Estrogen Receptor Pathway Activity Score to Predict Clinical Response or Resistance to Neoadjuvant Endocrine Therapy in Primary Breast Cancer


ABSTRACT

Endocrine therapy is important for management of patients with estrogen receptor (ER)-positive breast cancer; however, positive ER staining does not reliably predict therapy response. We assessed the potential to improve prediction of response to endocrine treatment of a novel test that quantifies functional ER pathway activity from mRNA levels of ER pathway–specific target genes. ER pathway activity was assessed on datasets from three neoadjuvant-treated ER-positive breast cancer patient cohorts: Edinburgh: 3-month letrozole, 55 pre-/posttreatment samples; TEAM IIa: 3- to 6-month exemestane, 49 pre-/28 posttreatment samples; and NEWEST: 16-week fulvestrant, 39 pretreatment samples. ER target gene mRNA levels were measured in fresh-frozen tissue (Edinburgh, NEWEST) with Affymetrix microarrays, and in formalin-fixed paraffin-embedded samples (TEAM IIa) with qRT-PCR. Approximately one third of ER-positive patients had a functionally inactive ER pathway activity score (ERPAS), which was associated with a nonresponding status. Quantitative ERPAS decreased significantly upon therapy (P<0.001 Edinburgh and TEAM IIa). Responders had a higher pretreatment ERPAS and a larger 2-week decrease in activity (P = 0.02 Edinburgh). Progressive disease was associated with low baseline ERPAS (P = 0.03 TEAM IIa; P = 0.02 NEWEST), which did not decrease further during treatment (P = 0.003 TEAM IIa). In contrast, the staining-based ER Allred score was not significantly associated with therapy response (P = 0.2). The ERPAS identified a subgroup of ER-positive patients with a functionally inactive ER pathway associated with primary endocrine resistance. Results confirm the potential of measuring functional ER pathway activity to improve prediction of response and resistance to endocrine therapy.

Introduction

Endocrine therapy is one of the mainstays in treatment of both early and metastatic breast cancer. Especially, the use of endocrine therapy has resulted in increased survival rates (1–3). Patients are currently selected for endocrine therapy using IHC analysis of estrogen receptor (ER) and progesterone receptor (PR) expression (4). Both the American Society for Clinical Oncology and the European Society for Medical Oncology advise a threshold of 1% ER-positive tumor cells (5, 6). In practice, many clinicians and countries choose a threshold of 10% (7). More quantifiable analyses like Allred scoring and H-score have been developed and suggested for clinical application, but are currently not routinely used (5).

Despite the success of endocrine therapy in ER-positive breast cancer, up to half of patients do not show the expected response (8, 9). In addition to cancer tissue heterogeneity, several mechanisms have been proposed to explain lack of therapy response, like emergence of ESR1-activating mutations or activation of other signal transduction pathways upon pharmacologic inhibition of the ER pathway (10, 11). Standard IHC analysis detects the presence of nuclear ER protein. However, to what extent positive nuclear ER staining indicates actual functional activity of the ER pathway has not been addressed satisfactorily. A test to predict response to endocrine therapy based on measuring functional activity of the ER pathway is expected to improve decision making regarding endocrine therapy and/or alternative therapies (12).

A number of data-driven RNA-based tests have been developed to assess recurrence risk and predict response to endocrine therapy (8, 12–14). However, no test is available to measure functional ER pathway activity. A knowledge-based Bayesian network computational model for the ER pathway has been developed to assess functional activity of this pathway in tumor tissues (15, 16). The model used mRNA expression levels of 27 high evidence target genes of the ER transcription factor to infer the activation state of the ER pathway. Earlier analysis using this ER pathway model showed that only part of ER-positive patients had an active pathway, which was associated with lower risk of relapse after adjuvant tamoxifen...
treatment (15). Here, we report results on use of this ER pathway formed. Tumor volumes were measured by three-dimensional ultrasound (14). Data from six patients were included in a study, in which postmenopausal women with operable, locally advanced (T2, 3, 4b; N0-3; M0) ER-positive breast tumors were assigned to an experimental (fulvestrant for 16 weeks) and a control group. Clinical response was determined by RECIST 1.1 criteria assessed by manual palpation, or by evaluation of tumor volume in three dimensions. The Affymetrix HGU133Plus2.0 GeneChip microarray analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Edinburgh breast cancer cohort

Datasets GSE10281 (22) and GSE20181 (14, 23) contain data from a cohort of postmenopausal patients with large strongly ER-positive (Allred score 5 or more) breast cancers recruited at the Edinburgh Breast Unit. Patients underwent neoadjuvant treatment with 2.5 mg daily letrozole for three months. Both sets contain data from fresh-frozen biopsies with at least 20% tumor content. The Affymetrix HGU133Plus2.0 GeneChip analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Affymetrix GeneChip microarrays

