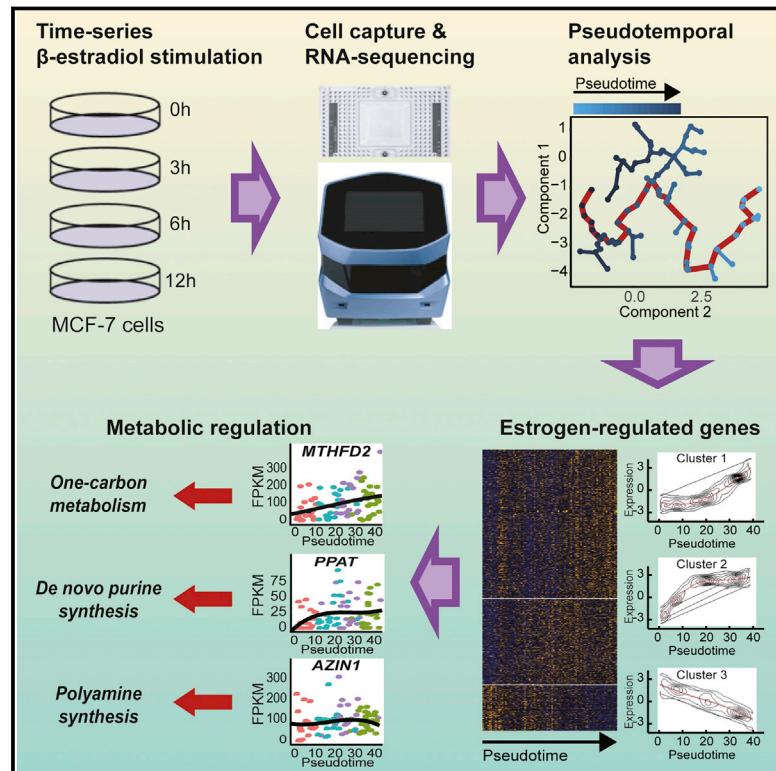


# Single-Cell Transcriptome Analysis Reveals Estrogen Signaling Coordinately Augments One-Carbon, Polyamine, and Purine Synthesis in Breast Cancer

## Graphical Abstract



## Authors

Detu Zhu, Zuxianglan Zhao,  
Guimei Cui, ..., Paul Robson, Yumei Luo,  
Edwin Cheung

**Correspondence**  
echeung@umac.mo

## In Brief

Zhu et al. perform single-cell RNA-seq to reveal a dynamic transcriptional network in ER $\alpha$ + breast cancer cells following estrogen stimulation and show that estrogen signaling promotes breast cancer cell survival and growth by mediating a metabolic switch in which folate-mediated one-carbon metabolism is reprogrammed via the mitochondrial folate pathway.

## Highlights

- Single-cell RNA-seq reveals estrogen-responsive genes in ER $\alpha$ + breast cancer cells
- Estrogen signaling induces a metabolic switch in breast cancer cells
- Estrogen signaling coordinately augments one-carbon, polyamine, and purine synthesis
- AZIN1 and PPAT are ER $\alpha$  targets that are essential for cell survival and growth



# Single-Cell Transcriptome Analysis Reveals Estrogen Signaling Coordinately Augments One-Carbon, Polyamine, and Purine Synthesis in Breast Cancer

Detu Zhu,<sup>1,2,5,10</sup> Zuxianglan Zhao,<sup>1,2,10</sup> Guimei Cui,<sup>1,2</sup> Shiehong Chang,<sup>1,2</sup> Lingling Hu,<sup>1,2</sup> Yi Xiang See,<sup>1,2,6</sup> Michelle Gek Liang Lim,<sup>3</sup> Dajiang Guo,<sup>1,2</sup> Xin Chen,<sup>1,2,7</sup> Barun Poudel,<sup>1,2,8</sup> Paul Robson,<sup>3,9</sup> Yumei Luo,<sup>4</sup> and Edwin Cheung<sup>1,2,3,11,\*</sup>

<sup>1</sup>Cancer Centre, University of Macau, Avenida da Universidade, Taipa, Macau, China

<sup>2</sup>Centre of Precision Medicine Research and Training, Faculty of Health Sciences, University of Macau, Avenida da Universidade, Taipa, Macau, China

<sup>3</sup>Genome Institute of Singapore, A\*STAR (Agency for Science, Technology and Research), Singapore, Singapore

<sup>4</sup>Key Laboratory for Major Obstetric Diseases of Guangdong Province, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong 510150, China

<sup>5</sup>Present address: Department of Genetic Medicine, Weill Cornell Medical College, New York, NY 10065, USA

<sup>6</sup>Present address: School of Biological Sciences, Nanyang Technological University, Singapore, Singapore

<sup>7</sup>Present address: School of Automation, Guangdong University of Technology, Guangzhou, China

<sup>8</sup>Present address: Department of Internal Medicine, University of Iowa, Iowa, USA

<sup>9</sup>Present address: The Jackson Laboratory for Genomic Medicine, Farmington, CT 06030, USA

<sup>10</sup>These authors contributed equally

<sup>11</sup>Lead Contact

\*Correspondence: echeung@umac.mo

<https://doi.org/10.1016/j.celrep.2018.10.093>

## SUMMARY

Estrogen drives breast cancer (BCa) progression by directly activating estrogen receptor  $\alpha$  (ER $\alpha$ ). However, because of the stochastic nature of gene transcription, it is important to study the estrogen signaling pathway at the single-cell level to fully understand how ER $\alpha$  regulates transcription. Here, we performed single-cell transcriptome analysis on ER $\alpha$ -positive BCa cells following 17 $\beta$ -estradiol stimulation and reconstructed the dynamic estrogen-responsive transcriptional network from discrete time points into a pseudotemporal continuum. Notably, differentially expressed genes show an estrogen-stimulated metabolic switch that favors biosynthesis but reduces estrogen degradation. Moreover, folate-mediated one-carbon metabolism is reprogrammed through the mitochondrial folate pathway and polyamine and purine synthesis are upregulated coordinately. Finally, we show AZIN1 and PPAT are direct ER $\alpha$  targets that are essential for BCa cell survival and growth. In summary, our study highlights the dynamic transcriptional heterogeneity in ER $\alpha$ -positive BCa cells upon estrogen stimulation and uncovers a mechanism of estrogen-mediated metabolic switch.

## INTRODUCTION

Breast cancer is the most prevalent cancer in women worldwide, with more than two-thirds diagnosed as estrogen receptor  $\alpha$  (ER $\alpha$ )

positive (De Santis et al., 2014). Estrogen is a steroid hormone that plays pivotal functions in the pathophysiology of ER $\alpha$ -positive breast cancers (Cheung and Kraus, 2010). Specifically, estrogen binds to ER $\alpha$ , which subsequently is recruited to estrogen response elements (EREs) in the genome. ER $\alpha$  then recruits a host of cofactor proteins to either activate or repress the transcription of target genes that are essential for tumor proliferation and progression (Liu and Cheung, 2014).

Global time-series gene expression studies have revealed a list of ER $\alpha$ -regulated genes and the underlying mechanisms of estrogen-mediated cancer development (Frasor et al., 2003). However, the initiation of ER $\alpha$  binding to chromatin is stochastic (Métivier et al., 2003). Moreover, an individual cell's response to stimuli is further complicated by a variety of intrinsic and extrinsic factors such as chromatin status, mitochondrial content, relative location within a population, and cell-to-cell contact (Battich et al., 2015). Hence, the classical time-series gene expression analysis based on bulk cell population averages may be incomplete because of the multi-layered stochasticity in cellular responses to estrogen stimulation. Thus, more technologies to analyze transcriptomes at the single-cell level are required to account for these variables.

Recent advances in live-cell imaging have enabled researchers to monitor the dynamic expression of target genes in individual cells and confirmed that gene expression at the single-cell level is highly stochastic and binary (Harper et al., 2011). However, these techniques are limited by the number of reporters and detection channels and are unable to facilitate *de novo* discovery (Leng et al., 2015). The availability of automated microfluidic platforms that can uniformly capture and process hundreds of single cells for RNA-seq analysis has enabled profiling of single-cell transcriptomes at discrete time points (Shalek et al., 2013). Even so, gene expression cannot be tracked continuously from the same



single cell, and the stochasticity of response to stimulation will exist across the time series (Shalek et al., 2014). Nevertheless, by leveraging single cell-based computational algorithms, we can construct a pseudotemporal continuum through the reorganization of single-cell transcriptomes from discrete time points (Trapnell et al., 2012). In these trajectory analyses, the transcriptome from each single cell represents an instance or a “pseudo-time point” along an artificial time vector that denotes the progress along the stimulus response (Durruthy-Durruthy and Heller, 2015). Thus, by constructing a pseudotemporal continuum, we can obtain an enhanced temporal resolution compared with analyzing each discrete experimental time point.

In this work, we performed time course studies with 17 $\beta$ -estradiol (E2) treatment followed by single-cell RNA sequencing (scRNA-seq) analysis on MCF-7 and T47D ER $\alpha$ -positive human breast cancer cell lines. We applied Monocle single-cell trajectory analysis and ordered cells along an artificial pseudotemporal continuum to allow the characterization of the unsynchronized response of breast cancer cells to estrogen stimulation. With an increased temporal resolution, we uncovered a mechanism of an estrogen-mediated metabolic switch and identified ER $\alpha$ -regulated genes that are essential for these reprogrammed metabolic pathways.

## RESULTS

### scRNA-seq Analysis of Breast Cancer Cells in Response to Estrogen Stimulation

To characterize the extent of gene expression variability in individual breast cancer cells in response to estrogen stimulation, we performed scRNA-seq analysis on ER $\alpha$ -positive MCF-7 and T47D cells harvested at 0, 3, 6, and 12 hr after E2 stimulation (Figure 1A). We captured a total of 177 (85 MCF-7 and 92 T47D) single cells and sequenced each cell to an averaged depth of 10 million reads (Table S1). After filtering for unique mapping reads (Figure S1A) and 3' or 5' coverage biases (Figure S1B), we retained 84 MCF-7 and 78 T47D single-cell profiles (Table S1) for further downstream analysis. We then estimated the expression level of all the UCSC-annotated genes and removed genes that were not appreciably expressed (fragments per kilobase of transcript per million mapped reads [FPKM] > 1 in more than half of the single-cell profiles). In the end, 6,867 genes were used for our analysis.

Next, we performed Pearson correlation coefficient (PCC) analysis followed by unsupervised clustering to measure the linear correlation between all the single-cell samples and the population replicates from the two cell lines. As shown in Figure 1B, MCF-7 and T47D cells are clearly separated, indicating that cell line differences are greater than cell-to-cell variances. Although the gene expression level of the bulk population replicates closely correlated with one another (Pearson  $r > 0.94$ ; Figures 1B and 1C), substantial differences in expression between individual cells exist ( $0.65 < r < 0.8$ ; Figures 1B, 1D, and S1C). Despite this extensive cell-to-cell variation, the averaged expression level of the single cells correlated well with the bulk population replicates (Figures 1E and 1F).

