ARTICLE  OPEN

SETER/PR: a robust 18-gene predictor for sensitivity to endocrine therapy for metastatic breast cancer

Bruno V. Sinn1,2, Chunxiao Fu1, Rosanna Lau1, Jennifer Litton3, Tsung-Heng Tsai1, Rashmi Murthy1, Alda Tam4, Eleni Andreopoulou5,5, Yun Gong1, Ravi Murthy4, Rebekah Gould1, Ya Zhang1, Tari A. King6, Agnes Viale7, Victor Andrade7,8, Dilip Giri7, Roberto Salgado9,10, Ioanna Laios11, Christos Sotiriou12, Esmeralda C. Marginean13, Danielle N. Kwiatkowski3, Rachel M. Layman3, Daniel Booser3, Christos Hatzis14, V. Vicente Valero3 and W. Fraser Symmans1

There is a clinical need to predict sensitivity of metastatic hormone receptor-positive and HER2-negative (HR+/HER2−) breast cancer to endocrine therapy, and targeted RNA sequencing (RNA-seq) offers diagnostic potential to measure both transcriptional activity and functional mutation. We developed the SETER/PR index to measure gene expression microarray probe sets that were correlated with hormone receptors (ESR1 and PGR) and robust to preanalytical and analytical influences. We tested SETER/PR index in biopsies of metastatic HR+/HER2− breast cancer against the treatment outcomes in 140 patients. Then we customized the SETER/PR assay to measure 18 informative, 10 reference transcripts, and sequence the ligand-binding domain (LBD) of ESR1 using droplet-based targeted RNA-seq, and tested that in residual RNA from 53 patients. Higher SETER/PR index in metastatic samples predicted longer PFS and OS when patients received endocrine therapy as next treatment, even after adjustment for clinical-pathologic risk factors (PFS: HR 0.534, 95% CI 0.299 to 0.955, p = 0.035; OS: HR 0.315, 95% CI 0.157 to 0.631, p = 0.001). Mutated ESR1 LBD was detected in 8/53 (15%) of metastases, involving 1–98% of ESR1 transcripts (all had high SETER/PR index). A signature based on probe sets with good preanalytical and analytical performance facilitated our customization of an accurate targeted RNA-seq assay to measure both phenotype and genotype of ER-related transcription. Elevated SETER/PR was associated with prolonged sensitivity to endocrine therapy in patients with metastatic HR+/HER2− breast cancer, especially in the absence of mutated ESR1 transcript.

npj Breast Cancer (2019) 5:16 ; https://doi.org/10.1038/s41523-019-0111-0

INTRODUCTION

Endocrine therapy is the principal treatment for metastatic HR+/HER2− breast cancer until resistance becomes clinically manifest.1,2 Molecular progression from reliance on estrogen is generally accepted as the basis of acquired resistance, and this can sometimes be identified as reduced hormone receptor expression (ER and PR loss in approximately 10% and 20%, respectively, at first metastatic relapse),3,4 upregulation of alternative growth pathways, acquisition of constitutively activating gene mutations in the ligand-binding domain (LBD) sequence of ESR1,5,6 or acquisition of other aberrations that accelerate growth and promote survival. Notably, the onset, rate and mechanisms of molecular progression vary for each patient.

Clinically, endocrine treatment resistance is recognized from short disease-free interval in the adjuvant or metastatic setting of endocrine treatment, development of visceral disease, or loss of ER or PR in metastatic breast cancer. However, these criteria are inexact. A quantitative biomarker of sensitivity to endocrine therapy (SET) in metastatic cancer might potentially contribute clinically useful information to address a clinical conundrum: whether to continue with endocrine therapy,2 combine this with another targeted therapy, or switch to chemotherapy-based treatment. Furthermore, it might inform a secondary concern: when in the course of therapies for metastatic breast cancer it might be optimal to add a cdk4/6 or PI3kinase/mTOR inhibitor to endocrine therapy. For example, it is still unclear whether addition of currently approved targeted agents to endocrine therapy in advanced disease improves progression-free survival (PFS) by reversing endocrine resistance or augmenting partial endocrine sensitivity.

