Clinical Cancer Research

A Comparison of PAM50 Intrinsic Subtyping with Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-Treated Estrogen Receptor–Positive Breast Cancer

Torsten O. Nielsen[1], Joel S. Parker[2], Samuel Leung[1], David Voduc[1], Mark Ebbert[3], Tammi Vickery[4], Sherri R. Davies[5], Jacqueline Snider[5], Inge J. Stijleman[3], Jerry Reed[4], Maggie C.U. Cheang[1,2], Elaine R. Mardis[4,6], Charles M. Perou[2], Philip S. Bernard[3], and Matthew J. Ellis[5,6]

Abstract

Purpose: To compare clinical, immunohistochemical (IHC), and gene expression models of prognosis applicable to formalin-fixed, paraffin-embedded blocks in a large series of estrogen receptor (ER)–positive breast cancers from patients uniformly treated with adjuvant tamoxifen.

Experimental Design: Quantitative real-time reverse transcription-PCR (qRT-PCR) assays for 50 genes identifying intrinsic breast cancer subtypes were completed on 786 specimens linked to clinical (median follow-up, 11.7 years) and IHC [ER, progesterone receptor (PR), HER2, and Ki67] data. Performance of predefined intrinsic subtype and risk-of-relapse scores was assessed using multivariable Cox models and Kaplan-Meier analysis. Harrell's C-index was used to compare fixed models trained in independent data sets, including proliferation signatures.

Results: Despite clinical ER positivity, 10% of cases were assigned to nonluminal subtypes. qRT-PCR signatures for proliferation genes gave more prognostic information than clinical assays for hormone receptors or Ki67. In Cox models incorporating standard prognostic variables, hazard ratios for breast cancer disease-specific survival over the first 5 years of follow-up, relative to the most common luminal A subtype, are 1.99 [95% confidence interval (CI), 1.09-3.64] for luminal B, 3.65 (95% CI, 1.64-8.16) for HER2-enriched subtype, and 17.71 (95% CI, 1.71-183.33) for the basal-like subtype. For node-negative disease, PAM50 qRT-PCR–based risk assignment weighted for tumor size and proliferation identifies a group with >95% 10-year survival without chemotherapy. In node-positive disease, PAM50-based prognostic models were also superior.

Conclusion: The PAM50 gene expression test for intrinsic biological subtype can be applied to large series of formalin-fixed, paraffin-embedded breast cancers, and gives more prognostic information than clinical factors and IHC using standard cut points. *Clin Cancer Res; 16(21); 5222–32. ©2010 AACR.*

Authors' Affiliations: [1]Genetic Pathology Evaluation Centre, Vancouver Coastal Health Research Institute, British Columbia Cancer Agency, and University of British Columbia, Vancouver, British Columbia, Canada; [2]Lineberger Comprehensive Cancer Center and Departments of Genetics, and Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; [3]Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, Utah; and [4]Department of Genetics, The Genome Center, Washington University School of Medicine, St Louis, Missouri; [5]Department of Medicine, Division of Oncology, Washington University School of Medicine, St Louis, Missouri; [6]Siteman Comprehensive Cancer Center, Washington University School of Medicine and Barnes-Jewish Hospital, St Louis, Missouri

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

Corresponding Author: Torsten O. Nielsen, Department of Pathology and Laboratory Medicine, University of British Columbia, Anatomical Pathology JP1401, Vancouver Hospital, 855 West 12th Avenue, Vancouver, British Columbia, Canada V5Z 1M9. Phone: 604-875-4111-66768; Fax: 604-875-4797; E-mail: torsten@interchange.ubc.ca.

doi: 10.1158/1078-0432.CCR-10-1282

Several gene expression technologies and statistical models have reported methodologies to identify breast cancer patients with estrogen receptor–positive (ER^+), node-negative (N0) disease that may be adequately managed with 5 years of tamoxifen monotherapy (1–5). However, these studies often included patients with tumors already associated with established low-risk biomarkers, for example, low-grade histology, low Ki67-based proliferation index, and favorable surgical stage. It therefore remains controversial whether genomic assays should be applied routinely, or whether surgical stage and a limited number of immunohistochemical (IHC) markers will, in most cases, be adequate and less costly (6).

The clinical significance of continued efforts in this area is relevant for decisions about both chemotherapy and endocrine agents, as patients at low risk after 5 years of tamoxifen monotherapy could be spared the morbidity

Translational Relevance

Molecular intrinsic subtyping reveals the major biological categories of breast cancer. Herein, we show adaptation of a 50-gene intrinsic subtyping signature for testing standard paraffin blocks. Using a large, homogeneously treated cohort of breast cancer patients, we directly compare gene expression results with high-quality clinical and central immunohistochemical data. We show the PAM50 approach to be superior as a prognostic test, specifically able to identify an ultralow-risk group who may not need chemotherapy. Based on these results, intrinsic subtyping tests are now being applied to randomized clinical trials series in Canada and the United States to assess predictive capacity (already under way for response to endocrine therapy, anthracyclines, and taxanes, with further studies under consideration). Should such studies prove a predictive value for intrinsic subtyping, this test could be clinically implemented in a similar form, as it has been designed for application on standard laboratory specimens.

associated with extended aromatase inhibitor therapy (7). Studies that address this issue are few because extremely long follow-up and information on breast cancer–specific mortality are required. Furthermore, because frozen tumor archives are unavailable from suitably large patient populations, gene expression technologies must be applicable to degraded RNA extracted from formalin-fixed, paraffin-embedded tissues that are necessarily more than a decade old.

Our group has assembled and published several technological and statistical approaches to address prognosis in ER^+ breast cancer. We therefore sought to compare clinicopathologic, IHC, and molecular methodologies in a single independent test set to identify the best approach. Importantly, we focused on fixed statistical models that were previously trained on independent data sets to avoid overoptimistic results. The models we report in this article include the use of standard pathologic factors, such as centrally reviewed histologic grade, as incorporated into Adjuvant! Online (8), models based on IHC for biomarkers of intrinsic subtypes (6), and a gene expression assay using 50 genes (PAM50). The latter represents a reduced gene set, amenable to assay by techniques such as quantitative real-time reverse transcription-PCR (qRT-PCR), which accurately identifies the major intrinsic biological subtypes of breast cancer and generates risk-of-relapse (ROR) scores (9). The investigation used a large independent cohort of formalin-fixed, paraffin-embedded pathology specimens from patients with ER^+ breast cancer, all M0 but otherwise representing a spectrum of T and N stages including a large fraction of node-positive (N^+) patients. All patients received adequate local treatment, 5 years of tamoxifen therapy but no adjuvant chemotherapy, and were followed for relapse-free survival (RFS) and disease-specific survival (DSS) for over a decade.

Materials and Methods

Patient and sample characteristics

The study cohort was accrued from female patients with invasive breast cancer, diagnosed in British Columbia between 1986 and 1992. Cancer tissue from these patients had been frozen and shipped to Vancouver Hospital for central ER and progesterone receptor (PR) testing by dextran-charcoal–coated (DCC) ligand-binding assay. The PAM50 assay was conducted on the portion of this tissue that was formalin fixed and paraffin embedded for histologic correlation. Characteristics of this cohort have been previously described (6), and the same source blocks were used to assemble tissue microarrays for previously published studies on ER (10), HER2 (11), PR (12), Ki67, cytokeratin 5/6, and epidermal growth factor receptor (6, 13). Quantitative ER was determined using the Ariol automated digital imaging system (14), and the same method was applied for PR. For this study, we selected samples from patients with ER^+ tumors by IHC who had received tamoxifen as their only adjuvant systemic therapy. Provincial guidelines from that time period recommended tamoxifen for women >50 years of age, whose ER status was positive or unknown, and who were either node positive or had lymphovascular invasion. Cohort identification and sample selection for this study are summarized as per REMARK criteria (15) in Supplementary Table S1.

RNA preparation, qRT-PCR, and assignment of biological subtype and ROR score

H&E sections from each block were reviewed by a pathologist (T.O.N.). Areas containing representative invasive breast carcinoma were selected and circled on the source block. Using a 1.0-mm punch needle, at least two tumor cores were extracted from the circled area. Details of RNA preparation from paraffin cores, the qRT-PCR assay for the PAM50 panel and reference genes, and how these results allow assignment into luminal A, luminal B, HER2-enriched, and basal-like subtypes, and the independently trained ROR-S (ROR based on subtype), ROR-T (ROR based on tumor size weighted model), ROR-P (ROR based on proliferation weighted model), and ROR-PT (ROR based on proliferation and tumor size weighted) risk score assignments are presented in Supplementary Materials and Methods. For clarity, the term ROR-T is now used for the same model described in our earlier publication as ROR-C ("clinical"; ref. 9).

