Estrogen Receptor pathway activity score to predict clinical response or resistance to neo-adjuvant endocrine therapy in primary breast cancer

Estrogen Receptor pathway activity score to predict clinical response or resistance to neo-adjuvant endocrine therapy in primary breast cancer

M. A. Inda[3], E. Blok[1,2], P. J. K. Kuppen[1], A. Charehbili[1], E. den Biezen[4], A. van Brussel[3], S. E. Fruytier[1], E. Meershoek-Klein Kranenbarg[1], S. Kloet[5], B. van der Burg[6], J. Martens[7], A. H. Sims[8], A. K. Turnbull[8,9], J. M. Dixon[9], W. Verhaegh[3], J.R. Kroep[2], C. van de Velde[1], A. van de Stolpe[3]

1. Department of Surgery, Leiden University Medical Center, Leiden, the Netherlands
2. Department of Medical Oncology, Leiden University Medical Center, Leiden, the Netherlands
3. Philips Research, Eindhoven, The Netherlands
4. Philips Molecular Pathway Dx, Eindhoven, The Netherlands
5. Leiden Genome Technology Center, Leiden University Medical Center, The Netherlands
6. BioDetection Systems, Amsterdam, The Netherlands
7. Medical Oncology, Erasmus MC, Rotterdam, The Netherlands
8. Applied Bioinformatics of Cancer, University of Edinburgh Cancer Research UK Centre, MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK
9. Edinburgh Breast Unit, Western General Hospital, Edinburgh, UK

Contribution:
M. A. Inda: design study / cell line experiments, pathway model development, results analysis, writing
E. Blok: study coordinator, trial results analysis, data interpretation, writing
P. J. K. Kuppen: design/supervision of the study
A. Charehbili: study coordinator, trial results analysis, data interpretation
E. den Biezen: development of qPCRs for the ER pathway model test; analysis
A. van Brussel: development of qPCRs for the ER pathway model test; analysis
S. E. Fruytier: samples, TEAM IIa study
E. Meershoek-Klein Kranenbarg: samples, TEAM IIa study
S. Kloet: sequencing
B. van der Burg: E2 determination
J. Martens: providing BrCa samples for estradiol measurement
A. Sims: neoadjuvant study data and analysis
A. Turnbull: neoadjuvant study data and analysis
M. J. Dixon: neoadjuvant study, clinical perspective
W. Verhaegh: pathway model development
J. Kroep: TEAM IIa study, data interpretation, clinical perspective.
C. van de Velde: TEAM IIa study, clinical perspective.
A. van de Stolpe: design study / cell line experiments, results analysis, writing

Running title (60 characters): ER pathway score to predict response to endocrine therapy

Keywords (5 keywords): Estrogen Receptor positive breast cancer, signal transduction pathways, endocrine therapy response, neoadjuvant treatment; Bayesian modeling

Additional information.

Corresponding Author: Márcia A. Inda,
Address: Precision Diagnostics Department,
Philips Research Laboratories,
High Tech Campus 11, room p.2.47
5656 AE Eindhoven - The Netherlands
Phone: +31 40 27 30293
Email: marcia.inda@philips.com

Potential conflict of interest statement and financial support

M. A. Inda, A. van Brussel, A. van de Stolpe, and W. Verhaegh are Philips research employees. A. van de Stolpe has Philips stocks. E. den Biezen is employee of Philips Molecular Pathway Dx. B. van der Burg is Employee BioDetection Systems. In the last 5 years, J. Martens received funding from Philips Research, Therawis. Pamgene, and Sanofi.

In the last 5 years, J.R. Kroep received funding from Philips. Genomic Health funded the measurements from GSE20181 via a research grant to the University of Edinburgh.

Patents relevant for this publication: 1) Patent applicant: Koninklijke Philips NV: WO2013/011479; granted; inventors: W. Verhaegh, A. van de Stolpe, H. van Ooijen, K. Dulla, M.A. Inda, R. Hoffmann; relation to publication: ER pathway model. 2) Patent applicants: Medical Research Council (GB), Lothian Health Board (GB), The University Court of the University of Edinburgh (GB): WO 2016/092299, granted, inventors A. H. Sims, A. K. Turnbull, J. M. Dixon.

word count: 5195

total number of figures and tables: 3 tables and 4 figures

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 months letrozole, 55 pre-/two-week/post-treatment matched samples; TEAM IIa: 3-6 months exemestane, 49 pre-/28 post-treatment paired samples; NEWEST: 16 weeks fulvestrant, 39 pre-treatment 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 RT-qPCR.

Approximately one third of ER-positive patients had a functionally inactive ER pathway activity score (ERPAS), which was associated with a non-responding status. Quantitative ERPAS decreased significantly upon therapy ($p<0.001$ Edinburgh and TEAM IIa). Responders had a higher pre-treatment ERPAS and a larger two-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 immunohistochemical 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 pharmacological inhibition of the ER pathway (10,11). Standard immunohistochemical 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 model to predict and assess response to endocrine therapy in patients with ER-positive breast cancer treated in a neo-adjuvant setting.

Materials and Methods

Cell line cultures and Estradiol levels in breast cancer tissue

Estradiol (E2) stimulation experiments, RNA isolation of MCF7, T47D, CAMA1, and BT474 breast cancer ER-positive cell lines (ATCC), and measurement of 17-beta-estradiol concentrations in eight fresh frozen ER-positive breast tissue samples supplied by Erasmus Medical Center (Rotterdam, The

Netherlands) were performed by BioDetection Systems (BDS, Amsterdam). See supplemental methods.

Affymetrix GeneChip microarrays

Affymetrix HGU133Plus2.0 GeneChip microarray analysis was performed on the samples of the cell line experiments as described before (15,17). See supplemental methods. Data were deposited in GEO (GSE127760). All Affymetrix (CEL) data (generated in this study or retrieved from GEO) were processed using R/BioConductor *affy* and *frma* (18–20) packages.

Edinburgh breast cancer cohort

Datasets GSE10281 (21) and GSE20181 (14,22) contain data from a cohort of post-menopausal patients with large strongly ER-positive (Allred score 5 or more) breast cancers recruited at the Edinburgh Breast Unit. Patients underwent neo-adjuvant 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 GSE10281 (AffyP2 Edinburgh) dataset concerns 18 patients with samples collected at baseline and after three months of treatment. The Affymetrix HGU133A GSE20181 (AffyA Edinburgh) dataset contains gene expression data from biopsies of 62 patients collected at baseline and after approximately two weeks and three months of treatment. Clinical response was defined as tumor volume reduction of at least 50%, assessed after three months using three-dimensional ultrasound (14). Data from six patients was included in both datasets (overlap).

NEWEST breast cancer cohort.

Dataset GSE48905 (23) contains data from a multi-center phase II study, in which post-menopausal women with operable, locally advanced (T2, 3, 4b; N0-3; M0) ER-positive breast tumors were randomized for neo-adjuvant treatment with 500 mg or 250 mg fulvestrant for 16 weeks. Tumor core biopsies (fresh frozen tissue) were obtained before start of treatment and Affymetrix HGU133Plus2.0 microarrays were performed. Tumor volumes were measured by three-dimensional ultrasound and clinical response was classified as complete response (CR, disappearance of all lesions), partial response (PR, at least 65% reduction in tumor volume), progressive disease (PD, at least 73% increase in tumor volume), or stable disease (SD, none of the previous).