Based on our previous development of a signature of ESR1-related transcripts in early breast cancer,7,8 we hypothesized that a combination of genes with expression related to both estrogen and progesterone receptors (gene symbols ESR1 and PGR), but not proliferation, might predict sensitivity to endocrine therapy in metastatic breast cancer.9 We also considered preanalytical and analytical effects on measurement of gene expression in our approach to select transcripts for our signature, in order to develop a technically robust signature of a few genes that we could then translate to a customized assay with strong analytical...
validity. In addition, we considered that elevated hormone receptor-related transcription might represent natural activity (and indicate sensitivity to endocrine therapy) or perversely result from constitutive activating mutation of ESR1 transcripts (already implicated in resistance to aromatase inhibitors12). Overall, we felt that the current evidence for altered biology of progressive breast cancer after relapse requires a more specialized approach to risk stratification than adoption of multi-gene assays that were developed for the earliest stages of hormone receptor-positive breast cancer.13-15 Hence, we aimed to combine both genotypic and phenotypic information, using a customized RNA sequencing (RNAseq) assay to measure sensitivity to endocrine therapy (SET).

RESULTS
Definition of the SETER/PR index
Eighteen informative transcripts (correlated with both ESR1 and PGR and without obvious association with proliferation) and ten reference transcripts were selected for inclusion in the SETER/PR index (Fig. 1, Supplementary Table 2). The reference genes were selected based on minimal variability and high reproducibility across 331 hormone receptor-positive, HER2-negative samples of the training set (Supplementary Fig. 1). SETER/PR was defined as:

$$\text{SETER/PR} = \frac{\sum_{i=1}^{18} T_i - \sum_{i=1}^{10} R_i}{10} + 2,$$

where $T_i$ is the expression of the $i$th of the 18 informative genes and $R_i$ the expression of the $i$th of the ten reference genes. The distribution of SETER/PR index scores was scaled to be above zero for most HR+/HER2—cancers and below zero for HR—cancers. Negative score values are assigned zero value to avoid confusion and variance from low expression of the target genes. We used the median value of SETER/PR in the clinically annotated dataset as a cut-off value to assign patients to groups with high vs. low SETER/PR.

Performance under preanalytical and analytical conditions used for development
SETER/PR was robust to technical replication (ICC = 0.990), intratumoral sampling (ICC = 0.953), type of cancer sample (cytology vs. tissue, $p = 0.952$), and type of microarray platform (U133A vs. Plus2.0 arrays, $p = 0.990$). Score values obtained from Plus2.0 arrays had a slight bias towards higher values when compared to U133A microarrays (Supplementary Fig. 2).

Performance under independent preanalytical and analytical conditions
Supplementary Fig. 3 demonstrates the performance of SETER/PR in preanalytical and analytical validation studies that were not previously used in the feature selection process. The cross-platform reproducibility was validated in an independent dataset of 32 cases profiled on both U133A and Plus2.0 microarrays with $\rho = 0.994$ for the corrected score and $\rho = 0.995$ for inter-laboratory reproducibility. The technical reproducibility of the assay on U133A microarrays was validated in an independent dataset of 63 data pairs ($\rho = 0.994$). SETER/PR was stable over relevant ranges of contamination with liver or normal breast tissue with negative score values regressing more rapidly to the baseline levels from normal liver or normal breast tissues. Categories of high vs. low SETER/PR index (relative to median of 0.82) were consistent ($κ = 0.881$ and 0.905, respectively) over a range of 0–90% RNA added from normal liver or breast. There was no statistically significant effect of time delay (ex vivo ischemic time) and sample preservation method (RNAlater versus snap frozen) on SETER/PR measurements (Supplementary Table 3).

Prognostic performance in metastatic breast cancer
The characteristics of 140 patients with hormone receptor-positive, HER2-negative metastatic breast cancers are summarized in Table 1. The observed range of SETER/PR was comparable in samples from different sites of metastasis (Supplementary Fig. 4). SETER/PR was positively associated with PR immunohistochemical staining and with negative score values regressing more rapidly to the baseline levels from normal liver or normal breast tissues. Categories of high vs. low SETER/PR index were significantly associated with the number of prior progression events ($p = 0.009$).