Relation of clinicopathologic factors, intrinsic subtypes, and ROR scores to clinical outcome

Statistical analyses were conducted using SPSS v16.0 and R v2.8.0. Univariable analyses of tumor subtype against breast cancer RFS and DSS were done by Kaplan-Meier analysis with log-rank test. Multivariable analyses

Table 1. Clinical characteristics of the whole cohort

Clinical parameter		Total	PAM50 subtype (all *N* = 786)				
			Luminal A	Luminal B	HER2	Basal	Normal
Sample size	*n*	786	372	329	64	5	16
Follow-up times in recurrence-free patients (y)	Median (min-max)	9.7 (0.12-18)	12 (0.25-18)	7.6 (0.12-18)	7.3 (0.47-18)	2.3 (0.6-4.1)	13 (3.2-18)
Follow-up-times in disease-specific surviving patients (y)	Median (min-max)	12 (0.55-18)	13 (0.57-18)	10 (0.64-18)	8.8 (0.55-18)	5 (1.6-16)	14 (3.2-18)
Age (y)	Median	67	67	68	66.5	65	68.5
Premenopausal	Yes	20	10	7	2	1	0
	No	752	358	314	62	4	14
	Unknown/pregnant	14	4	8	0	0	2
Surgery	Complete mastectomy	468	210	203	39	5	11
	Partial mastectomy	306	159	119	23	0	5
	Other	12	3	7	2	0	0
Axillary node dissection	Yes	745	349	313	62	5	16
	No	41	23	16	2	0	0
Radiation therapy	Yes	419	207	164	40	1	7
	No	367	165	165	24	4	9
Tumor size (cm)	Median	2.1	2.0	2.5	2.5	3.5	2.3
T stage (clinical)	T0/IS	1	0	0	0	0	1
	T1	331	180	118	27	3	3
	T2	380	165	179	28	2	6
	T3	18	9	5	3	0	1
	T4	34	10	17	3	0	4
	TX	22	8	10	3	0	1
No. positive nodes	0	222	95	97	19	1	10
	1-3	360	182	148	26	1	3
	4-9	125	55	53	12	2	3
	10+	26	10	14	2	0	0
	Unknown	53	30	17	5	1	0
Grade	Grade 1: well differentiated	34	25	5	1	1	2
	Grade 2: moderately differentiated	338	186	129	14	0	9
	Grade 3: poorly differentiated	370	135	179	48	3	5
	Unknown	44	26	16	1	1	0
Histologic subtype	Ductal NOS	708	329	302	60	4	13
	Lobular	61	32	21	4	1	3
	Mucinous	7	4	3	0	0	0
	Tubular	7	6	1	0	0	0
	Medullary	2	1	1	0	0	0
	Apocrine	1	0	1	0	0	0
Lymphovascular invasion	Yes	485	210	220	44	2	9
	No	262	139	94	19	3	7
	Unknown	39	23	15	1	0	0
Clinical ER status (DCC ligand-binding assay)	Missing	9	5	3	0	0	1
	Negative (0-9 fmol/mg)	9	3	2	4	0	0
	Positive (>10 fmol/mg)	768	364	324	60	5	15
	Median (fmol/mg)	254.5	255.5	327.0	74.0	32.0	54.0

(*Continued on the following page*)

Table 1. Clinical characteristics of the whole cohort (Cont'd)

Clinical parameter		Total	PAM50 subtype (all *N* = 786)				
			Luminal A	Luminal B	HER2	Basal	Normal
Clinical PR status (DCC ligand-binding assay)	Missing	161	84	53	15	2	7
	Negative (0-9 fmol/mg)	72	15	39	18	0	0
	Positive (>10 fmol/mg)	553	273	237	31	3	9
	Median (fmol/mg)	129	202	84.5	17	153	239
IHC HER2 with FISH correction on 2+ cases	Negative	696	348	294	34	4	16
	Positive	75	15	30	29	1	0
	Unknown	15	9	5	1	0	0

Abbreviation: NOS, not otherwise specified.

were done against the standard clinical parameters of tumor size, nodal status, histologic grade, patient age, and HER2 status. HER2 scores were centrally determined based on assay of adjacent cores from the same source blocks, assembled into tissue microarrays, and subjected to IHC and fluorescence *in situ* hybridization (FISH) analysis using clinical-equivalent protocols (11). Cox regression models (16) were built to estimate the adjusted hazard ratios of the qRT-PCR–assigned breast cancer subtypes, as well as ROR scores categorized by published cut points and as a continuous variable. IHC-based subtypes were assigned as previously defined (6). The online decision-making tool Adjuvant! Online (http://www.adjuvantonline.com), previously validated on the British Columbia population cohort (8), was used to generate breast cancer RFS and DSS estimates for each patient in this cohort. Only cases with information for all the covariates were included in the analyses.

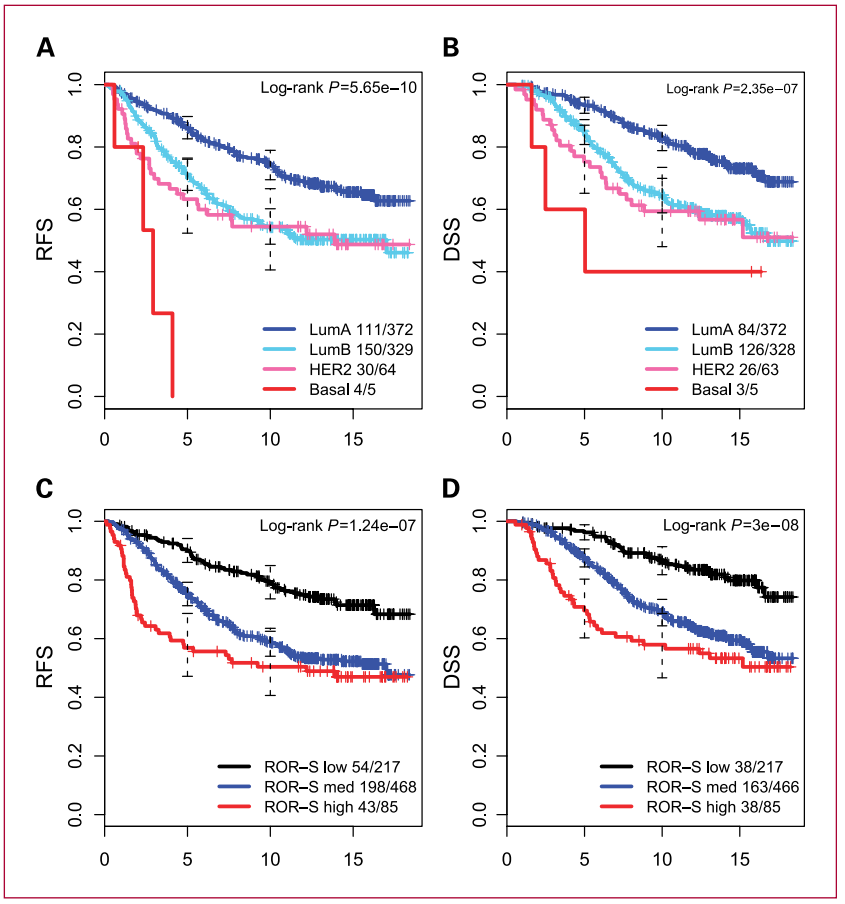

Fig. 1. Kaplan-Meier survival analysis of intrinsic subtype (A and B) and ROR-S (C and D), as determined by PAM50 gene expression measurement by qRT-PCR done on paraffin blocks from women with invasive breast carcinoma, treated with adjuvant tamoxifen. The number of events and total number of patients in each group are shown beside the description of each curve. B and D, breast cancer DSS (excludes two cases with unknown cause of death).

Smoothed plots of weighted Schoenfeld residuals were used to assess proportional hazard assumptions (17), and time stratifications were used where hazards were not proportional over the entire follow-up period.

The C-index (concordance index; ref. 18) is defined as the probability that risk assignments to members of a random pair are accurately ranked according to their prognosis. The number of concordant pairs (order of failure and risk assignment agree), discordant pairs (order of failure and risk assignment disagree), and uninformative pairs are tabulated to calculate the measure. C-index values of 0.5 indicate random prediction, and higher values indicate increasing prediction accuracy. Variability in the C-index for each predictor and *P* values from comparisons were estimated from 1,000 bootstrap samples of the risk assignments. Calculation was done using the rcorr.cens function implemented in the Hmisc (19) library for R statistical software version 2.8.1 (http://www.R-project.org).

Results

Intrinsic subtyping of ER$^+$, tamoxifen-treated breast cancer using the PAM50 assay

RNA was extracted from pathologist-guided tissue cores from 991 formalin-fixed, paraffin-embedded breast cancer specimens. Eight hundred and eleven samples yielded sufficient RNA for analysis (at least 1.2 μg total RNA at a concentration of ≥25 ng/μL). Template was technically sufficient in 786 cases, based on all internal housekeeper gene controls being expressed in the sample above background. Clinical characteristics for the patients included in the PAM50 analysis are presented in Table 1 (Supplementary Tables S2 and S3 provide details stratified by node status). Based on the nearest PAM50 centroid algorithm, intrinsic breast cancer subtypes were assigned using gene expression as follows: 372 samples (47.3%) were luminal A, 329 (41.9%) luminal B, 64 (8.1%) HER2 enriched, 5 (0.6%) basal-like, and 16 (2.0%) normal-like. Thus, although all cases in this study were positive for ER by centrally assessed IHC analysis on a tissue microarray (10), and 98.8% were also positive by DCC biochemical assay (Table 1), the gene expression panel nevertheless assigned 9% of cases into nonluminal subtypes, mostly HER2 enriched. This phenomenon has been previously observed when interrogating published data sets for expression of the PAM50 genes (9). For the 16 cases assigned as normal-like, histology was reviewed from adjacent tissue cores, and in 14 of 16 cases, invasive cancer cells were absent or rare. Normal-like cases were therefore excluded from outcome analyses, as a breast cancer subtype could not be confidently assigned due to insufficient tumor content.

The intrinsic biological subtypes were strongly prognostic by Kaplan-Meier analysis (Fig. 1A and B). In the British Columbia population at the time these samples were originally acquired, many patients with a clinically low-risk profile received no adjuvant systemic therapy (8). In contrast, those receiving adjuvant tamoxifen (the subjects of this study) had tumors that were mostly node positive and high grade, exhibited lymphovascular invasion, and therefore constitute a higher-risk group with overall 10-year RFS of 62% and DSS of 72%. Those assigned by the PAM50 assay to luminal A status had a significantly better outcome (10-year RFS, 74%; DSS, 83%) than luminal B, HER2-enriched, or basal-like tumors (Fig. 1A for RFS and Fig. 1B for DSS). The ROR algorithms (9) were originally trained on microarray data from N0 patients who received no adjuvant systemic therapy, and have not previously been applied to a population homogeneously treated with adjuvant tamoxifen, nor to a series containing large numbers of N$^+$ cases, nor to the endpoint of DSS. In this data set, ROR-S (a model based solely on gene expression) nevertheless showed performance consistent with our previous report (Fig. 1C and D). Multivariable Cox models were constructed to test the independent value of PAM50 subtyping against standard clinical and pathologic factors including age, histologic grade, lymphovascular invasion, HER2 expression, nodal status, and tumor size. To meet proportional hazard assumptions, multivariable models were assessed with the time axis split at 5 years (20), as HER2-enriched and basal-like tumors (Fig. 1A and B) and ROR-S high category tumors (Fig. 1C and D) had a much higher event rate in the first 5 years than subsequently. The intrinsic biological subtype and ROR-S remained significant in the multivariable models for DSS (Table 2) and RFS (Supplementary Table S4), particularly in the first 5 years, as did pathologic staging variables (tumor size and node status). However, histologic grade, lymphovascular invasion, and clinical HER2 status, significant in univariable analysis in this cohort, no longer contributed significant independent prognostic information when the multivariable analysis included the PAM50 assignments.

