

## EPRS is a critical regulator of cell proliferation and estrogen signaling in ER<sup>+</sup> breast cancer

Igor Katsyv<sup>1,2,3</sup>, Minghui Wang<sup>2,3</sup>, Won Min Song<sup>2,3</sup>, Xianxiao Zhou<sup>2,3</sup>, Yongzhong Zhao<sup>2,3</sup>, Sun Park<sup>4</sup>, Jun Zhu<sup>2,3</sup>, Bin Zhang<sup>2,3,5</sup>, Hanna Y. Irie<sup>4,5</sup>

<sup>1</sup>Medical Scientist Training Program, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>2</sup>Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>3</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>4</sup>Division of Hematology and Medical Oncology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>5</sup>Department of Oncological Sciences, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

**Correspondence to:** Bin Zhang, **email:** bin.zhang@mssm.edu  
Hanna Y. Irie, **email:** hanna.irie@mssm.edu

**Keywords:** breast cancer, ER<sup>+</sup>, EPRS, gene networks

**Received:** May 04, 2016

**Accepted:** August 25, 2016

**Published:** September 06, 2016

### ABSTRACT

**Aminoacyl tRNA synthetases (ARSs) are a class of enzymes with well-conserved housekeeping functions in cellular translation. Recent evidence suggests that ARS genes may participate in a wide array of cellular processes, and may contribute to the pathology of autoimmune disease, cancer, and other diseases. Several studies have suggested a role for the glutamyl prolyl tRNA synthetase (EPRS) in breast cancers, although none has identified any underlying mechanism about how EPRS contributes to carcinogenesis. In this study, we identified EPRS as upregulated in estrogen receptor positive (ER<sup>+</sup>) human breast tumors in the TCGA and METABRIC cohorts, with copy number gains in nearly 50% of samples in both datasets. EPRS expression is associated with reduced overall survival in patients with ER<sup>+</sup> tumors in TCGA and METABRIC datasets. EPRS expression was also associated with reduced distant relapse-free survival in patients treated with adjuvant tamoxifen monotherapy for five years, and EPRS-correlated genes were highly enriched for genes predictive of a poor response to tamoxifen. We demonstrated the necessity of EPRS for proliferation of tamoxifen-resistant ER<sup>+</sup> breast cancer, but not ER<sup>-</sup> breast cancer cells. Transcriptomic profiling showed that EPRS regulated cell cycle and estrogen response genes. Finally, we constructed a causal gene network based on over 2500 ER<sup>+</sup> breast tumor samples to build up an EPRS-estrogen signaling pathway. EPRS and its regulated estrogenic gene network may offer a promising alternative approach to target ER<sup>+</sup> breast cancers that are refractory to current anti-estrogens.**

### INTRODUCTION

ARSs are enzymes that charge tRNAs with their cognate amino acids. In higher eukaryotes, however, ARSs contain additional domains that allow them functions beyond their canonical roles in translation, and several studies have reported dysregulation of ARSs and their noncanonical functions in disease [1]. It is unclear, however, if this dysregulation contributes to or is a byproduct of disease-driving processes.

The glutamyl prolyl tRNA synthetase (EPRS) was identified as a tumor immunogen in human breast and gastrointestinal cancers [2], as well as in spontaneous

tumors in neu-transgenic mice [3]. Kim et al. identified EPRS copy number gains in lung, esophageal, hepatocellular, skin and breast cancers tumors [4]. EPRS has also been implicated in engrailed 1 (EN1)-mediated survival of triple-negative breast cancer cells [5], but not in ER<sup>+</sup> breast cancers. EPRS protein is comprised of N-terminus glutamyl-tRNA synthetase (ERS) domain, a C-terminus prolyl-tRNA synthetase (PRS) domain, and linker composed of three 50-amino-acid-long WHEP domains. While the ERS and PRS domains carry out the canonical and functions of EPRS – aminoacylation of cognate tRNAs – the WHEP domains have been implicated in transcript-specific regulation of translation, particularly

of the genes ceruloplasmin and VEGFA [6, 7]. While this suggests *EPRS* may play a role in tumor angiogenesis, a functional link between *EPRS* and angiogenesis has not been established in tumors.

In this study, we examined the role of *EPRS* in the pathogenesis of ER+ breast cancers, which comprise over 70% of breast cancers and which are treated primarily with anti-estrogen drugs, such as tamoxifen. We showed elevated *EPRS* transcript levels in ER+ breast cancer samples compared to adjacent normal breast tissue, and *EPRS* copy number gain in ER+ breast cancers. *EPRS* expression was associated with reduced overall survival in patients with ER+, but not ER- breast cancers and was associated with reduced distant relapse-free survival in patients treated with tamoxifen adjuvant monotherapy. *EPRS*-correlated transcripts in ER+ breast cancer samples were highly enriched for genes predictive of relative resistance to tamoxifen, and *EPRS* inhibition induced a G1/S arrest in tamoxifen-resistant MCF7 cells, but not in ER- breast cancer cells. Using RNA-sequencing and Bayesian network inference, we demonstrated that *EPRS* is a critical regulator of *ERα* expression and activity. Our findings raise the possibility that *EPRS* inhibition may be an alternative approach to suppressing the growth of ER+ breast cancers refractory to tamoxifen treatment. To our knowledge, we are the first to implicate *EPRS* in ER+ breast cancer and the first to describe an underlying, tumor cell-intrinsic mechanism through which *EPRS* may contribute to ER+ breast tumorigenesis.

## RESULTS

### ***EPRS* is upregulated in ER+ breast cancers and is associated with reduced overall survival**

We compared expression of ARS genes in ER+ tumor and adjacent normal samples in TCGA and METABRIC cohorts (Supplementary Figure S1). Twenty-eight of the 37 genes were differentially expressed in the TCGA cohort and 35 in the METABRIC cohorts at an FDR cutoff of 0.05 (Supplementary Figure S1). Meta-analysis of differential expression in the two cohorts, by summing  $-\log_{10} p$ -values, identified *EPRS* as the most significantly differentially expressed ARS (TCGA:  $\log_{2}FC = 0.39$ , student's  $t$ -test  $p = 6.70 \times 10^{-14}$ ; METABRIC:  $\log_{2}FC = 0.80$ , student's  $t$ -test  $p = 1.04 \times 10^{-87}$ ) (Supplementary Figure S1) (Figure 1A and 1B). Interestingly, among ER+ tumors, *EPRS* mRNA expression was higher in the relatively more endocrine therapy resistant Luminal B tumors compared to Luminal A tumors in TCGA ( $\log_{2}FC = 0.10$ , student's  $t$ -test  $p = 0.05$ ) and METABRIC ( $\log_{2}FC = 0.10$ , student's  $t$ -test  $p = 4.78 \times 10^{-5}$ ) cohorts (Figure 1A and 1B). *EPRS* expression is additionally upregulated in basal, Her2, and normal-like breast tumors (Figure 1A and 1B).

*EPRS* maps to chromosome 1q41, amplifications of which frequently occur in breast cancers [8, 9] and are

associated with bone metastasis [10]. Single nucleotide polymorphisms in 1q41 have been linked in genome-wide association studies with increased risk of colorectal cancer [11]. Using COSMIC, we identified *EPRS* copy number gains in 68 of 151 (45.0%) ER+ breast cancer samples from the TCGA cohort (Figure 1C); *EPRS* copy number gains are present in 672 of 1505 (44.7%) of METABRIC ER+ breast cancer samples (Figure 1D). *EPRS* copy number gains may thus partially account for elevated *EPRS* mRNA expression in breast cancers compared to adjacent normal breast, and amplifications of 1q41 involving *EPRS* may drive a subset of breast cancers.

*EPRS* expression has prognostic significance. We found elevated *EPRS* expression to be associated with poorer overall survival in patients with ER+ breast cancers in both TCGA (KM  $p = 5.29 \times 10^{-3}$ , Cox  $p = 1.04 \times 10^{-3}$ ; HR = 2.07) and METABRIC datasets (KM  $p = 7.73 \times 10^{-3}$ , Cox  $p = 1.67 \times 10^{-2}$ ; HR = 1.23) (Figure 1E). We did not observe a significant association between *EPRS* transcript levels and outcomes for patients with ER- breast cancers (Figure 1F) in either cohort. Thus, *EPRS* may play a role specifically in promoting ER+ tumor growth and treatment sensitivity.

