

# ENDORSE: a prognostic model for endocrine therapy in estrogen-receptor-positive breast cancers

Aritro Nath<sup>1</sup> , Adam L Cohen<sup>2</sup> & Andrea H Bild<sup>1,\*</sup> 

## Abstract

Advanced and metastatic estrogen receptor-positive (ER<sup>+</sup>) breast cancers are often endocrine resistant. However, endocrine therapy remains the primary treatment for all advanced ER<sup>+</sup> breast cancers. Treatment options that may benefit resistant cancers, such as add-on drugs that target resistance pathways or switching to chemotherapy, are only available after progression on endocrine therapy. Here we developed an endocrine therapy prognostic model for early and advanced ER<sup>+</sup> breast cancers. The endocrine resistance (ENDORSE) model is composed of two components, each based on the empirical cumulative distribution function of ranked expression of gene signatures. These signatures include a feature set associated with long-term survival outcomes on endocrine therapy selected using lasso-regularized Cox regression and a pathway-based curated set of genes expressed in response to estrogen. We extensively validated ENDORSE in multiple ER<sup>+</sup> clinical trial datasets and demonstrated superior and consistent performance of the model over clinical covariates, proliferation markers, and multiple published signatures. Finally, genomic and pathway analyses in patient data revealed possible mechanisms that may help develop rational stratification strategies for endocrine-resistant ER<sup>+</sup> breast cancer patients.

**Keywords** biomarker; cox proportional hazards model; endocrine resistance; lasso; survival analysis

**Subject Categories** Biomarkers; Cancer; Computational Biology

**DOI** 10.15252/msb.202110558 | Received 17 November 2021 | Revised 19 May 2022 | Accepted 20 May 2022

**Mol Syst Biol.** (2022) 18: e10558

## Introduction

Breast cancer is the most common form of cancer globally, with more than two million cases diagnosed in 2020 (Sung *et al.*, 2021). Pathogenesis and classification of breast cancer is based on the presence or absence of estrogen receptor alpha (ER), progesterone receptor (PR), and human growth factor-neu receptor (HER2). These

subtypes guide the selection of systemic therapy for breast cancer patients. More than 70% of breast cancers express ER and are negative for HER2 (ER<sup>+</sup>/HER2<sup>-</sup>) (Harvey *et al.*, 1999; Kohler *et al.*, 2015). The primary systemic therapy for ER<sup>+</sup>/HER2<sup>-</sup> breast cancer is endocrine therapy, which counters the growth of tumors by targeting their dependency on estrogen signaling (Waks & Winer, 2019). These therapies include selective estrogen receptor modulators (SERMs) such as tamoxifen and selective estrogen receptor degraders (SERDs) such as fulvestrant that directly prevent ER activation, or aromatase inhibitors like exemestane and anastrozole that reduce circulating levels of estrogen in the body (Smith & Dowsett, 2003; McDonnell & Wardell, 2010). Endocrine therapy substantially reduces the risk of recurrence within 5 years, although chemotherapy may be recommended for some patients with a high risk of recurrence. While clinicopathological features are not reliable predictors of recurrence risk, gene expression-based genomic tests that predict the risk of recurrence can aid in deciding whether the benefit of adding chemotherapy outweighs its side effects in certain patients (Cardoso *et al.*, 2016; Sparano *et al.*, 2018). These biomarkers have been validated and recommended for clinical use only in early stage, node-negative cancers based on guidelines from the American Society of Clinical Oncology and European Group on Tumor Markers (Duffy *et al.*, 2017; Krop *et al.*, 2017).

Locally advanced and metastatic ER<sup>+</sup> breast cancers often develop resistance to endocrine therapy with significantly higher rates of recurrence and death compared to early-stage disease. Despite these challenges, single-agent endocrine therapy or in combination with CDK4/6 inhibitors remains the primary systemic therapy recommended for locally advanced and metastatic breast cancers (McAndrew & Finn, 2020). Patients may benefit from the addition of a targeted inhibitor against the MTOR or PI3K pathways (Baselga *et al.*, 2012; André *et al.*, 2019) or switching to chemotherapy (McAndrew & Finn, 2020). However, these treatment options are recommended for consideration only upon progression on endocrine therapy, according to the American Society for Clinical Oncology (Rugo *et al.*, 2016), National Comprehensive Cancer Network (Gradishar *et al.*, 2017, 2020) and European Society for Medical Oncology (Cardoso *et al.*, 2020) clinical practice guidelines. Therefore, the ability to predict the potential benefit from first-line

1 Department of Medical Oncology and Therapeutics, City of Hope Comprehensive Cancer Center, Monrovia, CA, USA

2 Neuro Oncology Program, Inova Schar Cancer Institute, Fairfax, VA, USA

\*Corresponding author. Tel: +1 626 218 6052; E-mail: abild@coh.org

endocrine therapy may be crucial for locally advanced and metastatic ER<sup>+</sup> breast cancers that may benefit from continued endocrine therapy, a combination treatment or chemotherapy as the primary treatment strategy.

Unlike early stage, node-negative disease, genomic tests for endocrine therapy response are not available for advanced and metastatic ER<sup>+</sup> breast cancers. To address this limitation, a few studies have assessed the genomic signature of endocrine response in ER<sup>+</sup> metastatic breast cancers (ER<sup>+</sup> MBC) (Jeselsohn *et al*, 2016; Sinn *et al*, 2019). The TransCONFIRM trial evaluated the transcriptomes of 112 ER<sup>+</sup>/HER2<sup>-</sup> MBCs and identified a set of 37 genes that were associated with progression-free survival (PFS) of patients receiving fulvestrant (Jeselsohn *et al*, 2016). Another study analyzed the transcriptomes of 140 ER<sup>+</sup>/HER2<sup>-</sup> MBC on endocrine therapy to develop SET ER/PR, an 18-gene predictive score for endocrine therapy sensitivity (Sinn *et al*, 2019). While both the TransCONFIRM and SET ER/PR biomarkers predicted endocrine response in their respective training datasets, neither study performed systematic validation of their predictive signatures to demonstrate the reproducibility and accuracy in independent clinical datasets. This issue highlights a critical flaw in biomarker development pipelines and is one important reason why genomic biomarkers are infrequently translated into clinical practice (Boutros, 2015). Another pervasive issue hindering clinical translation arises from the reliance on a large number of predictive features in complex models that are difficult to interpret and often perform poorly in independent validation due to overfitting (Taylor *et al*, 2008; Witten & Tibshirani, 2010).

Here we developed ENDORSE: a low-dimensional gene expression-based prognostic model for endocrine therapy, and systemically tested its performance and predictive ability in multiple clinical trial datasets. ENDORSE was developed using the tumor transcriptomes and overall survival (OS) of more than 800 ER<sup>+</sup> breast cancers on endocrine therapy (Curtis *et al*, 2012; Pereira *et al*, 2016). We evaluated the performance of ENDORSE compared to clinical covariates, proliferation markers, and other published signatures. We validated the ENDORSE model in multiple independent clinical trial datasets, including the TransCONFIRM and SET ER/PR trials for endocrine therapy in metastatic ER<sup>+</sup> breast cancer. Our results show that ENDORSE consistently identified high-risk patients and outperformed all other prognostic models in ER<sup>+</sup> breast cancers.

## Results

### Developing a prognostic model for endocrine therapy

The objective of this study is to develop a prognostic model for all ER<sup>+</sup> cancers on endocrine therapy regardless of tumor stage, grade or node status. We developed a two-component prognostic model for endocrine therapy response using the tumor transcriptomes and long-term survival outcomes of 833 ER<sup>+</sup>/HER2<sup>-</sup> tumors that received endocrine therapy (Curtis *et al*, 2012; Pereira *et al*, 2016; Table 1, Fig 1A). About 2 in 5 tumors in this training cohort were node-positive, while more than a third of the tumors were poorly differentiated, grade 3 tumors (Table 1). The two components included an empirical gene signature modeled on OS (median = 10 years)

**Table 1. Training data patient characteristics**

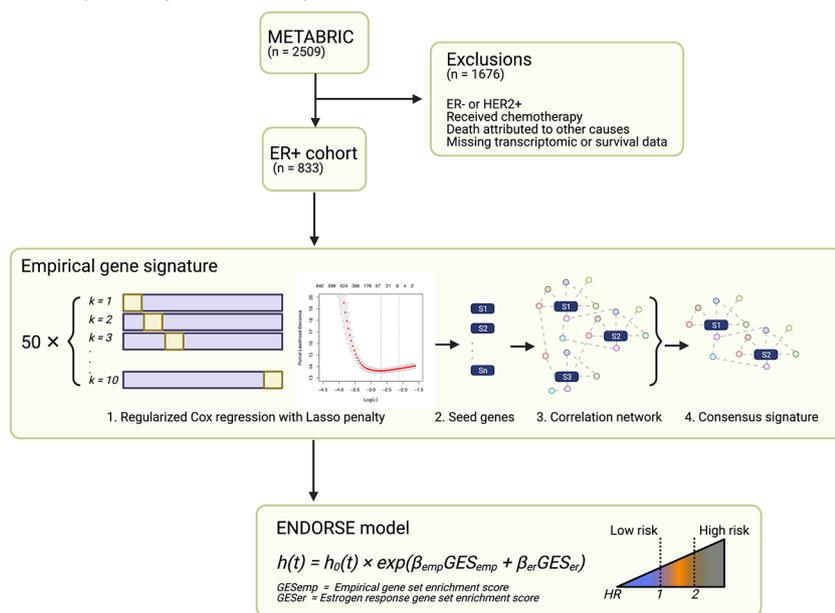
Variable	Mean	95% CI	N available
Time to event (in months)	135	130–140	833
Events (death due to disease)	0.409	0.376–0.443	833
Age at diagnosis	61.5	60.7–62.3	833
Mutation count	5.55	5.31–5.79	809
Tumor size	24.3	23.4–25.1	828
Tumor stage	1.64	1.59–1.69	634
Stage 0–1 (n = 270)			
Stage 2 (n = 324)			
Stage > 3 (n = 40)			
Tumor Grade	2.29	2.18–2.27	808
Grade 1 (n = 103)			
Grade 2 (n = 417)			
Grade 3 (n = 288)			
Number of positive lymph nodes detected	1.56	1.32–1.79	833
0 (n = 491)			
1–3 (n = 228)			
4–9 (n = 85)			
> 10 (n = 29)			

and a curated gene signature defining response to estrogen (Liberzon *et al*, 2015). Figure 1A outlines the inclusion criteria for the training dataset, method for developing the empirical gene signature and the final ENDORSE model based on the gene set enrichment scores (GES) of the two signatures. The empirical signature was developed by first performing a feature selection on the training dataset using a repeated cross-validation analysis of a lasso-regularized proportional hazards model. Each iteration yielded a core set of predictive features that were expanded to a correlation network. The final gene signature was derived from the consensus correlation network, defined as genes appearing in at least 50% of the iterations (Fig EV1, Dataset EV1). In a bivariate Cox proportional hazards model of the training data, the empirical signature was associated with a reduction in survival probability, while the estrogen response signature was associated with improved survival (Fig 1B). The coefficients for the ENDORSE model were calculated using the training cohort, resulting in ENDORSE = 1.54 × (empirical signature GES) – (2.72 × estrogen response GES). The ENDORSE model could also be used to stratify the tumors based on predicted risk. For example, a threshold of ≥ 2-fold relative risk of death as “high-risk” and ≤ 1 risk as “low-risk” resulted in significant difference between the strata in Kaplan–Meier analysis (log-rank test  $P = 3 \times 10^{-14}$ ; medium-risk  $P = 5.43 \times 10^{-12}$ , high-risk  $P = 5.99 \times 10^{-9}$ ; Fig 1C).