Comparisons between gene expression and clinical assays for hormone receptors and proliferation

In a case that is ER$^+$ by IHC, additional information about hormone receptor expression can be obtained in several ways, including DCC ligand-binding assay, quantitative IHC for ER, or equivalent measures of PR. Most published assays for breast cancer prognosis in ER$^+$ disease include tumor growth rate as one of the parameters in the statistical model, and this data set was previously assessed in detail for IHC Ki67 index (6). The PAM50 qRT-PCR data allow detailed quantitative assessment of the functionality of the estrogen response pathway (8-gene luminal signature) as well as a proliferation signature based on the mean expression of 11 genes linked to cell cycle progression (trained on published data, as per Supplementary Materials and Methods). The availability of all these measurements (10) provides an opportunity to determine which approach most accurately captures the prognostic effect of estrogen pathway biomarkers and tumor growth rate in a direct comparison (Fig. 2). Given a randomly selected pair of subjects, C-index is the probability that the patient assigned the more extreme risk score actually has a worse prognosis. A value of 0.5 indicates discrimination

Table 2. Cox model multivariable analysis of breast cancer DSS among ER$^+$, tamoxifen-treated women, incorporating standard clinicopathologic factors and (A) intrinsic subtype or (B) ROR-S, as determined by PAM50 qRT-PCR measurements

Clinical endpoint	Multivariable DSS (0-5 y of follow-up)		Multivariable DSS (5 y to end of follow-up)	
	Hazard ratio (95% CI)	*P*	Hazard ratio (95% CI)	*P*
A. Intrinsic subtype				
Age	1.02 (0.99-1.05)	0.2665	1.00 (0.98-1.02)	0.9786
Grade (1-2 vs 3)	1.51 (0.87-2.60)	0.1405	1.05 (0.71-1.56)	0.8109
Lymphovascular invasion	1.02 (0.58-1.81)	0.9421	1.16 (0.77-1.75)	0.4852
HER2 (IHC)	1.50 (0.77-2.91)	0.2314	0.82 (0.40-1.69)	0.5968
Node status (0 as reference group)		<0.0001		0.0012
1-3	2.07 (0.95-4.54)		1.54 (0.96-2.47)	
4+	5.80 (2.64-12.71)		2.78 (1.60-4.82)	
Tumor size (T1 as reference group)		0.049		0.0002
T2	1.22 (0.71-2.09)		1.62 (1.08-2.42)	
T3	3.92 (1.50-10.22)		5.11 (1.78-14.62)	
T4	1.31 (0.38-4.50)		4.02 (1.85-8.74)	
Subtype (luminal A as reference group)		0.0018		0.0381
Luminal B	1.99 (1.09-3.64)		1.70 (1.13-2.55)	
HER2 enriched	3.65 (1.64-8.16)		1.52 (0.72-3.18)	
Basal-like	17.71 (1.71-183.33)		NA	
B. ROR-S				
Age	1.02 (0.99-1.05)	0.2676	1.00 (0.98-1.02)	0.9089
Grade (1-2 vs 3)	1.36 (0.79-2.36)	0.2674	1.01 (0.68-1.51)	0.9588
Lymphovascular invasion	0.95 (0.54-1.66)	0.852	1.18 (0.78-1.79)	0.4299
HER2 (IHC)	1.46 (0.77-2.77)	0.2467	0.87 (0.43-1.77)	0.6964
Node status (0 as reference group)		<0.0001		0.0014
1-3	2.14 (1.00-4.60)		1.55 (0.97-2.48)	
4+	6.03 (2.79-13.05)		2.78 (1.59-4.86)	
Tumor size (T1 as reference group)		0.0647		0.0003
T2	1.19 (0.70-2.05)		1.64 (1.10-2.45)	
T3	3.34 (1.32-8.43)		3.69 (1.30-10.46)	
T4	0.90 (0.25-3.19)		4.44 (2.01-9.78)	
ROR-S (low as reference group)		<0.0001		0.0388
Med	2.04 (0.89-4.66)		1.86 (1.15-3.00)	
High	6.48 (2.56-16.40)		1.57 (0.71-3.46)	

NOTE: *P* values for multilevel categorical variables are derived from likelihood ratio tests between models with and without each these variables.
Abbreviation: 95% CI, 95% confidence interval.

that is no better than chance prediction, and a value of 1 indicates perfect discrimination of samples. Using the C-index to compare prognostic capacity in this uniformly tamoxifen-treated cohort, the combination of luminal genes measured by the PAM50 yields more prognostic information than other methods of hormone receptor analysis, but the differences are not significant. Although Ki67 index by IHC seems to outdo quantitative ER, the proliferation signature provides the most robust approach for the prediction of both RFS and DSS (Fig. 2; Supplementary Table S5). Multivariable analysis indicated that the Ki67 IHC assay did not contribute significant independent information to prognostic models for either N0 or N$^+$ breast cancer patients when information on the proliferation signature is included (Supplementary Table S6).

Comparison of fixed models of prognosis in N0 breast cancer

For formal model comparisons, data were generated on four fixed approaches, without any element of training within the test set: (*a*) clinical model based on Adjuvant! Online, (*b*) IHC-based (incorporating data on Ki67 and HER2), (*c*) the ROR-S approach based on PAM50 gene

expression alone, and (*d*) the proliferation signature alone and as incorporated into the ROR-P risk model using a β coefficient weighting for proliferation (described in Supplementary Materials and Methods). Adjuvant! Online incorporates full tumor size staging information; to account for the influence of tumor size, the biomarker models were also weighted by a β coefficient (T) that incorporated the prognostic information associated with T1 status versus higher T stage (the level of detail available in the independent training sets). This approach created IHC-T, ROR-T, and ROR-PT models. In addition, the strong independent influence of N stage was accounted for by conducting the analysis separately in the N0 and N^+ populations. C-index assessments showed superiority of the biomarker models over the purely clinical Adjuvant! Online model in the N0 population, with the ROR-PT approach providing the most prognostic information (Fig. 3A). In multivariable analysis, the addition of ROR-P to a model of ROR-S results in a significant increase in explained prognostic variation (RFS, $P = 0.0032$; DFS, $P = 0.0015$); ROR-PT is also significant after conditioning on ROR-S (RFS, $P = 0.0023$; DFS, $P = 0.0015$) but not ROR-P (RFS, $P = 0.12$; DFS, $P = 0.13$). A continuous score based on ROR-PT was generated to translate the data into an individual RFS and DSS risk assessment tool (Fig. 3B). Kaplan-Meier analysis illustrates the ability of the ROR-PT model to identify patients who have an extremely high chance (>95%) of remaining disease-free (Fig. 3C) and alive beyond 10 years (Fig. 3D). In contrast, our previously published IHC model (6) could not identify a group with sufficiently favorable outcomes that 5 years of tamoxifen might be considered adequate treatment (i.e., <90% 10-year RFS; Fig. 3E and F).

Comparison of fixed models of prognosis in N^+ breast cancer

For N^+ disease, C-index analysis (Fig. 4A) supports the conclusion that the ROR-T score produces the best prognostic model; in contrast to N0 disease, the proliferation signature added relatively little information and proliferation weighting (ROR-PT) did not yield a superior model. Adjuvant! Online performed almost as well, but had the advantage of incorporating the actual number of involved lymph nodes. This information was not available in the independent training sets used to build the ROR models, and so could not be used in the current analysis (which can, however, serve to train future models incorporating number of involved lymph nodes). The continuous score model for N^+ disease (Fig. 4B) produces a very broad range of prognosis, similar to N0 disease, although few patients have a prognosis in the range where tamoxifen monotherapy for 5 years would be considered sufficient treatment. Although there were large and highly significant differences in survival in ROR-defined risk groups, Kaplan-Meier analysis (Fig. 4C and D) illustrates that even patients in the lowest risk ROR group are still subject to relapses and late deaths from breast cancer, particularly after the 5th year of follow-up. The IHC-based risk model incorporating Ki67 and HER2 also produces a statistically significant prognostic effect for RFS (Fig. 4E) and DSS (Fig. 4F), although these differences are narrower than those achieved by the gene expression–based model.

Discussion

Previous studies have established that intrinsic biological signatures are present and have prognostic significance in breast cancer cohorts from multiple different institutions, profiled with several gene expression microarray platforms (21–24). To identify these subtypes on standard formalin-fixed, paraffin-embedded pathology specimens, we developed a qRT-PCR test based on a panel of 50 genes (9). The analysis reported here applied this test to a series of paraffin blocks with >15-year detailed follow-up.

