EarlyR - A Robust Gene Expression Signature for Predicting Outcomes of ER+ Breast Cancer.

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

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

Address for correspondence:
Sunil Badve, MD, FRCPath
Department of Pathology and Laboratory Medicine,
Indiana University School of Medicine,
635 Barnhill Drive, MS A128, Indianapolis, IN 46202
Email: sbadve@iupui.edu

Keywords: Breast Cancer; gene signature; estrogen receptor positive; prognosis.

Running title: *Early signature in breast cancer*

Conflict of interest:

SAB, YGP and SSB have equity interest in SYSGenomics, LLC. University of Notre Dame and Indiana University have submitted a patent application for the gene signature.

Microabstract:

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 utilities. 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.

ABSTRACT

Introduction: Early stage ER^+ 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 utilities.

Methods: EarlyR prognostic score was developed using integrative analysis of microarray datasets and FFPE-based qRT-PCR assay and validated in Affymetrix datasets 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 (DRFI) 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 $P < 10^{-16}$) datasets of ER^+ breast cancer. Similar results were observed for the breast cancer-free interval endpoint. At most 13% of patients were intermediate risk and at least 66% patients were low-risk in both ER^+ cohorts. EarlyR score was significantly prognostic (DRFI; $P<0.001$) in both LN^-, LN^+ patients and independent from clinical factors. EarlyR and surrogates of current molecular signatures were comparable in prognostic significance by concordance index.

Conclusions: The five-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.

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 which was treated by surgical excision followed by chemotherapy, with the addition of tamoxifen for estrogen receptor positive (ER^+) 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, Onco*type* DX®[3], Mammaprint®[4,5], Prosigna® (Risk of Recurrence)[6], EndoPredict® [7], Breast Cancer Index®[8,9], were developed to help inform physicians' treatment decisions for adjuvant therapy. National Comprehensive Cancer Network (NCCN) treatment guidelines for ER^+, $HER2^-$, > 0.5 cm tumors, with no lymph node involvement, recommend Onco*type* DX Recurrence Score (RS) testing, followed by hormone therapy alone for low-risk patients (RS ≤ 18), and hormone therapy and adjuvant chemotherapy for high-risk patients (RS ≥ 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 Onco*type* DX[10,11].

The impact of such a genomic assay on treatment decisions depends, in part, on the proportions of patients classified as low risk, high risk or intermediate risk. The tests listed above identify approximately 50% of lymph node negative ER^+ patients as low risk. The Onco*type* DX assay has a large intermediate risk group (38% and 40% in two 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^+, lymph node negative, $HER2^-$ tumors in METABRIC cohort[14], EarlyR identified 72%, 12%, 16% as low, intermediate and high risk, respectively. This is significantly more low risk, and significantly fewer intermediate risk patients than reported by currently available assays (e.g. Mammaprint and Onco*type* DX[15, 16]). To build evidence of the clinical utility of the test, we converted it to a "proof of principle" quantitative real time polymerase chain reaction (RT-PCR)-based assay for formalin-fixed, paraffin-embedded (FFPE) tissue.

METHODS

Microarray datasets used in this study

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 datasets: GSE12093[19], GSE6532[20], GSE2034[21], GSE11121[22], GSE17705[23]. The CEL files from all series were normalized together and expression values computed with GCRMA (GC Robust Multi-array 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 Suppl. Table S1. None of the patients in the Affymetrix cohorts had received chemotherapy and 46% 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 lymph nodes or large tumors (more than 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, and breast cancer free interval (BCFI), defined as the time from surgery to recurrent distant metastatic breast cancer or loco-regional invasive ipsilateral breast cancer. Data for both BCFI and DRFI were obtained for patients in the METABRIC cohort (unpublished). In Affymetrix cohorts the endpoints were described as distant metastasis-free survival (DMFS). It is unlikely that differences between study-specific endpoints and DRFI

would result in significant changes in the number of events in these datasets. Prognostic significance with respect to DRFI and BCFI was assessed up to 8 years following diagnosis. The threshold of eight years was chosen based on 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].

Translation of the microarray gene set to qRT-PCR platform using formalin-fixed, paraffin-embedded tissues

Sample Selection and Preparation

The Institutional Review Board of Indiana University approved the study. Informed consent waiver was obtained and only de-identified data was used in the analyses.

Archival formalin-fixed, paraffin-embedded (FFPE) tumor blocks were chosen from 72 patients with breast cancer at the Indiana University Simon Cancer Center based on their Onco*type* DX RS. Initial real-time quantitative RT-PCR (qRT-PCR) analysis was conducted using 10 samples of ER^+ breast cancers. This was followed by qRT-PCR 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 (Suppl. Table S2). 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 ($P = 0.88$).

RNA was extracted from 10μm-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 the TaqMan qRT-PCR Assays

Specific target sequences for each probe from Human Genome U133A 2.0 Array were obtained using NetAffx Analysis Center (http://www.affymetrix.com/analysis/index.affx). Target sequences were aligned to the appropriate mRNA reference sequence (REFSEQ) accession number using 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 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 (Applied Biosystems), were tested for the amplification efficiency based on the ABI defined criteria. Control RNA (Universal Human Reference RNA; Stratagene) and FFPE samples were used to test the efficiency of the probes. Based on the observed efficiency, probes were selected for custom array microfluidic cards (TaqMan assays; Suppl. Table S3).

qRT-PCR Analysis using Custom Arrays

TaqMan reactions were performed in triplicates using custom array microfluidic cards preloaded with TaqMan Gene Expression Assays containing 17 genes (12 discriminant genes and five reference genes) on an ABI Prism 7900HT Fast Real-Time platform according to the manufacturer's instructions (Suppl. Table S3). *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 Applied Biosystems DataAssist™ Software v3.0.

Construction of a genomic signature from gene expression measurements

The methodology for construction of a genomic signature is described in detail in Supplementary Methods.

Statistical analyses

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

The prognostic significance of genomic signatures was compared using concordance index[30, 31]. The concordance index estimates the probability that, for a random pair of patients, the patient with earlier recurrence has higher score than the patients with either later or no

recurrence. The concordance index is a number between 0 and 1, defined more formally in Supplementary Methods. A concordance index greater than 0.5 indicates that the prognostic score is more significant that 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 Onco*type* DX, Mammaprint and Risk of Recurrence (ROR) Score, we compute surrogates of these signatures in METABRIC using the Bioconductor **genefu** package[32].

To compute Onco*type* DX, we select probes in the IlluminaHumanv3 platform representing the 16 target genes in the panel[3]; for genes represented by multiple probes we select 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 Recurrence Score 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], and referred to as GENE70 in genefu. GENE70 was originally computed using 70 array probes from the Agilent Hu25K array platform, of which 56 were associated with 52 unique EntrezIDs. For these 52 EntrezIDs, 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 produces 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 transATAC[33], specifically, 55%, 25%, 20%, respectively.