To validate the inter-cellular heterogeneity observations from our scRNA-seq analysis, we performed droplet digital PCR

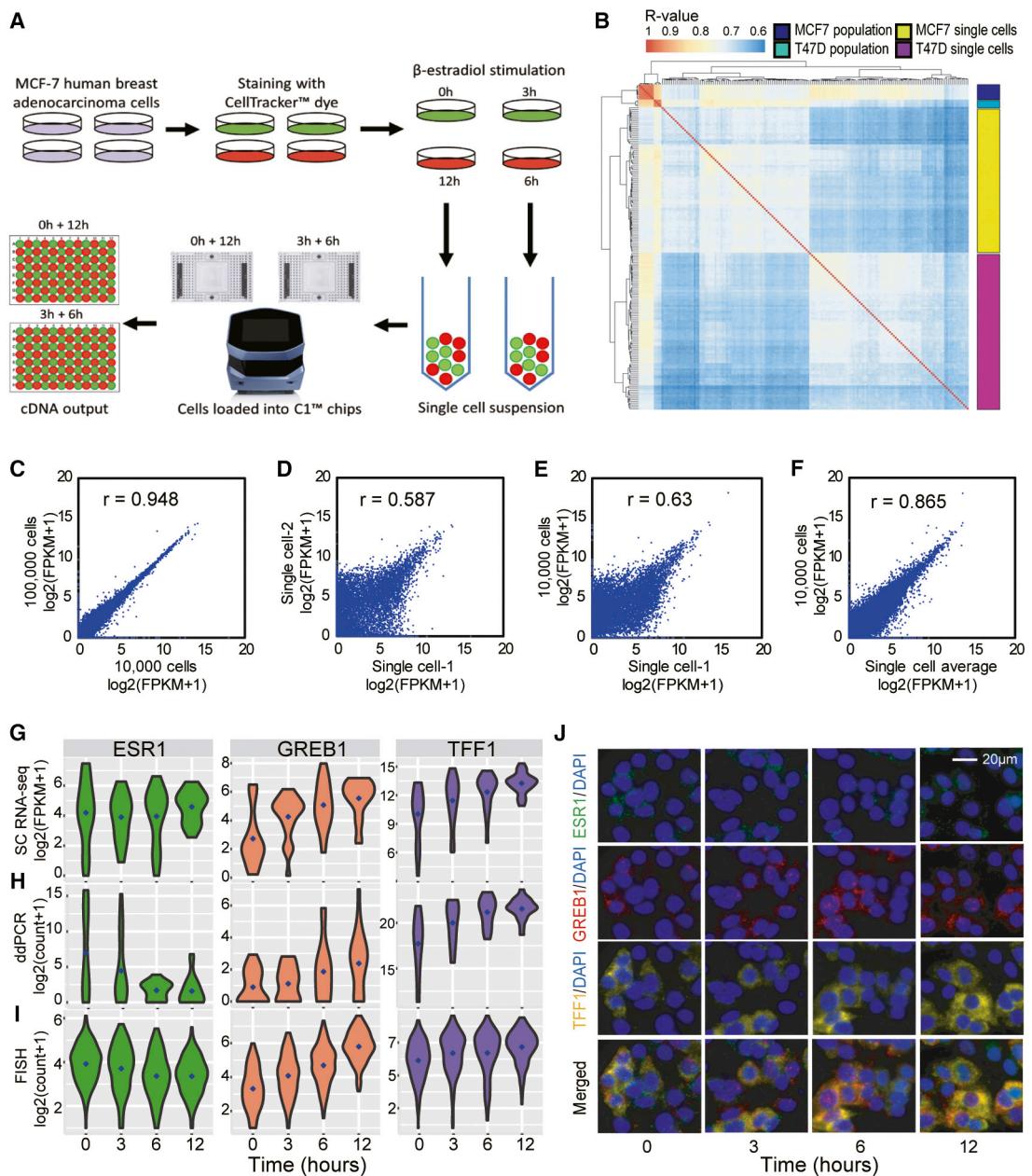
(ddPCR), a high-resolution approach for quantifying the absolute number of mRNA copies in a tiny amount of sample. We assayed for the expression level of TFF1, GREB1, and ESR1 (three well-known E2-responsive marker genes), as well as ACTB (a housekeeping gene) (Figure S2). Overall, our ddPCR results showed there is a large cell-to-cell variance for every gene, including the housekeeping gene (Figure S3). Moreover, the general expression profiles of the E2-responsive genes correlated well with the scRNA-seq results, with the exception of ESR1, which could be due to technical variability (Figures 1G, 1H, and S4A). Besides ddPCR, we also used single-molecule RNA fluorescence *in situ* hybridization (smRNA-FISH), an amplification-free imaging technique, to examine the expression of the above marker genes but in a greater number of cells. Similar to ddPCR, results from the smRNA-FISH analysis also mirrored the expression trends of the scRNA-seq data for both the MCF-7 (Figures 1I and 1J) and T47D (Figures S5A–S5C) cell lines. Taken together, our results showed that breast cancer cells displayed great transcriptional heterogeneity in response to estrogen stimulation.

### Bimodal Gene Expression Is Common across Single Cells and Produces Averaging Artifacts

The mRNA expression level of genes frequently operates in two modes: baseline and over- or under-expression, which is a major contributor of cell-to-cell variances (Shalek et al., 2013). Thus, to characterize the extent of bimodal gene expression in our breast cancer scRNA-seq data, we used a recently developed algorithm called SIBER (Tong et al., 2013). SIBER works by fitting the single-cell gene expression distribution into two log-normal distributions (components 1 and 2) and then calculating the mean values ( $\mu_1$  and  $\mu_2$ ) and the ratios ( $\pi_1$  and  $\pi_2$ ) for these two subpopulations (Figure 2A). Finally, a bimodal index (BI) value is calculated for each gene. Using SIBER, we calculated the BI value for every coding gene across all the single-cell profiles at each time point and determined any gene with a  $BI > 1.8$  as a bimodal gene (Figure 2B). As shown in Figure 2C, a large proportion (27%–41%) of the genes in both MCF-7 and T47D cells are bimodal in each time point. Moreover, the BI values of each gene changed across time points. For example, approximately 6%–8% of genes were constantly bimodal across the four time points, while >60% of the genes altered between bimodality and unimodality (Figure 2D).

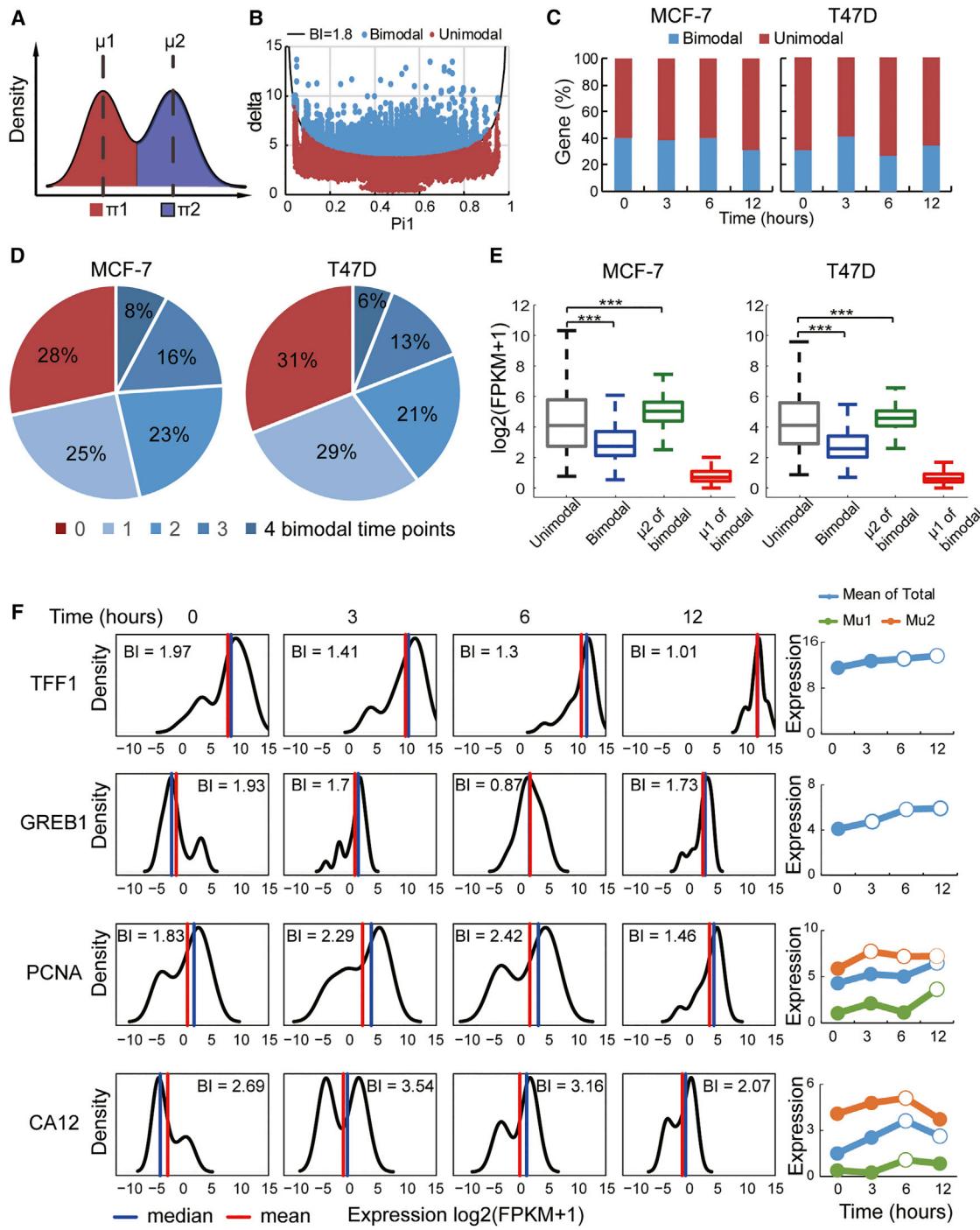
One of the major issues of a bimodal gene is that its averaged expression level value does not truly represent its two subpopulations. To address this, we compared the averaged expression level of bimodal and unimodal genes. As shown in Figure 2E, the average expression level of bimodal genes in general is significantly lower than the average expression level of unimodal genes. However, when we singled out component 2 (higher expressing subpopulation) of bimodal genes and compared it with unimodal genes, we found that the average expression level of component 2 ( $\mu_2$ ) is significantly higher than the average expression of the unimodal genes (Figure 2E). Thus, our findings suggest that highly expressed genes in a subpopulation of cells might be masked because of the averaging with the lower expressing subpopulation.

To explore this observation further, we examined the expression of two well-known E2-responsive genes (TFF1 and GREB1) with low bimodality and another two known E2-responsive genes



**Figure 1. Generation and Validation of scRNA-Seq Data from Breast Cancer Cells under Estrogen Stimulation**

- (A) Workflow depicting the rapid and high-throughput isolation and RNA sequencing (RNA-seq) of single-cells from breast cancer cell lines under time course E2 treatment.
- (B) Correlation between single-cell profiles and bulk population profiles.
- (C) Scatterplot of gene expression values ( $\log_2(\text{FPKM} + 1)$ ) between the 100k cell bulk profile and the 10k cell bulk profile ( $r = 0.948$ ).
- (D) Scatterplot between two random single-cell profiles with low correlation ( $r = 0.587$ ).
- (E) Scatterplot between the 10k cell bulk profile and a random single-cell profile showing low correlation ( $r = 0.63$ ).
- (F) Scatterplot between the 10k cell bulk profile and the average of all single-cell profile ( $r = 0.865$ ).
- (G) Violin plots showing the single-cell RNA sequencing (scRNA-seq) expression values ( $\log_2(\text{FPKM} + 1)$ ) of estrogen-responsive marker genes TFF1, GREB1, and ESR1 from the time course E2 treatment of MCF-7 cells.
- (H) Violin plots showing single-cell ddPCR values ( $\log_2(\text{count} + 1)$ ) of the marker genes from the same MCF-7 time course experiment.
- (I) Violin plots showing smRNA-FISH values ( $\log_2(\text{count} + 1)$ ) of the marker genes in a time course E2 treatment of MCF-7 cells. Two hundred cells were counted for each time point.
- (J) Multiplexed smRNA-FISH images showing the mRNA copies of the marker genes in MCF-7 cells under time course E2 treatment with nuclei counter stained by DAPI.

