Article

Identifying key regulatory genes in drug resistance acquisition: Modeling pseudotime trajectories of single-cell transcriptome

Keita Iida [1,*] and Mariko Okada [1]

[1] Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan

* Correspondence: kiida@protein.osaka-u.ac.jp

Summary: Despite recent advancements in experimental technology for genome-wide molecular profiling, our understanding of the dynamic mechanism underlying cancer drug resistance remains limited. In this study, we present an approach that combines mathematical modeling with the pseudotime analysis of single-cell time-series transcriptome data of drug-treated breast cancer cells. Our method identifies approximately 600 genes out of 6,000 exhibiting multistable expression states, including *RPS6KB1*, a predictor of poor prognosis, cell survival, and growth in estrogen receptor-positive breast cancers. The bifurcation analysis elucidates the regulatory mechanisms of the key regulatory genes, which can also be mapped into a molecular network based on cell survival and metastasis-related pathways, providing a comprehensive understanding of the interplay between signaling pathways and regulatory genes. Our method serves as a powerful tool for deciphering the complexities of drug resistance mechanisms in human diseases.

Abstract: Single-cell RNA-sequencing (scRNA-seq) technology has provided significant insights into cancer drug resistance at the single-cell level. However, understanding dynamic cell transitions at the molecular systems level remains limited, requiring a systems biology approach. We present an approach that combines mathematical modeling with pseudotime analysis using time-series scRNA-seq data obtained from the breast cancer cell line MCF-7 treated with tamoxifen. Our single-cell analysis identified five distinct subpopulations, including tamoxifen-sensitive and -resistant groups. Using a single-gene mathematical model, we discovered approximately 560–680 genes out of 6,000 exhibiting multistable expression states in each subpopulation, including key estrogen receptor-positive breast cancer cell survival genes, such as *RPS6KB1*. Bifurcation analysis elucidated their regulatory mechanisms, and we mapped these genes into a molecular network associated with cell survival and metastasis-related pathways. Our modeling approach comprehensively identifies key regulatory genes for drug resistance acquisition, enhancing our understanding of potential drug targets in breast cancer.

Keywords: single-cell transcriptome; pseudotime analysis; nonautonomous ordinary differential equation; drug resistance; breast cancer; intracellular signaling

1. Introduction

Breast cancer has been the most prevalent cancer among women and a leading cause of death since the late 20th century [1-3]. In the United States, approximately 75% of breast cancers are diagnosed as luminal A subtype [4], indicating their dependency on estrogen receptor (ER) for growth. Tamoxifen, a widely used endocrine therapy for ER-positive breast cancer, has resulted in improvements in overall survival [5, 6]. However, a significant proportion of patients treated with tamoxifen experience relapse within 5–15 years, often resulting in metastasis and mortality [5, 6]. Despite accumulating knowledge on genetic and nongenetic molecular mechanisms of endocrine resistance, such as *ESR1* mutation, transcriptomic and epigenetic changes, and activation of signaling pathways [6-8], our understanding of the dynamic processes remains limited, requiring a comprehensive and systemic approach.

Recently, single-cell RNA-sequencing (scRNA-seq) analysis using MCF-7 cells, a luminal A subtype breast cancer model, has provided insights into drug resistance acquisition, revealing cellular heterogeneity and dynamic cell transitions. Notably, a "pre-adapted" subpopulation exhibiting plasticity between drug tolerance and resistance has been identified in MCF-7 cells [9, 10]. These studies investigated the activation of cell survival signals, such as the NF-κB pathway. Furthermore, Magi et al. reported altered metabolic regulation, cell adhesion, and histone modification based on the analysis of time-series scRNA-seq data under tamoxifen treatment [11]. While these studies have uncovered important genes and pathways associated with drug resistance acquisition, gaps still exist in our understanding of regulatory mechanisms at the systems level. Hence, further exploration for potential drug targets is warranted.

To address these issues, we propose an approach that combines mathematical modeling with pseudotime analysis using time-series transcriptome data, which automatically identifies key regulatory genes. We first identified potential drug-sensitive and -resistant cells by analyzing the time-series scRNA-seq data of tamoxifen-treated MCF-7 cells. Next, we performed pseudotime analysis, from which we observed a transcriptional relay involved in cAMP, PI3K-AKT-mTOR, RHO-GTP, NF-κB, and NOTCH pathways. We identified genes exhibiting multistable expression states by modeling the individual gene expression patterns and conducting steady-state analysis. Finally, we propose a transcriptional relay network model that includes cell survival, focal adhesion, and epithelial–mesenchymal transition (EMT), along with the key regulatory genes. Our results suggest that the modeling approach through pseudotime analysis of single-cell transcriptome serves as a powerful tool for understanding drug-induced cellular transition.

2. Results

2.1. Overview of the analysis workflow

To theoretically address the challenge of identifying effective targets for regulating complex molecular systems during drug resistance acquisition (**Figure 1a**), we developed an approach combining mathematical modeling and pseudotime analysis using time-series transcriptome data, which automatically identifies key regulatory genes (**Figure 1b**). Previous research on cell differentiation revealed that transcription factor networks exhibiting multistable states simulate dynamic cell transitions of progenitor cells into different terminal states [12-14]. This suggests that genes with multistable expression states could serve as potential indicators of cell differentiation, potentially regulated by switch-like responding pathways. To identify these key regulatory genes, we modeled each gene expression using an ordinary differential equation, determining whether the expression state in the long-time limit is mono- or multistable. Although multistability was investigated in the steady-state, we considered these "multistable genes" key regulators, assuming cell fate is determined within the long-time limit. Subsequently, as discussed later, we constructed a transcriptional relay network model for cell survival, focal adhesion, and EMT.

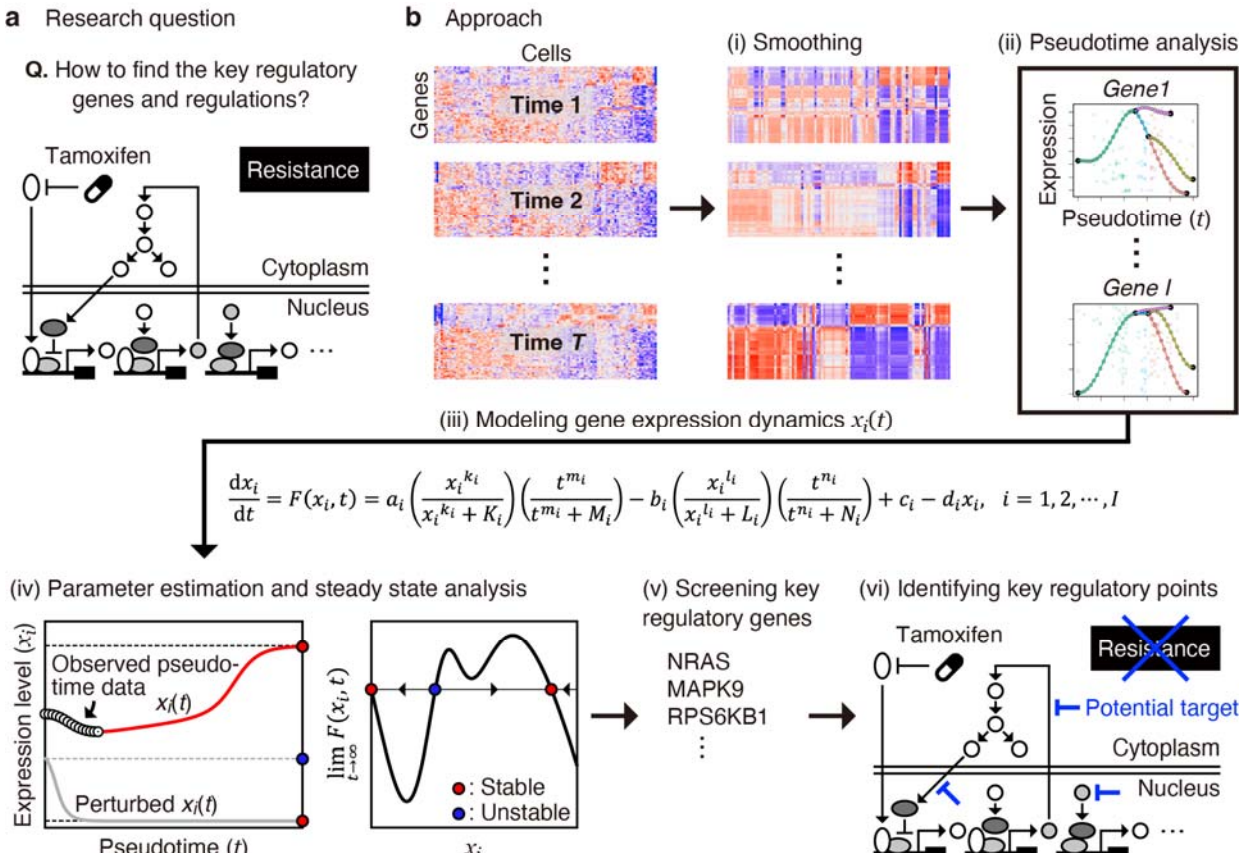

Figure 1. Research question and overview of our modeling approach. (**a**) Research question: How to identify the effective targets regulating complex molecular systems upon drug resistance acquisition. (**b**) Modeling approach: (i) Smoothing of single-cell data; (ii) pseudotime analysis, which outputs pseudotime gene expression data; (iii) mathematical modeling of each gene expression using an ordinary differential equation; (iv) parameter estimation by fitting the model (left) and steady-state analysis (right); (v) screening of key regulatory genes exhibiting multistable expression levels at equilibrium; and (vi) mapping of key regulatory genes into a molecular network associated with drug resistance acquisition and identifying potential regulatory points.