TEAM IIa breast cancer cohort

The Dutch TEAM IIa trial is a completed neo-adjuvant phase II prospective trial (24). Briefly, 102 patients with ER-positive ($\geq$50% nuclear ER staining) breast cancer were randomized to receive three or six months neo-adjuvant exemestane treatment. Due to slow accrual, the study changed to single arm design consisting of 6 months therapy. Standard clinicopathological characteristics, including PR and HER2 status, were assessed. Pre-treatment cancer tissue biopsies and post-neo-adjuvant

treatment tumor resection specimens were collected and analyzed retrospectively; all tissue samples were formalin-fixed paraffin-embedded (FFPE). A total of 49 biopsy and 48 resection samples were eligible for analysis, of which 28 samples were paired cases from the same patient (Supplemental Figure 5). Primary clinical outcome was reduction in tumor size using RECIST 1.1 criteria assessed by manual palpation.

LCM and RNA isolation for FFPE samples from TEAM IIa

Cancer tissue was separated from surrounding normal and stromal tissue using Laser Capture Microdissection (LCM) as described by Espina *et al* (24). See supplemental methods. In FFPE biopsy samples low tumor content limited the amount of available mRNA.

ER pathway models

The method used to develop the ER pathway model has been described before (15). The model uses mRNA expression levels of ER target genes measured in a tissue sample to infer the odds in favor of a transcriptionally active ER transcription factor and, as a consequence, odds in favor of an active ER pathway. A Bayesian network representing the ER pathway transcriptional program (15) describes (a) how target gene regulation depends on ER transcription complex activity and (b) how expression level intensities in turn depend on regulation of the respective target genes. The network consists of three types of nodes: (i) ER transcription complex activation node, (ii) target gene regulation node, with states 'down' and 'up', and (iii) expression intensity level nodes, with states 'low' and 'high', each corresponding to an ER target gene. The previously described ER pathway model was developed for analysis of Affymetrix HGU133Plus2.0 (AffyP2) data, i.e., target gene expression level nodes correspond to AffyP2 probesets (15). During ongoing development towards a diagnostic ER pathway test to further improve the sensitivity and specificity of the previously described ER pathway model , an additional informative ER target gene (PDZK1) was included (15,25,26). Following measurements of breast cancer tissue estrogen levels, the model was adjusted to estradiol levels in breast tissue by recalibration on data from MCF7 cultures exposed to 1 nM estradiol (E2) from GSE35428 (27, Table 1), instead of 25 nM E2 used to calibrate the original model (28, Table 1). Agreement between the original and optimized AffyP2 model was assessed on a collection of 5,395 breast cancer tissue and cell line samples (Supplemental Table 2). Sensitivity and specificity were assessed on 421 breast cancer cell line samples with well-established ER pathway activity (Supplemental Table 2). Subsequently, the sensitivity-optimized (1nM) ER pathway model was adapted for use on two other mRNA measurement platforms, i.e., Affymetrix HGU133A (AffyA) and 1-step RT-qPCR (PCR). Adaptation of the ER pathway Bayesian network model between AffyP2 and AffyA microarray platforms consisted of selecting probesets that were identical on both platforms

and re-calibrating on samples measured on the respective platforms. For lack of probe availability in the AffyA platform, nine probesets and the CDH26 gene were removed from the AffyA computational model. The resulting AffyA model retained 50 probesets, representing 27 of the 28 ER target genes. The model was calibrated using Affymetrix HGU133A mRNA expression data from dataset GSE9936 (29, Table 1). For the qPCR 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 qPCR-based model was intended use on FFPE samples. To avoid multiplexing of qPCRs leading to bad qPCR performance, multiple separate diagnostic grade qPCRs needed to be developed. Since frequently very small samples are available from biopsy FFPE material (as in the TEAMIIA study), the amount of RNA isolated from such samples is limiting with respect to the number of qPCRs 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 inactive and active MCF7 cell cultures. Subsequently this selection was validated by comparing ER pathway activity scores obtained with the 28 gene Affymetrix model with a reduced Affymetrix model adapted to this limited target gene set. Following this successful initial validation step, RT-qPCR (qPCR) assays were developed and validated according to standard procedures for each of the selected target genes and a set of reference genes. The Bayesian computational qPCR model (TEAMII PCR ER model) was constructed as previously described (15) and calibrated using the BrCa 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 the TEAMII PCR model. A research use prototype qPCR plate, containing reagents for performing the qPCR reactions, was used in this TEAM IIa study (Molecular Pathway Dx, Philips, Eindhoven, TEAMII PCR prototype plate). For all models, a normalized ER pathway activity score was used, as described under statistical methods.

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

Generation of ER target gene RT-qPCR expression data

Quantitative expression levels of target genes and reference genes were determined using one-step RT-qPCR. See supplemental methods. Normalized quantification cycle (Cq) values were used as input for the TEAMII PCR ER pathway model. For the TEAM IIa trial samples, inferred ER pathway activities

were calculated in a blinded manner at Philips and returned to Leiden University Medical Center 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 a high ER pathway activity score on the resection sample. DNA was extracted from whole samples and, if possible, from macrodissected samples. See supplemental methods.

Evaluation of agreement between models

Agreement between ER pathway models was evaluated by correlating activities 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 (18) and SPSS 23.0 (IBM). Odds in favor of an active ER pathway were transformed into a (base 2) logarithmic scale and then 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. This normalization facilitates comparison between activity scores obtained on different measurement platforms (Figure 4A/B). Paired t-tests were used to assess mean decrease in ER activity between baseline and post-treatment measurements, chi-squared tests to assess association between outcome categories and dichotomized ER pathway activity, and two sample t-tests and ANOVA to assess ER activity and decrease in activity associated with clinical outcome and baseline parameters. All tests assumed unequal variance. Unless indicated otherwise, 2-sided p-values and 95% confidence intervals 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 nM E2 (Supplemental Table 1), in line with concentrations reported by others (30,31). The new AffyP2 ER pathway model was calibrated on MCF7 breast cancer cells stimulated with 1 nM estradiol, which

lies between E2 concentrations in normal breast (< 1 nM) and in cancer tissue (7 nM in this study). Results obtained with this 1 nM-calibrated ER model strongly correlated with results obtained with the previously described 25 nM-calibrated ER model (17), corr=0.94, and showed the expected increase in sensitivity (from 83% to 91%), without loss in specificity (92%), Supplemental Figures 1 and 2. The 1 nM 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 qPCR data, as described in the methods section, and compared to the 1 nM AffyP2 reference model on sets of technical replicate samples (32, Table 1). Results obtained with the AffyA model strongly correlated with results obtained with the 1 nM E2 AffyP2 model on the dataset with 16x2 technical duplicates (corr=0.80, Supplemental Figure 3A). Differences in sample preparation protocols reduced the correlation, illustrating the importance of standardized sample preparation procedures (corr=0.63, Supplemental Figure 3B). The TEAMII PCR model correlated well with the AffyP2 model on 11 technical duplicates (corr=0.83, Supplemental Figure 4). Mean deviations (MD) from the fitted regression lines give an indication of the measurement error between two methods and were similar when comparing the AffyP2 1 nM model vs 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 neo-adjuvant AI treatment

Edinburgh cohort

For the AffyP2 Edinburgh cohort, ER activity score decreased significantly upon letrozole treatment (41.6 vs 28.4), mean decrease of 13.2 (CI=9.7-16.8, paired t-test p-value < 0.001, Figure 2A). ER activity scores at start of treatment were positively correlated to decrease in ER activity after three months of therapy, indicating that the higher the baseline ER activity, the higher the decrease in ER activity upon treatment (corr=0.88, Figure 2B). Clinical response data were not available for this cohort.