RESULTS

Discovery of EarlyR: a continuous score and prognostic stratification using expression values of ESPL1, MKI67, SPAG5, PLK1 and PGR

To identify the gene signature, an integrative approach consisting of analysis of *in silico* data and FFPE samples was used (Suppl. Figure S1). 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 qRT-PCR analysis of these 12 genes in a training set of 36 ER^+ breast cancer FFPE samples with known Onco*type* DX RS (Suppl Table S2). For each of the 12 target genes on the qRT-PCR array, risk scores were derived using the Δ-CT values from the training set of 36 samples (Supplementary Methods). The nine genes whose risk scores were significantly predictive of TAILORx risk group (p-value of the linear model < 0.05) were considered for further gene signature development.

The nine genes *(ESPL1, CDC45L, PLK1, CENPA, MKI67, SPAG5, CDT1, PGR, CXCL9)* identified by the above method, were further analyzed for inclusion in a multi-gene signature in the Affymetrix training dataset of 266 ER^+/lymph node negative (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 (Supplementary Methods). This process was continued until the multigene score with maximally

significant results was obtained (Supplementary Methods). This signature, EarlyR score, uses the genes *ESPL1, MKI67, SPAG5, PLK1,* and *PGR*. Specifically, 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*), 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 Onco*type* DX recurrence score RS is linearly dependent on the EarlyR stratification ($P = 0.001$, Suppl Figure S2) in the FFPE validation set (Table S2). Samples were further separated into risk groups with respect to RS using the Onco*type* DX thresholds[3] and the TAILORx thresholds[34]. We found a significant concordance between EarlyR risk strata and Onco*type* DX risk strata ($P = 0.004$) and TAILORx risk strata ($P = 0.002$). 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 blind to all clinical features of the samples (Supplementary Methods).

EarlyR score classified a large majority of samples as either low risk or high risk in 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 (Suppl. Table S4). Specifically, approximately 65% (63%-71%) of samples were classified as EarlyR-Low in ER^+, LN^- and LN^+ samples in both cohorts (Suppl. Table S4) 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- 88% of patients in all subgroups with only 12-15% of samples being classified as of intermediate risk.

The prognostic significance of EarlyR stratification and EarlyR continuous score were assessed in ER^+ overall and subgroups defined by lymph node status, HER2 status and tumor size for Affymetrix validation cohort and METABRIC cohort. EarlyR score and stratification were significantly prognostic in all subgroups (Table 1, Figure 2, Figure S3).

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 Affymetrix validation cohort (Table 1, Figure 3a,b, Figure S4a,b) . Also, among those ER^+ patients with LN^- disease who were treated with hormone therapy, EarlyR stratification and continuous score were both prognostic in Affymetrix validation cohort and METABRIC (Table 1, Figure 3c,d, Figure S4c,d). In ER^+, LN^- patients treated with hormone therapy in Affymetrix validation cohort, the EarlyR-Low patients (78%) had probability of distant relapse after 8 years 0.91 (95%CI 0.78 – 0.94).

In addition to being prognostic of DRFI, in METABRIC, EarlyR stratification was prognostic of 8-year BCFI in all ER^+ patients (HR of High v Low = 2.3 (95%CI 1.8-2.9), Figure 4a), ER^+, LN^- patients (HR of High v Low = 1.9 (95%CI 1.4-2.8), Figure 4b); ER^+, LN^+ (HR of

High v Low = 2.6 (95%CI 2.0-3.6), Figure 4c); and ER^+ patients, treated with hormone therapy (HR of High v Low = 2.3 (95%CI 1.7-3.0), Figure 4d).

EarlyR prognostic significance in multivariate analysis including clinico-pathological variables

The independence of EarlyR from clinico-pathological 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 (Suppl. Table S5) 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-squared statistic of the log-likelihoods). This provided strong evidence that EarlyR offers prognostic information that cannot be derived from clinico-pathological variables.

In multivariate Cox models (using 8-year DRFI) including stratified EarlyR and each of tumor size, age, and tumor grade, within ER^+, LN^- METABRIC samples, only binary tumor size is statistically significant ($P = 0.0045$) in addition to EarlyR. We further analyzed the effect of EarlyR prognostic significance separately in small and large tumors. First, there is no interaction effect for EarlyR stratification and tumor size; i.e., the hazard ratio for EarlyR–High versus EarlyR–Low is not significantly different between small (≤ 2 cm) and large (> 2 cm) tumors. In the cohort of ER^+, LN^-, METABRIC tumors of size ≤ 2cm, EarlyR stratification is prognostic of 8-year DRFI (HR High v Low 2.2 (95%CI 1.2-4.1)) with 8-year expected survival probabilities 0.89 (95%CI 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, EarlyR-High, respectively. In the corresponding set of patients with tumors > 2cm,

EarlyR stratification is prognostic of 8-year DRFI (HR High v 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), 0.71 (95%CI 0.61 – 0.82) for EarlyR-Low, EarlyR-Int, EarlyR-High, respectively.

Prognostic significance of EarlyR was superior or comparable to other genomic assays

Surrogates for Recurrence Score[3], GENE70 (a precursor of Mammaprint)[4, 5] and Risk of Recurrence Score[6], and stratified versions, were computed for the METABRIC cohort using the **genefu** R package ([32]and Methods.) 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 (Supplementary Figure S5) showed that the concordance index was highest for EarlyR followed by Recurrence Score, GENE70 and ROR scores, in that order, although this difference was not statistically significant by the 95% confidence intervals. This establishes that EarlyR is at least as prognostic as surrogates of these other signatures.

Expected survival probabilities (8-year DRFI) for stratified versions of EarlyR and the Recurrence Score, Gene70 and ROR were computed in the ER+, LN-, HER2- METABRIC cohort (Table S6). The expected survival probability in EarlyR-Low is nearly the same as for the other signatures, although EarlyR-Low contains 72% of the samples, and the low risk groups for Recurrence Score, GENE70 and ROR are 55%, 55%, 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 non-linear 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 EarlyR-Low) while dependably identifying a small subset of patients (EarlyR-Int) whose risk classification is uncertain. In contrast to Onco*type* 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 Onco*type* 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, using surrogates of Recurrence Score, Mammaprint and Risk of Recurrence 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; i.e., these are samples that straddle the boundary between good and poor prognosis. In multiple Kaplan-Meier analyses, we found that the expected survival probability for EarlyR-Int was comparable to that of EarlyR-High (Fig 2, 3). Further studies in well annotated clinical trial cohorts with 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 is markedly lower than that computed for Recurrence Score and Risk of Recurrence 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% (Suppl Table S6). This is likely due to the fact that 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 Risk of Recurrence[6], Recurrence Score[37], and EndoPredict[38]. The commercial Prosigna score combines Risk of Recurrence 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 ≤ 2 cm and tumors > 2 cm. We feel that conflating size and a genomic score into a single score confuses the independent effects of the two risk factors.