2.2. Dissecting the heterogeneity and defining pseudotime for MCF-7 cells

To understand transcriptional regulation during drug resistance acquisition, we analyzed a published dataset containing time-series scRNA-seq data of MCF-7 cells subjected to continuous tamoxifen treatment [11]. The data were collected at weeks 0 (before tamoxifen treatment), 3, 6, and 9, referred to as W0, W3, W6, and W9, respectively (**Figure 2a**). The authors reported that cell growth was inhibited by the fifth week, after which two resistant subpopulations emerged: one displaying altered metabolic regulation and the other exhibiting cell adhesion and histone modification. To provide a deeper insight into these findings, we initially analyzed W0 data to identify potential drug-sensitive and -resistant cells present before tamoxifen treatment. Subsequently, we analyzed W0–W9 data to derive pseudotime gene expression data.

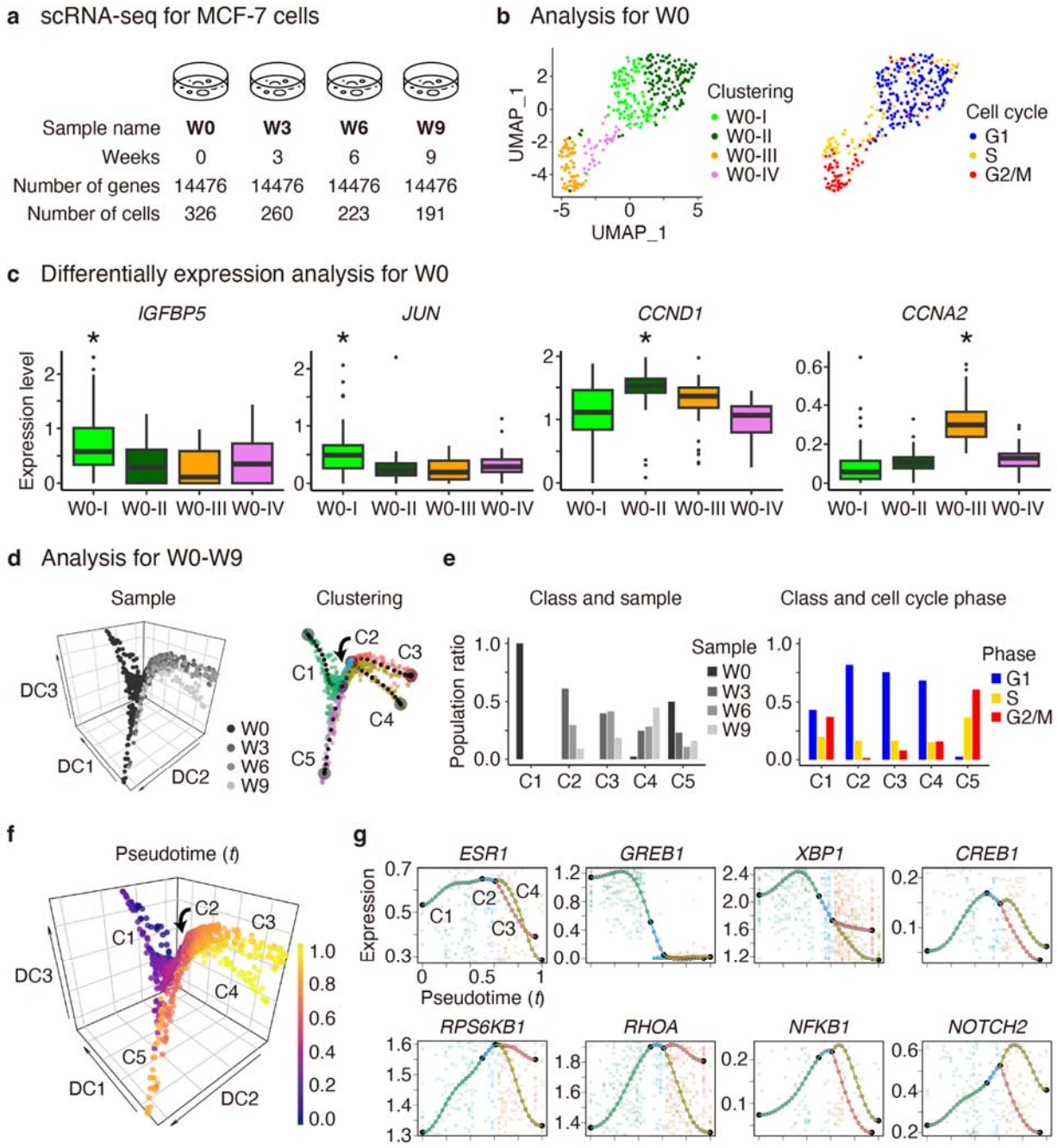

Figure 2. Statistical and pseudotime analysis of time-series single-cell RNA-seq data of MCF-7 cells under continuous tamoxifen treatment. (**a**) Overview of the dataset. (**b**) Uniform manifold approximation and projection (UMAP) plots for the control sample (W0), showing the clustering results (left) and cell cycle phases (right) computed using Seurat. (**c**) Box plots for W0, showing the normalized expression levels of differentially expressed genes. Adjusted *P*-values for the clusters marked with asterisks indicate statistical significance (*$P < 0.001$, Mann–Whitney U test by Seurat). (**d**) Diffusion map plots for all samples (W0–W9), showing the sample labels (left) and clustering labels (right) computed using MERLoT. (**e**) Population ratios across clustering versus sample labels (left) and clustering versus cell cycle phases (right). (**f**) Diffusion map plot showing pseudotime *t*, where we set $t = 0$ and $t = 1$ at the putative start and end points of branches C1 and C4, respectively. (**g**) Pseudotime gene expression profiles of key regulators in estrogen receptor pathway (such as *ESR1*), cAMP pathway (*GREB1*, *XBP1*, and *CREB1*), PI3K-AKT (*RPS6KB1*), RHOA-GTP (*RHOA*), NF-κB (*NFKB1*), and NOTCH (*NOTCH2*) pathways, which are sequentially upregulated from top left to bottom right, suggesting the existence of transcriptional relay.

After normalizing W0–W9 data (**Appendices A** and **B**), we initially focused on W0. Using Seurat [15], a standard software for scRNA-seq data analysis, we clustered the cells into four distinct subpopulations (W0-I, -II, -III, and -IV) and inferred the cell cycle phases (**Figure 2b**). We examined differentially expressed genes to assess potential drug-resistant states (**Figure 2c**). We observed significant upregulation of *IGFBP5* and *JUN* in the W0-I subpopulation. Previous studies have implicated IGFBP5 in cancer inhibition by temporally inactivating the insulin-like growth factor receptor [16], and JUN has been linked to increased tamoxifen sensitivity through protein kinase C activation [17, 18]. Conversely, a positive cell cycle regulator *CCND1* [19] and apoptosis inhibitor *XBP1* [20] were significantly upregulated in W0-II, while other positive cell cycle regulators *CCNA2* [21] and *CCNB1/2* [22] were significantly upregulated in W0-III (**Figure A1a**). A notable increase in mitochondrial read fractions [23] and results from Gene Ontology enrichment analysis [24] suggested that cells within W0-IV undergo apoptosis (**Figure A1b,c**).

Subsequently, we analyzed W0–W9 data using Seurat and conducted a principal component analysis (PCA) based on 5,000 highly variable genes, followed by a nonlinear dimensionality reduction using a diffusion map (**Figure 2d**, left). Thereafter, using MERLoT [25], pseudotime analysis software, we smoothed the gene expression profiles and clustered the cells into five classes C1–C5 (**Figure 2d**, right). We observed that cells in W0, W3–W6, and W9 predominantly belong to C1, C2–C3, and C4, respectively (**Figure 2e**, left). We noticed that the population ratio in the G2/M phase drastically decreased from C1 to C2 but gradually increased from C2 to C4 (**Figure 2e**, right). Since tamoxifen treatment induces G1 arrest followed by gradual cell growth recovery [11], we assumed that the C1-C2-C3-C4 axis approximately corresponds to the progression of tamoxifen resistance acquisition, whereas C5 predominantly consists of cells in S and G2/M phases.

Finally, we defined pseudotime t for all cells, where we set $t = 0$ and $t = 1$ at the putative start and end points in C1 and C4, respectively (**Figure 2f** and **Appendix C**). Consistent with our interpretation of pseudotime as the progression of tamoxifen resistance acquisition, we annotated W0-I and W0-II as the potential tamoxifen-sensitive and resistant populations, respectively (**Figure A1d**). According to the pseudotime gene expression profiles (**Figure 2g**), we observed that *ESR1* was downregulated for $t > 0.5$, suggesting that *ESR1*-independent mechanisms occur after tamoxifen treatment, as previously indicated [11]. We also noticed that key regulators in the cAMP (such as *GREB1*, *XBP1*, and *CREB1*), PI3K-AKT (*RPS6KB1*), RHOA-GTP (*RHOA*), NF-κB (*NFKB1*), and NOTCH (*NOTCH2*) pathways were sequentially upregulated, suggesting the existence of transcriptional relay.

2.3. Modeling pseudotime gene expression patterns

As extensively reviewed, several computational tools have been developed to construct gene regulatory networks from scRNA-seq data [26]. One common approach is transcription factor networks based on the assumption that mRNA levels correlate well with protein levels [27, 28]. However, questions arise from the intricate processes, including environmental changes (e.g., drug concentration [29]), interconnection of signaling pathways, chemical modification, and nuclear translocation of transcription factors, as well as cooperation with coactivators and corepressors [7]. These processes can significantly influence target gene expressions. To address such complexity, we modeled individual gene expressions using an autoregulatory gene circuit widely used for modeling various biochemical processes (**Figure 3a**) [30, 31], which is exemplified by an ER-related signaling pathway (**Figure 3b**) [32-35]. This model operates under the assumptions that all genes are controlled by positive and negative feedback mechanisms, and the strength of this feedback control is determined by the mRNA levels and interactions among intracellular signals. The model is described by the following nonautonomous ordinary differential equation:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = F(x_i, t), \quad i = 1, 2, \cdots, I, \#(1)$$

$$F(x_i, t) := a_i \left(\frac{{x_i}^{k_i}}{{x_i}^{k_i} + K_i} \right) \left(\frac{t^{m_i}}{t^{m_i} + M_i} \right) - b_i \left(\frac{{x_i}^{l_i}}{{x_i}^{l_i} + L_i} \right) \left(\frac{t^{n_i}}{t^{n_i} + N_i} \right) + c_i - d_i x_i,$$

where I is the number of genes, x_i is a normalized expression level ($x \geq 10^{-4}$) for gene i, a_i and b_i are the magnitudes of positive and negative feedback regulation, c_i and d_i are the transcription and degradation rates, K_i, L_i, M_i, and N_i are the half-saturation constants, and k_i, l_i, m_i, and n_i are Hill coefficients. All the parameter values are non-negative real numbers.