Affymetrix HGU133Plus2.0 GeneChip microarray analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Edinburgh breast cancer cohort

Datasets GSE10281 (22) and GSE20181 (14, 23) contain data from a cohort of postmenopausal patients with large strongly ER-positive (Allred score 5 or more) breast cancers recruited at the Edinburgh Breast Unit. Patients underwent neoadjuvant treatment with 2.5 mg daily letrozole for three months. Both sets contain data from fresh-frozen biopsies with at least 20% tumor content. The Affymetrix HGU133Plus2.0 GeneChip analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Affymetrix GeneChip microarrays

Affymetrix HGU133Plus2.0 GeneChip microarray analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Edinburgh breast cancer cohort

Datasets GSE10281 (22) and GSE20181 (14, 23) contain data from a cohort of postmenopausal patients with large strongly ER-positive (Allred score 5 or more) breast cancers recruited at the Edinburgh Breast Unit. Patients underwent neoadjuvant treatment with 2.5 mg daily letrozole for three months. Both sets contain data from fresh-frozen biopsies with at least 20% tumor content. The Affymetrix HGU133Plus2.0 GeneChip analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Affymetrix GeneChip microarrays

Affymetrix HGU133Plus2.0 GeneChip microarray analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Edinburgh breast cancer cohort

Datasets GSE10281 (22) and GSE20181 (14, 23) contain data from a cohort of postmenopausal patients with large strongly ER-positive (Allred score 5 or more) breast cancers recruited at the Edinburgh Breast Unit. Patients underwent neoadjuvant treatment with 2.5 mg daily letrozole for three months. Both sets contain data from fresh-frozen biopsies with at least 20% tumor content. The Affymetrix HGU133Plus2.0 GeneChip analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Affymetrix GeneChip microarrays

Affymetrix HGU133Plus2.0 GeneChip microarray analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Edinburgh breast cancer cohort

Datasets GSE10281 (22) and GSE20181 (14, 23) contain data from a cohort of postmenopausal patients with large strongly ER-positive (Allred score 5 or more) breast cancers recruited at the Edinburgh Breast Unit. Patients underwent neoadjuvant treatment with 2.5 mg daily letrozole for three months. Both sets contain data from fresh-frozen biopsies with at least 20% tumor content. The Affymetrix HGU133Plus2.0 GeneChip analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Affymetrix GeneChip microarrays

Affymetrix HGU133Plus2.0 GeneChip microarray analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.

Edinburgh breast cancer cohort

Datasets GSE10281 (22) and GSE20181 (14, 23) contain data from a cohort of postmenopausal patients with large strongly ER-positive (Allred score 5 or more) breast cancers recruited at the Edinburgh Breast Unit. Patients underwent neoadjuvant treatment with 2.5 mg daily letrozole for three months. Both sets contain data from fresh-frozen biopsies with at least 20% tumor content. The Affymetrix HGU133Plus2.0 GeneChip analysis was performed on the samples of all cell line experiments as described before (15, 19). See Supplementary Methods for more details. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R (https://www.R-project.org) BioConductor affy and frma (20, 21) packages.
BrCa cell line test AffyP2 and GSE8597 AffyP2 Calibration of original 25 nmol/L aTEAMII PCR prototype plate.

Abbreviations: AffyA: Affymetrix HGU133A; AffyP2: Affymetrix HGU133Plus2.0. *TEAMII PCR prototype plate.

Table 1. Datasets used in this study.