Each gene in the EarlyR panel, *ESPL1*, *MKI67, SPAG5, PLK1,* and *PGR*, has 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 a risk factor independent of PAM50, Recurrence Score, 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 (see NCCN Guidelines, Breast Cancer[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 2,775 samples. These studies had variable methods of pre-analytical 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, in spite of 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 qRT-PCR using FFPE tissues. Additional studies are planned using clinical trial samples to validate the results of the studies presented herein.

Conclusion

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 (Onco*type* 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.

Author Contributions

All authors (SAB, YG-P and SSB) designed the study, interpreted the data, and wrote this paper. SAB also developed the algorithm and performed the bioinformatical and statistical analyses.

Acknowledgements

Funding is partly supported by Indiana Clinical Translational Sciences Institute (Indiana CTSI) to SSB and SAB. SSB is also supported by Susan G Komen for the Cure Scholar Award.

Clinical Practice Points

- There is a need for better gene signatures in ER+ breast cancer as current assays identify a percent 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

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, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747-52.
3. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine*. 2004;351:2817-26.
4. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *The New England journal of medicine*. 2002;347:1999-2009.
5. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530-6.
6. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27:1160-7.
7. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, 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, Salunga R, Kesty NC, Erlander MG, 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, Jerevall PL, Jankowitz RC, Fornander T, 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, Blayney DW, Edge SB, Herman JG, 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, Laronga C, Lee MC, Loftus L, 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, Hurd S, Chagpar AB, Killelea BK, et al. Impacts of Early Guideline-Directed 21-Gene Recurrence Score Testing on Adjuvant Therapy Decision Making. *J Oncol Pract*. 2017;13:e1012-e20.
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, Turashvili G, Rueda OM, Dunning MJ, 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, Slaets L, Viale G, Delaloge S, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *The New England journal of medicine*. 2016;375:717-29.

16. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, 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, Vega VB, Vergara L, Ploner A, 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, Wang Y, Lallemand F, Haibe-Kains B, 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, Casey G, Cufer T, Paradiso A, 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, Lallemand F, Tutt AM, Gillet 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, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005;365:671-9.
22. Schmidt M, Bohm D, von Torne C, Steiner E, Puhl A, Pilch H, 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, Andre F, Peintinger F, Regitnig P, 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. *Journal of the American Statistical Association*. 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, Gray RJ, Pritchard KI, Chapman JA, 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, Lopez-Knowles E, Ferree S, Cowens JW, 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. *Journal of the American Statistical Association*. 2002;97:611-31. .
29. Fraley C, Raftery AE. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. . *Technical Report* 2012;No. 597, Department of Statistics, University of Washington.
30. Harrell FE. Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2013. xxii, 568 p. p.
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.

505 32. Gendoo DM, Ratanasirigulchai N, Schroder MS, Pare L, Parker JS, Prat A, et al. Genefu:
506 an R/Bioconductor package for computation of gene expression-based signatures in breast
507 cancer. *Bioinformatics*. 2016;32:1097-9.
508 33. Sestak I, Cuzick J, Dowsett M, Lopez-Knowles E, Filipits M, Dubsky P, et al. Prediction of
509 late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients
510 from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or
511 in combination randomized trials using the PAM50 risk of recurrence score. *J Clin Oncol*.
512 2015;33:916-22.
513 34. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Prospective
514 Validation of a 21-Gene Expression Assay in Breast Cancer. *The New England journal of*
515 *medicine*. 2015;373:2005-14.
516 35. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for
517 Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med*.
518 2012;9:e1001216.
519 36. Kelly CM, Krishnamurthy S, Bianchini G, Litton JK, Gonzalez-Angulo AM, Hortobagyi GN,
520 et al. Utility of oncotype DX risk estimates in clinically intermediate risk hormone receptor-
521 positive, HER2-normal, grade II, lymph node-negative breast cancers. *Cancer*. 2010;116:5161-7.
522 37. Tang G, Cuzick J, Costantino JP, Dowsett M, Forbes JF, Crager M, et al. Risk of recurrence
523 and chemotherapy benefit for patients with node-negative, estrogen receptor-positive breast
524 cancer: recurrence score alone and integrated with pathologic and clinical factors. *J Clin Oncol*.
525 2011;29:4365-72.
526 38. Buus R, Sestak I, Kronenwett R, Denkert C, Dubsky P, Krappmann K, et al. Comparison of
527 EndoPredict and EPclin With Oncotype DX Recurrence Score for Prediction of Risk of Distant
528 Recurrence After Endocrine Therapy. *J Natl Cancer Inst*. 2016;108.
529 39. Finetti P, Guille A, Adelaide J, Birnbaum D, Chaffanet M, Bertucci F. ESPL1 is a candidate
530 oncogene of luminal B breast cancers. *Breast Cancer Res Treat*. 2014;147:51-9.
531 40. Zhu C, Ma X. The prognostic role of SPAG5 gene expression in breast cancer patients
532 with systematic therapy. *The Breast*. 2017;32:109 Abst#P261.
533 41. Abdel-Fatah TMA, Agarwal D, Liu DX, Russell R, Rueda OM, Liu K, et al. SPAG5 as a
534 prognostic biomarker and chemotherapy sensitivity predictor in breast cancer: a retrospective,
535 integrated genomic, transcriptomic, and protein analysis. *Lancet Oncol*. 2016;17:1004-18.
536 42. Bhola NE, Jansen VM, Bafna S, Giltnane JM, Balko JM, Estrada MV, et al. Kinome-wide
537 functional screen identifies role of PLK1 in hormone-independent, ER-positive breast cancer.
538 *Cancer Res*. 2015;75:405-14.
539 43. King SI, Purdie CA, Bray SE, Quinlan PR, Jordan LB, Thompson AM, et al.
540 Immunohistochemical detection of Polo-like kinase-1 (PLK1) in primary breast cancer is
541 associated with TP53 mutation and poor clinical outcom. *Breast Cancer Res*. 2012;14:R40.
542 44. The National Comprehensive Cancer Network (NCCN) Guidelines. Breast Cancer, version
543 3.2017, November 10, 2017 [cited 2017 November 10]. Available from:
544 https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf
545

Table 1. Significance and expected survival of EarlyR strata with respect to 8-year DRFI in selected subgroups

	P value for EarlyR score	HR for EarlyR score*	P value 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+	$< 2.2 \times 10^{-16}$§	1.7 (1.5-2.0)	$< 2.2 \times 10^{-16}$ †	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 $\leq$ 2cm	5.2×10^{-7}	1.8 (1.4-	2.9×10^{-6}	2.8 (1.8-4.3)	0.88 (0.84-	0.74 (0.65-	0.70 (0.62-

