's C-index was also used to estimate performance.

**Results:** All signatures were prognostic in patients with ER+ node-negative tumors, whereas most were prognostic in ER+ node-positive disease. Among the signatures evaluated, PAM50-ROR, OncotypeDX, Mammprint and SET were consistently found to be independent predictors of relapse. A combination of all signatures significantly increased the performance prediction. Importantly, low-risk tumors (>90% DRFS at 8.5 years) were identified by the majority of signatures only within node-negative disease, and these tumors were mostly luminal A (78%–100%).

**Conclusions:** Most established genomic signatures were successful in outcome predictions in ER+ breast cancer and provided statistically independent information. From a clinical perspective, multiple signatures combined together most accurately predicted outcome, but a common finding was that each signature identified a subset of luminal A patients with node-negative disease who might be considered suitable candidates for adjuvant endocrine therapy alone.

**Key words:** breast cancer, genomics, luminal, mammprint, oncotype, PAM50

## Introduction

Gene expression-based assays have been developed that can successfully predict outcomes in early-stage ER-positive (ER+) breast cancer beyond standard clinicopathological variables [1–5]. OncotypeDX recurrence score (GHI)<sup>2</sup> and Mammprint® (NKI70)<sup>3</sup> are clinically available and currently being evaluated in two large prospective clinical trials (TAILORx and MINDACT) [6, 7]. Since then, other prognostic predictors such as the Rotterdam 76-gene signature (ROT76) [8, 9] and the risk of relapse (ROR) score based on the PAM50 subtype assay [10] have been developed using two

different node-negative and adjuvant treatment-naive populations.

Previous studies have also shown that many of these expression signatures are concordant for predicting outcomes [11, 12]. However, it is currently unknown if these findings are still valid in a more contemporary ER+ population treated with adjuvant endocrine therapy only [13]. Moreover, recent signatures specifically designed to track hormonal responsiveness, such as the estrogen-induced gene set (IE-IIE) [14] and the genomic index of sensitivity to endocrine therapy (SET) [15], can also predict survival in early-stage ER+ disease. Thus, estrogen-regulated gene signatures could be tracking ER+ tumors with high endocrine sensitivity.

From a clinical perspective, genomic assays are helping to identify patients with early-stage ER+ breast cancers who do

\*Correspondence to: Prof. C. M. Perou, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA. Tel: +1-919-843-5740; Fax +1-919-843-5718; E-mail: cperou@med.unc.edu

not need chemotherapy and are effectively treated with adjuvant endocrine agents alone. Alternatively, they could identify groups of patients with ER+ tumors who are more likely to be biologically homogenous and/or who might benefit from specific treatment strategies. In this report, we evaluated the relapse prediction abilities of six independent genomic signatures using a cohort of ER-positive breast cancer patients treated with adjuvant tamoxifen only.

## methods

### patients and samples

Four different publicly available microarray datasets were combined together to create a single large set of 594 ER+ patients, all of whom received appropriate local therapy and adjuvant tamoxifen only (see supplemental Figure S1, available at *Annals of Oncology* online). Thousand fifty-three Affymetrix U133A CEL files from various publicly available microarray datasets (GSE17705 [MDACC298] [15], GSE6532 [LOI327] [16, 17], GSE12093 [ZHANG136] [18], GSE1456 [PAWITAN159] [19] and MDACC133 [20]) were processed using MAS 5.0 (R/Bioconductor) to generate probe-level intensities with a median array intensity of 600, and each expression value was  $\log_2$  transformed. To batch correct the gene expression data [21, 22], the probeset medians in each individual dataset were adjusted to the MDACC133 reference set accounting for differences in the proportion of clinical ER+/- samples; after batch correction, all ER- tumors were removed, as were all ER+ tumors not treated with tamoxifen-only, thus leaving 594 tumors per microarrays.

### genomic predictors

The following gene expression signatures were evaluated using the combined microarray dataset: GHI [2], NKI70 [3], ROT76 [8], IE-IIE [14], SET [15] and PAM50 [10] (supplemental Table S1, available at *Annals of Oncology* online). Each signature was evaluated as a continuous variable and as group categories according to the published cut-offs [2, 3, 8, 10, 14, 15]. Briefly, the intrinsic subtypes, the risk of relapse based on subtype (PAM50-RORS), the ROR based on subtype and proliferation (PAM50-RORP) and the proliferation index (PAM50-PROLIF) were identified using the PAM50 subtype assay [10]. The PAM50-PROLIF index is the mean expression of 11 PAM50 proliferation-related genes of the PAM50 assay [23]. GHI and NKI70 were evaluated as previously described [12]. For the IE-IIE signature, we calculated the Spearman correlation to the two training centroids (IE and IIE) as described by Oh et al. [14]; samples with a correlation ratio to the IE centroid/IIE centroid  $>1.0$  were assigned to the IE group and the rest to the IIE group. Finally, for the ROT76 and SET signatures, all Affymetrix U133A probes were evaluated as described in both publications, respectively [8, 15]. The list of gene and/or probes, the scores and the group categories for each signature can be obtained from supplemental data, available at *Annals of Oncology* online.