The continuous SETER/PR index was prognostic for PFS and OS in patients receiving endocrine-based therapy (PFS: hazard ratio (HR) 0.51 (0.41–0.74), $p < 0.001$; OS: 0.40 (0.26–0.62), but not in patients receiving chemotherapy (PFS: HR 0.76 (0.45–1.27), $p = 0.290$). We selected the median value (0.82) as threshold to dichotomize SETER/PR index. Dichotomized SETER/PR was independently prognostic for PFS (Table 2) and OS (Table 3) in univariate and multivariate analyses with standard clinical-pathologic risk factors. We further analyzed the survival of patients whose biopsy was obtained at a time of recurrence (after prior systemic therapy) and whose next treatment included endocrine therapy. In patients who had previously demonstrated clinical evidence of sensitivity to endocrine therapy, the continuous SETER/PR index was independently prognostic for PFS in a multivariate model that included PR immunohistochemistry status of the metastasis, the number of prior relapse events, and the presence or absence of any visceral metastasis (Tables 2 and 3). Figure 2 shows Kaplan–Meier plots using the dichotomized SETER/PR index in the same cohort of patients. SETER/PR was significantly associated with patient outcome over a wide range of different possible cut-points (Supplementary Fig. 4D).

In addition to the multivariate analyses using standard clinical and pathological tumor characteristics, we evaluated if AURKA as marker of proliferation might add prognostic information. As illustrated in Supplementary Table 4, AURKA is prognostic for both PFS and OS in patients who received chemotherapy as next treatment, independent of SETER/PR, and also after adjustment for
Table 1. Patient characteristics

| Characteristic                        | Median | Range |
|--------------------------------------|--------|-------|
| Stage at initial diagnosis           |        |       |
| Stage IV                            | 45     | 32    |
| Stage I–III                         | 95     | 68    |
| Visceral metastases                 |        |       |
| Yes                                 | 80     | 57    |
| No                                  | 60     | 43    |
| Progesterone Receptor Status         |        |       |
| Positive                            | 80     | 57    |
| Negative                            | 60     | 43    |
| Prior sensitivity                   |        |       |
| Sensitive                           | 70     | 50    |
| Resistant                           | 39     | 28    |
| No prior endocrine therapy          | 31     | 22    |
| Number of events biopsed            |        |       |
| Initial diagnosis                   | 20     | 14    |
| 1st                                 | 42     | 30    |
| 2nd                                 | 26     | 19    |
| 3rd                                 | 14     | 10    |
| 4th or more                         | 38     | 27    |
| Treatment                           |        |       |
| Endocrine treatment                 | 97     | 69    |
| Chemotherapy                        | 33     | 24    |
| Other                               | 8      | 6     |
| Radiotherapy alone                  | 2      | 1     |
| Median                              |        |       |
| Range                               |        |       |
| Age                                 | 55     | 32–82 |
| Progression-free survival           |        |       |
| Months                              | 5.53   | 0.16–74 |
| Overall survival                    |        |       |
| Months                              | 24     | 0.16–126 |

Characteristics of the 140 patients with stage IV breast cancer

Clinical and pathological characteristics. If patients received endocrine therapy as next treatment, expression of AURKA did not add prognostic information when SETER/PR was included in bivariate and multivariate models, while SETER/PR retained its significance.

Customization of the SETER/PR assay using targeted RNA sequencing (RNAseq)
The customized RNAseq assay integrates measurements of ER and PR-related transcriptional activity (SETER/PR index) and the proportion of ESR1 transcript reads with activating LBD mutation. SETER/PR index was calibrated between microarray and customized RNAseq assays in 40 breast cancer samples analyzed in duplicate with both assays (Supplementary Fig. 5). There was excellent interassay agreement ($\rho = 0.965$ and $\kappa = 0.823$) in an independent test of 23 breast cancer samples.

Proportion of ESR1 transcript reads with LBD mutation related to the SETER/PR index
The customized RNAseq assay detected mutations in the LBD of ESR1 in 8/53 samples, with an average of 33,000-fold coverage depth. Metastases with an ESR1 mutation had high SETER/PR Index (Fig. 3). We confirmed that the customized RNAseq assay for SETER/PR index achieved a similar prognostic separation (Fig. 3) to the original microarray assay (Fig. 2) in patients treated with endocrine therapy. An exploratory analysis suggested that the prognosis among patients with an ESR1 LBD mutation (and consequently higher SETER/PR index) may be intermediate between those with low SETER/PR index and high SETER/PR index with wild-type ESR1 (Fig. 3).