Similarly, in the AffyA dataset of the Edinburgh cohort (Table 2, Figure 3A/B), ER pathway activity score decreased significantly between baseline and two weeks (46.9 vs 39, mean decrease of 7.9, paired t-test p-value <0.001) and between baseline and three months of letrozole (46.9 vs 39.5, mean decrease of 7.5, paired t-test p-value <0.001). No significant difference in activity 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 and decrease in ER activity were positively correlated, both after two weeks and three months of treatment (corr=0.74 and 0.83, respectively, Figure 3C-D). In this cohort, clinical response data were available and both baseline ER pathway activity and decrease in ER activity after two weeks of AI treatment were significantly higher in

clinical responders to letrozole than in non-responders (p=0.02, Table 2, Figure 3B/E). Upon letrozole treatment, average ER activity scores decreased in responders to the same low ER activity level of non-responders and remained at such level up to the three months measurement point (Figure 3A/B/F).

Using a threshold for ER 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 ER pathway score 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%. This preset value is very close to 41.5, the best cutoff point obtained by ROC analysis of the present dataset. Using the predefined 42.7 threshold, 69% of patients (n=38) had an active pre-treatment 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 non-responders (Table 2). Accordingly, a baseline ER active state was significantly associated with response to therapy (chi-square p-value=0.009, Table 2). Univariate logistic regression analysis reported an odds ratio (OR) of 5.9 (CI=1.4-24.5, p-value=0.01) on the dichotomous active/inactive ER status and a protective OR of 0.35 (CI=0.13-0.95, p-value 0.02) for an increase in 10 points in the ER pathway activity score. As reference, relation of response to established markers (baseline tumor grade and ER Allred score) was also analyzed (Supplemental Table 3). While tumor grade was associated with response to therapy (chi-square test p-value=0.03), ER Allred score was not (chi-square p-value=0.2). Sample size was small for reliable multivariate analysis, but initial analysis indicated that the ER pathway score remained significant after adjustment for grade and Allred score (Supplemental Table 4).

Analysis of this cohort not only illustrated the expected decrease in ER activity upon effective letrozole treatment, but also that roughly one third of ER-positive patients had a functionally inactive ER pathway, associated with non-responder status. Finally, in the first two weeks of letrozole treatment, ER pathway activity decreased in the majority of patients (85%, n=47), while between two weeks and three months ER activity slightly increased in the majority of patients (60%, n=33, Table 2). In line, the observed decrease in ER activity from baseline to two weeks was higher in responders than non-responders (9.3 vs 2.7, p-value=0.02), and remained higher in responders after three months of treatment (8.6 vs 2.9, p-value=0.13).

TEAM IIa cohort

For the TEAM IIa cohort (Table 3) the average ER pathway activity score 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 activity score was 9.1 (paired t-test p-value < 0.001). In this cohort, again a positive but weak correlation was found between baseline ER activity and decrease in ER activity during treatment (corr=0.29, Supplemental Figure 6).

To test the value of the TEAMII PCR ER pathway model to predict response to aromatase inhibitors (AI), ER pathway activity was correlated to decrease in tumor size assessed by manual palpation (Figure 4D, Supplemental Table 6). Patients with progressive disease (PD) during AI treatment had a significantly lower ER pathway activity at baseline (28.4, n=3) compared to patients with non-PD (44.6, n=35); t-test p-value=0.03 (Supplementary Table 4, response assessed at last measurement). A relation between pre-treatment ER pathway activity scores and clinical response was already present at three months exemestane (non-PD=44.6, n=27 vs PD=28.7, n=3; p-value=0.04, Supplemental Table 6, response assessed at 3 months).

Palpation-based response assessment was available for 23 patients with paired biopsy/resection ER pathway scores (Figure 4A/C). During AI treatment, ER pathway activity 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 ER activity was measured after neo-adjuvant 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, ER activity was low before and had not changed after treatment. Decrease in activity score was higher in non-PD (9.9, n=21) than PD (−0.16, n=2), paired t-test p-value=0.003 (Figure 4C). Though ER pathway activity did not differ significantly at baseline for CR, PR, and SD patients, a clear distinction was observed between CR/PR and PD patients (Figure 4A). In addition, decrease in ER activity was larger in CR patients, becoming gradually smaller in PR and SD patients (Figure 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 ER activity score remained associated to PD after adjustment.

NEWEST cohort

For the NEWEST cohort, Affymetrix data were only available from biopsies prior to start of neo-adjuvant fulvestrant. Data were analyzed using the AffyP2 ER model (Figure 4B). Baseline ER activity score was 28.7 in the 2 PD samples vs 36.6 in the 37 non-PD patients (difference of 8.0 CI=2.2-13.7, t-test p-value=0.02), similar to results of the TEAM IIa study. Again, there was no significant difference 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 ER activity than the ER+/PR- samples (48.7 vs. 34.3, two sample t-test p-value=0.003) and the decrease in ER activity after treatment was larger in the ER+/PR+ group, although this did not reach significance (Supplemental Table 3). 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 33).

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 (Supplemental Figures 7 and 8); remaining samples were of insufficient quality. Though non-synonymous ERS1 SNVs were only found in patients with partial response or stable disease, 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 neo-adjuvant endocrine treatment in ER-positive breast cancer patients was evaluated, analyzing two clinical studies from Edinburgh (14,21), the Dutch TEAM IIa cohort (24) and the NEWEST study (23). 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 ER pathway activity score.

AI treatment reduced ER pathway activity

In all three clinical studies for which pre- and post-treatment data were available, AI treatment induced a decrease in ER pathway activity scores, and the decrease was already maximal after two weeks of treatment. In the studies for which clinical response data were available, the ER activity score prior to treatment correlated with clinical response. In the AffyA Edinburgh cohort, a high ER activity score was associated with favorable clinical response to AI. In the TEAM IIa and NEWEST studies, low ER pathway activity score was associated with progressive disease under AI and fulvestrant treatment, respectively. In general, patients with higher activity scores responded better

to endocrine treatment than patients with lower activity scores, and treatment-induced decrease in ER pathway activity was largest in clinical responders and smallest in progressive disease patients. These observations are in accordance with broad evidence that AI and fulvestrant therapies are effective in patients with an active ER pathway driving tumor growth, by interfering with production of estradiol, and by blocking of ER transcriptional activity and degrading ER, respectively (9,34). An inactive or minimally active ER pathway is expected to be associated with primary resistance and progressive disease under endocrine therapy, as we observed. Results are also in line with our earlier finding that ER-positive breast cancer patients with low ER pathway scores, treated with adjuvant tamoxifen, have a shorter relapse free survival (15).

In the TEAM IIa study, ER activity scores were not significantly different between 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 non-fixed 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 (35). Additionally, in the NEWEST study, three-dimensional ultrasound assessment of tumor volume was used to assess clinical response, and again low ER pathway activity was associated with progressive disease, while ER pathway activity score 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 ER pathway activity.

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, while PR staining did correlate with ER pathway activity, 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 (36). This is likely due to the indirect relationship between ER transcriptional activity and expression level of the actual PR protein (37).

The Allred score for ER immunohistochemistry staining has also been described as a predictive marker for neoadjuvant hormonal therapy response, with improved response at Allred scores higher than 5 (3). In the AffyA Edinburgh cohort for which this score was available, no relation with response to therapy was found, probably due to intentional pre-selection bias towards high Allred scores to enrich the responders group (33).