### ***EPRS* is associated with tamoxifen resistance**

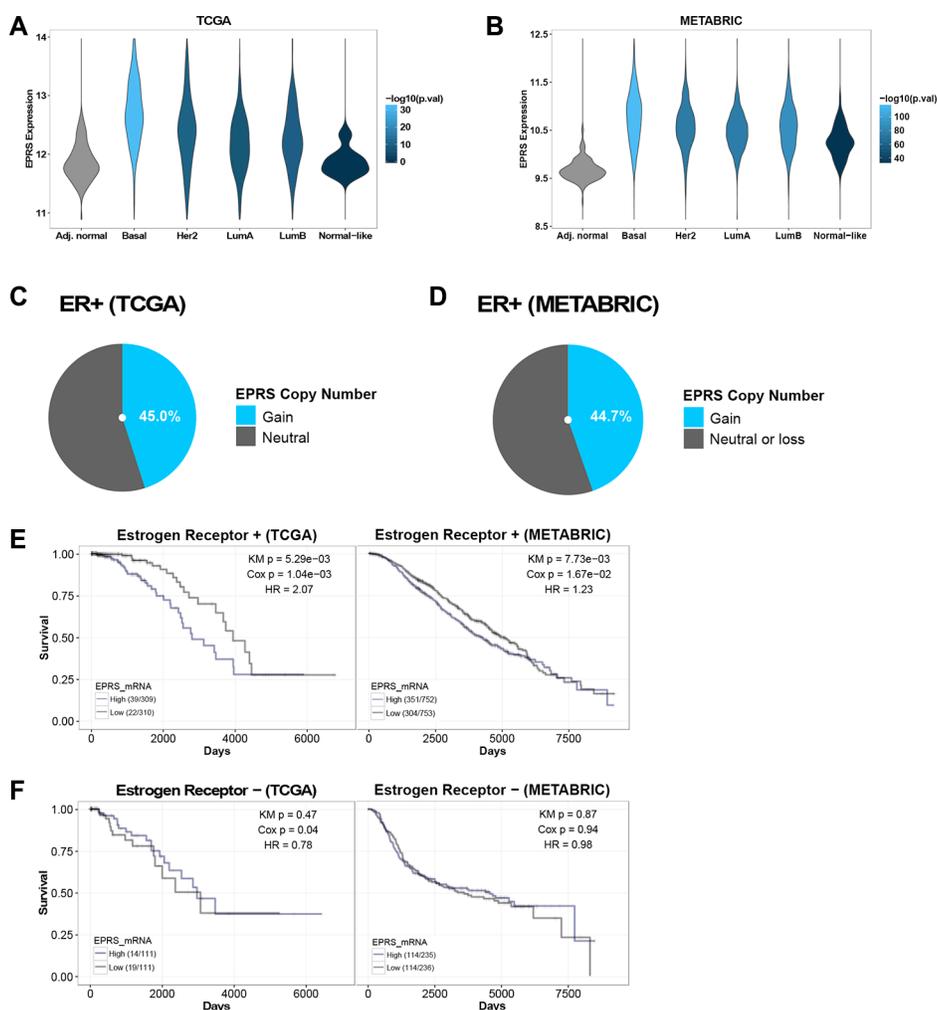
Endocrine therapies, such as the ER antagonist tamoxifen, are the mainstays of treatment for patients with ER+ breast cancer [12]. Given the prognostic significance of *EPRS* expression for ER+ breast tumors and the relatively higher expression of *EPRS* in Luminal B compared to Luminal A tumors, we hypothesized that *EPRS* expression may be associated with reduced sensitivity to endocrine therapy. In a cohort of 298 ER+ breast cancer patients who were treated with tamoxifen alone for 5 years [13], we observed a significant negative association between *EPRS* mRNA expression and recurrence-free survival (KM =  $2.82 \times 10^{-3}$ , Cox  $p = 9.42 \times 10^{-3}$ ; HR = 2.08) (Figure 2A). The genes whose expression correlates with that of *EPRS* in TCGA and METABRIC ER+ tumor samples are significantly enriched for predictors of nonresponse to tamoxifen therapy [13–15] (TCGA: corrected FET  $p = 3.96 \times 10^{-16}$ , FE = 3.78; METABRIC: corrected FET  $p = 2.52 \times 10^{-6}$ , FE = 3.08) (Figure 2B). In the TCGA cohort, the genes anticorrelated with *EPRS* are significantly enriched for genes predictive of favorable tumor response to tamoxifen (corrected FET  $p = 1.43 \times 10^{-3}$ , FE = 3.04). These data support a role for *EPRS* in endocrine therapy resistance of ER+ tumors.

### ***EPRS* is necessary for proliferation of ER+ breast cancer cells**

To determine the functional roles of *EPRS* in ER+ breast cancer, we downregulated *EPRS* expression in ER+ breast cancer cell lines using shRNA vectors or

siRNA. To determine if *EPRS* is critical for the growth of endocrine therapy-resistant ER+ breast cancer cells, we downregulated *EPRS* in MCF7 TamR cells, an MCF7 variant that was *in vitro* selected for Tamoxifen resistance by continuous culture in the presence of increasing concentrations of Tamoxifen. *EPRS* downregulation inhibited MCF7 TamR growth in 3D Matrigel™ cultures (Figure 3A). As the growth inhibition may be due to differences in survival and/or proliferation, we assessed cell cycle progression by flow cytometry. When compared to MCF7 TamR cells expressing vector control shRNA, cells expressing any of three independent *EPRS*-targeting shRNA vectors were found to be arrested in G1 (vector control: 58.65% G1; shEPRS-73: 83.33% G1,  $p = 0.013$ ; shEPRS-74: 79.40% G1,  $p = 2.2E-3$ ; shEPRS-83: 89.12% G1,  $p = 1.9E-3$ ) with a concomitant decrease of

cells in S phase (vector control: 26.77%; shEPRS-73: 8.76%,  $p = 5.1E-3$ ; shEPRS-74: 3.40%,  $p = 2.4E-4$ ; shEPRS-84: 3.77%,  $p = 1.2E-3$ ) (Figure 3B). Consistent with these data, we observed downregulation of proteins involved in the G1-to-S and S-to-G2 transition, such as CDK2, CCNB1, and phospho-Rb (Figure 3C, Supplementary Figure S2). We did not observe a consistent increase in the subG1 fraction (Figure 3B) or cleavage of PARP (not shown). Furthermore, *EPRS* depletion induced growth arrest in the presence of the apoptosis inhibitor Z-VAD (Figure 3C). Taken together, these data suggest that apoptosis is not a major mechanism for *EPRS* downregulation-mediated growth inhibition. *EPRS* downregulation similarly affected parental MCF7 cells (Figure 3D). *EPRS* downregulation did not significantly affect the growth of the ER- cell line MDA-MB-453



**Figure 1: *EPRS* is upregulated in ER+ breast cancers, and is associated with worse outcomes in ER+ tumors.** *EPRS* is upregulated in all breast cancer subtypes compared to adjacent normal breast in TCGA (A) and METABRIC (B) cohorts. Violin plot color represents significance of student's *t*-test comparing each subtype with adjacent normal breast. Fraction of TCGA ER+ breast cancer samples in COSMIC (C) and fraction of METABRIC (D) ER+ samples with *EPRS* copy number gains. High *EPRS* expression is associated with worse prognosis in patients with ER+ (E) but not ER- (F) breast cancer in TCGA and METABRIC cohorts. "High" and "Low" refer to stratification of patients by median *EPRS* expression: upper 50% were called "high," lower 50% were called "low." Numbers in parenthesis (e.g. 39/309) are [number of patients who died]/[number of patients at risk].

(Figure 3E and 3F). These data support a critical and specific role for *EPRS* in regulating proliferation of ER+ breast cancer cells.

### ***EPRS*-regulated transcriptome**

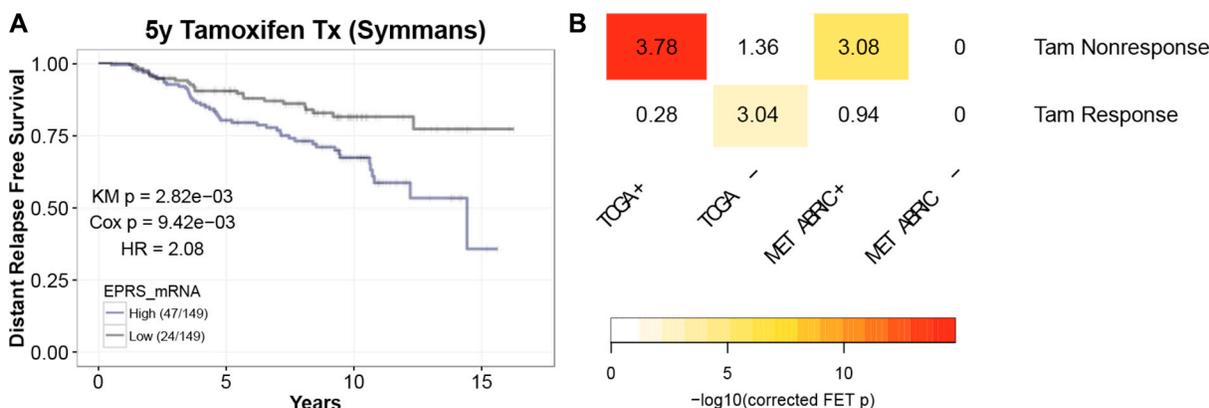
To begin to determine the mechanisms by which *EPRS* regulates G1-S transition and proliferation of ER+ breast cancers, we identified the transcriptional programs regulated by *EPRS* in ER+ tumors. We performed paired-end 100-nucleotide Illumina sequencing on MCF7 TamR expressing either vector control or one of three unique *EPRS* shRNA sequences. Using an FDR cutoff of 0.05 and a fold change cutoff of 1.3, we observed 7,129 differentially-expressed genes with shEPRS-73, 10,995 with shEPRS-74, and 5,114 with shEPRS-84 (Supplementary Figure S3). We used genes down- or upregulated by at least two of three shRNA vectors for downstream analysis yielding 3146 downregulated and 3715 upregulated genes (Supplementary Figure S3). We performed functional annotation analysis on this *EPRS* signature using the MSigDB Hallmark gene sets [16]. In agreement with the cell cycle arrest we observed, *EPRS* knockdown downregulated genes involved in cell proliferation: E2F targets (corrected FET  $p = 4.60e-40$ , FE = 2.96); G2M checkpoint (corrected FET  $p = 2.50e-24$ , FE = 2.51); Myc targets v1 (corrected FET  $p = 9.80e-21$ , FE = 3.72); Myc targets v2 (corrected FET  $p = 1.80e-14$ , FE = 2.15). Consistent with *EPRS*'s known role in inhibiting translation of interferon gamma-induced genes [17] upregulated genes were enriched for interferon gamma response (corrected FET  $p = 0.024$ , FE = 1.59), as well as bile acid metabolism (corrected FET  $p = 1e-04$ , FE = 2.13) and genes downregulated by KRAS-signaling (corrected FET  $p = 3e-03$ , FE = 1.68) (Figure 4A). We did not observe induction of *DDIT3* or downregulation of *COL1A1*, *COL1A2*, or *S100A4*, as has previously been reported in *EPRS*-inhibition-induced stress response