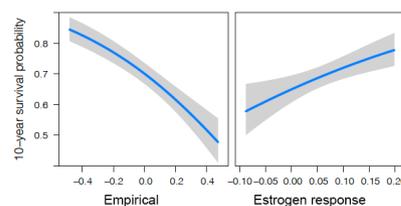
### ENDORSE model performance evaluation with clinical covariates and published signatures in METABRIC

To evaluate the performance of the ENDORSE model against clinical factors, proliferation index, and published signatures, we created a training and a hold-out test (validation) subset by performing a

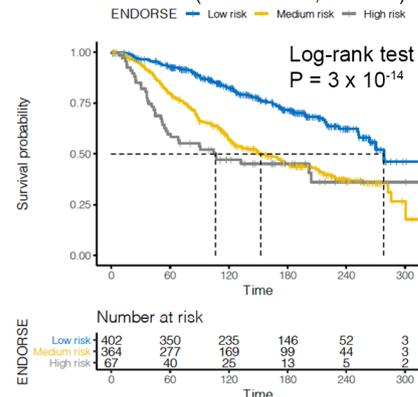
### A Empirical signature development and ENDORSE model



### B Predicted survival probabilities



### C METABRIC cohort (ER+/HER2-, n = 833)



**Figure 1. ENDORSE model development in METABRIC.**

- A Inclusion criteria and schematic of ENDORSE model development. Training samples were selected based on ER status and excluded from the analysis if they were either HER2<sup>+</sup>, received chemotherapy, died due to other causes besides breast cancer, or were missing transcriptomic or survival data. The empirical signature was developed using a repeated cross-validation analysis framework. Each iteration of the lasso-regularized proportional hazards model generated a feature set (seed genes) predictive of OS. The seed genes were expanded to a network of intercorrelated genes, and the final empirical signature was defined by identifying a consensus set across all iterations. The two-feature ENDORSE model was then constructed using the gene set enrichment scores of the empirical signature and estrogen response signature.
- B Predicted 10-year survival probabilities of the ER<sup>+</sup>/HER2<sup>-</sup> METABRIC breast cancers ( $n = 833$ ) based on a Cox proportional hazards model of gene signature enrichment scores of the empirical and estrogen response signatures as predictor variables. Grey shaded area indicates 95% confidence intervals of the Cox model predictions.
- C Kaplan–Meier curves and risk tables of METABRIC ER<sup>+</sup>/HER2<sup>-</sup> tumors stratified by ENDORSE. Dashed lines indicate median survival.  $P$ -value were obtained using the log-rank test for survival curves. The tumors were stratified according to an ENDORSE risk score (hazard ratio) threshold of  $\geq 2$  to define high-risk,  $\leq 1$  as low risk and all other intermediate values as medium risk.

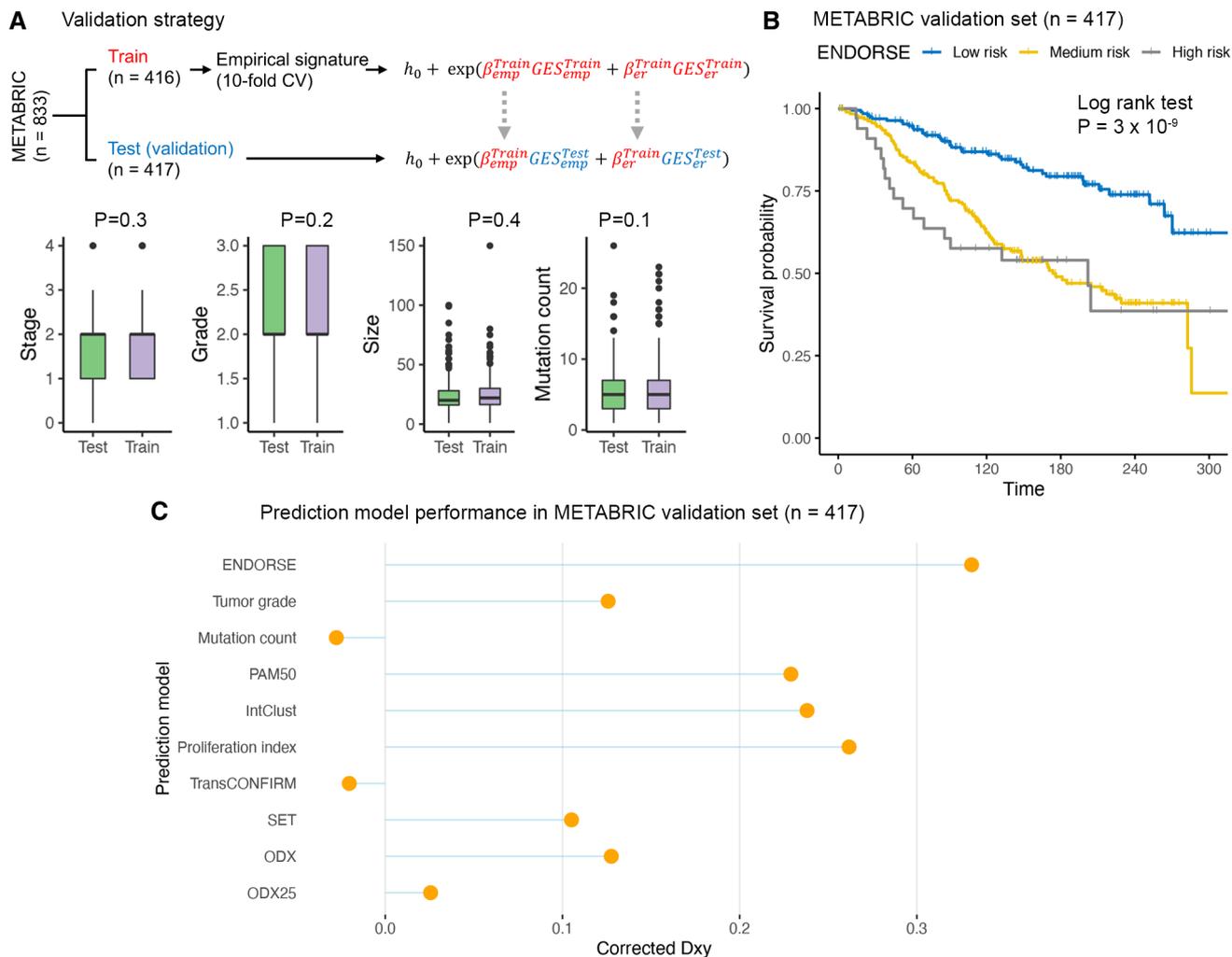
50/50 randomized data split, recalculating the empirical signature using only the training subset and obtaining the ENDORSE model coefficients (Fig 2A). The training and validation subsets had no significant differences in key pathological features, including tumor stage, grade, size and mutation count (Fig 2A). We predicted risk in the held-out validation subset using the training subset ENDORSE model for performance evaluation. Kaplan–Meier analyses of the validation subset showed a significant difference between the risk strata predicted using the ENDORSE training subset model (log-rank test  $P = 3 \times 10^{-9}$ ; medium risk  $P = 6.04 \times 10^{-8}$ , high risk  $P = 9.47 \times 10^{-8}$ ) (Fig 2B).

We performed bootstrap resampling analyses to validate the model in the held-out dataset and compared with other univariate prognostic models (Fig 2C). First, we compared ENDORSE with clinical factors, such as tumor grade and mutation burden. The ENDORSE model (Somers' D or  $D_{xy} = 0.351$ ) performed better than both tumor grade ( $D_{xy} = 0.129$ ) and mutation count ( $D_{xy} = -0.027$ ). We also compared the model with a “meta-PCNA” proliferation index that was reported to capture the prognostic ability of most published signatures of breast cancer (Venet *et al*, 2011; Ramaker *et al*, 2017). Again, the ENDORSE model outperformed the

proliferation index ( $D_{xy} = 0.265$ ),  $P = 4.42 \times 10^{-5}$ ), indicating its utility over measures of proliferation as a prognostic tool.

Next, we evaluated published prognostic signatures for breast cancers and compared their performance with ENDORSE. These signatures included PAM50, a 50-gene signature that was previously reported to be a better prognostic tool for ER<sup>+</sup> breast cancers on endocrine therapy than clinical factors, such as histopathological classification and tumor grade (Nielsen *et al*, 2010). A genomic classifier, IntClust, that was developed by the METABRIC consortium authors and trained on the same training dataset was also included in this comparison (Dawson *et al*, 2013). Both the PAM50 model ( $D_{xy} = 0.237$ ) and IntClust ( $D_{xy} = 0.246$ ) models were surpassed by ENDORSE.

Two previous clinical trials evaluating endocrine therapy response in metastatic ER<sup>+</sup> breast cancers developed prognostic signatures using tumor transcriptomes. The first signature developed in the TransCONFIRM trial included 37 genes that were associated with PFS of advanced ER<sup>+</sup> breast cancers on fulvestrant (Jeselson *et al*, 2016). We replicated the approach described in the study by performing hierarchical clustering of the samples based on the expression levels of the 37 genes and cutting the tree to obtain



**Figure 2. Model evaluation and comparison with other predictors in METABRIC.**

- A Schematic of validation strategy (top) and boxplots (bottom) comparing key pathological variables between the training and test subsets. The METABRIC cohort was split into a training subset ( $n = 416$ ) and test or validation subset ( $n = 417$ ), with each sample representing an independent biological replicate. The empirical signature was derived using the training split. Then, the ENDORSE model was trained only using the training subset. The coefficients from the training model were used to predict risk in the test subset. The colored boxes in the boxplots display interquartile range with median, while the whiskers show  $1.5 \times$  interquartile range.  $P$ -values indicate significance of difference in mean from two-tailed Welch's  $T$ -test.
- B Kaplan–Meier curves and risk tables of the validation subset ( $n = 417$ ) tumors stratified using ENDORSE risk predicted using the training subset model.  $P$ -values were obtained using log-rank test.
- C Lollipop plots displaying corrected Somer's  $D_{xy}$  indices of ENDORSE and various other univariate Cox proportional hazards models in the validation subset ( $n = 417$ ). The corrected  $D_{xy}$  indices were calculated using 150-fold bootstrap resampling of the validation subset.

two clusters. We referred to resultant clusters as the “TransCONFIRM” score. The TransCONFIRM score applied to the METABRIC dataset performed poorly ( $D_{xy} = -0.023$ ). The second signature (SET ER/PR) was developed using tumor transcriptomes of metastatic ER<sup>+</sup> breast cancers on endocrine therapy (Sinn *et al.*, 2019). This signature included 18 predictive genes that were correlated with *ESR1* or *PGR* expression and normalized using 10 reference transcripts. We implemented the methods described in the original study and referred to the resultant score as “SET.” The SET score ( $D_{xy} = 0.108$ ) performed better than TransCONFIRM; however, it was also easily outperformed by ENDORSE. Finally, we calculated a surrogate biomarker based on the published formula for the 21-gene

prognostic signature approved for early-stage, node-negative ER<sup>+</sup> breast cancers (Paik *et al.*, 2004). We referred to this score as ODX. We also compared a classifier that stratified samples based on the 25<sup>th</sup> percentile of ODX score as a proxy for the latest risk stratification threshold for this signature (Sparano *et al.*, 2018) and referred to this score as ODX25. We found that the ODX model ( $D_{xy} = 0.127$ ) was comparable to other published signatures like the SET score, but the stratified ODX25 score performed poorly ( $D_{xy} = 0.021$ ). In addition to the comparisons in the independent validation subset, we also performed similar comparisons between models fitted on the complete METABRIC dataset (Fig EV2A). Using partial likelihood ratio tests, we confirmed the superior performance of the

ENDORSE model compared to its components and all other clinical factors and prognostic signatures fitted on the complete dataset (Fig EV2B). These results demonstrate that ENDORSE is a significantly better prognostic model.