To further explore the PAM50, results were obtained from combining the microarray dataset with a quantitative RT-PCR (qRT-PCR) dataset of 786 ER+ breast cancer patients treated with adjuvant tamoxifen only from Nielsen et al. [23] (Nielsen series).

### statistical analysis

Distant relapse-free survival (DRFS) estimates were from the Kaplan–Meier curves and tests of differences by the log-rank test. The DRFS follow-up time was censored at 8.5 years since it was the longest follow-up time in the PAWI159 [19] dataset. Univariate and multivariable analyses (MVA) were calculated using a Cox proportional regression model.

MVA prognostic models including all the signatures as independent continuous variables were built and assessed using a Cox model with the

penalized least absolute shrinkage and selection operator (LASSO) method approach [24]. In each case, a training set (2/3 of the dataset) was randomly used to build a model, which was then applied to the testing set (i.e. the remaining 1/3). We repeated this procedure 200 times as previously carried out [5]. In each testing set, the prognostic performance of each model and each individual signature was estimated by calculating the concordance index (C-index) [25]. All statistical computations were carried out in R v.2.8.1 (<http://cran.r-project.org>).

## results

### clinicopathological characteristics of the combined microarray and qRT-PCR PAM50 dataset

We created a large dataset of 1380 ER-positive breast cancer patients treated with adjuvant tamoxifen only using publicly available microarray data ( $n = 594$ ) and PAM50 qRT-PCR data only ( $n = 786$ ) from the Nielsen series [4, 15–19]. Six hundred and ten and 699 patients were identified as having node-negative and node-positive disease, respectively (Table 1). As expected, luminal subtypes predominated ( $n = 1171$ , 84.9%). Non-luminal subtypes (HER2-enriched and basal-like) represented 9.1% ( $n = 125$ ) of the patients. The normal breast-like samples were not further considered as these specimens are predominantly composed of normal breast tissue, which precludes the correct assignment to a tumor subtype for meaningful outcome predictions [1, 10].

The PAM50 intrinsic subtypes were prognostic for DRFS within node-negative and node-positive patients (Figure 1A and B). In node-negative disease, luminal A tumors showed a better outcome than luminal B [hazard ratio (HR) = 0.313,  $P < 0.0001$ ], HER2-enriched (HR = 0.256,  $P < 0.001$ ) and basal-like (HR = 0.168,  $P < 0.001$ ) subtypes. However, no statistical significant differences in DRFS were observed among the poor prognostic luminal B, HER2-enriched and basal-like subtypes. In node-positive disease, the PAM50 subtypes were also prognostic; of note, DRFS of both luminal subtypes was significantly lower compared with their counterparts in node-negative disease (luminal A, HR = 3.29 and luminal B, HR = 2.26,  $P < 0.0001$  for both comparisons). Regardless of nodal status, both luminal subtypes had continued risk of relapse after 5 years; even the lowest risk node-negative luminal A subtype had 5-year DRFS of 96% that dropped to 91% by 8.5 years. A tendency for worse outcomes was also observed in node-positive HER2-enriched tumors compared with node-negative HER2-enriched tumors (HR = 1.91,  $P = 0.099$ ).

### genomic relationships and biological significance

For comparisons across different predictors, the combined dataset was confined to the 594 samples/tumor represented by Affymetrix microarray data. We first compared the gene overlap between any two signatures and found that  $\leq 25\%$  of the genes were shared between signatures (supplemental Table S2, available at *Annals of Oncology* online), except for 9 and 11 genes of the GHI signature ( $n = 21$ ) that were present in the IE-IIE and PAM50, respectively, and 15 genes of the IE-IIE signature that were present in PAM50. In spite of relatively little gene overlap, all predictors were significantly correlated