DISCUSSION
SETER/PR index is an unbiased calculation based on the straightforward concept of measuring transcription associated with ESR1 and PGR expression, which avoids over-fitting from modeling on outcome data. The assay was robust to critical preanalytical conditions (tissue and cytologic samples, ex vivo ischemia, preservation or fixation of tissue samples, and intratumoral spatial heterogeneity) and analytical conditions (technical reproducibility at all levels of the assay procedure, different technical platforms for the assay). We also describe how it was customized into an assay that also integrates measurement of mutated ESR1 transcripts.

To our knowledge, SETER/PR is the first multigene expression assay to be developed specifically for metastatic breast cancer. Higher SETER/PR index was associated with longer PFS and OS for patients treated by endocrine therapy, particularly for those who had previously demonstrated clinical sensitivity to hormonal therapy. Although we observed that SETER/PR was not associated with outcome in patients treated with chemotherapy, that cohort was too small to be able to make any conclusion. Additionally, the observation might be confounded because chemotherapy is usually offered when there is already clinical evidence for endocrine resistance. We also note that high expression of
SLC39A6 is observed in the SET\textsubscript{ER/PR} index (Supplementary Fig. 1, Supplementary Table 2). This transcript encodes LIV-1, the membrane target for the antibiotic-drug conjugate SGN-LIV1.16

ESR1 mutations occur within the LBD sequence, and are rare in primary cancer. They commonly occur in relapsed metastatic disease, and are possibly more frequent after treatment with aromatase inhibitors.17,18 These mutations induce constitutive receptor activity and have been identified as a mechanism of resistance to estrogen-depriving therapies, while patients might still benefit from selective estrogen receptor degradation (SERD) treatment, for example fulvestrant. In the FERGI and PALOMA-3 trials, ESR1 mutations had no effect on PFS in patients receiving chemotherapy and those that received endocrine treatment. Uni- and multivariate analyses are shown for the clinically relevant subgroups of patients that received endocrine treatment and presented with relapsed stage IV disease and the subset of patients with a prior history of endocrine sensitivity.

| SET\textsubscript{ER/PR} for prediction of overall survival | HR | 95 % CI | \( \text{p} \) |
|----------------------------------------------------------|---|--------|---|
| Chemotherapy \( (N = 33) \)                                    | 0.813 | 0.318–2.077 | 0.666 |
| Endocrine treatment \( (N = 97) \)                          | 0.391 | 0.239–0.638 | <0.001 |
| PR status                                                  | 0.524 | 0.267–1.029 | 0.061 |
| Visc. met.                                                  | 1.808 | 0.945–3.460 | 0.074 |
| Event \( >2 \)                                              | 4.463 | 1.943–10.25 | <0.001 |
| Prior Sens.                                                 | 0.331 | 0.156–0.700 | 0.04 |
| Endocrine treatment and relapsed stage IV \( (N = 46) \)    | 0.316 | 0.154–0.649 | 0.002 |
| PR status                                                  | 0.275 | 0.119–0.637 | 0.003 |
| Visc. met.                                                  | 0.433 | 0.189–0.995 | 0.049 |
| Event \( >2 \)                                              | 5.222 | 2.082–13.10 | <0.001 |

Cox regression analyses for prediction of overall survival using the dichotomized SET\textsubscript{ER/PR}. Results are shown for patients that received chemotherapy and those that received endocrine treatment. Uni- and multivariate analyses are shown for the clinically relevant subgroups of patients that received endocrine treatment and presented with relapsed stage IV disease and the subset of patients with a prior history of endocrine sensitivity.

HR hazard ratio, CI confidence interval

Metastatic breast cancer is a dynamic disease, prone to heterogeneity and evolving over time and under the selective pressure of different treatments.24,25 At this time the AURORA initiatives are aiming to characterize the molecular progression of metastatic breast cancer based on next-generation sequencing using serial biopsies taken over the course of the disease.23 This might lead to further insight into molecular evolution. Indeed, we don’t know yet whether the SET\textsubscript{ER/PR} index would change during successive progression events or in response to different classes of treatments.