<table>
<thead>
<tr>
<th>Set</th>
<th>Platform</th>
<th>Use</th>
<th>Set description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE8597</td>
<td>AffyA</td>
<td>Calibration of original 25 nmol/L E2</td>
<td>MCF7 cell line treated with 25 nmol/L E2 (n = 4) or DMSO (n = 4) for 24 hours.</td>
<td>(30)</td>
</tr>
<tr>
<td>GSE35428</td>
<td>AffyP2</td>
<td>Calibration of optimized 1 nmol/L E2</td>
<td>MCF7 cell line treated with 1 nmol/L E2 (n = 10) or ethanol (n = 10) for 24 hours.</td>
<td>(29)</td>
</tr>
<tr>
<td>GSE9936</td>
<td>AffyA</td>
<td>Calibration of AffyA 6 nmol/L E2</td>
<td>MCF7 cell cultures treated with 6 nmol/L E2 (n = 3) or vehicle (n = 3) for 24 hours.</td>
<td>(31)</td>
</tr>
<tr>
<td>BrCa cell line calibration</td>
<td>qRT-PCR*</td>
<td>Calibration of TEAMII PCR model</td>
<td>MCF7 cell cultures treated with 1 nmol/L E2 (n = 4) or vehicle (n = 4) for 16 hours.</td>
<td>This study</td>
</tr>
<tr>
<td>BrCa cell line test</td>
<td>qRT-PCR*</td>
<td>Agreement between AffyP2 and TEAMII PCR models</td>
<td>Biological replicate samples from 16 patients with breast cancer processed by both platforms in two different institutions (n = 16 × 2 matching samples).</td>
<td>(GSE17700)</td>
</tr>
<tr>
<td>GSE17700</td>
<td>AffyA</td>
<td>Agreement between AffyA and AffyP2</td>
<td>Evaluation of ERGAS decrease during AI and agreement with AffyA model.</td>
<td>(22)</td>
</tr>
<tr>
<td>AffyP2 Edinburgh (GSE10181)</td>
<td>AffyA</td>
<td>Correlation with AI response data</td>
<td>Biopsy samples of 55 patients of the Edinburgh cohort, collected at baseline and after 3 months neoadjuvant treatment with letrozole (n = 55 × 3 correlation with AI; n = 6 × 2 agreement with AffyA model).</td>
<td>(14, 23)</td>
</tr>
<tr>
<td>AffyA Edinburgh (GSE20181)</td>
<td>AffyP2</td>
<td>Correlation with AI response data</td>
<td>Forty-nine biopsy and 48 resection samples (28 paired samples) of the TEAM IIa cohort.</td>
<td>This study</td>
</tr>
<tr>
<td>TEAM IIa</td>
<td>qRT-PCR*</td>
<td>Correlation with AI therapy response</td>
<td>Forty-two samples from patients treated for 16 weeks with either 500 mg or 250 mg of the selective estrogen receptor degrader fulvestrant.</td>
<td>(24)</td>
</tr>
<tr>
<td>NEWEST</td>
<td>AffyP2</td>
<td>Correlation with fulvestrant response</td>
<td>Forty-two samples from patients treated for 16 weeks with either 500 mg or 250 mg of the selective estrogen receptor degrader fulvestrant.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Schematic representation of ER pathway model network. Left, schematic of the Affymetrix HGU133A (AffyA) model network; only genes and probe sets present in the AffyA platform were kept. Center, node type description. Right, schematic of TEAMII PCR model; qRT-PCR assays were developed for a subset of target genes of the AffyP2 model.

Adaptation of the ER pathway Bayesian network model between AffyP2 and AffyA microarray platforms consisted of selecting probe sets that were identical on both platforms and recalibrating on samples measured on the respective platforms. For lack of probe availability in the AffyA platform, nine probe sets and the CDH26 gene were removed from the AffyA computational model. The resulting AffyA model retained 50 probe sets, representing 27 of the 28 ER target genes. The model was calibrated using Affymetrix HGU133A mRNA expression data from dataset GSE9936 (Table 1; ref. 31). For the qRT-PCR platform, the 28-target gene ER pathway Bayesian network was reduced to a smaller network containing the most informative target genes from the AffyP2 model (15). The main reason for reducing the number of target genes in the qRT-PCR–based model was intended use on FFPE samples. To avoid multiplexing of qRT-PCRs leading to bad qRT-PCR performance, multiple separate diagnostic grade qRT-PCRs needed to be developed. Because frequently very small samples are available from biopsy FFPE material (as in the TEAM IIa study), the amount of RNA isolated from such samples is limiting with respect to the number of qRT-PCRs that can be performed. For this reason, a smaller selection of target genes was identified based on evidence of the gene being a direct ER target gene, as well as their capacity to differentiate between ER pathway inactive and active MCF7 cell cultures. Subsequently, this selection was validated by comparing inferred ER pathway activities obtained with the 28-gene Affymetrix model to inferred activities obtained with a reduced Affymetrix model adapted to this limited target gene set. Following this successful initial validation step, qRT-PCR assays were developed and validated according to standard procedures for each of the selected target genes and a
set of reference genes. A Bayesian computational qRT-PCR ER pathway model (TEAMII PCR ER pathway model) was constructed as described previously (15) and calibrated using the breast cancer cell line calibration set (Table 1). Target gene mRNA expression values from these cell line experiments were obtained as described below and used as input to calibrate and validate this TEAMII PCR model. A research use prototype qRT-PCR plate, containing reagents for performing the qRT-PCR reactions, was used in this TEAM IIa study (Molecular Pathway Dx, Philips, TEAMII PCR prototype plate). To facilitate comparison between results obtained on different measurement platforms, a normalized ER pathway activity score (ERPAS) was computed for all models: The inferred odds in favor of an active ER pathway were transformed into a (base 2) logarithmic scale and normalized to scores ranging from 0 to 100, where 0 corresponds to the lowest and 100 corresponds to the highest odds in favor of an active ER pathway that a specific model can infer.

In summary, the new 1 nmol/L E2 model for AffyU133P2.0 contains the PDKZ1 gene in addition to the gene list described before (15). The 6 nmol/L E2 AffyA model lost one gene for which no probe sets were available on the AffyA platform. For the qRT-PCR–based model, best performing target genes identified using AffyU133P2.0 results were selected for further development of the model toward a commercially available diagnostic qRT-PCR–based ER pathway test (www.Philips.com/Oncosignal).