		2.2)			0.91)	0.85)	0.79)
ER+, size > 2cm	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	1.9×10^{-6}	1.6 (1.3-	2.3×10^{-6}	2.5 (1.7-3.7)	0.83 (0.79-	0.5 (0.23-1)	0.63 (0.54-

hormone therapy		1.9)			0.87)		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)

* Hazard ratio for EarlyR score in increments of 50

§ chi-squared statistic with df = 1 is 76

† chi-squared statistic with df = 2 is 78

Figure legends

Figure 1. The continuous EarlyR score is plotted with respect to quantiles of the score in (a) the ER^+ samples in the Affymetrix validation cohort, and (b) ER^+ samples in the METABRIC cohort. Points are colored according to the risk strata EarlyR-Low (EarlyR $\leq$ 25), EarlyR-Int (25 < EarlyR $\leq$ 75), and EarlyR-High (75 < EarlyR). Dotted vertical lines indicating the boundaries between the strata are plotted, along with the percentages of samples in each stratum.

Figure 2. Kaplan-Meier survival curves with respect to distant recurrence are plotted for the EarlyR risk strata for cohorts (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). The numbers of samples in each stratum are reported in the legends. The 8-year expected survival with respect to DRFI for each cohort and stratum is reported in Table 1.

Figure 3. Kaplan-Meier survival curves with respect to distant recurrence are plotted for the EarlyR risk strata for the following cohorts treated with hormone therapy (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). The percentages of the subgroups in EarlyR strata are reported in figure legends. The 8-year distant relapse-free survival probabilities for the EarlyR risk strata are reported in Table 1.

Figure 4. Kaplan-Meier survival curves with respect to BCFI are plotted for the EarlyR risk strata for the following subgroups of METABRIC samples: (a) ER^+ (n = 1518); (b) ER^+, LN^- (n = 829); (c) ER^+, LN^+ (n = 689); (d) ER^+, HT treated (n = 1088). The 8-year breast cancer-free survival probabilities for the 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).