[5, 18, 19]. Similarly, enrichment of upregulated genes for “unfolded protein response” was not significant (corrected FET  $p = 0.17$ , FE = 1.55) (Figure 4A). We did not observe significant upregulation of genes associated with apoptosis (Figure 4A). Interestingly, *EPRS* shRNA expression also strongly downregulated early and late estrogen response genes (corrected FET  $p = 5.90e-10$ , FE = 1.95) suggestive of a role for *EPRS* in direct regulation of ER signaling (Figure 4A).

### ***EPRS* regulates expression of *ESR1* and *ESR1* target transcripts**

As *EPRS* knockdown downregulated estrogen response genes, we hypothesized that *EPRS* may regulate cell proliferation through regulation of ER signaling. Using the Broad Institute’s Connectivity Map (CMap) [20], we observed striking concordance between gene expression changes induced by *EPRS* shRNA expression and those induced in MCF7 cells by treatment with clinically used anti-endocrine therapies, such as fulvestrant (CMap score = 0.842,  $p = 8.76e-128$ ), clomifene (CMap score = 0.69,  $p = 1.39e-86$ ), raloxifene (CMap score = 0.559,  $p = 6.77e-52$ ), and tamoxifen (CMap score = 0.552,  $p = 3.56e-50$ ) (Figure 4B). We similarly observed a strong inverse association between gene expression changes induced by *EPRS* knockdown and those induced by estradiol treatment of MCF7 cells (CMap score = -0.691,  $p = 6.87e-87$ ) (Figure 4B).

To further validate the link between *EPRS* and ER signaling, we confirmed, by qRT-PCR, that *EPRS* shRNA expression downregulated ER $\alpha$  and the *ESR1* targets *CCND1*, *FOXM1*, *MYC*, and *PGR* (Figure 4C). Furthermore, downregulation of *ESR1* in MCF7 TamR cells, induced growth arrest and prevented phosphorylation of Rb and upregulation of the S/G2 proteins CDK2, and CCNB1 (Supplementary Figure S4A), phenocopying *EPRS* perturbation. This is consistent with persistent dependence of

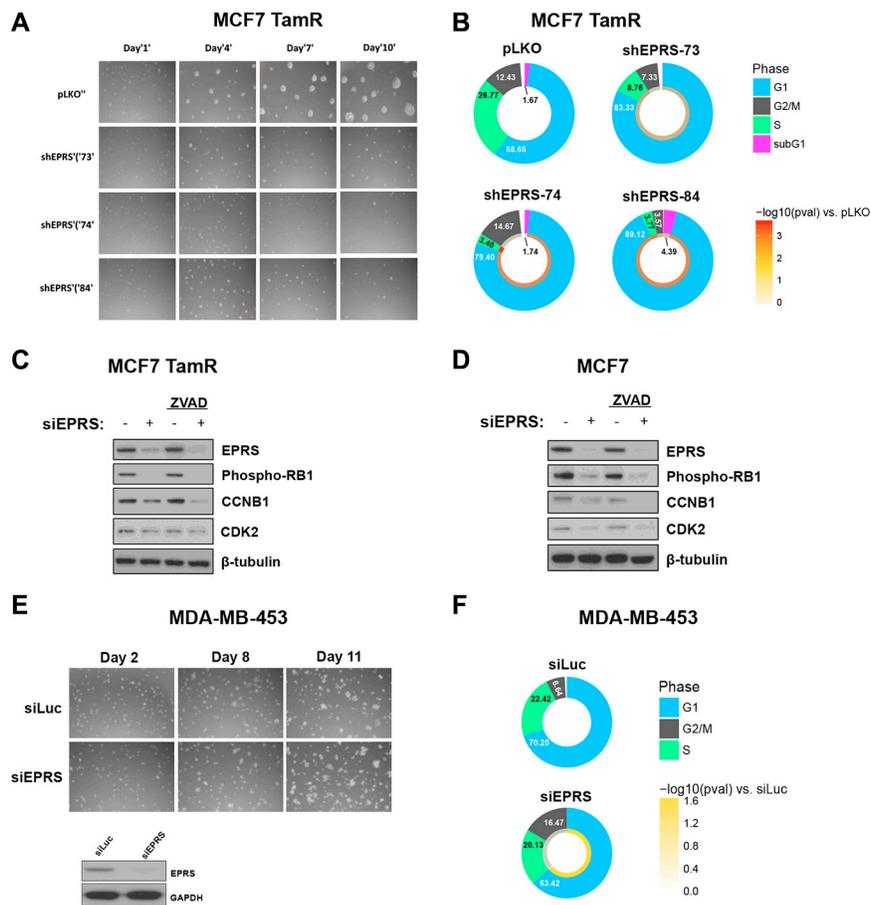


**Figure 2: *EPRS* is associated with tamoxifen resistance.** (A) Elevated *EPRS* expression is associated with reduced recurrence-free survival in patients treated with tamoxifen alone for five years (Symmans). (B) *EPRS*-correlated genes are enriched for gene signatures of nonresponse to tamoxifen in breast cancer patients. Numbers represent Fold Enrichment.

endocrine-resistant breast cancer cells on *ESR1* expression and activity. Based on these results, we hypothesized that reconstitution of *ESR1* expression may rescue the cell cycle arrest we observed with *EPRS* knockdown. Expression of exogenous *ESR1* with *EPRS* siRNA rescued *ESR1* expression and partially rescued siEPRS-induced downregulation of *CCNB1* and phospho-*RB* (Supplementary Figure S4B and S4C), however, it was not able to rescue G1/S arrest by flow cytometry (data not shown).

Based on these results, we hypothesized that *EPRS* may regulate *ESR1* coactivators and corepressors, as well as other regulators of estrogen signaling. To identify these potential ER regulators in an unbiased manner, we constructed a Bayesian gene regulatory network, as previously described [21], using ER+ breast cancer samples from the TCGA ( $n = 623$ ), METABRIC ( $n = 1505$ ), Miller ( $n = 213$ ) and Wang ( $n = 209$ ) datasets and transcription factor-target interactions from the ENCODE project [22] and MCF7 ChIP-chip data [23].

Bayesian network inference uses directed acyclic graphs (DAGs) to model the joint probability distribution of the states of multiple variables (e.g., genes), characterizing the dependent and independent relationships among genes, which is extremely useful for unraveling gene expression regulatory structures. Bayesian networks constructed from individual datasets were combined by the union of directed edges into a super Bayesian network comprised of 20,810 nodes connected by 73,283 edges. The gold standard for network validation compares a node's network neighborhoods with experimentally-derived perturbation signatures of that node. The *EPRS*-regulated subnetwork is highly enriched for genes downregulated by shEPRS (Figure 5), validating the ability of the Bayesian network to predict gene regulatory relationships. For example, the 5- and 6-layer downstream neighborhoods of *EPRS* are significantly enriched for the *EPRS* knockdown signature with FET  $p$  values =  $3.48e-11$  (1.29 fold) and  $2.92e-26$  (1.23 fold), respectively.



**Figure 3: *EPRS* is necessary for proliferation ER+, but not ER- breast cancer cells.** (A) *EPRS* knockdown inhibits growth of MCF7 TamR cells in 3D Matrigel™ culture. (B) Flow cytometric analysis of PI-stained DNA. Numbers are mean percentages from three independent experiments. Color ring represents  $-\log_{10}$  (student's  $t$ -test  $p$  values). C&D) Representative immunoblot of G1/S/G2 proteins in MCF7 TamR (C) and parental (D) cells treated with *EPRS* or control siRNA in the absence and presence of 20  $\mu$ M Z-VAD-FMK. (E) *EPRS* knockdown does not inhibit growth of ER- MDA-MB-453 cells in 3D Matrigel™ culture. Representative immunoblot of *EPRS* knockdown in MDA-MB-453 cells. (F) Flow cytometric cell cycle analysis of PI-stained MDA-MB-453 cells. Numbers are mean percentages from three independent experiments. siLuc: luciferase-targeting siRNA (control). siEPRS: pool of four unique siEPRS-targeting siRNAs.