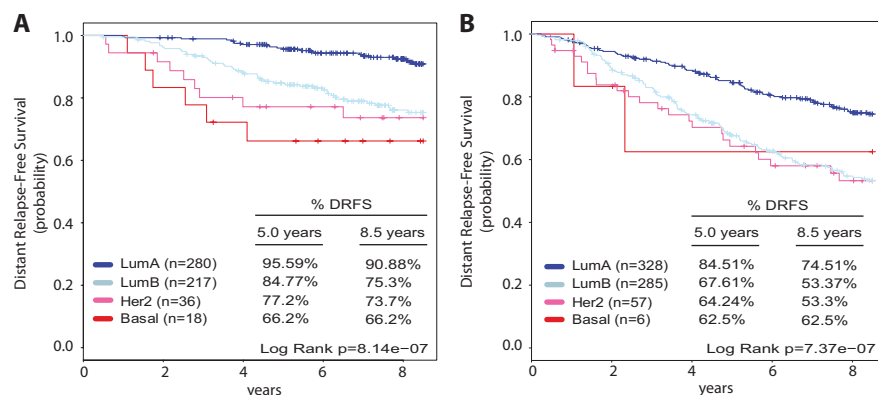
**Table 1.** Clinicopathological characteristics of the combined microarray and qRT-PCR patient dataset

	Symmans et al. [15]	%	Loi et al. [16, 17]	%	Zhang et al. [18]	%	Pawitan et al. [19]	%	Nielsen et al. [23]	%	Total	%
Dataset	MDACC298		LOI327		ZHANG136		PAWI159		NIELSEN			
GSE series	GSE17705	—	GSE6532	—	GSE12093	—	GSE1456	—	—	—	—	—
Patients <sup>a</sup>	298	100%	73	22%	136	100%	87	55%	786	100%	1380	—
Node-negative <sup>b</sup>	175	61%	20	29%	136	100%	57	69%	222	30%	610	47%
Node-positive <sup>b</sup>	112	39%	50	71%	0	0%	26	31%	511	70%	699	53%
PAM50 subtypes <sup>c</sup>												
Luminal A	132	44%	34	47%	66	49%	40	46%	372	47%	644	47%
Luminal B	100	34%	19	26%	48	35%	31	36%	329	42%	527	38%
HER2-enriched	16	5%	10	14%	5	4%	5	6%	64	8%	100	7%
Basal-like	8	3%	4	5%	4	3%	4	5%	5	1%	25	2%
Normal-like	42	14%	6	8%	13	10%	7	8%	16	2%	84	6%

<sup>a</sup>Only patients with ER+ disease treated with tamoxifen-only were selected from these datasets. In GSE6532, 103 samples have been removed since they overlap with GSE17705.

<sup>b</sup>GSE17705, GSE1456, GSE6532 and Nielsen et al. [23] have 11, 4, 3 and 53 patients without node status, respectively (total  $n = 71$ ).

<sup>c</sup>Subtype data in Nielsen et al. were obtained by the qRT-PCR PAM50 assay.  
qRT-PCR, quantitative RT-PCR.



**Figure 1.** Kaplan–Meier DRFS analysis of intrinsic subtype as determined by PAM50 gene expression measurement (quantitative reverse transcription–PCR and microarray-based) from women with (A) node-negative and (B) node-positive invasive breast carcinoma, treated with adjuvant tamoxifen only. The number of patients and the estimated DRFS at 8.5 years in each group are shown beside each curve’s description. DRFS, distant relapse-free survival.

(Pearson correlation range 0.36–0.79;  $P < 0.0001$  for each comparison), with PAM50-RORS, IE-IIE and GHI showing the highest correlation between them ( $>0.72$ ,  $P < 0.0001$ , Pearson correlation; supplemental Table S2, available at *Annals of Oncology* online).

The observed correlations suggested that most predictors are tracking tumors with similar biology. To further explore this hypothesis, we evaluated the scores of each signature as a continuous variable and as group categories across the four major intrinsic subtypes (as defined by the PAM50 assay [10]). As expected, each predictor discriminated luminal A tumors from the luminal B subtype and from the rest of the subtypes [ $P < 0.0001$ , Student’s  $t$ -test (supplemental Figure S3 and Table S3, available at *Annals of Oncology* online)]. The high hormonal sensitivity groups (SET-high and IE-like) and low risk of recurrence groups (PAM50-RORS-low, PAM50-RORP-low, GHI-low, ROT76-good and NKI-good) were largely composed of luminal A tumors ( $>71\%$ – $100\%$ ).

