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

Aritro Nath[1], Adam L. Cohen[2] and Andrea H. Bild[1*]

[1] Department of Medical Oncology and Therapeutics, City of Hope Comprehensive Cancer Institute, Monrovia, California

[2] Neuro Oncology Program, Inova Schar Cancer Institute, Fairfax, Virginia

*Correspondence: Andrea H. Bild, Ph.D.
Professor, Department of Medical Oncology and Therapeutics
City of Hope Comprehensive Cancer Institute
1218 S Fifth Ave, Monrovia, CA 91016
(626) 218-6052
abild@coh.org

Conflict of interest statement: The authors declare no potential conflicts of interest.

ABSTRACT

Endocrine therapy remains the primary treatment for advanced and metastatic estrogen receptor-positive (ER+) breast cancers. Patients who progress on endocrine therapy may benefit from add-on treatment targeting the PI3K/MTOR signaling pathways or by switching to chemotherapy. However, these options are only available after progression on first-line treatment with endocrine therapy. In the absence of reliable prognostic tests for advanced ER+ breast cancers, it is currently not possible to stratify patients into pertinent treatment arms at the baseline. To address this, we have developed a low-dimensional endocrine response signature (ENDORSE) model for advanced ER+ breast cancers. The ENDORSE model was developed using the baseline tumor transcriptomes and long-term survival outcomes of >800 invasive ER+ breast cancers and predicts the risk of death on endocrine therapy. ENDORSE was validated in multiple clinical trial datasets for endocrine therapy response in metastatic ER+ breast cancers and demonstrated superior predictive performance over clinical factors and published gene signatures. Our results show that ENDORSE is a reproducible and accurate prognostic model for endocrine therapy response in advanced and metastatic ER+ breast cancers.

INTRODUCTION

Breast cancer is the most common form of cancer globally, with more than two million cases diagnosed in 2020[1]. 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+/HER2-)[2,3]. The primary systemic therapy for ER+/HER2- breast cancer is endocrine therapy, which counters the growth of tumors by targeting their dependency on estrogen signaling[4]. This includes 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 [5,6]. Endocrine therapy substantially reduces the risk of recurrence within 5-years, although chemotherapy may be recommended for some patients with 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[7,8]. These biomarkers are have been validated and recommended for clinical use only in early stage, node-negative cancers based on guidelines from the American Society of Clinical Onoclogy and European Group on Tumor Markers [9,10].

Locally advanced and metastatic ER+ 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[11]. Patients may benefit from the addition of targeted inhibitor against the mTOR or PI3K pathways[12,13] or switching to chemotherapy[11]. However, these treatment options are recommended for consideration only upon progression on endocrine therapy, according to the American Society for Clinical Oncology [14], National Comprehensive Cancer Network [15,16] and European Society for Medical Oncology [17] clinical practice guidelines. Therefore, the ability to predict the potential benefit from first-line endocrine therapy may be crucial for locally advanced and metastatic ER+ 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+ breast cancers. To address this limitation, a few attempts have been made so far to develop a genomic signature of endocrine response in ER+ metastatic breast cancers (ER+ MBC)[18,19]. The TransCONFIRM trial evaluated the transcriptomes of 112 ER+/HER2- MBCs and identified a set of 37 genes that were associated with progression-free survival (PFS) of patients receiving fulvestrant[19]. Another study analyzed the transcriptomes of 140 ER+/HER2- MBC on endocrine therapy to develop SET ER/PR, an 18-gene predictive score for endocrine therapy sensitivity [18]. 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[20]. 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[21,22].

Here we developed ENDORSE: a low-dimensional expression-based prognostic model for endocrine therapy and systemically tested its performance and predictive ability in multiple-independent clinical trials against other diagnostic models and genomic signatures. ENDORSE was developed and trained using the tumor transcriptomes and overall survival (OS) of more than 800 ER+ breast cancers on endocrine therapy[23,24]. 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+ breast cancer. Our results show that ENDORSE reproducibly predicts endocrine response in independent validation clinical studies, and consistently outperforms all other models of endocrine therapy response in metastatic ER+ breast cancers, clinical factors, and proliferation signatures.

RESULTS

Developing a low-dimensional prognostic model for endocrine therapy

We developed a two-component prognostic model for endocrine therapy response using the tumor transcriptomes and long-term survival outcomes of 833 ER+/HER2-

tumors that received endocrine therapy[23,24] (Table 1, Figure 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) and a curated gene signature defining response to estrogen[25]. Figure 1a outlines the inclusion criteria for the training dataset, method for developing the empirical gene signature and the final Cox proportional hazards model based on the gene set enrichments 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 cross-validation iterations (Supplementary Figure 1, Supplementary Table 1). 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 (Figure 1b). The coefficients for the endocrine response, or ENDORSE, model was calculated using the training cohort, resulting in ENDORSE = 1.54 x (empirical signature GES) – (2.72 x estrogen response GES). The ENDORSE model could also be used to stratify the tumors based on predicted risk, for example by setting a threshold of ≥2-fold relative risk of death as "high-risk" and ≤1 risk as "low-risk", resulting in significant differences in the Kaplan-Meier survival curves across the strata ($P = 3 \times 10^{-14}$) (Figure 1c).

127

Table 1: Training data patient characteristics

Variable	**Mean**	**95% C.I.**	**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)			

128

129

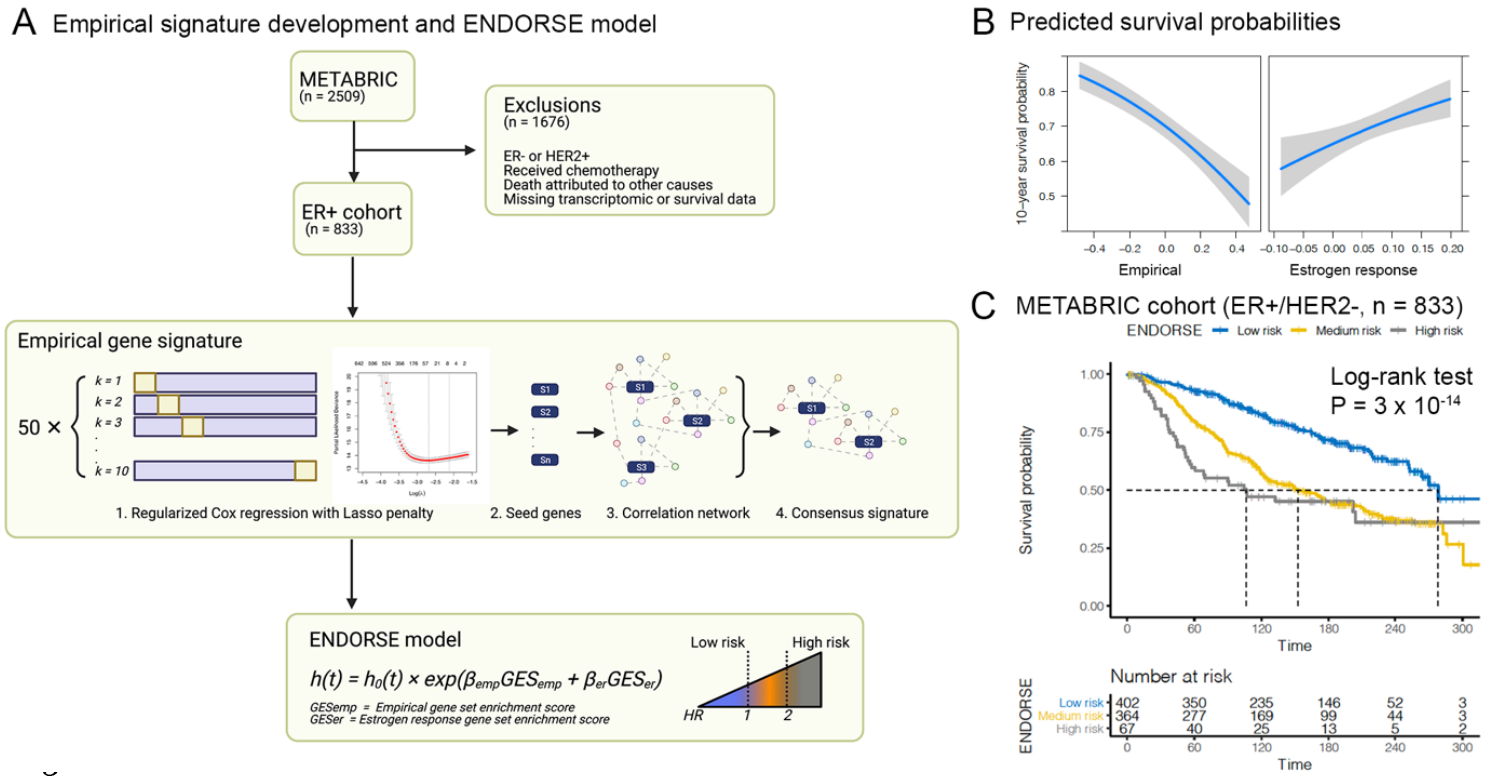


Figure 1: ENDORSE model development in METABRIC. A. Inclusion criteria and overall schematic of ENDORSE model development. Samples for training were selected based on ER+ status and excluded from the analysis if they were either HER2+, received chemotherapy in addition to hormone therapy, died due to 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 833

144 ER+/HER2- METABRIC breast cancers based on a Cox proportional hazards model of
145 gene signature enrichment scores of the empirical and estrogen response signatures as
146 predictor variables. C. Kaplan-Meier curves and risk tables of METABRIC ER+/HER2-
147 tumors stratified by ENDORSE. The tumors were stratified according to an ENDORSE
148 risk score (hazard ratio) threshold of ≥2 to define high-risk, ≤1 as low risk and all other
149 intermediate values as medium risk.
150

Internal performance evaluation, comparison with clinical covariates and published breast cancer signatures

We performed bootstrap resampling analyses to validate the Cox model in the training dataset (Figure 2a) and performed likelihood ratio tests (Figure 2b) to compare with other univariate prognostic models including clinical factors, proliferation index and published prognostic signatures for ER+ breast cancers. First, we compared the ENDORSE model to the univariate models based on the individual components of ENDORSE, i.e., the empirical signature and estrogen response signature. The ENDORSE model (Somer's D or D_{xy} = 0.301) was a better fit than the empirical signature (D_{xy} = 0.296, P = 1.09 x 10^{-3}) and the estrogen response signature (D_{xy} = 0.141, P = 3.93 x 10^{-14}) univariate models.