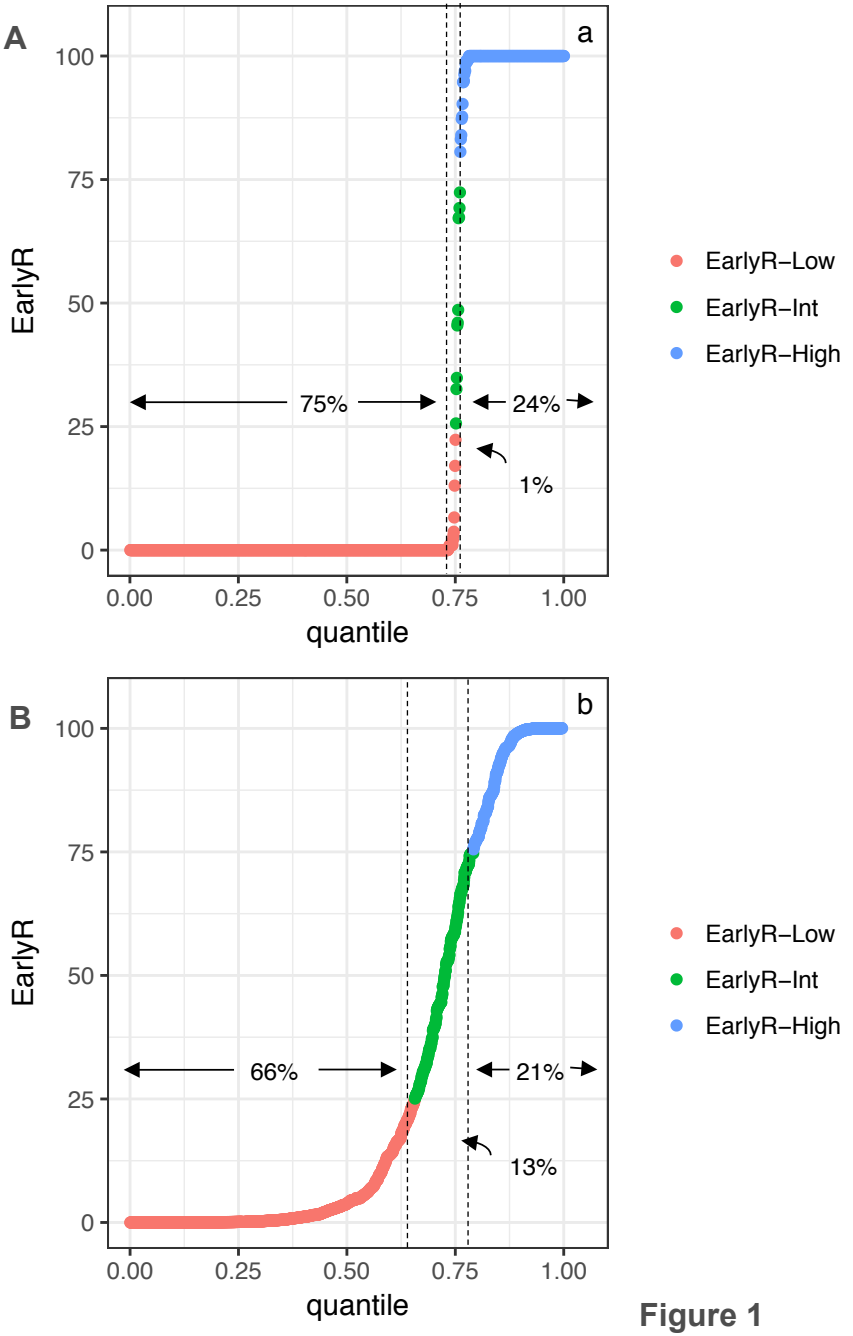
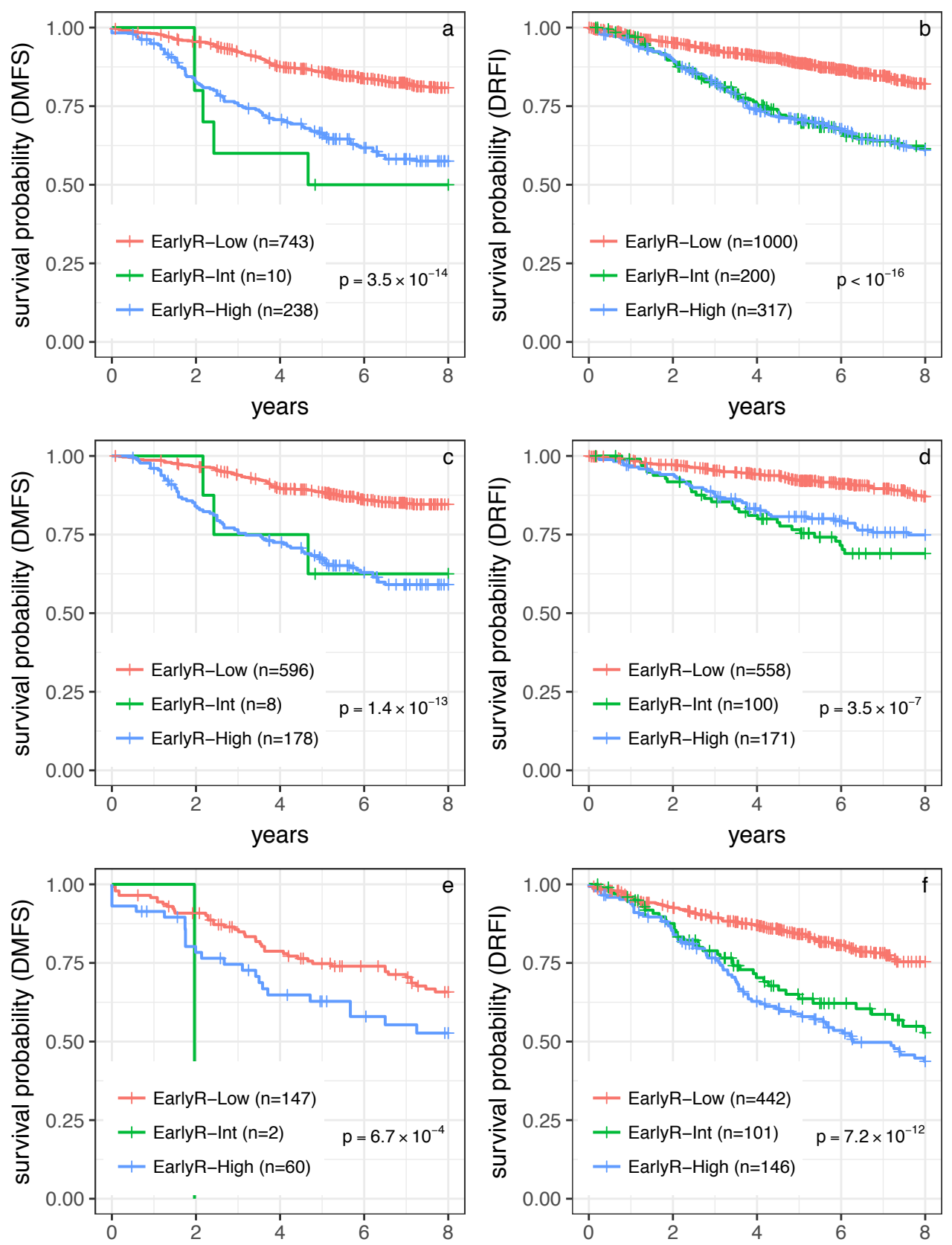
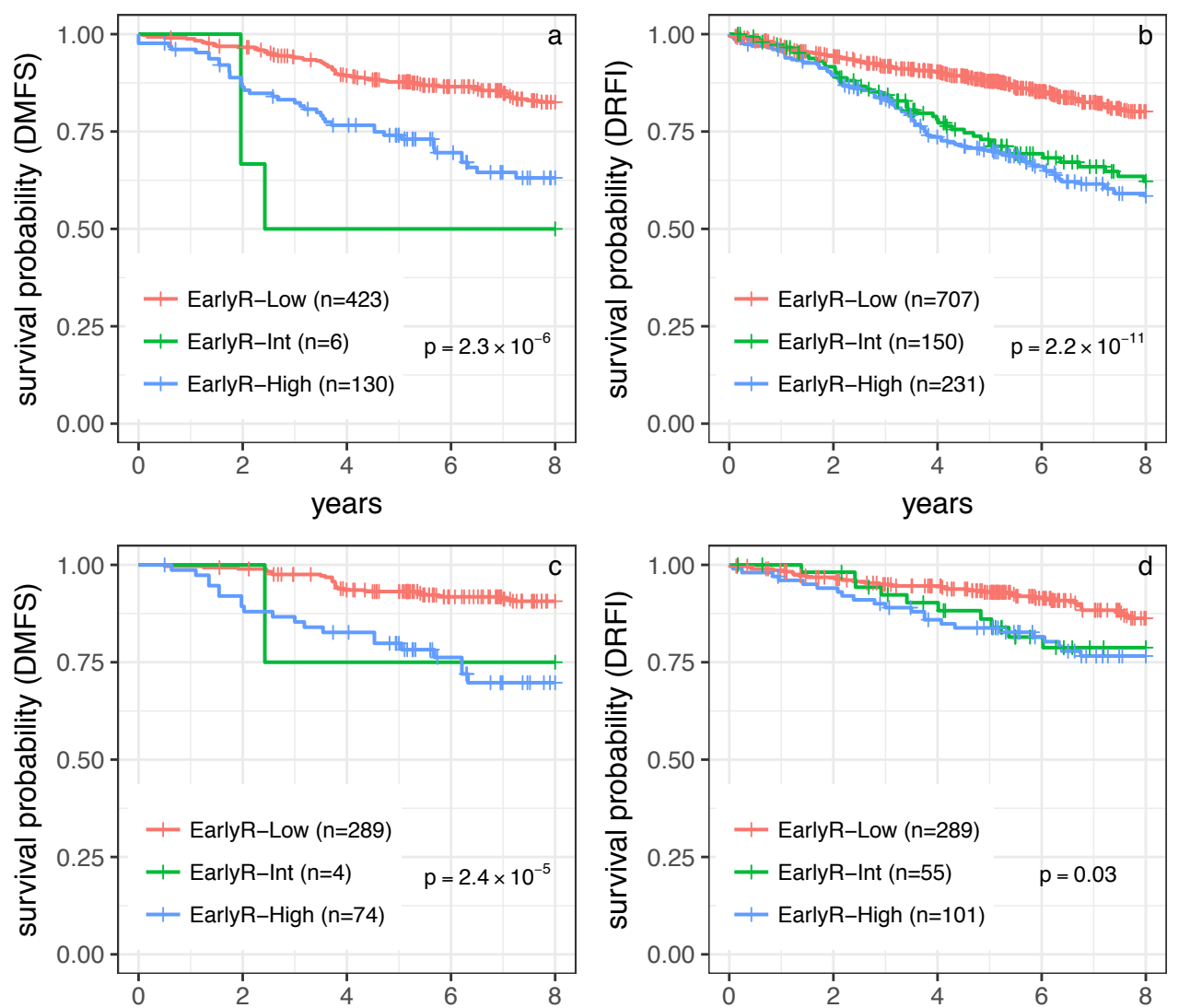