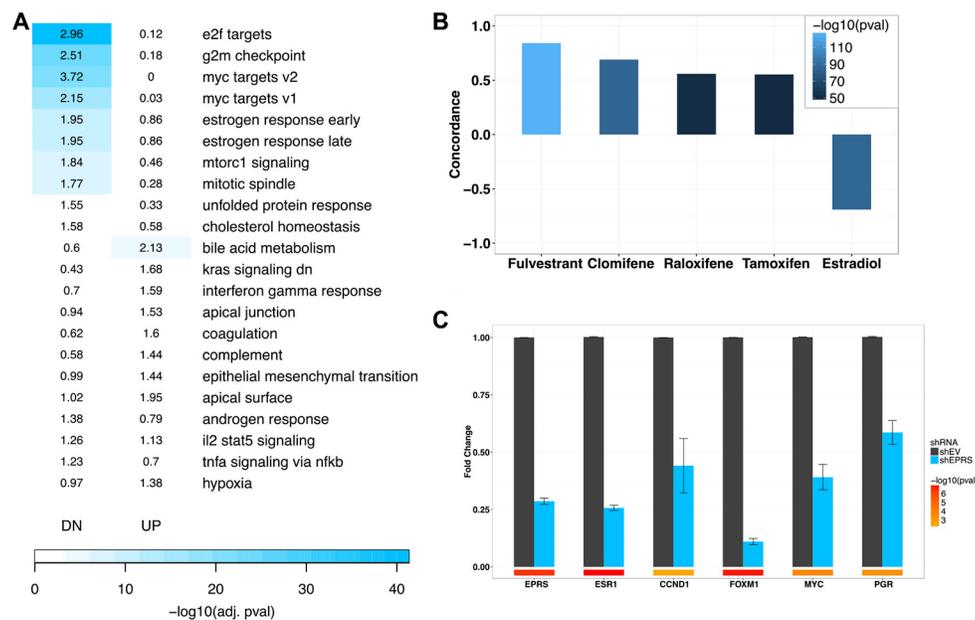
To map the direct and indirect regulation of *ESR1* by *EPRS*, we calculated the shortest network paths between *EPRS* and 452 known *ESR1* targets and pathway genes, as described in methods. The genes comprising these paths were then used to derive an *EPRS-ESR1* subnetwork (EES). The EES consists of the 452 known *ESR1* transcriptional target genes and 1079 intermediate nodes connected by 2902 directed edges (Figure 6A).

We performed Key Driver Analysis (KDA) [24, 25] on the EES to identify its master regulators downstream of *EPRS*. This yielded 80 genes (Figure 6) (Supplementary Table S1) predicted to be the drivers of the EES. Many predicted drivers, such as *ESR1* itself, *GATA3*, *BRD4*, *PGR*, and *PTGES3* are known regulators of *ESR1* signaling or are *ESR1* effector genes. To confirm that these predicted network drivers are contributing to *EPRS*-mediated regulation of estrogen signaling, we utilized the LINCS dataset to compare the network drivers' perturbation signatures to those of *EPRS* and *ESR1*. For the 41 EES network drivers that had differentially expressed genes in the corresponding perturbation experiments in LINCS their perturbation signatures all significantly overlap with the *EPRS* (corrected FET  $p \leq 5.04e-04$ ) and *ESR1* (corrected FET  $p \leq 1.28e-24$ ) signatures (Figure 6B). The overlap between the *EPRS* and *ESR1* LINCS signatures is highly significant (corrected FET  $p = 3.6e-143$ , OR = 8.86) (Figure 6B). When compared to LINCS shRNA signatures of genes not in the EES, LINCS shRNA signatures of EES keydrivers and of EES genes show significantly greater overlaps with the *EPRS* signature (Kolmogorov-Smirnov  $p = 9.98e-03$  and  $1.61e-03$ , respectively) (Figure 6C).

Twenty of the 80 EES drivers are downregulated by sh*EPRS*, and 11 are upregulated (Figure 6D). Nine are previously reported protein-protein interactors of *EPRS* (*HSPA8*, *ESR1*, *HSPD1*, *GATA3*, *PGR*, *PTGES3*, *KIF5B*, *PTPN11*, *FOS*) [26], and 14, including *EPRS*, were previously-reported protein-protein interactors of *ESR1* (*EPRS*, *SCYL2*, *UPF1*, *GLYR1*, *MED16*, *ACTR3*, *CCT2*, *HSPA8*, *PGR*, *PTGES3*, *RP20*, *BCLAF1*, *POLR2E*, *FOS*) [27] (Figure 6D). *HSPA8*, *PGR*, *PTGES3*, and *FOS* are interactors common to both *EPRS* and *ESR1* (Figure 6D). Of the 80 EES drivers, 45 are upregulated in breast tumors compared to adjacent normal breast in both TCGA and METABRIC cohorts, and 61 are upregulated in at least one cohort (Figure 6D). Thirteen are downregulated in both TCGA and METABRIC, and 25 are downregulated in at least one dataset (Figure 6D). Sixteen of the EES drivers are significantly (FDR < 0.10) associated with overall survival in both Cox and KM models, and 43 in at least one model. Furthermore, 34 predicted EES drivers are known ER signaling regulators or downstream targets (Supplementary Table S2), while the remaining 46 have not, to our knowledge, been described in the context of estrogen signaling.

## DISCUSSION

*EPRS* has been previously implicated in breast tumorigenesis, but the underlying mechanisms remained unclear. We have demonstrated that *EPRS* is the most significantly upregulated ARS gene in ER+ breast cancers, partly attributable to *EPRS* copy number gains, and that



**Figure 4: *EPRS*-regulated transcriptome.** (A) Enrichment of sh*EPRS* differentially-expressed genes for MSigDB Hallmark genesets. Numbers represent Fold Enrichment. (B) sh*EPRS* differential expression signature overlaps with estrogen receptor modulators. (C) QPCR validation of estrogen receptor target downregulation. Error bars represent SEM. Color bars represent  $-\log_{10}$  (student's *t*-test *p* values). *EPRS* shRNA-74 shown.

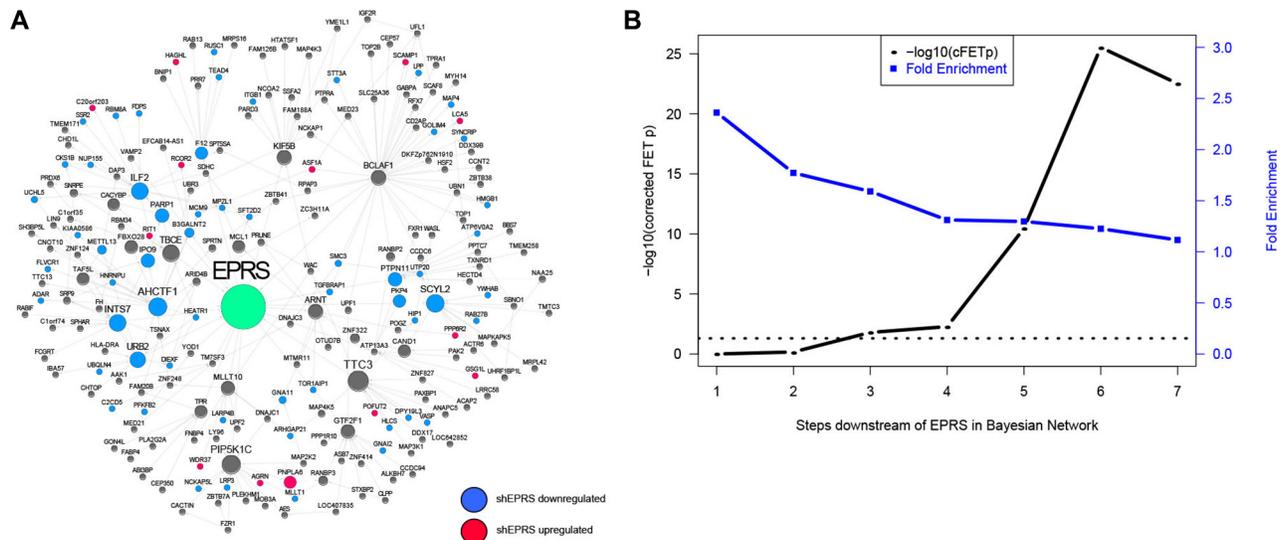
elevated *EPRS* expression is associated with reduced overall survival in patients with ER+ breast cancers. Higher *EPRS* transcript levels are also associated with reduced distant relapse-free survival in patients treated with adjuvant tamoxifen monotherapy. We further show that *EPRS*-correlated genes are highly enriched for signatures predictive of nonresponse to tamoxifen therapy. Depletion of *EPRS* resulted in mitotic arrest of tamoxifen resistant and parental ER+ breast cancer cells, but not of ER- cells. Transcriptomic profiling confirmed downregulation of cell cycle genes, and showed strong enrichment for downregulation of estrogen response genes. Applying a Bayesian network approach to 2550 ER+ breast cancer samples, we subsequently constructed an *EPRS*-centered directed estrogen signaling network, identified its driver genes, and validated them using publicly-available MCF7 shRNA perturbation signatures. This unbiased approach enabled us to identify 34 known regulators of estrogen signaling, such as *ESR1* itself, *GATA3*, *BRD4*, *PGR*, and *PTGES3*, and 46 genes which, to our knowledge, have not been previously associated with endocrine signaling. We have shown these drivers to share strikingly similar perturbation signatures, but further work must be done to assess their functional roles in ER+ breast cancer cells. In addition to gene knockdown and knockout experiments, it will be interesting to see if drugs that reverse the *EPRS* gene signature are able to inhibit the growth of ER+ breast cancer cells *in vitro* and *in vivo*. We are thus the first, to our knowledge, to identify a regulatory link between *EPRS* and estrogen signaling and the first to provide mechanistic insight into this relationship.