We then compared ENDORSE with clinical factors, such as tumor grade and mutation burden. The ENDORSE model performed better than both tumor grade (D_{xy} = 0.141, P = 2.08 x 10^{-3}) and mutation count (D_{xy} = 0.059, P = 9.76 x 10^{-6}). 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[26,27]. Again, the ENDORSE model performed significantly better than the proliferation index (D_{xy} = 0.235, P = 4.42 x 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

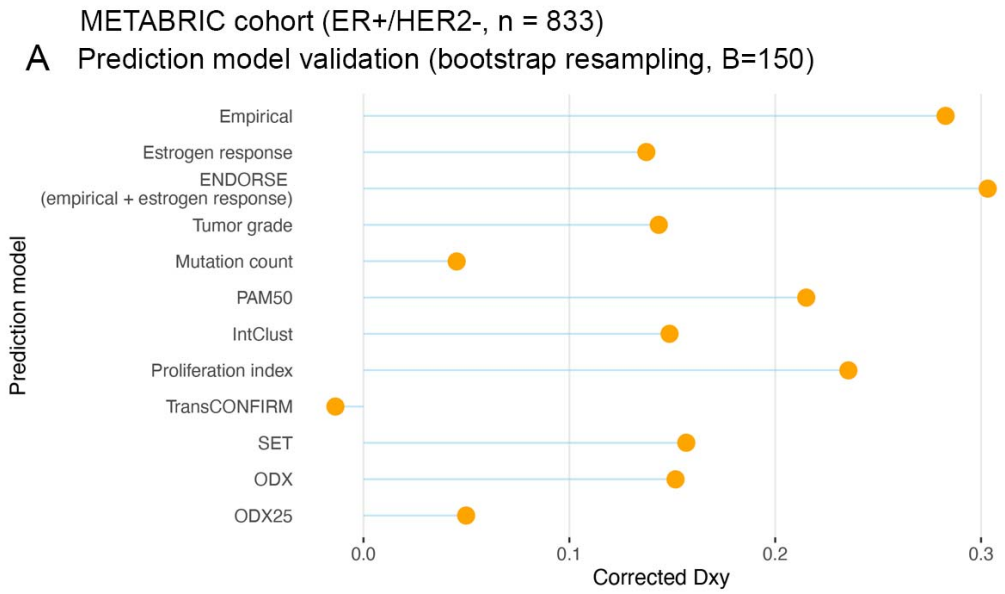
signature that was previously reported to be a better prognostic tool for ER+ breast cancers on endocrine therapy than clinical factors, such as histopathological classification and tumor grade[28]. A genomic classifier, IntClust, that developed by the METABRIC consortium authors and trained on the same training dataset was also included in this comparison[29]. The PAM50 model (D_{xy} = 0.220) performed better than IntClust (D_{xy} = 0.153), however the ENDORSE model outperformed both PAM50 (P = 0.033) and IntClust (P = 0.02) models.

Two previous clinical trials evaluating endocrine therapy response in metastatic ER+ 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+ breast cancers on fulvestrant[19]. 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 two clusters. We referred to resultant clusters as the 'TransCONFIRM' score. The TransCONFIRM score applied to the METABRIC dataset performed poorly (D_{xy} = -.002), suggesting that the signature performed no better than a random set of genes and was unsurprisingly outperformed by ENDORSE ($P = 9.73 \times 10^{-6}$).

The second signature (SET ER/PR) was developed using tumor transcriptomes of metastatic ER+ breast cancers on endocrine therapy[18]. 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 original study

and referred to the resultant score as 'SET'. The SET score (D_{xy} = 0.152) performed better than TransCONFIRM; however, it was also easily outperformed by ENDORSE (P = 3.53 x 10^{-5}).

Finally, we calculated a surrogate based on the published formula for the 21-gene prognostic signature approved for early-stage, node-negative ER+ breast cancers [30]. We referred to this score as ODX. We also compared a classifier that stratified samples based on 25th percentile of ODX score as a proxy for the latest risk stratification threshold for this signature[8], and referred to this score as ODX25. We found that the ODX model (D_{xy} = 0.159) was comparable to other published signatures like the SET score but the stratified ODX25 score performed poorly (D_{xy} = 0.056). Again, the ENDORSE model performed significantly better than both ODX (P = 6.15 x 10^{-5}) and ODX25 (P = 1.32 x 10^{-5}) models. These results show that ENDORSE is significantly better prognostic model than available gene signatures, clinical factors and proliferation index for endocrine therapy in the METABRIC dataset.



B Model comparisons

Model 1	***Model 2***	***Variance test*** *H1: Model 1 and Model 2 are distinguishable*	***Likelihood ratio test (nested)*** *H1A: Full model fits better than reduced model*	***Partial likelihood ratio test (non-nested)*** *H1A: Model 1 fits better than Model 2*
ENDORSE	Empirical	1.47×10^{-3}	1.09×10^{-3}	
	Estrogen response	3.83×10^{-6}	3.93×10^{-14}	
	METABRIC tumor characteristics and signatures			
	Tumor grade	$< 2 \times 10^{-16}$		2.08 x 10-3
	Mutation count	$< 2 \times 10^{-16}$		9.76 x 10-6
	PAM50	2.22 x 10-8		0.033
	IntClust	$< 2 \times 10^{-16}$		0.02
	Proliferation index	3.73×10^{-5}		4.42×10^{-5}
	External signatures			
	TransCONFIRM	1.67×10^{-7}		9.73×10^{-6}
	SET	$< 2 \times 10^{-16}$		3.53×10^{-5}
	ODX	2.18×10^{-7}		6.15×10^{-5}
	ODX25	1.49×10^{-6}		1.32×10^{-5}

Figure 2: Model evaluation and comparison with other predictors a. Lollipop plots displaying corrected Somer's D_{xy} indices of ENDORSE and various other univariate Cox proportional hazards models. The indices were calculated using 150-fold bootstrap resampling of the training dataset. b. Table comparing the ENDORSE model with various other univariate Cox models using partial likelihood ratio tests. The comparison

between the nested ENDORSE model and its two components were performed using a likelihood ratio test, while other non-nested univariate models were compared using a partial likelihood ratio test.

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+ tumors from three 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 also across other independent datasets.

The TransCONFIRM trial evaluated fulvestrant response in 112 advanced metastatic ER+ breast cancers previously treated with an antiestrogen[19]. 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 (Figure 3). The percentage of cells positive of Ki67 were significantly correlated with the ENDORSE estimated risk ($P = 2.5 \times 10^{-5}$) (Figure 3a), while stratification of the patients

242 based on the risk thresholds also showed significant difference in Ki67 staining
243 percentage between the strata (P = 1.2 x 10-3) (Figure 3b). However, the SET score
244 was not correlated with Ki67 staining (P = 0.3) (Figure 3c). The TransCONFIRM score
245 that was developed on this dataset was significant (P = 0.05) but performed worse than
246 the ENDORSE score trained on an independent dataset.
247
248 Next, we evaluated the performance of the signatures in the SET ER/PR cohort. This
249 clinical trial reported the PFS and OS of 140 stage IV ER+ metastatic breast cancers on
250 endocrine therapy. We compared the survival curves of the patients by stratifying them
251 based on the ENDORSE predicted risk, median SET scores, as described in the original
252 study, and the TransCONFIRM score. The stratification based on ENDORSE (Figure
253 4a) and SET (Figure 4b) scores both resulted in significant differences in the survival
254 curves (ENDORSE P = $2 \times 10^{-4,}$ SET P = 3×10^{-3}). However, the TransCONFIRM score
255 (Figure 4c) was not significant (P = 0.9). Similarly, we observed that ENDORSE (Figure
256 4d) and SET (Figure 4e) scores both resulted in significant differences in the PFS
257 curves (ENDORSE P = $1 \times 10^{-6,}$ SET P = 5×10^{-3}), while TransCONFIRM was not
258 significant (P = 0.2). Additionally, we compared the model fits using partial likelihood
259 ratio tests. The SET model that was trained using the same dataset was not a better fit
260 than the ENDORSE mode (OS P = 0.667, PFS P=0.258). The ENDORSE model was a
261 better fit than the TransCONFIRM model in each case (OS P = 0.046, PFS P=0.038).
262 In addition to the two metastatic ER+ breast cancer trials, we also evaluated the
263 performance of the signatures examined data from the ACOSOG Z1031B clinical trial
264 which evaluated neoadjuvant aromatase inhibitor (AI) treatment in Stage II or III ER+

265 breast cancers[31]. This study reported percentage of Ki67 staining both at the study
266 baseline and at the end of treatment (2-4 weeks). We compared the percentage of Ki67
267 positive cells across cancers stratified by the ENDORSE score and found significant
268 difference across the classes at both the baseline ($P = 4.9 \times 10^{-9}$) and at the end of
269 treatment ($P = 3 \times 10^{-18}$) (Figure 5a). Similarly, the continuous ENDORSE scores were
270 significantly correlated with both the baseline ($P = 3.3 \times 10^{-15}$) and end of treatment ($P =$
271 1.1×10^{-17}) Ki67 percentage (Figure 5b). The ENDORSE scores were also significantly
272 higher in the tumors that were classified as resistant based on clinical response ($P = 4.6$
273 $\times 10^{-6}$) (Figure 5c). In this cohort, the SET score was also significantly correlated with
274 Ki67 percentage at the baseline ($P = 2.8 \times 10^{-5}$) and end of treatment ($P = 2.2 \times 10^{-4}$)
275 (Figure 5d), with significant difference in the SET scores between the resistant and
276 sensitive tumors ($P = 0.05$) (Figure 5e). The transCONFIRM scores were not significant
277 at the baseline ($P = 0.5$) and end of treatment (Figure 5f) or between resistant and
278 sensitive tumors ($P = 0.7$) (Figure 5g).
279
280 In addition to the endocrine therapy trials in ER+ breast cancer, we also applied the
281 ENDORSE risk estimates to stratify 429 ER-negative METABRIC breast cancers as
282 negative controls. Kaplan-Meier analyses show no significant difference between the
283 strata ($P = 0.26$, Supplementary Figure 2). This suggests that the ENDORSE model is
284 specific to the ER+ cohort and not a general prognostic model.
285