Treatment of stage IV HR-/HER2- breast cancer typically relies on available endocrine treatments9,26 until more rapidly progressive disease favors a switch to chemotherapy.22,23 However, this treatment strategy increasingly requires nuanced clinical judgment, as the selection of treatment options continues to expand to include additional endocrine agents, alone or combined with targeted molecular agents, chemotherapy, and other molecularly targeted approaches. So an index of tumoral sensitivity to endocrine therapy might become a clinically useful metric, alone or in combination with proven biomarkers to select among the other treatment alternatives.27 In this context, the SET\textsubscript{ER/PR} index might inform the selection of next treatment: switch endocrine therapy, augment endocrine therapy with a targeted molecular therapy (such as mTOR, PI3K, or cdk4/6 inhibition), include an SERD agent to target emergent mutated ESR1 clone, or switch to a different treatment strategy (such as chemotherapy, immune therapy). Of course, any definitive statement on such clinical utility would require testing the SET\textsubscript{ER/PR} index using samples from randomized trials and goes beyond the scope of this first description of the assay. But even within those trials, we might gain insight as to whether the addition of different targeted therapies might augment sensitivity, or reverse resistance to endocrine therapy—questions that are difficult to answer without a biomarker for endocrine sensitivity.
There are several important caveats to the interpretation and generalizability of our results. Despite an overall sample size of 140 prospective biopsies of relapsed metastatic disease, the clinical and treatment subsets are small, requiring cautious interpretation of these results. This is a limitation of the combined analysis of SETER/PR index and percent mutated ESR1 transcripts.

Another limitation is the lack of an independent clinically annotated cohort to validate the findings that would also allow the definition and validation of an optimized cut-point for patient stratification.

Overall, this manuscript introduces a novel approach to assay development and this assay appears to be analytically valid. The promising clinical performance is still exploratory, and further independent clinical validation studies of the assay and its cut-point will still be required.

**METHODS**

All patients gave informed written consent to take part in the study and for the use of tissue material for research purposes. Protocols were approved by the MD Anderson Institutional Review Board (IRB). The microarray and accompanying data are available on NCBI GEO and summarized under a figshare metadata record.

**Discovery cohort**

The discovery cohort of Affymetrix U133A microarrays (N = 389) from invasive hormone receptor-positive breast cancers included 242 cases from our published dataset and 147 additional samples (GSE129551), all derived from fresh tissue or FNA biopsy samples obtained prior to any systemic therapy and stored frozen at −80 °C in RNAlater (approved IRB protocols LAB99-402, LAB04-0093). Receptor status, tumor stage and type of tumor samples are described in Supplementary Table 1.

ER- and PR-positivity was defined as nuclear immunostaining in ≥10% of tumor cells. Antibody clones 6F11, dilution 1:35, and PGR1294, dilution 1:200, were used on a Leica Bond-Max instrument according to standard procedures. HER2-positivity was defined as immunohistochemistry score of 3+ membrane staining and/or gene amplification (HER2/CEP17 ratio >2.2) by fluorescence in situ hybridization.

Gene expression profiling for target and reference transcripts RNA was extracted, processed and hybridized to Affymetrix human genome U133A microarrays (U133A GeneChip, Affymetrix, Santa Clara, CA, USA) as described previously. In brief, the raw intensity files were
processed using the MA5.0 algorithm to generate probe set-level intensities, normalized to a median array intensity of 600, log2-transformed and scaled using the expression of 1322 breast cancer reference genes within each sample. Target probe sets for gene transcripts in the 389 cases of the discovery cohort were identified based on Spearman’s rank correlation coefficient for coexpression with ESR1 and PGR (probe sets 205225_at and 208305_at) in hormone receptor-positive breast cancer samples. Reference probe sets were selected based on consistency and range of expression values. This manuscript follows REMARK guidelines.

Studies of preanalytical and analytical robustness
We conducted a series of studies to evaluate the reproducibility of gene expression measurements in breast cancer samples according to replication of technical, intratumoral, interplatform, and inter-sample type conditions (IRB protocols LAB08-0823, LAB08-0824). These included 6 technical (analytical) replicates from 20 breast cancers (GSE129558), 3 tumor samples from each of 51 breast cancers (GSE129557), inter-sample type comparisons of 116 matched cytology and tissue samples (GSE129559) that were collected from multiple institutions, and interplatform comparisons of Affymetrix U133A and Plus2 array platforms from 88 breast cancers (GSE129556). Figure 1 provides an overview of how these studies were used to select the probe sets for the final gene signature. We tested the robustness of the final SET\textsubscript{ER\textsubscript{PR}}/PR gene expression index in other studies: 11 breast cancers contaminated with increasing known amounts of liver RNA (GSE33116); 10 other breast cancers diluted with increasing known amounts of normal breast RNA (GSE124648); 17 other breast cancers with increasing duration of ischemic delay at room temperature, testing two sample preservation methods (GSE25011); matched U133A and Plus2 arrays in two different laboratories (MDACC and JBI; GSE17700); and technical replicates using U133A arrays in another 63 breast cancers from MDACC (GSE129560).