Persistent ER pathway activity under AI treatment

In a subgroup of patients from the AffyA Edinburgh and TEAM IIa cohorts (16% and 25%, respectively), ER pathway activity 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 non-compliance 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 (38). A side study comparing ER pathway activity in different areas of resected tumor from six TEAM IIa patients showed a variation in the ER pathway score of 9 points or more in four patients (Supplemental Figure 9). 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 with aromatase inhibitors (39,40). The observed increase in ER pathway activity score between two weeks and three months treatment in the AffyA Edinburgh study, especially in patients with an initial reduction in ER 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 (41).

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 (15,17,42,43). In brief, data-driven efforts to generate pathway gene expression signatures, such as Gene Set Enrichment Analysis (GSEA) and DAVID, have been typically based on discovery of non-causal associations between expressed genes and a pathway, by analyzing datasets with more or less specified relationships to the pathway (44–46). 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 (47). In contrast, the Bayesian ER pathway model was biologically validated can calculate the ER pathway activity score of an individual sample despite variations in the expressed target gene subset, and performance is relatively independent of the used mRNA measurement platform. The latter is illustrated by the highly comparable results obtained by analysis of the different patient cohorts, where ER pathway activity was measured using different mRNA measurement platforms. For example in the TEAM IIa and NEWEST cohorts, low ER activity scores predicted progressive disease under endocrine therapy, irrespective of being measured on fresh frozen samples with Affymetrix microarrays, or on FFPE samples with qPCR.

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 ER pathway activity score. 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 issue of non-responders. Analysis of three independent clinical breast cancer cohorts confirmed the potential value of measuring functional ER pathway activity to predict response to endocrine therapy. The Bayesian computational models for ER pathway activity, adapted to different measurement platforms, revealed comparable clinical results on fresh frozen tissue as well as FFPE samples. The ER pathway activity score has clinical value as a continuous measure: the higher the score, the more likely that the ER pathway was active in the analyzed sample and the more likely that the patient responded to anti-hormonal treatment. In a clinical sample consisting of a mixture of cells, an absolute ER pathway activity threshold is arbitrary, since the activity score represents averaged ER pathway activity in the cell mixture, thus depending on percentage of cancer cells and their level of ER pathway activity. In the future, a clinically useful approach might be the use of tertiles of activity scores, allowing middle tertile scores to indicate uncertain treatment response.

The frequently extremely small FFPE samples from pretreatment biopsies reflect the routine diagnostic situation and are a point of concern. Current developments are directed towards improvement of signaling pathway analysis on such samples. Further clinical validation of the ER pathway test is ongoing, as well as elucidating the role of other signaling pathways, such as the PI3K pathway, in hormonal resistance (17,42).

References

1. Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. Lancet. 2013;381:805–16.

2. Blok EJ, Derks MGM, van der Hoeven JJM, van de Velde CJH, Kroep JR. Extended adjuvant endocrine therapy in hormone-receptor positive early breast cancer: current and future evidence. Cancer Treat Rev. 2015;41:271–6.

3. Abrial C, Durando X, Mouret-Reynier M-A, Thivat E, Bayet-Robert M, Nayl B, et al. Role of neo-adjuvant hormonal therapy in the treatment of breast cancer: a review of clinical trials. Int J Gen Med. 2009;2:129–40.

4. Harvey JM, Clark GM, Osborne CK, Allred DC. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol. 1999;17:1474–81.

5. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol. 2010;28:2784–95.

6. Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2015;26 Suppl 5:v8-30.

7. Yi M, Huo L, Koenig KB, Mittendorf EA, Meric-Bernstam F, Kuerer HM, et al. Which threshold for ER positivity? a retrospective study based on 9639 patients. Ann Oncol. 2014;25:1004–11.

8. Selli C, Dixon JM, Sims AH. Accurate prediction of response to endocrine therapy in breast cancer patients: current and future biomarkers. Breast Cancer Res. 2016;18:118.

9. Spring LM, Gupta A, Reynolds KL, Gadd MA, Ellisen LW, Isakoff SJ, et al. Neoadjuvant Endocrine Therapy for Estrogen Receptor-Positive Breast Cancer. JAMA Oncol. 2016;2:1477–86.

10. Jeselsohn R, Buchwalter G, De Angelis C, Brown M, Schiff R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. Nat Rev Clin Oncol. 2015;12:573–83.

11. Ciruelos Gil EM. Targeting the PI3K/AKT/mTOR pathway in estrogen receptor-positive breast cancer. Cancer Treat Rev. 2014;40:862–71.

12. Ma CX, Bose R, Ellis MJ. Prognostic and Predictive Biomarkers of Endocrine Responsiveness for Estrogen Receptor Positive Breast Cancer. Adv Exp Med Biol. 2016;882:125–54.

13. Renoir J-M, Marsaud V, Lazennec G. Estrogen receptor signaling as a target for novel breast cancer therapeutics. Biochem Pharmacol. 2013;85:449–65.

14. Turnbull AK, Arthur LM, Renshaw L, Larionov AA, Kay C, Dunbier AK, et al. Accurate Prediction and Validation of Response to Endocrine Therapy in Breast Cancer. J Clin Oncol. 2015;33:2270–8.

15. Verhaegh W, van Ooijen H, Inda MA, Hatzis P, Versteeg R, Smid M, et al. Selection of personalized patient therapy through the use of knowledge-based computational models that identify tumor-driving signal transduction pathways. Cancer Res. 2014;74:2936–45.

16. Verhaegh W, van de Stolpe A. Knowledge-based computational models. Oncotarget. 2014;5:5196–7.

17. Ooijen H van, Hornsveld M, Dam-de Veen C, Velter R, Dou M, Verhaegh W, et al. Assessment of Functional Phosphatidylinositol 3-Kinase Pathway Activity in Cancer Tissue Using Forkhead Box-O Target Gene Expression in a Knowledge-Based Computational Model. Am J Pathol. 2018;188:1956–72.

18. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available from: https://www.R-project.org/

19. affy—analysis of Affymetrix GeneChip data at the probe level | Bioinformatics | Oxford Academic [Internet]. [cited 2018 Sep 19]. Available from: https://academic.oup.com/bioinformatics/article/20/3/307/185980

20. McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). Biostatistics. 2010;11:242–53.

21. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc Natl Acad Sci USA. 2009;106:13820–5.

22. Miller WR, Larionov A, Anderson TJ, Evans DB, Dixon JM. Sequential changes in gene expression profiles in breast cancers during treatment with the aromatase inhibitor, letrozole. Pharmacogenomics J. 2012;12:10–21.

23. Knudsen S, Jensen T, Hansen A, Mazin W, Lindemann J, Kuter I, et al. Development and Validation of a Gene Expression Score That Predicts Response to Fulvestrant in Breast Cancer Patients. PLoS One [Internet]. 2014 [cited 2018 Dec 10];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3914825/

24. Fontein DBY, Charehbili A, Nortier JWR, Meershoek-Klein Kranenbarg E, Kroep JR, Putter H, et al. Efficacy of six month neoadjuvant endocrine therapy in postmenopausal, hormone receptor-positive breast cancer patients--a phase II trial. Eur J Cancer. 2014;50:2190–200.

25. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012;481:389–93.

26. Lin Z, Reierstad S, Huang CC, Bulun SE. Novel estrogen receptor-alpha binding sites and estradiol target genes identified by chromatin immunoprecipitation cloning in breast cancer. Cancer Res. 2007;67:5017–24.

27. Wardell SE, Kazmin D, McDonnell DP. Research resource: Transcriptional profiling in a cellular model of breast cancer reveals functional and mechanistic differences between clinically relevant SERM and between SERM/estrogen complexes. Mol Endocrinol. 2012;26:1235–48.