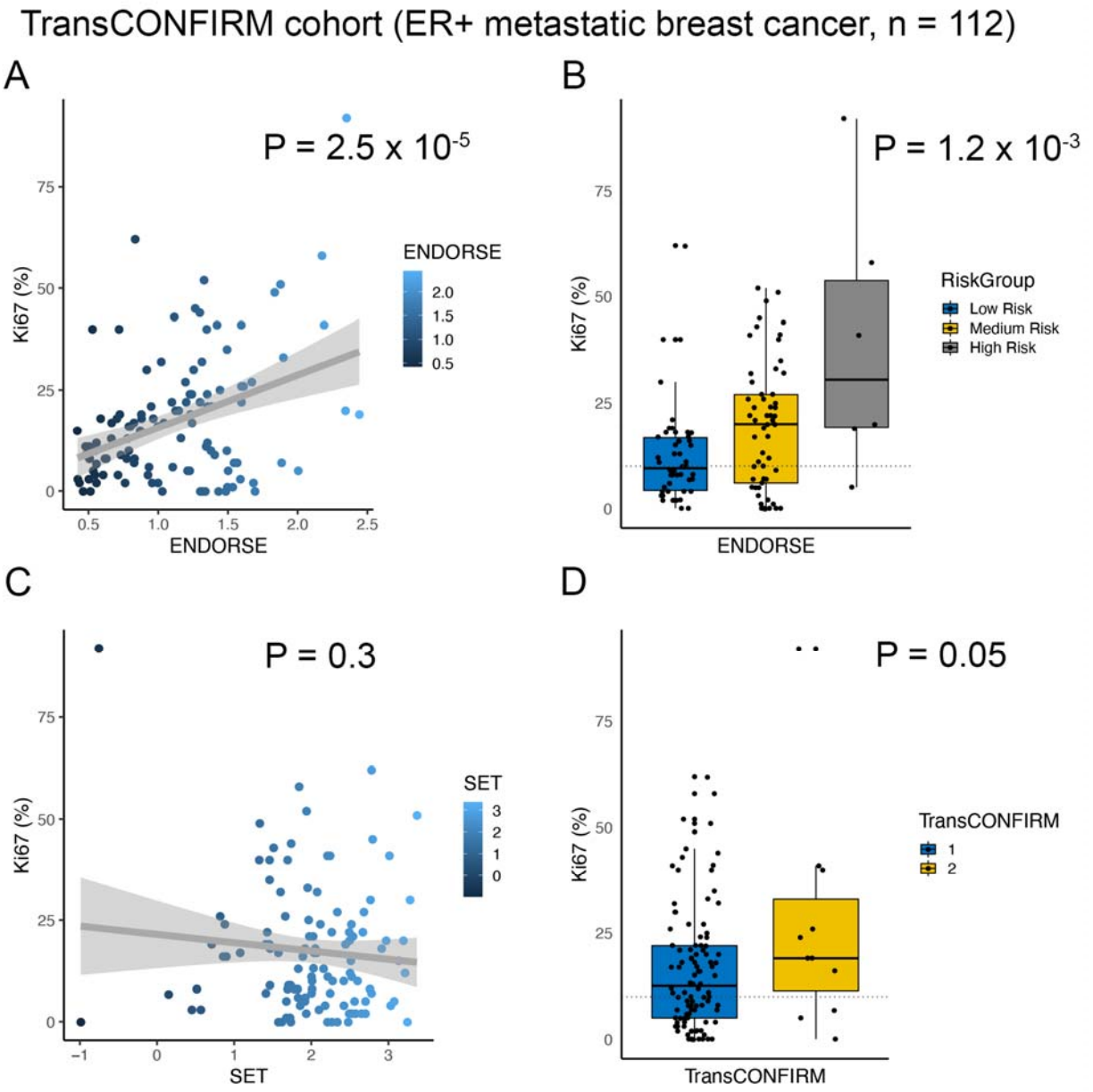
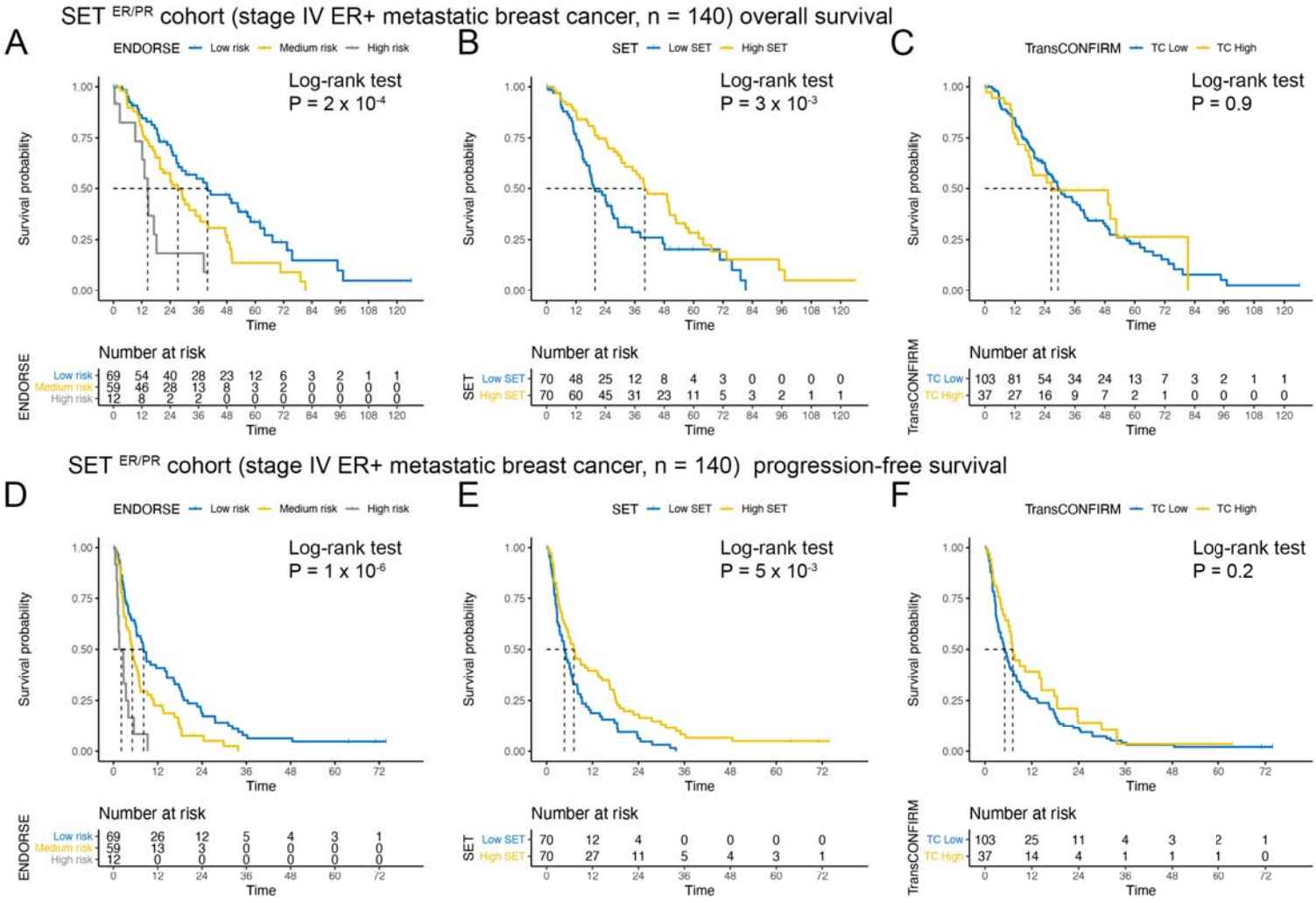


Figure 3: Model validation in TransCONFIRM cohort. A. Scatter plot comparing ENDORSE scores (X-axis) with trial-reported percentage of cells stained positive for Ki67 (Y-axis). Linear fit is shown as a grey line with shaded region showing 95% confidence intervals (C.I.). P-value indicates significance of the linear fit. B. Boxplot comparing Ki67 % across ENDORSE-guided patient strata. P-value indicates significance of the ANOVA model and the horizontal dotted line at 10% indicates

293 threshold of resistance. C. Scatter plot comparing SET scores (X-axis) Ki67 % (Y-axis).
294 Linear fit is shown as a grey line with shaded region showing 95% confidence intervals
295 (C.I.). P-value indicates significance of the linear fit. B. Boxplot comparing Ki67 %
296 across TransCONFIRM predicted patient strata. P-value indicates significance of the
297 ANOVA model and the horizontal dotted line at 10% indicates threshold of resistance.
298



G

Model 1	Model 2	Variance test H1: Model 1 and Model 2 are distinguishable	Partial likelihood ratio test (non-nested) H1A: Model 1 fits better than Model 2
	Overall survival		
ENDORSE	SET	2.1×10^{-4}	0.667
	TransCONFIRM	2.6×10^{-3}	0.046
	Progression free survival		
	SET	1.2×10^{-5}	0.258
	TransCONFIRM	9×10^{-6}	0.038

Figure 4: Model validation in SET ER/PR cohort. A-C. OS Kaplan-Meir curves and risk tables of SET ER/PR patients. The patients were stratified according to A. ENDORSE B. SET and C. TransCONFIRM predicted scores. P-values indicate significance of difference in survival curves based on log-rank tests. D-F PFS Kaplan-Meir curves and risk tables of SET ER/PR patients. The patients were stratified

according to A. ENDORSE B. SET and C. TransCONFIRM scores. P-values indicate significance of difference in survival curves based on 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.