A
a
100
75
50
25
0
EarlyR
75%
24%
1%
0.00
0.25
0.50
0.75
1.00
quantile
EarlyR–Low
EarlyR–Int
EarlyR–High
B
b
66%
21%
13%

Figure 1





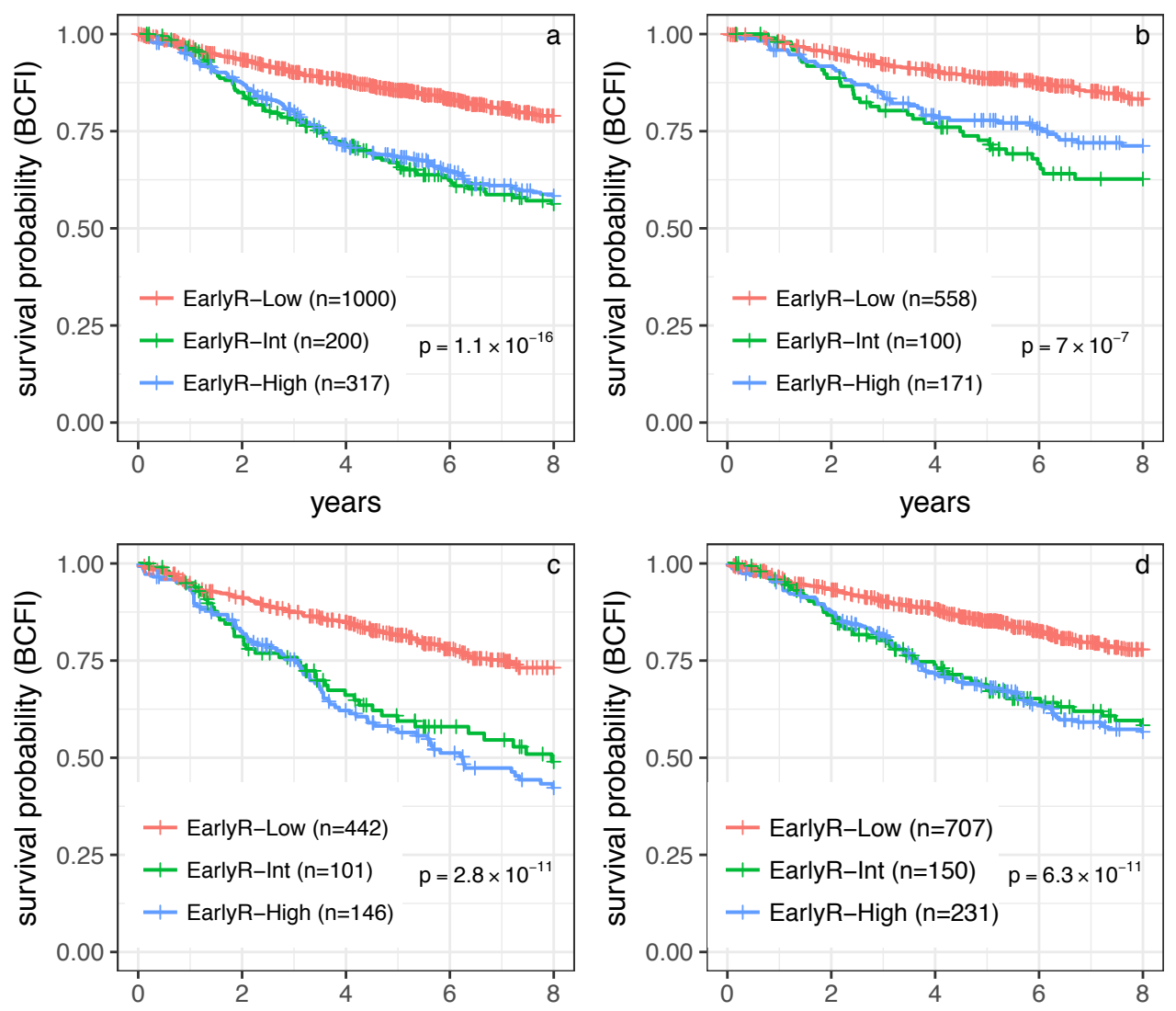


Figure 4

Supplementary Methods

Multistate method for construction of a signature score from a panel of genes.

Gene risk scores. Foundational to our approach to calculating a multigene signature is 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 BIC. M partitions the expression values into intervals, most typically, two intervals consisting of the expression values above and below a threshold. Distinguish as the *high-risk component* the interval which 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 two 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, 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 ($25 < S \leq 75$), High Risk ($75 < S$).

Going forward, it is important to bear in mind that

- 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, thus are also independent of clinical data.

Discovery of a 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 (see item 2 above). 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, 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 the 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 pane 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 1 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 prior to 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, the formula is adjusted [2].

References

1. Buechler SA. Low expression of a few genes indicates good prognosis in estrogen receptor positive breast cancer. *BMC Cancer*. 2009;9:243. doi:10.1186/1471-2407-9-243.

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(13):2109-2123. doi:10.1002/sim.1802.

Supplementary Tables

Table S1. Characteristics of the patients in the microarray datasets used in this study

	Affymetrix		METABRIC (ER+)
	training*	validation†	
number	266	991	1518
lymph node (-/+)	266/0	782/209	829/689
grade (1/2/3/NA)	80/141/43/2	32/99/37/823	166/712/570/70
size (≤ 2 cm/> 2 cm/NA)	162/102/0	90/113/788	669/835/14
Age (< 50/≥ 50/NA)	110/156/0	42/161/788	247/1271/0
8-year distant relapse event (no/yes/NA)	220/46/0	757/226/8	1190/327/1
BCFI event (no/yes/NA)	NA	NA	1142/375/1
HER2 status (+/-/NA)	NA	NA	268/1245/5
Hormone therapy (yes/no)	15/251	559/432	1088/430
Chemotherapy (yes/no)	0/266	0/991	164/1354

* GSE3494, GSE7390

† GSE12093, GSE6532 (Oxford cohort), GSE2034, GSE11121, GSE17705

Table S2. Characteristics of the patients in the FFPE datasets used in this study

	FFPE training set	FFPE validation set
Number	36	36
Age (< 50/ ≥ 50	15/21	15/21
Grade (1/2/3)	8/21/7	3/24/9
Size (≤ 2cm / > 2 cm)	24/12	29/7
TAILORx risk groups (LR/IR/HR)	4/18/14	6/15/15
Oncotype Dx risk groups (LR/IR/HR)	9/16/11	9/15/12

Table S3 Probes used for the development of quantitative PCR based EarlyR assay development (TaqMan Custom Array Format)

Gene Symbol	Assay ID	Amplicon length
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 the array	

Table S4. Distributions of EarlyR risk strata in clinically defined subsets

	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 ≤ 2cm	669	70%	12%	18%
Size > 2cm	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%