28. Bourdeau V, Deschenes J, Laperriere D, Aid M, White JH, Mader S. Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. Nucleic Acids Res. 2008;36:76–93.

29. Chang EC, Charn TH, Park S-H, Helferich WG, Komm B, Katzenellenbogen JA, et al. Estrogen Receptors alpha and beta as determinants of gene expression: influence of ligand, dose, and chromatin binding. Mol Endocrinol. 2008;22:1032–43.

30. Honma N, Saji S, Hirose M, Horiguchi S, Kuroi K, Hayashi S, et al. Sex steroid hormones in pairs of tumor and serum from breast cancer patients and pathobiological role of androstene-3beta, 17beta-diol. Cancer Sci. 2011;102:1848–54.

31. Geisler J. Breast cancer tissue estrogens and their manipulation with aromatase inhibitors and inactivators. J Steroid Biochem Mol Biol. 2003;86:245–53.

32. Symmans WF, Hatzis C, Sotiriou C, Andre F, Peintinger F, Regitnig P, et al. Genomic Index of Sensitivity to Endocrine Therapy for Breast Cancer. JCO. 2010;28:4111–9.

33. Miller WR, White S, Dixon JM, Murray J, Renshaw L, Anderson TJ. Proliferation, steroid receptors and clinical/pathological response in breast cancer treated with letrozole. Br J Cancer. 2006;94:1051–6.

34. Nathan MR, Schmid P. A Review of Fulvestrant in Breast Cancer. Oncol Ther. 2017;5:17–29.

35. Blok EJ, Charehbili A, Kroep JR, Seynaeve CM, Velde C van de, Kuppen PJK. Abstract P6-01-05: Radiological evaluation of neo-adjuvant endocrine therapy in hormone-receptor positive early breast cancer. Cancer Res. 2016;76:P6-01-05-P6-01–5.

36. Ma CX, Sanchez CG, Ellis MJ. Predicting endocrine therapy responsiveness in breast cancer. Oncology (Williston Park, NY). 2009;23:133–42.

37. Reiner GCA, Katzenellenbogen BS. Characterization of Estrogen and Progesterone Receptors and the Dissociated Regulation of Growth and Progesterone Receptor Stimulation by Estrogen in MDA-MB-134 Human Breast Cancer Cells. Cancer Res. 1986;46:1124–31.

38. Verhaegh W, van Brussel A, Moelans C, Alves de Inda M, Gil J, Bikker JW, et al. Heterogeneity in signaling pathway activity within primary breast cancer and between primary and metastases. JCO. 2019;37:589–589.

39. Jeselsohn R, Buchwalter G, De Angelis C, Brown M, Schiff R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. Nat Rev Clin Oncol. 2015;12:573–83.

40. Robinson DR, Wu Y-M, Vats P, Su F, Lonigro RJ, Cao X, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat Genet. 2013;45:1446–51.

41. Massard C, Michiels S, Ferté C, Le Deley M-C, Lacroix L, Hollebecque A, et al. High-Throughput Genomics and Clinical Outcome in Hard-to-Treat Advanced Cancers: Results of the MOSCATO 01 Trial. Cancer Discov. 2017;7:586–95.

42. van de Stolpe A, Holtzer L, Ooijen H van, Inda MA de, Verhaegh W. Enabling precision medicine by unravelling disease pathophysiology: quantifying signal transduction pathway activity across cell and tissue types. Scientific Reports. 2019;9:1603.

43. van de Stolpe A. Quantitative Measurement of Functional Activity of the PI3K Signaling Pathway in Cancer. Cancers. in press.

44. van de Stolpe A, Holtzer L, van Ooijen H, Inda MA, Verhaegh W. Enabling precision medicine by unraveling disease pathophysiology: quantifying signal transduction pathway activity across cell and tissue types. Scientific Reports. in press.

45. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102:15545–50.

46. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.

47. Hartskamp M van, Consoli S, Verhaegh W, Petković M, van de Stolpe A. Use of Artificial Intelligence in Clinical Healthcare Applications. IJMR. in press.

48. Hochstenbach K, van Leeuwen DM, Gmuender H, Gottschalk RW, Løvik M, Granum B, et al. Global gene expression analysis in cord blood reveals gender-specific differences in response to carcinogenic exposure in utero. Cancer Epidemiol Biomarkers Prev. 2012;21:1756–67.

49. Lemmen JG, van den Brink CE, Legler J, van der Saag PT, van der Burg B. Detection of oestrogenic activity of steroids present during mammalian gestation using oestrogen receptor alpha- and oestrogen receptor beta-specific in vitro assays. J Endocrinol. 2002;174:435–46.

50. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009;10:R25.

51. Holman TJ, Wilson MH, Kenobi K, Dryden IL, Hodgman TC, Wood AT, et al. Statistical evaluation of transcriptomic data generated using the Affymetrix one-cycle, two-cycle and IVT-Express RNA labelling protocols with the Arabidopsis ATH1 microarray. Plant Methods. 2010;6:9.

52. Wilson CL, Pepper SD, Hey Y, Miller CJ. Amplification protocols introduce systematic but reproducible errors into gene expression studies. BioTechniques. 2004;36:498–506.

Tables

Table 1: Datasets used in this study.

AffyA: Affymetrix HGU133A; AffyP2: Affymetrix HGU133Plus2.0.

Set	Platform	Use	Set description	References
GSE8597	AffyP2	Calibration of original 25 nM E2 AffyP2 model	MCF7 cell line treated with 25 nM E2 (N=4) or DMSO (N=4) for 24h.	(28)
GSE35428	AffyP2	Calibration of optimized 1 nM E2 AffyP2 model	MCF7 cell line treated with 1 nM E2 (N=10) or ethanol (N=10) for 24h.	(27)
GSE9936	AffyA	Calibration of AffyA 6 nM E2 model	MCF7 cell cultures treated with 6 nM E2 (N=3) or vehicle (N=3) for 24h.	(29)
BrCa cell line calibration	RT-qPCR	Calibration of qPCR TEAM IIa model	MCF7 cell cultures treated with 1 nM E2 (N=4) or vehicle (N=4) for 16h.	This study
BrCa cell line test	AffyP2 and RT-qPCR	Agreement between AffyP2 and TEAMII PCR models	MCF7, T47D, CAMA1 and BT474 cell lines treated with 1 nM E2, 10 nM E2 or DMSO (control) for 16h (N=11 matching samples).	This study (GSE127760)
GSE17700	AffyP2 and AffyA	Agreement between AffyP2 and AffyA models	Biological replicate samples from 16 breast cancer patients processed by both platforms in two different institutions (N=16x2 matching samples).	
AffyP2 Edinburgh (GSE10181)	AffyP2	Correlation with ER activity decrease during AI and agreement with AffyA model	Biopsy samples of 18 patients of the Edinburgh cohort, collected at baseline and after 3 months neoadjuvant treatment with letrozole (N=18x2 correlation with AI; N=6x2 agreement with AffyA model).	(21)
AffyA Edinburgh (GSE20181)	AffyA	Correlation with AI response data and agreement with AffyP2 model	Biopsy samples of 55 patients of the Edinburgh cohort, collected at baseline and after 3 months neoadjuvant treatment with letrozole (N=55x3 correlation with AI; N=6x2 agreement with AffyP2 model).	(14,22)
TEAM IIa	RT-qPCR (TEAMII PCR prototype plate)	Correlation with AI therapy response	49 biopsy and 48 resection samples (28 paired samples) of the TEAM IIa cohort.	This study
NEWEST (GSE48905)	AffyP2	Correlation with fulvestrant response	42 samples from patients treated for 16 weeks with either 500 mg or 250 mg of the selective estrogen receptor degrader fulvestrant	(23)

Table 2: Changes in ER pathway activity scores upon AI treatment.
ER pathway activity score and decrease in activity score for the 55 complete cases of AffyA Edinburgh dataset (GSE20181, 44 responders and 11 non-responders). R: responders, NR: non-responders; av: average, sd: standard deviation. Base-2w: change in ER pathway activity between initiation and two weeks of treatment; Base-3m: change in ER pathway activity between initiation and three months of treatment; 2w-3 m: change in ER pathway activity between two weeks and three months of treatment. T-tests reported as difference (CI), p-value.

