EarlyR: A Robust Gene Expression Signature for Predicting Outcomes of Estrogen Receptor–Positive Breast Cancer

Steven A. Buechler,[1] Yesim Gökmen-Polar,[2] Sunil S. Badve[2,3,4]

Abstract

Currently available molecular signatures assess the risk of recurrence and the benefit of chemotherapy; however, these tests may have large intermediate risk groups, limiting their usefulness. We describe a novel 5-gene signature that is a robust prognostic assay that performed similarly to currently available signatures in concordance analyses. However, it identified significantly fewer patients as intermediate risk and more as low risk than currently available assays.

Introduction: Early stage estrogen receptor (ER)-positive breast cancer may be treated with chemotherapy in addition to hormone therapy. Currently available molecular signatures assess the risk of recurrence and the benefit of chemotherapy; however, these tests may have large intermediate risk groups, limiting their usefulness. **Methods:** The EarlyR prognostic score was developed using integrative analysis of microarray data sets and formalin-fixed, paraffin-embedded–based quantitative real-time PCR assay and validated in Affymetrix data sets and METABRIC cohort using Cox proportional hazards models and Kaplan-Meier survival analysis. Concordance index was used to measure the probability of prognostic score agreement with outcome. **Results:** The EarlyR score and categorical risk strata (EarlyR-Low, EarlyR-Int, EarlyR-High) derived from expression of *ESPL1*, *MKI67*, *SPAG5*, *PLK1* and *PGR* was prognostic of 8-year distant recurrence-free interval in Affymetrix (categorical $P = 3.5 \times 10^{-14}$; continuous $P = 8.8 \times 10^{-15}$) and METABRIC (categorical $P < 2.2 \times 10^{-16}$; continuous $P < 10^{-16}$) data sets of ER^+ breast cancer. Similar results were observed for the breast cancer–free interval end point. At most 13% of patients were intermediate risk and at least 66% patients were low risk in both ER^+ cohorts. The EarlyR score was significantly prognostic (distant recurrence-free interval; $P < .001$) in both lymph node–negative and lymph node–positive patients and was independent from clinical factors. EarlyR and surrogates of current molecular signatures were comparable in prognostic significance by concordance index. **Conclusion:** The 5-gene EarlyR score is a robust prognostic assay that identified significantly fewer patients as intermediate risk and more as low risk than currently available assays. Further validation of the assay in clinical trial–derived cohorts is ongoing.

Clinical Breast Cancer, Vol. 19, No. 1, 17-26
Keywords: Estrogen receptor-positive, Gene signature, Molecular diagnostic, Prognosis

Introduction

Classification and management of a disease significantly reflects understanding of the disease condition. Until recently, breast cancer was believed to be a single disease that was treated by surgical excision followed by chemotherapy, with the addition of tamoxifen for estrogen receptor (ER)-positive disease. Molecular analysis of breast cancers using gene expression microarrays resulted in the recognition of breast cancer as a heterogeneous disease in which different subtypes respond to distinct therapeutic regimens.[1,2] In recent years, numerous genomic assays, including Oncotype DX,[3] Mammaprint,[4,5] Prosigna (Risk of Recurrence, ROR),[6] EndoPredict,[7] and Breast Cancer Index,[8,9] were developed to help inform physicians' treatment decisions for adjuvant therapy. National Comprehensive Cancer Network treatment guidelines for ER^+, human epidermal growth factor receptor 2 (HER2) negative, > 0.5 cm tumors, with no lymph node (LN) involvement, recommend Oncotype DX Recurrence Score (RS) testing, followed by hormone

[1]Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN
[2]Department of Pathology
[3]Department of Medicine, Indiana University School of Medicine, Indianapolis, IN
[4]Indiana University Melvin and Bren Simon Cancer Center, Indianapolis, IN

Submitted: Jan 7, 2018; Revised: May 1, 2018; Accepted: Jul 9, 2018; Epub: Jul 17, 2018

Address for correspondence: Sunil S. Badve, MD, FRCPath, Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, 635 Barnhill Dr, MS A128, Indianapolis, IN 46202
E-mail contact: sbadve@iupui.edu

1526-8209/$ - see frontmatter
https://doi.org/10.1016/j.clbc.2018.07.011

therapy alone for low-risk patients (RS $\leq$ 18), and hormone therapy and adjuvant chemotherapy for high-risk patients (RS $\geq$ 31). Physicians may alternatively use prognostic information from other molecular signatures to guide treatment. Multiple studies have reported a reduction in the proportion of ER^+ patients receiving chemotherapy concurrent with adoption of Oncotype DX.[10,11]

The impact of such a genomic assay on treatment decisions depends in part on the proportions of patients with disease classified as low risk, high risk, or intermediate risk. The tests listed above identify approximately 50% of LN^-, ER^+ patients as low risk. The Oncotype DX assay has a large intermediate risk group (38% and 40% in 2 clinical use studies,[10,12] respectively), for which a treatment recommendation is unclear.

Herein, we describe a gene signature "EarlyR" using a novel probe expression analysis methodology[13] for the prognostication of ER^+ breast cancer. EarlyR may be applied as a continuous score or in low-, intermediate-, and high-risk strata. In the ER^+, LN^-, $HER2^-$ tumors in METABRIC cohort,[14] EarlyR identified 72%, 12%, and 16% as low, intermediate, and high risk, respectively. This is significantly more low-risk patients, and significantly fewer intermediate-risk patients, than reported by currently available assays (eg, Mammaprint and Oncotype DX[15,16]). To build evidence of the clinical utility of the test, we converted it to a proof-of-principle quantitative real-time PCR (qPCR)-based assay for formalin-fixed, paraffin-embedded (FFPE) tissue.

Methods

Microarray Data Sets

The Affymetrix training set used in this study was obtained from the LN^- samples in GSE3494[17] and GSE7390[18] (Gene Expression Omnibus; http://www.ncbi.nlm.nih.gov). The validation set was derived from patients from the following data sets: GSE12093,[19] GSE6532,[20] GSE2034,[21] GSE11121,[22] and GSE17705.[23] The CEL files from all series were normalized together and expression values computed with the GCRMA (GC Robust Multiarray Average) package.[24] Batch effects were eliminated with the ComBat tools.[25]

The patient characteristics of the Affymetrix (training and validation) and METABRIC cohorts are described in Supplemental Table 1 in the online version. None of the patients in the Affymetrix cohorts had received chemotherapy, and 46% had received hormone therapy. Patients in METABRIC cohort[14] were treated with hormone therapy and/or chemotherapy as directed by the treating physician. In contrast to current practice, hormone therapy was predominately prescribed only for patients with positive LNs or large tumors (> 2 cm); only 54% of LN^- patients received hormone therapy. A tumor in METABRIC cohort was considered $HER2^+$ if there was gain in the number of copies of *ERBB2* as assessed using microarray-based copy number analysis.[14]

The prognostic significance of EarlyR was studied as per the STEEP guidelines.[26] Distant recurrence-free interval (DRFI) was defined as the time from surgery to recurrent distant metastatic breast cancer; breast cancer–free interval was defined as the time from surgery to recurrent distant metastatic breast cancer or locoregional invasive ipsilateral breast cancer. Data for both breast cancer–free interval and DRFI were obtained for patients in the METABRIC cohort (unpublished data). In Affymetrix cohorts, the end points were described as distant metastasis-free survival. It is unlikely that differences between study-specific end points and DRFI would result in significant changes in the number of events in these data sets. Prognostic significance with respect to DRFI and breast cancer–free interval was assessed up to 8 years after diagnosis. The threshold of 8 years was chosen on the basis of a prior publication from the Cuzick group showing that the prognostic utility of current genomic signatures for ER^+ breast cancer deteriorates after 8 years.[27]

Sample Selection and Preparation

The institutional review board of Indiana University approved the study. An informed consent waiver was obtained, and only deidentified data were used in the analyses.

Archival FFPE tumor blocks were chosen from 72 patients with breast cancer at the Indiana University Simon Cancer Center based on their Oncotype DX RS. Initial real-time qPCR analysis was conducted using 10 samples of ER^+ breast cancers. This was followed by qPCR analysis using customized arrays of 23 cases with high RS, 26 cases with intermediate RS, and 23 cases with low RS. Demographic and clinical characteristics of the patients were acquired from medical charts (Supplemental Table 2 in the online version). The cases were equally divided into training and validation sets, each of 36 cases. The distribution of RS in the training set was shown to be significantly equivalent to the distribution of RS in the validation set using the Kolmogorov-Smirnoff test ($P = .88$).

RNA was extracted from 10 μm thick sections of archival paraffin blocks using RecoverAll Total Nucleic Acid Isolation Kit (Life Technologies, Grand Island, NY) according to the manufacturer's instructions. The quality of RNA was assessed using the Nanodrop ND-1000 spectrophotometer (ThermoScientific, Wilmington, DE). Total RNAs were reverse transcribed using the High Capacity cDNA Reverse Transcription kit (Life Technologies) according the manufacturer's instructions.

Selection of TaqMan qPCR Assays

Specific target sequences for each probe from the Human Genome U133A 2.0 Array were obtained using the NetAffx Analysis Center (http://www.affymetrix.com/analysis/index.affx). Target sequences were aligned to the appropriate messenger RNA reference sequence (REFSEQ) accession number using National Center for Biotechnology Information (NCBI) BLAST (Basic Local Alignment Search Tool) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and accessed the consensus sequence through the NCBI Entrez nucleotide database.