Generation of ER target gene qRT-PCR expression data
Quantitative expression levels of target genes and reference genes were determined using one-step qRT-PCR. See Supplementary Methods for more details. Normalized quantification cycle (Cq) values were used as input for the TEAMII PCR ER pathway model. For the TEAM IIa trial samples, inferred ERPASs were calculated in a blinded manner at Philips and returned to Leiden University Medical Center (Leiden, the Netherlands) for correlation to therapy response.

DNA sequencing
Targeted sequencing was used to perform mutation analysis of ESR1 and PIK3CA genes on 28 available TEAM IIa resection samples with either a paired biopsy sample or an ERPAS on the resection sample. DNA was extracted from whole samples and, if possible, from macrodissected samples. See Supplementary Methods and ref. 32.

Evaluation of agreement between models
Agreement between ER pathway models was evaluated by correlating activity scores obtained by the different models on identical sample sets (Table 1). Linear relationship between inferred activity scores was assessed by orthogonal regression, assuming equal residual variance. Agreement was assessed by Pearson correlation and mean deviations (MD) from the fitted orthogonal regression line were used as indication of the measurement error between two models.

Statistical analysis
Statistical analysis was performed using R (https://www.R-project.org) and SPSS 23.0 (IBM). Paired t tests were used to assess mean decrease in ERPAS between baseline and posttreatment measurements. ë2 tests were used to assess association between outcome categories and dichotomized ERPAS. Two sample t tests and ANOVA were used to assess ERPAS and decrease in ERPAS association with clinical outcome and baseline parameters. All tests assumed unequal variance. Unless indicated otherwise, two-sided P values and 95% confidence intervals (CI) are reported.

Results
Development and validation of ER pathway models
Optimization of the previously described ER pathway model included adjusting sensitivity to match estradiol (E2) levels present in breast cancer tissue. For this, E2 levels were measured in tissue samples of ER-positive breast cancers. Measured concentrations varied between 2.2 and 7 nmol/L E2 (Supplementary Table S2), in line with concentrations reported by others (33, 34). The new AffyP2 ER pathway model was calibrated on MCF7 breast cancer cells stimulated with 1 nmol/L estradiol, which lies between E2 concentrations in normal breast (<1 nmol/L) and in cancer tissue (7 nmol/L in this study). Results obtained with this 1 nmol/L–calibrated ER model strongly correlated with results obtained with the previously described 25 nmol/L–calibrated ER model (15), corr = 0.94, and showed the expected increase in sensitivity (from 83% to 91%), without loss in specificity (92%; Supplementary Figs. S2 and S3). The 1 nmol/L model was used for analysis of AffyP2 datasets in this study. Because the ER pathway models are measurement platform specific, the AffyP2 model was adapted for use on AffyA and qRT-PCR data, as described in the Materials and Methods section, and compared with the 1 nmol/L AffyP2 reference model on sets of technical replicate samples (Table 1, ref. 35). Results obtained with the AffyA model strongly correlated with results obtained with the 1 nmol/L E2 AffyP2 model on the dataset with 16 × 2 technical duplicates (corr = 0.80, Supplementary Fig. S4A). Differences in sample preparation protocols reduced the correlation (corr = 0.63, Supplementary Fig. S4B), illustrating the importance of standardized sample preparation procedures (36, 37). The TEAMII PCR model correlated well with the AffyP2 model on 11 technical duplicates (corr = 0.83, Supplementary Fig. S5). MDs from the fitted regression lines give an indication of the measurement error between two methods and were similar when comparing the AffyP2 I nmol/L model versus the AffyA and TEAMII PCR models and the two AffyP2 models (MD = 3.1, 5.6, and 2.3, respectively).

Functional ER pathway activity before and after neoadjuvant endocrine treatment
Edinburgh cohort
For the AffyP2 Edinburgh cohort, ERPASs decreased significantly upon letrozole treatment (41.6 vs. 28.4), mean decrease of 13.2 (CI = 9.7–16.8; paired t test P < 0.001; Fig. 2A). ERPAS at the start of the treatment was positively correlated to decrease in ERPAS after three months of therapy, indicating that the higher the baseline ER pathway activity, the higher the decrease in ER pathway activity upon treatment (corr = 0.88; Fig. 2B). Clinical response data were not available for this cohort.