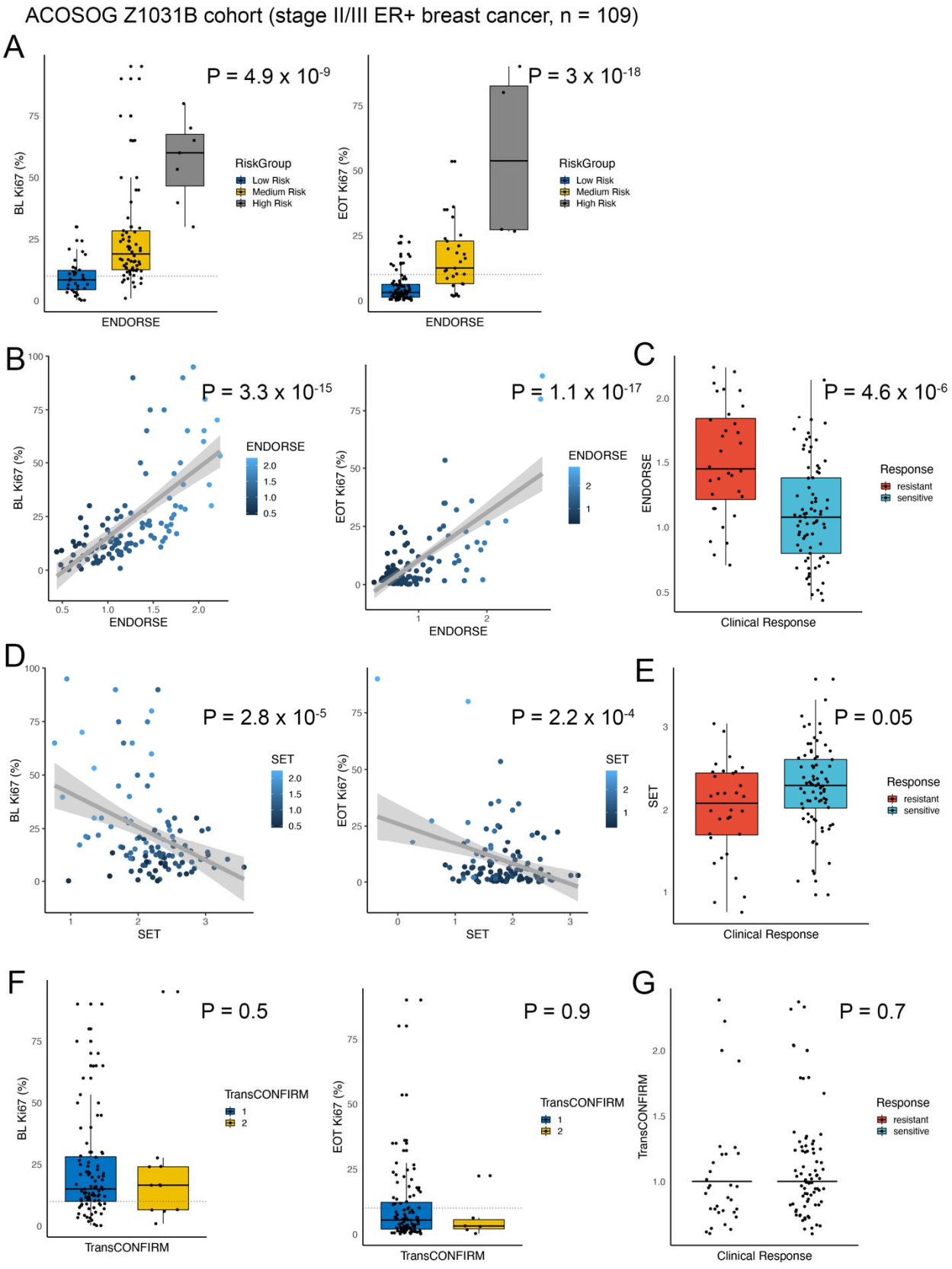


Figure 5: Model validation in ACOSOG Z1031B cohort. A. Boxplots comparing Ki67 % at the baseline (left panel) and end of treatment (right panel) across ENDORSE-predicted patient strata. P-value indicates significance of the ANOVA model and the horizontal dotted line at 10% indicates threshold of resistance. 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 classified as resistant or sensitive clinical response. 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 classified as resistant or sensitive clinical response. 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. P-value indicates significance of the ANOVA model and the horizontal dotted line at 10% indicates threshold of resistance. G. Boxplots comparing TransCONFIRM predictions between patients classified as resistant or sensitive clinical response. P-value indicates significance of ANOVA model

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, 4690 curated and 189 oncogenic signatures from the METABRIC transcriptomes and fitted a generalized additive model for ENDORSE scores with each signature as the predictor (Supplementary Tables 2-4). We found multiple hallmark signatures and oncogenic pathways to be significantly associated with the ENDORSE scores (Supplementary Tables 2-4). Key enriched hallmark signatures included MTOR signaling ($P = 1.03x10^{-72}$) and MYC targets (v2, $P = 2.66x10^{-83}$), while key oncogenic signatures included gain in E2F1 target expression ($P = 8.06x10^{-302}$) and loss of RB1 activity via p107 and p130 ($P = 9.51x10^{-137}$, $1.31x10^{-67}$) (Supplementary Tables 2, 4). Next, we calculated the GES for the hallmark and oncogenic signatures in the three validation datasets (Supplementary Tables 5-10). We observed that pathways associated with cell-cycle progression and proliferation, along with signatures for the loss of RB-1 activity and activation of the PI3K/AKT/MTOR signaling pathways were generally enriched across the METABRIC and all the three validation datasets (Figure 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 (Figure 6a, Supplementary Tables 5-10). The commonality of the signatures enriched across different datasets suggested 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+ tumors. We found a statistically significant association (FDR < 0.05) between the ENDORSE scores and SNVs of only five genes (Figure 6b, Supplementary Table 11). 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 (Supplementary Figure 3). Of the five significant genes, only tumors with *TP53* mutations showed a significantly higher ENDORSE score (Figure 6b). 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 (Figure 6b).

Several gene-level amplifications were also associated with significant differences in ENDORSE scores (Figure 6c, Supplementary Figure 4). Interestingly, the significant amplifications were localized at chromosome 1q, 8p, 8q, or 11q, suggesting different genetic alterations affecting a recurring set of loci may be correlated with the emergence of resistance in the high-risk tumors (Figure 6c, Supplementary Table 12). 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 (Figure 6d).

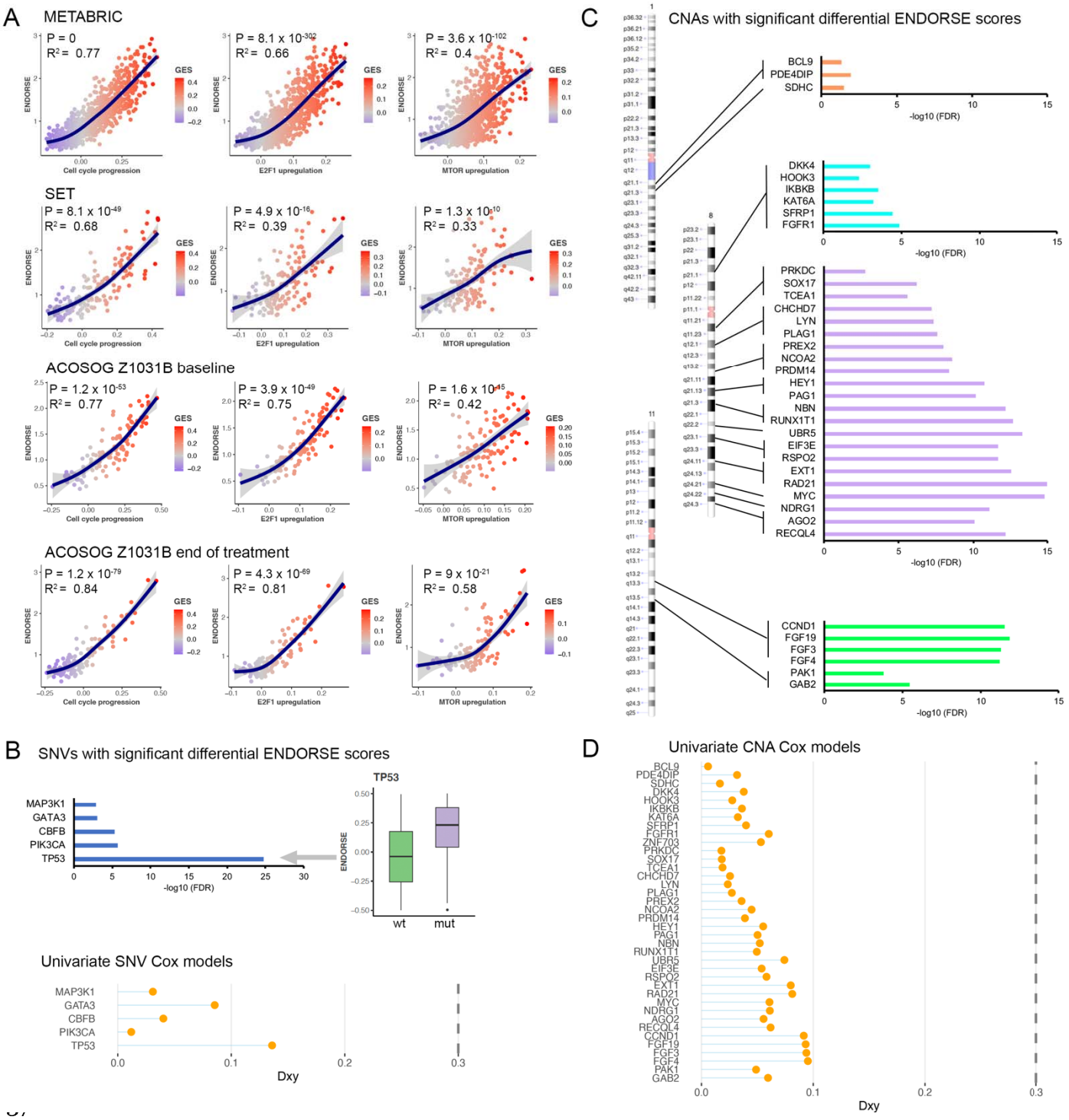


Figure 6: Biology of the high-risk tumors. A. Scatter plots displaying gene set enrichment scores (GES) of key pathways (X-axis) and ENDORSE scores (Y-axis). The cell cycle progression panel represents the hallmark G2M checkpoint signature, the

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. Barplots showing 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 and wildtype tumors. 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. Barplots on the right show 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.

DISCUSSION

The criteria for classifying tumor as ER+ is based on a broad criteria of positive immunohistochemical staining of 1-100% of cell nuclei for the estrogen receptor[14,32]. However, ER+ tumors are heterogeneous, both in terms of dependence on estrogen signaling for growth and survival and intrinsic or acquired resistance to endocrine therapy [33,34]. Therefore, optimal clinical management of each ER+ 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 [35] or assessing the likelihood of benefit from chemotherapy in early-stage, node-negative ER+ breast cancers [7,30]. Results from the MINDACT and TAILORx studies[7,8] show that it is possible 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 proven to be useful in the advanced and metastatic ER+ breast cancer setting. The default primary treatment for advanced ER+ breast cancer remains endocrine therapy, despite proven benefit from add-on targeted therapy or potential switch to chemotherapy. Therefore, the key challenge in advanced ER+ 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[36].

To address this challenge, we have developed a new prognostic model to predict endocrine response in advanced ER+ breast cancers. We developed our model using

417 invasive tumors from the METABRIC study that were ER+ and included node-positive,
418 high-grade tumors. Our model addressed several challenges associated with the
419 development of genomic biomarkers. Since the number of available features to train the
420 genomic models tend to be much larger than the number of available samples (p >> n),
421 it is quite easy to create complex prediction models that contain a large number of
422 predictor variables. Often, such models perform very well in the training datasets, but
423 the performance cannot be replicated in independent test datasets due to overfitting. A
424 number of approaches have been proposed to address this issue. Broadly, these can
425 be classified into unsupervised and supervised approaches. The unsupervised
426 approach typically relies on grouping or clustering the samples into based on similarity
427 of gene expression profiles, followed by analysis of association with survival outcomes
428 [37]. Alternatively, a supervised approach is to perform dimensionality reduction prior to
429 modelling the survival outcome or drug response using univariate or multivariate models
430 [38]. Our model utilized the later strategy by using a regularized Cox model for feature
431 selection, effectively reducing the dimensionality of the gene expression data. We
432 further collapsed the genes into a signature and parameterized the final Cox model on
433 the GES of the signatures. The rank-based approach to calculate GES also helped
434 mitigate issues associated with batch effects and differences in methods for
435 transcriptome profiling. We performed extensive performance evaluation of our model
436 against other published signatures and clinical factors. Consistently, we found that the
437 ENDORSE model was a better predictor than all other models in the training dataset
438 (Figure 2). Moreover, ENDORSE clearly outperformed all other published signatures
439 when they were applied to external validation datasets (Figures 3-5). Our results show

that ENDORSE is a highly accurate and reproducible model that outperforms current approaches to predict endocrine response in metastatic ER+ breast cancer.