Using the UMapIt mapping tool of Applied Biosystems (ABI, Foster City, CA), the Affymetrix probe IDs were mapped to TaqMan assays specific to each sequence. TaqMan assays, where necessary custom designed using Primer Express (ABI), were tested for the amplification efficiency on the basis of the ABI-defined criteria. Control RNA (Universal Human Reference RNA; Stratagene, San Diego, CA) and FFPE samples were used to test the efficiency of the probes. On the basis of the observed efficiency, probes were selected for custom array microfluidic cards (TaqMan assays; Supplemental Table 3 in the online version).

qPCR Analysis Using Custom Arrays

TaqMan reactions were performed in triplicate using custom array microfluidic cards preloaded with TaqMan Gene Expression

Assays containing 17 genes (12 discriminant genes and 5 reference genes) on an ABI Prism 7900HT Fast Real-Time platform according to the manufacturer's instructions (Supplemental Table 3 in the online version). *ACTB, TFRC, GUS, RPLPO,* and *GAPDH* were used as endogenous reference controls for normalization. Delta threshold cycle values for each of the 12 genes of interest were normalized using these endogenous controls according to the method of ABI DataAssist v3.0 software.

Construction of Genomic Signature From Gene Expression Measurements

The methodology for construction of a genomic signature is described in detail in the Supplemental Methods in the online version.

Statistical Analyses

All statistical analyses were performed using R (R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org/). Mixture models were fit using the package *mclust*,[28,29] and survival analysis was performed with the package *survival*. The significance of a Cox proportional hazard (CPH) model was assessed with the *P* value of the log-rank score test. The significance of a multivariate CPH over a CPH using a subset of the variables was measured with a chi-square test of the log likelihoods. The proportional hazard condition was tested with the cox.zph function.

The prognostic significance of genomic signatures was compared using the concordance index.[30,31] The concordance index estimates the probability that, for a random pair of patients, the patient with earlier recurrence has a higher score than the patients with either later or no recurrence. The concordance index is a number between 0 and 1 and is defined more formally in the Supplemental Methods in the online version. A concordance index greater than 0.5 indicates that the prognostic score is more significant than random chance. A confidence interval for concordance index was computed by resampling. A function from the *survcomp* package was used to compute concordance index.

Computation of Alternative Genomic Signatures

To compare the prognostic significance of EarlyR with that of Oncotype DX, Mammaprint, and ROR score, we computed surrogates of these signatures in METABRIC using the Bioconductor *genefu* package.[32]

To compute Oncotype DX, we selected probes in the IlluminaHuman-v3 platform (Illumina, San Diego, CA) representing the 16 target genes in the panel;[3] for genes represented by multiple probes, we selected the probe with the highest variance in the ER^+ METABRIC cohort, the recommended method in *genefu*. Probes representing all 16 target genes were identified, and the RS was computed using the package's function for that purpose. To accommodate for possible differences between expression values assessed by the Illumina platform and the native RT-PCR platform, we computed surrogate low-risk, intermediate-risk, and high-risk strata so that the percentages in ER^+, LN^- patients match those found in NSABP B-14.[3]

Mammaprint was derived from a 70-gene signature,[4] referred to as GENE70 in *genefu*. GENE70 was originally computed using 70 array probes from the Agilent Hu25K array platform (Agilent Technologies, Santa Clara, CA), of which 56 were associated with 52 unique Entrez IDs. For these 52 Entrez IDs, Illumina probes were selected that had the highest variance in METABRIC. From these probes, and the appropriate *genefu* function, a continuous GENE70 score was computed. The GENE70 stratification into low-risk and high-risk groups was computed using a GENE70 score threshold that produced a low-risk group containing 50% of the ER^+, LN^- METABRIC cohort.

The ROR score[6] was computed in METABRIC using the appropriate *genefu* function with the default arguments. The ROR stratification was created with the same percentages in low-, intermediate-, and high-risk groups for ER^+ METABRIC as for trans-ATAC,[33] specifically, 55%, 25%, and 20%, respectively.

Results

Discovery of EarlyR

To identify the gene signature, an integrative approach consisting of analysis of *in silico* data and FFPE samples was used (Supplemental Figure 1 in the online version). This was undertaken to ensure stability of the probes in fresh and frozen tissue and across multiple analytical platforms. Prior analysis of GSE4922 (UPPS), GSE6532 (OXFD, GUYT), GSE7390 (TRANSBIG), GSE9195 (GUYT2), and GSE11121 (MZ)[13] led to the identification of a set of 12 genes (*ESPL1, CDC45L, PLK1, CENPA, MKI67, SPAG5, CDT1, PGR, CXCL9, PHLPP1, CDC6, PRPF4*) that provided prognostic information in these ER^+ breast cancer samples. To determine the feasibility of using these probes for a prognostic signature with FFPE tissue, we performed a qPCR analysis of these 12 genes in a training set of 36 ER^+ breast cancer FFPE samples with known Oncotype DX RS (Supplemental Table 2 in the online version). For each of the 12 target genes on the qPCR array, risk scores were derived using the $\Delta-CT$ values from the training set of 36 samples (Supplemental Methods). The 9 genes whose risk scores were significantly predictive of TAILORx risk group (*P* value of the linear model $< .05$) were considered for further gene signature development.

The 9 genes (*ESPL1, CDC45L, PLK1, CENPA, MKI67, SPAG5, CDT1, PGR, CXCL9*) identified by the above method were further analyzed for inclusion in a multigene signature in the Affymetrix training data set of 266 ER^+, LN^- breast cancers obtained from GSE3494[17] and GSE7390.[18] The incremental impact of addition of each gene to the prognostic score was analyzed starting with the most prognostic gene, *ESPL1*. Next, the top 2 genes (*ESPL1* and *SPAG5*) were combined to derive a 2-gene signature score (Supplemental Methods in the online version). This process was continued until the multigene score with maximally significant results was obtained (Supplemental Methods in the online version). This signature, EarlyR score, uses the genes *ESPL1, MKI67, SPAG5, PLK1,* and *PGR*. Specifically, the EarlyR score was computed in the Affymetrix-based cohorts using a gene signature derived from the expression of the following 5 probes: 204817_at (*ESPL1*), 212022_s_at (*MKI67*), 203145_at (*SPAG5*), 202240_at (*PLK1*), and 208305_at (*PGR*).

Concordance of EarlyR and RS in FFPE Samples

To reconfirm the ability of EarlyR to assess risk in FFPE tissues, we showed that the Oncotype DX RS is linearly dependent on the

EarlyR stratification ($P = .001$, Supplemental Figure 2 in the online version) in the FFPE validation set (Supplemental Table 2 in the online version). Samples were further separated into risk groups with respect to RS using the Oncotype DX thresholds[3] and the TAILORx thresholds.[34] We found a significant concordance between EarlyR risk strata and Oncotype DX risk strata ($P = .004$) and TAILORx risk strata ($P = .002$). This confirms the feasibility of using EarlyR for the analysis of FFPE samples.

Validation of EarlyR in Affymetrix and METABRIC Cohorts

The EarlyR score and the EarlyR strata, EarlyR-Low (EarlyR $\leq$ 25), EarlyR-Int (25 $<$ EarlyR $\leq$ 75), and EarlyR-High (75 $<$ EarlyR) were computed from the score values in all validation cohorts (GSE12093,[19] GSE6532,[20] GSE2034,[21] GSE11121,[22] GSE17705,[23] and METABRIC). This computation was performed in a manner blinded to all clinical features of the samples (Supplemental Methods in the online version).

EarlyR score classified a large majority of samples as either low-risk or high-risk in the Affymetrix (Figure 1A) and METABRIC (Figure 1B) validation cohorts. Moreover, the percentage of samples in each stratum was comparable across subgroups defined by clinical traits (Supplemental Table 4 in the online version). Specifically, approximately 65% (63-71%) of samples were classified as EarlyR-Low in ER^+, LN^-, and LN^+ samples in both cohorts (Supplemental Table 4 in the online version), and 71% of ER^+, $HER2^-$ samples in METABRIC were EarlyR-Low. Together, low-risk and high-risk categories defined by EarlyR accounted for over 85% to 88% of patients in all subgroups, with only 12% to 15% of samples being classified as of intermediate risk.

The prognostic significance of EarlyR stratification and EarlyR continuous score was assessed in ER^+ overall and subgroups defined by LN status, HER2 status, and tumor size for the Affymetrix validation cohort and METABRIC cohort. EarlyR score and stratification were significantly prognostic in all subgroups (Table 1, Figure 2, Supplemental Figure 3 in the online version).

In contemporary treatment regimens, almost all ER^+ breast cancer patients are treated with hormone therapy. EarlyR stratification and continuous score were both prognostic in hormone therapy—treated patients in the Affymetrix validation cohort (Table 1, Figure 3A,B, Supplemental Figure 4A,B in the online version). Also, among those ER^+ patients with LN^- disease who were treated with hormone therapy, EarlyR stratification and continuous score were both prognostic in the Affymetrix validation cohort and METABRIC (Table 1, Figure 3C,D, Supplemental Figure 4C,D in the online version). In ER^+, LN^- patients treated with hormone therapy in the Affymetrix validation cohort, the EarlyR-Low patients (78%) had a probability of distant relapse after 8 years 0.91 (95% confidence interval [CI], 0.78-0.94).