We separately estimated the model parameters for 6,082 genes, which passed a filter removing low-expressed genes—for C1, C3, and C4 subpopulations—wherein we neglected C2 and C5 due to the lack of sufficient pseudotime lengths (**Appendices D** and **E**). Numerical data fitting in mechanistic models with tens to hundreds of unknown parameters remains challenging [36]. Nevertheless, we addressed this problem using a simplified model (Eq. 1). Here, we used a multi-starts method [37], one of the successful approaches to avoid suboptimal estimation using different initial guesses, followed by a method using smooth and match estimators [38, 39], a fast optimization method fitting a curve $(x_i, t_i, F(x_i, t))$ to a given data (**Figure 3c**). Furthermore, we applied our original l_2-norm penalty method—a parameter set minimizing $\|(a_i, b_i, c_i, d_i)\|_2$ is likely to be selected—by assuming that these parameters represent energy costs of activation and inhibition for gene expressions. This restriction improved the robustness of the estimation.

We observed that fitting errors were generally small (< 0.1) for most genes, indicating a good fit of our model to the pseudotime gene expression data (**Figure 3d**). However, suboptimal parameter sets, derived from different initial guesses during local optimization, also yielded small fitting errors but resulted in different predictions (**Figure A2**). This suggests the complexity of our model, which may involve an excessive number of parameters compared with the pseudotime data and a possibility of overfitting. Thus, we evaluated the robustness of parameter optimization against small differences in gene expression profiles, controlling the number of representative pseudotime points, denoted as n_{pc} (i.e., the number of "nodes" in **Figure 2d**, right), which impacts the pseudotime gene expression profiles [25].

To test whether gene-wide median parameter values were consistent across different values of n_{pc}, we independently conducted a nonparametric one-way ANOVA for C1, C3, and C4 (**Figure 3e**). Notably, all 12 parameters for C1 showed no significant differences between $n_{pc} = 55, 56, \cdots, 60$, whereas parameters a_i, b_i, and N_i exhibited significant differences between $n_{pc} = 55, 56, \cdots, 65$ (**Figure 3e**, left), which may be attributed to fluctuations in parameter values observed for certain genes (**Figure 3e**, right). Robustness was also observed for C3 and C4 (**Figure A3** and **Appendix F**). Moreover, we verified that all 12 parameters were essential for our model, confirming that every principal component computed from all the estimated parameter values has almost equal contributions (**Figure A4**). Consequently, we demonstrated statistically accurate and robust parameter estimation, offering a reliable method for predicting cellular transition states.

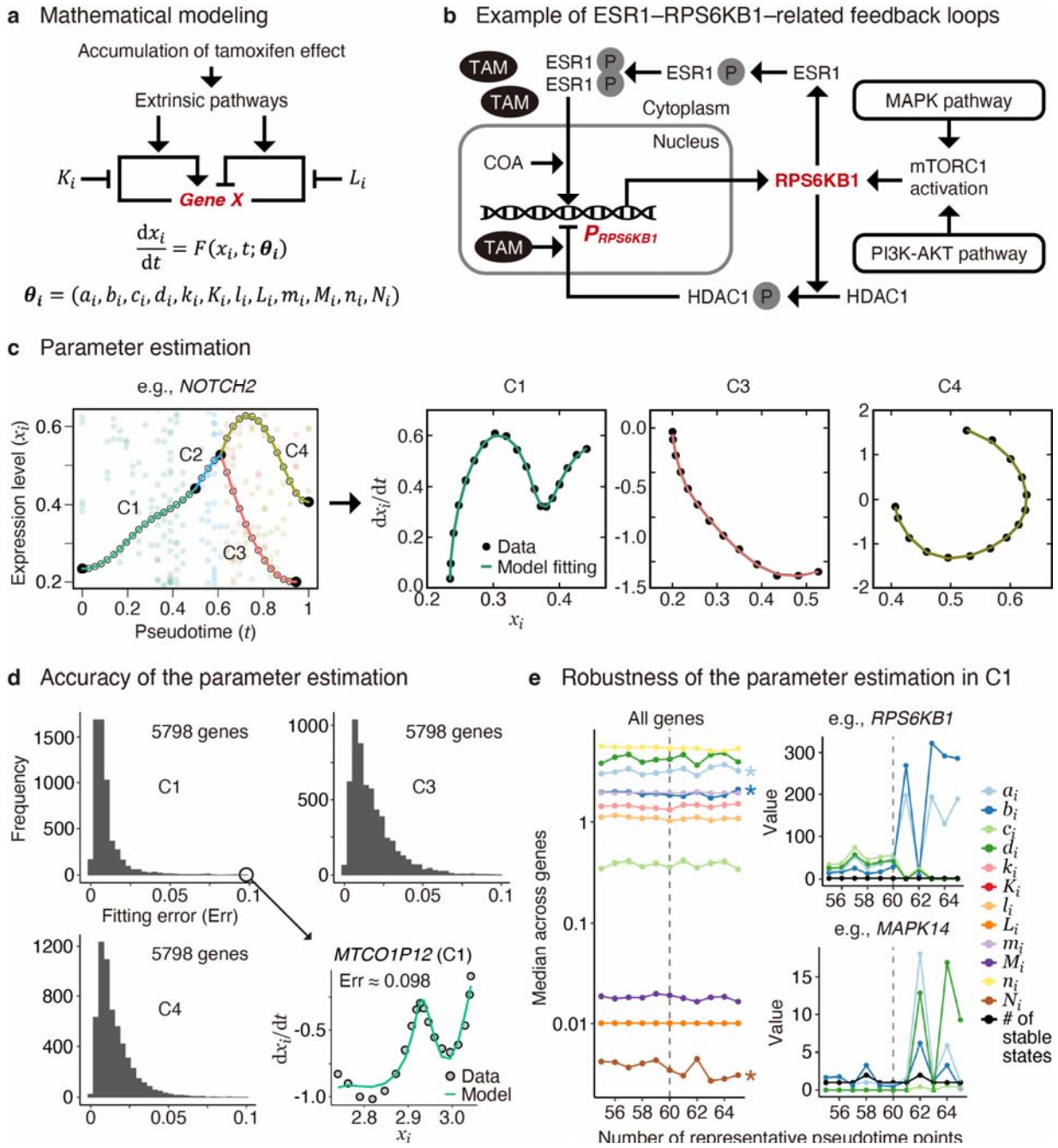

Figure 3. Mathematical modeling, parameter estimation, and evaluation of accuracy and robustness. (**a**) Schematic representation of the mathematical model in which intracellular tamoxifen levels affect gene regulation via extrinsic pathways. The arrows and blunt arrows represent positive and negative regulation, respectively. Particularly, the K_i and L_i parameters represent inhibitory effects on positive and negative regulation, respectively. (**b**) Schematic representation of *RPS6KB1* regulation, an example of estrogen receptor-related signaling. (**c**) Demonstration of parameter estimation using *NOTCH2* pseudotime trajectories. Parameters were independently optimized for subpopulations C1, C3, and C4, whereas C2 and C5 were neglected due to insufficient pseudotime length. (**d**) Histograms of the fitting errors showing that the errors are generally small ($\leq$ 0.1) for most genes (5,798 genes out of 6,082), as exemplified by the result for *MTCO1P12* in C1 at the right bottom. (**e**) Gene-wide median parameter values for C1 (left), along with individual parameter values and the number of stable fixed points for *RPS6KB1* (top right) and *MAPK14* (bottom right) across different numbers of representative pseudotime points (n_{pc}). In this study, $n_{pc} = 60$ was used (indicated by vertical broken lines). According to the Kruskal–Wallis test, all 12 parameters showed no significant differences between $n_{pc} = 55, 56, \cdots, 60$ (P-value $> 10^{-4}$), whereas a_i, b_i, and N_i exhibited significant differences between $n_{pc} = 55, 56, \cdots, 65$: *$P$-values $< 10^{-10}$.

2.4. Transient and bifurcation analysis for key regulatory genes

We conducted steady-state analysis to identify genes exhibiting multistable expression states at equilibrium. We counted the number of stable fixed points by deriving the steady-state equation from Eq. (1) and computing its zero points with a negative differential coefficient. Our analysis revealed that 664, 558, and 680 genes exhibited multistability in C1, C3, and C4 subpopulations, respectively (**Figure 4a**). Notably, RPS6KB1, a predictor of poor prognosis, cell survival, and growth in ER-positive breast cancers [40], exhibited bistability in both C1 and C4, implying that its low and high expression levels are regulated by a binary switch governing distinct cell fates.

To further investigate transient gene expression dynamics dependent on initial conditions, we computed a potential landscape analogous to Waddington's landscape [41]. This landscape visually represents how gene expression states evolve over time, analogous to marbles rolling on hills and valleys. Derived from Eq. (1), the potential function $U(x_i, t)$ is given by the following gradient form of the landscape:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -\nabla U(x_i, t), \quad i = 1, 2, \cdots, I, \#(2)$$

$$U(x_i, t) := \frac{a_i {x_i}^{k_i+1}}{K_i(k_i+1)}\, {}_2F_1\left(1, 1+\frac{1}{k_i}, 2+\frac{1}{k_i}; -\frac{{x_i}^{k_i}}{K_i}\right)\left(\frac{t^{m_i}}{t^{m_i}+M_i}\right)$$
$$-\frac{b_i {x_i}^{l_i+1}}{L_i(l_i+1)}\, {}_2F_1\left(1, 1+\frac{1}{l_i}, 2+\frac{1}{l_i}; -\frac{{x_i}^{l_i}}{L_i}\right)\left(\frac{t^{n_i}}{t^{n_i}+N_i}\right) + c_i x_i - \frac{d_i}{2}{x_i}^2,$$

where ${}_2F_1$ is the Gauss hypergeometric function. Our observations indicated that the long-time behavior of *RPS6KB1* expression is predominantly determined early in the pseudotime in C1, regardless of initial perturbations (**Figure 4b**, left). Furthermore, the potential landscape drove most initial states toward higher expression levels in both C1 and C4 (**Figure 4b**, right), suggesting challenges in controlling *RPS6KB1* expression post-cell fate determination.