The *EPRS* depletion phenotype we observed is distinct from that recently reported in the literature. Beltran et al. reported that *EPRS* inhibition by shRNA or

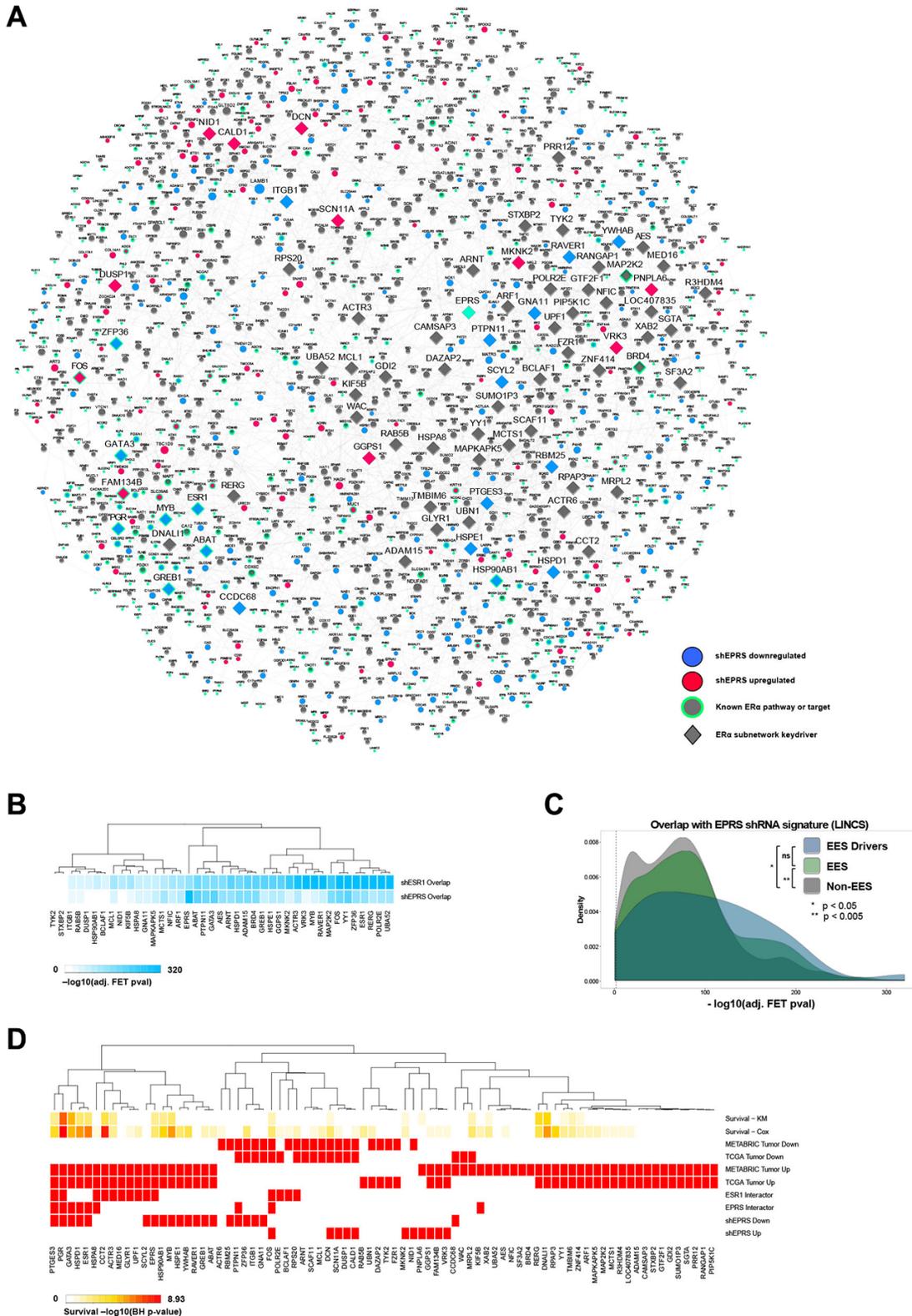
the ProRS inhibitor halofuginone (HF) induced a stress response and cell death in basal breast cancer cells [5]. Their work is consistent with another paper in which primary mouse CD4+ CD25- cells were treated with HF, resulting in downregulation of *COL1A1*, *COL1A2*, and *S100A4*, and upregulation of *DDIT3* mRNA [18]. Microarray profiling of HF-treated mouse mammary epithelial cells similarly revealed induction of *Ddit3*, *Trib3*, *Nrdg1*, *Gadd45a*, *Slc1a4*, and other genes implicated in cellular response to stress [19]. In contrast, we did not observe differential expression of these genes following *EPRS* knockdown (not shown), suggesting depletion of total *EPRS* protein affects distinct processes compared to specific inhibition of the *EPRS* ProRS domain. An alternate explanation is that ER+ breast cancer cells are more dependent on *EPRS* to maintain ER-mediated growth-promoting signaling. In support of this, *EPRS* depletion in the ER-/HER2+ breast cancer cell line MDA-MB-453 did not inhibit growth. However, we do not discount the possibility of ER signaling-independent mechanisms through which *EPRS* may affect cell growth.

*EPRS* has previously been shown to regulate transcript-specific translation in macrophages and hepatocellular carcinoma cells through the gamma interferon-activated inhibitor of translation (GAIT) system [6, 7, 17, 28–33]. Although GAIT activity has not been demonstrated in other cell types, including breast cancer cells, several genes predicted to be regulated by GAIT, such as *RORA*, *NRIP1*, and *DZIP3* have been shown to mediate ER $\alpha$ -dependent proliferation of breast cancer cells [34–39].

Further work is required to validate and fully develop the molecular interactome linking *EPRS* to estrogen signaling and cell proliferation in ER+ breast



**Figure 5: Validation of *EPRS* Bayesian network.** (A) An network neighborhood around *EPRS*. Genes differentially-expressed following *EPRS* knockdown are highlighted in blue (downregulated) or red (upregulated). (B) Enrichment of *EPRS* downstream network for the *EPRS*-downregulated RNA-seq signature. Horizontal dashed line represents  $-\log_{10}(0.05)$ .



**Figure 6: *EPRS*-regulated *ESR1* network.** (A) *EPRS* neighborhood encompassing *ESR1* target and pathway genes. (B) Overlap of keydriver shRNA signatures with those of *EPRS* and *ESR1* for genes with perturbations available in the LINCS shRNA dataset. (C) As a group, EES driver and EES perturbation signatures show more significant overlaps with the sh*EPRS* signatures than do non-EES genes. (D) *EPRS* estrogen network keydriver association with survival using Cox and KM models, differential expression in TCGA and METABRIC ER+ tumors compared to adjacent normal breast, protein interactions with *ESR1* and *EPRS*, and differential expression following *EPRS* knockdown.

cancer cells. The necessity of *EPRS* in non-malignant cells, as well as the therapeutic window of *EPRS* inhibition must also be determined, although this is beyond the scope of this paper. *EPRS* and the EES it regulates may be a promising target for development of novel therapies, or the repurposing of existing therapies, to treat patients with ER+ breast cancers whose tumors do not respond to currently-used endocrine modulators.

## MATERIALS AND METHODS

### Breast tumor expression data

#### TCGA

The level 3 IlluminaHiSeq-RNASeqV2 (RSEM-normalized) expression data from the TCGA data portal was first log-transformed and quantile-normalized. The quantile-normalized data was then split into the tumor and adjacent normal groups. The tumor and adjacent normal data were then corrected by linear regression for confounding factors including batch, tissue source site, center and plate, race and age. Male samples were excluded.

#### METABRIC

Normalized Illumina HT12v3 mRNA microarray data was downloaded from the European Bioinformatics Institute and corrected for age.

#### Miller and wang

Miller et al. 2005 [40] and Wang et al. 2005 [41] Affymetrix Human Genome U133A array expression datasets were obtained from Bioconductor [42] through the packages “breastCancerUPP” (Miller et al. 2005) and “breastCancerVDX” (Wang et al. 2005) and corrected for covariates as described above.

#### Symmans

Affymetrix Human Genome U133A array expression data was downloaded from the Gene Expression Omnibus (GEO) (GSE17705) and corrected for age and batch.

### CNA analysis

Copy number alterations in the TCGA breast cancer cohort were obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC) [43] and ER+ samples were extracted. METABRIC copy number alterations were called by Curtis et al. 2012 [44]. From the somatic copy number aberration (SCNA) segments called from Curtis *et al.* we first subset segments for ER+ patients (1505 out of 2000 samples combining discovery and validation data sets), and counted the number of segments called with copy number gain (denoted as GAIN) or amplification

(denoted as AMP) within 2Mbps of transcription starting site of *EPRS* from hg18 (NCBI build 36).