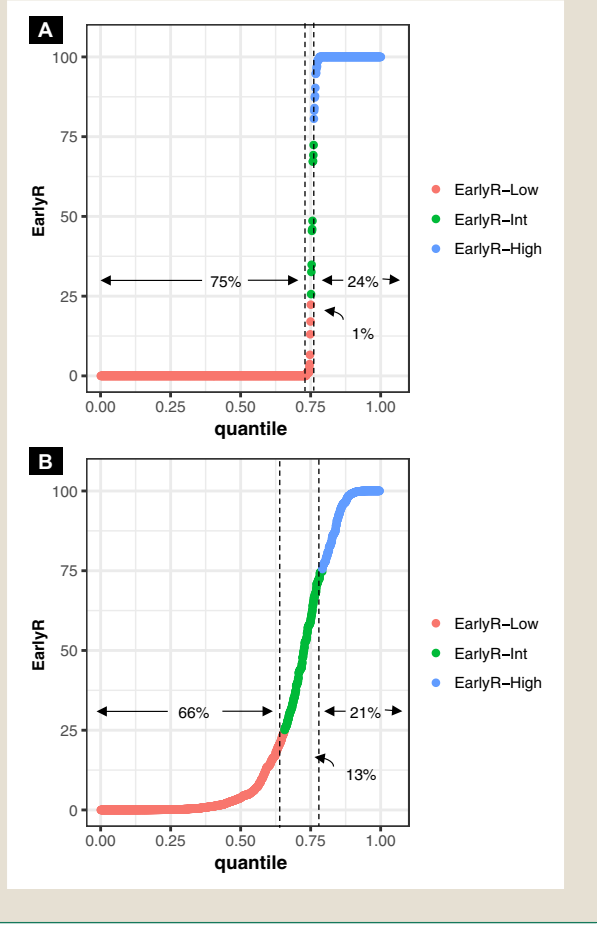
In addition to being prognostic of DRFI, in METABRIC, EarlyR stratification was prognostic of 8-year breast cancer—free interval in all ER^+ patients (hazard ratio [HR] of High vs. Low = 2.3 [95% CI, 1.8-2.9], Figure 4A), ER^+, LN^- patients (HR of High vs. Low = 1.9 [95% CI, 1.4-2.8], Figure 4B); ER^+, LN^+ (HR of High vs. Low = 2.6 [95% CI, 2.0-3.6], Figure 4C); and ER^+ patients, treated with hormone therapy (HR of High vs. Low = 2.3 [95% CI, 1.7-3.0], Figure 4D).

EarlyR Prognostic Significance in Multivariate Analysis Including Clinicopathologic Variables

The independence of EarlyR from clinicopathologic variables was assessed using multivariate Cox models in accordance with the REMARK recommendations.[35] Adding EarlyR to each of LN status, tumor size (binary and continuous), patient age (binary), and tumor grade significantly increased the prognostic significance of the clinical variable (Supplemental Table 5 in the online version) in METABRIC. In a Cox model including LN, binary tumor size, binary age and grade, the addition of EarlyR significantly improved on the prognostic significance of the clinically based model ($P = 1.1 \times 10^{-12}$ for the chi-square statistic of the log likelihoods). This provided strong evidence that EarlyR offers prognostic information that cannot be derived from clinicopathologic variables.

In multivariate Cox models (using 8-year DRFI) including stratified EarlyR and each of tumor size, age, and tumor grade,

Figure 1 **Continuous EarlyR Score Plotted With Respect to Quantiles of Score. (A) ER^+ Samples in Affymetrix Validation Cohort and (B) ER^+ Samples in METABRIC Cohort. Points are Colored according to Risk Strata EarlyR-Low (EarlyR ≤ 25), EarlyR-Int (25 < EarlyR ≤ 75), and EarlyR-High (75 < EarlyR). Dotted Vertical Lines Indicate Boundaries Between Plotted Strata, Along With Percentages of Samples in Each Stratum**



Abbreviation: ER = estrogen receptor.

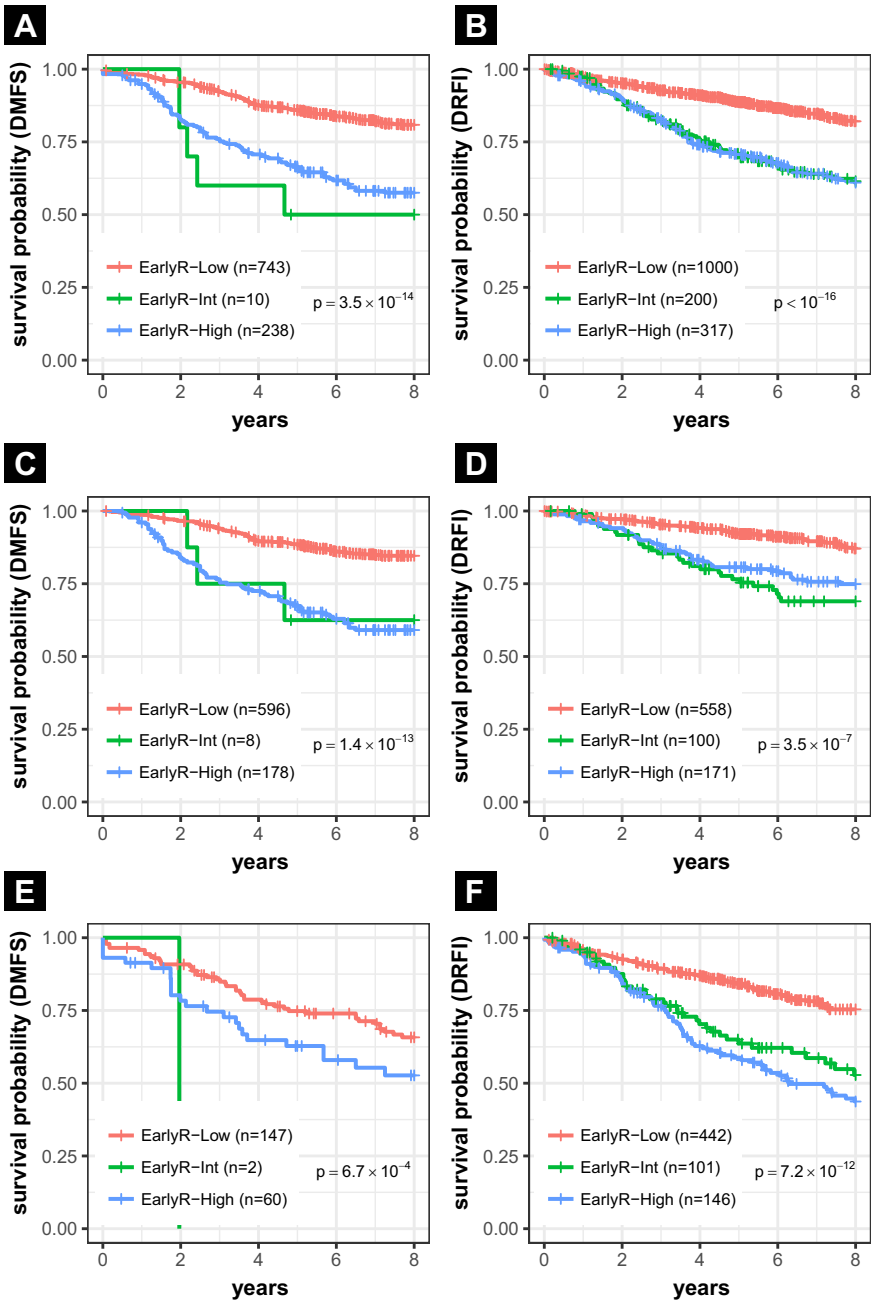
Table 1 Significance and Expected Survival of EarlyR Strata With Respect to 8-Year DRFI in Selected Subgroups

Characteristic	*P* for EarlyR Score	HR for EarlyR Score[a]	*P* for EarlyR Strata	HR of EarlyR-High to EarlyR-Low	Expected Survival With Respect to 8-Year DRFI (95% CI)		
					EarlyR-Low	EarlyR-Int	EarlyR-High
METABRIC							
All ER^+	$< 2.2 \times 10^{-16}$[b]	1.7 (1.5-2.0)	$< 2.2 \times 10^{-16}$[c]	2.6 (2.0-3.3)	0.82 (0.79-0.85)	0.62 (0.54-0.70)	0.61 (0.56-0.67)
ER^+, LN^-	3.6×10^{-6}	1.6 (1.3-1.9)	3.5×10^{-7}	2.3 (1.5-3.4)	0.87 (0.84-0.90)	0.69 (0.60-0.79)	0.75 (0.69-0.82)
ER^+, LN^+	8.3×10^{-14}	1.8 (1.5-2.1)	7.3×10^{-12}	2.9 (2.1-3.9)	0.75 (0.71-0.80)	0.53 (0.42-0.66)	0.44 (0.36-0.54)
ER^+, $HER2^-$	4.0×10^{-15}	1.8 (1.5-2.1)	2.2×10^{-14}	2.6 (2.0-3.5)	0.83 (0.80-0.86)	0.61 (0.53-0.71)	0.63 (0.56-0.71)
ER^+, size $\leq$ 2 cm	5.2×10^{-7}	1.8 (1.4-2.2)	2.9×10^{-6}	2.8 (1.8-4.3)	0.88 (0.84-0.91)	0.74 (0.65-0.85)	0.70 (0.62-0.79)
ER^+, size $>$ 2 cm	8.5×10^{-11}	1.6 (1.4-1.9)	5.1×10^{-10}	2.3 (1.7-3.1)	0.77 (0.83-0.81)	0.53 (0.44-0.65)	0.54 (0.47-0.62)
ER^+, with hormone therapy	1.5×10^{-12}	1.7 (1.4-1.9)	2.2×10^{-11}	2.5 (1.9-3.3)	0.80 (0.77-0.84)	0.62 (0.54-0.72)	0.58 (0.52-0.66)
ER^+, LN^-, with hormone therapy	0.01	1.4 (1.1-1.9)	0.03	2.0 (1.1-3.4)	0.86 (0.82-0.91)	0.79 (0.68-0.92)	0.77 (0.68-0.86)
Affymetrix Validation							
All ER^+	8.8×10^{-15}	1.7 (1.5-1.9)	3.5×10^{-14}	2.7 (2.1-3.5)	0.81 (0.78-0.84)	0.5 (0.27-0.93)	0.58 (0.51-0.65)
ER^+, LN^-	7.3×10^{-15}	1.8 (1.6-2.1)	1.4×10^{-13}	3.2 (2.4-4.5)	0.85 (0.82-0.88)	0.63 (0.37-1)	0.59 (0.52-0.67)
ER^+, LN^+	0.048	1.3 (1-1.6)	6.7×10^{-4}	1.6 (1-2.7)	0.66 (0.58-0.75)	NA	0.53 (0.40-0.69)
ER^+, with hormone therapy	1.9×10^{-6}	1.6 (1.3-1.9)	2.3×10^{-6}	2.5 (1.7-3.7)	0.83 (0.79-0.87)	0.5 (0.23-1)	0.63 (0.54-0.73)
ER^+, LN^-, with hormone therapy	3.6×10^{-6}	1.9 (1.4-2.6)	2.4×10^{-5}	3.7 (2.0-6.6)	0.91 (0.78-0.94)	0.75 (0.43-1.0)	0.70 (0.59-0.82)