We also explored the biology of the ER+ tumors to identify possible mechanisms that are commonly shared by high-risk tumor. We found that high-risk tumors showed a consistent enrichment of pathways associated cell cycle progression and gain of PI3K/MTOR signaling pathways (Figure 6a). In addition, we observed consistent gain of the E2F1 signature, which may be associated with metastatic progression of breast cancers[39,40]. We also observed loss of Rb1 activity, which has been associated with therapeutic resistance in ER+ breast cancers [41,42]

In addition to common pathway phenotypes shared across high-risk tumors, mutations in the *TP53* tumor suppressor genes were also significant (Figure 6b). Loss of function *TP53* variants have long been associated with aggressiveness and chemotherapeutic resistance in hormone-receptor negative breast cancers [43,44]. However, recent studies show that even though *TP53* mutations are infrequent in ER+ breast cancers, they have similar negative impact on patient outcome as hormone-receptor negative breast cancers [45]. We also found recurrent copy number gains at chromosomes 8 and 11 to be associated with high-risk tumors (Figure 6c). 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 [46,47]. The survival models based on genomic alterations were clearly outperformed by ENDORSE;

462 however, the recurrent nature of these alterations in high-risk tumors suggests further
463 studies to investigate their role in promoting endocrine resistance are warranted.
464
465 Drugs that target CDK4/6 to inhibit cell cycle activation[48], PI3K-inhibitors that target
466 tumor with activating *PIK3CA* mutations[12] and mTOR-inhibitors that prevent the
467 activation of mTOR signaling and cell proliferation[13] have been studied and approved
468 for the treatment of advanced ER+ breast cancers in combination with endocrine
469 therapy. However, patients must first advance on primary endocrine therapy, with or
470 without additional CDK4/6 inhibitors, before they can be stratified in a different treatment
471 arm. Therefore, identifying high-risk tumors with the ENDORSE model prior to first-line
472 administration of single-agent endocrine therapy could help identify which cancers may
473 be better suited for an add-on regimen or switching to chemotherapy. Thus, future
474 clinical trials applying ENDORSE model may benefit from early and accurate prediction
475 of endocrine response in advanced, metastatic ER+ breast cancers. This could
476 ultimately help prolong survival of patients by stratifying in more appropriate treatment
477 group.
478

METHODS

Data retrieval and pre-processing

METABRIC gene expression, phenotypic and survival data were retrieved using cBioPortal for cancer genomics[49]. Independent validation datasets used in this study were retrieved from the NCBI Gene Expression Omnibus database with the following accession IDs: SET ER/PR GSE124647[18], TransCONFIRM GSE76040[19], and ACOSOG Z1031B GSE87411[31]. For each gene expression dataset (log2 transformed), we removed genes with zero variance and summarized genes with multiple probes by mean expression and scaling of the expression levels of each gene to a mean of zero and standard deviation of one.

Inclusion criteria for METABRIC training cohort

The METABRIC cohort contained a total of 2509 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[50]. 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 (β) for each predictor in the model can be estimated by maximizing the partial likelihood function L(β), defined as:

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

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 λ to the log of the likelihood function.

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

Where, l(β) = log L(β). The λ penalty parameter was determined using 10-fold cross-validation implemented in R package glmnet [36,37]. The optimal λ minimized model deviance.

We applied the model in a repeated (50 x 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 λ 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[53] using the ssGSEA method[54]. The parameters for the ENDORSE model were obtained by fitting the model to the full training cohort of 833 samples, resulting in β_{emp} = 1.54 and β_{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[27].

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 categoties. 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.

Cox model performance evaluation in training data

The predictive ability of ENDORSE and various other models were evaluated in the METABRIC training dataset using 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 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.

To compare each of the external signatures and clinical feature models with the ENDORSE model, we applied Vuong's[55] partial likelihood ratio test for non-nested Cox regression models calculated using the R package 'nonnestcox' (https://github.com/thomashielscher/nonnestcox/). The individual components of the

566 ENDORSE model were compared to the full model using likelihood ratio tests for nested
567 Cox models.
568
569 **Model validation in independent datasets**
570 We compared the predictive performance of ENDORSE in multiple independent
571 datasets. First, we integrated the training (METABRIC) and test (independent validation)
572 datasets to perform batch correction using the ComBat function of the R package 'sva
573 '[56]. Next, we calculated the GES for the ENDORSE signatures in the training
574 (METABRIC) and test (independent validation) splits of the batch-corrected gene
575 expression dataset. Then, the parameters of the ENDORSE Cox model were calculated
576 on the batch-corrected training split with the OS information as the response variable.
577 Finally, the parameterized Cox model was applied to the test split to obtain a predicted
578 risk score.
579
580 In case of the SET ER/PR cohort, we used the predicted ENDORSE risk scores to
581 stratify the patients into risk categories, with an ENDORSE score ≥2 representing the
582 high-risk group, ≤1 representing the low-risk group and other intermediate values
583 representing the medium risk group. We compared the significance of stratification of
584 both OS and PFS curves based on ENDORSE, SET and TransCONFIRM scores using
585 log-rank tests. Further, we compared the models using partial likelihood ratio tests for
586 non-nested Cox models.
587

For the TransCONFIRM and ACOSOG cohorts, we compared the ENDORSE risk scores, 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.

Biological features associated with ENDORSE scores

To determine the possible biological mechanisms associated with 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[57]. 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.

Gene-level somatic SNV and CNV analyses were performed using data reported by the METABRIC study. SNV's were retained based on a mutation frequency of ≥5 across all samples and limited to genes that are known cancer-related genes. Pathogenic *PIK3CA* variants associated with PI3K inhibitor sensitivity were obtained from the drug labels for alpelisib based on the SOLAR1 clinical trial[12]. Significant SNVs and CNVs were

obtained using a one-way ANOVA analysis of the ENDORSE scores with mutation status as the factor.

Data availability and code

All training and validation datasets used in this study are publicly available and listed under “data retrieval, preprocessing and analysis”. All analyses were performed in RStudio (1.2.5033, R 3.6.3). The sample code for reproducing the analyses in this study are available at https://osf.io/bd3m7/?view_only=da4f860bd2474745880944fce1d433b1

ACKNOWLEDGEMENTS

Funding for this research was provided by the National Cancer Institute of the National Institutes of Health through the U54 grant 1U54CA209978.

AUTHOR CONTRIBUTIONS

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.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interest.

REFERENCES

1. Sung, H. *et al.* 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 (2021).
2. Harvey, J. M., Clark, G. M., Osborne, C. K. & Allred, D. C. 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 (1999).
3. Kohler, B. A. *et al.* Annual Report to the Nation on the Status of Cancer, 1975-2011, Featuring Incidence of Breast Cancer Subtypes by Race/Ethnicity, Poverty, and State. *JNCI: Journal of the National Cancer Institute* **107**, (2015).
4. Waks, A. G. & Winer, E. P. Breast Cancer Treatment: A Review. *JAMA* **321**, 288 (2019).
5. McDonnell, D. P. & Wardell, S. E. The molecular mechanisms underlying the pharmacological actions of ER modulators: implications for new drug discovery in breast cancer. *Current Opinion in Pharmacology* **10**, 620–628 (2010).
6. Smith, I. E. & Dowsett, M. Aromatase Inhibitors in Breast Cancer. *N Engl J Med* **348**, 2431–2442 (2003).
7. Cardoso, F. *et al.* 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *https://doi.org/10.1056/NEJMoa1602253* https://www.nejm.org/doi/10.1056/NEJMoa1602253 (2016) doi:10.1056/NEJMoa1602253.

8. Sparano, J. A. *et al.* Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *New England Journal of Medicine* **379**, 111–121 (2018).

9. Krop, I. *et al.* 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 (2017).

10. Duffy, M. J. *et al.* Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *European Journal of Cancer* **75**, 284–298 (2017).

11. McAndrew, N. P. & Finn, R. S. Management of ER positive metastatic breast cancer. *Seminars in Oncology* **47**, 270–277 (2020).

12. André, F. *et al.* Alpelisib for PIK3CA-Mutated, Hormone Receptor–Positive Advanced Breast Cancer. *New England Journal of Medicine* **380**, 1929–1940 (2019).

13. Baselga, J. *et al.* Everolimus in Postmenopausal Hormone-Receptor–Positive Advanced Breast Cancer. *N Engl J Med* **366**, 520–529 (2012).

14. Rugo, H. S. *et al.* Endocrine Therapy for Hormone Receptor–Positive Metastatic Breast Cancer: American Society of Clinical Oncology Guideline. *JCO* **34**, 3069–3103 (2016).

15. Gradishar, W. J. *et al.* NCCN Guidelines Insights: Breast Cancer, Version 1.2017. *J Natl Compr Canc Netw* **15**, 433–451 (2017).

16. Gradishar, W. J. *et al.* Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network* **18**, 452–478 (2020).

17. Cardoso, F. *et al.* 5th ESO-ESMO international consensus guidelines for advanced breast cancer (ABC 5). *Annals of Oncology* **31**, 1623–1649 (2020).

18. Sinn, B. V. *et al.* SET ER/PR: a robust 18-gene predictor for sensitivity to endocrine therapy for metastatic breast cancer. *NPJ Breast Cancer* **5**, 16 (2019).

19. Jeselsohn, R. *et al.* TransCONFIRM: Identification of a Genetic Signature of Response to Fulvestrant in Advanced Hormone Receptor-Positive Breast Cancer. *Clin. Cancer Res.* **22**, 5755–5764 (2016).

20. Boutros, P. C. The path to routine use of genomic biomarkers in the cancer clinic. *Genome Res* **25**, 1508–1513 (2015).

21. Witten, D. M. & Tibshirani, R. Survival analysis with high-dimensional covariates. *Stat Methods Med Res* **19**, 29–51 (2010).

22. Taylor, J. M. G., Ankerst, D. P. & Andridge, R. R. Validation of Biomarker-Based Risk Prediction Models. *Clin Cancer Res* **14**, 5977–5983 (2008).

23. Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nature Communications* **7**, 1–16 (2016).

24. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).

25. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems* **1**, 417–425 (2015).

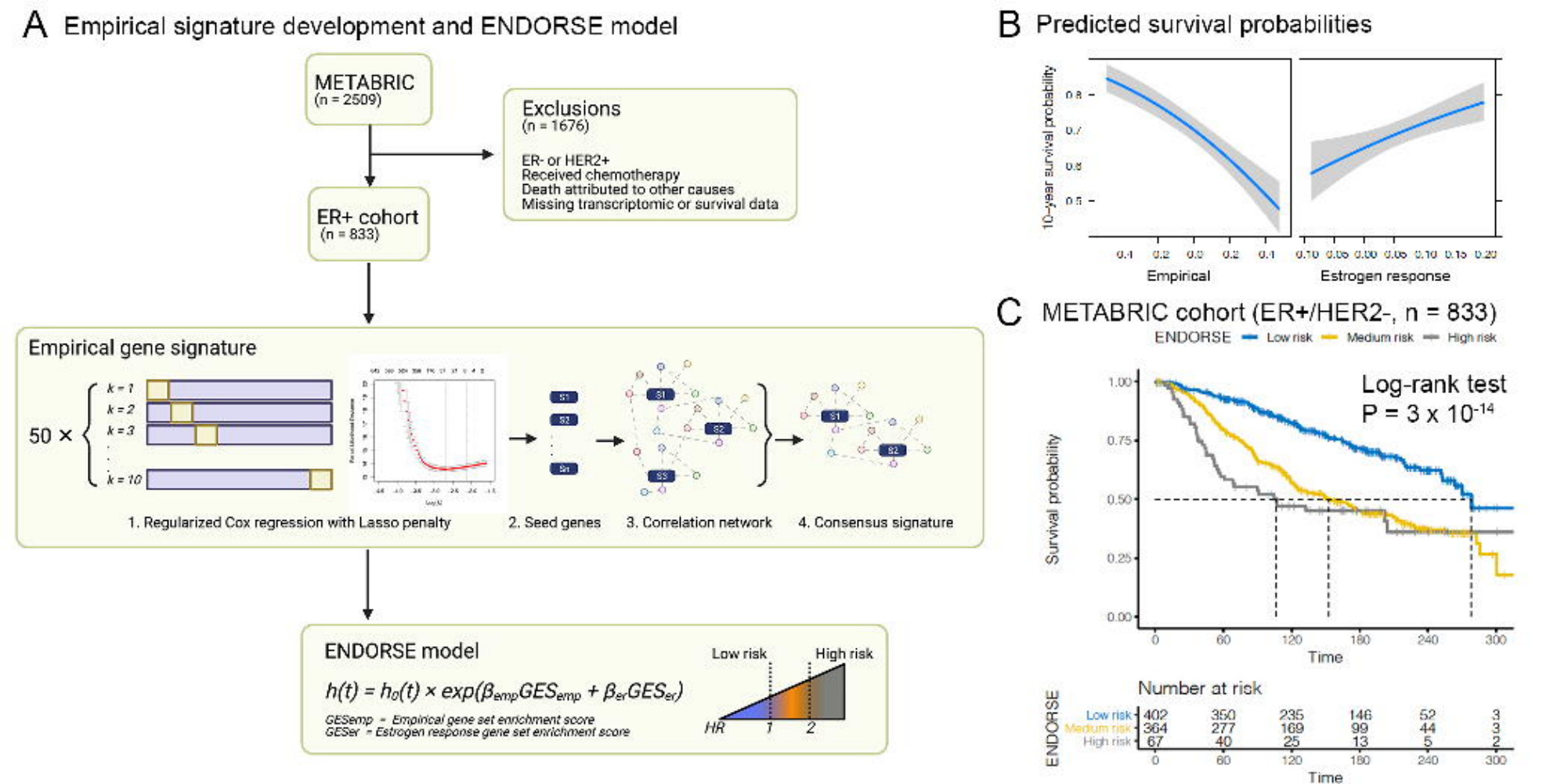
26. Venet, D., Dumont, J. E. & Detours, V. Most Random Gene Expression Signatures Are Significantly Associated with Breast Cancer Outcome. *PLOS Computational Biology* **7**, e1002240 (2011).

699 27. Ramaker, R. C. *et al.* RNA sequencing-based cell proliferation analysis across 19
700 cancers identifies a subset of proliferation-informative cancers with a common
701 survival signature. *Oncotarget* **8**, 38668–38681 (2017).
702 28. Nielsen, T. O. *et al.* A Comparison of PAM50 Intrinsic Subtyping with
703 Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-Treated
704 Estrogen Receptor–Positive Breast Cancer. *Clin Cancer Res* **16**, 5222–5232 (2010).
705 29. Dawson, S.-J., Rueda, O. M., Aparicio, S. & Caldas, C. A new genome-driven
706 integrated classification of breast cancer and its implications. *EMBO J* **32**, 617–628
707 (2013).
708 30. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated,
709 node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
710 31. Ellis, M. J. *et al.* Ki67 Proliferation Index as a Tool for Chemotherapy Decisions
711 During and After Neoadjuvant Aromatase Inhibitor Treatment of Breast Cancer:
712 Results From the American College of Surgeons Oncology Group Z1031 Trial
713 (Alliance). *J. Clin. Oncol.* **35**, 1061–1069 (2017).
714 32. Allison, K. H. *et al.* Estrogen and Progesterone Receptor Testing in Breast
715 Cancer: ASCO/CAP Guideline Update. *JCO* **38**, 1346–1366 (2020).
716 33. Musgrove, E. A. & Sutherland, R. L. Biological determinants of endocrine
717 resistance in breast cancer. *Nature Reviews Cancer* **9**, 631–643 (2009).
718 34. Spoerke, J. M. *et al.* Heterogeneity and clinical significance of ESR1 mutations in
719 ER-positive metastatic breast cancer patients receiving fulvestrant. *Nature*
720 *Communications* **7**, 1–10 (2016).

721 35. Parker, J. S. *et al.* Supervised Risk Predictor of Breast Cancer Based on Intrinsic
722 Subtypes. *J Clin Oncol* **27**, 1160–1167 (2009).
723 36. Hart, C. D. *et al.* Challenges in the management of advanced, ER-positive,
724 HER2-negative breast cancer. *Nat Rev Clin Oncol* **12**, 541–552 (2015).
725 37. Sotiriou, C. *et al.* Breast cancer classification and prognosis based on gene
726 expression profiles from a population-based study. *PNAS* **100**, 10393–10398 (2003).
727 38. Paul, D., Bair, E., Hastie, T. & Tibshirani, R. “Preconditioning” for feature
728 selection and regression in high-dimensional problems. *Ann. Statist.* **36**, 1595–1618
729 (2008).
730 39. Hollern, D. P., Honeysett, J., Cardiff, R. D. & Andrechek, E. R. The E2F
731 Transcription Factors Regulate Tumor Development and Metastasis in a Mouse
732 Model of Metastatic Breast Cancer. *Molecular and Cellular Biology* **34**, 3229–3243
733 (2014).
734 40. Hollern, D. P. *et al.* E2F1 Drives Breast Cancer Metastasis by Regulating the
735 Target Gene FGF13 and Altering Cell Migration. *Sci Rep* **9**, 10718 (2019).
736 41. Bosco, E. E. *et al.* The retinoblastoma tumor suppressor modifies the therapeutic
737 response of breast cancer. *J Clin Invest* **117**, 218–228 (2007).
738 42. Witkiewicz, A. K. & Knudsen, E. S. Retinoblastoma tumor suppressor pathway in
739 breast cancer: prognosis, precision medicine, and therapeutic interventions. *Breast*
740 *Cancer Research* **16**, 207 (2014).
741 43. Cattoretti, G., Rilke, F., Andreola, S., D’Amato, L. & Delia, D. P53 expression in
742 breast cancer. *International Journal of Cancer* **41**, 178–183 (1988).

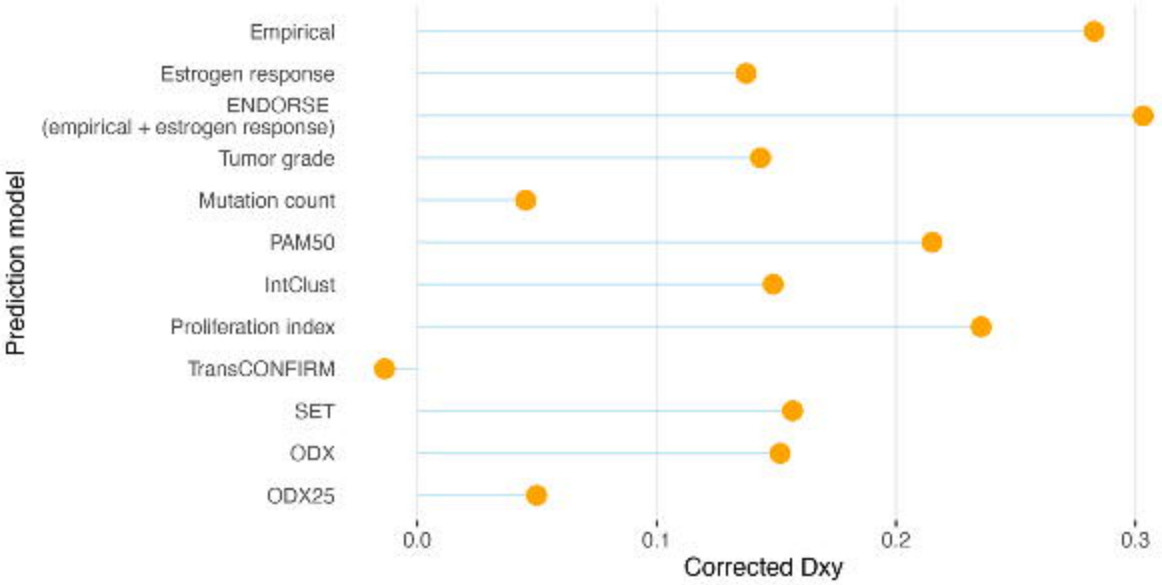
743 44. Elledge, R. M. *et al.* Prognostic significance of p53 gene alterations in node-
744 negative breast cancer. *Breast Cancer Res Tr* **26**, 225–235 (1993).
745 45. Ungerleider, N. A. *et al.* Breast cancer survival predicted by TP53 mutation status
746 differs markedly depending on treatment. *Breast Cancer Research* **20**, 115 (2018).
747 46. Baslan, T. *et al.* Novel insights into breast cancer copy number genetic
748 heterogeneity revealed by single-cell genome sequencing. *eLife* **9**, e51480 (2020).
749 47. Lundgren, K., Holm, K., Nordenskjöld, B., Borg, Å. & Landberg, G. Gene
750 products of chromosome 11q and their association with CCND1gene amplification
751 and tamoxifen resistance in premenopausal breast cancer. *Breast Cancer Res* **10**,
752 R81 (2008).
753 48. Hortobagyi, G. N. *et al.* Ribociclib as First-Line Therapy for HR-Positive,
754 Advanced Breast Cancer. *N Engl J Med* **375**, 1738–1748 (2016).
755 49. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical
756 profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).
757 50. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the*
758 *Royal Statistical Society: Series B (Methodological)* **58**, 267–288 (1996).
759 51. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized
760 Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**, 1–22
761 (2010).
762 52. Simon, N., Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for
763 Cox's Proportional Hazards Model via Coordinate Descent. *Journal of Statistical*
764 *Software* **39**, 1–13 (2011).