We also applied the ENDORSE risk estimates to stratify 133 METABRIC ER<sup>+</sup> tumors that received a combination of endocrine therapy and chemotherapy and found a significant difference in the predicted strata ( $P = 3.7 \times 10^{-7}$ ) (Fig EV3A). On the other hand, applying the ENDORSE model to 429 ER-negative METABRIC breast cancers as negative control showed no significant difference between the strata ( $P = 0.09$ , Fig EV3B). These data suggest that the ENDORSE model is specific to the ER<sup>+</sup> breast cancers and not a general prognostic model to identify aggressive tumors.

### Validation and performance evaluation in independent clinical trial datasets

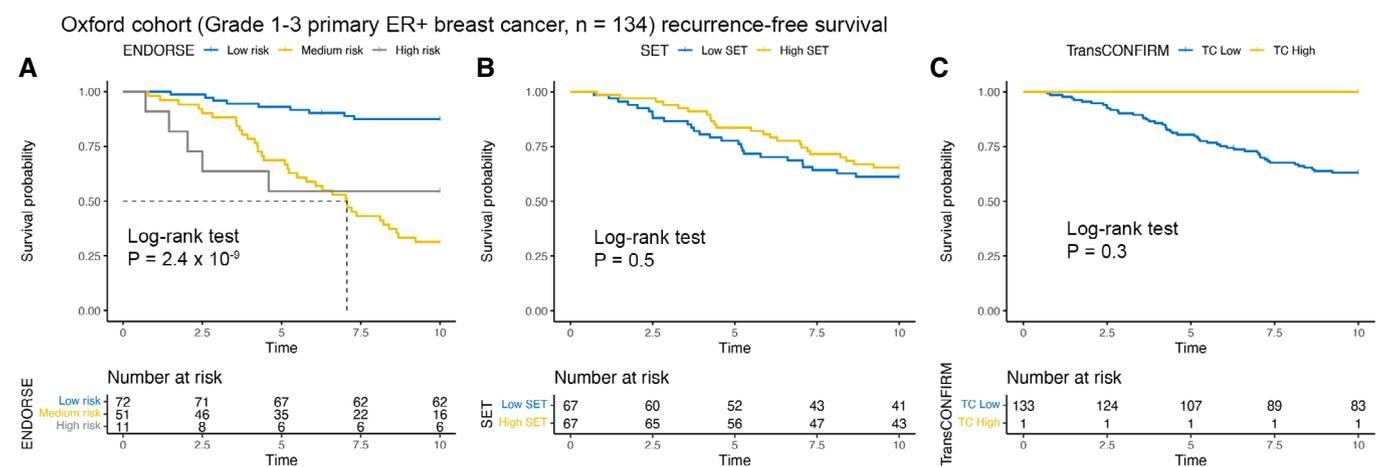
To test the reproducibility and validate the performance of ENDORSE, we applied the model to the baseline transcriptomes of ER<sup>+</sup> tumors from four independent clinical trials and compared the ENDORSE-predicted risk or strata with the outcomes reported in each trial. These independent trials also included the TransCONFIRM and SET ER/PR studies discussed earlier. So, we also compared the performance of TransCONFIRM and SET scores in their respective training datasets and across other independent datasets.

First, we evaluated the performance of ENDORSE in the Oxford cohort, an independent validation dataset with similar key histopathological features as the METABRIC dataset. This cohort included 134 ER<sup>+</sup> breast cancers and about 31% of the tumors were grade III, comparable to the 35% grade III tumors in the METABRIC cohort. The Oxford study reported 10-year recurrence-free survival data of the patients. We evaluated the performance of the ENDORSE model and compared it against the SET and TransCONFIRM scores in stratifying the patients. Kaplan–Meier analyses using ENDORSE predicted risk show significant difference in the strata (log-rank test  $P = 2 \times 10^{-9}$ ; medium-risk  $P = 5.23 \times 10^{-8}$ , high-risk  $P = 2.38 \times$

$10^{-3}$ ) (Fig 3A). In comparison, both the SET scores ( $P = 0.5$ ) (Fig 3B) and TransCONFIRM scores ( $P = 0.3$ ) (Fig 3C) failed to meaningfully stratify the tumors. Thus, only the ENDORSE model could successfully stratify tumors in a comparable independent dataset.

The TransCONFIRM trial evaluated fulvestrant response in 112 advanced metastatic ER<sup>+</sup> breast cancers previously treated with an anti-estrogen (Jeselson *et al*, 2016). While the original study developed and evaluated the performance of their 37-gene signature based on PFS, this survival data was not made available with the publication (the authors did not respond to our requests for this data). However, the study reported the post-therapy resistant or sensitive states of the tumors based on histopathological staining (Ki67 staining). Therefore, we compared the percentage of cells positive for Ki67 staining reported by in study with risk predictions from ENDORSE and other signatures (Fig 4). The percentage of cells positive of Ki67 was significantly correlated with the ENDORSE estimated risk ( $P = 2.5 \times 10^{-5}$ ) (Fig 4A), while stratification of the patients based on the risk thresholds also showed a significant difference in Ki67 staining percentage between the strata (ANOVA  $P = 1.2 \times 10^{-3}$ ; medium-risk vs. low-risk  $P = 0.04$ ; high-risk vs. low-risk  $P = 1.753 \times 10^{-4}$ ) (Fig 4B). However, the SET score was not correlated with Ki67 staining ( $P = 0.3$ ) (Fig 4C). The TransCONFIRM score that was developed on this dataset was significant ( $P = 0.05$ ) but performed worse than the ENDORSE score trained on an independent dataset (Fig 4D).

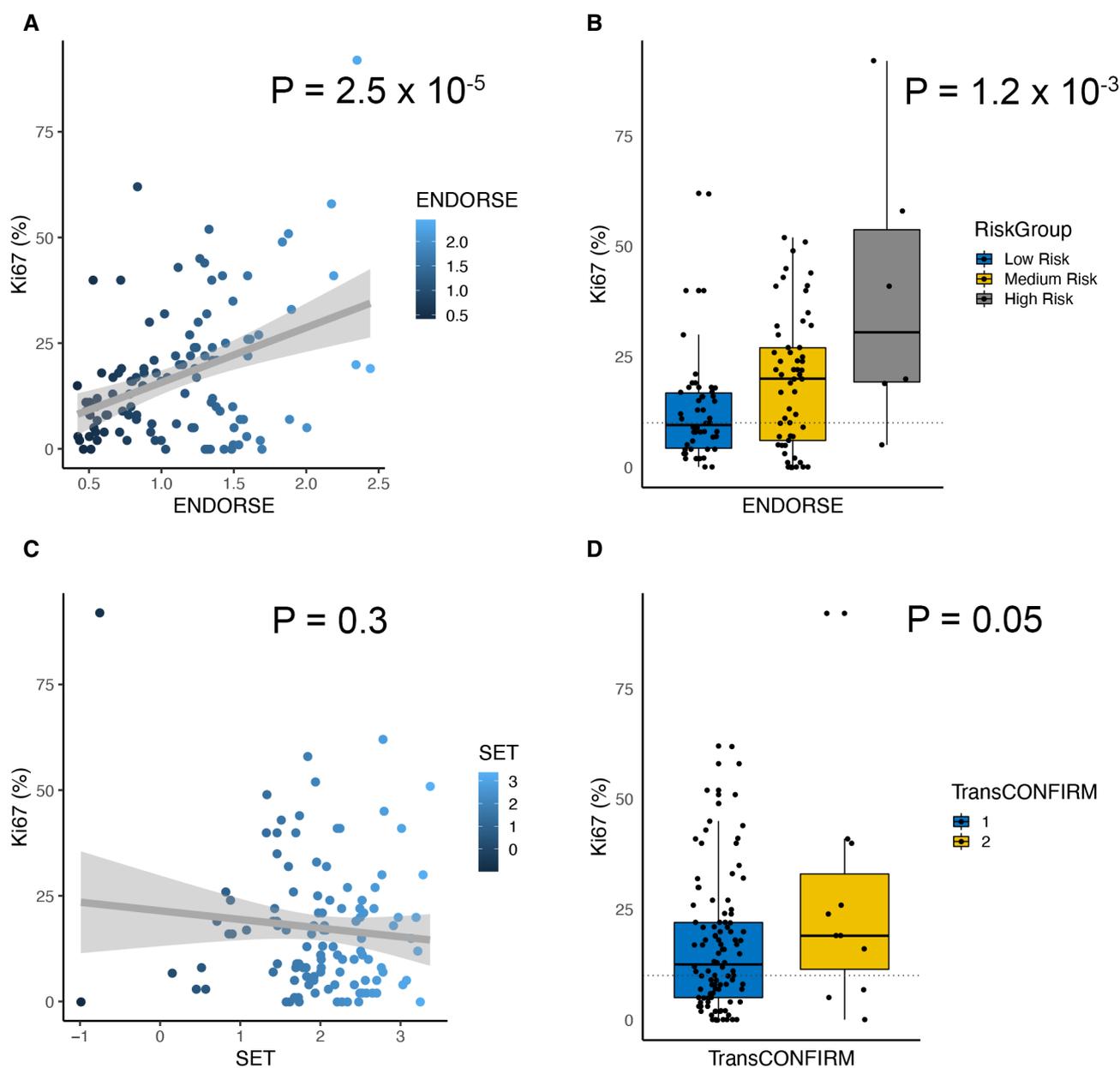
Next, we evaluated the performance of the signatures in the SET ER/PR cohort. This clinical trial reported the PFS and OS of 140 stage IV ER<sup>+</sup> metastatic breast cancers on endocrine therapy. We compared the survival curves of the patients by stratifying them based on the ENDORSE predicted risk, median SET scores, as described in the original study, and the TransCONFIRM score. The stratification based on ENDORSE (log-rank test  $P = 2 \times 10^{-4}$ ; medium-risk  $P = 0.016$ , high-risk  $P = 1.88 \times 10^{-4}$ ) (Fig 5A) and SET ( $P = 3 \times 10^{-3}$ ) (Fig 5B) scores both resulted in significant differences in the survival curves. (However, the TransCONFIRM score (Fig 5C) was not significant;  $P = 0.9$ .) Similarly, we observed that



**Figure 3. Model validation in Oxford cohort.**

A–C Kaplan–Meier recurrence-free survival curves and risk tables of Oxford cohort ER<sup>+</sup> patients ( $n = 134$ ). Dashed lines indicate median survival. The patients were stratified according to (A) ENDORSE (B) SET and (C) TransCONFIRM predicted scores.  $P$ -values were obtained using log-rank test.