Abbreviations: CI = confidence interval; DRFI = distant recurrence-free interval; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; LN = lymph node; NA = not applicable.
[a]Hazard ratio for EarlyR score in increments of 50.
[b]Chi-square statistic ($df = 1$) is 76.
[c]Chi-square statistic ($df = 2$) is 78.

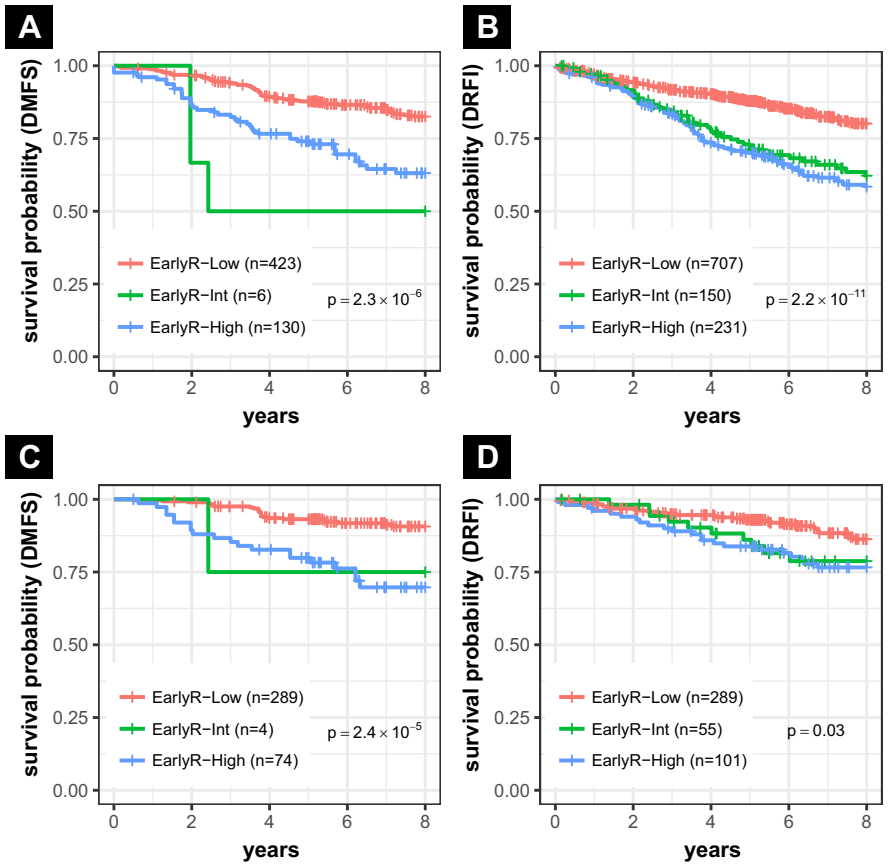
within ER^+, LN^- METABRIC samples, only binary tumor size was statistically significant ($P = .0045$) in addition to EarlyR. We further analyzed the effect of EarlyR prognostic significance separately in small and large tumors. First, there was no interaction effect for EarlyR stratification and tumor size; ie, the HR for EarlyR-High versus EarlyR-Low was not significantly different between small (≤ 2 cm) and large (> 2 cm) tumors. In the cohort of ER^+, LN^-, METABRIC tumors sized ≤ 2 cm, EarlyR stratification were prognostic of 8-year DRFI (HR High vs. Low 2.2 [95% CI, 1.2-4.1]), with 8-year expected survival probabilities 0.89 (95% CI,

Figure 2 Kaplan-Meier Survival Curves With Respect to Distant Recurrence Plotted for EarlyR Risk Strata. (A) ER^+ Affymetrix Validation (n = 991), (B) ER^+ METABRIC (n = 1518), (C) ER^+, LN^- Affymetrix Validation (n = 782), (D) ER^+, LN^- METABRIC (n = 829), (E) ER^+, LN^+ Affymetrix Validation (n = 209), (F) ER^+, LN^+ METABRIC (n = 689). Numbers of Samples in Each Stratum are Reported. Eight-Year Expected Survival With Respect to DRFI for Each Cohort and Stratum Is Reported in Table 1



Abbreviations: DRFI = distant recurrence-free interval; ER = estrogen receptor; LN = lymph node.

Figure 3 Kaplan-Meier Survival Curves With Respect to Distant Recurrence Plotted for EarlyR Risk Strata for Cohorts Treated With HT. (A) ER^+ Affymetrix Validation, HT Treated (n = 559); (B) ER^+ METABRIC, HT Treated (n = 1088); (C) ER^+, LN^- Affymetrix Validation, HT Treated (n = 369); (D) ER^+, LN^- METABRIC, HT Treated (n = 445). Percentages of Subgroups in EarlyR Strata are provided. Eight-Year Expected Survival With Respect to DRFI for Each Cohort and Stratum Is Reported in Table 1



Abbreviations: DRFI = distant recurrence-free interval; ER = estrogen receptor; HT = hormone therapy; LN = lymph node.

0.86-0.93), 0.78 (95% CI, 0.67-0.91), 0.79 (95% CI, 0.70-0.89) for EarlyR-Low, EarlyR-Int, and EarlyR-High, respectively. In the corresponding set of patients with tumors > 2 cm, EarlyR stratification was prognostic of 8-year DRFI (HR High vs. Low 2.1 [95% CI, 1.2-3.6]) with 8-year expected survival probabilities 0.84 (95% CI, 0.79-0.89), 0.63 (95% CI, 0.50-0.79), and 0.71 (95% CI, 0.61-0.82) for EarlyR-Low, EarlyR-Int, and EarlyR-High, respectively.

Prognostic Significance of EarlyR Was Superior or Comparable to Other Genomic Assays

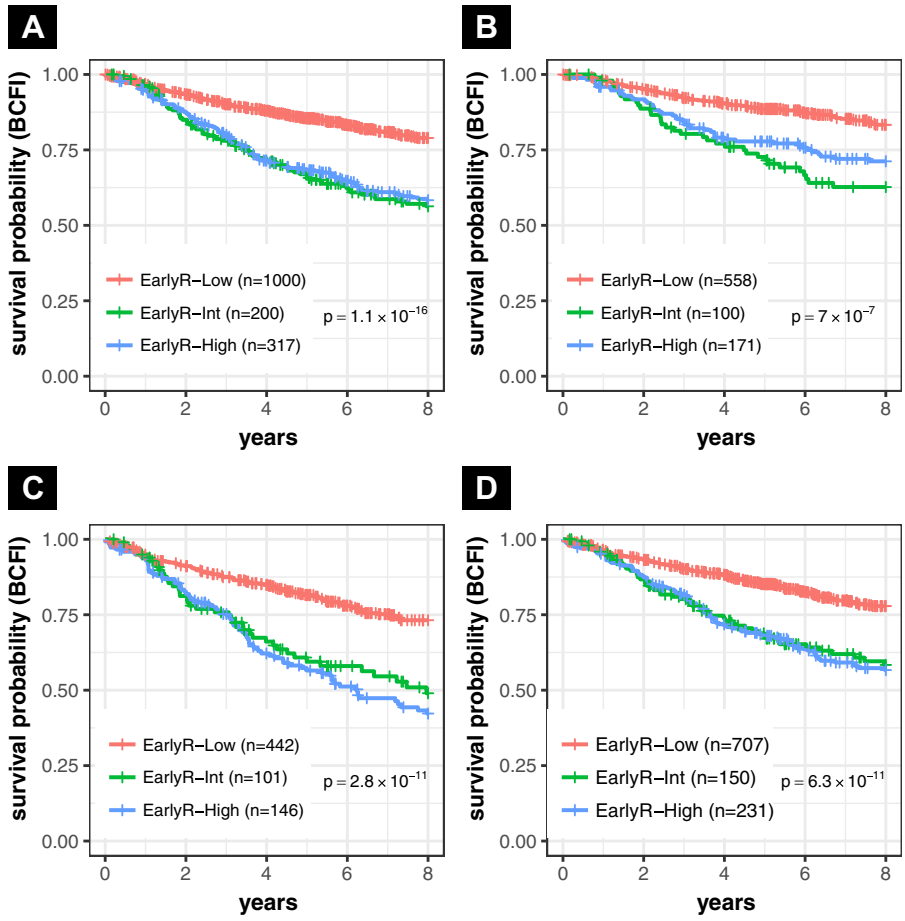
Surrogates for RS,[3] GENE70 (a precursor of Mammaprint),[4,5] and ROR,[6] as well as stratified versions, were computed for the METABRIC cohort using the genefu R package.[32] The concordance index[30,31] was used to compare the prognostic significance of these tests as continuous scores, as has been previously done for prognostic assays.[16] These comparisons (Supplemental Figure 5 in the online version) showed that the concordance index was highest for EarlyR followed by RS, GENE70, and ROR scores, in that order, although this difference was not statistically significant by the 95% confidence intervals. This established that EarlyR was at least as prognostic as surrogates of these other signatures.