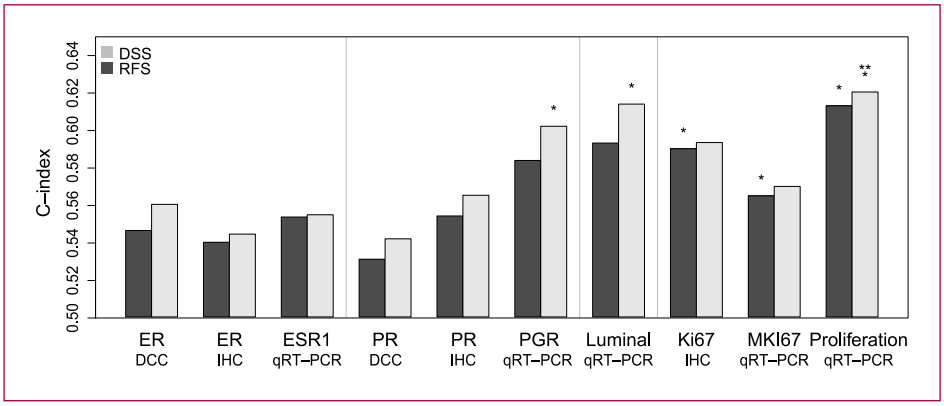

Fig. 2. C-index estimates of RFS and DSS for different measures of hormone receptors and proliferation. The luminal and proliferation measures are the means of normalized qRT-PCR values across 8- and 11-signature genes, respectively, as described in Supplementary Materials and Methods. *P* values were estimated from 1,000 bootstrap samples. Single asterisk (*) designates significant improvement ($P < 0.05$) in C-index relative to clinical quantitative ER by DCC ligand-binding assay. Double asterisk (**) designates significant improvement ($P < 0.05$) in C-index relative to visual quantitative Ki67 index.

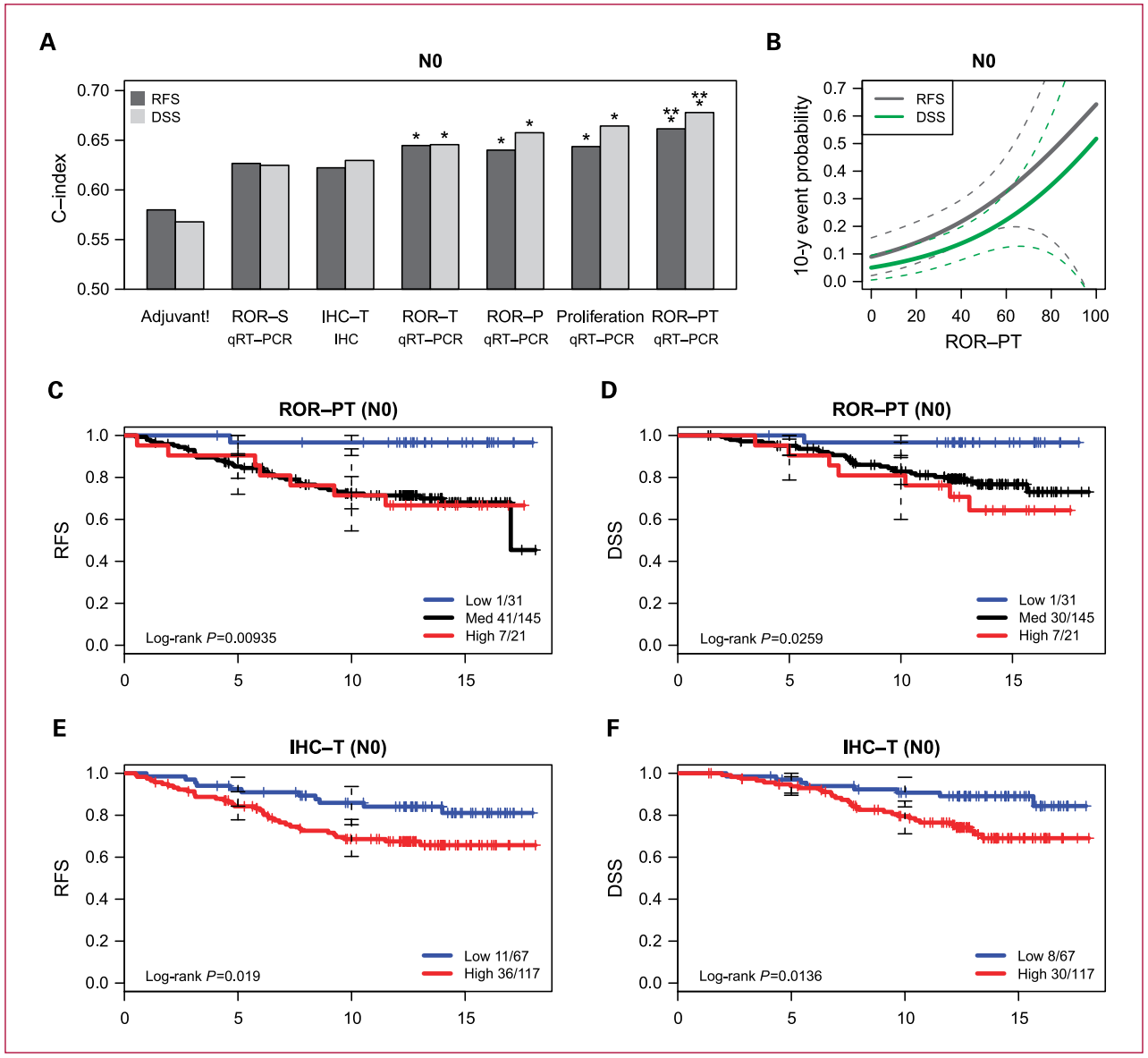

Fig. 3. Comparison of prognostic classifiers in N0 subjects. A, the C-index is used to compare accuracy of the prognostic classifiers (Supplementary Table S5). Asterisks denote significant improvement ($P < 0.05$) in C-index relative to the clinical model (Adjuvant!) (*), or relative to the IHC-T model (**). B, taking the best-performing model, ROR-PT values are related to actual 10-year event probabilities using a Cox proportional hazard model (dotted lines are 95% confidence interval). C and D, Kaplan-Meier survival analysis of the size and proliferation weighted ROR (ROR-PT) assignments. E and F, comparable information provided by a model of IHC subtype and tumor size. D and F, breast cancer DSS (excludes two cases with unknown cause of death).

Whereas previously assessed cohorts consisted mainly of low-risk women receiving no adjuvant systemic therapy, or were heterogeneously treated, the cases in the current study are all women with ER^+ breast cancer who received endocrine therapy as their sole adjuvant treatment, a group of particular clinical importance and contemporary relevance. In this analysis, we sought to compare different technologies for predicting long-term outcomes for such patients. In this study cohort, patients were diagnosed with N^+ or higher-risk N0 disease. Only 8% of the N0 population had grade 1 disease and 55% exhibited lymphovascular invasion (Table S2). Under the current standard of care in most countries, the majority of these patients would now be treated with adjuvant chemotherapy (25) and extended endocrine therapy. Using a series of fixed models trained in independent data sets, we compared a standard approach using clinicopathologic information (Adjuvant! Online) with our published luminal B discriminator based on Ki67 and HER2 IHC additionally weighted for T stage (IHC-T), and with PAM50 gene expression–based ROR models weighted for T stage (ROR-T and ROR-PT). In N0 patients, the ROR-PT approach was the most accurate

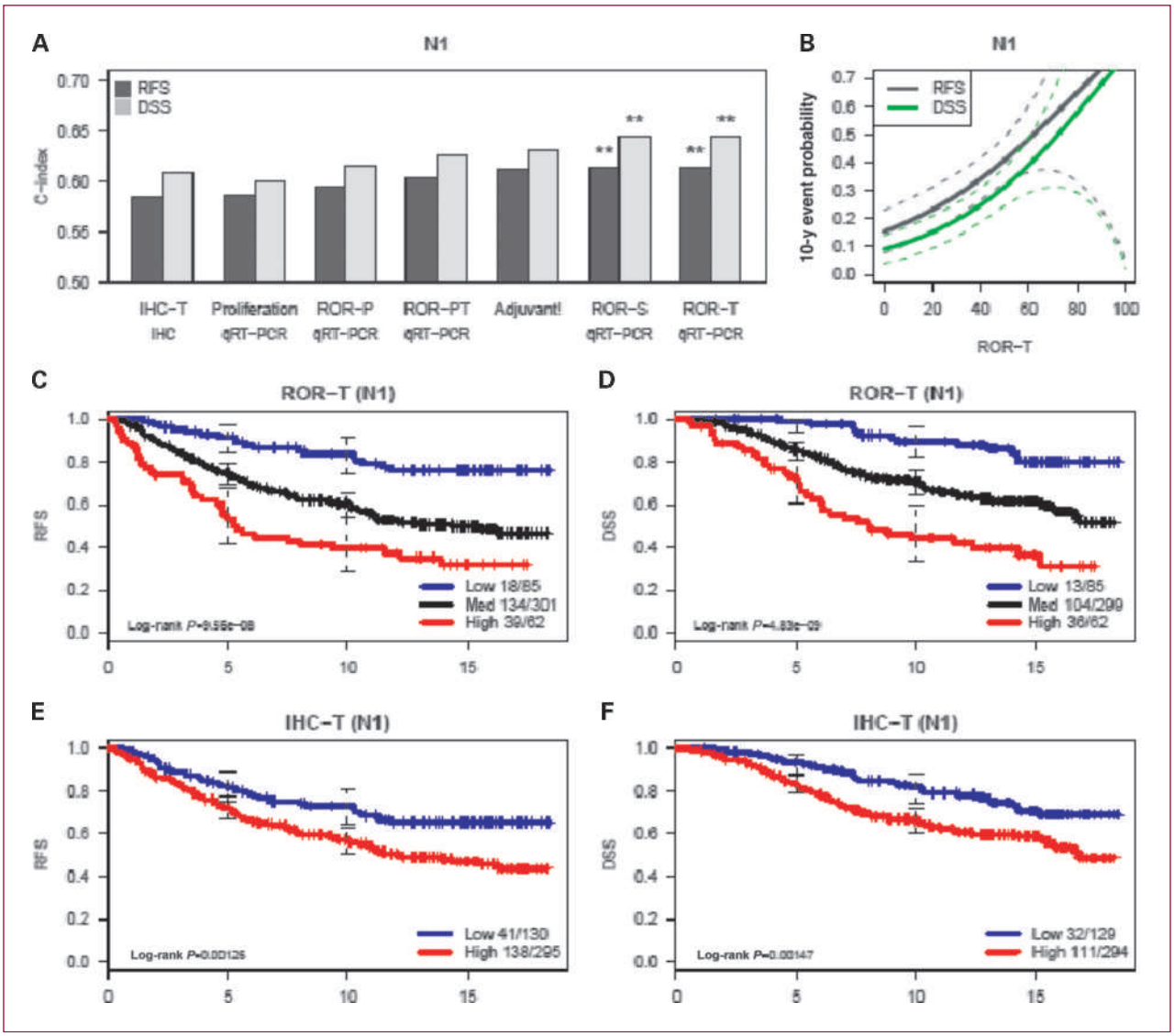

Fig. 4. Comparison of prognostic classifiers in N^+ subjects. A, C-index comparison of the accuracy of prognostic classifiers as described in Fig. 3. B, Cox proportional hazard model relating the best-performing model (ROR-T) to actual 10-year event probabilities. C and D, Kaplan-Meier survival analysis of the ROR-T assignments for RFS (C) and DFS (D). E and F, comparable information as provided by a model of IHC subtype and tumor size.