To interpret the association of *RPS6KB1* with tamoxifen sensitivity and resistance, we partitioned C1 based on normalized expression levels < 1 or $\geq$ 1. Tamoxifen-sensitive cells in the W0-I subpopulation predominantly belonged to the *RPS6KB1*-low population, while tamoxifen-resistant and growing cells in W0-II and W0-III primarily belonged to the *RPS6KB1*-high population (**Figure 4c**). Thus, *RPS6KB1* expression serves as a potential indicator of tamoxifen-resistant states regulated by binary switching pathways.

In bifurcation analysis, we investigated the relationship between equilibrium expression states and parameters K_i and L_i, which represent inhibitory effects on positive and negative regulation, respectively. The bistable region for *RPS6KB1* was quite asymmetric with respect to K_i and L_i: Changes in K_i were more efficient than those in L_i for driving the bistable state into a low-expression state (**Figure 4d**). This suggests that positive regulation inhibitors effectively control *RPS6KB1* expression, offering potential targets for drug resistance interventions.

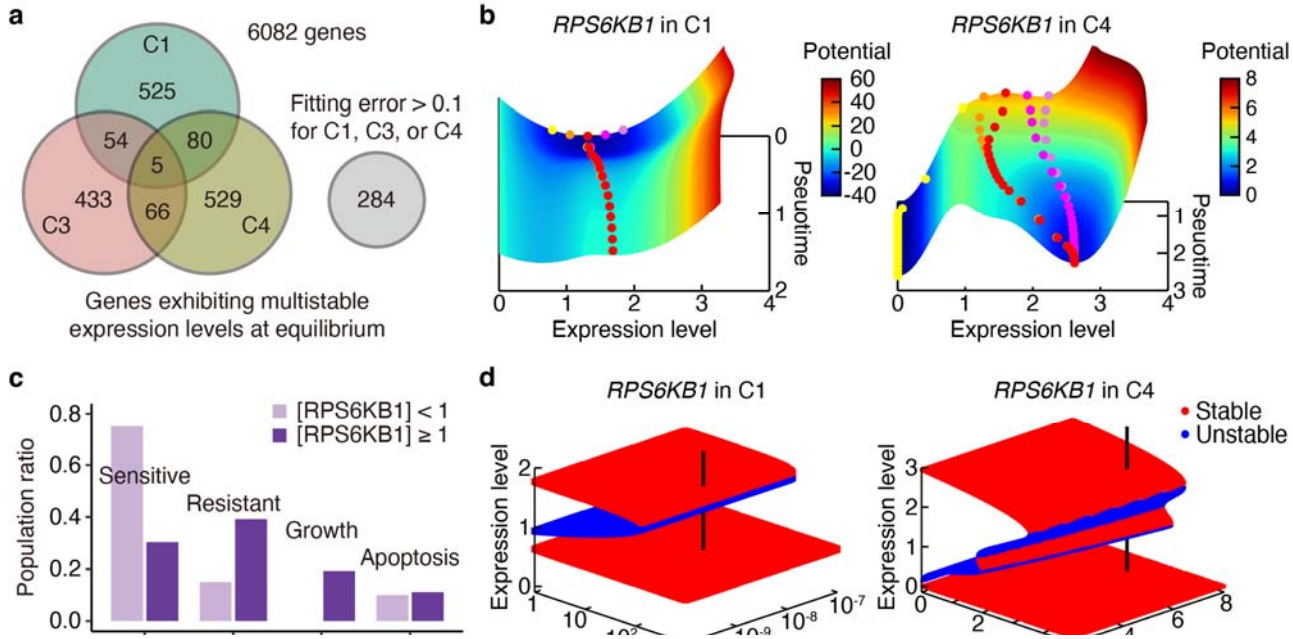

Figure 4. Transient and bifurcation analysis for "multistable genes" exhibiting multistable expression levels at equilibrium. (**a**) The number of multistable genes identified in C1, C3, and C4 subpopulations and those whose pseudotime gene expression profiles could not be well fitted by the model. (**b**) Potential RPS6KB1 expression landscape in C1 (left) and C4 (right), providing a visual representation of how gene expression states evolve over pseudotime. Colored, filled circles are snapshots of inferred expression dynamics given several initial values: Red circles correspond to the pseudotime data, while yellow, orange, magenta, and violet circles correspond to initial perturbations with 0.6-, 0.8-, 1.2-, and 1.4-fold changes, respectively. (**c**) Population ratios across W0-I (tamoxifen-sensitive cells), W0-II (tamoxifen-resistant cells), W0-III (growing cells), and W0-IV (apoptotic cells) versus RPS6KB1-low and -high populations. (**d**) Bifurcation diagrams showing stable (red) and unstable (blue) expression states dependent on parameters K_i and L_i, which represent inhibitory effects on positive and negative regulation, respectively. The vertical black lines represent the parameter set estimated from the pseudotime data.

2.5. Targeting key regulatory genes in survival and metastasis-related pathways

To comprehensively identify key regulatory genes in the C1 subpopulation, we computed bifurcation diagrams for the genes exhibiting mono- and multistable expression states involved in cAMP, PI3K-AKT-mTOR, MAPK, NOTCH, and several other pathways, with K_i and L_i serving as parameters controlling the bifurcation (**Figure 5a**). Notably, we observed significant asymmetry in the bistable regions; genes such as *RPS6KB1*, *NRAS*, *ITGB1BP1*, and *ARID1A*, K_i exhibited higher efficacy for inducing bifurcation, whereas for *TBL1XR1*, L_i exhibited higher efficacy. Conversely, for *JUND*, controlling both K_i and L_i proved efficient for inducing bifurcation. On the other hand, for *MYC*, neither K_i nor L_i induced multistability.

Additionally, we integrated the genes exhibiting multistable expression states into a schematic representation of cell survival and metastasis-related pathways, focusing on those involved in focal adhesion and EMT (**Figure 5b**), which was manually constructed based on the results of pseudotime analysis and previous reports (**Appendix G**). This comprehensive mapping provides a method for further understanding the complex interplay between signaling pathways and genes for therapeutic intervention. Furthermore, the observed peaks in *RPS6KB1*, *NFKB1*, *RHOA*, and *NOTCH2* expressions (**Figure 1g**) suggest cell survival, focal adhesion, and EMT processes in C2 (~3 weeks), C3 (~3–6 weeks), and C4 (~3-9 weeks), respectively, adding further information to the dynamic processes of drug resistance acquisition.