**Figure 2. Bimodal Variation of Gene Expression Levels across Single Cells Causes Averaging Artifacts**

(A) The bimodality index (BI) of each gene across single cells at every time point was calculated using the SIBER method, which is based on the ratio and distance between two components.  $\pi_1$  is the ratio of the lower component,  $\pi_2$  is the ratio of the higher component,  $\mu_1$  is the mean of the lower component, and  $\mu_2$  is the mean of the higher component.

(B) Genes with  $BI > 1.8$  (blue dots above the black curve) are defined as bimodal expression.

(C) Percentage of genes with bimodal expression at every time point in the MCF-7 and T47D profiles.

(D) Percentage of genes showing bimodal gene expression across zero, one, two, three, and four time points among all genes in the MCF-7 and T47D profiles. Genes without any bimodal time point (red sector) are defined as unimodal genes; genes with at least one bimodal time point are defined as bimodal genes.

(legend continued on next page)

(PCNA and CA12) (Yamaga et al., 2013) with high bimodality (Figure 2F). We applied the Mann-Whitney U test to examine whether their gene expression levels at 3, 6, and 12 hr after E2 stimulation changed significantly compared with that at 0 hr. As expected, the expression levels of both TFF1 and GREB1 changed significantly along the time course. For PCNA and CA12, we examined the expression levels of all the cells in both components (Mu1 and Mu2). Intriguingly, Mu2 for PCNA changed significantly at 3 and 6 hr, but because of the averaging with Mu1 (which was not significantly changed), the mean expression level of all cells was not altered significantly. This example illustrates how a typical upregulated gene in a subpopulation is masked in bulk. More interestingly, for CA12, even though both Mu1 and Mu2 were not significantly altered at 12 hr, the mean expression for all the cells was significantly changed. In this case, this is an example of Simpson's paradox, a phenomenon in which the average of two groups displays an inverse trend compared with the two individual groups (Simpson, 1951). Similar averaging artifacts were also observed for known E2-responsive genes (PGR and CCND1) (Yamaga et al., 2013) with high bimodality in T47D cells (Figure S5D). Taken together, our findings show that bimodal gene expression introduces different types of averaging artifacts and greatly affects the identification of DEGs.

### Pseudotemporal Analysis Identifies More Differentially Expressed Genes with High Bimodality

To resolve the averaging artifacts introduced by the bimodal expression of genes and at the same time obtain a better temporal resolution of the E2-stimulated transcriptional dynamics, we applied a recently developed computational method, Monocle (Trapnell et al., 2012), to reorder all the MCF-7 single-cell profiles along a pseudotemporal continuum (Figure 3A). We first confirmed from cell cycle analysis (Buettnner et al., 2015) that this ordering was not dominated by cell cycle phase differences (Figure S4B). Next, on the basis of the pseudotemporal continuum profile, we identified 1,102 differentially expressed genes (DEGs) (Figure 3B). In comparison, 1,011 DEGs were obtained on the basis of real time points. We compared the DEGs from the two approaches and found 552 common genes (Figure 3B). We also found more bimodal DEGs based on pseudotime (813 genes) compared with real time (660 genes) (Figure 3C). Similar findings were also observed for T47D cells (Figures S6A–S6D). Among the pseudotime DEGs were well-known E2-responsive genes including TFF1, GREB1, PCNA, and CA12 (Figure 3D). Moreover, other E2-responsive genes such as CCNB2 and TOP2A (Walker et al., 2007) were among the top bimodal DEGs (Figure S6E) that were not identified as DEGs on the basis of real-time analysis, which highlights the power of the pseudotemporal method in solving the averaging artifact problem.

Next, we performed clustering analysis on the pseudotime DEGs and identified three distinct groups: cluster 1 represents

late upregulated genes, cluster 2 represents early upregulated genes, and cluster 3 represents repressed genes (Figure 3E). To characterize how bimodal DEGs change across the pseudotemporal continuum, we selected the constantly bimodal genes (63 in cluster 1 and 23 in cluster 2) for further analysis. Cluster 3 was not included, as it has only four constantly bimodal genes. Surprisingly, we observed that Pi2 values but not Mu2 values significantly changed at 12 hr (Figures 3F and 3G). This finding suggests that most upregulated bimodal DEGs are “switched on” to a similar expression level in more cells rather than elevated to a higher expression level in all the cells.

### Estrogen Stimulates Metabolic Reprogramming that Favors Biosynthesis and Cell Proliferation but Reduces Degradation of Estrogen

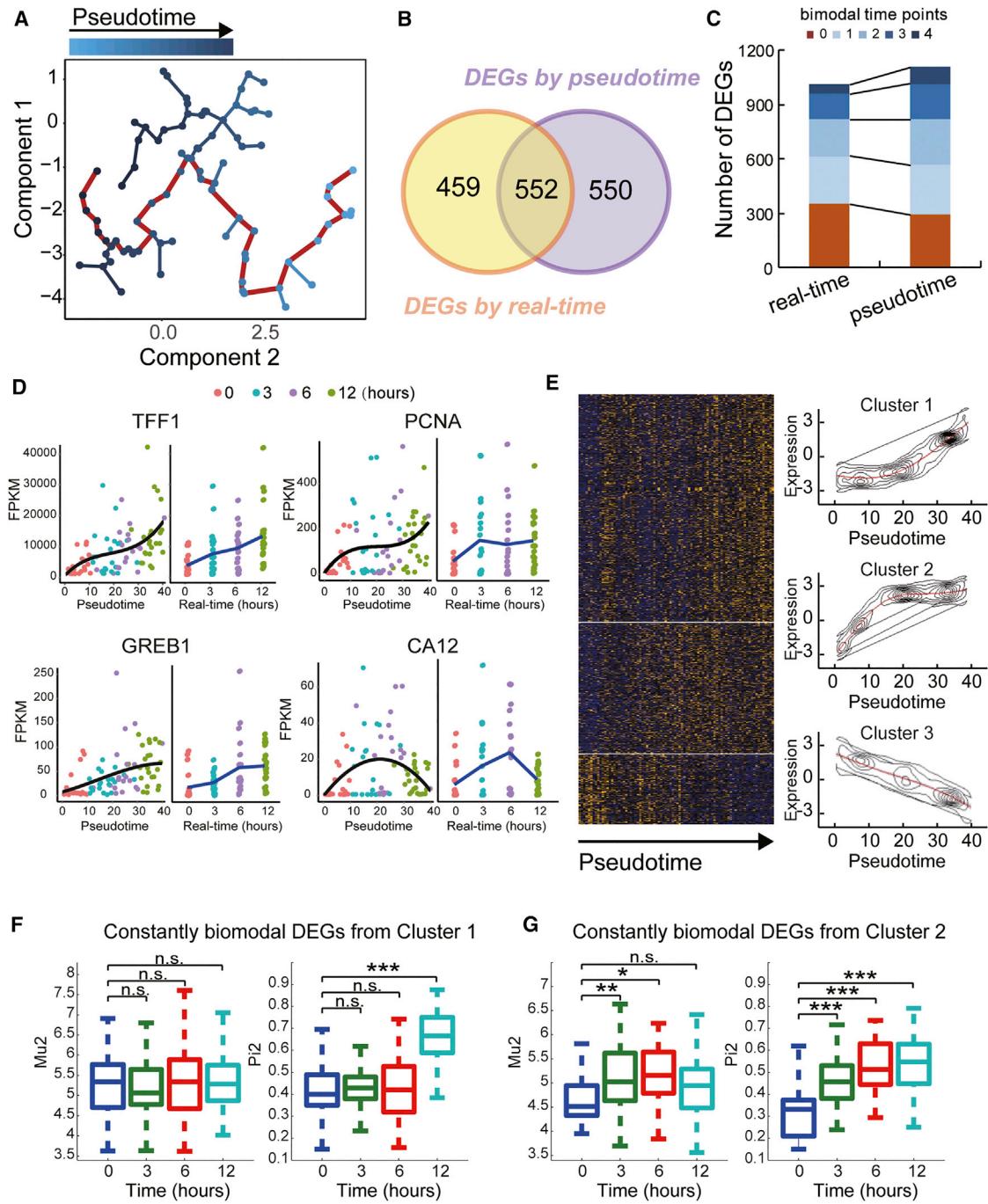
To gain functional insights into the DEGs identified by pseudotemporal analysis, we performed pathway enrichment analysis. As shown in Figure 4A, upregulated DEGs are enriched in pathways that favor cell proliferation, including cell cycle, DNA repair, and anabolic pathways. In contrast, downregulated DEGs are enriched in the xenobiotic and glutathione metabolism pathways (Figure 4A). These two pathways are associated with the degradation of estrogen-derived byproducts and reactive oxygen species (ROS) clearance (Santen et al., 2015). Thus, the downregulation of these two pathways could potentially lead to an accumulation of genotoxic estrogen metabolites and an elevation of ROS, which are known to promote tumorigenesis (Yager, 2015). Taken together, our findings indicate that E2 stimulation induces a metabolic switch that favors biosynthesis and cell proliferation but impairs estrogen catabolism and ROS clearance.

Among the above enriched pathways, glycolysis, TCA cycle, aminoacyl-tRNA biosynthesis, one-carbon metabolism, and folate biosynthesis pathways were identified only by pseudotime analysis but not by real-time analysis. By examining the expression trend of these pathways on the basis of real time points (Figure 4B) and pseudotime (Figure 4C), we found that in general the pseudotime analysis showed much greater alterations in these pathways. In addition, we found genes in the one-carbon metabolism and folate biosynthesis pathways having very high BI values; therefore, these two pathways were missed by real-time analysis, possibly because of the averaging artifacts of gene bimodality (Figure 4D). To the best of our knowledge, the regulation of these two pathways by estrogen in breast cancer cells has not been reported in previous time-series studies of bulk MCF-7 cells using either gene expression microarray (Fraser et al., 2003), RNA-seq (Yamaga et al., 2013), or proteomics (Drabovich et al., 2012) technologies.

To validate the above findings on a larger number of cells, we repeated the MCF-7 time course E2 stimulation experiment at 0 and 12 hr and profiled >500 single cells for each time point using C1 high-throughput (HT) integrated fluidic circuit (IFC) chips.

(E) Average level ( $\log_2[\text{FPKM} + 1]$ ) of unimodal genes, bimodal genes, lower (mu1) and higher (mu2) components of bimodal genes across all MCF-7 and T47D single-cell profiles.