Expected survival probabilities (8-year DRFI) for stratified versions of EarlyR and the RS, GENE70, and ROR were computed in the ER^+, LN^-, $HER2^-$ METABRIC cohort (Supplemental Table 6 in the online version). The expected survival probability in EarlyR-Low was nearly the same as for the other signatures, although EarlyR-Low contained 72% of samples; the low-risk groups for RS, GENE70, and ROR were 55%, 55%, and 62%, respectively.

Discussion

The decision to use adjuvant chemotherapy to treat early stage breast cancer must balance the reduced risk of metastasis with chemotherapy's toxic effects. Increasingly, tests that analyze gene expression patterns in primary tumors are being used to guide this decision. Herein, we have developed an assay wherein the EarlyR score (0-100) is defined by combining the risk scores of the 5 panel genes using a nonlinear formula. The computation of the EarlyR score, and the resulting stratification into risk groups, is intended to offer the convenience of discrete classification (EarlyR-High or

Figure 4 Kaplan-Meier Survival Curves With Respect to BCFI Plotted for EarlyR Risk Strata for METABRIC Samples. (A) ER^+ (n = 1518); (B) ER^+, LN^- (n = 829); (C) ER^+, LN^+ (n = 689); (D) ER^+, HT Treated (n = 1088). Eight-Year Breast Cancer–Free Survival Probabilities for EarlyR Risk Strata are as Follows: (A) EarlyR-Low: 0.79 (95% CI, 0.76-0.82), EarlyR-Int: 0.56 (95% CI, 0.49-0.64), EarlyR-High: 0.58 (95% CI, 0.53-0.64); (B) EarlyR-Low: 0.83 (95% CI, 0.80-0.87), EarlyR-Int: 0.63 (95% CI, 0.53-0.74), EarlyR-High: 0.71 (95% CI, 0.65-0.79); (C) EarlyR-Low: 0.73 (95% CI, 0.69-0.78), EarlyR-Int: 0.49 (95% CI, 0.39-0.62), EarlyR-High: 0.42 (95% CI, 0.34-0.52); (D) EarlyR-Low: 0.78 (95% CI, 0.75-0.81), EarlyR-Int: 0.58 (95% CI, 0.50-0.68), EarlyR-High: 0.57 (95% CI, 0.50-0.64)



Abbreviations: BCFI = breast cancer–free interval; CI = confidence interval; ER = estrogen receptor; HT = hormone therapy; LN = lymph node.

EarlyR-Low) while dependably identifying a small subset of patients (EarlyR-Int) whose risk classification is uncertain. In contrast to Oncotype DX, which has an intermediate-risk group of at least 35% in most studies, EarlyR-Int consistently contains at most 15% of samples. Thus, EarlyR offers a definitive prognosis for significantly more patients than Oncotype DX.

EarlyR is further distinguished by identifying a large majority of ER^+, LN^-, $HER2^-$ patients as low risk (72% in METABRIC). In contrast, in several studies, at most 59% of ER^+, LN^- patients are classified as low risk by Oncotype DX.[12,16,36] In spite of the larger low risk stratum for EarlyR, by using surrogates of RS, Mammaprint, and ROR in METABRIC, we showed that EarlyR is at least as significant as a prognostic tool using concordance index and expected survival.

EarlyR-Int consists of samples in which EarlyR score is rising sharply from the low-risk group to the high-risk group; ie, these are samples that straddle the boundary between good and poor prognoses. In multiple Kaplan-Meier analyses, we found that the expected survival probability for EarlyR-Int was comparable to that of EarlyR-High (Figures 2 and 3). Further studies in well-annotated clinical trial cohorts will determine the need for this intermediate-risk category.

In the ER^+, LN^-, $HER2^-$ METABRIC samples, the 8-year distant recurrence-free survival estimate for the EarlyR low risk stratum was 88%. This estimated survival percentage was markedly lower than that computed for RS and ROR in transATAC[16] or for Mammaprint in the MINDACT trial.[15] However, the low-risk strata for surrogates of these other signatures in the same subset of METABRIC are between 87% and 89% (Supplemental Table 6 in the online version). This is likely because only 52% of these patients received hormone therapy, and 54% of them had tumors >

2 cm in diameter. In contrast, in MINDACT, all ER^+ patients were recommended for hormone therapy, and only 28% of tumors (including both LN^- and LN^+) were > 2 cm.

Tumor size has been found to be significantly prognostic independent of ROR,[6] RS,[37] and EndoPredict.[38] The commercial Prosigna score combines ROR with tumor size to form a single score, and EPclin combines EndoPredict and tumor size. We found that tumor size was also significantly prognostic independent of EarlyR in the ER^+, LN^- METABRIC cohort. To elucidate the combined prognostic significance of EarlyR and tumor size, we reported the prognostic significance of EarlyR separately in tumors $\leq$ 2 cm and tumors > 2 cm. We think that conflating size and a genomic score into a single score confuses the independent effects of the 2 risk factors.

Each gene in the EarlyR panel, *ESPL1*, *MKI67, SPAG5, PLK1,* and *PGR*, plays a role in multiple processes related to ER^+ breast cancer progression and treatment response. *ESPL1*, which is critical for the timely separation of sister chromatids during anaphase, has been found to be disproportionately elevated in luminal B tumors and has also been found to be a risk factor independent of PAM50, RS, Mammaprint, and EndoPredict.[39] *MKI67* is a well-studied biomarker for proliferation. Elevated expression of *SPAG5*, which is associated with the mitotic spindle apparatus, is predictive of sensitivity to cytotoxic chemotherapy in breast cancer.[40,41] *PLK1* is known to promote hormone-independent ER transcription and growth,[42] as well as being associated with mutations of *TP53*.[43] The role of the hormone receptor *PGR* in progression of breast cancer is well established.

Prognostic signatures for ER^+ breast cancer, including EarlyR, were developed to assist physicians in selecting patients for hormone therapy alone or combined with systemic chemotherapy.[44] Studies are planned to build evidence that patients identified as high risk by EarlyR are good candidates for chemotherapy, while those in EarlyR-Low are unlikely to benefit from chemotherapy.

There are a number of limitations of the current study. The major limitation is that all of the analyses were performed in a retrospective manner using *in silico* data obtained from several studies with only 2775 samples. These studies had variable methods of preanalytical tissue preparation, analytical techniques (U133A and IlluminaHuman-v3) and statistical analytic methods. Moreover, the samples were from patients not treated under current standards for ER^+ breast cancer in that many did not receive hormone therapy or chemotherapy. However, despite these, the EarlyR score showed remarkable stability in predicting outcomes. Another important issue is the small number of FFPE samples used in the study. This analysis was meant to provide a proof of principle for an assay to execute EarlyR testing with qPCR using FFPE tissues. Additional studies are planned using clinical trial samples to validate the results of the studies presented herein.

Conclusion

The EarlyR assay is a risk score that classified at least 85% of ER^+ patients as high or low risk. The intermediate-risk category contained at most 15% of patients, approximately half that observed in other assays. EarlyR classified significantly more patients (72% of ER^+, LN^-, $HER2^-$) as low risk compared to other signatures (Oncotype DX RS, Mammaprint, and PAM50 ROR), without apparent loss in prognostic significance. We showed that the prognostic significance of EarlyR is not improved by the addition of age or grade in ER^+, LN^- tumors, but tumor size is independently significant. Further independent validation in well-annotated cohorts of patients treated with current standards for hormone therapy is necessary to determine EarlyR's clinical utility.

Clinical Practice Points

- There is a need for better gene signatures in ER^+ breast cancer because current assays identify a percentage of patients as having uncertain risk of recurrence.
- The goal of this study was to establish the utility of a novel 5-gene signature for ER^+ breast cancer and compare it with existing assays.
- The 5-gene signature, EarlyR, performs similarly to existing commercial assays in concordance analyses.
- EarlyR assay is a risk score that classified at least 85% of ER^+ patients as high or low risk. The intermediate-risk category contained at most 15% of patients, approximately half that observed in other assays

Acknowledgments

Supported in part by Indiana Clinical Translational Sciences Institute (Indiana CTSI) to S.S.B. and S.A.B. S.S.B. is also supported by a Susan G. Komen for the Cure Scholar Award.