Table S5. Prognostic significance (8-year DRFI) of EarlyR in excess of clinical features in multivariate analysis in ER+ METABRIC cohort

Feature	p-value of feature	p-value of EarlyR in excess of feature*
LN	$< 10^{-16}$ §	3.3×10^{-16}
size (continuous)	$< 10^{-16}$ ¶	8.6×10^{-14}
size (≤ 2 cm / > 2 cm)	9.3×10^{-11}	1.9×10^{-14}
age (< 50 / ≥ 50)	0.45	3.3×10^{-16}
grade	3.2×10^{-5}	2.5×10^{-12}
LN + size + age + grade	$< 10^{-16}$ †	1.6×10^{-12}

* p-value of likelihood ratio of Cox proportional hazard model additively including EarlyR strata in comparison to model with only the clinical feature(s).
§ Chi-squared statistic (df = 1) is 127
¶ Chi-squared statistic (df = 1) is 78
† Chi-squared statistic (df = 5) is 87

Table S6. Expected survival probabilities (8-year DRFI) 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)

Supplementary Figures

Figure S1.
Flowchart detailing the steps associated with the development of the EarlyR gene signature.

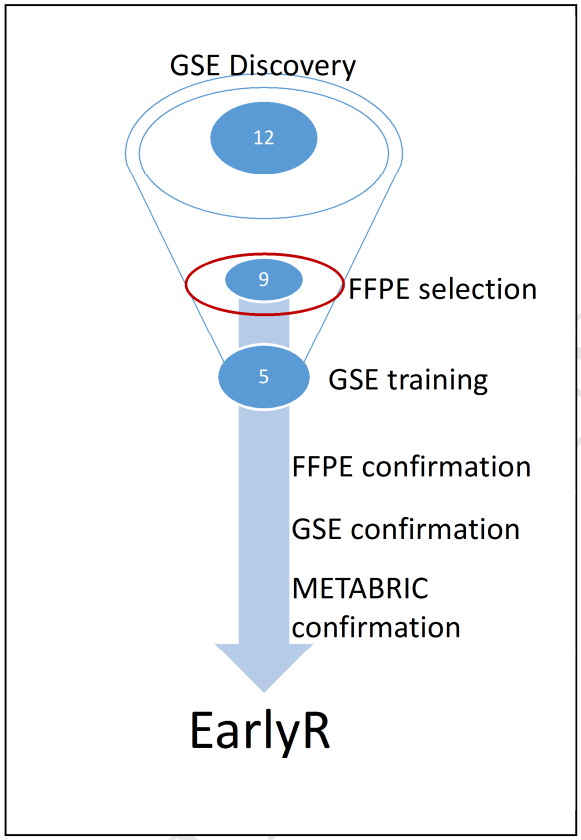


Figure S2.
For samples from the FFPE validation set (n = 36), Oncotype DX Recurrence Score (RS) is linearly dependent on EarlyR strata (p = 0.001). Samples are also colored by TAILORx risk group.

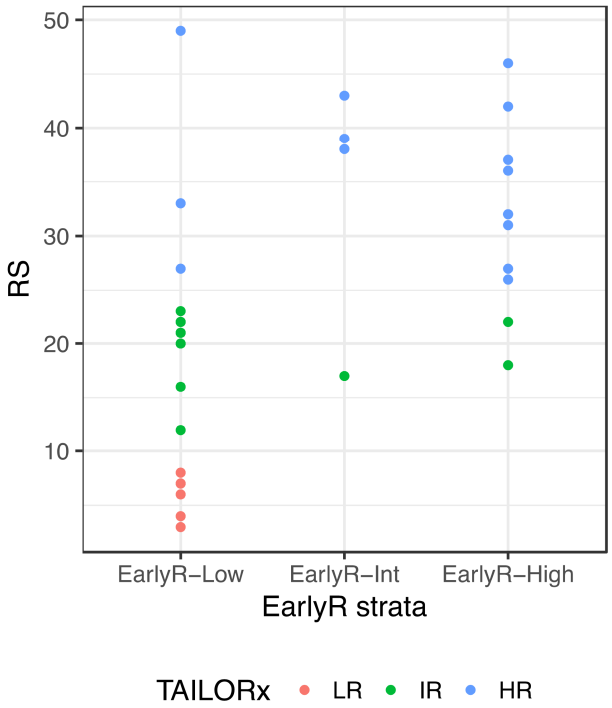


Figure S3. Estimated distant relapse-free survival eight years after diagnosis is plotted by quantiles of the continuous EarlyR score for (a) ER^+ Affymetrix validation (n = 991), (b) ER^+ METABRIC (n = 1518), (c) ER^+, LN^- Affymetrix validation (n = 782), (d) ER^+, LN^- METABRIC (n = 829), (e) ER^+, LN^+ Affymetrix validation (n = 209), (f) ER^+, LN^+ METABRIC (n = 689). EarlyR stratum membership is indicated by the color of the point.