## TransCONFIRM cohort (ER+ metastatic breast cancer, n = 112)



**Figure 4. Model validation in TransCONFIRM cohort.**

A Scatter plot comparing ENDORSE scores (X-axis) with TransCONFIRM trial-reported ( $n = 112$ ) percentage of cells stained positive for Ki67 (Y-axis). Linear fit is shown as a grey line with shaded region showing 95% confidence intervals (C.I.).  $P$ -value indicates significance of the linear fit.

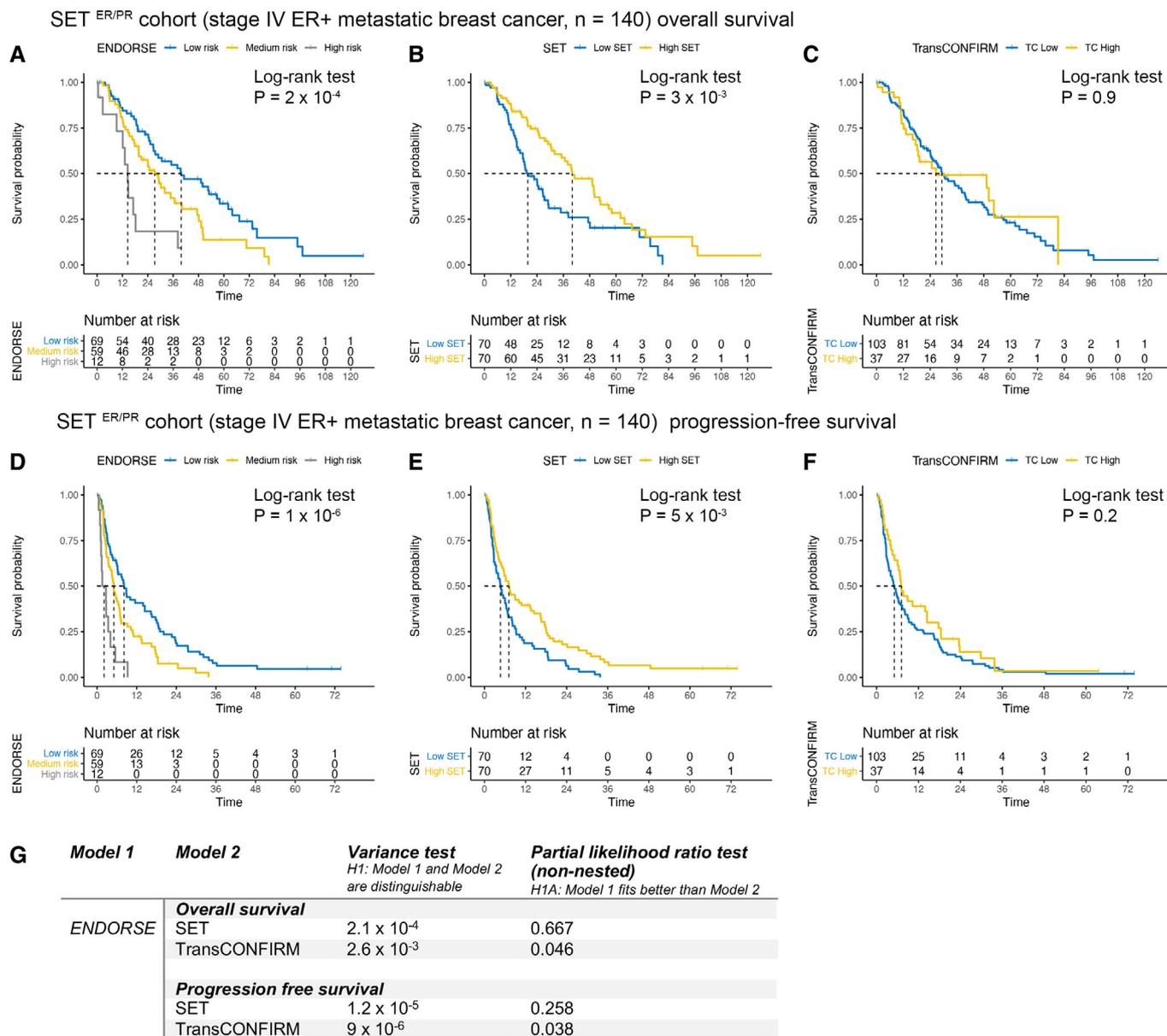
B Boxplot comparing Ki67% across ENDORSE-guided patient strata. The colored boxes display interquartile range with median, while the whiskers show  $1.5 \times$  interquartile range.  $P$ -value indicates significance of the ANOVA model and the horizontal dotted line at 10% indicates threshold of resistance.

C Scatter plot comparing SET scores (X-axis) Ki67% (Y-axis). Linear fit is shown as a grey line with shaded region showing 95% confidence intervals (C.I.).  $P$ -value indicates significance of the linear fit.

D Boxplot comparing Ki67% across TransCONFIRM predicted patient strata. The colored boxes display interquartile range with median, while the whiskers show  $1.5 \times$  interquartile range.  $P$ -value indicates significance of the ANOVA model and the horizontal dotted line at 10% indicates threshold of resistance.

ENDORSE (log-rank test  $P = 1 \times 10^{-6}$ ; medium-risk  $P = 0.004$ , high-risk  $P = 2.35 \times 10^{-6}$ ) (Fig 5D) and SET ( $P = 5 \times 10^{-3}$ ) (Fig 5E) scores both resulted in significant differences in the PFS curves,

while TransCONFIRM was not significant ( $P = 0.2$ ) (Fig 5F). Additionally, we compared the model fits using partial likelihood ratio tests (Fig 5G). The SET model that was trained using the same



**Figure 5. Model validation in SET ER/PR cohort.**

A–C OS Kaplan-Meier curves and risk tables of SET ER/PR patients (n = 140). Dashed lines indicate median survival. The patients were stratified according to (A) ENDORSE (B) SET and (C) TransCONFIRM predicted scores. P-values were obtained using log-rank tests.  
 D–F PFS Kaplan-Meier curves and risk tables of SET ER/PR patients. Dashed lines indicate median survival. The patients were stratified according to (A) ENDORSE (B) SET and (C) TransCONFIRM scores. P-values were obtained using log-rank tests.  
 G Table comparing the ENDORSE overall and PFS models with SET and TransCONFIRM models using partial likelihood ratio tests for non-nested Cox models.

dataset was not a better fit than the ENDORSE model (OS  $P = 0.667$ , PFS  $P = 0.258$ ). The ENDORSE model was a better fit than the TransCONFIRM model in each case (OS  $P = 0.046$ , PFS  $P = 0.038$ ).

In addition to the two metastatic ER+ breast cancer trials, we also evaluated the performance of the signatures in the ACOSOG Z1031B clinical trial which evaluated neoadjuvant aromatase inhibitor (AI) treatment in Stage II or III ER+ breast cancers (Ellis et al, 2017). This study reported percentage of Ki67 staining both at the study baseline

and at the end of treatment (2–4 weeks). We compared the percentage of Ki67 positive cells across cancers stratified by the ENDORSE score and found significant difference across the classes at both the baseline (ANOVA  $P = 4.9 \times 10^{-9}$ ; medium risk vs. low risk  $P = 8.58 \times 10^{-5}$ , high risk vs. low risk  $P = 0$ ) and at the end of treatment (ANOVA  $P = 3 \times 10^{-18}$ ; medium risk vs. low risk  $P = 1.6 \times 10^{-6}$ , high risk vs. low risk  $P = 0$ ) (Fig 6A). Similarly, the continuous ENDORSE scores were significantly correlated with both

**A** ACOSOG Z1031B cohort (stage II/III ER+ breast cancer, n = 109)

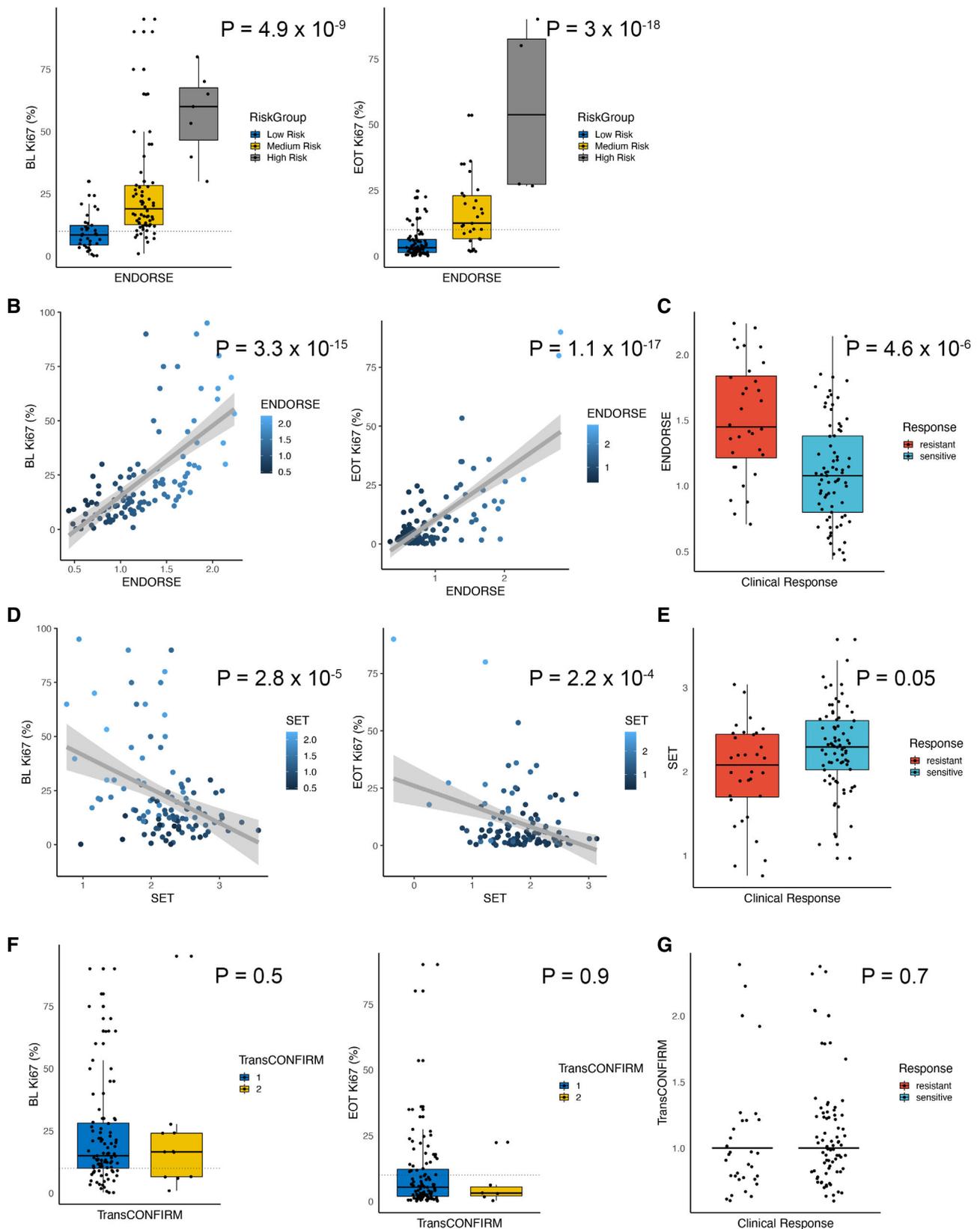


Figure 6.