(F) Density plots showing the expression distribution of known estrogen-responsive genes TFF1, GREB1, PCNA, and CA12 across single cells in every time point. Line charts showing the average expression level ( $\log_2[\text{FPKM} + 1]$ ) in all single-cell profiles across time points; average level of lower (Mu1) and higher (Mu2) components for PCNA and CA12 are also shown. Open circle indicates significant alteration ( $p < 0.05$  by Mann-Whitney U test) compared with 0 hr; closed circle indicates no significant alteration.



**Figure 3. Pseudotime Ordering of Estrogen-Stimulated Single-Cell Profiles by Monocle**

(A) Single-cell profile (points) of MCF-7 cells in a two-dimensional independent component space. Lines connecting points represent edges of the minimal spanning tree (MST) constructed by Monocle. Solid red line indicates the main diameter of MST and provides the backbone of Monocle's pseudotime ordering of the cells.

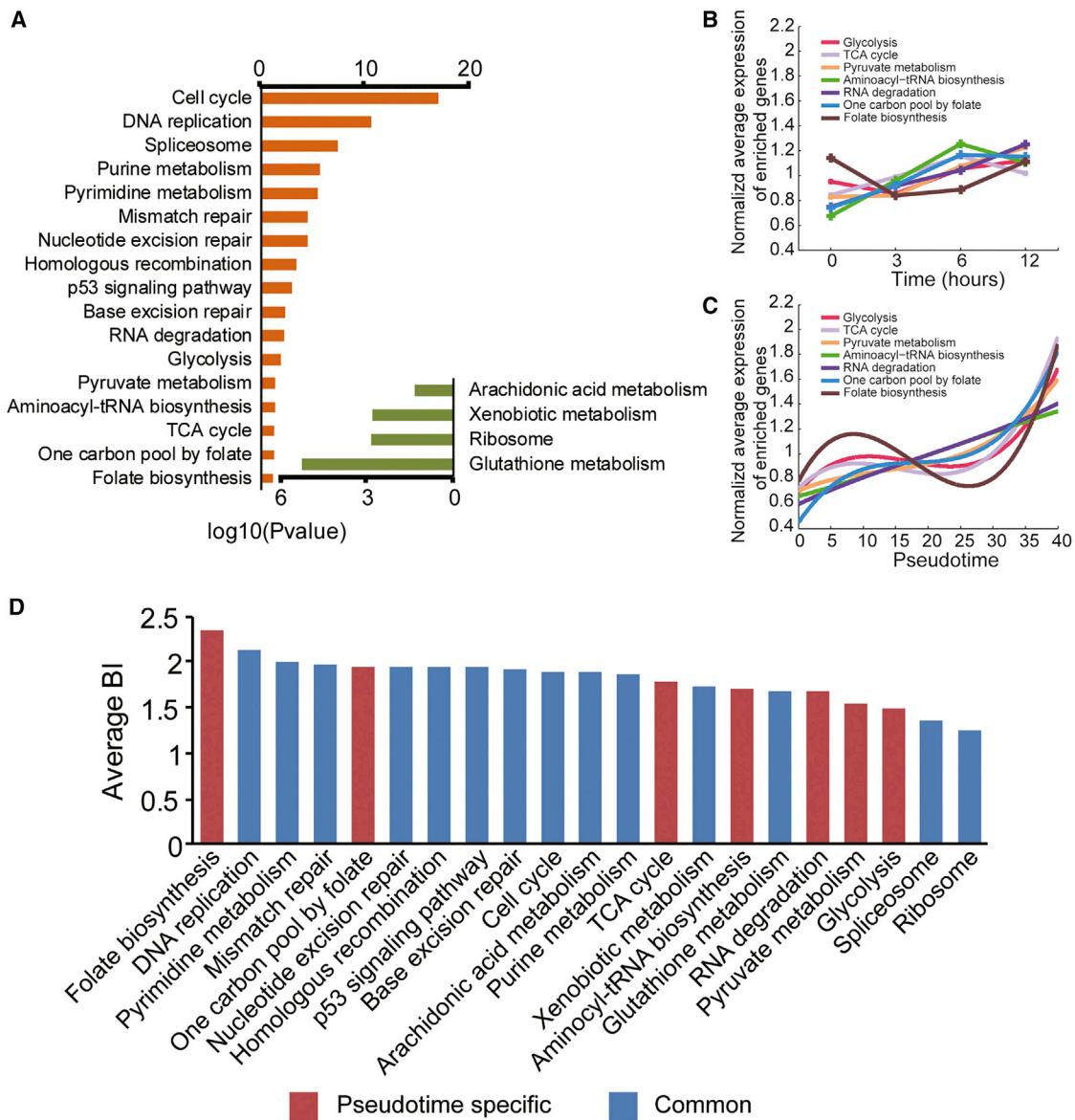
(B) Number of differentially expressed genes (DEGs) identified on the basis of real-time analysis and pseudotime analysis.

(C) Number of DEGs showing bimodal gene expression at zero, one, two, three, and four original time points identified by Monocle on the basis of real time and pseudotime.

(D) Expression of marker genes TFF1, GREB1, PCNA, and CA12 ordered by real time and pseudotime.

(E) Expression profiles for DEGs identified by Monocle on the basis of pseudotime order. The DEGs were grouped into three clusters by k-means clustering.

(F) Alterations of the component ratios and the average expression levels of constantly bimodal DEGs (bimodal at all four time points) in clusters 1 and 2. Pi2 is the ratio of the higher expressing subpopulation; Mu2 is the average expression level of the higher expressing subpopulation.

**Figure 4. Estrogen-Stimulated Metabolic Reprogramming in MCF-7 Cells**

(A) KEGG pathways significantly enriched among the DEGs on the basis of pseudotime.

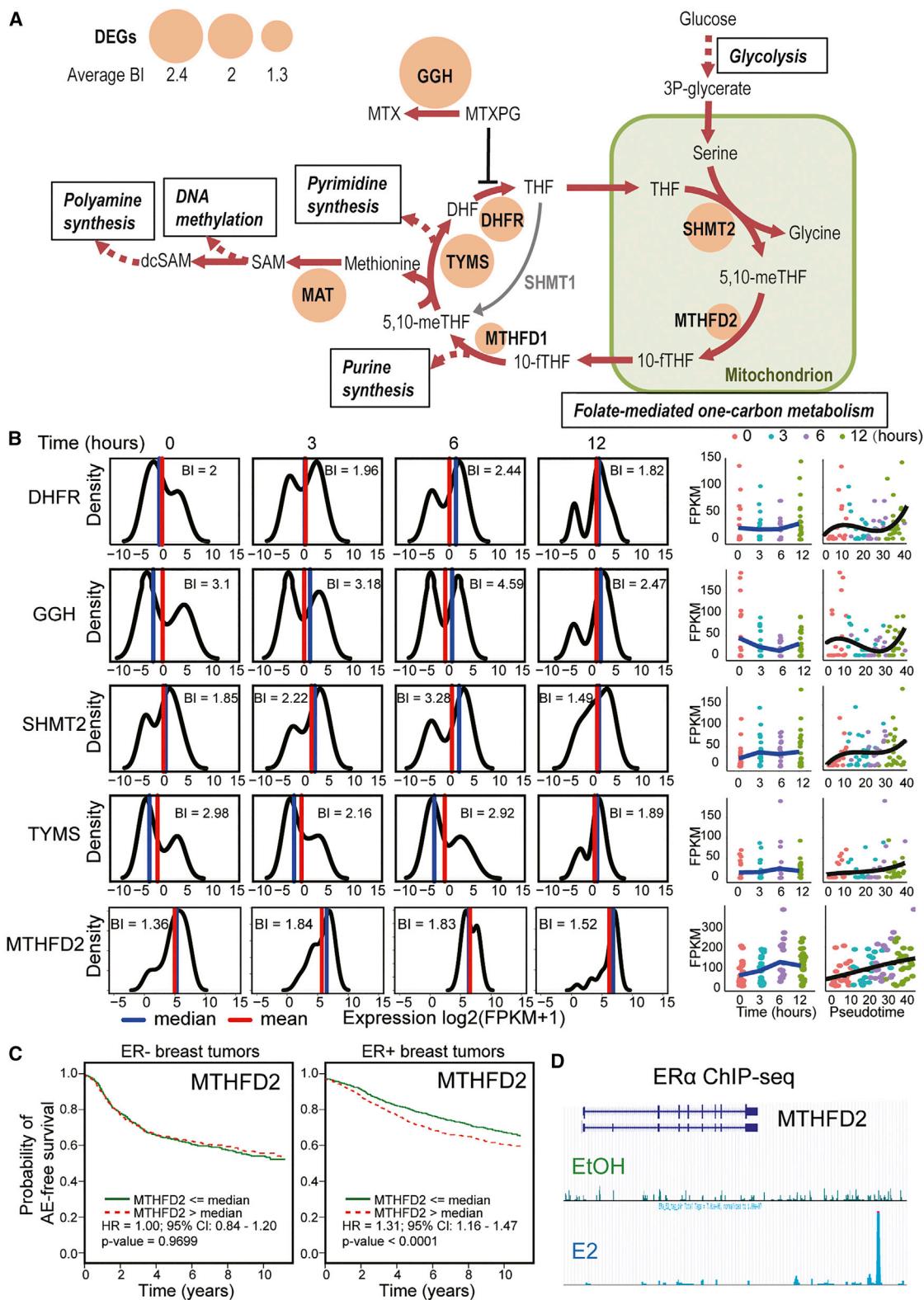
(B) Expression trend of upregulated pathways on the basis of real time.

(C) Expression trend of upregulated pathways on the basis of pseudotime. Expression level of a pathway was calculated by averaging the normalized levels of DEGs within the pathway.

(D) Average BI values of the upregulated pathways. Average BI value of a pathway is calculated by averaging the BI values of all DEGs at all time points in the pathway. Pathways identified by pseudotime only are indicated in red.

The HT scRNA-seq was able to profile much more MCF-7 single-cell transcriptomes but yielded lower unique mapping reads in each cell (Figure S7A). Despite the lower reads per cells, the HT MCF-7 single-cell profiles still displayed cell-to-cell variation comparable with our previous single-cell profiles (Figures S7B–S7E). We repeated pseudotemporal analysis on this larger single-cell dataset (Figure S7F) and obtained similar results as the smaller dataset in that the DEGs could also be grouped into three clusters: early responsive, late responsive, and downregulated

genes (Figure S7G). In addition, the expression of well-known E2-regulated marker genes was consistent with our previous observation (Figure S7H). In gene set enrichment analysis (GSEA), E2 up- and downregulated genes identified from the smaller single-cell dataset were significantly up- and downregulated in the larger dataset as well (Figures S8A and S8B). Finally, GSEA of KEGG pathways showed that most of the upregulated pathways identified from the smaller single-cell dataset were also enriched in the larger dataset, including one-carbon pool



(legend on next page)

by folate and purine metabolism (Figure S8C). Taken together, our results show that our single-cell analysis is reproducible and consistent with both small and large numbers of single cells.