Disclosure

S.A.B., Y.G.P., and S.S.B. have equity interest in SYSGenomics LLC. University of Notre Dame and Indiana University have submitted a patent application for the gene signature.

Supplemental Data

The Supplemental tables, figures, and data accompanying this article can be found in the online version at https://doi.org/10.1016/j.clbc.2018.07.011.

References

1. Mehta R, Jain RK, Badve S. Personalized medicine: the road ahead. *Clin Breast Cancer* 2011; 11:20-6.
2. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000; 406:747-52.
3. 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.
4. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; 347:1999-2009.
5. 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.
6. 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.
7. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res* 2011; 17:6012-20.
8. Jerevall PL, Ma XJ, Li H, et al. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *Br J Cancer* 2011; 104:1762-9.
9. Zhang Y, Schnabel CA, Schroeder BE, et al. Breast cancer index identifies early-stage estrogen receptor-positive breast cancer patients at risk for early- and late-distant recurrence. *Clin Cancer Res* 2013; 19:4196-205.
10. Hassett MJ, Silver SM, Hughes ME, et al. Adoption of gene expression profile testing and association with use of chemotherapy among women with breast cancer. *J Clin Oncol* 2012; 30:2218-26.
11. Joh JE, Esposito NN, Kiluk JV, et al. The effect of Oncotype DX recurrence score on treatment recommendations for patients with estrogen receptor-positive early stage breast cancer and correlation with estimation of recurrence risk by breast cancer specialists. *Oncologist* 2011; 16:1520-6.

12. Dzimitrowicz H, Mougalian S, Storms S, et al. Impacts of early guideline-directed 21-gene recurrence score testing on adjuvant therapy decision making. *J Oncol Pract* 2017; 13:e1012-20.
13. Buechler S. Low expression of a few genes indicates good prognosis in estrogen receptor positive breast cancer. *BMC Cancer* 2009; 9:243.
14. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012; 486:346-52.
15. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* 2016; 375:717-29.
16. Dowsett M, Sestak I, Lopez-Knowles E, et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol* 2013; 31:2783-90.
17. Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* 2005; 102:13550-5.
18. 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.
19. Zhang Y, Sieuwerts AM, McGreevy M, et al. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat* 2009; 116:303-9.
20. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor–positive breast carcinomas through genomic grade. *J Clin Oncol* 2007; 25:1239-46.
21. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node–negative primary breast cancer. *Lancet* 2005; 365:671-9.
22. Schmidt M, Bohm D, von Torne C, et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* 2008; 68:5405-13.
23. Symmans WF, Hatzis C, Sotiriou C, et al. Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol* 2010; 28:4111-9.
24. Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A model-based background adjustment for oligonucleotide expression arrays. *J Am Stat Soc* 2004; 99:909-17.
25. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; 8:118-27.
26. Hudis CA, Barlow WE, Costantino JP, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol* 2007; 25:2127-32.
27. Sestak I, Dowsett M, Zabaglo L, et al. Factors predicting late recurrence for estrogen receptor–positive breast cancer. *J Natl Cancer Inst* 2013; 105:1504-11.
28. Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Soc* 2002; 97:611-31.
29. Fraley C, Raftery AE, Scrucca L. mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. Technical Report 597, Available at: https://www.stat.washington.edu/sites/default/files/files/reports/2012/tr597.pdf 2012. Accessed: July 19, 2018.
30. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY: Springer; 2013.
31. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004; 23:2109-23.
32. Gendoo DM, Ratanasirigulchai N, Schroder MS, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* 2016; 32:1097-9.
33. Sestak I, Cuzick J, Dowsett M, et al. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. *J Clin Oncol* 2015; 33:916-22.
34. Sparano JA, Gray RJ, Makower DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. *N Engl J Med* 2015; 373:2005-14.
35. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med* 2012; 9:e1001216.
36. Kelly CM, Krishnamurthy S, Bianchini G, et al. Utility of Oncotype DX risk estimates in clinically intermediate risk hormone receptor-positive, HER2-normal, grade II, lymph node–negative breast cancers. *Cancer* 2010; 116:5161-7.
37. Tang G, Cuzick J, Costantino JP, et al. Risk of recurrence and chemotherapy benefit for patients with node-negative, estrogen receptor–positive breast cancer: recurrence score alone and integrated with pathologic and clinical factors. *J Clin Oncol* 2011; 29:4365-72.
38. Buus R, Sestak I, Kronenwett R, et al. Comparison of EndoPredict and EPclin with Oncotype DX recurrence score for prediction of risk of distant recurrence after endocrine therapy. *J Natl Cancer Inst* 2016; 108. https://doi.org/10.1093/jnci/djw149.
39. Finetti P, Guille A, Adelaide J, Birnbaum D, Chaffanet M, Bertucci F. ESPL1 is a candidate oncogene of luminal B breast cancers. *Breast Cancer Res Treat* 2014; 147:51-9.
40. Zhu C, Ma X. The prognostic role of *SPAG5* gene expression in breast cancer patients with systematic therapy [abstract #P261]. *Breast* 2017; 32:109.
41. Abdel-Fatah TMA, Agarwal D, Liu DX, et al. *SPAG5* as a prognostic biomarker and chemotherapy sensitivity predictor in breast cancer: a retrospective, integrated genomic, transcriptomic, and protein analysis. *Lancet Oncol* 2016; 17:1004-18.
42. Bhola NE, Jansen VM, Bafna S, et al. Kinome-wide functional screen identifies role of PLK1 in hormone-independent, ER-positive breast cancer. *Cancer Res* 2015; 75:405-14.
43. King SI, Purdie CA, Bray SE, et al. Immunohistochemical detection of Polo-like kinase-1 (PLK1) in primary breast cancer is associated with TP53 mutation and poor clinical outcom. *Breast Cancer Res* 2012; 14:R40.
44. National Comprehensive Cancer Network (NCCN). NCCN guidelines—breast cancer, version 3.2017, Available at: https://www.nccn.org/professionals/physician_gls/recently_updated.aspx. Accessed: July 29, 2018.

Supplemental Methods:

Multistate Method for Construction of Signature Score From Panel of Genes

Gene Risk Scores

Foundational to our approach to calculating a multigene signature is the concept of a gene risk score, derived from that of a multistate gene,[1] as follows. Given the expression values of a gene in a sample set S, let M be the Gaussian mixture model fit with minimal Bayesian Information Criterion. M partitions the expression values into intervals—most typically 2 intervals consisting of the expression values above and below a threshold. Distinguish as the high-risk component the interval that has the greater proportion of cases that recur. Define as the gene's risk score the probability that a sample is in the high-risk component, as determined by the model M.

When the mixture model M defines more than 2 intervals, it defines more than one possible threshold between high and low expression values. In this case, there are several possible risk scores for this gene. In defining a prognostic signature using these methods, the discovery process will select the risk score that results in the most significant signature. In the discovery process, if the model M for a specific gene has only one component, then that gene will be eliminated from consideration for the panel.

It bears emphasizing that the gene risk score is derived from fitting a model to the gene's expression values. There is no algebraic formula for computing the risk score. The computer program for executing the model fit is proprietary.

Multistate Gene Signatures

Given panel genes $g_1,\ldots,g_n$ for a multistate gene signature, derived through the discovery process given below, and a cohort of patient samples, C, for which expression values of $g_1,\ldots,g_n$ have been assayed, the multistate gene signature score is computed as follows.

1. For each panel gene g_i, let r_i be the gene risk score for g_i in C.
2. The signature score S is 1 minus the product of all numbers of the form $(1 - r_i r_j)$, as (i,j) range over all possible distinct pairs from 1 to n. For convenience, S is scaled to 0-100. (If we interpret r_i as the probability that a sample is in a high-risk state due to gene g_i, then S is the probability that some pair of panel genes are in high-risk states.)
3. Given the continuous score S, discrete risk strata for the signature are defined as low risk ($S \leq 25$), intermediate risk ($25 < S \leq 75$), and high risk ($75 < S$).

Going forward, it is important to bear in mind the following:

- The computation of the score S in a cohort of patient samples is independent of the technology used to measure gene expression and all clinical data.
- The signature risk strata are computed directly from the score values and thus are also independent of clinical data.

Discovery of Multistate Signature

To discover a multistate gene signature, a training cohort of samples with whole-genome expression data is selected. From the expression values for all genes assayed, all possible gene risk scores are computed and are individually evaluated for prognostic significance using the score statistic of a Cox proportional hazards model. Ranking these by individual significance, sets of genes are combined as possible panels and the resulting signatures computed. A set of panel genes is selected whose signature is maximally prognostic, as computed for Cox proportional hazards models. More specifically, if P_i is the signature produced with the i highest ranked genes, then we select as the signature the minimal i such that the Cox proportional hazards model with variables P_i and P_{i+1} is not statistically more significant than that with the variable P_i, compared using log likelihoods.