### survival analyses within node-negative and node-positive disease

Univariate DRFS analyses revealed that each signature, evaluated as a continuous variable or as group categories, was highly prognostic in patients with node-negative disease (supplemental Figure S4 and Table S4, available at *Annals of Oncology* online). As expected, Kaplan–Meier survival analyses showed highly significant differences in DRFS across the groups predicted to have good or intermediate or poor prognosis (PAM50-RORS, PAM50-RORP, GHI, ROT76 and NKI70) or the groups predicted to have high or intermediate versus low expression of ER-regulated genes (SET and IE-IIE). Importantly, all predictors identified groups of node-negative patients with 93.7%–97.9% and 88.4%–96.2% DRFS at 5.0 and 8.5 years, respectively, although the number of patients in each group differed (Table 2); when limited to the combined microarray dataset and across the predictors with three risk categories (GHI, SET, PAM50-RORS and PAM50-RORP), the

**Table 2.** Low-risk group comparison among signatures

	Node-negative			Node-positive		
	Low-risk group			Low-risk group		
	N	% of luminal A	DRFS at 8.5 years	N	% of luminal A	DRFS at 8.5 years
RORP (PAM50)	82 (24%)	100%	94%	38 (22%)	100%	80%
RORS (PAM50)	140 (41%)	100%	90%	63 (37%)	100%	74%
PROLIF <sup>a</sup> (PAM50)	72 (22%)	100%	93%	33 (19%)	100%	82%
GHI	47 (14%)	94%	95%	14 (8%)	93%	81%
ROT76 <sup>b</sup>	164 (48%)	85%	92%	81 (47%)	77%	76%
IE-IIE <sup>b</sup>	235 (69%)	72%	88%	100 (58%)	71%	69%
NKI70 <sup>b</sup>	136 (40%)	78%	91%	53 (31%)	79%	76%
SET	26 (8%)	81%	96%	21 (12%)	91%	89%
RORP (PAM50) <sup>c</sup>	116 (21%)	100%	96%	116 (17%)	100%	84%
RORS (PAM50) <sup>c</sup>	197 (36%)	100%	91%	197 (29%)	100%	79%
PROLIF (PAM50) <sup>c</sup>	142 (26%)	99%	95%	166 (24%)	99%	80%

<sup>a</sup>Since proliferation (PROLIF) index does not have previously defined cut-offs, patients in the low-risk group are the ones with the lowest quartile expression.

<sup>b</sup>ROT76, IE-IIE and NKI70 signatures have two risk categories.

<sup>c</sup>qRT-PCR PAM50 data from the Nielsen series have been included. N, number of patients in the low-risk group and the percentage from the total number of patients based on node status.

DRFS, distant relapse-free survival.

PAM50-RORS identified the largest number of low-risk patients ( $n = 140$ , 41%), followed by PAM50-RORP ( $n = 82$ , 24%), GHI ( $n = 47$ , 14%) and SET ( $n = 27$ %). Inclusion of the 786 ER+ patient qRT-PCR PAM50 Nielsen series data confirmed that both PAM50-RORP and PAM50-RORS identify 21%–36% of all node-negative patients ( $n = 551$ ) as low risk [or alternatively they identify 41%–70% of all node-negative luminal A tumors ( $n = 280$ ) as low risk], and the PAM50-RORP-low and PAM50-RORS-low groups showed a DRFS at 8.5 years of 96.09% and 91.21%, respectively (Table 2 and supplemental Figure S5, available at *Annals of Oncology* online).

In node-positive disease, univariate DRFS analyses revealed that most signatures were barely significant when evaluated as continuous variables (supplemental Figure S6 and Table S4, available at *Annals of Oncology* online). When evaluated as group categories, low risk of relapse or high expression of ER-regulated gene groups showed either no statistical significance or borderline significance in terms of DRFS compared with the predicted poor prognostic or low expressers of ER-regulated gene groups. More importantly, no predictor identified a clear node-positive group of patients treated with tamoxifen alone with a DRFS at 8.5 years >90%. Although these results could be related to the sample size, data for PAM50-RORS and PAM50-RORP in node-positive disease confirmed this finding when the qRT-PCR PAM50 Nielsen series was included for a total of 676 patients (supplemental Figure S5, available at *Annals of Oncology* online). Finally, similar to node-negative disease, the predicted low-risk outcome groups in node-positive disease were predominantly comprised of luminal A tumors (71%–100%; Table 2).