### Estrogen Stimulates Reprogramming of Folate-Mediated One-Carbon Metabolism through the Mitochondrial Folate Pathway

Because our analysis revealed folate and one-carbon metabolism are estrogen-responsive pathways, we decided to investigate these two pathways in further detail. Folate metabolism is a pivotal process for one-carbon metabolism and thus they are often linked together and called folate-mediated one-carbon metabolism (Ducker and Rabinowitz, 2017). Folate metabolism can fuel one-carbon metabolism via either the mitochondrial pathway or the cytosolic pathway (Ducker et al., 2016). As shown in Figure 5A, E2 upregulated the expression of key enzymes in the mitochondrial pathway, including SHMT2, MTHFD2, and MTHFD1, but not SHMT1, a key enzyme in the cytosolic pathway (data not shown). Hence, our results suggest folate-mediated one-carbon metabolism is upregulated preferentially via the mitochondrial folate pathway in response to E2 stimulation. E2 also stimulated other important enzymes that are associated with this pathway, such as DHFR, GGH, TYMS, and MAT2B (Figure 5A), which likely contributes to the upregulation of nucleotide synthesis, DNA repair, DNA methylation, and polyamine synthesis pathways. Interestingly, most of the above genes exhibited a bimodal expression patterns (Figures 5B and S9A, left). Moreover, the expression trend based on pseudotime changed more significantly than those based on real time points (Figures 5B and S9A, right). Taken together, our observations for the folate-mediated one-carbon metabolism pathway are consistent with our earlier conclusion that the increase in expression of bimodal genes is due to the ratio of higher expressing subpopulation and not by the average expression level in the total cells.

Next, we asked whether any of these genes play an important role in ER $\alpha$ -positive breast tumor progression. For this, we performed prognostic analysis on the key enzymes in the folate-mediated one-carbon metabolism pathway using bc-GenExMiner 4.0 (Jézéquel et al., 2012), a webtool based on a collection of 36 breast cancer cohort studies. Our results showed that these genes could potentially serve as good prognostic markers for ER $\alpha$ -positive but not ER $\alpha$ -negative breast cancer patients (Figures 5C and S10). Moreover, we asked whether any of these genes are direct targets of ER $\alpha$ . Using our previous ER $\alpha$  chromatin immunoprecipitation sequencing (ChIP-seq) data in MCF-7 (Joseph et al., 2010), we found an ER $\alpha$  binding site located 12 kb upstream of the transcription start site (TSS) of MTHFD2, suggesting that this gene is likely a direct target of ER $\alpha$  (Figure 5D). MTHFD2 and the mitochon-

drial folate pathway have been shown to be important for various solid tumors (Nilsson et al., 2014), but its dependence on estrogen regulation in ER $\alpha$ -positive breast tumors has not been reported yet. Taken together, our pseudotemporal analysis revealed that estrogen reprograms one-carbon metabolism through the mitochondrial folate pathway.

### Estrogen Stimulation Augments the Mitochondrial Folate Pathway to Coordinate Fuel Polyamine and De Novo Purine Synthesis

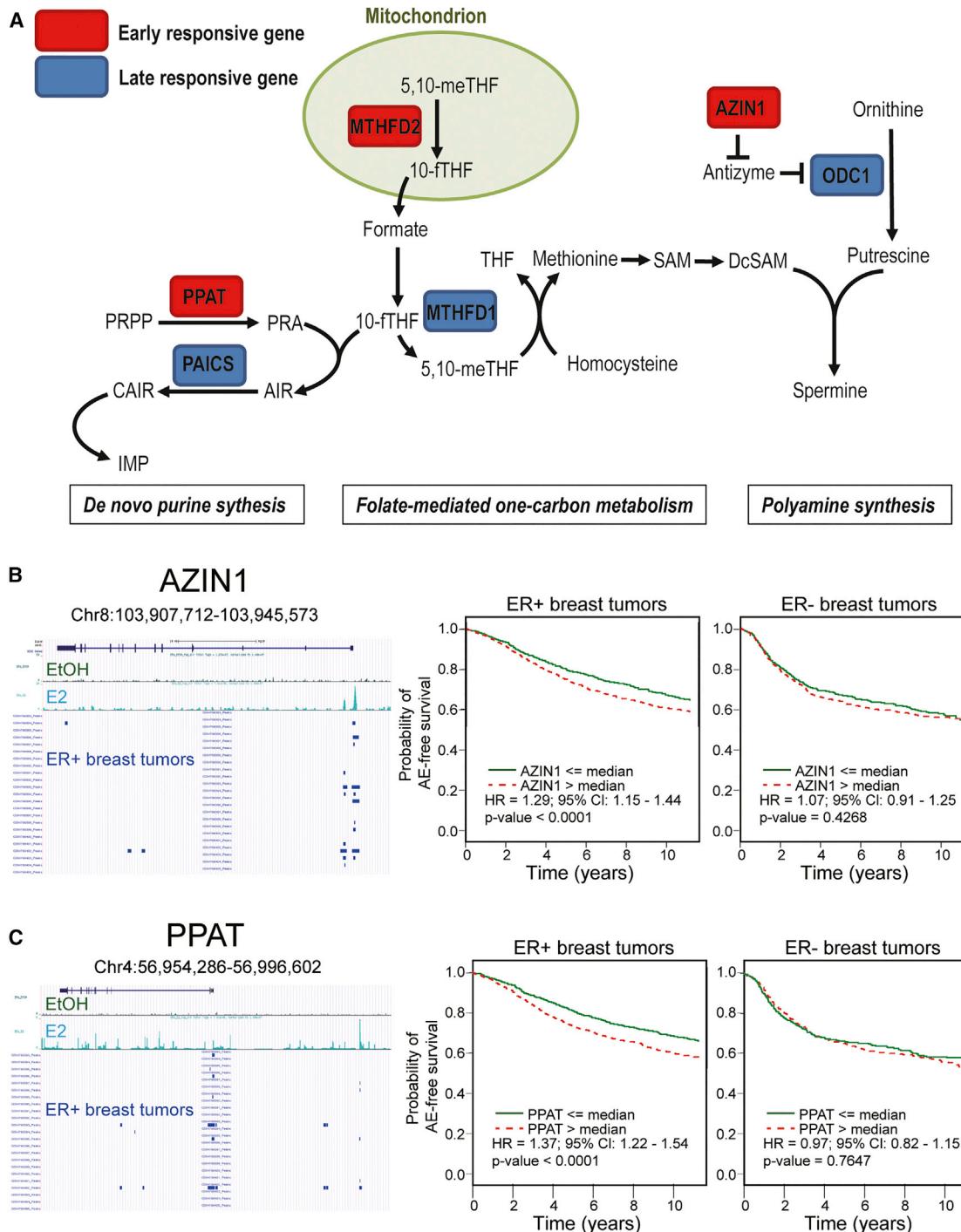
When we examined the metabolic pathways associated with one-carbon metabolism in further detail, we found both the polyamine and *de novo* purine synthesis pathways are also coordinately regulated by estrogen stimulation (Figure 6A). Specifically, the upstream enzymes and regulators of these pathways, including MTHFD2, PPAT, and AZIN1, are early E2 upregulated genes (Figures 6A and S9B). In contrast, several of the enzymes in the downstream part of the pathways, such as MTHFD1, PAICS, and ODC1, are late E2 upregulated genes (Figures 6A, S9A, and S9B). The early responsive genes AZIN1 and PPAT are also bimodal DEGs discovered by pseudotemporal analysis but not by real-time point analysis (Figure S9B). Integrative analysis with ER $\alpha$  ChIP-seq data (Ross-Innes et al., 2012) showed that both genes are associated with ER $\alpha$  binding sites in MCF-7 cells and ER $\alpha$ -positive breast tumors (Figures 6B and 6C, left). In survival analysis, we also found these genes are significant prognostic markers for ER $\alpha$ -positive but not ER $\alpha$ -negative breast cancer patients (Figures 6B and 6C, right). Taken together, these findings suggest that AZIN1 and PPAT are important ER $\alpha$  targets in ER $\alpha$ -positive breast tumors.

To examine whether AZIN1 and PPAT have important functional roles in ER $\alpha$ -positive breast cancer, we conducted dsRNA knockdown assays for AZIN1 and PPAT in MCF-7 cells. MTT cell proliferation assays showed that knockdown of both genes dramatically reduced MCF-7 cell proliferation (Figures 7A and 7B). In addition, we performed sub-G1 assay and observed a significant increase in cell death in MCF-7 cells after gene knockdown (Figures 7C and 7D). Finally, to determine the reason for the cell death, we performed annexin V/propidium iodide assay and showed that gene knockdown induced strong cell apoptosis in MCF-7 cells (Figures 7E and 7F). Therefore, AZIN1 and PPAT are essential genes for ER $\alpha$ -positive breast cancer cell proliferation and survival.

In conclusion, we have revealed a metabolic switch process orchestrated by E2 regulation in breast cancer cells. Upon E2 stimulation, ER $\alpha$  directly upregulates expression of MTHFD2, AZIN1, and PPAT to coordinate fuel polyamine and purine synthesis via the mitochondrial folate pathway.

**Figure 5. Estrogen Stimulates Reprogramming of One-Carbon Metabolism through the Mitochondrial Folate Pathway**

- (A) Schematic presentation showing significantly altered key regulators in the folate-mediated one-carbon metabolism pathway and their average BI values. 5,10-meTHF, 5,10-methylenetetrahydrofolate; 10-fTHF, 10-formyltetrahydrofolate; dcSAM, decarboxylated S-adenosylmethionine; DHF, dihydrofolate; MTX, methotrexate; MTXPG, methotrexate polyglutamates; SAM, S-adenosylmethionine; THF, tetrahydrofolate.
- (B) DEGs with high BI values in the folate biosynthesis and one-carbon metabolism pathways. Density plots show the expression level distribution in every time point. Lines charts show expression trend of the genes on the basis of real time and pseudotime.
- (C) Kaplan-Meier survival analysis of MTHFD2 for ER $\alpha$ -positive and negative breast cancer patients.
- (D) ER $\alpha$  ChIP-seq data in MCF-7 cells shows binding sites near the MTHFD2 gene body.