### Survival analysis

*EPRS* expression from the previously described TCGA, METABRIC, and Symmans breast cancer cohorts, was split into ER+ and ER- groups. Within each ER status-specific *EPRS* profile, we defined “*EPRS*-high” and “*EPRS*-low” subgroups of patients by median *EPRS* expression value. We performed the Kaplan-Meier analysis for *EPRS*-high and *EPRS*-low in both ER+ and ER- cohorts as well as Cox regression analysis to evaluate the prognostic significance of the subgroups. Correction for multiple hypothesis testing was done using the Benjamini-Hochberg (BH) method [45].

### Tamoxifen response signature

A tamoxifen response/nonresponse signature was identified by combining published gene signatures predictive of response [13–15]. Our final signature consisted of 364 genes comprised of 124 predictive of tamoxifen response and 240 predictive of tamoxifen nonresponse (Supplementary Table S1).

### *EPRS*-correlated genes

Spearman correlations between *EPRS* and all other genes in TCGA and METABRIC datasets were computed and *p*-values were corrected using the BH method [45]. Significant correlations were defined by BH corrected *p*-value < 0.05 and absolute value of Spearman’s rho > 0.2.

### Quantitative real-time PCR

Total RNA was isolated from cells using the Qiagen RNeasy kit. RNA concentration was determined using a NanoDrop 8000 spectrophotometer. Fifty nanograms of total RNA were used per triplicate qRT-PCR reaction and detected using SYBR Green fluorescence. *GAPDH* was used as an internal control. QRT-PCR data was analyzed using the delta-Ct method.

### Cell culture

MCF7 parental cells were cultured in RPMI media with 10% FBS (Life Technologies). MCF7 TamR cells, obtained from the Rachel Schiff Lab (Houston, TX) [46, 47], were cultured in phenol red-free RPMI media (Life Technologies) with 10% charcoal-stripped fetal bovine serum (FBS) (Sigma), one percent penicillin/streptomycin (Life Technologies), and 100 nM 4-hydroxytamoxifen (4-OHT). 4-OHT was withdrawn for functional assays. MDA-MB-453 cells were cultured in DMEM media (Life

Technologies) with 10% FBS (Life Technologies) and one percent penicillin/streptomycin (Life Technologies).

### 3D Matrigel™ cultures

Eight-well chamber slides (BD Biosciences) were coated with Matrigel™. Cells were plated at a density of  $3\text{--}5 \times 10^3$  cells per well in the middle four wells, and cells were allowed to grow for two to three weeks.

### RNAi

For siRNA, cells were reverse-transfected with 20 nM siRNA using Lipofectamine RNAiMAX (Life Technologies) according to the manufacturer's protocol. Media was changed after overnight incubation. *EPRS* shRNAs in the pLKO.1 vector were obtained from Sigma. For shRNA lentivirus production, 293T cells were transfected with viral plasmids using Lipofectamine 2000 according to standard protocols. Virus-containing supernatant was collected at 48, 72, 96, and 120 hours after transfection, pooled, and frozen at  $-80^\circ\text{C}$  to eliminate carry-over 293T cells. For shRNA lentivirus infection of target cells, lentivirus was thawed on ice and concentrated using centrifugal filters (Amicon). ShRNA lentivirus and target cells were simultaneously added to six-well plates, centrifuged at 2250 rpm for 30 minutes at room temperature, then incubated at  $37^\circ\text{C}$  overnight. Media was changed the next morning and cells were allowed to recover for six to eight hours, after which  $2 \mu\text{g}/\text{mL}$  puromycin was added to select shRNA-expressing cells. Puromycin was reduced to a maintenance dose of  $1 \mu\text{g}/\text{mL}$  after 24–48 h. For Z-VAD-FMK (BD Biosciences; San Jose, CA) treatment, cells were treated with  $20 \mu\text{M}$  Z-VAD-FMK or DMSO.

### Cell cycle analysis

Cells were infected with shRNA or transfected with siRNA as described above. Twenty-four hours after siRNA transfection or 48 hours after shRNA transfection and puromycin selection, cells were serum-starved in serum-free growth media for 24 hours, after which serum was reintroduced and cells are incubated for an additional 48 hours. Five days after RNAi introduction, cells were trypsinized, split, and lysed for total protein for Western blot analysis or fixed in 80% ethanol and stored at  $4^\circ\text{C}$  for flow cytometric analysis. Cells were then stained with propidium iodide/RNAase (BD Pharmingen) for 30 minutes at  $37^\circ\text{C}$  and analyzed on a BD FACSCanto flow cytometer. Cell cycle analysis was performed with FlowJo (TreeStar Inc) using both Watson-Pragmatic [48] and Dean/Jett/Fox [49] algorithms.

### RNA-seq analysis

Cells expressing *EPRS* shRNA vectors were generated. Seventy-two hours after infection, cells were lysed and total RNA was purified using a Qiagen RNeasy kit according to the manufacturer's protocol. RNA was stored at  $-80^\circ\text{C}$  until submission to the Mount Sinai Genomics Core Facility for ribosomal RNA depletion, cDNA library preparation, and sequencing using paired-end, 100 nt reads on an Illumina HiSeq 2000. The paired-end sequencing reads were aligned to human genome hg38 using star aligner (version 2.5.0 b) [50]. Following read alignment, featureCounts [51] was used to quantify gene expression at the gene level based on GENCODE gene model release 22. Gene expression was normalized as counts per million (TPM) using trimmed mean of M-values normalization (TMM) method [52] to adjust for sequencing library size difference. Differential gene expression was predicted using the Bioconductor package limma [53]. The false discovery rate (FDR) of the differential expression test was estimated using the Benjamini–Hochberg method. We used a cutoff of corrected  $p < 0.05$  and fold change  $> 1.3$  in at least two of three *EPRS*-targeting shRNAs. Overlap between independent *EPRS*-targeted shRNAs was tested for statistical significance and visualized using the SuperExactTest [54]. Total RNA used for RNA-seq was subsequently used for validation by qRT-PCR.

### Connectivity map query

The differentially expressed genes (DEGs) between the sh*EPRS* and control samples were compared with the drug treatment gene expression data from the Connectivity MAP (CMap) database, a reference collection of gene expression data from 5 cultured human cell lines treated with 1309 drug compounds [55]. In the original CMap study, the Kolmogorov-Smirnov (KS) statistic was employed to compute the similarity between an input DEG signature and a drug's gene signature. A significant KS statistic suggests some functional connection between drug treatment and the biology represented by a given input DEGs [55]. In this study, we utilized a weighted KS test which was previously proposed for gene set enrichment analysis [56] to calculate the drug connectivity score. We described the analytic approach below.

Let  $R$  be the  $N$  genes which are rank sorted in descending order of the expression fold change in a CMap drug instance. Next the ranks are converted to weighted gene-drug correlations using a formula  $r = \frac{\text{Rank} - \text{mean}(\text{Rank})}{\text{mean}(\text{Rank})}$ ,

through which the genes ranked in the top or bottom (i.e. the most induced or suppressed) had absolute drug

correlation scores close to 1, while genes ranked in the middle (possibly the least targeted by the drug) had gene-drug correlation measures close to 0. Let  $G$  be a set of  $n$  input DEGs, which is comprised of two subsets,  $G_1$  and  $G_2$ , representing the significantly up- and down-regulated genes, respectively. For each subset  $G_x$  ( $x = 1$  or  $2$ ), we then compute a drug connectivity scores  $S_x$  as the maximum derivation from zero, i.e.,  $f_1(i) - f_2(i)$ , where  $i$  ( $= 1, \dots, N$ ) indices a position in the rank of  $R$ , and

$$f_1(i) = \sum_{\substack{g_j \in G_x \\ j \leq i}} \frac{|r_j|^p}{M}, \text{ where } M = \sum_{g_j \in G_x} |r_j|^p,$$

$$f_2(i) = \sum_{\substack{g_j \notin G_x \\ j \leq i}} \frac{1}{N - n_x}$$

Here,  $j$  is an index with a value between 1 and  $i$ , and  $g_j$  denotes the  $j$ th gene in the rank sorted list  $R$ . Finally, the drug connectivity score  $S = 0$  if  $S_1 \times S_2 \geq 0$ , and  $S = S_1 - S_2$  otherwise. When  $p = 0$ , this method reduces to standard KS test. We set  $p = 1$  for the analyses in this paper.

The statistical significance of  $S$  can be estimated from permutation. The sign of  $S$  reflects the direction of the drug treatment with respect to the input gene signatures. A positive  $S$  implies that the drug has a gene signature that is concordant to the input signature and that drug could induce the same transcriptional regulation pattern of the input genes, while a negative  $S$  implies that the drug has a gene expression signature that is opposite to the input signature and the drug could potentially reverse the disease trait.