Development of customized RNAseq assay
We employed a digital PCR-based RNAseq strategy with three steps: (1) droplet-generation using RainDance Source system (BioRad, Hercules, CA) and one-step RT-PCR reaction (first PCR) to target the regions of interest with our custom multiplex primer set; (2) second PCR to incorporate RainDance DirecSeq primers for sample indexing and Illumina specific adapters for cluster generation/sequencing; (3) library quantification, QC, and Illumina MiSeq sequencing (Illumina, San Diego, CA). We perform pooled sequencing of up to 40 sequence libraries per flow cell. The read
count of each targeted sequence was log2 transformed, and the sequence reads of the LBD of the ESR1 transcript were analyzed for single nucleotide variants and reported as the percent of ESR1 reads and type of mutation.

Clinical cohort with stage IV breast cancer

Patients with metastatic HR+ breast cancer were offered participation in a prospective research protocol to obtain a research sample at the time of their clinical biopsy of metastasis at MD Anderson (protocol LAB04-0093) between 2004 and 2013, obtained as fine-needle aspiration (FNA) or core biopsy (CBX). Their next treatment was recorded and was at the discretion of their oncologist. A total of 234 samples were profiled using Affymetrix U133A gene expression microarrays, 212 microarrays passed our quality control analysis. We excluded 32 HER2-positive and 26 hormone receptor-negative cases based on immunohistochemistry and (where appropriate) HER2 in situ hybridization testing of the metastatic samples. Fourteen additional cases were excluded for other reasons (no follow-up data after biopsy, diagnosis other than breast cancer), resulting in 140 eligible cases with quality microarray data in this study (GSE124647). Median PFS and OS were 5.5 and 24.0 months, respectively (Table 1). PR positivity was defined as ≥10% nuclear immunostaining. Proliferation (Ki-67 immunohistochemistry) is not usually assessed in metastatic samples, so we evaluated Aurora kinase-A (AURKA; probe set 208079_s_at) as a reliable genomic marker for proliferation in multivariate survival analyses. Statistical analysis showed variable sensitivity was defined as a history of at least 6 months of freedom from progression while on endocrine therapy for metastatic disease or 5 years adjuvant endocrine therapy for primary breast cancer without recurrence. A subset of 53 cases was available for analysis of ESR1 gene mutations by RNAseq.

Statistical methods

Pearson’s correlation coefficient (ρ) was used to compare cross-platform and cross-tissue reproducibility of each candidate probe set on the array. The intraclass correlation coefficient (ICC) was used to evaluate intra-assay and intratumoral reproducibility. A linear mixed-effects model (LME) with random within-group intercept was used to estimate the effect of sample preservation method (RNAlater vs. fresh frozen) and time delay (0 vs. 40 min) using the r package lme4. The effect of sample stabilization delay (cold ischemic time) was assessed using a similar model with fixed slope (for the cold ischemic time effect) and random intercept (for biological variation among tumors). The statistical significance of the coefficients was evaluated by comparing the likelihood ratio test to compare the full model with a reduced model that did not include the term of interest. To examine the impact of contamination with normal breast tissue and liver tissue, SETERRPR values were plotted against the percentage of contaminant. Fleiss’ κ statistic for multiple raters was used to evaluate the reproducibility of risk class assignment. We used the R package survival for survival analyses. PFS was defined as the time from the start of new treatment after the biopsy of relapsed disease, until disease progression or death from any cause. The endpoint definition for overall survival was death from any cause. We used Cox regression to model relationship between the continuous SETERRPR and survival outcomes. The Kaplan—Meier method and log-rank test were used to evaluate survival outcomes using the dichotomized score. All statistical analyses and computations were performed in R v. 3.1.2 and Bioconductor.15

Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

DATA AVAILABILITY

The data generated and analyzed during this study are described in the following metadata record: https://doi.org/10.6084/m9.figshare.7998809.26 Datasets are available on NCBI Gene Expression Omnibus (GEO) summarized under SuperSeries GSE124648.