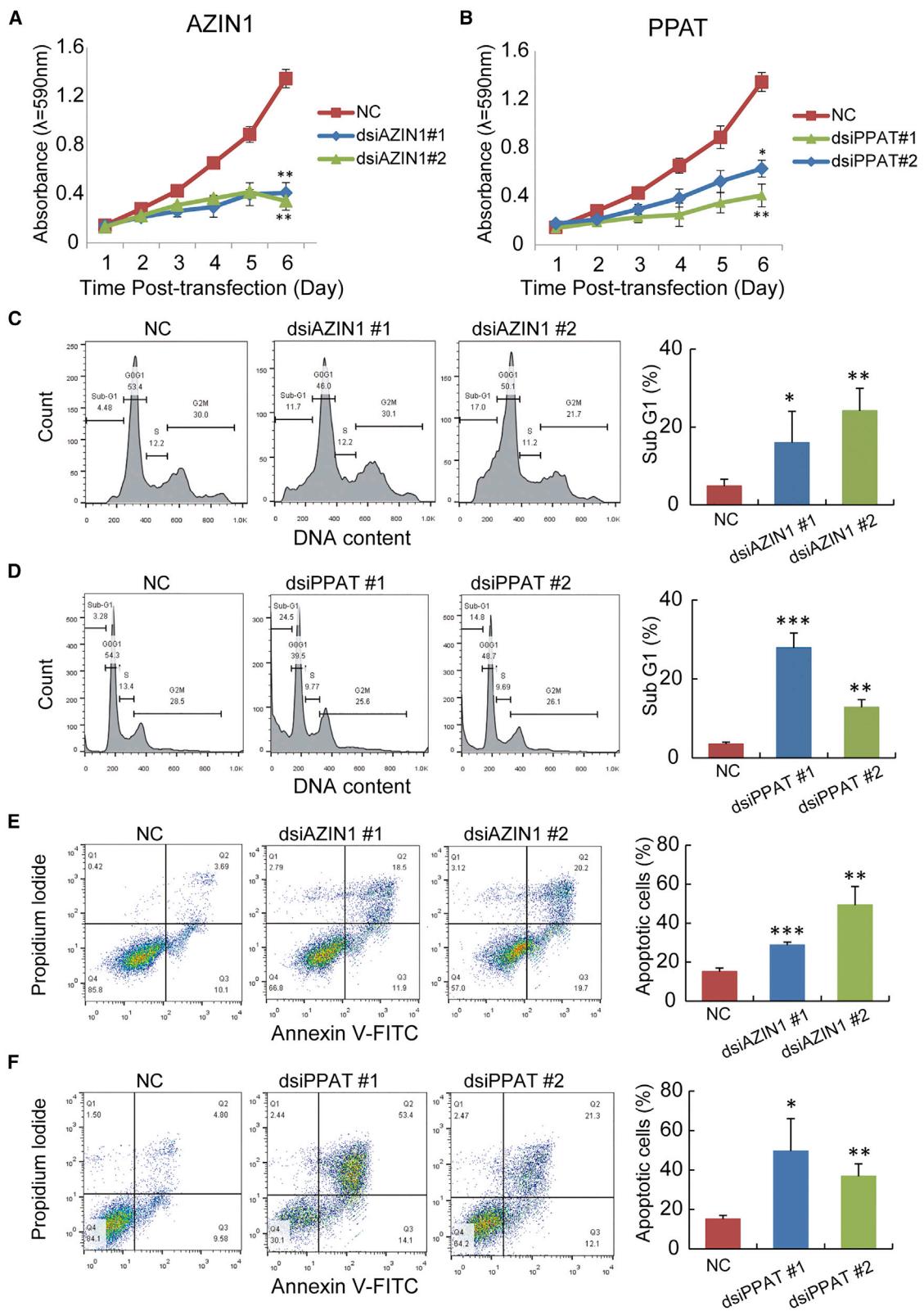


**Figure 6. Estrogen Stimulation Coordinates Upregulates One-Carbon Metabolism, De Novo Purine Synthesis, and Polyamine Synthesis**

(A) Schematic presentation showing estrogen stimulation coordinates upregulates one-carbon metabolism, *de novo* purine synthesis, and polyamine synthesis. AIR, aminoimidazole ribotide; CAIR, carboxyaminoimidazole ribotide; PRA, phosphoribosylamine; PRPP, phosphoribosyl pyrophosphate.

(B) ChIP-seq data showing ER $\alpha$  binding sites near the AZIN1 gene in MCF-7 cells and in clinical tumor samples. Kaplan-Meier survival analysis of AZIN1 for ER $\alpha$ -positive and negative breast cancer patients.

(C) ChIP-seq data showing ER $\alpha$  binding sites near PPAT gene bodies in MCF-7 cells and in clinical tumor samples. Kaplan-Meier survival analysis of PPAT for ER $\alpha$ -positive and negative breast cancer patients.



(legend on next page)

## DISCUSSION

In this study, we performed scRNA-seq analysis to profile the transcriptional landscape of E2 time course treatment on ER $\alpha$ -positive breast cancer cell lines. We used the Fluidigm C1 platform to facilitate HT automated single-cell capture and cDNA preparation of more than 1,000 single-cell libraries from MCF-7 and T47D cells. With these data, we characterized the cellular transcriptional response to estrogen stimulation at the single-cell level.

As expected, we observed a greater variation in gene expression between individual cell transcriptomes compared with bulk population of cells. However, the inter-cellular correlation of our datasets ( $0.65 < r < 0.80$ ) was higher than previously published datasets derived from *in vivo* cells ( $0.29 < r < 0.62$ ) (Shalek et al., 2013). We believe this difference could be due in part to the more uniformed *in vitro* cell culture environment in our experiment. As a validation of our scRNA-seq results, we used ddPCR and smRNA-FISH, two independent techniques that are capable of accurate absolute quantification. Our results suggest the cell-to-cell heterogeneity that we observed in our scRNA-seq is real and not due to technical noise because of sequencing minute amounts of material.

A recent single-cell study examining the response of lipopolysaccharide (LPS) stimulation on mouse bone marrow-derived dendritic cells showed that most LPS-responsive genes are highly variable and that more than 75% of the highly variable genes exhibit a bimodal expression pattern (Shalek et al., 2013). These findings are consistent with our observations in breast cancer cells, in which we also identified a large proportion of E2-responsive genes with bimodal expression. Specifically, we found that more than two-thirds of the genes showed bimodal expression in at least one time point. The gene expression bimodality is possibly caused by confounding effects of multiple factors, including both biological factors, such as the unsynchronized response of individual cells to external stimuli (Métivier et al., 2003) and the innate bistability of the estrogen signaling pathway (Battich et al., 2015), as well as technical issues such as stochastic dropout events and amplification bias (Wagner et al., 2016). More importantly, the gene expression bimodality in many instances produced averaging artifacts. First, it masked the higher expression subpopulation by averaging it with the lower expression subpopulation. Second, the expression trend of the subpopulation was masked. In some extreme examples, Simpson's paradox happened; that is, the expression trend of the total cells and the two subpopulations were inverse to each other. From our work, we not only observed the gene expression bimodality but also explained some of the possible

artifacts that caused by it when using conventional methods for DEG analysis.

To circumvent this problem, we used the Monocle algorithm, which orders single-cell profiles along an artificial pseudotemporal continuum (Trapnell et al., 2012). Each single-cell profile was considered as a pseudo-time point on the continuum and thus substantially enhanced the temporal resolution. The DEGs that were identified on the basis of real time points and pseudotime were quite different (Figure 3B). In particular, significantly more bimodal genes were identified by pseudotime compared with real time (Figure 3C). Moreover, analysis of the constantly bimodal DEGs showed that these highly bimodal genes were altered significantly by the ratios of subpopulations instead of the averaged expression level of the subpopulations (Figures 3F and 3G). These results highlight a dynamic model of transcriptional regulation by nuclear receptors in response to hormone stimulation. Unlike unimodal genes that change in expression levels, the activated levels of these bimodal genes are tightly restricted at a narrow range in individual cells, but the frequencies of "switch-on" increase, showing a higher ratio of the switched-on subpopulation. Monocle successfully identified more highly bimodal responsive genes, probably because of the higher temporal resolution by a continuous arrangement of single-cell profiles instead of discrete grouping by a few experimental time points. Therefore, we have demonstrated that the single-cell trajectory analysis method is a good solution for dissecting the averaging artifacts produced by gene expression bimodality.

From our single-cell analysis, we identified folate synthesis and one-carbon metabolism as estrogen-regulated metabolic pathways that are enriched for bimodal genes (Figure 5B). Folate metabolism can fuel one-carbon metabolism via either the mitochondrial folate pathway or the cytosolic folate pathway (Ducker et al., 2016). Interestingly, we found that estrogen stimulation did not upregulate the cytosolic folate pathway but instead selectively activated the mitochondrial folate pathway (Figure 5A). Moreover, we identified MTHFD2 as an ER $\alpha$  target and an early responsive regulator for the reprogramming. MTHFD2 and the mitochondrial folate pathway have recently been highlighted as markers for metabolic reprogramming in a wide spectrum of solid tumors (Nilsson et al., 2014). Thus, the link between ER $\alpha$  regulation and MTHFD2 overexpression in our finding underscores an important event for the establishment of metabolic reprogramming in ER $\alpha$ -positive breast tumors. On the other hand, glycolysis and TCA cycle pathways, which are known estrogen-regulated pathways in breast cancer (Forbes et al., 2006), were also identified by pseudotemporal analysis but not by real-time point analysis. These pathways are enriched for unimodal genes; hence, increased temporal resolution not only helped identify

**Figure 7. AZIN1 and PPAT Are Essential for Breast Cancer Cell Proliferation and Survival**

- (A) Knockdown of AZIN1 with dsRNAs reduces MCF-7 cell proliferation. Experiments are performed in triplicate.
- (B) Knockdown of PPAT with dsRNAs reduces MCF-7 cell proliferation. Experiments are performed in triplicate.
- (C) Knockdown of AZIN1 with dsRNAs increases MCF-7 sub-G1 cell fraction.
- (D) Knockdown of PPAT with dsRNAs increases MCF-7 sub-G1 cell fraction.
- (E) Knockdown of AZIN1 with dsRNAs increases MCF-7 cell apoptosis.
- (F) Knockdown of PPAT with dsRNAs increases MCF-7 cell apoptosis.

All experiments are repeated three independent times. Error bars represent SD. \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ .

highly bimodal responsive genes but also aided in the identification of unimodal responsive genes with insignificant changes between discrete time points.

Our single-cell analysis also showed the polyamine and *de novo* purine synthesis pathways are coordinately upregulated by estrogen with the mitochondrial folate pathway. In these pathways, we identified the bimodal genes, AZIN1 and PPAT, as early E2-responsive genes and ER $\alpha$  targets. AZIN1 is an inhibitor of antizyme that induces degradation of ODC1 and interferes with the uptake of external polyamine (Kahana, 2009), while PPAT is the first-step rate-limiting enzyme for *de novo* purine synthesis (Pedley and Benkovic, 2017). Other important enzymes downstream of AZIN1 and PPAT, such as ODC1 (Kahana, 2009) and PAICS (Pedley and Benkovic, 2017) are among the late E2-responsive genes. Hence, the transcriptional dynamics illustrated by our pseudotemporal analysis is consistent with the flow of the metabolic pathways (Figure 6A). Finally, we were able to validate both AZIN1 and PPAT are essential for MCF-7 proliferation and migration (Figure 7). Although the mitochondrial folate pathway is adopted by various tumors (Nilsson et al., 2014), recent studies showed that it can be compensated by the cytosolic folate pathway if it is disrupted (Ducker et al., 2016). Therefore, targeting the polyamine and *de novo* purine synthesis pathways may be a better choice than targeting the folate mitochondrial pathway.

In summary, the dynamic heterogeneity and metabolic switch of cellular response to E2 stimulation revealed from our work may serve as a potential strategy for cells to eventually acquire resistance to hormone therapy. Thus, we believe that further single-cell transcriptome analysis of additional breast cancer cell lines in long-term anti-estrogen and aromatase inhibition treatments will produce invaluable insight for understanding how therapy-resistant cells emerge from a given population of cancer cells.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
- METHOD DETAILS
  - E2 stimulation and scRNA-seq library generation
  - scRNA-seq data processing
  - Gene expression bimodality analysis
  - Mann-Whitney rank sum test
  - Pseudotemporal analysis
  - Single-cell droplet digital PCR
  - Single-molecule RNA fluorescent *in situ* hybridization
  - Functional enrichment analysis and public datasets used in this study
  - Cell proliferation assays
  - Cell apoptosis assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources

## SUPPLEMENTAL INFORMATION

Supplemental Information includes ten figures and three tables and can be found with this article online at <https://doi.org/10.1016/j.celrep.2018.10.093>.