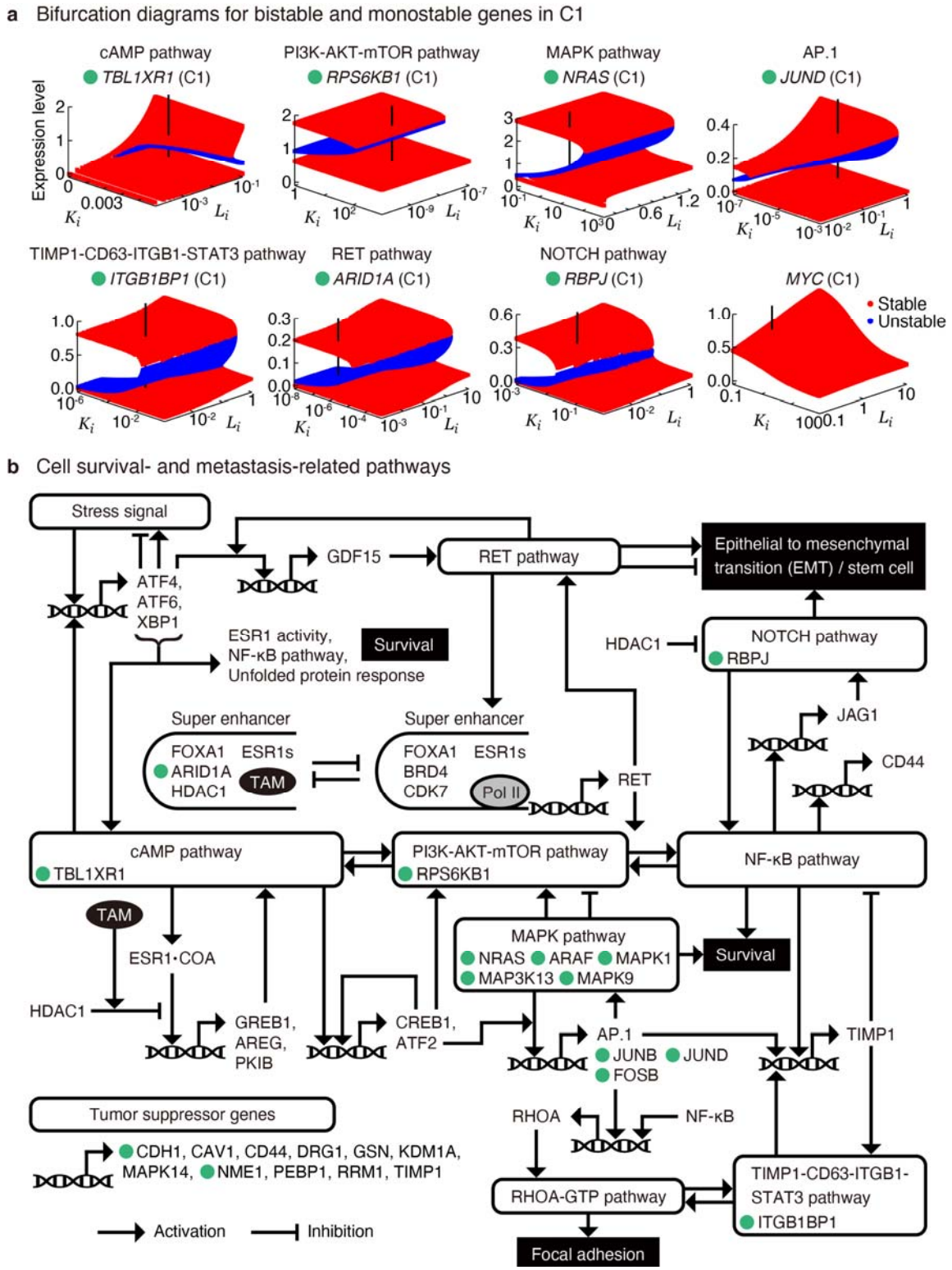

Figure 5. Integration of key regulatory genes into cell survival and metastasis-related pathways involved in focal adhesion and epithelial–mesenchymal transition (EMT). (**a**) Bifurcation diagrams for genes exhibiting bistable (*TBL1XR1*, *RPS6KB1*, *NRAS*, *JUND*, *ITGB1BP1*, *ARID1A*, and *RBPJ*) and monostable (*MYC*) expressions in C1 population that are involved in cAMP, PI3K-AKT-mTOR, MAPK, AP.1, TIMP-1-CD63-ITGB1-STAT3, RET, and NOTCH pathways. (**b**) Comprehensive mapping of the key regulatory genes, marked by green-filled circles, into a schematic representation of cell survival and metastasis-related pathways, focusing on those involved in focal adhesion and EMT. The green-labeled genes are shown in (**a**). TAM, tamoxifen.

3. Discussion

We proposed an approach that combines mathematical modeling with pseudotime analysis using single-cell time-series transcriptome data to identify key regulatory genes involved in drug resistance acquisition in breast cancer. In contrast to previous systems biology models that rely on predefined kinetic networks [42-45], our approach offers a widely applicable and computationally efficient method for modeling each gene expression. We demonstrated statistically accurate and robust parameter estimation for approximately 6,000 genes and identified 560–680 genes exhibiting multistable expression states, including *RPS6KB1*, a predictor of poor prognosis, cell survival, and growth in ER-positive breast cancers [40]. Furthermore, our analysis revealed that the long-time behavior of *RPS6KB1* expression is predominantly determined early in the pseudotime and can be effectively modulated by inhibitors for positive regulation. Nevertheless, experimental validation is important and potentially involves targeting drug design focusing on the key genes.

Despite yielding small fitting errors, suboptimal parameter sets produced divergent gene expression dynamics, indicative of the sensitivity of our model, which may involve an excessive number of parameters. Thus, future work should focus on model parameter reduction, possibly by sharing parameters among specific genes and mapping the original and reduced parameter spaces. Further model reduction will also mitigate the risk of overfitting. Furthermore, cell-to-cell heterogeneity should not be disregarded. While some genes, such as *NFKB1* and *NOTCH2*, exhibited similar pseudotime gene expression patterns (**Figure 2g**), the Pearson correlation coefficient across cells was < 0.05. This low correlation coefficient indicates considerable variability in gene expression among cells. Oral cancer cells exhibit heterogeneous responses to transforming growth factor β stimuli via multiple EMT pathways [46]. This emphasizes the critical role of heterogeneous signaling states in affecting drug response, which should be incorporated into future modeling approaches.

Finally, several limitations are worth noting. We identified potential drug-sensitive and -resistance subpopulations in the tamoxifen-untreated cells and found that genes exhibiting bistability were concentrated in the MAPK-related pathways associated with cell survival (**Figure 5b**), suggesting that cell survival is regulated by switch-like responding pathways. Additionally, a "pre-adapted" MCF-7 subpopulation exhibiting drug tolerance has been identified [9, 10]. Although these findings are useful, our analysis is currently limited to only the published scRNA-seq data [11]. This does not ensure its generalizability to other cancer datasets. Therefore, further analysis using multiple datasets such as time-series transcriptome [9, 10] and proteome data [47] is necessary to improve our data-driven model and thoroughly understand cancer drug resistance acquisition.

5. Conclusions

Our modeling approach, combined with single-cell pseudotime analysis, revealed genes with multistable expression states and elucidated their regulatory mechanisms through bifurcation analysis. Mapping the key regulatory genes into a molecular network considering cell survival and metastasis-related pathways provided a comprehensive understanding of the interplay between signaling pathways and genes, serving as a powerful tool for deciphering the complexities of drug resistance mechanisms in breast cancer. Future analyses should aim to uncover additional insights into the mechanisms of cancer drug resistance by leveraging larger datasets.

Supplementary Materials: Figure A1: Differentially expressed gene (DEG) analysis and annotation of MCF-7 cells; Figure A2: Model-based prediction of pseudotime gene expressions with the optimal and suboptimal parameter sets, exemplified by the results for *RPS6KB1* in C1, C3, and C4 subpopulations; Figure A3: Evaluation of

robustness in parameter estimation; Figure A4: Principal component analysis (PCA) of the estimated parameter values for 5,798 genes; Table S1: List of genes associated with cell survival, focal adhesion, and epithelial–mesenchymal transition-related pathways.

Author Contributions: Conceptualization, K.I.; methodology, K.I.; data analysis, K.I.; writing—original draft preparation, K.I.; writing—review and editing, K.I. and M.O.; project administration, M.O.; funding acquisition, K.I. and M.O. All authors have read and agreed to the published version of the manuscript.

Funding: K.I. was supported by JST Moonshot R&D [JPMJMS2021]. M.O. was supported by JST CREST [JPMJCR21N3], JSPS KAKENHI [18H04031] and the Uehara Memorial Foundation.

Acknowledgments: We thank Takeya Kasukawa and Masato Tsutsui for their assistance in identifying mathematical errors and improving the study design. We also thank Masahiro Inoue and Sho Soeda for improving the figures.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Removing low-quality genes and cells from single-cell RNA-sequencing data

Low-quality genes and cells were removed from all single-cell RNA-sequencing (scRNA-seq) datasets through two steps: (i) removing genes for which the number of non-zero expressing cells is less than 10; (ii) removing cells whose read counts $\geq$ 1,000, number of genes expressed with non-zero read counts $\geq$ 0, and percent reads mapped to mitochondrial genes is within 0–55. The remaining read count table contained 14,476 genes and 1,000 cells.

Appendix B. Identifying potential drug-sensitive and -resistant cells in week 0 data

The W0–W9 data were normalized using the R packages RECODE (v1.0.0) [48] followed by Seurat NormalizeData() (v4.4.0) [49], under default settings, after which only W0 data were selected. Thereafter, highly variable genes were selected using FindVariableFeatures() based on a variance-stabilizing transformation with 800 genes. Dimensionality reduction was conducted using ScaleData() and RunPCA() with features = Seurat::VariableFeatures(). The principal components that explained 50% of the total variability were used for FindNeighbors() computations. Cells were clustered into four distinct groups (W0-I, -II, -III, and -IV) by FindClusters() at a resolution of 0.5. Cell cycle phases were inferred by CellCycleScoring() with s.features = Seurat::cc.genes$s.genes and g2m.features = Seurat::cc.genes$g2m.genes. Differentially expressed genes were detected by FindAllMarkers() based on adjusted *P*-values < 0.001 (**Figure 2c** and **Figure A1a**) and were subjected to subsequent Gene Ontology enrichment analysis. We performed bitr() inputting org.Hs.eg.db (v3.15.0) using clusterProfiler (v4.4.4) [24], followed by compareCluster() with fun = "enrichGO", ont = "BP", pAdjustMethod = "BH", and pvalueCutoff = 0.001 (**Figure A1c**).

Appendix C. Defining pseudotime in weeks 0–9 data

The W0–W9 data were normalized using the R packages bayNorm (v1.14.0) [50] with Condition = [sample labels, W0, W3, W6, and W9], Prior_type = "LL", and mode_version = TRUE, followed by Seurat NormalizeData(). Afterward, highly variable genes were selected using FindVariableFeatures() based on a variance-stabilizing transformation with 5,000 genes. Dimensionality reduction was conducted using ScaleData() and RunPCA() with features = Seurat::VariableFeatures(), followed by further dimensionality reduction using DiffusionMap() in destiny package (v3.13.0) [51]. Cell cycle phases were inferred by CellCycleScoring() with s.features = Seurat::cc.genes$s.genes and g2m.features = Seurat::cc.genes$g2m.genes. Based on the first four components of diffusion map coordinates, cells were clustered into five distinct groups (C1, C2, C3, C4, and C5) using MERLoT (v0.2.2) [25] as follows: CalculateScaffoldTree() with NEndpoints = 4 and random_seed = 1, CalculateElasticTree() with N_yk = 60, and GenesSpaceEmbedding() with NCores = 3, which was subjected to subsequent pseudotime analysis. Pseudotime (t) was computed using CalculatePseudotimes() and scaled to fall within the range of 0 to 1, setting $t = 0$ and $t = 1$ at the putative start and end points of branches C1 and C4, respectively. The pseudotime distribution was significantly lower in W0-I compared with that in W0-II, which were annotated as tamoxifen-sensitive and resistant subpopulations, respectively (**Figure A1d**).