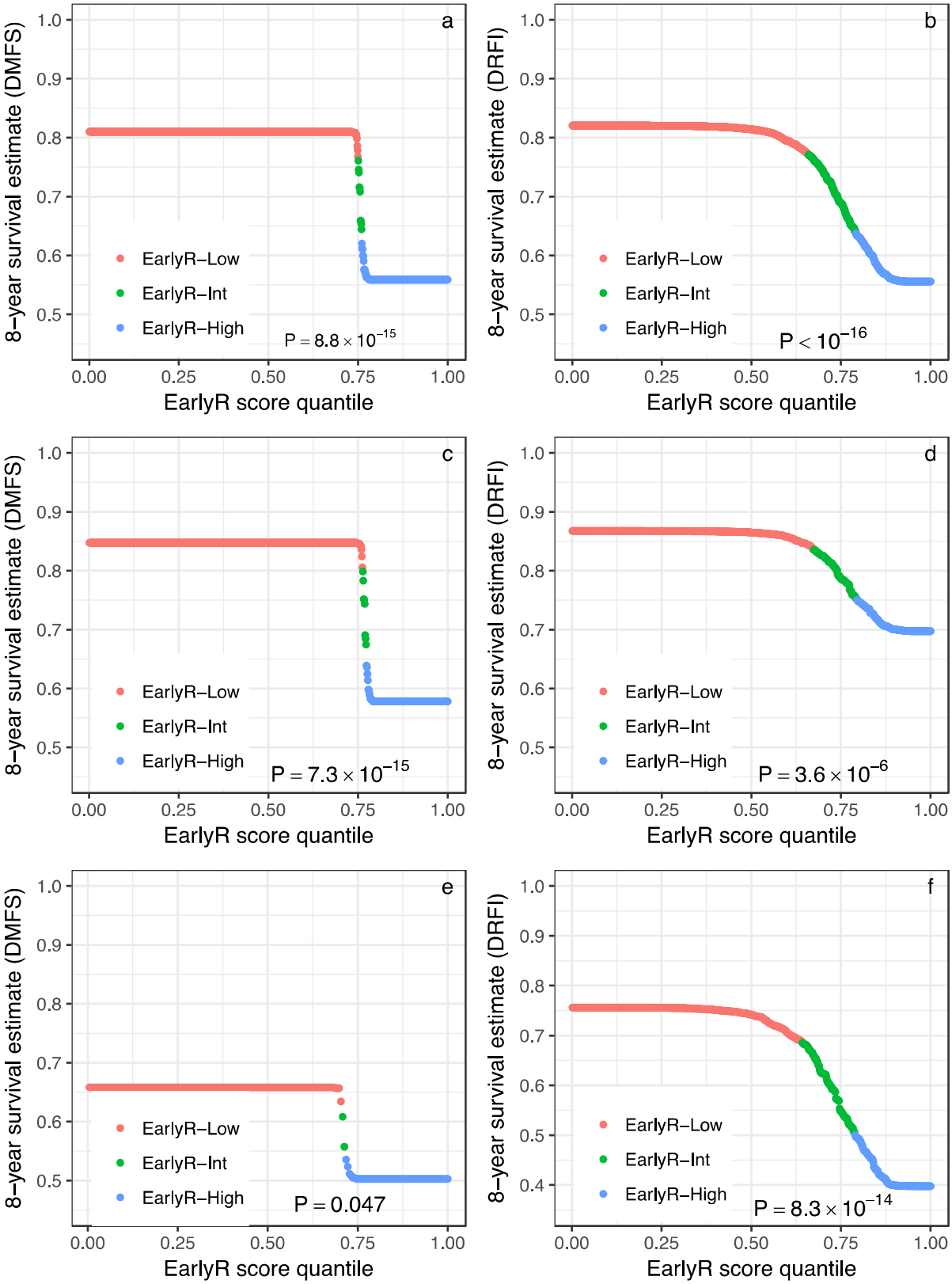


Figure S4. Estimated distant relapse-free survival eight years after diagnosis is plotted by quantiles of the continuous EarlyR score for the following cohorts treated with hormone therapy (HT): (a) ER^+ Affymetrix validation, HT treated (n = 559); (b) ER^+ METABRIC, HT treated (n = 1088); (c) ER^+, LN^- Affymetrix validation, HT treated (n = 369); (d) ER^+, LN^- METABRIC, HT treated (n = 445).

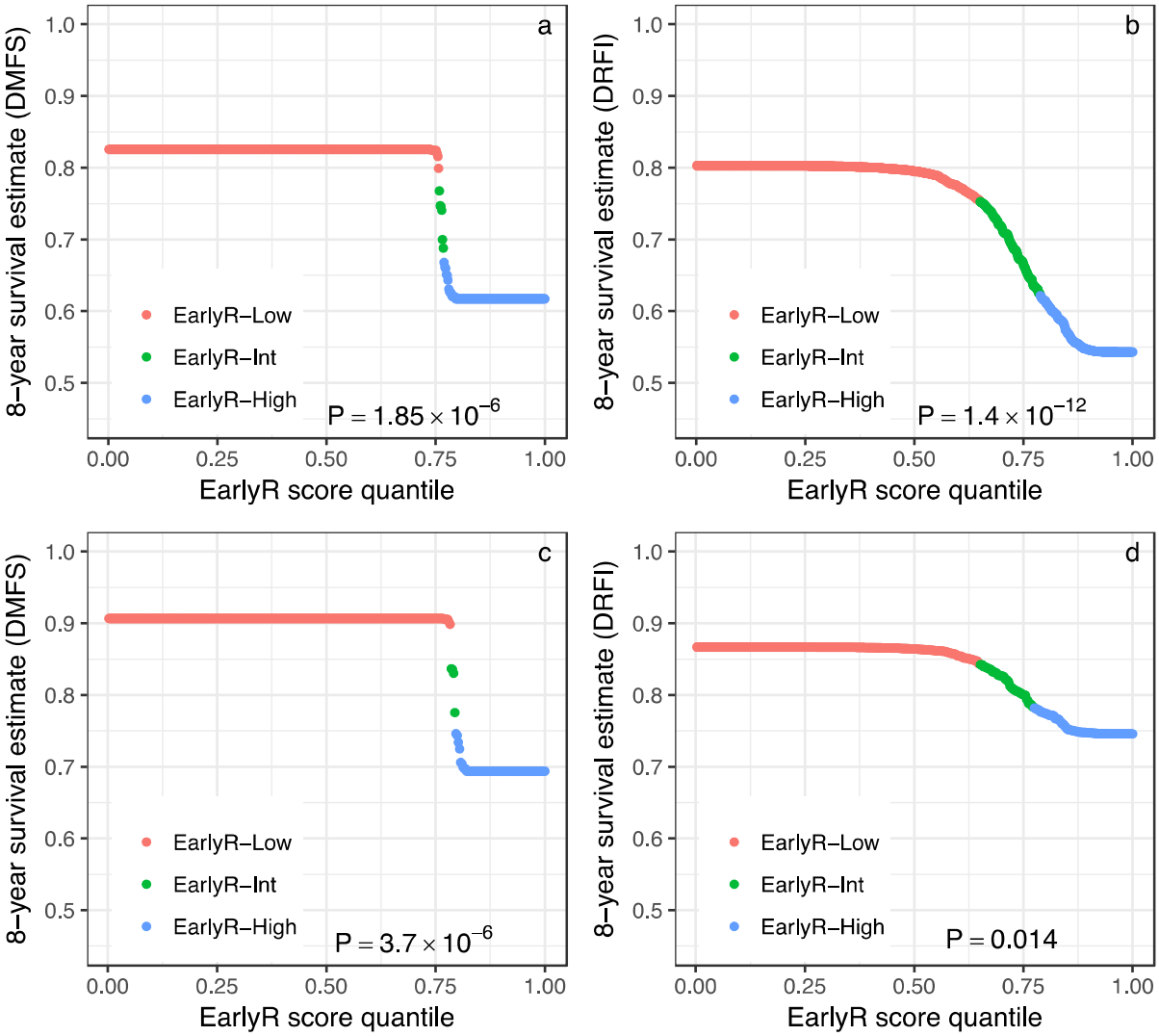


Figure S5. Concordance indices of genomic signatures with respect to 8-year DRFI are plotted for ER$^+$ METABRIC cohort. Square points indicate the concordance indices and lines are the 95% confidence intervals. Continuous scores were evaluated for each of EarlyR, Recurrence Score, GENE70 and Risk of Recurrence. All of these tests are statistically significant since the concordance index confidence intervals are all entirely greater than 0.5. The highest concordance index is for EarlyR (0.664).