## ACKNOWLEDGMENTS

This work was supported by the Macau Science and Technology Development Fund (FDCT102/2015/A3 and FDCT/023/2014/A1), a University of Macau Multi-Year Research Grant (MYRG2015-00196-FHS), and a University of Macau Start-up Research Grant (SRG2014-00004-FHS) to E.C.; and by a Guangdong Science and Technology Department grant (2017KTSCX151) and a Guangzhou Science, Technology and Innovation Commission grant (201804010340) to Y.L. We would also like to acknowledge the Genomics and Bioinformatics Core at FHS. This work was performed in part at the High Performance Computing Cluster (HPCC), which is supported by the Information and Communication Technology Office (ICTO) of the University of Macau. We thank Jacky Chan Hoi Kei and William Pang from HPCC for their support and help.

## AUTHOR CONTRIBUTIONS

The study was designed by E.C. Experiments were performed by D.Z., G.C., S.C., Y.X.S., M.G.L.L., D.G., P.R., L.H., and Y.L. Data were analyzed by D.Z., Z.Z., and X.C. The manuscript was written by D.Z. and E.C. All authors read and approved the final manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 13, 2018

Revised: September 18, 2018

Accepted: October 24, 2018

Published: November 20, 2018; corrected online: May 28, 2019

## REFERENCES

- Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of transcript variability in single mammalian cells. *Cell* 163, 1596–1610.
- Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33, 155–160.
- Cheung, E., and Kraus, W.L. (2010). Genomic analyses of hormone signaling and gene regulation. *Annu. Rev. Physiol.* 72, 191–218.
- DeSantis, C., Ma, J., Bryan, L., and Jemal, A. (2014). Breast cancer statistics, 2013. *CA Cancer J. Clin.* 64, 52–62.
- Drabovich, A.P., Pavlou, M.P., Dimitromanolakis, A., and Diamandis, E.P. (2012). Quantitative analysis of energy metabolic pathways in MCF-7 breast cancer cells by selected reaction monitoring assay. *Mol. Cell. Proteomics* 11, 422–434.
- Ducker, G.S., and Rabinowitz, J.D. (2017). One-carbon metabolism in health and disease. *Cell Metab.* 25, 27–42.
- Ducker, G.S., Chen, L., Morscher, R.J., Gherguovich, J.M., Esposito, M., Teng, X., Kang, Y., and Rabinowitz, J.D. (2016). Reversal of cytosolic one-carbon flux compensates for loss of the mitochondrial folate pathway. *Cell Metab.* 24, 640–641.
- Durruthy-Durruthy, R., and Heller, S. (2015). Applications for single cell trajectory analysis in inner ear development and regeneration. *Cell Tissue Res.* 361, 49–57.
- Forbes, N.S., Meadows, A.L., Clark, D.S., and Blanch, H.W. (2006). Estradiol stimulates the biosynthetic pathways of breast cancer cells: detection by metabolic flux analysis. *Metab. Eng.* 8, 639–652.

- Frasor, J., Danes, J.M., Komm, B., Chang, K.C., Lytle, C.R., and Katzenellenbogen, B.S. (2003). Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. *Endocrinology* **144**, 4562–4574.
- Harper, M.A., Chen, Z., Toy, T., Machado, I.M., Nelson, S.F., Liao, J.C., and Lee, C.J. (2011). Phenotype sequencing: identifying the genes that cause a phenotype directly from pooled sequencing of independent mutants. *PLoS ONE* **6**, e16517.
- Hastie, T., and Tibshirani, R. (1986). Generalized additive models. *Stat. Sci.* **1**, 297–310.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.
- Jézéquel, P., Campone, M., Gouraud, W., Guérin-Charbonnel, C., Leux, C., Ricolleau, G., and Campion, L. (2012). bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast Cancer Res. Treat.* **131**, 765–775.
- Joseph, R., Orlov, Y.L., Huss, M., Sun, W., Kong, S.L., Ukil, L., Pan, Y.F., Li, G., Lim, M., Thomsen, J.S., et al. (2010). Integrative model of genomic factors for determining binding site selection by estrogen receptor- $\alpha$ . *Mol. Syst. Biol.* **6**, 456.
- Kahana, C. (2009). Antizyme and antizyme inhibitor, a regulatory tango. *Cell. Mol. Life Sci.* **66**, 2479–2488.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36.
- Leng, N., Chu, L.F., Barry, C., Li, Y., Choi, J., Li, X., Jiang, P., Stewart, R.M., Thomson, J.A., and Kendziorski, C. (2015). Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments. *Nat. Methods* **12**, 947–950.
- Liu, M.H., and Cheung, E. (2014). Estrogen receptor-mediated long-range chromatin interactions and transcription in breast cancer. *Mol. Cell. Endocrinol.* **382**, 624–632.
- Métivier, R., Penot, G., Hübner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763.
- Nilsson, R., Jain, M., Madhusudhan, N., Sheppard, N.G., Strittmatter, L., Kampf, C., Huang, J., Asplund, A., and Mootha, V.K. (2014). Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nat. Commun.* **5**, 3128.
- Pedley, A.M., and Benkovic, S.J. (2017). A new view into the regulation of purine metabolism: the purinosome. *Trends Biochem. Sci.* **42**, 141–154.
- Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393.
- Santen, R.J., Yue, W., and Wang, J.P. (2015). Estrogen metabolites and breast cancer. *Steroids* **99 (Pt A)**, 61–66.
- Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240.
- Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369.
- Simpson, E.H. (1951). The interpretation of interaction in contingency tables. *J. R. Stat. Soc. Series B Stat. Methodol.* **13**, 238–241.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* **102**, 15545–15550.
- Tong, P., Chen, Y., Su, X., and Coombes, K.R. (2013). SIBER: systematic identification of bimodally expressed genes using RNAseq data. *Bioinformatics* **29**, 605–613.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578.
- Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**, 1145–1160.
- Walker, G., MacLeod, K., Williams, A.R., Cameron, D.A., Smyth, J.F., and Langdon, S.P. (2007). Estrogen-regulated gene expression predicts response to endocrine therapy in patients with ovarian cancer. *Gynecol. Oncol.* **106**, 461–468.
- Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185.
- Yager, J.D. (2015). Mechanisms of estrogen carcinogenesis: The role of E2/E1-quinone metabolites suggests new approaches to preventive intervention—a review. *Steroids* **99 (Pt A)**, 56–60.
- Yamaga, R., Ikeda, K., Horie-Inoue, K., Ouchi, Y., Suzuki, Y., and Inoue, S. (2013). RNA sequencing of MCF-7 breast cancer cells identifies novel estrogen-responsive genes with functional estrogen receptor-binding sites in the vicinity of their transcription start sites. *Horm. Cancer* **4**, 222–232.

## STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE   | SOURCE  | IDENTIFIER  |
|---|---|---|
| Chemicals, Peptides, and Recombinant Proteins                 |   |   |
| 17 $\beta$ -estradiol (E2)                                    | Sigma   | Cat# 50-28-2  |
| Critical Commercial Assays                                    |   |   |
| RNeasy Mini Kit   | QIAGEN  | Cat# 74106  |
| SMARTer Ultra Low RNA Kit                                     | Clontech  | Cat# 634936   |
| C1 Single-Cell mRNA Seq IFC                                   | Fluidigm  | Cat# 100-6041   |
| C1 Single-Cell Reagent Kit for mRNA Seq                       | Fluidigm  | Cat# 100-6201   |
| C1 Single-Cell mRNA Seq HT IFC                                | Fluidigm  | Cat# 101-4982   |
| C1 Single-Cell mRNA Seq HT Reagent Kit v2                     | Fluidigm  | Cat# 101-3473   |
| NexTera XT Sample Prep Kit                                    | Illumina  | Cat# FC-131-1096  |
| NexTera XT DNA Sample Preparation Index Kit                   | Illumina  | Cat# FC-131-1002  |
| KAPA Library Quantification Kit                               | KAPA Biosystems   | Cat# KR0405   |
| NexTera XT Index Kit v2                                       | Illumina  | Cat# FC-131-2001(set A) and FC-131-2002(set B).   |
| ddPCR Supermix for Probes (no dUTP)                           | BioRad  | Cat# 186-3024   |
| ddPCR Droplet Reader Oil                                      | BioRad  | Cat# 186-3004   |
| DG8 Cartridges and Gaskets                                    | BioRad  | Cat# 186-4007   |
| ddPCR 96-Well PCR Plates                                      | BioRad  | Cat# 12001925   |
| RNAscope Multiplex Fluorescent v2 Kit                         | ACDbio  | Cat# 323110   |
| Cell apoptosis kit  | ThermoFisher  | Cat# V13245   |
| MTT   | ThermoFisher  | Cat# M6494  |
| Deposited Data  |   |   |
| Raw and analyzed data   | This paper  | GEO: GSE107864, GSE119455   |
| Genome reference UCSC hg19                                    | UCSC  | <a href="http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/">http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/</a>                                 |
| Gene annotation reference Gencode V18                         | Gencode database  | <a href="https://www.gencodegenes.org/18.html">https://www.gencodegenes.org/18.html</a>   |
| ER $\alpha$ ChIP-seq data                                     | Ross-Innes et al., 2012   | GEO: GSE32222   |
| Experimental Models: Cell Lines                               |   |   |
| Human: MCF7   | ATCC  | ATCC HTB-22   |
| Human: T47D   | ATCC  | ATCC HTB-133  |
| Oligonucleotides  |   |   |
| Primer sets and probes used in droplet digital PCR (Table S2) | This paper  | N/A   |
| DsiRNAs used for knockdown assays (Table S3)                  | This paper  | N/A   |
| Software and Algorithms                                       |   |   |
| Tophat v2.0.6   | Kim et al., 2013  | <a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>   |
| Cufflink v2.0.2   | Trapnell et al., 2012   | <a href="http://cole-trapnell-lab.github.io/cufflinks">http://cole-trapnell-lab.github.io/cufflinks</a>   |
| FastQC v0.11.5  | <a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a> | <a href="https://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc">https://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc</a> |
| RSeQC v2.6.4  | Wang et al., 2012   | <a href="http://rseqc.sourceforge.net/">http://rseqc.sourceforge.net/</a>   |
| SIBER (OOMPA v2.1.0)  | Tong et al., 2013   | <a href="http://bioinformatics.mdanderson.org/main/OOMPA:Overview">http://bioinformatics.mdanderson.org/main/OOMPA:Overview</a>                               |
| bc-GenExMiner 4.0   | Jézéquel et al., 2012   | <a href="http://bcgenex.centregauducheau.fr/BC-GEM/GEM-Accueil.php?js=1">http://bcgenex.centregauducheau.fr/BC-GEM/GEM-Accueil.php?js=1</a>                   |
| DAVID website v6.8  | Huang et al., 2009  | <a href="https://david.ncifcrf.gov/">https://david.ncifcrf.gov/</a>   |
| GSEA software (MSigDB 6.2)                                    | Subramanian et al., 2005  | <a href="http://software.broadinstitute.org/gsea/index.jsp">http://software.broadinstitute.org/gsea/index.jsp</a>   |

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Edwin Cheung ([echeung@umac.mo](mailto:echeung@umac.mo)).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Cell Lines

MCF-7 and T47D ER $\alpha$ -positive human breast cancer cell lines were obtained from ATCC and maintained in DMEM and RPMI 1640 media (GIBCO), respectively. Both media were supplemented with 10% fetal bovine serum (FBS) (GIBCO) and grown in a 37°C incubator with 5% CO<sub>2</sub>.