ACKNOWLEDGEMENTS

This work was supported in part by The Breast Cancer Research Foundation (W.F.S.) and in part by Federal Funds from the National Cancer Institute, National Institutes of Health, under contract no. HHSN261200800001E, grant no. R21CA18156 (W.F.S.) and no. 1U01CA215547 (W.F.S. and C.H.). B.V.S. was supported by funds from the German Research Foundation (DFG), grant no. SI 1919/1-1 and by the BIH Charité Clinician Scientist Program funded by the Charité—Universitätsmedizin Berlin and the Berlin Institute of Health. The Affymetrix GeneChip DX2 instruments were provided to the research laboratory as a gift from the Stading Family. The RainDance Source instrument was provided to the research laboratory as a gift from the Toomim Family Fund (J.L.). We thank Dr. Michael Samuels for technical advice during assay development.

AUTHOR CONTRIBUTIONS

B.V.S., W.F.S. and C.H. conceived and planned the experiments and T.-H.T., J.L. and C.S. were involved in the study design. C.F., R.L., A.V., V.A., R.S., I.L. conducted experiments. B.V.S., C.H. and T.-H.T. analyzed the data. J.L., R.M.L., Rashmi M., A.T., E.A., Y.G., Ravi M., R.G., Y.Z., T.A.K., V.A., D.G., R.S., C.S., E.C.M., D.N.K., R.L., D.B., V.V. and W.F.S. provided material. B.V.A. and W.F.S. wrote the manuscript. All authors interpreted the results, edited and approved the manuscript.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the npj Breast Cancer website (https://doi.org/10.1038/s41523-019-0111-0).

Competing interests: W.F.S., B.V.S., C.H., C.F. and R.L. are co-inventors on patent WO2017189976A1 ‘Targeted measure of transcriptional activity related to hormone receptors’. W.F.S., C.F. and R.L. own shares of Delphi Diagnostics, without employment or administrative position. W.F.S. owns shares of IONIS Pharmaceuticals and has received honoraria from Merck and from Almac Diagnostics during the past 12 months. J.L. received grants or research support from Pfizer, Astra Zeneca, Genentech, EMDSerono and GSK, is member of the speaker’s bureau at Medilearning group and PER and advisory committee member (uncompensated) at Pfizer and Astra-Zeneca. A.T. received research grants from Guebert LCC and consulted Merit Medical, Jounce Therapeutics, AbbVie. R.S. participated in an advisory board at BMS, received travel grants and research support from Merck, Roche. T.A.K. served as speaker for Genomic Health. C.H. is currently an employee of Bristol-Myers Squibb. The other authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Rugo, H. S. et al. Endocrine therapy for hormone receptor-positive metastatic breast cancer: American Society of Clinical Oncology Guideline. J. Clin. Oncol. 34, 3069–3103 (2016).
2. Cardoso, F. et al. ESO-ESMO 2nd international consensus guidelines for advanced breast cancer (ABC2). Ann. Oncol. 25, 1871–1888 (2014).
3. Lower, E. E., Glass, E. L., Bradley, D. A., Blau, R. & Heffelfinger, S. Impact of metastatic estrogen receptor and progesterone receptor status on survival. Breast Cancer Res. Treat. 90, 65–70 (2005).
4. Hoeffnagel, L. D. et al. Receptor conversion in distant breast cancer metastases. Breast Cancer Res. 12, R75 (2010).
5. Amir, E. et al. Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. J. Clin. Oncol. 30, 587–592 (2012).
6. Thompson, A. M. et al. Prospective comparison of switches in biomarker status between primary and recurrent breast cancer: the Breast Recurrence In Tissues Study (BRITS). Breast Cancer Res. 12, R92 (2010).
7. Toy, W. et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. Nat. Genet. 45, 1439–1445 (2013).
8. Robinson, D. R. et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat. Genet. 45, 1446–1451 (2013).
9. Barrios, C. et al. The sequential use of endocrine treatment for advanced breast cancer: where are we? Ann. Oncol. 23, 1378–1386 (2012).
10. Symmans, W. F. et al. Genomic index of sensitivity to endocrine therapy for breast cancer. J. Clin. Oncol. 28, 4111–4119 (2010).
11. Ravidn, B. P. M. et al. Estrogen receptor-positive patients with Metastatic Breast Southwest Oncology Group Study. J. Clin. Oncol. 10, 1284–1291 (1992).
12. Angus, L., Beije, N., Jager, A., Martens, J. W. M. & Sleijfer, S. ESR1 mutations: moving towards guiding treatment decision-making in metastatic breast cancer patients. Cancer Treat. Rev. 35, 33–40 (2017).
13. Veer, L. J. et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–536 (2002).
14. Paik, S. et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N. Engl. J. Med. 351, 2817–2826 (2004).