### prognostic prediction performance

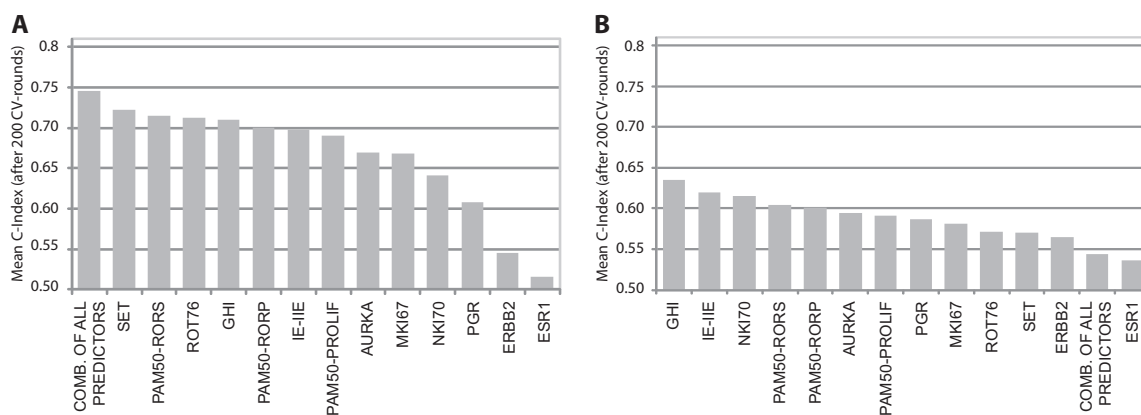
C-index values were calculated to estimate the performance of each genomic signature for predicting outcome (Figure 2). The C-index is a measure of the probability of concordance

between the predicted and the observed survival, ranging from 0.5 (random) to 1 (perfect). Although no clear cut-off value has been defined, values >0.70 are indicative of good prediction accuracy [25]. In node-negative disease, the vast majority of signatures showed similar predictive abilities (mean C-index range of 0.70–0.73), except PAM50-PROLIF index (mean C-index of 0.69) and NKI70 (mean C-index of 0.64). Conversely, in node-positive disease, all predictors carried out worse than in node-negative (mean C-index range of 0.56–0.63).

Despite comparable prognostic performance of these signatures and high correlation values among them, we observed that these signatures generally retained their prognostic significance independent of each other when testing two signatures at a time in multivariate analyses (Table 3). Thus, we sought to determine if we could improve prognostic performance by integrating information from all signatures into a single model; we determined that the combined model was better at predicting outcome than individual signatures in node-negative disease (Figure 2A) but failed in node-positive disease (Figure 2B). However, the absolute increase in performance of the combined model within node-negative disease was modest (range 0.02–0.11).

### prognostic predictions within the intrinsic subtypes

We explored the predictive ability of each signature within the predominant luminal A and B subtypes. In node-negative luminal A disease ( $n = 185$ ), ROT76 and SET (Figure 3A) were found to be prognostic in univariate analyses, and patients with the low-risk group showed a DRFS at 8.5 years of 94%–96% (supplemental Table S5, available at *Annals of Oncology* online). When limited to the microarray dataset, the PAM50-RORS and PAM50-RORP were trending toward significance (supplemental Table S5, available at *Annals of Oncology* online) and both were significant when the Nielsen series was included



**Figure 2.** Comparison of prognostic classifiers and single genes in (A) node-negative and (B) node-positive subjects. The C-index is used to compare accuracy of the prognostic classifiers and single genes. Signatures have been ranked ordered from highest to lowest mean C-index. In node-negative disease, the C-index of the combined model was superior to the C-index of each individual signature in at least 75% of the 200 total estimations.

**Table 3.** Multivariate Cox proportional hazards analyses of distant relapse-free survival among predictors<sup>a</sup>

		Adjusted on the following predictor							
		PAM50-RORP	PAM50-RORS	PAM50-PROLIF	GHI	ROT76	IE-IIE	NKI70	SET
Predictor $\chi^2$ statistic and <i>P</i> value	PAM50-RORP		10.6; 0.005	7.2; 0.027	11.5; 0.003	9.6; 0.008	10.45; 0.005	16.2; <0.001	17.4; <0.001
	PAM50-RORS	0.0; 0.990		1.0; 0.617	5.4; 0.067	2.9; 0.240	3.1; 0.220	7.1; 0.029	9.2; 0.012
	PAM50-PROLIF	0.2; 0.890	4.6; 0.099		7.6; 0.023	4.6; 0.100	5.8; 0.056	10.7; 0.005	13; 0.002
	GHI	9.1; 0.010	13.6; 0.001	12.2; 0.002		13.5; 0.001	13.4; 0.001	14.4; <0.001	20.0; <0.001
	ROT76	4.3; 0.031	8.29; 0.004	6.4; 0.012	10.7; 0.001		10.4; 0.001	13.0; <0.001	11.0; 0.0013
	IE-IIE	3.2; 0.072	6.4; 0.011	5.5; 0.019	8.6; 0.003	8.4; 0.004		9.2; 0.002	12.0; 0.001
	NKI70	5.7; 0.022	7.3; 0.007	7.24; 0.013	9.2; 0.002	7.8; 0.005	6.1; 0.014		14.0; <0.001
	SET	6.6; 0.042	9.0; 0.011	8.7; 0.012	13.6; 0.001	9.7; 0.008	8.5; 0.015	13.6; 0.001	

<sup>a</sup>Each square denotes the change in the likelihood ratio statistic ( $\chi^2$ ) of the signature in each row and its *P* value when conditioned on a signature in the column.

for a total of 280 luminal A patients (supplemental Table S5, available at *Annals of Oncology* online and PAM50-RORP in Figure 3B).