Similarly, in the AffyA dataset of the Edinburgh cohort (Table 2; Fig. 3A and B), ERPAS decreased significantly between baseline and two weeks (46.9 vs. 39, mean decrease of 7.9, paired t test P < 0.001) and between baseline and three months of letrozole (46.9 vs. 39.5, mean decrease of 7.5, paired t test P < 0.001). No significant difference in ERPAS was seen between two weeks and three months of treatment, indicating an early and maintained response to letrozole for the patient group as a whole. Again, baseline ERPAS and decrease in ERPAS were positively correlated, both after two weeks and three months of treatment (corr = 0.74 and 0.83, respectively, Fig. 3C and D). In this cohort, clinical response data were available and both baseline ERPAS and decrease in ERPAS after two weeks of letrozole treatment were significantly higher in clinical responders to letrozole than in nonresponders (P = 0.02, Table 2; Fig. 3B and E). Upon letrozole treatment, average ERPAS decreased in responders to the
same low activity score level of nonresponders and remained at such level up to the 3-month measurement point (Fig. 3A, B, and F).

Using a threshold for ER pathway activity, as described before (15), enables analysis at individual patient level with respect to clinical response to letrozole. To evaluate the predictive value of the AffyA ER pathway activity model in an unbiased way, a threshold of 42.7 was set prior to analysis. This value corresponds to an odds of 1:1, the point where the inferred probability of the ER pathway being active is 50%.

Table 2. Changes in ER pathway activity scores upon AI treatment.

<table>
<thead>
<tr>
<th>ER pathway activity</th>
<th>Status</th>
<th>Score</th>
<th>Paired t test (baseline vs. time point)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>Active (%)</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Baseline</td>
<td>38 (69%)</td>
<td>46.9</td>
<td>8.9</td>
</tr>
<tr>
<td>Two weeks</td>
<td>15 (24%)</td>
<td>39.0</td>
<td>6.2</td>
</tr>
<tr>
<td>Three months</td>
<td>17 (31%)</td>
<td>39.4</td>
<td>5.3</td>
</tr>
<tr>
<td>Stratified by response</td>
<td>Responders (n = 44)</td>
<td>34 (77%)</td>
<td>48.2</td>
</tr>
<tr>
<td></td>
<td>Nonresponders (n = 11)</td>
<td>4 (36%)</td>
<td>41.8</td>
</tr>
<tr>
<td>Two weeks</td>
<td>Responders</td>
<td>11 (25%)</td>
<td>39.0</td>
</tr>
<tr>
<td></td>
<td>Nonresponders</td>
<td>2 (18%)</td>
<td>39.1</td>
</tr>
<tr>
<td>Three months</td>
<td>Responders</td>
<td>14 (32%)</td>
<td>39.6</td>
</tr>
<tr>
<td></td>
<td>Nonresponders</td>
<td>3 (27%)</td>
<td>38.8</td>
</tr>
<tr>
<td>Decrease in ER pathway activity</td>
<td>Direction</td>
<td>Decrease in score</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>Decrease</td>
<td>Increase</td>
<td>Mean</td>
</tr>
<tr>
<td>Baseline – two weeks</td>
<td>47 (85%)</td>
<td>8 (15%)</td>
<td>7.9</td>
</tr>
<tr>
<td>Two weeks – three months</td>
<td>22 (40%)</td>
<td>33 (60%)</td>
<td>−0.47</td>
</tr>
<tr>
<td>Stratified by response</td>
<td>Responders (n = 44)</td>
<td>39 (89%)</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>Nonresponders (n = 11)</td>
<td>8 (73%)</td>
<td>2.7</td>
</tr>
<tr>
<td>Two weeks – three months</td>
<td>Responders</td>
<td>38 (86%)</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td>Nonresponders</td>
<td>6 (55%)</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Note: ER pathway activity score and decrease in activity score for the 55 complete cases of AffyA Edinburgh dataset (GSE20181), 44 responders and 11 nonresponders. t Tests reported as difference (CI), P value.
This preset value is very close to 41.5, the best cut-off point obtained by ROC analysis of the present dataset. Using the predefined 42.7 threshold, 69% of patients (n = 38) had an active pretreatment ER pathway. This proportion decreased to 24% (n = 13) after two weeks and slightly increased to 31% (n = 17) after three months of treatment (Table 2). At baseline, 77% (n = 34) of clinical responders had an active ER pathway, against only 36% (n = 4) of nonresponders (Table 2). Accordingly, a baseline ER pathway active state was significantly associated with response to therapy ($\chi^2$ test $P = 0.009$; Table 2). Univariate logistic regression analysis reported a protective OR of 0.17 (CI = 0.04–0.69; $P = 0.01$) on the dichotomous active/inactive ER pathway activity status and a corresponding protective OR of 0.35 (CI = 0.13–0.95; $P = 0.02$) for an increase in 10 points in the ERPAS. As reference, relation of response to established markers (baseline tumor grade and ER Allred score) was also analyzed (Supplementary Table S3). Although tumor grade was associated with response to therapy ($\chi^2$ test $P = 0.03$), ER Allred score was not ($\chi^2$ test $P = 0.2$). Sample size was small for reliable multivariate analysis, but initial analysis indicated that the ERPAS remained significant after adjustment for grade and Allred score (Supplementary Table S4).