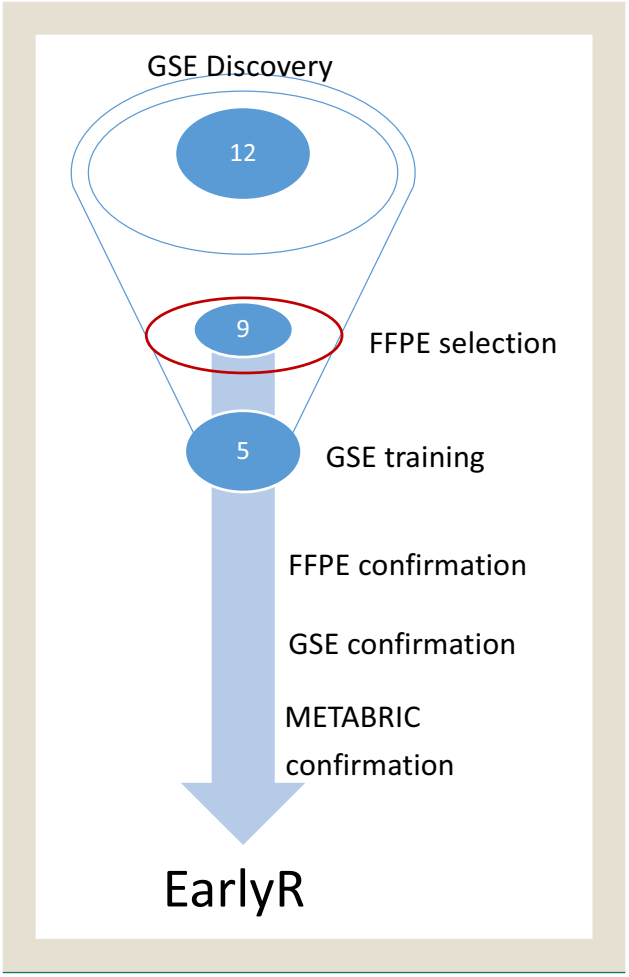
Computation of Multistate Gene Signature Score for Samples Not in Training Cohort

In the training cohort, gene risk scores are computed using the model fitting process described above. To compute the signature score for a new sample, the expression values for the panel genes are assayed and compared to the expression values in a reference set of samples (such as the training cohort). A lookup table is used to estimate the risk score values for each of the panel genes. Subsequently, the signature score values are computed as described above.

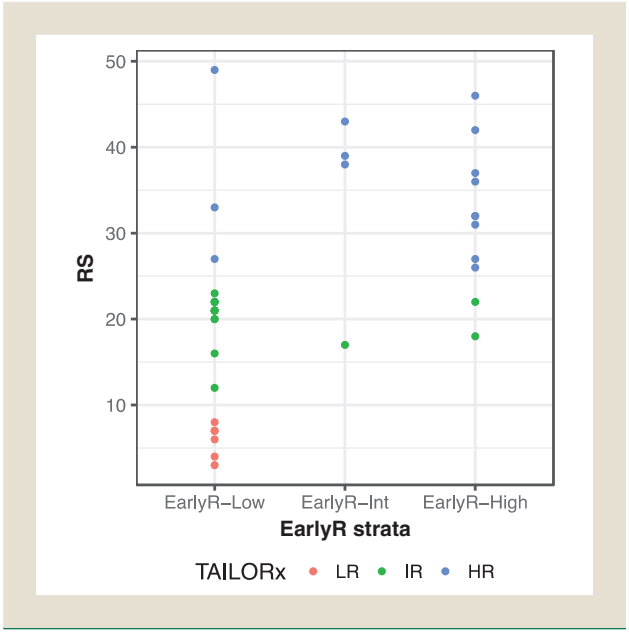
Concordance Index

The concordance index for a continuous score S in a set of samples X with survival data Event and Time, is computed as follows. A pair, i, j, from X is called evaluable if at least one incurred an event, and if only one incurs an event (say i), then the censoring time of j is later than the event time for i. For each evaluable pair, i and j compute a number c(i,j) to be 1 if i relapses before the relapse or follow-up time of j and $S(i) < S(j)$; c(i,j) is also 1 if the preceding clause is true after switching i and j. The concordance index for S in X is then the mean of the numbers c(i,j) over all evaluable pairs. If S is a stratification rather than a continuous score, then the formula is adjusted.[2]

Supplemental Figure 1 Flowchart Detailing Steps Associated With Development of EarlyR Gene Signature

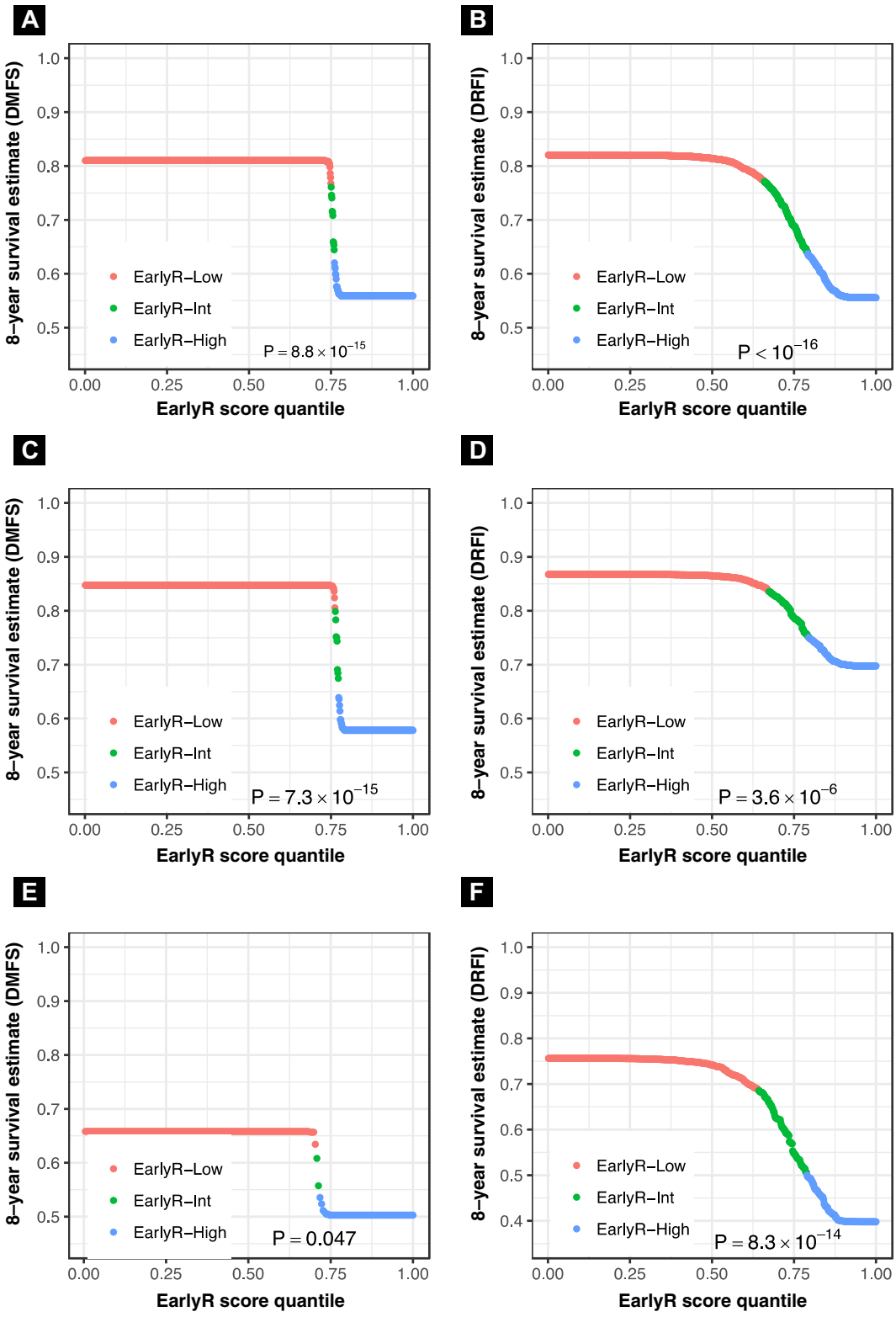


Supplemental Figure 2 For Samples From FFPE Validation Set (N = 36), Oncotype DX RS Is Linearly Dependent on EarlyR Strata (P = .001). Samples are Also Colored by TAILORx Risk Group



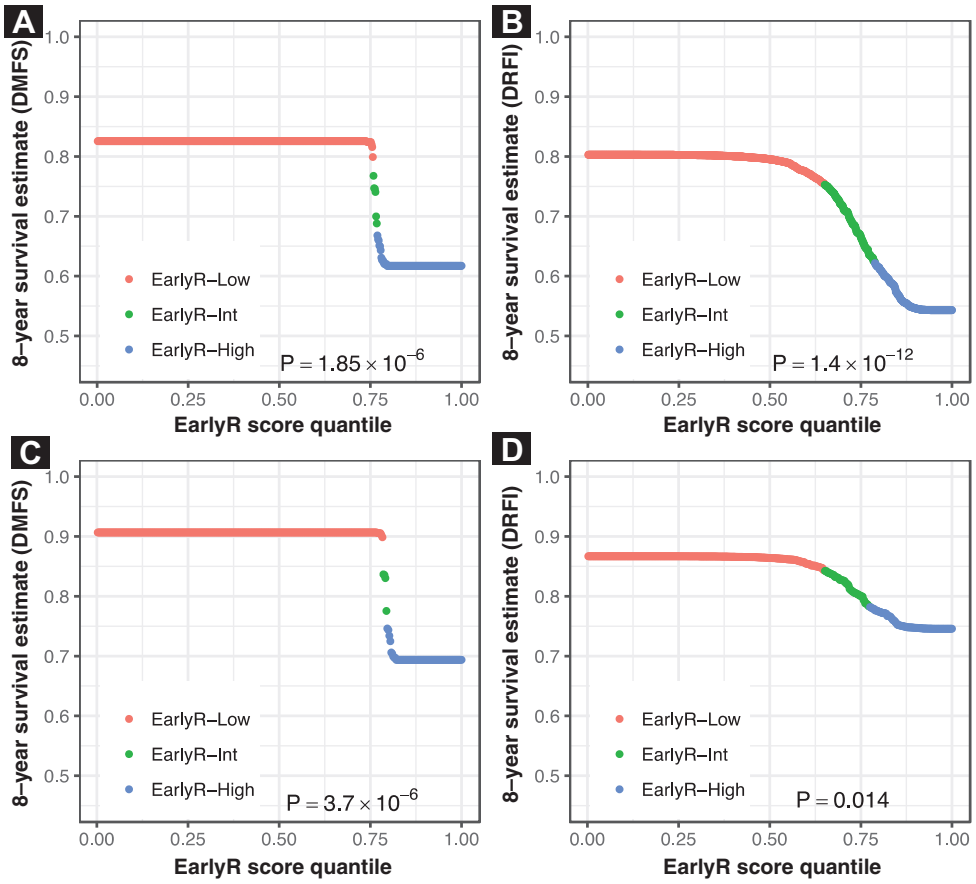
Abbreviations: FFPE = formalin-fixed, paraffin-embedded; RS = recurrence score.