In node-positive luminal A disease ( $n = 81$ ) on the microarray dataset, GHI, NKI70 and IE-IIE were prognostic when evaluated as a continuous variable, and the combined low and intermediate risk GHI groups identified a group of significantly low-risk node-positive luminal A tumors ( $n = 30$ , 37%) with an outstanding DRFS at 8.5 years (96%,  $P < 0.01$ ; Figure 3C). When we included the qRT-PCR PAM50 Nielsen series dataset ( $n = 326$ ), PAM50-RORS and PAM50-RORP were found prognostic as a continuous variable and as group categories, with the low-risk PAM50-RORP group achieving a DRFS at 8.5 years of 84.02% ( $P < 0.01$ ; Figure 3D).

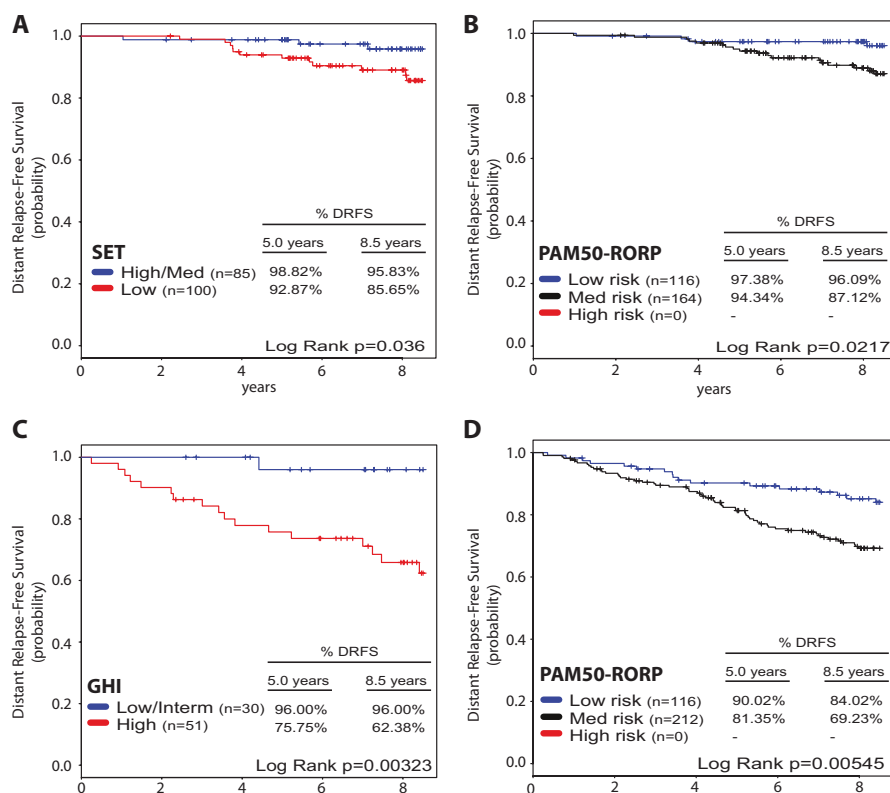
Within the luminal B subtype ( $n = 120$ ), the vast majority of signatures were found to be prognostic when evaluated as a continuous variable in node-negative disease (supplemental Table S6, available at *Annals of Oncology* online); however, no statistically significant group of patients with >90% DRFS at 8.5 years was identified by any of the predictors (supplemental Table S6, available at *Annals of Oncology* online); similar findings were obtained when we included the qRT-PCR PAM50 Nielsen series dataset. Finally, no significant prognostic ability

was found within node-positive luminal B tumors, with ( $n = 285$ ) or without ( $n = 70$ ) the Nielsen series, respectively (supplemental Table S6, available at *Annals of Oncology* online).