### METHOD DETAILS

#### E2 stimulation and scRNA-seq library generation

Three days prior to E2 stimulation, MCF-7 and T47D cells were transferred into phenol-red free DMEM/F12 and RPMI media (GIBCO), respectively. Both media were supplemented with 10% charcoal/dextran-stripped FBS (Hyclone). Cells were then stained with CellTracker Green CMFDA or Orange CMTMR (Molecular Probes) 24 h before drug treatment. Stained cells were treated with either ethanol (vehicle) or E2 at 4 time points (0 and 3 h: Green, 6 and 12 h: Orange), before being dissociated into single-cell suspension for single-cell and bulk population analyses.

Single-cell cDNA samples were obtained using the Fluidigm C1 Single-Cell Auto Prep System. Briefly, 125,000 cells from two time points (one labeled red and the other labeled green) were mixed together in 1 mL of media and then loaded into a 10–17  $\mu$ m Fluidigm C1 Integrated Fluidic Circuit (IFC). The IFC was loaded into the C1 Single-Cell Auto Prep System for cell capture. After cell capture, the IFC was imaged using an automated inverted fluorescence microscope to identify the captured cells based on their fluorescence. Then cell lysis and cDNA synthesis processes were continued on the chip. Finally, single-cell cDNA samples were harvested as per manufacturer's instructions.

Cells from each time point were also used in bulk population cDNA generation. For these cells, RNA was extracted using the RNeasy Mini Kit (QIAGEN) and then reverse transcribed using the SMARTer Ultra Low RNA Kit (Clontech) according to the manufacturer's protocol.

Microscope images of the single-cells were examined to exclude wells with multiple cells or cells with multiple staining. Single-cell and bulk population samples were barcoded and pooled together for library preparation using the Nextera XT Sample Prep Kit (Illumina). The pooled libraries were quantified through quantitative PCR using the KAPA Library Quantification Kit (KAPA Biosystems), followed by sequencing on the Illumina platform (2  $\times$  76 bp for the MCF-7 libraries or 2  $\times$  101 bp for the T47D libraries).

For HT single-cell mRNA analysis, we used the 10–17  $\mu$ m HT IFCs. Briefly, 5,200 MCF-7 cells were loaded into a HT IFC and capture efficiency was measured under a microscope after cell capture. Capture sites that contained no cell or with more than one cell were marked and excluded from further analysis. For the HT IFC, a cell-specific barcode is applied during cDNA conversion and a Nextera index is added during the 3' end enrichment step using the Nextera XT Index Kit v2 (Illumina). The libraries are then pooled, quantified, and sequenced as described above.

#### scRNA-seq data processing

All RNA-seq reads were first trimmed to 75 bp. Next, the adaptor sequences were trimmed. Reads were mapped as paired-end sequencing reads with the Tophat package (v.2.0.6) against UCSC hg19 and the gene annotation reference Gencode V18. The FPKM (Fragments Per Kilobase of exon model per Million mapped fragments) expression values were calculated from the BAM file generated by Tophat using the Cufflink package (v.2.0.2). Genes with FPKM greater than one in at least one sample were retained for further analysis. Sequencing quality was evaluated using the FastQC and RSeQC ([Wang et al., 2012](#)) packages to assess the single-cell libraries on the total sequenced reads, uniquely mapped reads, 5' or 3' coverage bias, as well as the percentage of intronic bases and intergenic bases. A sample was deemed biased if the coverage coefficient of the median gene body position was less than 0.8 ([Figure S1B and C](#)). Single-cell libraries with a low sequencing depth of fewer than 1 million reads or exhibiting 5' or 3' coverage bias were filtered out. In total, 84 MCF-7 and 78 T47D single-cell samples ([Table S1](#)) were used for downstream analyses. We performed unsupervised clustering on all the bulk population and single-cell samples based on Pearson correlation coefficients of the global expression profiles between every two samples.

For processing the HT samples, demultiplexing was performed using Fluidigm's HT Demultiplexer Perl script. All reads were trimmed to 75 bp and then aligned to the reference genome using Tophat2 ([Kim et al., 2013](#)). Gene expression was calculated using cufflinks. Single-cells with less than 20,000 reads or less than 4,000 detected genes were filtered out. In total, 507 and 594 cells from 0 and 12 h were used for downstream analyses.

**Gene expression bimodality analysis**

A total of 5,975 coding genes were assessed for bimodality expression using the R package, SIBER (Tong et al., 2013), which is a bioinformatic approach for identifying bimodal expressed genes from RNA-seq data. First, we specified a *normal mixture model* (model = 'NL') on the log2 transformed FPKM values to fit the single-cell gene expression distribution into a two-component mixture model (component 1 and 2) and then calculated the mean values ( $\mu_1$  and  $\mu_2$ ), variance values ( $\sigma_1^2$ ,  $\sigma_2^2$ ), and the proportion of samples ( $p_1$  and  $p_2$ ) for these two components. Next, a bimodal index (BI) value was calculated to assess the potential bimodality for each gene using the following equation:

$$BI = \frac{\sqrt{p_1 \cdot p_2} \cdot |\mu_1 - \mu_2|}{\sqrt{p_2 \cdot \sigma_1^2 + p_1 \cdot \sigma_2^2}}.$$

BI rewards balanced allocation of the samples into two components (as it is largest when  $p_1 = p_2$ ) and the effect size between the two components was defined as:

$$\delta = |\mu_1 - \mu_2| / \sqrt{p_2 \cdot \sigma_1^2 + p_1 \cdot \sigma_2^2}.$$

We repeated this procedure 1,000 times to generate a set of random statistics and averaged the top 5%. We defined a gene with a BI value  $> 1.8$  (FDR = 0.05) in at least one time point as a bimodally expressed gene (generated from 1,000 times random repeats and top 5% confident degree).

**Mann-Whitney rank sum test**

Genes with FPKM  $< 1$  in more than median number of single-cell samples were filtered. We compared samples from the 3, 6, and 12 h time points with those from the 0 h time point. We performed Mann-Whitney rank sum test (U-test) on the genes and considered  $p < 0.05$  as significant DEGs.

**Pseudotemporal analysis**

The Monocle (Trapnell et al., 2012) package is a recently developed algorithm for scRNA-seq data analysis, which highlights the dynamic transcriptome with increased temporal resolution along time points by an artificial temporal curve. Briefly, the Monocle algorithm implements a dimension reduction by applying independent component analysis (ICA) prior to cell ordering. Artificial temporal curve are generated by minimal spanning trees (MST) that connect cells along the longest possible path involving as many cells as possible. We selected the differential expression genes ( $q < 0.05$ ) based on real time to construct the MST and calculated the pseudotime for each single sample. DEGs across the pseudotime axis were identified using a generalized additive model (GAM) (Hastie and Tibshirani, 1986). Testing for differential expression was performed with an approximate  $\chi^2$  likelihood ratio test. We then measured DEG list based on the pseudotime using a threshold  $p$  value  $< 0.05$ .

**Single-cell droplet digital PCR**

Single-cell cDNA samples were diluted by 100 times and 1  $\mu$ L was used for detecting the estrogen-responsive genes, ESR1, TFF1, and GREB1, and the house-keeping gene, ACTB. Duplexed droplet digital PCR (ddPCR) was performed according to the manufacturer's protocol and the results were reported as copies of mRNA per  $\mu$ L sample. In brief, 1  $\mu$ L of DNA template was added to 20  $\mu$ L of ddPCR Supermix for Probes (Bio-Rad) and primer/probe mixtures of ESR1 (FAM probe) + ACTB (HEX probe) or TFF1 (FAM probe) + GREB1 (HEX probe). The 20  $\mu$ L samples and 70  $\mu$ L of Droplet Generation Oil for Probes (Bio-Rad) were used for droplet generation. Droplets were then thermal cycled with the following conditions: 5 min at 95°C, 40 cycles of 94°C for 30 s, 55°C for 1 min followed by 98°C for 10 min (ramp rate 2°C/sec). Samples were then transferred to a QX200 Droplet Reader (Bio-Rad) for fluorescent measurement of FAM and HEX signals. Gating was performed on the basis of negative controls and the copies of mRNA were calculated for each sample using QuantaSoft software (Bio-Rad). Experiments were performed in 2 technical replicates.

**Single-molecule RNA fluorescent *in situ* hybridization**

Multiplexed single-molecule RNA fluorescent *in situ* hybridization (smRNA-FISH) was performed using RNAscope probes (ACDBio) according to the manufacturer's protocol. MCF-7 and T47D cells were treated with E2 treatment at 0, 3, 6 and 12 hr. TFF1, GREB1, and ESR1 mRNAs were all probed at the same time. At least 200 cells per condition were counted for the mRNA copies for each gene.

**Functional enrichment analysis and public datasets used in this study**

KEGG pathway enrichment analysis of the DEGs was performed using the DAVID database (Huang et al., 2009). Gene set enrichment analysis was performed using the Broad Institute GSEA software (Subramanian et al., 2005). Prognostic analysis of selected genes in ER $\alpha$ -positive and negative tumors was performed using bc-GenExMiner 4.0 with 5,861 clinical cases from a collection of 36 previous breast cancer cohort studies (Jézéquel et al., 2012). ER $\alpha$  ChIP-seq data from breast tumor tissues were retrieved from GEO: GSE32222 (Ross-Innes et al., 2012).

### Cell proliferation assays

To assess the effects of AZIN1 and PPAT depletion on MCF-7 cell proliferation, cells were first transfected with dsRNAs targeting AZIN1, PPAT, or a non-targeting (NC) control. 24 h after transfection, cells were then seeded into 96-well plates at  $3 \times 10^3$  cells/well. Cells were incubated with 5 mg/mL sterilized MTT (Life Technologies) in PBS for 3 h at 37°C. The plate was then incubated for a further 30 min at 37°C. The absorbance was measured at an OD of 590 nm using a plate reader. Cell proliferation was monitored every day over a period of 6 days. Experiments were performed in 3 technical replicates and repeated 3 independent times.

### Cell apoptosis assays

Cell apoptosis assay was performed using the Dead Cell Apoptosis Kit (Invitrogen) according to the manufacturer's instruction. In brief, MCF-7 cells were transfected with dsRNAs targeting AZIN1, PPAT or a non-targeting (NC) control. After 72 h of transfection, the cells were harvested, centrifuged at 1,200 rpm for 5 min at 4°C, and then resuspended in binding buffer containing Annexin V-FITC and propidium iodide. The cells were incubated for 15 min in the dark at room temperature. After staining, the cells were mixed with 400  $\mu$ L binding buffer and immediately analyzed by flow cytometry. Experiments were repeated 3 independent times.

## QUANTIFICATION AND STATISTICAL ANALYSIS

For the smRNA-FISH experiments, at least 200 cells per condition were counted for the mRNA copies for each gene, and the results are presented as violin plots. For the other assays, the ddPCR standard assays were performed in 2 technical replicates, the cell proliferation assays were performed in 3 technical replicates and repeated 3 independent times, and the cell apoptosis assays were repeated 3 independent times, and the results are represented as mean  $\pm$  s.d. The statistical significance of differences is determined by paired, two-tailed Student's t-Test. A p value  $< 0.05$  is considered to be statistically significant.

## DATA AND SOFTWARE AVAILABILITY

### Data Resources

The accession numbers for the scRNA-seq data reported in this paper are GEO: GSE107864 and GSE119455.