**Figure 6. Model validation in ACOSOG Z1031B cohort.**

- A Boxplots comparing Ki67% at the baseline (left panel) and end of treatment (right panel) across ENDORSE-predicted patient strata in the ACOSOG Z1031B cohort. The colored boxes display interquartile range with median, while the whiskers show 1.5 × interquartile range. The boxplots represent 109 biological replicates at the baseline (low-risk  $n = 35$ , medium-risk  $n = 67$  and high-risk  $n = 7$ ) or at the end of treatment (low-risk  $n = 76$ , medium-risk  $n = 29$  and high-risk  $n = 4$ ).  $P$ -value indicates significance of the ANOVA model and the horizontal dotted line at 10% indicates threshold of resistance.
- B Scatter plot comparing ENDORSE scores (X-axis) and Ki67% (Y-axis) at the baseline (left panel) and end of treatment (right panel). Linear fit is shown as a grey line with shaded region showing 95% confidence intervals (C.I.).  $P$ -value indicates significance of the linear fit.
- C Boxplots comparing ENDORSE scores between patients grouped by clinical response with 32 resistant and 77 sensitive biological replicates. The colored boxes display interquartile range with median, while the whiskers show 1.5 × interquartile range.  $P$ -value indicates significance of the ANOVA model.
- D Scatter plot comparing SET scores (X-axis) and Ki67% (Y-axis) at the baseline (left panel) and end of treatment (right panel). Linear fit is shown as a grey line with shaded region showing 95% confidence intervals (C.I.).  $P$ -value indicates significance of the linear fit.
- E Boxplots comparing SET scores between patients grouped by clinical response with 32 resistant and 77 sensitive biological replicates. The colored boxes display interquartile range with median, while the whiskers show 1.5 × interquartile range.  $P$ -value indicates significance of ANOVA model.
- F Boxplots comparing Ki67% at the baseline (left panel) and end of treatment (right panel) across TransCONFIRM-predicted patient strata. The colored boxes display interquartile range with median, while the whiskers show 1.5 × interquartile range. The boxplots represent 109 biological replicates at the baseline (cluster 1  $n = 97$ , cluster 2  $n = 12$ ) or at the end of treatment (cluster 1  $n = 103$ , cluster 2  $n = 6$ ).  $P$ -value indicates significance of the ANOVA model and the horizontal dotted line at 10% indicates threshold of resistance.
- G Jitter plot comparing TransCONFIRM predictions between patients grouped by clinical response including 32 resistant and 77 sensitive biological replicates with horizontal bars indicating median.  $P$ -value indicates significance of the ANOVA model.

the baseline ( $P = 3.3 \times 10^{-15}$ ) and end of treatment ( $P = 1.1 \times 10^{-17}$ ) Ki67 percentage (Fig 6B). The ENDORSE scores were also significantly higher in the tumors that were classified as resistant based on clinical response ( $P = 4.6 \times 10^{-6}$ ) (Fig 6C). In this cohort, the SET score was also significantly correlated with Ki67 percentage at the baseline ( $P = 2.8 \times 10^{-5}$ ) and end of treatment ( $P = 2.2 \times 10^{-4}$ ) (Fig 6D), with a significant difference in the SET scores between the resistant and sensitive tumors ( $P = 0.05$ ) (Fig 6E). The transCONFIRM scores were not significant at the baseline ( $P = 0.5$ ) and end of treatment ( $P = 0.9$ ) (Fig 6F) or between resistant and sensitive tumors ( $P = 0.7$ ) (Fig 6G).

### Common pathway phenotypes and somatic alterations enriched in high-risk tumors

We analyzed the pathway phenotypes enriched in each dataset to identify potential mechanisms that defined the high-risk tumors. First, we calculated the GES for 50 hallmark, 4,690 curated, and 189 oncogenic signatures from the METABRIC transcriptomes and fitted a generalized additive model for ENDORSE scores with each signature as the predictor (Datasets EV2–EV4). We found multiple hallmark signatures and oncogenic pathways to be significantly associated with the ENDORSE scores (Datasets EV2–EV4). Key enriched hallmark signatures included MTOR signaling ( $P = 1.03 \times 10^{-72}$ ) and MYC targets (v2,  $P = 2.66 \times 10^{-83}$ ), while key oncogenic signatures included gain in E2F1 target expression ( $P = 8.06 \times 10^{-302}$ ) and loss of RB1 activity via p107 and p130 ( $P = 9.51 \times 10^{-137}$ ,  $1.31 \times 10^{-67}$ ) (Datasets EV2 and EV4). Next, we calculated the GES for the hallmark and oncogenic signatures in the three validation datasets (Datasets EV5–EV10). We observed that pathways associated with cell-cycle progression and proliferation, along with signatures for the loss of RB1 activity and activation of the PI3K/AKT/MTOR signaling pathways, were generally enriched across the METABRIC and all the three validation datasets (Fig 6A). Similar to the training dataset, we also found gain in cell cycle progression along with MTOR signaling and E2F1 target expression to be associated with high ENDORSE scores across all datasets (Fig 7A, Datasets EV5–EV10). The commonality of the signatures enriched across different datasets suggested that similar underlying phenotypes were acquired by the high-risk tumors.

We also analyzed the association between gene-level somatic mutations, including non-synonymous single-nucleotide variants (SNV) and copy number alterations, with the ENDORSE scores of the METABRIC ER<sup>+</sup> tumors. We found a statistically significant association (FDR < 0.05) between the ENDORSE scores and SNVs of only five genes (Fig 7B, Dataset EV11). While *PIK3CA* mutations were found in ~50% of all tumors, we found that ENDORSE scores were not significantly higher in tumors with non-synonymous *PIK3CA* variants or activating *PIK3CA* variants that guide the use of PI3K inhibitors (Fig EV4). Of the five significant genes, only tumors with *TP53* mutations showed a significantly higher ENDORSE score (Fig 7B). We then performed bootstrap analyses with the univariate Cox models of the significant genes and found that none of the SNV Cox models performed better than the ENDORSE model (Fig 7B).

Several gene-level amplifications were also associated with significant differences in ENDORSE scores (Figs 7C and EV5). Interestingly, the significant amplifications were localized at chromosome 1q, 8p, 8q, or 11q, suggesting that different genetic alterations affecting a recurring set of loci may be correlated with the emergence of resistance in the high-risk tumors (Fig 7C, Dataset EV12). Like the univariate SNV models above, the univariate copy number alteration models also performed poorly when compared to the ENDORSE model in bootstrap resampling analyses (Fig 7D).

Finally, we evaluated whether the ENDORSE model reflects different key phenotypes associated with early (within 5 years) vs. late (beyond 5 years) mortality. For patients with an ENDORSE risk > 1, we found that tumors associated with late mortality showed elevated estrogen response signatures, whereas tumors linked with early mortality showed enrichment of epithelial to mesenchymal transition, stemness, angiogenesis, and related signaling pathways that are known to contribute to these phenotypes including TGF $\beta$ , YAP, MEK, and AKT pathways (Datasets EV13–EV15).

## Discussion

Breast cancers are classified as ER<sup>+</sup> based on a broad criterion of positive immunohistochemical staining of 1–100% of cell nuclei for the estrogen receptor (Rugo et al, 2016; Allison et al, 2020).

However, ER<sup>+</sup> tumors are heterogeneous, both in terms of dependence on estrogen signaling for growth and survival and intrinsic or acquired resistance to endocrine therapy (Musgrove & Sutherland, 2009; Spoerke et al, 2016). Therefore, optimal clinical management of each ER<sup>+</sup> breast cancer depends on accurate prediction of response to endocrine therapy and selection of companions for

endocrine therapy. Several genomic tests are available for classifying breast cancers into molecular subtypes (Parker et al, 2009) or assessing the likelihood of benefit from chemotherapy in early-stage, node-negative ER<sup>+</sup> breast cancers (Paik et al, 2004; Cardoso et al, 2016). Results from the MINDACT and TAILORx studies (Cardoso et al, 2016; Sparano et al, 2018) show that it is possible

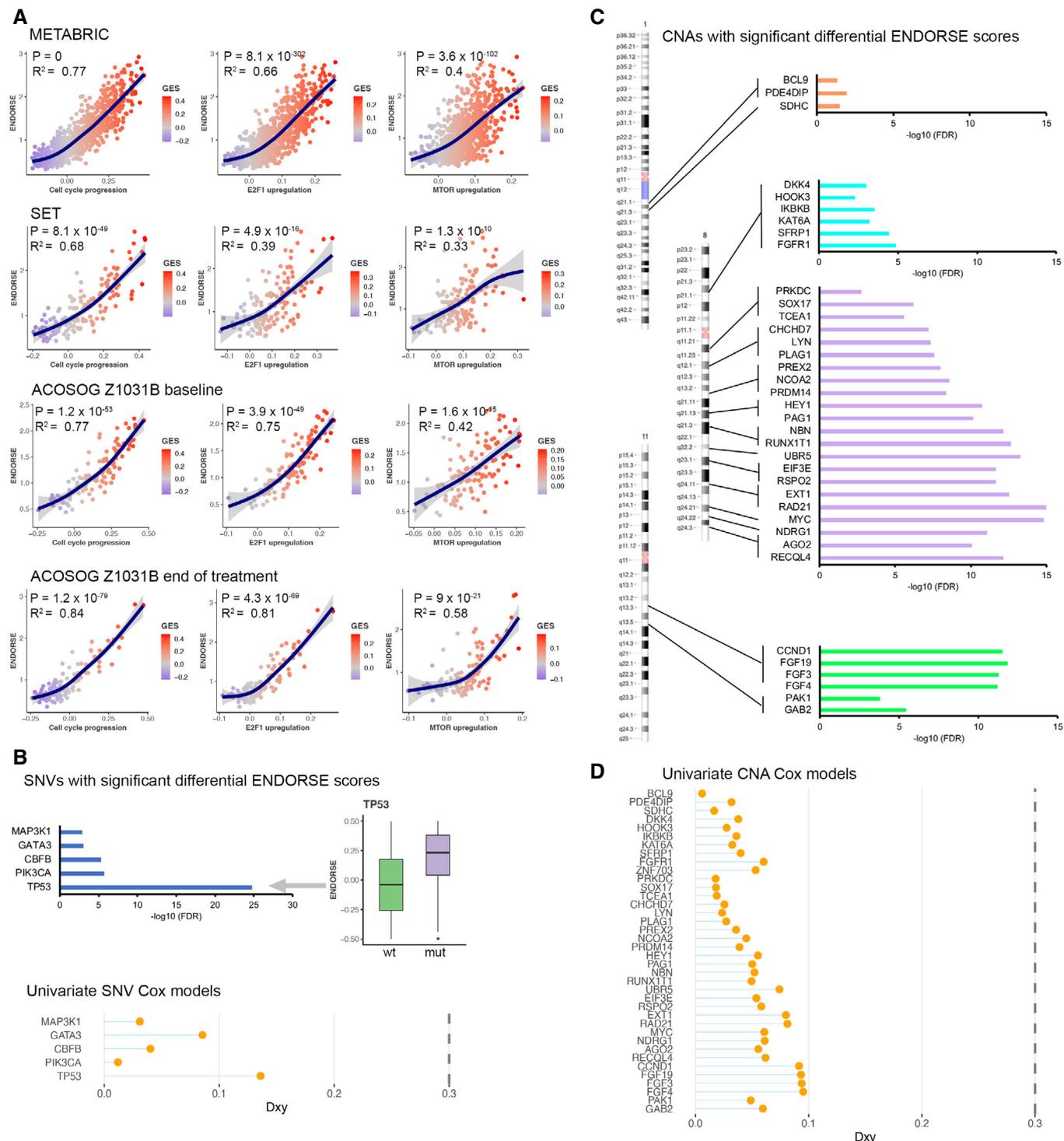


Figure 7.