## Bayesian network construction

We employed a Monte Carlo Markov Chain (MCMC) simulation process based approach [57] to infer probabilistic regulatory relationship between genes. A uniform prior was used for the regulatory relation between pair of genes. As a uniform prior is unable to break Markov equivalence, we need additional data to assist with the identification of directionality in the Bayesian network construction. For this purpose, known transcription factor (TF)-target pairs were downloaded from the ENCODE project [22] and a dataset of nuclear receptor binding sites in breast cancer cells [58]. We allowed TF nodes to be parent nodes of their targets, but targets were not allowed to be the parent nodes of their TF in the network. As in [57], we followed a network averaging strategy in which 1,000 networks were generated by this MCMC process starting with different random structure, and links that appeared in more than 30% of the networks were used to define a final consensus network. If loops were present in the consensus network, the weakly supported link involved in a loop was removed to ensure the final network structure was a directed acyclic graph. Bayesian networks constructed from independent datasets were then combined by the union of directed edges, and loops were removed.

## EPRS subnetwork

Estrogen signaling pathway gene sets were obtained from the Pathway Interaction Database (ERalpha pathway) [59], MSigDB Hallmark gene sets (Estrogen response early; Estrogen response late) [16], KEGG (Estrogen Signaling) [60], Gene Ontology (Intracellular estrogen receptor signaling pathway) [60], and recent literature findings [61–65] for a total of 452 unique estrogen signaling genes (ESG). The shortest paths between *EPRS* and all genes in ESG in the super Bayesian network were computed. *EPRS*, ESGs, and all intermediate connecting nodes were called the *EPRS-ESR1* subnetwork (EES). Keydriver Analysis (KDA) was performed on the EES as previously described [24].

## LINCS differential expression signature calling

Normalized level 3 expression data was downloaded from the Library of Integrated Network-based Cellular Signatures (LINCS) cloud website (<http://www.lincscloud.org/11000/>). LINCS data includes the directly measured expression levels of 978 landmark genes and the inferred expression levels of more than 21 K other genes. Gene expression profiles of both directly measured landmark transcripts and inferred genes were normalized using an 80-gene invariant set scaling followed by the quantile-normalization. Differentially-expressed gene signatures for each shRNA were called using the *limma* R package based on the criteria of significance  $p < 0.05$  and fold-change  $> 1.2$ .

## Reagents

### Antibodies

*EPRS* (Abcam ab31531); *ESR1* (Cell Signaling 8644); phospho-*RB* (Cell Signaling 9301); total *RB* (Cell Signaling 9309); *CDK2* (Cell Signaling 2546); *CCNB1* (Cell Signaling 12231); *GAPDH* (Cell Signaling 2118),  $\beta$ -*tubulin* (Cell Signaling 2128), *HSP70* (Cell Signaling 4872).

### QPCR primers

*EPRS* (sense: 5' -GCCTTCAGGGACAGTAAGCA - 3', antisense: 5' - ATGAAGTTGCTGCACAAGGG - 3'); *ESR1* (sense: 5' - AGAGGGTGCCAGGCTTTGT - 3', antisense: 5' - CAGACGAGACCAATCATCAGG - 3'); *CCND1* (QuantiTect Primer Assay # QT00495285, Qiagen); *MYC* (sense: 5' - CACCGAGTCGTAGTCG AGGT - 3', antisense: 5' - TTTCGGGTAGTGGAAAA CCA - 3'); *PGR* (sense: 5' - AGCCAGAGCCCACAAT ACAG - 3', antisense: 5' - GACCTTACAGCTCCC ACAGG - 3'); *FOXMI* (sense: 5' -TGCAGCTAGGGATGT GAATCTTC - 3', antisense: 5' - GGAGCCCAGTCCAT CAGA ACT - 3').

### RNAi

*EPRS* and *ESR1* siGENOME SMARTpool™ pools of four siRNA sequences (Dharmacon GE Healthcare). shRNA: pLKO-empty vector, shEPRS-73

(TRCN0000293873, 5' – CCGGGCCAAGTACTACACCT TATTTCTCGAGAAATAAGGTGTAGTACTTGGCTT TTTG – 3'), shEPRS-74 (TRCN0000293874, 5' – CCG GATGAACCTGTTAGCCCATATACTCGAGTATATGG GCTAACAGGTTCAATTTTTG – 3'), and shEPRS-84 (TRCN0000286384, 5' - CCGGGCCTGGCAAGAACAG TTGAAACTCGAGTTTCAACTGTTCTTGCCAGGCTT TTG - 3').

### Cell lines

Parental and TamR MCF7 cells were obtained from the Rachel Schiff Lab (Baylor College of Medicine, Houston, TX) [46, 47]; MDA-MB-453 cells were purchased from ATCC (Manassas, VA).

### Exogenous ESR1-expressing MCF7 TamR cell line

*ESR1* cDNA was obtained from Addgene (#11351) and was cloned into pBABE-puro. MCF7 TamR cells were stably infected with empty vector (EV)- or *ESR1*-pBABE-puro retrovirus and selected as described above.

## ACKNOWLEDGMENTS AND FUNDING

This work was supported by in part the grants R01AG046170 (to B.Z., W.S., M.W. and J. Z.) from the NIH/National Institute on Aging (NIA), R01CA163772 (to B.Z. and J.Z.) from NIH/National Cancer Institute (NCI), and U01AI111598-01 (to B.Z., W.S., M.W. and J. Z.) from NIH/National Institute of Allergy and Infectious Diseases (NIAID). Tamoxifen-resistant and parental MCF7 cells were kindly provided by Rachel Schiff at the Baylor College of Medicine (Houston, TX).

## CONFLICTS OF INTEREST

The authors declare no conflicts of interests.

## REFERENCES

1. Kim D, Kwon NH, Kim S. Association of aminoacyl-tRNA synthetases with cancer. *Top Curr Chem.* 2014; 344:207–45.
2. Line A, Slucka Z, Stengrevics A, Silina K, Li G, Rees RC. Characterisation of tumour-associated antigens in colon cancer. *Cancer Immunol Immunother.* 2002; 51:574–82.
3. Lu H, Knutson KL, Gad E, Disis ML. The tumor antigen repertoire identified in tumor-bearing neu transgenic mice predicts human tumor antigens. *Cancer Res.* 2006; 66:9754–61.
4. Kim S, You S, Hwang D. Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nat Rev Cancer.* 2011; 11:708–18.
5. Beltran AS, Graves LM, Blancafort P. Novel role of Engrailed 1 as a prosurvival transcription factor in basal-like breast cancer and engineering of interference peptides block its oncogenic function. *Oncogene.* 2013.
6. Jia J, Arif A, Ray PS, Fox PL. WHEP domains direct noncanonical function of glutamyl-Prolyl tRNA synthetase in translational control of gene expression. *Mol Cell.* 2008; 29:679–90.
7. Sampath P, Mazumder B, Seshadri V, Gerber CA, Chavatte L, Kinter M, Ting SM, Dignam JD, Kim S, Driscoll DM, Fox PL. Noncanonical function of glutamyl-prolyl-tRNA synthetase: gene-specific silencing of translation. *Cell.* 2004; 119:195–208.
8. Forozan F, Veldman R, Ammerman CA, Parsa NZ, Kallioniemi A, Kallioniemi OP, Ethier SP. Molecular cytogenetic analysis of 11 new breast cancer cell lines. *Br J Cancer.* 1999; 81:1328–34.
9. Bieche I, Champeme MH, Lidereau R. Loss and gain of distinct regions of chromosome 1q in primary breast cancer. *Clin Cancer Res.* 1995; 1:123–7.
10. Liu Y, Zhou R, Baumbusch LO, Tsavachidis S, Brewster AM, Do KA, Sahin A, Hortobagyi GN, Taube JH, Mani SA, Aaroe J, Warnberg F, Borresen-Dale AL, et al. Genomic copy number imbalances associated with bone and non-bone metastasis of early-stage breast cancer. *Breast Cancer Res Treat.* 2014; 143:189–201.
11. Spain SL, Carvajal-Carmona LG, Howarth KM, Jones AM, Su Z, Cazier JB, Williams J, Aaltonen LA, Pharoah P, Kerr DJ, Cheadle J, Li L, Casey G, et al. Refinement of the associations between risk of colorectal cancer and polymorphisms on chromosomes 1q41 and 12q13.13. *Hum Mol Genet.* 2012; 21:934–46.
12. Schnitt SJ. Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Mod Pathol.* 2010; 23:S60–4.
13. Symmans WF, Hatzis C, Sotiriou C, Andre F, Peintinger F, Regitnig P, Daxenbichler G, Desmedt C, Domont J, Marth C, Delaloge S, Bauernhofer T, Valero V, et al. Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol.* 2010; 28:4111–9.
14. Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AM, Gillet C, Ellis P, Ryder K, Reid JF, Daidone MG, Pierotti MA, Berns EM, et al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics.* 2008; 9:239.
15. Chanrion M, Negre V, Fontaine H, Salvétat N, Bibeau F, Mac Grogan G, Mauriac L, Katsaros D, Molina F, Theillet C, Darbon JM. A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. *Clin Cancer Res.* 2008; 14:1744–52.
16. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015; 1:417–425.