Supplemental Figure 3 Estimated Distant RFS 8 Years After Diagnosis. RFS Is Plotted by Quantiles of Continuous EarlyR Score for (A) ER^+ Affymetrix Validation (n = 991), (B) ER^+ METABRIC (n = 1518), (C) ER^+, LN^- Affymetrix Validation (n = 782), (D) ER^+, LN^- METABRIC (n = 829), (E) ER^+, LN^+ Affymetrix Validation (n = 209), and (F) ER^+, LN^+ METABRIC (n = 689). EarlyR Stratum Membership Is Indicated by Color of Point



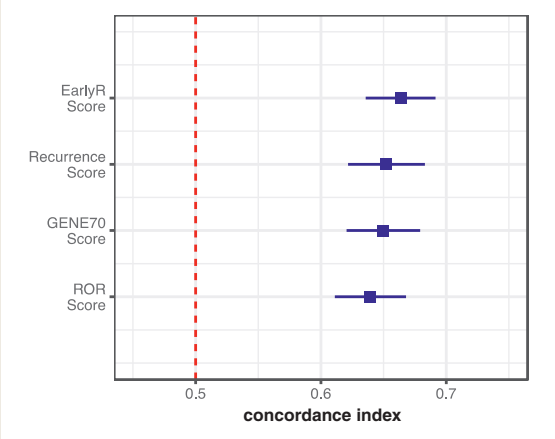
Abbreviations: ER = estrogen receptor; LN = lymph node; RFS = recurrence-free survival.

Supplemental Figure 4 Estimated Distant Relapse-Free Survival 8 Years After Diagnosis Plotted by Quantiles of Continuous EarlyR Score. Cohorts Were Treated With HT: (A) ER^+ Affymetrix Validation, HT Treated (n = 559); (B) ER^+ METABRIC, HT Treated (n = 1088); (C) ER^+, LN^- Affymetrix Validation, HT Treated (n = 369); (D) ER^+, LN^- METABRIC, HT Treated (n = 445)



Abbreviations: ER = estrogen receptor; HT = hormone therapy; LN = lymph node.

Supplemental Figure 5 **Concordance Indices of Genomic Signatures With Respect to 8-Year DRFI Plotted for ER$^+$ METABRIC Cohort. Square Points Indicate Concordance Indices; Lines are 95% CIs. Continuous Scores Were Evaluated for Each of EarlyR, Recurrence Score, GENE70, and Risk of Recurrence. All Tests Were Statistically Significant because Concordance Index CIs are all > 0.5. Highest Concordance Index Was for EarlyR (0.664)**



Abbreviations: CI = confidence interval; DRFI = distant recurrence-free interval; ER = estrogen receptor.

Supplemental Table 1 **Characteristics of Patients in Microarray Data Sets Used in Study**

Characteristic	Affymetrix		METABRIC (ER$^+$)
	Training[a]	Validation[b]	
Number	266	991	1518
Lymph Node			
Negative	266	782	829
Positive	0	209	689
Grade			
1	80	32	166
2	141	99	712
3	43	37	570
NA	2	823	70
Size			
$\leq$ 2 cm	162	90	669
$>$ 2 cm	102	113	835
NA	0	788	14
Age			
$<$ 50 years	110	42	247
$\geq$ 50 years	156	161	1271
NA	0	788	0
8-Year Distant Relapse Event			
No	220	757	1190
Yes	46	226	327
NA	0	8	1
BCFI Event			
No	NA	NA	1142
Yes	NA	NA	375
NA	NA	NA	1
HER2 Status			
Positive	NA	NA	268
Negative	NA	NA	1245
NA	NA	NA	5
Hormone Therapy			
Yes	15	559	1088
No	251	432	430
Chemotherapy			
Yes	0	0	164
No	266	991	1354

Abbreviations: BCFI = breast cancer–free interval; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; NA = not applicable.
[a]GSE3494, GSE7390.
[b]GSE12093, GSE6532 (Oxford cohort), GSE2034, GSE11121, GSE17705.

Supplemental Table 2 Characteristics of Patients in FFPE Data Sets

Characteristic	FFPE Training Set	FFPE Validation Set
Number	36	36
Age		
< 50 years	15	15
≥ 50 years	21	21
Grade		
1	8	3
2	21	24
3	7	9
Size		
≤ 2 cm	24	29
> 2 cm	12	7
TAILORx Risk Group		
Low	4	6
Intermediate	18	15
High	14	15
Oncotype DX Risk Group		
Low	9	9
Intermediate	16	15
High	11	12

Abbreviation: FFPE = formalin fixed, paraffin embedded.

Supplemental Table 3 Probes Used for Development of Quantitative PCR–Based Earlier Assay Development (TaqMan Custom Array Format)

Gene Symbol	Assay ID	Amplicon Length (bp)
MKI67	Hs04260396_g1	64
SPAG5	Hs04260397_s1	60
ESPL1	Hs00901789_g1	62
CDC6	Hs00154374_m1	77
CDC45L	Hs00907337_m1	62
CDT1	Hs00368864_m1	59
PLK1	Hs00983233_g1	61
PHLPP1	Hs01597874_m1	90
CENPA	Hs00903938_g1	62
CXCL9	Hs00171065_m1	60
PGR	Hs01556792_m1	77
PRPF4	Hs00992013_g1	74
*ACTB***	Hs00357333_g1	77
*TFRC***	Hs00951083_m1	66
*GUS***	Hs99999908_m1	81
*RPLPO***	Hs99999902_m1	105
*GAPDH***	Control in array	—

Supplemental Table 4 Distributions of Earlier Risk Strata in Clinically Defined Subsets

Characteristic	N	EarlyR-Low (%)	EarlyR-Int (%)	EarlyR-High (%)
METABRIC				
All ER^+	1518	66	13	21
LN−	829	67	12	21
LN^+	689	64	15	21
Size $\leq$ 2 cm	669	70	12	18
Size $>$ 2 cm	835	63	14	23
HER2−	1245	71	12	16
Affymetrix Validation				
All ER+	991	75	1	24
LN−	782	76	1	23
LN+	209	70	1	29

Abbreviations: ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; LN = lymph node.

Supplemental Table 5 Prognostic Significance (8-Year DRFI) of EarlyR in Excess of Clinical Features in Multivariate Analysis in ER^+ METABRIC Cohort

Feature	*P* of Feature	*P* of EarlyR in Excess of Feature[a]
LN	$< 10^{-16}$,[b]	3.3×10^{-16}
Size (continuous)	$< 10^{-16}$,[c]	8.6×10^{-14}
Size ($\leq$2 cm/$>$ 2 cm)	9.3×10^{-11}	1.9×10^{-14}
Age ($<$50/$\geq$ 50)	0.45	3.3×10^{-16}
Grade	3.2×10^{-5}	2.5×10^{-12}
LN + size + age + grade	$< 10-16$,[d]	1.6×10^{-12}

Abbreviations: DRFI = distant recurrence-free interval; ER = estrogen receptor; LN = lymph node.
[a] *P* value of likelihood ratio of Cox proportional hazard model additively including EarlyR strata in comparison to model with only clinical features.
[b] Chi-square statistic ($df = 1$) is 127.
[c] Chi-square statistic ($df = 1$) is 78.
[d] Chi-square statistic ($df = 5$) is 87.

Supplemental Table 6 Expected Survival Probabilities for Strata of Surrogates of Genomic Assays in LN−, HER2− METABRIC Cohort

Test	Expected 8-Year Survival Probability (DRFI) (95% CI)		
	Low Risk	Intermediate Risk	High Risk
EarlyR	0.88 (0.84—0.91)	0.68 (0.58-0.80)	0.77 (0.69-0.86)
Recurrence Score	0.89 (0.86-0.93)	0.82 (0.76-0.89)	0.72 (0.65-0.80)
GENE70	0.89 (0.85-0.92)	NA	0.77 (0.73-0.83)
Risk of Recurrence	0.87 (0.84-0.91)	0.79 (0.72-0.86)	0.76 (0.67-0.86)

Abbreviations: CI = confidence interval; DRFI = distant recurrence-free interval; HER2 = human epidermal growth factor receptor 2; LN = lymph node; NA = not applicable.

References

1. Buechler SA. Low expression of a few genes indicates good prognosis in estrogen receptor positive breast cancer. *BMC Cancer* 2009; 9:243.
2. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004; 23:2109-23.