and was able to identify patients in whom 5 years of tamoxifen may be adequate treatment based on the very low late relapse rate in the 5- to 10-year window (Fig. 3C). In N^+ disease, the PAM50 approach represents an advance in prognostication, but late relapses and deaths were seen even in the lowest risk group identified using the best ROR model. Unlike in N0 disease, proliferation signature weighting did not improve the C-index in N^+ disease.

On this cohort, detailed centrally determined IHC analyses have previously been done and published (6, 10–13, 26). C-index, Kaplan-Meier, and Cox model analyses show that IHC approaches do work and provide significant prognostic information. However, the PAM50-based models are superior in terms of adding significant additional information and in their capacity to identify a particularly low-risk group of women.

We view these PAM50 models, derived from archival formalin-fixed RNA, as a potential replacement for grade-, hormone receptor–, Ki67-, and HER2-based prognostic models, but not as a replacement for pathologic stage (as tumor size and nodal status remain independent predictors in multivariable models that include PAM50-based prognostic information). One weakness of our approach is that our current accounting for pathologic stage is oversimplified due to the limited stage distributions and clinical information in our training sets. We analyzed the data as either N0 or N^+, and accounted for T

stage by categorizing the samples as either T1 or greater. A future aim is to integrate the PAM50 data into the Adjuvant! Online approach (27) to more completely account for the prognostic influence of pathologic stage. To achieve this, we would need to construct a training set that adequately includes all the five categories of T size and four categories of N stage used in Adjuvant! Online to gauge the prognostic weight of these pathologic stage categories in the setting of PAM50 information. Additionally, incorporation of all IHC data as continuous variables in a combined model may improve its prognostic value. The current series contains sufficiently detailed clinical and IHC information to contribute to such detailed comparisons as a training set requiring further validation.

An additional caveat to our study is that the population was strongly biased toward higher-risk breast cancers and so likely underestimates of the number of patients in the broader, N0 population for whom adjuvant tamoxifen would represent adequate treatment. The current generation of adjuvant aromatase inhibitor trials would be an appropriate setting to address the value of our approach further. We accept the possibility that a better model using Ki67 at a different cut point could be developed. However, because we were focused on comparing fixed models, we used our published approach. Further work on the Ki67 model and cut-point optimization will require independent data sets.

In comparison with other signatures such as the recurrence score and genomic grade index (1, 28, 29), the PAM50 has the potential advantage of discriminating high-risk patients into luminal B, HER2-enriched, and basal-like subtypes, who are likely to respond differently to the main systemic therapy options (endocrine, anti-HER2, and anthracycline versus nonanthracycline versus taxane chemotherapy regimens). The assay requires neither frozen tissue (30) nor manual microdissection of cut sections (1), and can be readily applied to standard paraffin blocks including archival tissues from clinical trials. Currently available assays such as Mammaprint (31) and Oncotype DX (32) were optimized to recognize particularly low-risk patients from among a N0 early-stage population who did not receive chemotherapy. Because intrinsic subtyping is designed to identify discriminative biological features of breast cancer, rather than being derived around clinical outcome in a specific population, this approach is particularly likely to extrapolate well onto other patient cohorts (33). The current study shows the ability of PAM50 to recognize a very low-risk prognostic group among women receiving tamoxifen and no chemotherapy, similar to the Oncotype DX assay (34, 35). A direct comparison of different expression profile approaches may become possible in the future through a reanalysis of cohorts with the PAM50 that have already been analyzed by Oncotype DX, because both assays can be applied to the same source material.

Our inability to identify a group of patients with N^+ disease in whom 5 years of tamoxifen is adequate is reminiscent of the recent findings from the Southwest Oncology Group, who also found that a molecular signature for good outcome in N0 disease failed in N^+ disease in this regard (35). It would be relevant to study a series of patients treated with extended adjuvant aromatase inhibitor therapy, who will have even lower residual risk, as some of the patients in the low-risk N^+ group may simply require longer treatment with modern endocrine therapy rather than chemotherapy. The development of new approaches for defining prognosis in N^+ disease is also warranted. We have already established the preoperative endocrine prognostic index, which showed that the "on endocrine treatment" Ki67 value is more effective than baseline Ki67 for the identification of patients with clinical stage II and III disease who have excellent long-term outcomes after neoadjuvant endocrine therapy (36). A comparison between Ki67 and the PAM50-based proliferation signature in the neoadjuvant endocrine therapy setting is therefore one logical next step. The applicability of this test to formalin-fixed, paraffin-embedded tissues will make possible its use on large clinical trial archives that address this issue (37). The results of our study highlight the feasibility of measuring multigene expression panels on such series as a means for showing clinical utility using a method readily applicable to prospective clinical samples that provides more prognostic information than clinical or standard IHC approaches.

Disclosure of Potential Conflicts of Interest

T.O. Nielsen, C.M. Perou, M.J. Ellis, P.S. Bernard: ownership interest, Bioclassifier LLC; U.S. Patent No. 61/057,508.

Acknowledgments

We thank current and former members of the British Columbia Cancer Agency's Breast Cancer Outcomes Unit, including S. Chia, K. Gelmon, H. Kennecke, I. Olivotto, and C. Speers, for maintaining the clinical database.

Grant Support

T. Nielsen is a Senior Scholar of the Michael Smith Foundation for Health Research. Grant support was provided by National Cancer Institute (NCI) Strategic Partnering to Evaluate Cancer Signatures grant U01 CA114722-01, Canadian Cancer Society, Huntsman Cancer Institute/Foundation (P.S. Bernard), ARUP Institute for Clinical and Experimental Pathology (P.S. Bernard), NCI Breast Specialized Program of Research Excellence grant P50-CA58223-09A1 (C.M. Perou), St. Louis Affiliate of the Susan G. Komen Foundation CRAFT (M.J. Ellis), Breast Cancer Research Foundation (C.M. Perou and M.J. Ellis), and Sanofi-Aventis Canada unrestricted educational grant. Additional support provided by the TRAC facility and Informatics at the Huntsman Cancer Center, supported in part by NCI Cancer Center Support grant P30 CA42014-19, and the tissue procurement facility at the Alvin J. Siteman Cancer Center at Washington University School of Medicine, which is funded in part by the NCI Cancer Center Support grant P30 CA91842.

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 05/13/2010; revised 07/29/2010; accepted 08/25/2010; published OnlineFirst 09/13/2010.

References

1. 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.
2. 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.
3. Goetz MP, Suman VJ, Ingle JN, et al. A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. Clin Cancer Res 2006;12:2080–7.
4. Ross JS. Multigene classifiers, prognostic factors, and predictors of breast cancer clinical outcome. Adv Anat Pathol 2009;16:204–15.
5. Tutt A, Wang A, Rowland C, et al. Risk estimation of distant metastasis in node-negative, estrogen receptor-positive breast cancer patients using an RT-PCR based prognostic expression signature. BMC Cancer 2008;8:339.
6. Cheang MC, Chia SK, Voduc D, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst 2009;101:736–50.
7. Goss PE, Ingle JN, Martino S, et al. A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. N Engl J Med 2003;349:1793–802.
8. Olivotto IA, Bajdik CD, Ravdin PM, et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. J Clin Oncol 2005;23:2716–25.
9. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160–7.
10. Cheang MC, Treaba DO, Speers CH, et al. Immunohistochemical detection using the new rabbit monoclonal antibody SP1 of estrogen receptor in breast cancer is superior to mouse monoclonal antibody 1D5 in predicting survival. J Clin Oncol 2006;24:5637–44.
11. Chia S, Norris B, Speers C, et al. Human epidermal growth factor receptor 2 overexpression as a prognostic factor in a large tissue microarray series of node-negative breast cancers. J Clin Oncol 2008;26:5697–704.
12. Liu S, Chia SK, Mehl E, et al. Progesterone receptor is a significant factor associated with clinical outcomes and effect of adjuvant tamoxifen therapy in breast cancer patients. Breast Cancer Res Treat 2009;119:53–61.
13. Cheang MC, Voduc D, Bajdik C, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res 2008;14:1368–76.
14. Turbin DA, Leung S, Cheang MC, et al. Automated quantitative analysis of estrogen receptor expression in breast carcinoma does not differ from expert pathologist scoring: a tissue microarray study of 3,484 cases. Breast Cancer Res Treat 2008;110:417–26.
15. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumor MARKer prognostic studies (REMARK). Nat Clin Pract Oncol 2005;2:416–22.
16. Cox D, Oakes D. Analysis of survival data. Monographs on statistics and probability. London (United Kingdom): Chapman and Hall; 1984.
17. Grambsch P, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994;81:515–26.
18. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87.
19. Harrell FE, Jr. Design Package, in R package version 2.3-0. 2009.
20. Schemper M. Cox analysis of survival data with non-proportional hazard functions. The Statistician 1992;41:455–65.
21. Calza S, Hall P, Auer G, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res 2006;8:R34.
22. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 2006;355:560–9.
23. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006;7:96.
24. Kapp AV, Jeffrey SS, Langerod A, et al. Discovery and validation of breast cancer subtypes. BMC Genomics 2006;7:231.
25. Goldhirsch A, Wood WC, Gelber RD, Coates AS, Thurlimann B, Senn HJ. Progress and promise: highlights of the international expert consensus on the primary therapy of early breast cancer 2007. Ann Oncol 2007;18:1133–44.
26. Jensen KC, Turbin DA, Leung S, et al. New cutpoints to identify increased HER2 copy number: analysis of a large, population-based cohort with long-term follow-up. Breast Cancer Res Treat 2008;112:453–9.
27. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J Clin Oncol 2001;19:980–91.
28. Ivshina AV, George J, Senko O, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res 2006;66:10292–301.
29. 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.
30. Glas AM, Floore A, Delahaye LJ, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. BMC Genomics 2006;7:278.
31. 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.
32. Paik S. Development and clinical utility of a 21-gene recurrence score prognostic assay in patients with early breast cancer treated with tamoxifen. Oncologist 2007;12:631–5.
33. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 2005;11:5678–85.
34. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 2006;24:3726–34.
35. 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 2009;11:55–65.
36. 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;100:1380–8.
37. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009;101:1446–52.