**Figure 7. Biology of the high-risk tumors.**

- A Scatter plots displaying gene set enrichment scores (GES) of key pathways (X-axis) and ENDORSE scores (Y-axis) in the METABRIC ER<sup>+</sup> cohort ( $n = 833$ ). The cell cycle progression panel represents the hallmark G2 M checkpoint signature, the E2F1 upregulation panel represents E2F1\_UP.V1\_UP oncogenic (C6) signature and the MTOR upregulation panel represents MTOR\_UP.V1\_UP oncogenic (C6) signature. Blue lines with shading indicate generalized additive model fits with 95% C.I., with  $R^2$  and  $P$ -values of the significant of the fit annotated on the panels.
- B Bar plots showing false discovery rate-adjusted  $P$ -values from the ANOVA analysis of ENDORSE scores with mutation status as the grouping variable. The boxplot on the right shows difference in the ENDORSE scores between *TP53* mutant ( $n = 159$ ) and wild type tumors ( $n = 674$ ). The colored boxes display interquartile range with median, while the whiskers show  $1.5 \times$  interquartile range, with each sample representing a biological replicate. The lollipop plot below shows Somer's  $D_{xy}$  of the univariate Cox models for the SNVs, with the vertical dotted line indicating  $D_{xy}$  of the ENDORSE model.
- C Ideograms showing mapped regions with copy number gains that are significant in ANOVA analysis of ENDORSE scores with copy number gain status as the grouping variables. Bar plots on the right show false discovery rate-adjusted  $P$ -values from the ANOVA analysis.
- D Lollipop plot showing Somer's  $D_{xy}$  of the univariate Cox models for the copy number gains, with the vertical dotted line indicating  $D_{xy}$  of the ENDORSE model.

for node-negative, early-stage breast cancers to safely waive additional chemotherapy if they are predicted to be at a low risk of recurrence based on genomic signatures. However, these tests have not been proved to be useful in the advanced and metastatic ER<sup>+</sup> breast cancer setting. The default primary treatment for advanced ER<sup>+</sup> breast cancer remains endocrine therapy, despite proven benefits from add-on targeted therapy or potential switch to chemotherapy. Therefore, the key challenge in advanced ER<sup>+</sup> breast cancer is to stratify patients that will likely benefit from continued endocrine therapy and patients that are likely resistant to single-agent endocrine therapy and will benefit from selecting a different treatment strategy (Hart *et al*, 2015).

To address this challenge, we have developed a new prognostic model to predict endocrine response in advanced ER<sup>+</sup> breast cancers. We developed our model using invasive tumors from the METABRIC study that were ER<sup>+</sup> and included node-positive, high-grade tumors. Our model addressed several challenges associated with the development of genomic biomarkers. Since the number of available features to train the genomic models tend to be much larger than the number of available samples ( $P \gg n$ ), it is quite easy to create complex prediction models that contain many predictor variables. Often, such models perform very well in the training datasets, but the performance cannot be replicated in independent test datasets due to overfitting. Several approaches have been proposed to address this issue. Broadly, these can be classified into unsupervised and supervised approaches. The unsupervised approach typically relies on grouping or clustering the samples based on the similarity of gene expression profiles, followed by analysis of association with survival outcomes (Sotiriou *et al*, 2003). Alternatively, a supervised approach is to perform dimensionality reduction prior to modeling the survival outcome or drug response using univariate or multivariate models (Paul *et al*, 2008). Our model utilized the later strategy by using a regularized Cox model for feature selection, effectively reducing the dimensionality of the gene expression data. We further collapsed the genes into a signature and parameterized the final Cox model on the GES of the signatures. The rank-based approach to calculate GES also helped mitigate issues associated with batch effects and differences in methods for transcriptome profiling. We performed an extensive performance evaluation of our model against other published signatures and clinical factors. Consistently, we found that the ENDORSE model was a better predictor than all other models in the training dataset (Fig 2). Moreover, ENDORSE clearly outperformed all other published signatures when they were applied to external validation datasets (Figs 3–6). Our results show that ENDORSE is a highly

accurate and reproducible model that outperforms current approaches to predict endocrine response in metastatic ER<sup>+</sup> breast cancer.

We also explored the biology of the ER<sup>+</sup> tumors to identify possible mechanisms that are commonly shared by high-risk tumors. We found that high-risk tumors showed a consistent enrichment of pathways-associated cell cycle progression and gain of PI3K/MTOR signaling pathways (Fig 7A). In addition, we observed consistent gain of the E2F1 signature, which may be associated with metastatic progression of breast cancers (Hollern *et al*, 2014, 2019). We also observed a loss of Rb1 activity, which has been associated with therapeutic resistance in ER<sup>+</sup> breast cancers (Bosco *et al*, 2007; Witkiewicz & Knudsen, 2014). In the METABRIC cohort, tumors with early mortality (within 5 years of sample collection) showed consistent enrichment of pathways that are known to contribute to epithelial to mesenchymal transition, angiogenesis, and stemness. Enriched pathways, including TGF-beta (Band & Laiho, 2011) and YAP (Ma *et al*, 2022), have been suggested to directly crosstalk and repress the estrogen receptor signaling pathway. Concurrently, MEK (Fujii *et al*, 2011) and PI3K/AKT/MTOR (Ciruelos Gil, 2014) have been implicated as key primary resistance mechanisms in endocrine-resistant ER<sup>+</sup> breast cancers. In contrast, tumors with late mortality did not show enrichment of specific signaling pathways, suggesting that diverse resistance mechanisms may evolve long term and contribute to the late relapse of these tumors. Characterizing the heterogeneity in late relapse tumors would be necessary to understand the diverse mechanisms that lead to the evolution of resistance in these tumors (Zardavas *et al*, 2015).

In addition to common pathway phenotypes shared across high-risk tumors, mutations in the *TP53* tumor suppressor genes were also significant (Fig 7B). Loss-of-function *TP53* variants have long been associated with aggressiveness and chemotherapeutic resistance in hormone-receptor-negative breast cancers (Cattoretti *et al*, 1988; Elledge *et al*, 1993). However, recent studies show that even though *TP53* mutations are infrequent in ER<sup>+</sup> breast cancers, they have similar negative impact on patient outcome as hormone-receptor-negative breast cancers (Ungerleider *et al*, 2018). We also found recurrent copy number gains at chromosomes 8 and 11 to be associated with high-risk tumors (Fig 7C). Amplifications at these loci have been previously associated with aggressive and drug-resistant cancers and included several oncogenes such as *MYC*, *CCND1*, and multiple fibroblast growth factors (Lundgren *et al*, 2008; Baslan *et al*, 2020). The survival models based on genomic alterations were clearly outperformed by ENDORSE; however, the recurrent nature of these alterations in high-risk tumors suggests

that further studies to investigate their role in promoting endocrine resistance are warranted.

Drugs that target CDK4/6 to inhibit cell cycle activation (Hortobagyi *et al*, 2016), PI3K-inhibitors that target tumor with activating *PIK3CA* mutations (André *et al*, 2019), and MTOR-inhibitors that prevent the activation of MTOR signaling and cell proliferation (Baselga *et al*, 2012) have been studied and approved for the treatment of advanced ER<sup>+</sup> breast cancers in combination with endocrine therapy. However, patients first advance on primary endocrine therapy, with or without additional CDK4/6 inhibitors before they are stratified in a different treatment arm. Therefore, identifying high-risk tumors with the ENDORSE model prior to first-line administration of single-agent endocrine therapy could help identify which cancers may be better suited for an add-on regimen or switching to chemotherapy. Thus, future clinical trials applying the ENDORSE model may benefit from early and accurate prediction of endocrine response in advanced, metastatic ER<sup>+</sup> breast cancers. This could ultimately help prolong the survival of patients by stratifying into more appropriate treatment groups.

## Materials and Methods

### Data retrieval and pre-processing

METABRIC microarray gene expression, mutations, copy number alterations, and clinical annotations survival data were retrieved using cBioPortal for cancer genomics ([https://www.cbioportal.org/study/summary?id=brca\\_metabric](https://www.cbioportal.org/study/summary?id=brca_metabric)) (Gao *et al*, 2013; Data ref: Gao *et al*, 2013). Independent validation datasets used in this study were: SET ER/PR microarray gene expression: Gene Expression Omnibus, accession number: GSE124647 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124647>). (Sinn *et al*, 2019; Data ref: Sinn *et al*, 2019), TransCONFIRM microarray gene expression: Gene Expression Omnibus: accession number: GSE76040 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76040>). (Jeselson *et al*, 2016; Data ref: Jeselson *et al*, 2016), ACOSOG Z1031B microarray gene expression: Gene Expression Omnibus, accession number: GSE87411 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87411>). (Ellis *et al*, 2017; Data ref: Ellis *et al*, 2017), and Oxford cohort microarray gene expression: Gene Expression Omnibus, accession number: GSE22219 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22219>). (Buffa *et al*, 2011; Data ref: Buffa *et al*, 2011).

To ensure that gene expression datasets generated using diverse platforms are comparable and appropriate for predictive analyses, we uniformly applied the following pre-processing steps on all datasets. First, we removed genes with zero variance from each dataset and summarized genes with multiple probes by mean expression. Next, we log<sub>2</sub> transformed the expression levels, unless the data was already provided as log-transformed values. Finally, we scaled and standardized the expression levels such that each gene had a mean expression level of zero and a standard deviation of one.

### Inclusion criteria for METABRIC training cohort

The METABRIC cohort contained a total of 2,509 samples. Samples that met all of the following criteria were included in the training

cohort: patients that were ER-positive and HER2-negative based on immunohistochemistry, patients that received hormone therapy but did not receive additional chemotherapy, patients that were either alive or died due to the disease and no other causes, and patients with complete survival and transcriptomic data. After filtering, 833 samples were retained for model construction.