17. Mukhopadhyay R, Jia J, Arif A, Ray PS, Fox PL. The GAIT system: a gatekeeper of inflammatory gene expression. *Trends Biochem Sci.* 2009; 34:324–31.
18. Keller TL, Zocco D, Sundrud MS, Hendrick M, Edenius M, Yum J, Kim YJ, Lee HK, Cortese JF, Wirth DF, Dignam JD, Rao A, Yeo CY, et al. Halofuginone and other febrifugine derivatives inhibit prolyl-tRNA synthetase. *Nat Chem Biol.* 2012; 8:311–7.
19. Kamberov YG, Kim J, Mazitschek R, Kuo WP, Whitman M. Microarray profiling reveals the integrated stress response is activated by halofuginone in mammary epithelial cells. *BMC Res Notes.* 2011; 4:381.
20. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science.* 2006; 313:1929–35.
21. Zhu J, Sova P, Xu Q, Dombek KM, Xu EY, Vu H, Tu Z, Brem RB, Bumgarner RE, Schadt EE. Stitching together Multiple Data Dimensions Reveals Interacting Metabolomic and Transcriptomic Networks That Modulate Cell Regulation. *PLoS Biol.* 2012; 10:e1001301.
22. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489:57–74.
23. Kittler R, Zhou J, Hua S, Ma L, Liu Y, Pendleton E, Cheng C, Gerstein M, White KP. A comprehensive nuclear receptor network for breast cancer cells. *Cell Rep.* 2013; 3:538–51.
24. Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, Zhang C, Xie T, Tran L, Dobrin R, Fluder E, Clurman B, Melquist S, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell.* 2013; 153:707–20.
25. Huan T, Zhang B, Wang Z, Joehanes R, Zhu J, Johnson AD, Ying S, Munson PJ, Raghavachari N, Wang R, Liu P, Courchesne P, Hwang SJ, et al. A systems biology framework identifies molecular underpinnings of coronary heart disease. *Arterioscler Thromb Vasc Biol.* 2013; 33:1427–34.
26. Lee JH, You S, Hyeon do Y, Kang B, Kim H, Park KM, Han B, Hwang D, Kim S. Comprehensive data resources and analytical tools for pathological association of aminoacyl tRNA synthetases with cancer. *Database (Oxford).* 2015.
27. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, Reguly T, Nixon J, Ramage L, et al. The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* 2015; 43:D470–8.
28. Arif A, Chatterjee P, Moodt RA, Fox PL. Heterotrimeric GAIT complex drives transcript-selective translation inhibition in murine macrophages. *Mol Cell Biol.* 2012; 32:5046–55.
29. Arif A, Jia J, Moodt RA, DiCorleto PE, Fox PL. Phosphorylation of glutamyl-prolyl tRNA synthetase by cyclin-dependent kinase 5 dictates transcript-selective translational control. *Proc Natl Acad Sci USA.* 2011; 108:1415–20.
30. Arif A, Jia J, Mukhopadhyay R, Willard B, Kinter M, Fox PL. Two-site phosphorylation of EPRS coordinates multimodal regulation of noncanonical translational control activity. *Mol Cell.* 2009; 35:164–80.
31. Mukhopadhyay R, Ray PS, Arif A, Brady AK, Kinter M, Fox PL. DAPK-ZIPK-L13a axis constitutes a negative-feedback module regulating inflammatory gene expression. *Mol Cell.* 2008; 32:371–82.
32. Ray PS, Fox PL. A post-transcriptional pathway represses monocyte VEGF-A expression and angiogenic activity. *EMBO J.* 2007; 26:3360–72.
33. Yao P, Potdar AA, Arif A, Ray PS, Mukhopadhyay R, Willard B, Xu Y, Yan J, Saidel GM, Fox PL. Coding region polyadenylation generates a truncated tRNA synthetase that counters translation repression. *Cell.* 2012; 149:88–100.
34. Dong C, Yuan L, Dai J, Lai L, Mao L, Xiang S, Rowan B, Hill SM. Melatonin inhibits mitogenic cross-talk between retinoic acid-related orphan receptor alpha (RORalpha) and ERalpha in MCF-7 human breast cancer cells. *Steroids.* 2010; 75:944–51.
35. Odawara H, Iwasaki T, Horiguchi J, Rokutanda N, Hirooka K, Miyazaki W, Koibuchi Y, Shimokawa N, Iino Y, Takeyoshi I, Koibuchi N. Activation of aromatase expression by retinoic acid receptor-related orphan receptor (ROR) alpha in breast cancer cells: identification of a novel ROR response element. *J Biol Chem.* 2009; 284:17711–9.
36. Xiong G, Wang C, Evers BM, Zhou BP, Xu R. RORalpha suppresses breast tumor invasion by inducing SEMA3F expression. *Cancer Res.* 2012; 72:1728–39.
37. Zhu Y, McAvoy S, Kuhn R, Smith DI. RORA, a large common fragile site gene, is involved in cellular stress response. *Oncogene.* 2006; 25:2901–8.
38. Docquier A, Harmand PO, Fritsch S, Chanrion M, Darbon JM, Cavailles V. The transcriptional coregulator RIP140 represses E2F1 activity and discriminates breast cancer subtypes. *Clin Cancer Res.* 2010; 16:2959–70.
39. Purcell DJ, Chauhan S, Jimenez-Stinson D, Elliott KR, Tsewang TD, Lee YH, Marples B, Lee DY. Novel CARM1-Interacting Protein, DZIP3, Is a Transcriptional Coactivator of Estrogen Receptor-alpha. *Mol Endocrinol.* 2015; 29:1708–19.
40. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET, Bergh J. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA.* 2005; 102:13550–5.
41. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet.* 2005; 365:671–9.

42. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004; 5:R80.
43. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 2011; 39:D945–50.
44. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012; 486:346–52.
45. Benjamini, Y. and Hochberg Y, Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological.* 1995; 57:289–300.
46. Morrison G, Fu X, Shea M, Nanda S, Giuliano M, Wang T, Klinowska T, Osborne CK, Rimawi MF, Schiff R. Therapeutic potential of the dual EGFR/HER2 inhibitor AZD8931 in circumventing endocrine resistance. *Breast Cancer Res Treat.* 2014; 144:263–72.
47. Zhang Y, Tseng CC, Tsai YL, Fu X, Schiff R, Lee AS. Cancer cells resistant to therapy promote cell surface relocalization of GRP78 which complexes with PI3K and enhances PI(3,4,5)P3 production. *PLoS One.* 2013; 8:e80071.
48. Watson JV, Chambers SH, Smith PJ, A pragmatic approach to the analysis of DNA histograms with a definable G1 peak. *Cytometry.* 1987; 8:1–8.
49. Fox MH, A model for the computer analysis of synchronous DNA distributions obtained by flow cytometry. *Cytometry.* 1980; 1:71–7.
50. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29:15–21.
51. Liao Y, Smyth GK, Shi W, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014; 30:923–930.
52. Robinson MD, D.J. McCarthy, Smyth GK, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26:139–140.
53. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research.* 2015; 43:e47.
54. Wang M, Zhao Y, Zhang B, Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep.* 2015; 5:16923.
55. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet J-P, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, et al. The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science.* 2006; 313:1929–1935.
56. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–15550.
57. Zhu J, Wiener MC, Zhang C, Fridman A, Minch E, Lum PY, Sachs JR, Schadt EE. Increasing the Power to Detect Causal Associations by Combining Genotypic and Expression Data in Segregating Populations. *PLoS Comput Biol.* 2007; 3:e69.
58. Kittler R, Zhou J, Hua S, Ma L, Liu Y, Pendleton E, Cheng C, Gerstein M, White Kevin P. A Comprehensive Nuclear Receptor Network for Breast Cancer Cells. *Cell Reports.* 2013; 3:538–551.
59. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. PID: the Pathway Interaction Database. *Nucleic Acids Res.* 2009; 37:D674–9.
60. Kanehisa, M. and Goto S, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28:27–30.
61. Feng Q, Zhang Z, Shea MJ, Creighton CJ, Coarfa C, Hilsenbeck SG, Lanz R, He B, Wang L, Fu X, Nardone A, Song Y, Bradner J, et al. An epigenomic approach to therapy for tamoxifen-resistant breast cancer. *Cell Res.* 2014; 24:809–19.
62. Buchwalter G, Hickey MM, Cromer A, Selfors LM, Gunawardane RN, Frishman J, Jeselsohn R, Lim E, Chi D, Fu X, Schiff R, Brown M, Brugge JS. PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. *Cancer Cell.* 2013; 23:753–67.
63. Patani N, Jiang WG, Newbold RF, Mokbel K. Histone-modifier gene expression profiles are associated with pathological and clinical outcomes in human breast cancer. *Anticancer Res.* 2011; 31:4115–25.
64. Daniel AR, Gaviglio AL, Knutson TP, Ostrander JH, D'Assoro AB, Ravindranathan P, Peng Y, Raj GV, Yee D, Lange CA. Progesterone receptor-B enhances estrogen responsiveness of breast cancer cells via scaffolding PELP1- and estrogen receptor-containing transcription complexes. *Oncogene.* 2015; 34:506–15.
65. Fenne IS, Helland T, Flageng MH, Dankel SN, Mellgren G, Sagen JV. Downregulation of steroid receptor coactivator-2 modulates estrogen-responsive genes and stimulates proliferation of mcf-7 breast cancer cells. *PLoS One.* 2013; 8:e70096.