Supplementary methods, *Nielsen et al.*, A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor positive breast cancer**.**

RNA preparation from extracted paraffin cores:

RNA was recovered using the High Pure RNA Paraffin Kit including an on-column DNase I treatment to remove any residual DNA (Roche Applied Science, Indianapolis IN). RNA yields were assessed using an ND-1000 Spectrophotometer (NanoDrop Technologies, Rockland DE), and samples used for analysis where the concentration of total RNA exceeded 25 ng/uL and yield exceed 1.2 ug.

qRT-PCR:

cDNA synthesis was completed using a mixture of random hexamers and gene-specific primers, and qPCR was performed with the Roche LightCycler 480 instrument as previously described (1, 2). Each 384-well plate contained samples and a calibrator in triplicate with 2.5 ng cDNA and 10 ng cDNA, respectively per reaction. A tumor sample was considered of insufficient quality if any of the reference controls (*ACTB*, *PSMC4*, *RPLP0*, *MRPL19*, or *SF3A1*) failed. Since the expression of one of the 50 discriminator genes may not be detected by PCR due to a technical failure or due to the biology of the sample (i.e. that gene is not expressed by the tumor at a measureable level) we developed validation rules to distinguish between these possibilities. A technical failure occurred if 1) the calibrator failed in all replicates, 2) the calibrator Cps were out of range for that gene, or 3) all calibrators had secondary melting peaks greater than three-quarter the height of the specific melting peak. If the negative and calibrators passed, then a gene could be assigned a low copy value. The low copy was a Cp=40 for that gene. This value was assigned if either 1) all 3 sample replicates had Cp=0, or 2) all sample replicates had a Cp>35 with non-specific melting peaks, or 3) there was PCR amplification with a Cp>38. If both low copy and a measurable copy number value were assigned for sample replicates, then the measurable copy value was used.

Subtype assignment by the 50 gene classifier

Gene expression microarrays have been extensively used to characterize gene expression variation in breast cancer. The predominant variation has been characterized as 5 “intrinsic” subtypes of breast cancer that have been repeatedly observed (3-5). A 50 gene qRT-PCR assay has been constructed to provide stable and highly repeatable subtype classification in relevant tissue specimens (FFPE and frozen). A complete description of the development and initial evaluation of this assay was described in Parker et al.(1).

Development of the classifier began with a set of over 200 samples with 20k expression measurements by microarray and 160 genes measured by qRT-PCR. The microarray data allowed identification of gold

standard specimens representing each of the five primary breast cancer subtypes. The set of 160 qRT-PCR genes was then optimized to identify a reduced subset of genes from which accurate subtype classification could still be performed. A set of 50 genes was finalized using cross-validation. The 50 genes, corresponding centroids, and other parameters of the nearest centroid breast cancer intrinsic subtype classifier are available in Parker et al. (1).

The qRT-PCR assay consists of these 50 genes and an additional 5 housekeeper genes used for sample normalization. Analysis by qRT-PCR is performed by first normalizing the raw Ct values to gene specific technical controls followed by normalization to the sample controls (housekeepers) (6). The distance to each centroid is calculated using Spearman's rank correlation. The centroid associated with the largest positive correlation value is assigned as the subtype of the sample. This protocol produces reproducible subtype correlations and subtype assignment independently for each sample.

Risk models of the 50 gene based subtype

Subtype classification of a breast tumor may provide insight into the prognosis and illuminate potential treatment strategies. However, variation exists within each subtype, and this variation carries useful prognostic information. Thus the prognostic value of the gene expression information is not limited to the final subtype assignment.

Development of the ROR risk of relapse models consisted of assignment of the five subtype correlations to each sample as described above. This was performed using set of node negative patients with no adjuvant systemic therapy. The four tumor subtype (normal-like excluded) correlations were related to relapse free survival as additive terms in a Cox model. Ridge regression was performed in the context of cross-validation to optimize the weights for each term. This model of the four tumor subtype correlations (equation 1) provides the subtype based risk of relapse (ROR-S):

(equation 1) ROR-S = 0.05*Basal + 0.12*Her2 -0.34*LumA + 0.23*LumB (1)

Tumor size was found to be an independent prognostic factor and was included in the model as an additive term. Tumor size was limited to binary information (<=2cm vs >2cm) as this was the data available in the training set. The weights were re-learned using the same process to produce a combined model of tumor size and subtype. This model was previously termed ROR-C (1) and is given in equation 2. The current manuscript has renamed this metric from ROR-C to ROR-T to distinguish this combined model from other combined models that are evaluated.

(equation 2) ROR-T = ROR-C = 0.05*Basal + 0.11*Her2 -0.23*LumA + 0.09*LumB + 0.17*T (1)

Additional models were explored using the same process as was used to develop these models. The process of model development and the training data for these novel models is identical to that described above (and more completely in (1)) for ROR-S and ROR-T. The motivation behind these models

was to incorporate proliferation information as an independent factor with respect to subtype. The 50 gene set contains many genes that are known markers of proliferation. A set of 11 proliferation genes (table M1) was highly correlated (r > 0.8) in the qRT-PCR and microarray training sets described in Parker et al [1].

Table M1

Genes in proliferation index		
GeneName	UniGene	EntrezGene
BIRC5	Hs.514527	332
CCNB1	Hs.23960	891
CDC20	Hs.524947	991
NUF2	Hs.651950	83540
CEP55	Hs.14559	55165
NDC80	Hs.414407	10403
MKI67	Hs.80976	4288
PTTG1	Hs.350966	9232
RRM2	Hs.226390	6241
TYMS	Hs.592338	7298
UBE2C	Hs.93002	11065

These 11 genes were summarized by averaging the normalized expression estimates in each sample. This proliferation signature index was modeled as an additional term to the four subtypes to produce ROR-P (equation 3), and in addition to the model of four subtypes and tumor size to produce ROR-PT (equation 4).

(equation 3) ROR-P = -0.001*Basal + 0.7*Her2 -0.95*LumA + 0.49*LumB + 0.34*Prolif

(equation 4) ROR-PT = -0.001*Basal + 0.73*Her2 -0.9*LumA + 0.05*LumB + 0.13*T + 0.33*Prolif

Thresholds for all four models were identified in the training set using the same process. The distribution of risk scores from each model was stratified by the subtype assignment. The low risk threshold for each model was assigned as the minimum of all Luminal B scores, and the high risk threshold was the maximum of all Luminal A scores.

Luminal gene signature

Evaluation of the information content of ER protein and *ESR1* gene expression are compared to a summarized luminal gene expression signature. This gene expression signature was constructed in a similar fashion to the proliferation signature. Specifically, a set of 8 genes (table M2) was highly correlated (r > 0.8) in qRT-PCR and microarray training sets from Parker et al [1] and clearly

distinguished luminal tumors. These 8 genes were summarized by averaging the normalized expression estimates in each sample.

Table M2

Genes in luminal signature		
GeneName	UniGene	EntrezGene
BAG1	Hs.377484	573
ESR1	Hs.208124	2099
FOXA1	Hs.163484	3169
GPR160	Hs.231320	26996
NAT1	Hs.591847	9
MAPT	Hs.101174	4137
MLPH	Hs.102406	79083
PGR	Hs.368072	5241

Risk models based on standard clinical data and immunohistochemically-determined intrinsic subtype

The accuracy of risk classification is an important measure for comparing the various models. Other variables such as standard clinical risk factors and immunohistochemistry (IHC) based subtype may also be compared in this context if appropriate models are available. Pathological grade and IHC based subtype assignments are correlated with the 50 gene based subtype assignments, and independent of pathological stage with respect to prognostic information (1, 7). Formal models of these variables are needed to evaluate which of these three biologic measures best compliments pathologic stage to produce the most accurate classification.

The online decision making tool Adjuvant! Online (www.adjuvantonline.com) was used to generate breast cancer specific survival (BCSS) estimates for each patient in this cohort (8). The clinical and pathologic factors used to generate the risk estimates were age, tumor size, tumor grade, ER status, number of positive lymph nodes, and lymphovascular invasion.