### Empirical signature and ENDORSE model construction

The empirical gene signature was developed using a LASSO-regularized Cox proportional hazards models, with OS as the outcome variable (Tibshirani, 1996). The hazard function in the Cox model is defined as:

$$h_i(t) = h_0(t) \exp(\beta x_i^T)$$

where  $X$  is a set of predictive gene expression features and  $h_0$  is an arbitrary baseline hazard function. The coefficient ( $\beta$ ) for each predictor in the model can be estimated by maximizing the partial likelihood function  $L(\beta)$ , defined as:

$$L(\beta) = \prod_i \frac{\exp(\beta x_{j(i)}^T)}{\sum_{l \in R_i} \exp(\beta x_l^T)}$$

where  $R_i$  is the set of indices of observations failing (events) at time  $t_i$ . In the LASSO Cox model, the regularized coefficient is obtained by adding a penalty parameter  $\lambda$  to the log of the likelihood function.

$$\hat{\beta} = \min -\frac{1}{N} l(\beta) + \lambda \|\beta\|_1$$

where  $l(\beta) = \log L(\beta)$ . The  $\lambda$  penalty parameter was determined using 10-fold cross-validation implemented in R package *glmnet* (Friedman *et al*, 2010; Simon *et al*, 2011). The optimal  $\lambda$  minimized model deviance.

We applied the model in a repeated (50 × 10-fold) cross-validation framework. In each iteration, a set of “seed genes” or features with positive coefficients in the regularized Cox model at a  $\lambda$  equal to one standard error from the minimum model deviance were identified. The seed genes were expanded to a redundant correlation network by adding all genes in the training transcriptome dataset that had Pearson’s correlation > 0.75 with any of the seed genes. Across all iterations, we identified the common set of features that were present in at least 50% of the correlation networks and defined this set of features as the empirical signature.

The ENDORSE model was defined as the hazard’s ratio of the Cox proportional hazards model fitted on OS data of the training cohort with two components: GES for the empirical gene signature and GES for the hallmark estrogen early response signature.

$$h(t) = h_0(t) \times \exp(\beta_{emp} GES_{emp} + \beta_{er} GES_{er}).$$

where *emp* represents the empirical signature and *er* represents the estrogen response signature.

For each signature, the GES were calculated for individual samples using the GSVA package for R (Hänzelmann *et al*, 2013) using the ssGSEA method (Barbie *et al*, 2009). The ssGSEA method

is especially well suited for calculating GES from data from diverse platforms, as the method uses empirical cumulative distribution functions of the ranked values of probe or gene expression to calculate enrichment scores. By using ranked expression values, this method mitigates challenges encountered with applying predictive models on datasets with vastly different distribution profiles than the training data. The final parameters for the ENDORSE model were obtained by fitting the model to the full training cohort of 833 samples, resulting in  $\beta_{\text{emp}} = 1.54$  and  $\beta_{\text{er}} = -2.72$ .

### Models based on external signatures and clinical factors

Clinical features such as tumor grade and mutation count, along with scores from PAM50 and IntClust analyses, were obtained directly from the METABRIC clinical annotations accompanying the transcriptome data and were directly utilized in univariate Cox models. Proliferation index based on the metaPCNA signature was calculated using the R-package ProliferativeIndex (Ramaker *et al*, 2017). We replicated the signatures and algorithms developed in the TransCONFIRM, SET ER/PR, and 21-gene prognostic signature studies by following the methods described in the respective studies. The TransCONFIRM signature composed of 37 genes was implemented by performing hierarchical clustering of the gene expression data using these genes and cutting the tree ( $k = 2$ ) to stratify samples in high or low TransCONFIRM score categories. The SET signature was implemented by calculating (the average expression of the 18-genes in the signature) – (the expression of 10 house-keeping genes) + 2. The 21-gene signature (ODX) score was calculated by following the unscaled risk score calculation reported by the study. *BAG1* transcript was missing from the METABRIC cohort and was not included in the unscaled score calculation. Since this transcript was uniformly missing on all samples, the relative risk scores could be compared across the samples.

### ENDORSE model performance evaluation in METABRIC

The predictive ability of ENDORSE and various other models were evaluated in the METABRIC training dataset using two approaches. In the first approach, we split the METABRIC cohort (50/50) into a training and hold-out test or validation subset. We applied the LASSO-regularized Cox proportional hazards model on only the training subset to obtain an empirical signature using 10-fold cross-validation. Next, we calculated GES of the training subset-derived empirical signature and the estrogen response signature and fit a Cox proportional hazards model on the OS data of the training subset. Using this model, we predicted the OS of the held-out test subset. To compare the performance of the model against other independent signatures and clinical features associated with each sample, we performed a bootstrap resampling analysis of the Cox regression models. The resampling was repeated 150 times for each model and a Somer's  $D_{xy}$  rank correlation was calculated in each repeat. A final bias-corrected index of Somer's  $D_{xy}$  was obtained as a measure of the model's predictive ability. The bootstrap resampling and calculations of the Somer's  $D_{xy}$  were performed using the R package "rms." Models based on SNVs and CNAs significantly associated with ENDORSE scores were also evaluated by obtaining Somer's  $D_{xy}$  rank correlation metric of the univariate Cox model. In the second approach, we calculated the ENDORSE scores of the full cohort and

compared each of the external signatures and clinical feature models using Vuong's (Vuong, 1989) partial likelihood ratio test for non-nested Cox regression models. The non-nested partial likelihood ratio tests were implemented using the R package "nonnestcox" (<https://github.com/thomashielscher/nonnestcox/>). The individual components of the ENDORSE model were compared to the full model using likelihood ratio tests for nested Cox models.

### Model validation in independent datasets

We compared the predictive performance of ENDORSE in multiple independent datasets. First, we integrated the training (METABRIC) and test (independent validation) datasets to perform batch correction using the ComBat function of the R package "sva" (Leek *et al*, 2012). Next, we calculated the GES for the ENDORSE signatures in the batch corrected METABRIC and independent validation gene expression datasets. Then, the parameters of the ENDORSE Cox model were calculated on the batch-corrected training dataset with the OS information as the response variable. Finally, the parameterized Cox model was applied to the independent validation dataset to obtain a predicted risk score.

In case of the Oxford and SET ER/PR cohort, we used the predicted ENDORSE risk scores to stratify the patients into risk categories, with an ENDORSE score  $\geq 2$  representing the high-risk group,  $\leq 1$  representing the low-risk group, and other intermediate values representing the medium risk group. We compared the significance of stratification of recurrence-free survival curves in the Oxford cohort and both OS and PFS curves in the SET ER/PR cohort based on ENDORSE, SET, and TransCONFIRM scores using log-rank tests. Further, we compared individual strata using Cox models and partial likelihood ratio tests to compare non-nested Cox models.

For the TransCONFIRM and ACOSOG cohorts, we compared the ENDORSE (with an ENDORSE score  $\geq 2$  representing the high-risk group,  $\leq 1$  representing the low-risk group and other intermediate values representing the medium risk group), SET, and TransCONFIRM predictions with reported clinical variables, such as percentage of cells positive for Ki67 at the end of treatment and clinical outcomes using generalized linear models for continuous outcome variables or one-way ANOVA analysis for categorical outcomes. *Post-hoc* analyses comparing significance of difference in means between different ENDORSE risk groups were obtained using Tukey's HSD test.

### Biological features associated with ENDORSE scores

To determine the possible biological mechanisms associated with the emergence of endocrine resistance and high ENDORSE risk scores, we evaluated the enrichment scores of various biological pathway and oncogenic signatures across the training and independent validation cohorts. We used the ssGSEA method to obtain GES for hallmark, curated (C2), and oncogenic signature (C6) gene sets from the molecular signatures database (Liberzon *et al*, 2011). For each signature, we fitted a generalized additive model against the predicted ENDORSE score to obtain significance of the fit,  $R^2$ , and proportion of variance explained by the model. None of the curated signatures were significant in the METABRIC analyses and were excluded from further consideration in the independent validation

datasets. Pathways enriched in early mortality (< 5 years OS) vs. late mortality (> 5 years OS) ER<sup>+</sup> METABRIC tumors with an ENDORSE risk score > 1 were determined using Welch's *t*-test.

Gene-level somatic SNV and CNV analyses were performed using data reported by the METABRIC study. SNVs were retained based on a mutation frequency of  $\geq 5$  across all samples and limited to genes that are known cancer-related genes, obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC) cancer gene census (Sondka et al, 2018; Tate et al, 2019). Pathogenic *PIK3CA* variants associated with PI3K inhibitor sensitivity were obtained from the drug labels for alpelisib based on the SOLARI clinical trial (André et al, 2019). Significant SNVs and CNVs were obtained using a one-way ANOVA analysis of the ENDORSE scores with mutation status as the factor.

## Data availability

Scripts for developing ENDORSE, validation using METABRIC and other independent clinical trials datasets, and analysis to determine significant pathways and somatic variants associated with high ENDORSE score tumors are available on GitHub: <https://github.com/aritronath/ENDORSE>. The gene set components of the ENDORSE model and training datasets are also available on the GitHub repository under releases.

**Expanded View** for this article is available online.

## Acknowledgements

Funding for this research was provided by the National Cancer Institute of the National Institutes of Health through the U54 grant 1U54CA209978. Synopsis figure and Fig 1A were created with BioRender.com.

## Author contributions

**Aritro Nath:** Conceptualization; Data curation; Software; Formal analysis; Validation; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing. **Adam L Cohen:** Conceptualization; Resources; Supervision; Funding acquisition; Investigation; Methodology; Project administration; Writing—review and editing. **Andrea H Bild:** Conceptualization; Resources; Supervision; Funding acquisition; Investigation; Methodology; Project administration; Writing—review and editing.

In addition to the CRediT author contributions listed above, the contributions in detail are:

Conceptualization, Methodology, Investigation—AN, ALC, AHB; Data Curation, Formal Analysis, Software, Visualization, Validation—AN; Funding Acquisition, Resources, Supervision, Project Administration—ALC, AHB; Writing—original draft—AN; Writing—review & editing—AN, ALC, AHB.

## Disclosure and competing interests statement

The authors AN and AHB are listed as co-inventors on a patent application (filed, pending) covering the ENDORSE biomarker model described in this manuscript.