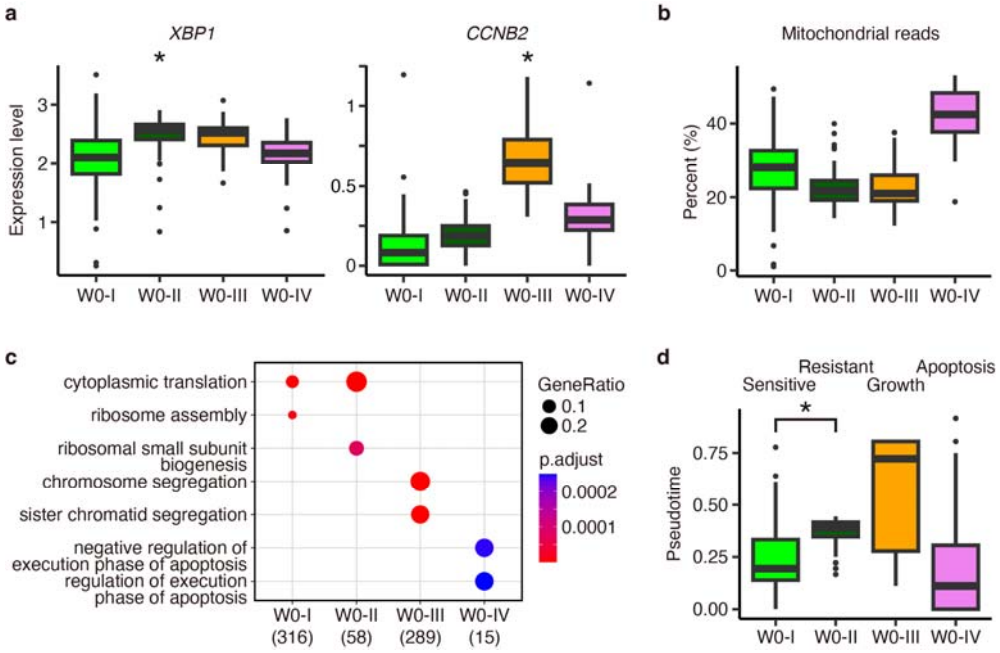

Figure A1. Differentially expressed gene (DEG) analysis and annotation of MCF-7 cells. (**a**) Box plots for W0 data showing the normalized DEG expression levels. Adjusted P-values for the clusters marked with asterisks indicate statistical significance (*$P < 0.001$, Mann–Whitney U test by Seurat). (**b**) Percentage of reads mapped to mitochondrial genes across subpopulations W0-I, -II, -III, and -IV. (**c**) Gene Ontology enrichment analysis based on the DEGs, highlighting the top two enriched terms. (**d**) Box plots for W0, showing pseudotimes across cells. An asterisk indicates statistical significance (*$P < 10^{-10}$, Mann–Whitney U test by R stats). Subpopulations W0-I, -II, -III, and -IV are manually annotated as tamoxifen-sensitive, tamoxifen-resistant, cell growth, and apoptosis, respectively, based on the DEG analysis and the interpretation of pseudotime distribution, which reflects the progression of tamoxifen resistance acquisition (see also the main text).

Appendix D. Processing pseudotime gene expression data

We excluded low-expressed genes with maximal expression levels along all pseudotime trajectories that were less than 0.1. Subsequently, 8,448 genes out of 14,476 were excluded, leaving 6,028 genes. The pseudotime ranges for C1, C2, C3, C4, and C5 were approximately [0, 0.5], [0.5, 0.61], [0.61, 0.94], [0.61, 1], and [0.5, 0.81], respectively. In this study, we neglected C2 and C5 due to insufficient pseudotime length. The pseudotime data computed by MERLoT contained small negative expression values, with a minimum value of approximately -0.024 across all genes. Since the pseudotime gene expression data will be used in our mathematical model analysis, we set the negative expression values to 10^{-4} by adding the required values.

Appendix E. Parameter estimation of the mathematical model

Let us denote the pseudotime gene expression profile of the ith gene by $x_i(t)$ and the parameter set by:

$$\boldsymbol{\theta_i} = (a_i, b_i, c_i, d_i, k_i, K_i, l_i, L_i, m_i, M_i, n_i, N_i), \quad i = 1, 2, \cdots, I,$$

where I is the number of genes. We separately estimated $\boldsymbol{\theta_i}$ for the C1, C3, and C4 subpopulations. For the parameter estimation, we set multiple initial guesses as:

$$(\alpha_i^0, b_i^0, c_i^0, d_i^0, k_i^0, K_i^0, l_i^0, L_i^0, m_i^0, M_i^0, n_i^0, N_i^0) = (1, 1, 1, \phi, \alpha, \beta, \alpha, \beta, \gamma, \delta, \gamma, \delta),$$

where $\alpha \in \{0.1, 1\}, \beta \in \{0.1, 1\}, \gamma \in \{0.1, 1\}, \delta \in \{0.1, 1\}$, and $\phi = (c_i^0 - dx_i/dt(0))/x_i(0)$. To constrain the parameter values to be positive, we rewrote

$$\boldsymbol{\theta_i} = \left(\tilde{a}_i{}^2, \tilde{b}_i{}^2, \tilde{c}_i{}^2, \tilde{d}_i{}^2, \tilde{k}_i{}^2, \tilde{K}_i{}^2, \tilde{l}_i{}^2, \tilde{L}_i{}^2, \tilde{m}_i{}^2, \tilde{M}_i{}^2, \tilde{n}_i{}^2, \tilde{N}_i{}^2\right), \quad i = 1, 2, \cdots, I,$$

and fitted the curve $(x_i, t_i, dx_i/dt)$ to the pseudotime data using Gnuplot software (v6.0) for estimating $\tilde{a}_i, \tilde{b}_i, \tilde{c}_i, \tilde{d}_i, \tilde{k}_i, \tilde{K}_i, \tilde{l}_i, \tilde{L}_i, \tilde{m}_i, \tilde{M}_i, \tilde{n}_i$, and $\tilde{N}_i$, which implements the Levenberg–Marquardt algorithm, a nonlinear least-squares method. Our parameter estimation method provided multiple candidates of the optimal parameter set for each gene, along with the fitting error and l_2-norm $\|(a_i, b_i, c_i, d_i)\|_2$. We then selected the optimal parameter sets as follows: when the absolute difference between two parameter candidates is less than 0.01, we selected the parameter set minimizing $\|(a_i, b_i, c_i, d_i)\|_2$ as the optimal one; otherwise, we selected the parameter set with the minimal fitting error. Although this restriction improved the robustness of the estimation, suboptimal parameter sets, derived from different initial guesses during local optimization, also yielded small fitting errors but resulted in different predictions (**Figure A2**).

Appendix F. Evaluation of robustness in parameter estimation

To evaluate the robustness of parameter optimization against small differences in gene expression profiles, we controlled the number of representative pseudotime points, denoted as n_{pc} (i.e., the number of "nodes" in **Figure 2d**, right), which impacts the pseudotime gene expression profiles [25]. We independently tested whether gene-wide median parameter values were consistent across different values of n_{pc} for C1, C3, and C4 using a nonparametric one-way ANOVA (**Figure 3e** and **Figure A3**). Additionally, to verify the necessity of all 12 parameters in our model, we performed PCA on the scaled gene-by-parameter matrix. This matrix comprises all estimated parameter values for C1, C3, and C4, resulting in a size of 17,394 (5,798 × 3) by 12. Our analysis confirmed that each principal component contributes almost equally (**Figure A4**).

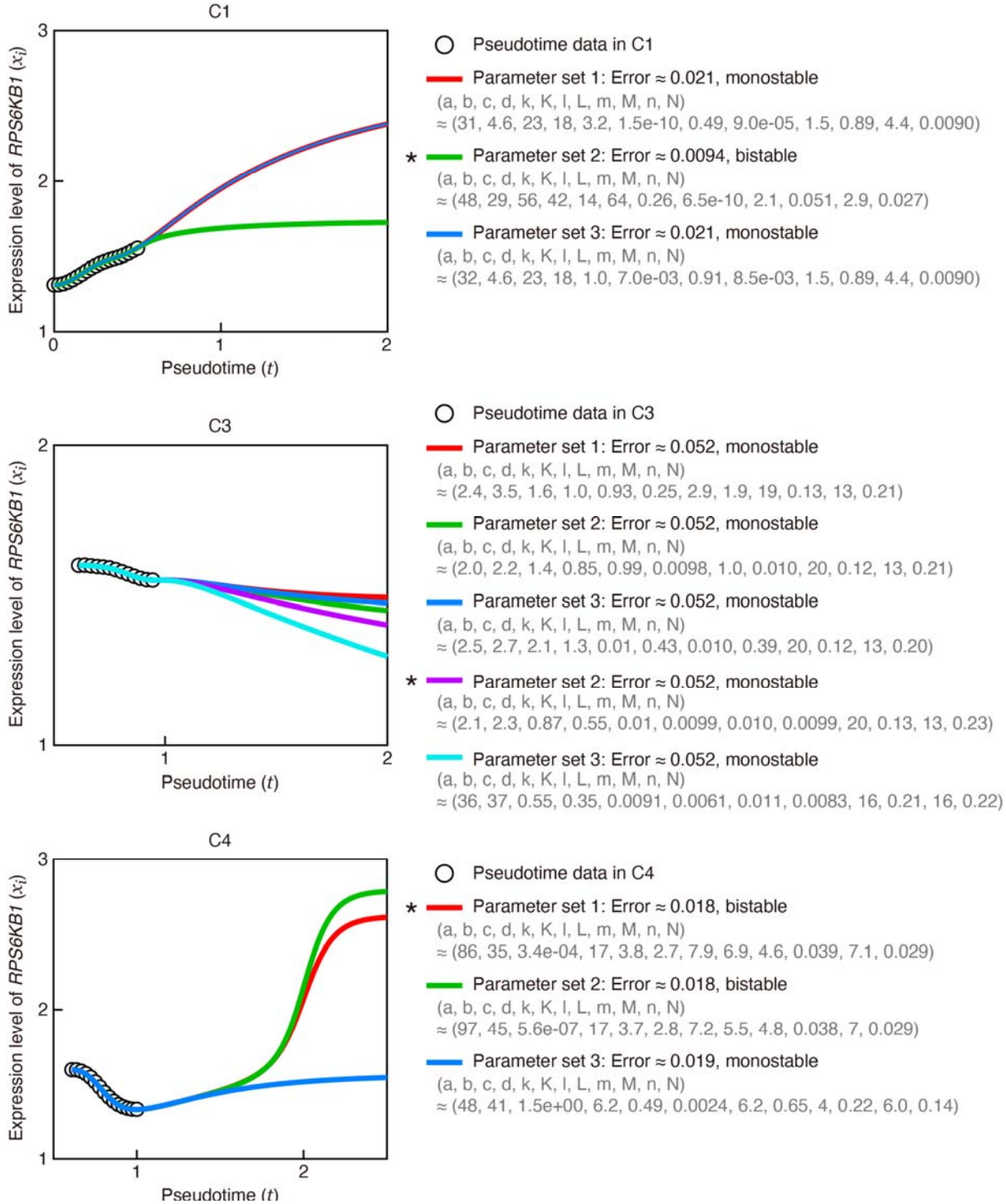

Figure A2. Model-based prediction of pseudotime gene expressions with the optimal and suboptimal parameter sets, exemplified by the results for RPS6KB1 in C1, C3, and C4 subpopulations. Trajectories computed using the optimal parameter set are marked with asterisks. Information regarding the estimated parameter values, fitting errors, and stability (indicating whether the expression levels at equilibrium are mono- or bistable) is provided on the right.