ER activity	**status**		**score**			**paired t–test (baseline**
All	**active (%)**	**inactive (%)**	**av**	**sd**	**Range**	**vs time point)**
Baseline	38 (69%)	17 (31%)	46.9	8.9	[32.1, 69.9]	
Two weeks	13 (24%)	42 (75%)	39.0	6.2	[26.9, 53.8]	7.9 (5.7, 10.2), <0.001
Three months	17 (31%)	38 (69%)	39.5	5.3	[27.6, 49.4]	7.5 (5.0, 9.9), <0.001
Stratified by response						**t–test (R vs NR)**
Baseline						
R (n=44)	34 (77%)	10 (23%)	48.2	8.9	[32.1, 69.9]	**6.5 (1.2, 11.7), 0.02**
NR (n=11)	4 (36%)	7 (64%)	41.8	7.0	[33.6, 53.2]	
Two weeks						
R	11 (25%)	33 (75%)	39.0	6.2	[26.9, 53.8]	−0.1 (−3.9, 3.7), 0.95
NR	2 (18%)	9 (82%)	39.1	5.1	[28.1, 46.3]]	
Three months						
R	14 (32%)	30 (68%)	39.6	4.8	[27.6, 48.3]	0.8 (−4.0, 5.7), 0.72
NR	3 (27%)	8 (72%)	38.8	7.0	[28.3, 49.4]	
Decrease in ER activity	**direction**		**decrease in score**			
All	**decrease**	**increase**	**av**	**sd**	**Range**	
Base − 2 w	47 (85%)	8 (15%)	7.9	8.3	[−9.1, 25.0]	
Base − 3 m	46 (84%)	9 (16%)	7.5	8.9	[−15.0, 28.0]	
2 w − 3 m	22 (40%)	33 (60%)	−0.47	4.9	[−12.0, 8.7]	
Stratified by response			**av**	**Sd**	**Range**	**t–test (R vs NR)**
Base − 2 w						
R	39 (89%)	5 (11%)	9.3	8.0	[−7.0, 25.0]	**6.6 (1.3, 11.8), 0.02**
NR	8 (73%)	3 (27%)	2.7	7.2	[−9.1, 13.0]	
Base − 3 m						
R	38 (86%)	6 (14%)	8.6	8.1	[−9.7, 28.0]	5.6 (−1.8, 13), 0.13
NR	6 (55%)	5 (45%)	2.9	11.0	[−15.0, 21.0]	
2 w − 3 m						
R	17 (39%)	27 (61%)	−0.65	4.8	[−12.0, 8.7]	-0.91 [-4.9, 3.0], 0.6
NR	5 (45%)	6 (55%)	0.26	5.6	[−7.6, 8.7]	

Table 3: ER pathway score at baseline and tumor resection for the TEAM IIa cohort.

Measurements are from the TEAMII PCR platform. CR: complete remission, PR: partial remission, SD: stable disease, PD: progressive disease; av: average, sd: standard deviation. N: number of samples available for pathway analysis. 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.

	Sample availability			Baseline ER activity score				Decrease in ER activity score				Resection ER activity score			
	baseline only	paired	resection only	N	av	sd	range	N	av	sd	Range	N	av	sd	range
All	21	28	20	49	42.2	11.6	[9.9, 62.2]	28	9.1	12.8	[−11.2, 31.5]	48	33.9	14.0	[5.6, 59.8]
	Stratified by response assessed by palpation at last measurement														
All	15	23	17	38	43.3	10.9	[18.9, 62.2]	23	9.0	13.3	[−11.2, 31.5]	40	32.6	13.5	[5.6, 59.8]
CR	5	6	3	11	47.6	7.8	[38.9, 62.2]	6	20.0	9.0	[7.1, 27.7]	9	28.7	10.1	[13.4, 46.1]
PR	5	7	8	12	41.8	13.0	[18.9, 56.3]	7	6.4	11.7	[−9.2, 25.4]	15	31.8	16.5	[5.6, 59.8]
SD	4	8	5	12	44.5	9.1	[25.8, 56.0]	8	5.4	15.2	[−11.2, 31.5]	13	38.9	10.2	[20.3, 54.4]
PD	1	2	1	3	28.4	6.5	[23.7, 35.7]	2	−0.16	0.28	[0.35, 0.04]	3	20.7	7.3	[12.4, 26.1]

Figure Legends

Figure 1: Schematic representation of ER pathway model network.

Left: Schematic of Affymetrix HGU133A (AffyA) model network, only genes and probesets present in the AffyA platform were kept. Center: Node type description. Right: Schematic of TEAMII PCR model, RT-qPCR assays were developed for a subset of target genes of the AffyP2 model.

Figure 2: ER pathway activity scores significantly decreased following aromatase inhibitor treatment.

Scores on AffyP2 Edinburgh dataset (GSE10281, n=18) were calculated using the 1nM E2 AffyP2 model. (A) individual patients' line plots of ER activity at baseline and at three months (mean decrease = 13.2, CI=9.7-16.8), *** indicates paired t-test p-value < 0.001). (B) ER activity at baseline versus decrease in activity after three months of therapy (corr=0.88, CI=0.71-0.96, p<0.001).

Figure 3: ER pathway activity scores are significantly lower at baseline and two weeks in non-responding than responding tumors in the AffyA Edinburgh (GSE20181) dataset.

ER pathway activity was measured with the AffyA model. (A) Individual patients' line plot of ER pathway activity at baseline and after two weeks and three months of treatment. Solid lines: activity went down during treatment; dashed lines: activity went up during treatment. (B) ER activity score as a function of treatment time for responders (left panel, blue) and non-responders (right panel, red). C/D: Relationship between ER activity at baseline and decrease in ER activity (C) at two weeks (corr=0.74, CI=0.6-0.84, p-value<0.001) and (D) at three months (corr=0.83, CI=0.72-0.89, p-value<0.001) of letrozole treatment (n=55). (E) Box plots of ER pathway activity in responders and non-responders at baseline. (F) Decrease in ER pathway activity score, between baseline and two weeks (left) and between baseline and three months (right) of treatment. Resp/blue: responders; Non-Resp/red: non-responders. ***: t-test p-value < 0.001; *: t-test p-value < 0.05.