## discussion

Our data indicates that (i) clinically used signatures and ER-regulated gene signatures are tracking tumors with similar underlying biology (luminal A versus not) and show significant agreement in outcome predictions; (ii) the performance of these signatures is most relevant in node-negative disease; and (iii) some single genomic signatures can perform nearly as well as a combination of two or more signatures, although a combination of multiple signatures was statistically the best. Importantly, this is the first report to show that groups of patients with >95% DRFS at 8.5 years might only be consistently identified within node-negative and luminal A disease. Alternatively, for patients with luminal B cancer treated only with tamoxifen, additional therapies should be offered, which, as of today, would suggest chemotherapy.

These results also demonstrate that most of the signatures evaluated in this study can provide similar outcome predictions, although significant differences across predictors



**Figure 3.** Kaplan–Meier DRFS analysis of selected gene signatures within luminal A disease treated with adjuvant tamoxifen only. (A) SET index within node-negative luminal A tumors; (B) PAM50-RORP within node-negative luminal A tumors (Nielsen series included); (C) GHI within node-positive luminal A tumors; (D) PAM50-RORP within node-positive luminal A tumors (Nielsen series included). The complete survival analyses can be found in supplemental Tables S5 and S6, available at *Annals of Oncology* online. DRFS, distant relapse-free survival.

are present. This result is harmonious with our previous observation of concordance between intrinsic subtypes, NKI70 and GHI in a cohort of heterogeneously treated ER+ and ER– breast cancer patients [12]. Importantly, here, we show that these and other signatures are tracking ER+ tumors with a similar biology. Indeed, the vast majority of ER+ tumors identified here as having either basal-like, HER2-enriched or luminal B subtypes were correctly classified by the other signatures as having a poor prognosis. On the other hand, luminal A tumors were mostly identified as having good outcome and showing high expression of ER-regulated signatures. Interestingly, a recent neoadjuvant aromatase inhibitor clinical trial reported that the luminal A subtype benefits the most from endocrine therapy [26].

The performance of each predictor in node-positive disease was significantly worse when compared with node-negative disease, and almost no group of patients with node positivity had a DRFS >90% at 8.5 years by any predictor; the only exceptions being GHI within luminal A disease. In two previously published node-positive ER+ cohorts receiving adjuvant endocrine treatment only (TransATAC and SWOG-8814), the 9-year DRFS and 10-year disease-free survival estimates were 83% and 60% for the low-risk groups of the GHI, respectively [27, 28]. A plausible biological explanation is that in advanced luminal A primaries, a small percentage of cells within the bulk of the tumor have already metastasized and/or acquired endocrine resistance. Indeed, the presence of these subclones is supported by data from a neoadjuvant

endocrine trial [29]. However, within node-positive luminal A tumors, some patients with the low and intermediate risk score of GHI had a DRFS at 8.5 years >90%. Hence, future studies are warranted to determine if these or other predictors can identify, within the luminal A subtype, a group of node-positive patients whose survival with endocrine therapy could preclude the administration of adjuvant chemotherapy. The MINDACT [6] trial, which has completed accrual, and the RxPONDER trial (NCT01272037) will address this issue, particularly for patients with one to three positive lymph nodes.

Multivariate analyses including two predictors at a time revealed that, in most cases, many of these correlated predictors, in particular the PAM50-RORP, GHI, NKI70 and SET, remained statistically independent of each other (Table 3). This finding suggests that these predictors are not the same. In fact, at the individual level, the risk group assignment concordance among these predictors was found to be 36% for PAM50-RORP versus GHI, 54% for PAM50-RORP (low/med versus high) versus NKI70 and 74% for GHI (low/intermediate versus high) versus NKI70. Cohen’s kappa coefficients between risk group assignments were also indicative of slight to fair agreement (range 0.11–0.42) [30, 31]. The clinical relevance of this finding is currently unknown. However, a plausible explanation is that these signatures might be tracking different poor outcome luminal/ER+ subtypes; support for this heterogeneity comes from Parker et al. [10], where five statistically significant groups of luminal tumors were identified. Nonetheless, when we built a model here using all

predictors, we only observed modest improvements in performance. This finding suggests that gene expression profiling may be reaching its maximum prognostic power.

There are several important caveats to our analyses that must be recognized and always kept in mind when interpreting 'across platform' genomic studies. First, although we strove to implement each predictor as published, signatures developed on platforms other than the Affymetrix U133A were suboptimally implemented. This is because when taking a predictor from one technology and applying it to another platform, different oligonucleotide probes/sequences are used to represent each gene (and thus may not behave identically), and each technology has unique normalization methods. Second, changing platforms/technologies almost always causes a loss of genes (see supplemental Table S1, available at *Annals of Oncology* online), and this loss was significantly present for PAM50 (6/50) and NKI70 (12/60), which likely explains the observed lower performance of this predictor with respect to others. Nonetheless, many of the across platform evaluated predictors carried out well including the PAM50-ROR and GHI; the survival outcomes of the GHI low-risk group within node-negative disease was highly concordant to previous publications [32] despite that the absolute survival rates are highly dataset dependent. Finally, we could not compare the prognostic ability of these signatures versus standard clinicopathological variables since these variables were not available from most microarray datasets. This highlights the need for centralized collection of clinical and pathology data in all genomic studies.