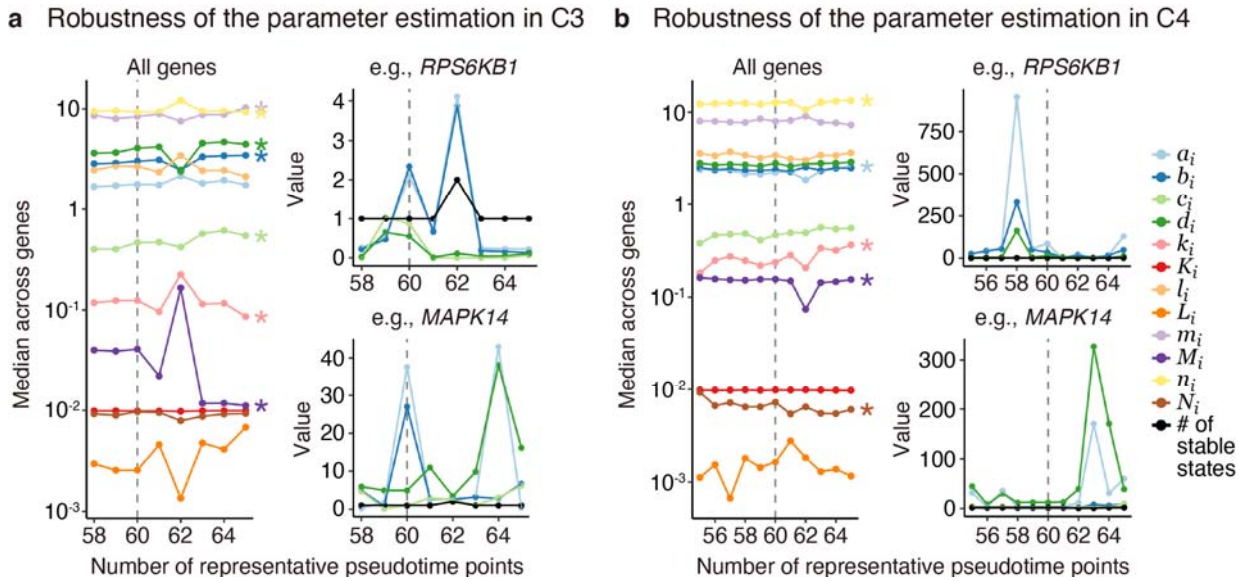

Figure A3. Evaluation of robustness in parameter estimation. (**a**) Gene-wide median parameter values for C3 (left), along with individual parameter values and the number of stable fixed points for *RPS6KB1* (top right) and *MAPK14* (bottom right) across different numbers of representative pseudotime points (n_{pc}). In this study, $n_{pc} = 60$ was used (indicated by vertical broken lines). According to the Kruskal–Wallis test, all 12 parameters showed no significant differences between $n_{pc} = 58, 59, \cdots, 60$ (P-value $> 10^{-6}$), whereas b_i, c_i, d_i, k_i, m_i, M_i, and n_i exhibited significant differences between $n_{pc} = 58, 59, \cdots, 65$: *$P$-values $< 10^{-10}$. (**b**) Similar to (**a**), all 12 parameters showed no significant differences between $n_{pc} = 55, 56, \cdots, 60$ (P-value $> 10^{-9}$), whereas a_i, k_i, M_i, n_i, and N_i exhibited significant differences between $n_{pc} = 55, 56, \cdots, 65$: *$P$-values $< 10^{-10}$.

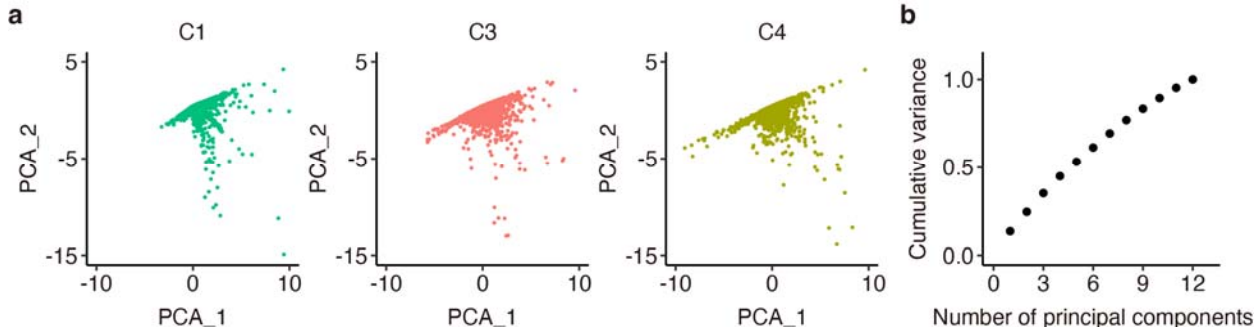

Figure A4. Principal component analysis (PCA) of the estimated parameter values for 5,798 genes. (**a**) PCA plots for C1, C3, and C4 subpopulations, with 6, 4, and 13 outlier genes, respectively, removed. (**b**) Cumulative variance across the number of principal components, demonstrating approximately equal contributions from each component.

Appendix G. Construction of a transcriptional relay network model

We manually constructed a transcription relay network model based on the result of pseudotime analysis, genes exhibiting multistable expression states, and bibliographic survey for cell survival, focal adhesion, and epithelial–mesenchymal transition (**Table S1**).

Table S1. List of genes associated with cell survival, focal adhesion, and epithelial–mesenchymal transition-related pathways.

Pathway name	Related gene	Reference
cAMP	*HDAC1*, *HDAC2*, *NCOR1*, *TBL1X*, *TBL1XR1*, *CREBBP*, *NCOR1*, *TRAM1*, *GREB1*, *AREG*, *PKIB*, *CREB1*, *ATF2*	[6, 7, 52-59]
Cellular stress	*ATF2*, *ATF4*, *ATF6*, *XBP1*, *DDIT3*, *EIF2A*	[60-62]
PI3K-AKT-mTOR	*PIK3R1*, *PTEN*, *PIK3C2A*, *AKT1*, *AKTIP*, *MTOR*, *RPS6KB1*, *RPS6*	[56, 63, 64]
MAPK	*NRAS*, *MAP3K2*, *MAP3K7*, *TAB2*, *MAP3K13*, *MAP3K20*, *RAF1*, *ARAF*, *MAP2K2*, *MAPK3*, *MAPK1*, *MAPK14*, *MAPK9*, *MAPKAP1*	[55-57]
Survival	*ESR1*, *MYC*	
AP.1	*JUN*, *JUNB*, *JUND*, *FOS*, *FOSB*	[57, 65, 66]
NF-κB	*RELA*, *NFKB1*, *NFKBIA*	[66-73]
TIMP-1-CD63-ITGB1-STAT3	*TIMP1*, *CD63*, *ITGB1*, *ITGB1BP1*, *STAT3*	[65, 67, 68, 73-75]
RHO-GTP	*RHOA*, *PTK2*, *ARHGAP26*, *ROCK1*, *ROCK2*, *TRIOBP*	[74, 75]
RET	*GDF15*, *RET*, *YY1*, *FOXA1*, *ARID1A*, *BRD4*, *CDK7*	[63, 76-83]
NOTCH	*JAG1*, *NOTCH2*, *RBPJ*	[70-72, 84, 85]

References

1. Silverberg, E., *Cancer statistics, 1984.* CA Cancer J Clin, 1984. **34**(1): p. 7-23.
2. Jemal, A., et al., *Cancer statistics, 2004.* CA Cancer J Clin, 2004. **54**(1): p. 8-29.
3. Siegel, R.L., et al., *Cancer statistics, 2023.* CA Cancer J Clin, 2023. **73**(1): p. 17-48.
4. Society, A.C., *Breast Cancer Facts & Figures 2019-2020.* Atlanta: American Cancer Society, Inc., 2019.
5. Pan, H., et al., *20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years.* N Engl J Med, 2017. **377**(19): p. 1836-1846.
6. Saatci, O., K.T. Huynh-Dam, and O. Sahin, *Endocrine resistance in breast cancer: from molecular mechanisms to therapeutic strategies.* J Mol Med (Berl), 2021. **99**(12): p. 1691-1710.
7. Ring, A. and M. Dowsett, *Mechanisms of tamoxifen resistance.* Endocr Relat Cancer, 2004. **11**(4): p. 643-58.
8. Hanker, A.B., D.R. Sudhan, and C.L. Arteaga, *Overcoming Endocrine Resistance in Breast Cancer.* Cancer Cell, 2020. **37**(4): p. 496-513.
9. Hong, S.P., et al., *Single-cell transcriptomics reveals multi-step adaptations to endocrine therapy.* Nat Commun, 2019. **10**(1): p. 3840.
10. Semina, S.E., et al., *Selective pressure of endocrine therapy activates the integrated stress response through NFkappaB signaling in a subpopulation of ER positive breast cancer cells.* Breast Cancer Res, 2022. **24**(1): p. 19.