A model of immunohistochemical subtype and tumor size (IHC-T) was constructed following the same process outlined for development of the ROR scores. Specifically, the two categorical variables were treated as additive terms in the Cox models, fit with Ridge regression, and optimized with cross-validation. Cross-validation was performed by randomly selecting 2/3 of the samples for training with the remaining 1/3 used for validation. The process was repeated 100 times to evaluate model performance. IHC subtype definitions for Luminal A and B follow the protocol described in Cheang et al. (7), and incorporate the information from ER, PR, HER2 and Ki67 immunohistochemistry: Luminal A = (ER or PR) positive and HER2 negative and Ki67 <= 13%; Luminal B = (ER or PR) positive and (HER2 positive or Ki67 > 13%); HER2 = ER negative and PR negative and HER2 positive; basal = ER negative and PR negative and HER2 negative and (ck5/6 or EGFR) positive.

The data set used for training these models has been previously described (6) and was derived concurrently with the test set data on the same tissue microarray series (making the technical aspects of immunohistochemistry performance and scoring identical between training and test set). This training

cohort consists of 1545 node negative samples from patients who received no adjuvant systemic therapy. Tumor size, IHC subtype, pathological grade, and lymphovascular invasion were significant predictors ($p < 0.05$) of relapse free survival in standard Cox multivariable analysis. Grade-T and IHC-T models were evaluated during cross-validation. As with the ROR models, relapse free survival, where any relapse was considered an event, was used for training. Accuracy estimates from cross-validation are illustrated in Figure M1. The final risk scores assigned to each possible level of a sample are provided in table M3. These values were assigned to the estrogen receptor positive, tamoxifen-treated series samples (described in the main paper) as appropriate for the comparative evaluations presented.

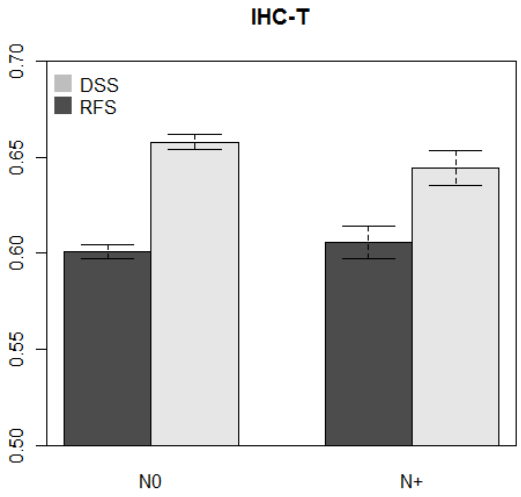

Figure M1. Accuracy estimates from cross-validation of pathological models in the training set. The C-index was calculated during each of 100 rounds of cross-validation from the "left-out" node negative samples. The bar height represents the mean C-index and corresponding standard error.

Table M3

	IHC-T	
IHC subtype	Tumor size	Risk Score
Luminal A	<= 2cm	0
Luminal A	> 2cm	0.23
Luminal B	<= 2cm	0.4
Luminal B	> 2cm	0.64
Her2	<= 2cm	0.44
Her2	> 2cm	0.67
Basal	<= 2cm	0.15
Basal	> 2cm	0.38

Supplementary References

1. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160-7.
2. Mullins M, Perreard L, Quackenbush JF, et al. Agreement in breast cancer classification between microarray and quantitative reverse transcription PCR from fresh-frozen and formalin-fixed, paraffin-embedded tissues. Clin Chem 2007;53:1273-9.
3. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52.
4. 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;100:8418-23.
5. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006;7:96.
6. Perreard L, Fan C, Quackenbush JF, et al. Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. Breast Cancer Res 2006;8:R23.
7. Cheang MC, Voduc D, Bajdik C, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res 2008;14:1368-76.
8. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J Clin Oncol 2001;19:980-91.

Supplemental Table S1. Cohort identification and sample selection.

Selection criteria	Sample Size
British Columbia breast cancer cases submitted for central ER testing (1986-1992). Cases with complete outcome and FFPE block containing histologically representative invasive breast cancer.	4046
Cases receiving Tamoxifen as sole adjuvant systemic therapy	1305
IHC ER positive on tissue microarray	1075
Samples sent for qRT-PCR	991
RNA extraction with 1.2 ug total RNA yield at a concentration of at least 25 ng/uL	811
Samples with PAM50 assignments	786
Luminal A/B, HER2-enriched or Basal-like (Normal-like breast cancer – excluded from analyses)	770 (16)

Supplemental Table S2. Clinical characteristics of the node negative (N0) cohort.

Clinical parameter		Total	PAM50 subtype (N0: n=222)				
			Luminal A	Luminal B	Her2	Basal	Normal
Sample Size	**N**	222	95	97	19	1	10
follow-up times in recurrence-free patients	**median (min. – max.)**	12 (0.5-18)	13 (0.97-18)	12 (0.56-18)	12 (0.5-16)	4.1 (4.1-4.1)	14 (3.7-17)
followup-times in disease specific surviving patients	**median (min. – max.)**	13 (1.4-18)	13 (2.1-18)	12 (1.4-18)	12 (2.3-16)	16 (16-16)	14 (4.1-17)
age (in years)	**median**	66	66	68	64	44	71
pre-menopausal	**yes**	6	1	3	1	1	0
	no	211	94	90	18	0	9
	unknown / pregnant	5	0	4	0	0	1
surgery	**complete mastectomy**	115	43	55	10	1	6
	partial mastectomy	107	52	42	9	0	4
	other	0	0	0	0	0	0
axillary node dissection	**yes**	222	95	97	19	1	10
	no	0	0	0	0	0	0
radiation therapy	**yes**	112	54	44	10	1	3
	no	110	41	53	9	0	7
tumour size (cm)	**median**	2.0	1.9	2.1	2.5	9.9	2.3
T stage (clinical)	**T0 / IS**	1	0	0	0	0	1
	T1	99	52	39	5	0	3
	T2	105	39	51	11	1	3
	T3	2	1	0	1	0	0
	T4	9	2	3	1	0	3
	TX	6	1	4	1	0	0
# positive nodes	**0**	222	95	97	19	1	10
	1-3	0	0	0	0	0	0
	4-9	0	0	0	0	0	0
	10+	0	0	0	0	0	0
	unknown	0	0	0	0	0	0
grade	**grade 1: well differentiated**	18	13	2	0	1	2
	grade 2: moderately differentiated	100	50	40	4	0	6
	grade 3: poorly differentiated	97	29	51	15	0	2
	unknown	7	3	4	0	0	0
histologic subtype	**ductal NOS**	189	77	86	18	0	8
	lobular	25	11	10	1	1	2

	mucinous	2	2	0	0	0	0
	tubular	5	5	0	0	0	0
	medullary	1	0	1	0	0	0
	apocrine	0	0	0	0	0	0
lymphovascular invasion	**yes**	118	42	59	12	0	5
	no	97	48	36	7	1	5
	unknown	7	5	2	0	0	0
clinical estrogen receptor status (DCC)	**missing**	3	2	1	0	0	0
	negative (0-9 fmol/mg)	2	2	0	0	0	0
	positive (>10 fmol/mg)	217	91	96	19	1	10
	median (fmol/mg)	261.0	279.0	324.5	178.5	32.0	53.0
immunohistochemical HER2 with FISH correction on 2+ cases	**negative**	200	91	87	11	1	10
	positive	18	1	9	8	0	0
	unknown	4	3	1	0	0	0

Supplemental Table S3. Clinical characteristics of the node positive cohort.

Clinical parameter		Total	PAM50 subtype (N+: n=511)				
			Luminal A	Luminal B	Her2	Basal	Normal
Sample Size	**N**	511	247	215	40	3	6
follow-up times in recurrence-free patients	**median (min. – max.)**	8.2 (0.12-18)	11 (0.25-18)	6.3 (0.12-18)	6.5 (0.47-17)	2.0 (0.6-2.3)	7.8 (3.2-18)
followup-times in disease specific surviving patients	**median (min. – max.)**	11 (0.55-18)	13 (0.57-18)	8.3 (0.64-18)	7.9 (0.55-17)	2.5 (1.6-16)	11 (3.2-18)
age (in years)	**median**	67	67	68	67	65	63.5
pre-menopausal	**yes**	13	9	3	1	0	0
	no	490	234	209	39	3	5
	Unknown/ pregnant	8	4	3	0	0	1
surgery	**Complete mastectomy**	336	158	144	26	3	5
	partial mastectomy	171	87	69	14	0	1
	other	4	2	2	0	0	0
axillary node dissection	**yes**	510	247	214	40	3	6
	no	1	0	1	0	0	0
radiation therapy	**yes**	283	140	111	28	0	4
	no	228	107	104	12	3	2
tumour size (cm)	**median**	2.4	2.0	2.5	2.5	3.5	2.3
T stage (clinical)	**T0 / IS**	0	0	0	0	0	0
	T1	206	112	73	19	2	0
	T2	262	118	123	17	1	3
	T3	13	6	5	1	0	1
	T4	17	5	10	1	0	1
	TX	13	6	4	2	0	1
N stage (pathological)	**N0**	0	0	0	0	0	0
	N1	496	242	208	37	3	6
	N2	14	5	6	3	0	0
	N3	1	0	1	0	0	0
	NX	0	0	0	0	0	0
# positive nodes	**0**	0	0	0	0	0	0
	1-3	360	182	148	26	1	3
	4-9	125	55	53	12	2	3
	10+	26	10	14	2	0	0
	unknown	0	0	0	0	0	0
grade	**grade 1: well differentiated**	12	9	2	1	0	0
	grade 2: moderately differentiated	216	123	81	9	0	3
	grade 3: poorly differentiated	251	95	122	29	2	3
	unknown	32	20	10	1	1	0
histologic subtype	**ductal NOS**	473	227	201	37	3	5
	lobular	33	18	11	3	0	1
	mucinous	3	1	2	0	0	0
	tubular	1	1	0	0	0	0
	medullary	0	0	0	0	0	0
	apocrine	1	0	1	0	0	0
lymphovascular invasion	**yes**	343	158	152	27	2	4
	no	139	73	51	12	1	2
	unknown	29	16	12	1	0	0
clinical estrogen receptor status (DCC)	**missing**	4	1	2	0	0	1
	negative (0-9 fmol/mg)	7	1	2	4	0	0
	positive (>10 fmol/mg)	500	245	211	36	3	5
	median (fmol/mg)	253.0	252.5	311.0	67.0	29.0	73.0

immunohistochemical HER2 with FISH correction on 2+ cases	**negative**	450	231	190	20	3	6
	positive	51	11	21	19	0	0
	unknown	10	5	4	1	0	0

Supplementary Table S4. Cox model multivariable analysis of breast cancer relapse free survival among ER positive, tamoxifen-treated women, incorporating standard pathological and clinical factors and (A) intrinsic subtype or (B) Risk of Relapse (ROR-S), as determined by PAM50 qRT-PCR measurements. p-values for multilevel categorical variables are derived from likelihood ratio tests between models with and without each of these variables.