Analysis of this cohort not only illustrated the expected decrease in ER pathway activity upon effective letrozole treatment, but also that roughly one third of ER-positive patients had a functionally inactive ER pathway, associated with nonresponder status. Finally, in the first two weeks of letrozole treatment, ERPAS decreased in the majority of patients (60%, n = 33; Table 2). In line, the observed decrease in ERPAS from baseline to two weeks was higher in responders than nonresponders (9.3 vs. 2.7; $P = 0.02$), and remained higher in responders after three months of treatment (8.6 vs. 2.9; $P = 0.13$).
TEAM Ila cohort

For the TEAM Ila cohort (Table 3), the average ERPAS was 42.2 at baseline (n = 49) and decreased to 33.9 (n = 48) after exemestane therapy. In the 28 paired samples (before and after therapy), mean decrease in ERPAS was 9.1 (paired t test P < 0.001). In this cohort, again a positive but weak correlation was found between baseline ERPAS and decrease in ERPAS during treatment (corr = 0.29; Supplementary Fig. S6).

To test the value of the TEAMII PCR ER pathway model to predict response to aromatase inhibitors (AI), activity scores were correlated to decrease in tumor size assessed by manual palpation (Fig. 4D; Supplementary Table S5). Patients with PD during exemestane treatment had a significantly lower ERPAS at baseline (28.4, n = 3) compared with patients with non-PD (44.6, n = 35); t test P = 0.03 (Supplementary Table S5, response assessed at last measurement). A relation between pretreatment ERPAS and clinical response was already present at three months exemestane (non-PD, n = 27 vs. PD = 28.7, n = 3; P = 0.04, Supplementary Table S5, response assessed at 3 months).

Palpation-based response assessment was available for 23 patients with corresponding paired biopsy/resection activity scores (Fig. 4A and C). During exemestane treatment, ERPAS decreased in all six patients with complete remission (CR). In two of seven patients with partial response (PR) and in three of eight patients with stable disease (SD), an increase in ERPAS was measured after neoadjuvant therapy, indicating failure of blocking ER pathway activity at least at the time of surgical resection. For the two patients with PD with available paired samples, ERPAS was low before and had not changed after treatment. Decrease in ERPAS was higher in non-PD (9.9, n = 21) than PD (–0.16, n = 2), paired t test P = 0.003 (Fig. 4C). Although ERPAS did not differ significantly at baseline for CR, PR, and SD patients, a clear distinction was observed between CR/PR and PD patients (Fig. 4A). In addition, decrease in ERPAS was larger in CR patients, becoming gradually smaller in PR and SD patients (Fig. 4D).

For this cohort, available clinical data and sample size was too limited for extensive analysis. However, brief examination of the association of response to established risk markers did not indicate any of the available markers were associated to PD, and ERPAS remained associated to PD after adjustment.

NEWEST cohort

For the NEWEST cohort, Affymetrix data were only available from biopsies prior to start of neoadjuvant fulvestrant. Data were analyzed using the AffyP2 ER model (Fig. 4B). Baseline ERPAS was 28.7 in the 2

Table 3. ER pathway activity score at baseline and tumor resection for the TEAM Ila cohort.

<table>
<thead>
<tr>
<th>Sample availability</th>
<th>Baseline only</th>
<th>Paired</th>
<th>Resection only</th>
<th>n</th>
<th>Mean ± SD</th>
<th>Mean ± SD</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ER pathway activity score baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessed by palpation at last measurement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>23</td>
<td>7</td>
<td>6</td>
<td>15</td>
<td>38 ± 10.9</td>
<td>38 ± 10.9</td>
<td>23 ± 13.3</td>
</tr>
<tr>
<td>CR</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>47.6 ± 7.8</td>
<td>47.6 ± 7.8</td>
<td>20.0 ± 7.1</td>
</tr>
<tr>
<td>PR</td>
<td>7</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>41.8 ± 13.0</td>
<td>41.8 ± 13.0</td>
<td>7 ± 11.7</td>
</tr>
<tr>
<td>SD</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>44.5 ± 9.1</td>
<td>44.5 ± 9.1</td>
<td>5.4 ± 15.2</td>
</tr>
<tr>
<td>PD</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>28.4 ± 6.5</td>
<td>28.4 ± 6.5</td>
<td>2 ± 0.28</td>
</tr>
</tbody>
</table>

Note: Measurements are from the TEAMII PCR model. Decrease in ER pathway activity score was calculated for paired samples from the same patient. Sample availability indicated as baseline only: number of patients with only a biopsy sample; paired: number of patients with paired biopsy and resection sample; resection only: number of patients with only resection sample.

Abbreviation: n, number of samples available for pathway analysis.
PD samples versus 36.6 in the 37 non-PD patients (difference of 8.0 CI = 2.2–13.7; t-test \(P = 0.02\)), similar to results of the TEAM IIa study. Again, there was no significant difference in ERPAS between PR and SD patients.