To conclude independently derived genomic predictors of breast cancer recurrence perform similarly and are tracking tumors with similar biology. However, most predictors were statistically independent from the others and thus, these should not be considered to be interchangeable assays. From a clinical perspective, adding genomic signatures together provided modest improvements in outcome prediction, but may not be practical given cost.

## acknowledgements

This work was presented as an oral presentation (Abstract #502) at the American Society for Clinical Oncology Annual Meeting in Chicago, 2011.

## funding

NCI Breast SPORE program (P50-CA58223-09A1); (RO1-420 CA138255) Breast Cancer Research Foundation, the Sociedad Española de Oncología Médica (SEOM) and the V Foundation for Cancer Research. AP is affiliated to the Medicine PhD program of the Autonomous University of Barcelona, Spain.

## disclosure

CMP, PSB, TON and MJE are equity stock holders of University Genomics and BioClassifier LLC. CMP, PSB, MCUC, TON, MJE and JSP have filed a patent on the PAM50 assay. AP, CF, LDM, JB, SKLC and LAC have declared no conflicts of interest.

## references

- Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Mol Oncol* 2010; 5: 5–23.
- 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: 2817–2826.
- 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: 1183–1192.
- Cheang MCU, Chia SK, Voduc D et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst* 2009; 101: 736–750.
- Fan C, Prat A, Parker J et al. Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. *BMC Medical Genomics* 2010; 4: 3.
- Cardoso F, Van't Veer L, Rutgers E et al. Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol* 2008; 26: 729–735.
- Hammond MEH, Hayes DF, Dowsett M et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol* 2010; 28: 2784–2795.
- Wang Y, Klijn J, Zhang Y et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005; 365: 671–679.
- 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: 3207–3214.
- Parker JS, Mullins M, Cheang MCU et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009; 27: 1160–1167.
- Wirapati P, Sotiriou C, Kunkel S 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.
- Fan C, Oh DS, Wessels L et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006; 355: 560–569.
- Oesterreich S, Lee AV, Davidson NE. Is it time to ReSET the standard for estrogen receptor testing in breast cancer? *J Clin Oncol* 2010; 28: 4101–4103.
- Oh DS, Troester MA, Usary J et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol* 2006; 24: 1656–1664.
- Symmans WF, Hatzis C, Sotiriou C et al. Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol* 2010; 28: 4111–4119.
- 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: 1239–1246.
- Loi S, Haibe-Kains B, Desmedt C et al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* 2008; 9: 239.
- Zhang Y, Sieuwerts A, McGreevy M et al. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat* 2009; 116: 303–309.
- 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; 7: R953–R964.
- Hess KR, Anderson K, Symmans WF et al. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* 2006; 24: 4236–4244.
- Perou C, Parker J, Prat A et al. Clinical implementation of the intrinsic subtypes of breast cancer. *Lancet Oncol* 2010; 11: 718–719.
- Lusa L, McShane LM, Reid JF et al. Challenges in projecting clustering results across gene expression profiling datasets. *J Natl Cancer Inst* 2007; 99: 1715–1723.
- Nielsen TO, Parker JS, Leung S et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res* 2010; 16: 5222–5232.

24. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997; 16: 385–395.
25. Harrell F, Lee K, Califf R et al. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984; 3: 143–152.
26. Ellis M, Suman V, Hoog J et al. ACOSOG Z1031, a randomized phase 2 neoadjuvant comparison between letrozole, anastrozole and exemestane for postmenopausal women with ER rich stage 2/3 breast cancer: clinical and biomarker outcomes and the predictive value of the baseline PAM50-based intrinsic subtype. *J Clin Oncol* 2011; 29: 2342–2349.
27. Albain KS, Barlow WE, Shak S et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 2010; 11: 55–65.
28. Dowsett M, Cuzick J, Wale C et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol* 2010; 28: 1829–1834.
29. Ellis MJ, Tao Y, Luo J et al. Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. *J Natl Cancer Inst* 2008; 100: 1380–1388.
30. Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol* 2011; 9(1): 48–57.
31. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
32. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. *J Clin Oncol* 2008; 26: 721–728.