Published in partnership with the Breast Cancer Research Foundation
15. Filipits, M. et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin. Cancer Res.* **17**, 6012–6020 (2011).

16. Sussman, D. et al. SGN-LIV1A: a novel antibody-drug conjugate targeting LIV-1 for the treatment of metastatic breast cancer. *Mol. Cancer Ther.* **13**, 2991–3000 (2014).

17. Jeselsohn, R. et al. Emergence of constitutively active estrogen receptor-α mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin. Cancer Res.* **20**, 1757–1767 (2014).

18. Spoerke, J. M. et al. Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant. *Nat. Commun.* **7**, 11579 (2016).

19. Friibbens, C. et al. Plasma ESR1 mutations and the treatment of estrogen receptor-Positive advanced breast cancer. *J. Clin. Oncol.* **34**, 2961–2968 (2016).

20. Chandraflapaty, S. et al. Prevalence of ESR1 mutations in cell-free DNA and outcomes in metastatic breast cancer. *JAMA Oncol.* **2**, 1310–1315 (2016).

21. Bardou, V. J., Arpino, G., Elledge, R. M., Osborne, C. K. & Clark, G. M. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *J. Clin. Oncol.* **21**, 1973–1979 (2003).

22. Mohammed, H. et al. Progesterone receptor modulates ERα action in breast cancer. *Nature* **523**, 313–319 (2015).

23. Viale, G. et al. Prognostic and predictive value of centrally reviewed expression of estrogen and progesterone receptors in a randomized trial comparing letrozole and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1-98. *J. Clin. Oncol.* **25**, 3846–3852 (2007).

24. Gerlinger, M. & Swanton, C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br. J. Cancer* **103**, 1139–1143 (2010).

25. Zardavas, D., Irrthum, A., Swanton, C. & Piccart, M. Clinical management of breast cancer heterogeneity. *Nat. Rev. Clin. Oncol.* **12**, 381–394 (2015).

26. Dodwell, D., Wardley, A. & Johnston, S. Postmenopausal advanced breast cancer: Options for therapy after tamoxifen and aromatase inhibitors. *Breast* **15**, 584–594 (2006).

27. Beslija, S. et al. Second consensus on medical treatment of metastatic breast cancer. *Ann. Oncol.* **18**, 215–225 (2007).

28. Amedos, M. et al. Precision medicine for metastatic breast cancer—limitations and solutions. *Nat. Rev. Clin. Oncol.* **12**, 693–704 (2015).

29. Sinn, B. V. et al. Metadata record supporting files in the related publication: “SET (ER/PR)—a robust 18-gene Predictor for Sensitivity to Endocrine Therapy for Metastatic Breast Cancer”. https://doi.org/10.6084/m9.figshare.7998809. (2019).

30. Hatzis, C. et al. A genomic predictor of response and survival following taxane- and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1–98. *Clin. Cancer Res.* **20**, 1767–1773 (2014).

31. Altman, D. G., McShane, L. M., Sauerbrei, W. & Taube, S. E. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *PloS Med.* **9**, e1001216 (2012).

32. Hatzis, C. et al. Effects of tissue handling on RNA integrity and microarray measurements from resected breast cancers. *J. Natl. Cancer Inst.* **103**, 1871–1883 (2010).

33. Halbe-Kains, B. et al. A three-gene model to robustly identify breast cancer molecular subtypes. *J. Natl. Cancer Inst.* **104**, 311–325 (2012).

34. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2015). http://www.R-project.org/.

35. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).

© The Author(s) 2019 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.