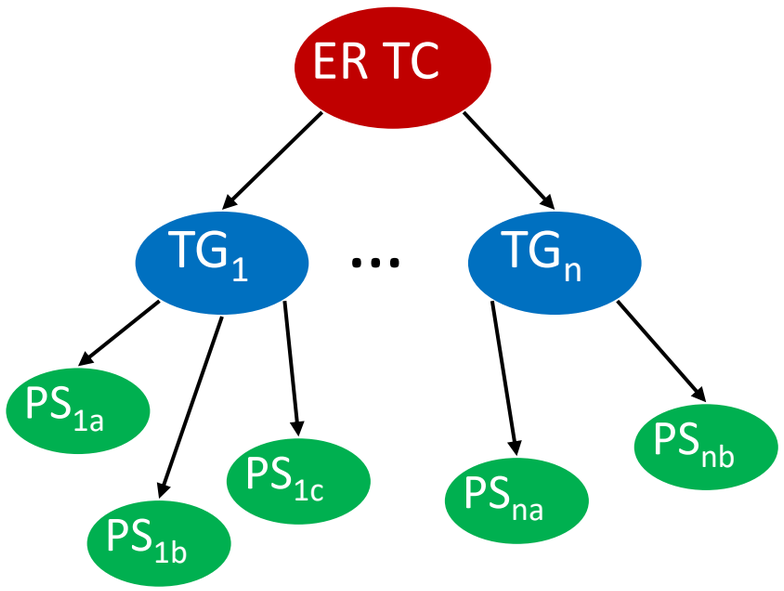
Figure 4: In patients with progressive disease ER pathway activity scores are significantly lower and do not change after treatment.

(A) ER activity score at baseline in patients of the TEAM-IIA cohort measured with the TEAMII qPCR model, (B) ER activity score at baseline in patients of the NEWEST cohort measured with the AffyP2 model, (C) Team IIA cohort, individual ER pathway activity score as function of treatment time. (D) Team IIA cohort, decrease in pathway activity score from baseline to tumor resection, stratified by response assessed by palpation at last measurement. CR, blue: complete remission; PR, green: partial remission; SD, orange: stable disease; PD, red: progressive disease. Solid lines:

activity went down during treatment; dashed lines: activity went up during treatment. **: t-test p-value < 0.01; *: t-test p-value < 0.05.

Figure 1

AffyA model



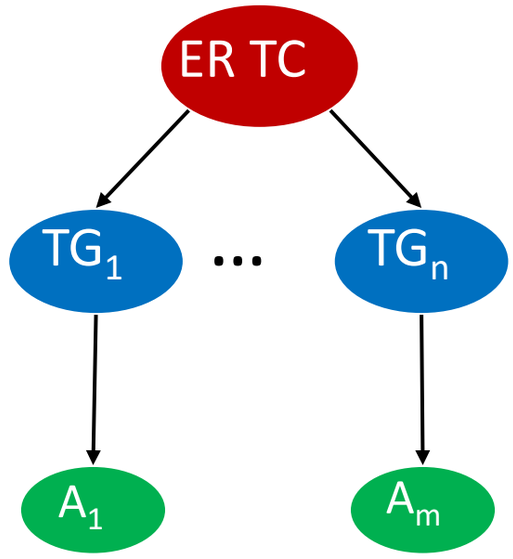
Expression levels measured by Affymetrix HGU133A probesets

(i) ER transcription complex activation node

(ii) Target gene (down/up) regulation node

(iii) Expression intensity (low/high) level nodes

PCR model



Expression levels measured by RT-qPCR assays

Figure 2

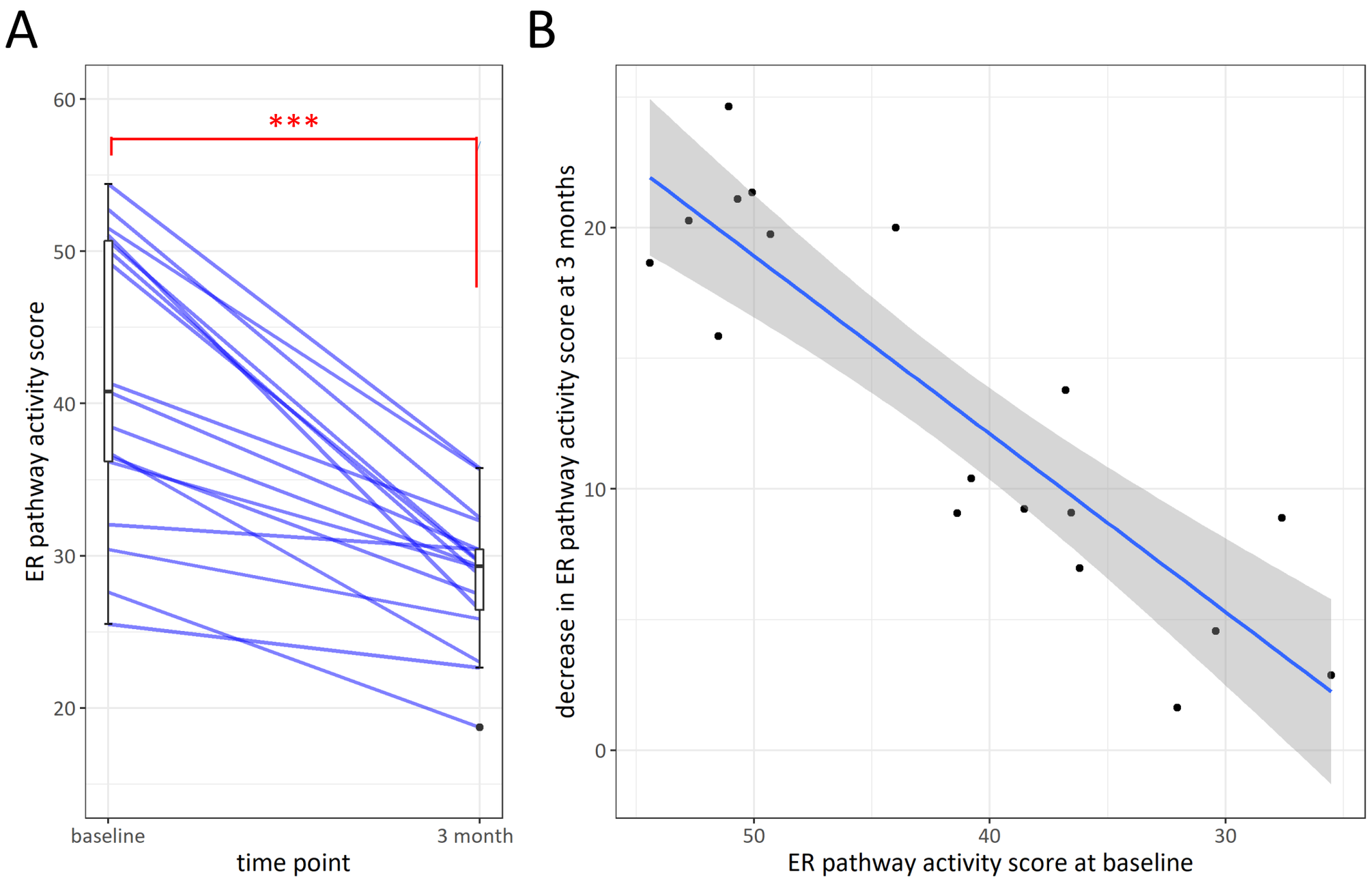

A
B
***
60
50
40
30
20
ER pathway activity score
baseline
3 month
time point
20
10
0
decrease in ER pathway activity score at 3 months
50
40
30
ER pathway activity score at baseline