765 53. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for
766 microarray and RNA-Seq data. *BMC Bioinformatics* **14**, 7 (2013).
767 54. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS -
768 driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
769 55. Vuong, Q. H. Likelihood Ratio Tests for Model Selection and Non-Nested
770 Hypotheses. *Econometrica* **57**, 307 (1989).
771 56. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva
772 package for removing batch effects and other unwanted variation in high-throughput
773 experiments. *Bioinformatics* **28**, 882–883 (2012).
774 57. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics*
775 **27**, 1739–1740 (2011).
776


A Empirical signature development and ENDORSE model
METABRIC
(n = 2509)
Exclusions
(n = 1676)
ER- or HER2+
Received chemotherapy
Death attributed to other causes
Missing transcriptomic or survival data
ER+ cohort
(n = 833)
Empirical gene signature
50 ×
k = 1
k = 2
k = 3
k = 10
1. Regularized Cox regression with Lasso penalty
2. Seed genes
3. Correlation network
4. Consensus signature
ENDORSE model
h(t) = h0(t) × exp(βempGESemp + βerGESer)
GESemp = Empirical gene set enrichment score
GESer = Estrogen response gene set enrichment score
Low risk
High risk
HR
1
2
B Predicted survival probabilities
10−year survival probability
Empirical
Estrogen response
C METABRIC cohort (ER+/HER2-, n = 833)
ENDORSE
Low risk
Medium risk
High risk
Log-rank test
P = 3 x 10-14
Survival probability
Time
Number at risk
Low risk 402 350 235 146 52 3
Medium risk 364 277 169 99 44 3
High risk 67 40 25 13 5 2
0 60 120 180 240 300

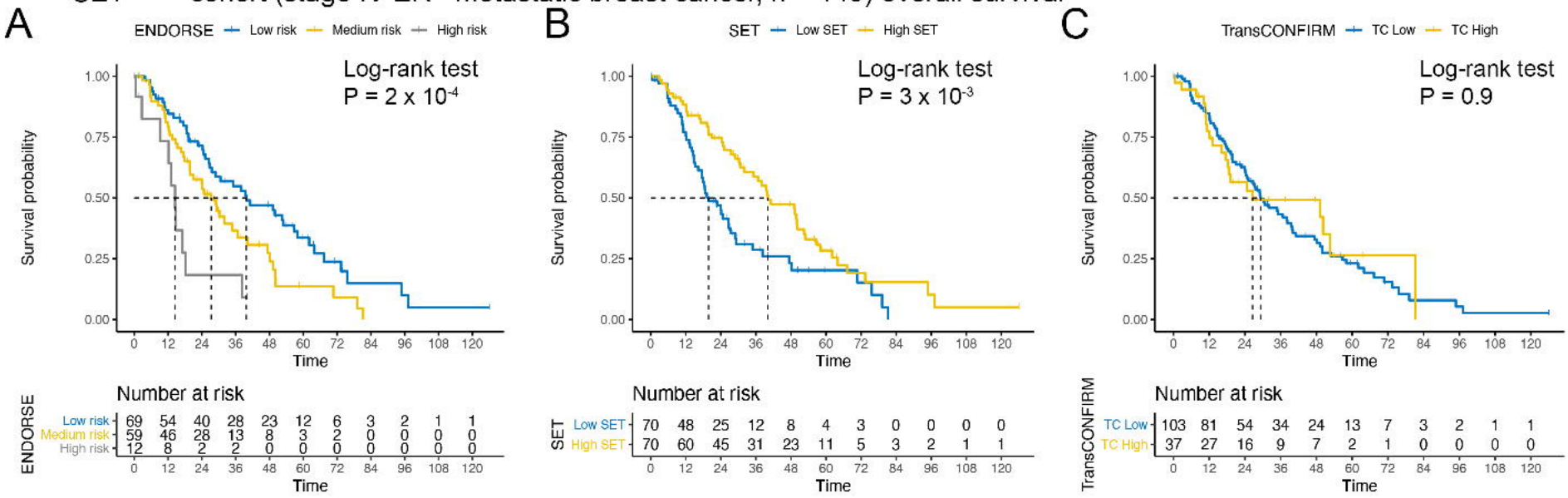
METABRIC cohort (ER+/HER2-, n = 833)

A Prediction model validation (bootstrap resampling, B=150)