**Relation between ER pathway activity and PR IHC staining**

Combined ER/PR positive IHC staining is generally thought to be a better indicator of response to endocrine treatment than ER status alone. Therefore, combined positive staining might be a more accurate predictor of an active ER pathway. Indeed, the PR positive biopsy samples of the TEAM IIa cohort (\(n = 36, 73\%\)) had significantly higher ERPAS than the ER\(^+\)/PR\(^−\) samples (48.7 vs. 34.3, two sample \(t\)-test \(P = 0.003\)) and the decrease in ERPAS after treatment was larger in the ER\(^+\)/PR\(^−\) group, although this did not reach significance (Supplementary Table S6). However, baseline PR protein staining alone did not correlate with response to AI therapy, neither in the TEAM IIa cohort (this study) nor in the Edinburgh cohort (reported in ref. 38).

**Mutation analysis of ESR1 and PIK3CA genes in TEAM IIa cohort**

ESR1 and PIK3CA mutations could be assessed in 15 samples (13 patients) of the TEAM IIa study (Supplementary Figs. S7 and S8); remaining samples were of insufficient quality. Although nonsynonymous ERS1 SNVs were only found in patients with partial response or stable disease (Supplementary Table S7), no literature evidence as to their functionality was found. A known PIK3CA (H1047R) activating mutation was detected in one CR and one SD patient.

**Discussion**

We previously described the development of a new method to quantitatively measure activity of the ER pathway in a tissue sample using a computational model that interprets measured ER target gene mRNA levels (15). In the current study, its use to predict response to neoadjuvant endocrine treatment in patients with ER-positive breast cancer was evaluated, analyzing two clinical studies from Edinburgh (14, 22), the Dutch TEAM IIa cohort (25), and the NEWEST study (24). For this purpose, the ER pathway model was adjusted to match sensitivity to E2 concentrations in cancer tissue, and adapted to the different mRNA measurement platforms used in these clinical studies. After freezing the models, analysis of the independent patient cohorts was performed to provide per sample a quantitative ERPAS.

**Endocrine treatment reduced ER pathway activity**

In all three clinical studies for which pre- and posttreatment data were available, endocrine treatment induced a decrease in ERPAS, and the decrease was already maximal after two weeks of treatment. In the studies for which clinical response data were available, the ERPAS prior to treatment correlated with clinical response. In the AffyA Edinburgh cohort, a high ERPAS was associated with favorable clinical response to letrozole. In the TEAM IIa and NEWEST studies, a low ERPAS was associated with progressive disease under endocrine therapy, as we observed. Results are also in line with our earlier finding that patients with ER-positive breast cancer with low ERPAS, treated with adjuvant tamoxifen, have a shorter relapse-free survival (15).

In the TEAM IIa study, ERPAS was not significantly different among the CR, PR, and SD groups. A contributing factor may have been inaccuracy in assessment of tumor size by manual palpation, which was performed by different MDs at nonfixed time points during treatment. Nevertheless, a detailed comparison of tumor size at resection versus estimated size by palpation, mammography, US, and MRI in the TEAM IIa cohort suggested palpation as most accurate measurement method (40). In addition, in the NEWEST study, threedimensional ultrasound assessment of tumor volume was used to assess clinical response, and again, a low ERPAS was associated with PD, while ERPAS could not statistically separate PR and SD. However, in both clinical studies, best responder patients (respectively CR and PR) were clearly distinguishable from PD patients by higher activity scores.

Positive ER/PR IHC staining is generally thought to be a more reliable indicator of response to endocrine therapy than positive ER staining alone. However, although PR staining did correlate ERPAS, it did not correlate with response to AI, indicating that combined ER/PR IHC alone is insufficient to decide on an active ER pathway and endocrine treatment, in agreement with clinical experience (41). This is likely due to the indirect relationship between ER transcriptional activity and expression level of the actual PR protein (42).

The Allred score for ER IHC staining has also been described as a predictive marker for neoadjuvant hormonal therapy response, with improved response at Allred scores higher than 5 (1). In the AffyA Edinburgh cohort for which this score was available, no relation with response to therapy was found, probably due to intentional preselection bias toward high Allred scores to enrich the responders group (38).