Figure 3

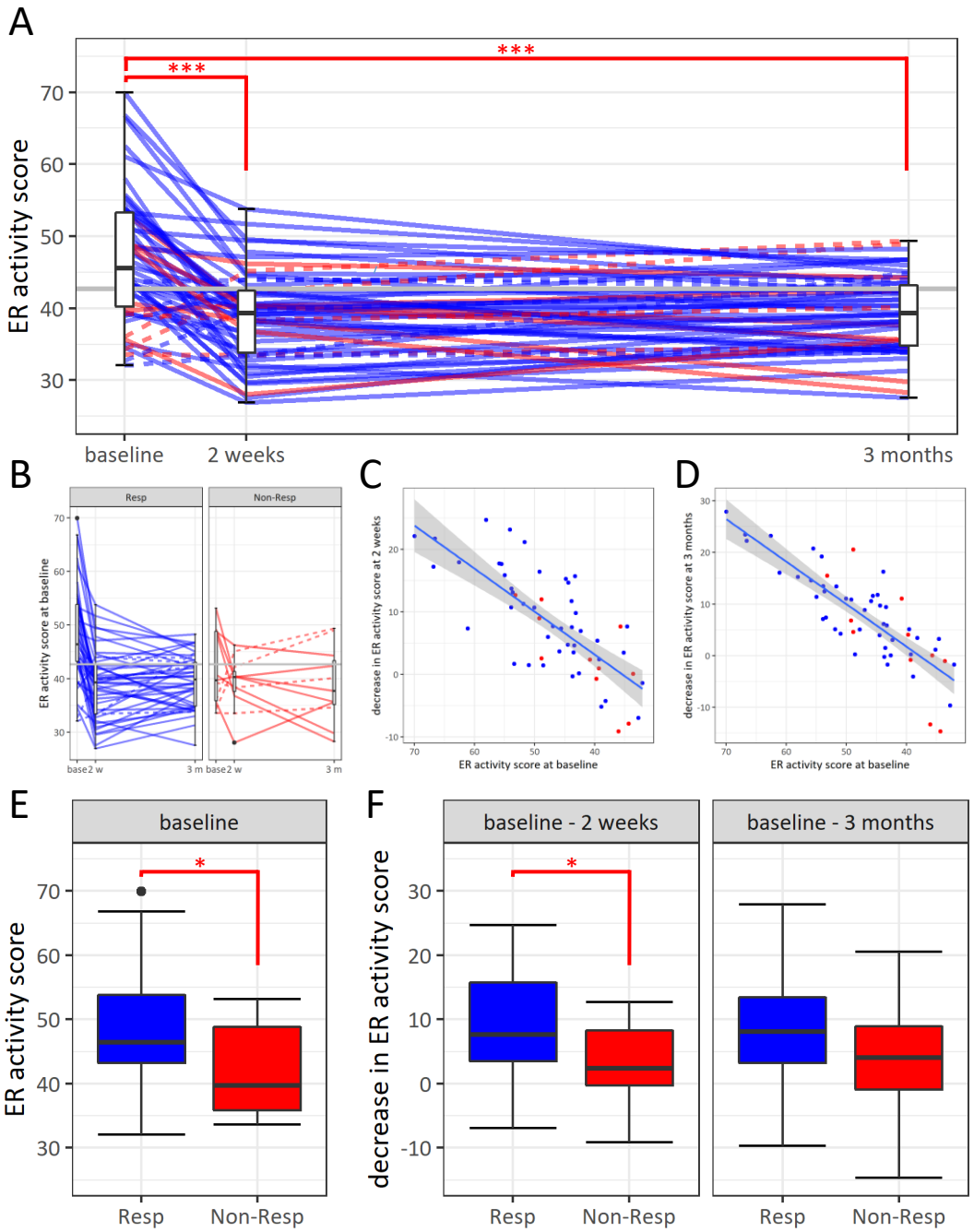
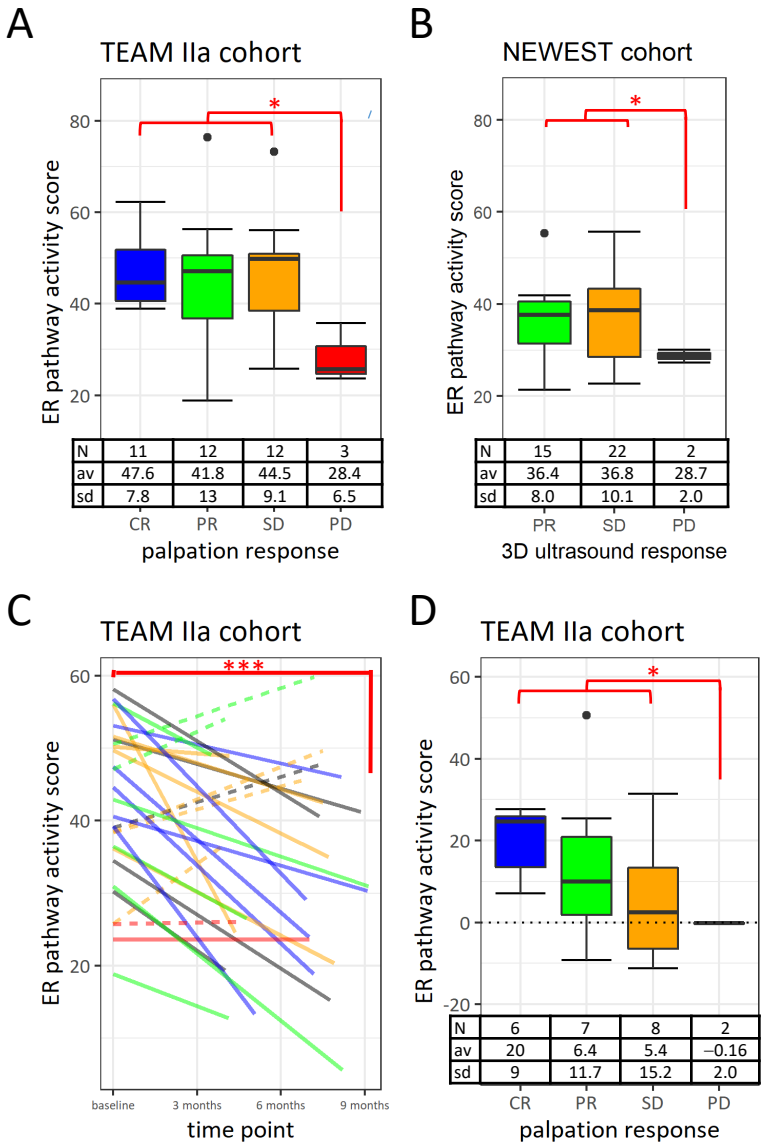

A
***
***
ER activity score
70
60
50
40
30
baseline
2 weeks
3 months
B
Resp
Non-Resp
ER activity score at baseline
base
2 w
3 m
C
decrease in ER activity score at 2 weeks
ER activity score at baseline
D
decrease in ER activity score at 3 months
ER activity score at baseline
E
baseline
*
ER activity score
Resp
Non-Resp
F
baseline - 2 weeks
baseline - 3 months
*
decrease in ER activity score
Resp
Non-Resp
Resp
Non-Resp

Figure 4


A
TEAM IIa cohort
ER pathway activity score
palpation response
CR PR SD PD
N 11 12 12 3
av 47.6 41.8 44.5 28.4
sd 7.8 13 9.1 6.5
B
NEWEST cohort
ER pathway activity score
3D ultrasound response
PR SD PD
N 15 22 2
av 36.4 36.8 28.7
sd 8.0 10.1 2.0
C
TEAM IIa cohort
***
ER pathway activity score
baseline 3 months 6 months 9 months
time point
D
TEAM IIa cohort
ER pathway activity score
palpation response
CR PR SD PD
N 6 7 8 2
av 20 6.4 5.4 −0.16
sd 9 11.7 15.2 2.0