## References

Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, Hayes DF, Lakhani SR, Chavez-MacGregor M, Perlmutter J et al (2020)

- Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. *JCO* 38: 1346–1366
- André F, Ciruelos E, Rubovszky G, Campone M, Loibl S, Rugo HS, Iwata H, Conte P, Mayer IA, Kaufman B et al (2019) Alpelisib for *PIK3CA*-mutated, hormone receptor-positive advanced breast cancer. *N Engl J Med* 380: 1929–1940
- Band AM, Laiho M (2011) Crosstalk of TGF- $\beta$  and estrogen receptor signaling in breast cancer. *J Mammary Gland Biol Neoplasia* 16: 109–115
- Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C et al (2009) Systematic RNA interference reveals that oncogenic *KRAS*-driven cancers require *TBK1*. *Nature* 462: 108–112
- Baselga J, Campone M, Piccart M, Burris HA, Rugo HS, Sahmoud T, Noguchi S, Gnant M, Pritchard KI, Lebrun F et al (2012) Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N Engl J Med* 366: 520–529
- Baslan T, Kendall J, Volyanskyy K, McNamara K, Cox H, D'Italia S, Ambrosio F, Riggs M, Rodgers L, Leotta A et al (2020) Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing. *eLife* 9: e51480
- Bosco EE, Wang Y, Xu H, Zilfou JT, Knudsen KE, Aronow BJ, Lowe SW, Knudsen ES (2007) The retinoblastoma tumor suppressor modifies the therapeutic response of breast cancer. *J Clin Invest* 117: 218–228
- Boutros PC (2015) The path to routine use of genomic biomarkers in the cancer clinic. *Genome Res* 25: 1508–1513
- Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, Taylor M, Harris AL, Ragoussis J (2011) microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res* 71: 5635–5645
- Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, Taylor M, Harris AL, Ragoussis J (2011) Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22219>). [DATASET]
- Cardoso F, Paluch-Shimon S, Senkus E, Curigliano G, Aapro MS, André F, Barrios CH, Bergh J, Bhattacharyya GS, Biganzoli L et al (2020) 5th ESO-ESMO international consensus guidelines for advanced breast cancer (ABC 5). *Ann Oncol* 31: 1623–1649
- Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, Pierga J-Y, Brain E, Causeret S, DeLorenzi M, Glas AM, Golfopoulos V (2016) 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* 375(8): 717–729
- Cattoretti G, Rilke F, Andreola S, D'Amato L, Delia D (1988) P53 expression in breast cancer. *Int J Cancer* 41: 178–183
- Ciruelos Gil EM (2014) Targeting the PI3K/AKT/mTOR pathway in estrogen receptor-positive breast cancer. *Cancer Treat Rev* 40: 862–871
- Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y et al (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486: 346–352
- Dawson S-J, Rueda OM, Aparicio S, Caldas C (2013) A new genome-driven integrated classification of breast cancer and its implications. *EMBO J* 32: 617–628
- Duffy MJ, Harbeck N, Nap M, Molina R, Nicolini A, Senkus E, Cardoso F (2017) Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *Eur J Cancer* 75: 284–298
- Elledge RM, Fuqua SAW, Clark GM, Pujol P, Allred DC, McGuire WL (1993) Prognostic significance of p53 gene alterations in node-negative breast cancer. *Breast Cancer Res Treat* 26: 225–235

- Ellis MJ, Suman VJ, Hoog J, Goncalves R, Sanati S, Creighton CJ, DeSchryver K, Crouch E, Brink A, Watson M et al (2017) Ki67 proliferation index as a tool for chemotherapy decisions during and after neoadjuvant aromatase inhibitor treatment of breast cancer: results from the American college of surgeons oncology group Z1031 trial (Alliance). *J Clin Oncol* 35: 1061–1069
- Ellis MJ, Suman VJ, Hoog J, Goncalves R, Sanati S, Creighton CJ, DeSchryver K, Crouch E, Brink A, Watson M et al (2017) Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87411>). [DATASET]
- Friedman JH, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33: 1–22
- Fujii S, Tokita K, Wada N, Ito K, Yamauchi C, Ito Y, Ochiai A (2011) MEK–ERK pathway regulates EZH2 overexpression in association with aggressive breast cancer subtypes. *Oncogene* 30: 4118–4128
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E et al (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6: pl1
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E et al (2013) cBioPortal ([https://www.cbioportal.org/study/summary?id=brca\\_metabric](https://www.cbioportal.org/study/summary?id=brca_metabric)). [DATASET]
- Gradishar WJ, Anderson BO, Balassanian R, Blair SL, Burstein HJ, Cyr A, Elias AD, Farrar WB, Forero A, Giordano SH et al (2017) NCCN guidelines insights: breast cancer, version 1.2017. *J Natl Compr Canc Netw* 15: 433–451
- Gradishar WJ, Anderson BO, Abraham J, Aft R, Agnese D, Allison KH, Blair SL, Burstein HJ, Dang C, Elias AD et al (2020) Breast cancer, version 3.2020, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 18: 452–478
- Hänzelmann S, Castelo R, Guinney J (2013) GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* 14: 7
- Hart CD, Migliaccio I, Malorni L, Guarducci C, Biganzoli L, Di Leo A (2015) Challenges in the management of advanced, ER-positive, HER2-negative breast cancer. *Nat Rev Clin Oncol* 12: 541–552
- Harvey JM, Clark GM, Osborne CK, Allred DC (1999) 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* 17: 1474–1481
- Hollern DP, Honeysett J, Cardiff RD, Andrechek ER (2014) The E2F transcription factors regulate tumor development and metastasis in a mouse model of metastatic breast cancer. *Mol Cell Biol* 34: 3229–3243
- Hollern DP, Swiatnicki MR, Rennhack JP, Misk SA, Matson BC, McAuliff A, Gallo KA, Caron KM, Andrechek ER (2019) E2F1 drives breast cancer metastasis by regulating the target gene FGF13 and altering cell migration. *Sci Rep* 9: 10718
- Hortobagyi GN, Stemmer SM, Burris HA, Yap Y-S, Sonke GS, Paluch-Shimon S, Campone M, Blackwell KL, André F, Winer EP et al (2016) Ribociclib as first-line therapy for HR-positive, advanced breast cancer. *N Engl J Med* 375: 1738–1748
- Jeselsohn R, Barry WT, Migliaccio I, Biagioni C, Zhao J, De Tribolet-Hardy J, Guarducci C, Bonechi M, Laing N, Winer EP et al (2016) TransCONFIRM: identification of a genetic signature of response to fulvestrant in advanced hormone receptor-positive breast cancer. *Clin Cancer Res* 22: 5755–5764
- Jeselsohn R, Barry WT, Migliaccio I, Biagioni C, Zhao J, De Tribolet-Hardy J, Guarducci C, Bonechi M, Laing N, Winer EP et al (2016) Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76040>). [DATASET]
- Kohler BA, Sherman RL, Howlander N, Jemal A, Ryerson AB, Henry KA, Boscoe FP, Cronin KA, Lake A, Noone A-M et al (2015) Annual report to the nation on the status of cancer, 1975–2011, featuring incidence of breast cancer subtypes by Race/Ethnicity, poverty, and state. *J Natl Cancer Inst* 107(6): 1975–2011
- Krop I, Ismaila N, Andre F, Bast RC, Barlow W, Collyar DE, Hammond ME, Kuderer NM, Liu MC, Mennel RG et al (2017) Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American society of clinical oncology clinical practice guideline focused update. *JCO* 35: 2838–2847
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28: 882–883
- Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P (2015) The molecular signatures database hallmark gene set collection. *Cell Syst* 1: 417–425
- Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27: 1739–1740
- Lundgren K, Holm K, Nordenskjöld B, Borg A, Landberg G (2008) Gene products of chromosome 11q and their association with CCND1 gene amplification and tamoxifen resistance in premenopausal breast cancer. *Breast Cancer Res* 10: R81
- Ma S, Tang T, Probst G, Konradi A, Jin C, Li F, Silvio Gutkind J, Fu X-D, Guan K-L (2022) Transcriptional repression of estrogen receptor alpha by YAP reveals the Hippo pathway as therapeutic target for ER<sup>+</sup> breast cancer. *Nat Commun* 13: 1061
- McAndrew NP, Finn RS (2020) Management of ER positive metastatic breast cancer. *Semin Oncol* 47: 270–277
- McDonnell DP, Wardell SE (2010) The molecular mechanisms underlying the pharmacological actions of ER modulators: implications for new drug discovery in breast cancer. *Curr Opin Pharmacol* 10: 620–628
- Musgrove EA, Sutherland RL (2009) Biological determinants of endocrine resistance in breast cancer. *Nat Rev Cancer* 9: 631–643
- Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, Davies SR, Snider J, Stijleman IJ, Reed J et al (2010) A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res* 16: 5222–5232
- Paik S, Shak S, Tang C, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T et al (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351: 2817–2826
- Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z et al (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27: 1160–1167
- Paul D, Bair E, Hastie T, Tibshirani R (2008) “Preconditioning” for feature selection and regression in high-dimensional problems. *Ann Statist* 36: 1595–1618
- Pereira B, Chin S-F, Rueda OM, Vollan H-K, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut S-J et al (2016) The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun* 7: 1–16
- Ramaker RC, Lasseigne BN, Hardigan AA, Palacio L, Gunther DS, Myers RM, Cooper SJ (2017) RNA sequencing-based cell proliferation analysis across 19 cancers identifies a subset of proliferation-informative cancers with a common survival signature. *Oncotarget* 8: 38668–38681
- Rugo HS, Rumble RB, Macrae E, Barton DL, Connolly HK, Dickler MN, Fallowfield L, Fowble B, Ingle JN, Jahanzab M et al (2016) Endocrine therapy for hormone receptor-positive metastatic breast cancer: American society of clinical oncology guideline. *JCO* 34: 3069–3103

- Simon N, Friedman JH, Hastie T, Tibshirani R (2011) Regularization paths for cox's proportional hazards model via coordinate descent. *J Stat Softw* 39: 1–13
- Sinn BV, Fu C, Lau R, Litton J, Tsai T-H, Murthy R, Tam A, Andreopoulou E, Gong Y, Murthy R et al (2019) SETER/PR: a robust 18-gene predictor for sensitivity to endocrine therapy for metastatic breast cancer. *NPJ Breast Cancer* 5: 16
- Sinn BV, Fu C, Lau R, Litton J, Tsai T-H, Murthy R, Tam A, Andreopoulou E, Gong Y, Murthy R et al (2019) Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124647>). [DATASET]
- Smith IE, Dowsett M (2003) Aromatase inhibitors in breast cancer. *N Engl J Med* 348: 2431–2442
- Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA (2018) The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* 18: 696–705
- Sotiriou C, Neo S-Y, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *PNAS* 100: 10393–10398
- Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, Geyer CE, Dees EC, Goetz MP, Olson JA et al (2018) Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med* 379: 111–121
- Spoerke JM, Gendreau S, Walter K, Qiu J, Wilson TR, Savage H, Aimi J, Derynck MK, Chen M, Chan IT et al (2016) Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant. *Nat Commun* 7: 1–10
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clin* 71: 209–249
- Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E et al (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 47: D941–D947
- Taylor JMG, Ankerst DP, Andridge RR (2008) Validation of biomarker-based risk prediction models. *Clin Cancer Res* 14: 5977–5983
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Roy Stat Soc: Ser B (Methodol)* 58: 267–288
- Ungerleider NA, Rao SG, Shahbandi A, Yee D, Niu T, Frey WD, Jackson JG (2018) Breast cancer survival predicted by TP53 mutation status differs markedly depending on treatment. *Breast Cancer Res* 20: 115
- Venet D, Dumont JE, Detours V (2011) Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol* 7: e1002240
- Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57: 307
- Waks AG, Winer EP (2019) Breast cancer treatment: a review. *JAMA* 321: 288
- Witkiewicz AK, Knudsen ES (2014) Retinoblastoma tumor suppressor pathway in breast cancer: prognosis, precision medicine, and therapeutic interventions. *Breast Cancer Res* 16: 207
- Witten DM, Tibshirani R (2010) Survival analysis with high-dimensional covariates. *Stat Methods Med Res* 19: 29–51
- Zardavas D, Irrthum A, Swanton C, Piccart M (2015) Clinical management of breast cancer heterogeneity. *Nat Rev Clin Oncol* 12: 381–394



**License:** This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.