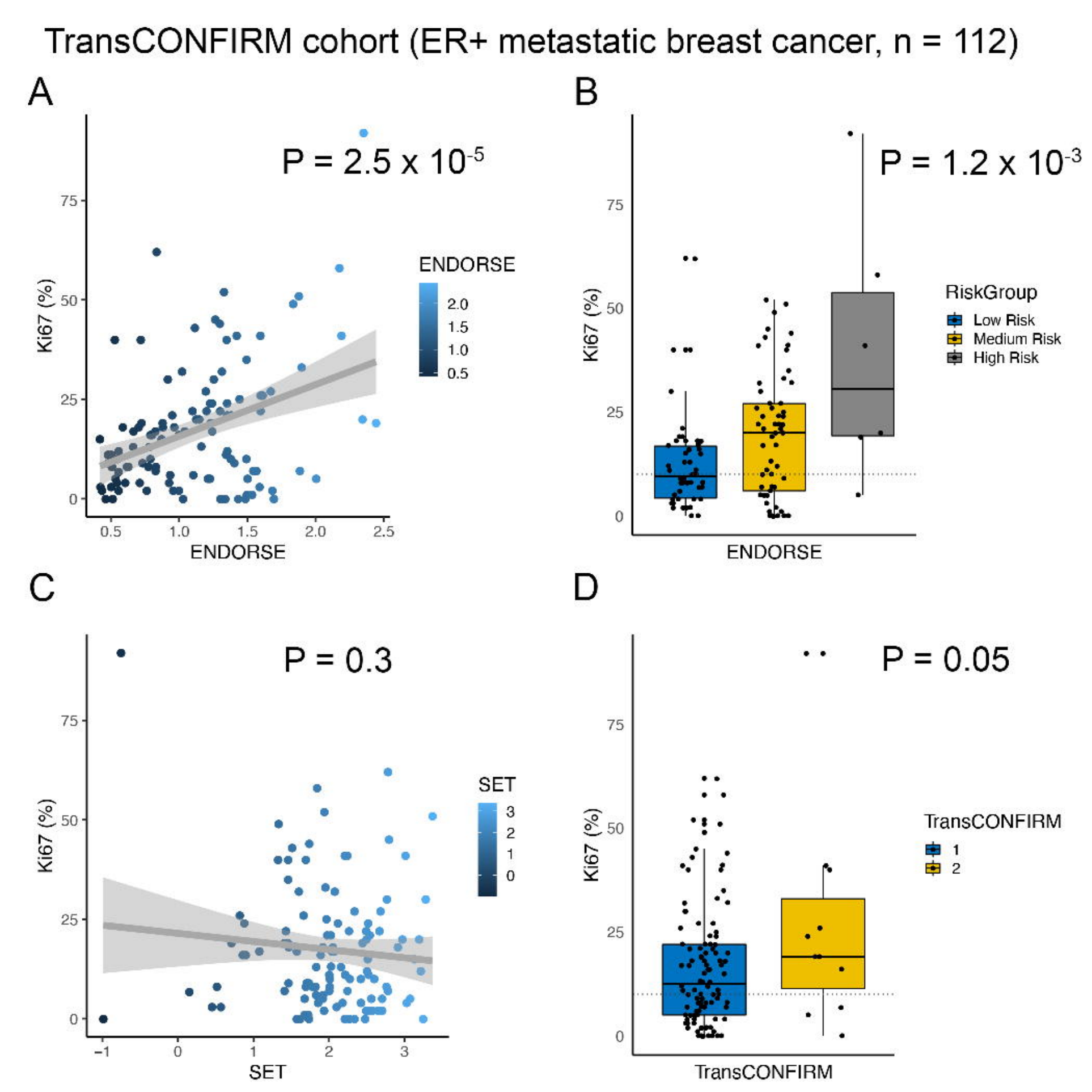
B Model comparisons

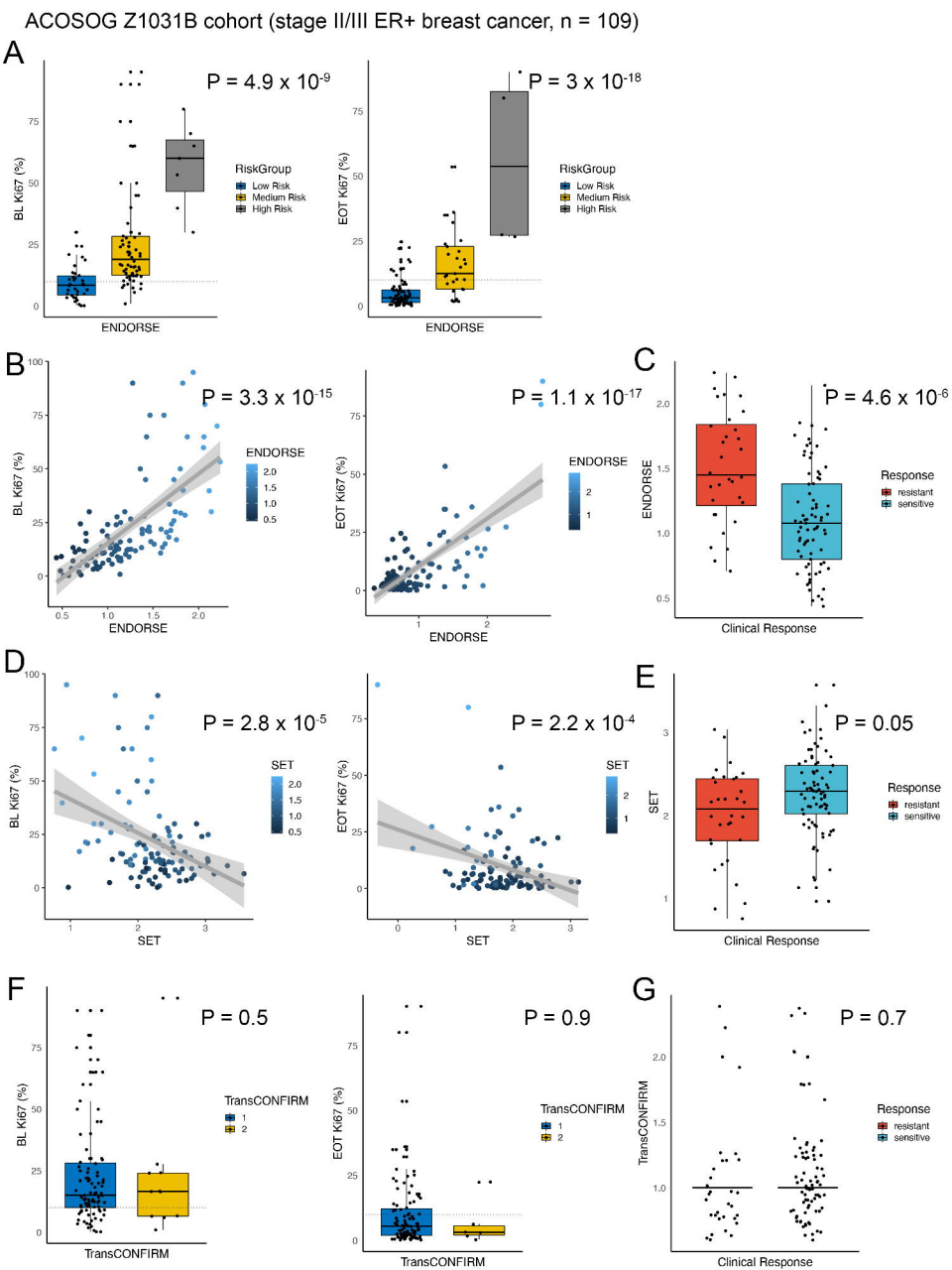
Model 1	***Model 2***	***Variance test*** *H1: Model 1 and Model 2 are distinguishable*	***Likelihood ratio test (nested)*** *H1A: Full model fits better than reduced model*	***Partial likelihood ratio test (non-nested)*** *H1A: Model 1 fits better than Model 2*
ENDORSE	Empirical	1.47 x 10^{-3}	1.09 x 10^{-3}	
	Estrogen response	3.83 x 10^{-6}	3.93 x 10^{-14}	
	METABRIC tumor characteristics and signatures			
	Tumor grade	< 2 x 10^{-16}		2.08 x 10-3
	Mutation count	< 2 x 10^{-16}		9.76 x 10-6
	PAM50	2.22 x 10-8		0.033
	IntClust	< 2 x 10^{-16}		0.02
	Proliferation index	3.73 x 10^{-5}		4.42 x 10^{-5}
	External signatures			
	TransCONFIRM	1.67 x 10^{-7}		9.73 x 10^{-6}
	SET	< 2 x 10^{-16}		3.53 x 10^{-5}
	ODX	2.18 x 10^{-7}		6.15 x 10^{-5}
	ODX25	1.49 x 10^{-6}		1.32 x 10^{-5}

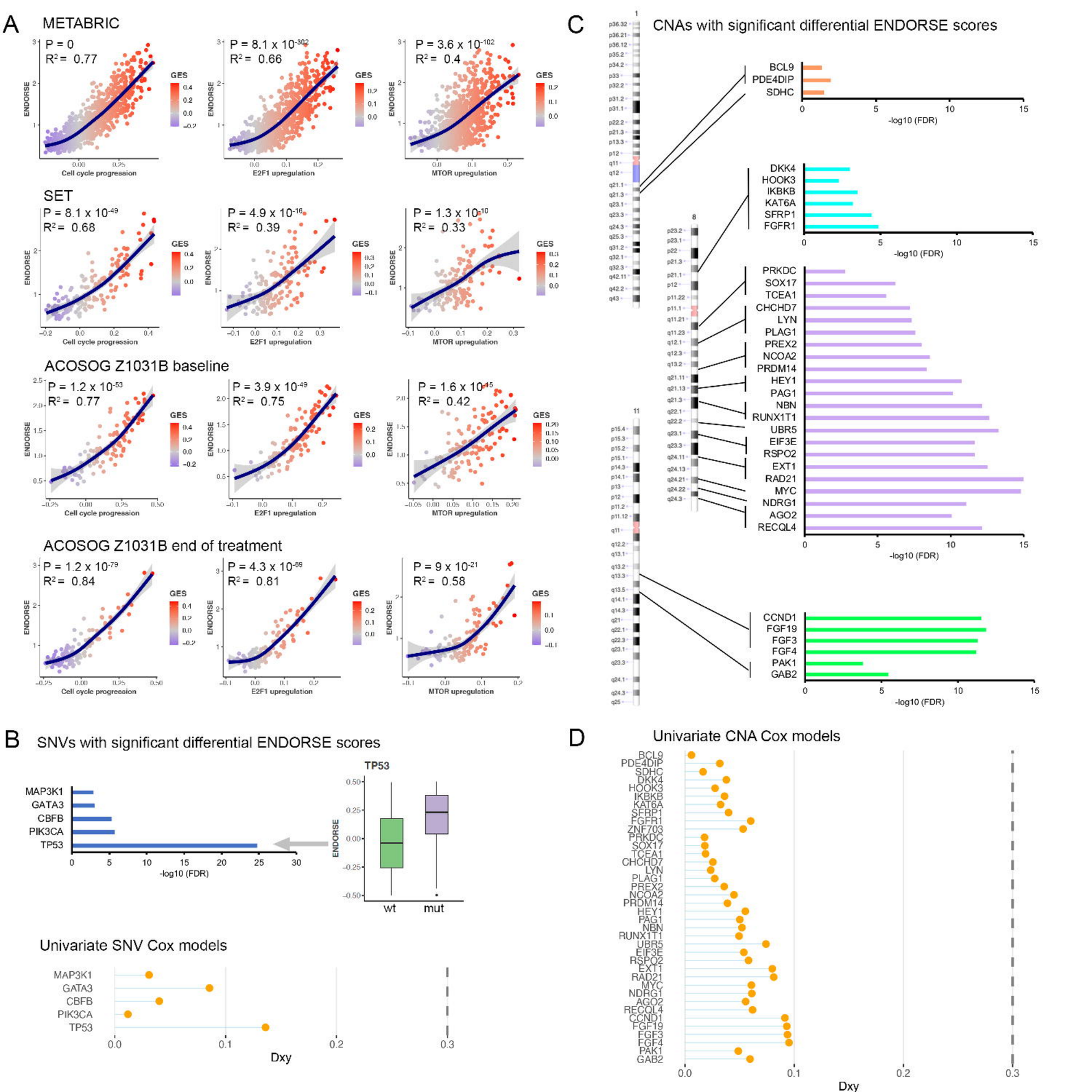


G

Model 1	Model 2	Variance test *H1: Model 1 and Model 2 are distinguishable*	Partial likelihood ratio test (non-nested) *H1A: Model 1 fits better than Model 2*
	Overall survival		
ENDORSE	SET	2.1×10^{-4}	0.667
	TransCONFIRM	2.6×10^{-3}	0.046
	Progression free survival		
	SET	1.2×10^{-5}	0.258
	TransCONFIRM	9×10^{-6}	0.038


TransCONFIRM cohort (ER+ metastatic breast cancer, n = 112)
A
P = 2.5 x 10-5
Ki67 (%)
ENDORSE
ENDORSE
2.0
1.5
1.0
0.5
B
P = 1.2 x 10-3
Ki67 (%)
ENDORSE
RiskGroup
Low Risk
Medium Risk
High Risk
C
P = 0.3
Ki67 (%)
SET
SET
3
2
1
0
D
P = 0.05
Ki67 (%)
TransCONFIRM
TransCONFIRM
1
2


ACOSOG Z1031B cohort (stage II/III ER+ breast cancer, n = 109)
A
P = 4.9 x 10-9
BL Ki67 (%)
ENDORSE
RiskGroup
Low Risk
Medium Risk
High Risk
P = 3 x 10-18
EOT Ki67 (%)
ENDORSE
RiskGroup
Low Risk
Medium Risk
High Risk
B
P = 3.3 x 10-15
BL Ki67 (%)
ENDORSE
ENDORSE
P = 1.1 x 10-17
EOT Ki67 (%)
ENDORSE
ENDORSE
C
P = 4.6 x 10-6
ENDORSE
Clinical Response
Response
resistant
sensitive
D
P = 2.8 x 10-5
BL Ki67 (%)
SET
SET
P = 2.2 x 10-4
EOT Ki67 (%)
SET
SET
E
P = 0.05
SET
Clinical Response
Response
resistant
sensitive
F
P = 0.5
BL Ki67 (%)
TransCONFIRM
TransCONFIRM
P = 0.9
EOT Ki67 (%)
TransCONFIRM
TransCONFIRM
G
P = 0.7
TransCONFIRM
Clinical Response
Response
resistant
sensitive


A
METABRIC
P = 0
R² = 0.77
P = 8.1 x 10-302
R² = 0.66
P = 3.6 x 10-102
R² = 0.4
SET
P = 8.1 x 10-49
R² = 0.68
P = 4.9 x 10-16
R² = 0.39
P = 1.3 x 10-10
R² = 0.33
ACOSOG Z1031B baseline
P = 1.2 x 10-53
R² = 0.77
P = 3.9 x 10-49
R² = 0.75
P = 1.6 x 10-15
R² = 0.42
ACOSOG Z1031B end of treatment
P = 1.2 x 10-79
R² = 0.84
P = 4.3 x 10-89
R² = 0.81
P = 9 x 10-21
R² = 0.58
ENDORSE
GES
Cell cycle progression
E2F1 upregulation
MTOR upregulation
B
SNVs with significant differential ENDORSE scores
MAP3K1
GATA3
CBFB
PIK3CA
TP53
-log10 (FDR)
TP53
wt
mut
Univariate SNV Cox models
Dxy
C
CNAs with significant differential ENDORSE scores
BCL9
PDE4DIP
SDHC
DKK4
HOOK3
IKBKB
KAT6A
SFRP1
FGFR1
PRKDC
SOX17
TCEA1
CHCHD7
LYN
PLAG1
PREX2
NCOA2
PRDM14
HEY1
PAG1
NBN
RUNX1T1
UBR5
EIF3E
RSPO2
EXT1
RAD21
MYC
NDRG1
AGO2
RECQL4
CCND1
FGF19
FGF3
FGF4
PAK1
GAB2
-log10 (FDR)
D
Univariate CNA Cox models
Dxy