**Persistent ER pathway activity under AI treatment**

In a subgroup of patients from the AffyA Edinburgh and TEAM IIa cohorts (16% and 25%, respectively), ERPAS had increased at the end of AI treatment, indicating failure of blocking ER pathway activity, at least at the time of surgical resection or last biopsy. Possible explanations for measuring persistent ER pathway activity under AI are lack of compliance, stopping therapy prior to surgical resection, tumor heterogeneity, and emerging endocrine resistance. In both studies, patients were followed closely and compliance was estimated as high by the treating oncologists; however, noncompliance cannot be excluded with certainty. Tumor heterogeneity with respect to ER active and ER inactive cancer cell clones is another explanation, as shown recently in a preliminary study (43). A side study comparing activity scores in different areas of resected tumor from six TEAM IIa patients showed a variation in the ERPAS of 9 points or more in four patients (Supplementary Fig. S9). Resistance mechanisms such as ER activating mutations and activation of other tumor driving signaling pathways are well-known consequences of endocrine therapy, and the incidence of ER-activating mutations lies around 12% after treatment (44, 45). The observed increase in ERPAS between two weeks and three months treatment in the AffyA Edinburgh study, especially in patients with an initial reduction in activity score, is suggestive of ER-activating ESR1 mutations, unfortunately, no mutation information was available. In the TEAM IIa study, due to lack of sufficient tissue, ESR1 mutations could only be assessed in a few samples by targeted sequencing and only ESR1 SNVs with unknown functionality and unrelated to AI resistance or clinical response were...
identified. Sequencing of PIK3CA mutations revealed two known PIK3CA activating mutations, also without relation to response. These results emphasize the challenge of interpreting mutations with respect to functional impact, in line with recent clinical findings (46).

**ER pathway model performance using different mRNA measurement techniques**

An important reason for successful adaptation of the ER pathway model to other mRNA measurement platforms than the original AffymetrixU133Plus2.0, is the Bayesian network model approach based on causal relationships between the ER transcription factor and well-established ER target genes (15, 16). The differences between this knowledge-based approach to develop signaling pathway tests, and data-driven methods, have been described and discussed before (15, 19, 47, 48). In brief, data-driven efforts to generate pathway gene expression signatures, such as gene set enrichment analysis and DAVID, have been typically based on discovery of noncausal associations between expressed genes and a pathway, by analyzing datasets with more or less specified relationships to the pathway (49, 50). Such methods are prone to finding spurious associations and carry a high risk of overfitting, which interferes with biological validation and performance on independent datasets and with adaptation to other gene measurement platforms (51). In contrast, the Bayesian ER pathway model was biologically validated, can calculate the ERPAS of an individual sample despite variations in the expressed target gene subset, and its performance is relatively independent of mRNA measurement platform used. The latter is illustrated by the highly comparable results obtained by analysis of the different patient cohorts, where ERPAS was measured using different mRNA measurement platforms. For example in the TEAM Ia and NEWEST cohorts, a low ERPAS predicted progressive disease under endocrine therapy, irrespective of being measured on fresh-frozen samples with Affymetrix microarrays, or on FFPE samples with qRT-PCR.

**Predicting response to endocrine therapy**

Despite strongly positive ER staining, some patients fail to respond to neoadjuvant endocrine therapy, potentially resulting in progressive disease. By measuring ER pathway activity, we were able to identify such patients based on their low ERPAS. By providing additional evidence that ER-positive breast cancers do not always have high ER pathway activity, we provide a rational explanation for the clinical evidence that ER-positive breast cancers do not always have high ER expression (15, 19, 47–49). In brief, data-driven efforts to generate pathway gene expression signatures, such as gene set enrichment analysis and DAVID, have been typically based on discovery of noncausal associations between expressed genes and a pathway, by analyzing datasets with more or less specified relationships to the pathway (49, 50). Such methods are prone to finding spurious associations and carry a high risk of overfitting, which interferes with biological validation and performance on independent datasets and with adaptation to other gene measurement platforms (51). In contrast, the Bayesian ER pathway model was biologically validated, can calculate the ERPAS of an individual sample despite variations in the expressed target gene subset, and its performance is relatively independent of mRNA measurement platform used. The latter is illustrated by the highly comparable results obtained by analysis of the different patient cohorts, where ERPAS was measured using different mRNA measurement platforms. For example in the TEAM Ia and NEWEST cohorts, a low ERPAS predicted progressive disease under endocrine therapy, irrespective of being measured on fresh-frozen samples with Affymetrix microarrays, or on FFPE samples with qRT-PCR.

**References**


ERPS to Predict Response to Endocrine Therapy


Molecular Cancer Therapeutics

Estrogen Receptor Pathway Activity Score to Predict Clinical Response or Resistance to Neoadjuvant Endocrine Therapy in Primary Breast Cancer


Updated version  Access the most recent version of this article at: doi:10.1158/1535-7163.MCT-19-0318

Supplementary Material  Access the most recent supplemental material at: http://mct.aacrjournals.org/content/suppl/2019/11/14/1535-7163.MCT-19-0318.DC1

Cited articles  This article cites 48 articles, 13 of which you can access for free at: http://mct.aacrjournals.org/content/19/2/680.full#ref-list-1

Citing articles  This article has been cited by 6 HighWire-hosted articles. Access the articles at: http://mct.aacrjournals.org/content/19/2/680.full#related-urls

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link http://mct.aacrjournals.org/content/19/2/680. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.