A. Intrinsic subtype

Clinical endpoint	relapse-free survival (0 – 5 years followup)		relapse-free survival (5 years to end of followup)	
	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value
Age	1.00 (0.98-1.02)	0.9813	0.99 (0.97-1.02)	0.579
Grade (1-2 vs 3)	1.17 (0.80-1.71)	0.4112	1.10 (0.73-1.68)	0.6424
LVI	1.01 (0.67-1.52)	0.9699	1.31 (0.84-2.06)	0.2382
HER2 (IHC)	1.54 (0.93-2.56)	0.0941	0.32 (0.10-1.04)	0.0587
Node status		***< 0.0001***		***0.0104***
1-3	***1.71 (1.03-2.83)***		1.49 (0.90-2.46)	
4+	***4.12 (2.44-6.96)***		***2.56 (1.39-4.72)***	
Tumor size		***0.0077***		***< 0.0001***
T2	***1.05 (0.72-1.53)***		***2.03 (1.30-3.15)***	
T3	***3.55 (1.69-7.42)***		***7.69 (2.32-25.56)***	
T4	1.07 (0.38-3.00)		***5.61 (2.44-12.89)***	
Subtype		***0.0003***		0.1444
Luminal B	***1.82 (1.20-2.76)***		1.52 (0.99-2.34)	
Her2-enriched	***2.82 (1.55-5.13)***		1.07 (0.44-2.64)	
Basal-like	***9.75 (1.82-52.34)***		NA	

B. Risk of Relapse (ROR-S)

Clinical endpoint	relapse-free survival (0 – 5 years followup)		relapse-free survival (5 years to end of followup)	
	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value
Age	1.00 (0.98-1.02)	0.8668	0.99 (0.97-1.02)	0.5048
Grade (1-2 vs 3)	1.07 (0.73-1.57)	0.7317	1.10 (0.71-1.68)	0.672
LVI	0.92 (0.62-1.38)	0.697	1.29 (0.82-2.04)	0.2647
HER2 (IHC)	1.57 (0.98-2.53)	0.0607	0.32 (0.10-1.05)	0.0593
Node status		***< 0.0001***		***0.0244***
1-3	***1.67 (1.02-2.73)***		1.43 (0.86-2.36)	
4+	***4.32 (2.57-7.27)***		***2.37 (1.27-4.42)***	
Tumor size		***0.0008***		***< 0.0001***
T2	1.03 (0.70-1.51)		***2.08 (1.33-3.24)***	
T3	***3.81 (1.94-7.50)***		***5.94 (1.78-19.84)***	
T4	0.91 (0.32-2.61)		***6.05 (2.61-14.02)***	
ROR-S		***<0.0001***		0.2495
Med	***2.00 (1.17-3.44)***		1.46 (0.90-2.38)	
High	***4.54 (2.34-8.79)***		1.01 (0.41-2.45)	

Supplemenatary Table S5: C-index mean values and 95% confidence intervals from Figures 2, 3A and 4A, in tabular form. Confidence intervals for the C-index estimates and p-values for pairwise model comparisons were calculated empirically from 1000 bootstrap samples.

Figure 2		Relapse Free Survival	Comparison to ER DCC	Comparison to Ki67 IHC%	Disease Specific Survival	Comparison to ER DCC	Comparison to Ki67 IHC%
		C-index (95% CI)	p-value	p-value	C-index (95% CI)	p-value	p-value
ER	DCC	0.54 (0.50-0.58)	---	0.94	0.56 (0.51-0.60)	---	0.76
ER	IHC%	0.55 (0.51-0.59)	0.34	0.89	0.56 (0.52-0.60)	0.44	0.74
PR	DCC	0.57 (0.52-0.60)	0.17	0.75	0.59 (0.55-0.63)	0.16	0.43
PR	IHC%	0.55 (0.51-0.59)	0.35	0.90	0.57 (0.52-0.61)	0.40	0.73
Luminal	qRT-PCR	0.58 (0.54-0.62)	0.073	0.62	0.60 (0.56-0.65)	0.058	0.22
Ki67	IHC%	0.59 (0.54-0.62)	0.062	---	0.59 (0.54-0.63)	0.24	---
Proliferation	qRT-PCR	0.62 (0.58-0.66)	0.006	0.031	0.62 (0.58-0.66)	0.035	0.025

Figure 3A		Relapse Free Survival	Comparison to Adjuvant!	Comparison to IHC-T	Disease Specific Survival	Comparison to Adjuvant!	Comparison to IHC-T
		C-index (95% CI)	p-value	p-value	C-index (95% CI)	p-value	p-value
Proliferation	qRT-PCR	0.65 (0.58-0.72)	0.042	0.21	0.67 (0.58-0.75)	0.029	0.19
Adjuvant!	clinical	0.57 (0.49-0.65)	---	0.92	0.56 (0.46-0.65)	---	0.96
IHC-T	IHC	0.62 (0.53-0.69)	0.081	---	0.63 (0.54-0.71)	0.037	---
ROR-S	qRT-PCR	0.63 (0.55-0.71)	0.11	0.41	0.63 (0.53-0.71)	0.12	0.52
ROR-T	qRT-PCR	0.65 (0.57-0.72)	0.003	0.11	0.66 (0.57-0.73)	0.004	0.13
ROR-P	qRT-PCR	0.65 (0.57-0.72)	0.048	0.24	0.66 (0.57-0.75)	0.034	0.24
ROR-PT	qRT-PCR	0.67 (0.59-0.74)	0.001	0.047	0.69 (0.60-0.76)	0.002	0.033

Figure 4A		Relapse Free Survival	Comparison to Adjuvant!	Comparison to IHC-T	Disease Specific Survival	Comparison to Adjuvant	Comparison to IHC-T
		C-index (95% CI)	p-value	p-value	C-index (95% CI)	p-value	p-value
Proliferation	qRT-PCR	0.58 (0.54-0.62)	0.87	0.60	0.60 (0.55-0.64)	0.82	0.65
Adjuvant!	clinical	0.61 (0.57-0.65)	---	0.17	0.63 (0.58-0.67)	---	0.24
IHC-T	IHC	0.59 (0.54-0.63)	0.83	---	0.61 (0.56-0.65)	0.76	---
ROR-S	qRT-PCR	0.61 (0.57-0.65)	0.50	0.15	0.65 (0.60-0.69)	0.28	0.088
ROR-T	qRT-PCR	0.61 (0.56-0.65)	0.58	0.10	0.64 (0.59-0.68)	0.24	0.035
ROR-P	qRT-PCR	0.59 (0.55-0.63)	0.77	0.42	0.62 (0.57-0.66)	0.66	0.42
ROR-PT	qRT-PCR	0.60 (0.55-0.64)	0.72	0.31	0.62 (0.57-0.67)	0.59	0.30

Supplementary Table S6: Cox model multivariable analysis of breast cancer relapse free survival and disease specific survival among ER positive, tamoxifen-treated women, incorporating standard pathological and clinical factors in addition to proliferation measures to evaluate (A) node negative and (B) node positive subjects. p-values for multilevel categorical variables are derived from likelihood ratio tests between models with and without each of these variables.

A. Node Negative Subjects

Outcome	Relapse free survival		Disease specific survival	
	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value
Age	1.01 (0.97-1.05)	0.5657	1.03 (0.98-1.07)	0.24773
Grade (1-2 vs 3)	1.22 (0.62-2.43)	0.56	0.93 (0.43-2.01)	0.85
LVI	1.17 (0.60-2.29)	0.65	1.37 (0.64-2.94)	0.42
HER2 (IHC)	0.30 (0.07-1.27)	0.10	0.18 (0.02-1.39)	0.10
Tumor size		0.14		0.085
T2	1.66 (0.85-3.21)		1.94 (0.91-4.11)	
T3	NA		NA	
T4	NA		NA	
Proliferation (qRT-PCR)	***4.12 (1.32-12.85)***	***0.0148***	***6.04 (1.55-23.55)***	***0.00959***
Ki67 (IHC)	0.99 (0.96-1.02)	0.48	0.99 (0.96-1.02)	0.46

B. Node Positive Subjects

Outcome	Relapse free survival		Disease specific survival	
	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value
Age	0.99 (0.97-1.01)	0.4032	1.00 (0.98-1.02)	0.8607
Grade (1-2 vs 3)	1.08 (0.78-1.49)	0.64	1.24 (0.87-1.78)	0.23
LVI	1.15 (0.80-1.65)	0.44	1.08 (0.73-1.60)	0.71
HER2 (IHC)	1.67 (1.02-2.74)	0.40	1.89 (1.12-3.16)	0.02
Tumor size		***<0.0001***		***<0.0001***
T2	***1.45 (1.04-2.02)***		***1.53 (1.06-2.22)***	
T3	***4.66 (2.42-9.00)***		***4.62 (2.22-9.60)***	
T4	***3.00 (1.52-5.96)***		***3.33 (1.65-6.73)***	
Proliferation (qRT-PCR)	***2.64 (1.50-4.62)***	***0.0007***	***3.04 (1.61-5.72)***	***0.0006***
Ki67 (IHC)	0.99 (0.97-1.01)	0.20	0.98 (0.96-1.01)	0.14