11. Magi, S., et al., *A combination approach of pseudotime analysis and mathematical modeling for understanding drug-resistant mechanisms.* Sci Rep, 2021. **11**(1): p. 18511.

12. Huang, S., et al., *Bifurcation dynamics in lineage-commitment in bipotent progenitor cells.* Dev Biol, 2007. **305**(2): p. 695-713.

13. Ferrell, J.E., Jr., *Bistability, bifurcations, and Waddington's epigenetic landscape.* Curr Biol, 2012. **22**(11): p. R458-66.

14. Marco, E., et al., *Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape.* Proc Natl Acad Sci U S A, 2014. **111**(52): p. E5643-50.

15. Satija, R., et al., *Spatial reconstruction of single-cell gene expression data.* Nat Biotechnol, 2015. **33**(5): p. 495-502.

16. Dittmer, J., *Biological effects and regulation of IGFBP5 in breast cancer.* Front Endocrinol (Lausanne), 2022. **13**: p. 983793.

17. Qi, X., et al., *p38gamma mitogen-activated protein kinase (MAPK) confers breast cancer hormone sensitivity by switching estrogen receptor (ER) signaling from classical to nonclassical pathway via stimulating ER phosphorylation and c-Jun transcription.* J Biol Chem, 2012. **287**(18): p. 14681-91.

18. Xu, Y., et al., *Inhibition of proliferation of estrogen receptor•positive MCF•7 human breast cancer cells by tamoxifen through c•Jun transcription factors.* Mol Med Rep, 2013. **7**(4): p. 1283-7.

19. Marks, B.A., et al., *GDNF-RET signaling and EGR1 form a positive feedback loop that promotes tamoxifen resistance via cyclin D1.* BMC Cancer, 2023. **23**(1): p. 138.

20. Shi, W., et al., *Unravel the molecular mechanism of XBP1 in regulating the biology of cancer cells.* J Cancer, 2019. **10**(9): p. 2035-2046.

21. Gao, T., et al., *CCNA2 is a prognostic biomarker for ER+ breast cancer and tamoxifen resistance.* PLoS One, 2014. **9**(3): p. e91771.

22. Meng, C., et al., *Estrogen-regulated PTTG1 promotes breast cancer progression by regulating cyclin kinase expression.* Mol Med, 2020. **26**(1): p. 33.

23. Luecken, M.D. and F.J. Theis, *Current best practices in single-cell RNA-seq analysis: a tutorial.* Mol Syst Biol, 2019. **15**(6): p. e8746.

24. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* OMICS, 2012. **16**(5): p. 284-7.

25. Parra, R.G., et al., *Reconstructing complex lineage trees from scRNA-seq data using MERLoT.* Nucleic Acids Res, 2019. **47**(17): p. 8961-8974.

26. Fiers, M., et al., *Mapping gene regulatory networks from single-cell omics data.* Brief Funct Genomics, 2018. **17**(4): p. 246-254.

27. Ocone, A., et al., *Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data.* Bioinformatics, 2015. **31**(12): p. i89-96.

28. Matsumoto, H., et al., *SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation.* Bioinformatics, 2017. **33**(15): p. 2314-2321.

29. Bachmann, N., et al., *Intracellular distribution of tamoxifen in resistant human breast adenocarcinoma cells using tamoxifen-eosin association.* Cell Biol Toxicol, 1998. **14**(6): p. 429-35.

30. Becskei, A. and L. Serrano, *Engineering stability in gene networks by autoregulation.* Nature, 2000. **405**(6786): p. 590-3.

31. Jia, C., et al., *Single-cell stochastic gene expression kinetics with coupled positive-plus-negative feedback.* Phys Rev E, 2019. **100**(5-1): p. 052406.

32. Giuliano, M., et al., *Biological mechanisms and clinical implications of endocrine resistance in breast cancer.* Breast, 2011. **20 Suppl 3**: p. S42-9.

33. Citro, S., et al., *PI3K/mTOR mediate mitogen-dependent HDAC1 phosphorylation in breast cancer: a novel regulation of estrogen receptor expression.* J Mol Cell Biol, 2015. **7**(2): p. 132-42.

34. Wu, X., et al., *Beyond controlling cell size: functional analyses of S6K in tumorigenesis.* Cell Death Dis, 2022. **13**(7): p. 646.

35. in *Breast Cancer*, H.N. Mayrovitz, Editor. 2022: Brisbane (AU).

36. Villaverde, A.F., et al., *Benchmarking optimization methods for parameter estimation in large kinetic models.* Bioinformatics, 2019. **35**(5): p. 830-838.

37. Frohlich, F., et al., *Scalable Parameter Estimation for Genome-Scale Biochemical Reaction Networks.* PLoS Comput Biol, 2017. **13**(1): p. e1005331.

38. Richard Bellman, R.S.R., *The use of splines with unknown end points in the identification of systems.* Journal of Mathematical Analysis and Applications, 1971. **34**(1): p. 26-33.

39. Gugushvili, S. and C.A.J. Klaassen, √
-consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing. Bernoulli, 2012. **18**(3): p. 1061-1098.

40. Mayrovitz, H.N., H.N. Mayrovitz, and H.N. Mayrovitz, *Breast Cancer*. 2022, Brisbane, Australia: Exon Publications.

41. Waddington, C.H., *The Strategy of the Genes.* Routledge Library Editions: 20th Century Science. 2014, London ; New York: Routledge, Taylor and Francis Group.

42. Nakakuki, T., et al., *Ligand-specific c-Fos expression emerges from the spatiotemporal control of ErbB network dynamics.* Cell, 2010. **141**(5): p. 884-96.

43. Tian, T. and J. Song, *Mathematical modelling of the MAP kinase pathway using proteomic datasets.* PLoS One, 2012. **7**(8): p. e42230.

44. Imoto, H., S. Zhang, and M. Okada, *A Computational Framework for Prediction and Analysis of Cancer Signaling Dynamics from RNA Sequencing Data-Application to the ErbB Receptor Signaling Pathway*. Cancers (Basel), 2020. **12**(10).

45. Imoto, H., S. Yamashiro, and M. Okada, *A text-based computational framework for patient -specific modeling for classification of cancers.* iScience, 2022. **25**(3): p. 103944.

46. Takahashi, K., et al., *TGF-beta generates a population of cancer cells residing in G1 phase with high motility and metastatic potential via KRTAP2-3.* Cell Rep, 2022. **40**(13): p. 111411.

47. Madden, S.F., et al., *Proteomic time course of breast cancer cells highlights enhanced sensitivity to Stat3 and Src inhibitors prior to endocrine resistance development*. Cancer Gene Ther, 2023. **30**(2): p. 324-334.

48. Imoto, Y., et al., *Resolution of the curse of dimensionality in single-cell RNA sequencing data analysis.* Life Sci Alliance, 2022. **5**(12).

49. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587 e29.

50. Tang, W., et al., *bayNorm: Bayesian gene expression recovery, imputation and normalization for single-cell RNA-sequencing data.* Bioinformatics, 2020. **36**(4): p. 1174-1181.

51. Angerer, P., et al., *destiny: diffusion maps for large-scale single-cell data in R.* Bioinformatics, 2016. **32**(8): p. 1241-3.

52. Bischoff, P., et al., *Estrogens Determine Adherens Junction Organization and E-Cadherin Clustering in Breast Cancer Cells via Amphiregulin.* iScience, 2020. **23**(11): p. 101683.

53. Mohammed, H., et al., *Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor.* Cell Rep, 2013. **3**(2): p. 342-9.

54. Ahmed, M.B., et al., *cAMP Signaling in Cancer: A PKA-CREB and EPAC-Centric Approach.* Cells, 2022. **11**(13).

55. Gazon, H., et al., *Hijacking of the AP-1 Signaling Pathway during Development of ATL.* Front Microbiol, 2017. **8**: p. 2686.

56. Ye, N., et al., *Small molecule inhibitors targeting activator protein 1 (AP-1).* J Med Chem, 2014. **57**(16): p. 6930-48.

57. Eferl, R. and E.F. Wagner, *AP-1: a double-edged sword in tumorigenesis.* Nat Rev Cancer, 2003. **3**(11): p. 859-68.

58. Wang, Y.W., et al., *Understanding the CREB1-miRNA feedback loop in human malignancies.* Tumour Biol, 2016. **37**(7): p. 8487-502.

59. Zhang, H., et al., *Complex roles of cAMP-PKA-CREB signaling in cancer.* Exp Hematol Oncol, 2020. **9**(1): p. 32.

60. B'Chir, W., et al., *The eIF2alpha/ATF4 pathway is essential for stress-induced autophagy gene expression.* Nucleic Acids Res, 2013. **41**(16): p. 7683-99.

61. Wang, D., et al., *GDF15: emerging biology and therapeutic applications for obesity and cardiometabolic disease.* Nat Rev Endocrinol, 2021. **17**(10): p. 592-607.

62. Barua, D., A. Gupta, and S. Gupta, *Targeting the IRE1-XBP1 axis to overcome endocrine resistance in breast cancer: Opportunities and challenges.* Cancer Lett, 2020. **486**: p. 29-37.

63. Morandi, A., I. Plaza-Menacho, and C.M. Isacke, *RET in breast cancer: functional and therapeutic implications.* Trends Mol Med, 2011. **17**(3): p. 149-57.

64. Rozengurt, E., H.P. Soares, and J. Sinnet-Smith, *Suppression of feedback loops mediated by PI3K/mTOR induces multiple overactivation of compensatory pathways: an unintended consequence leading to drug resistance.* Mol Cancer Ther, 2014. **13**(11): p. 2477-88.

65. Kwak, H.J., et al., *Transforming growth factor-beta1 induces tissue inhibitor of metalloproteinase-1 expression via activation of extracellular signal-regulated kinase and Sp1 in human fibrosarcoma cells.* Mol Cancer Res, 2006. **4**(3): p. 209-20.

66. Kim, J.G., et al., *Regulation of RhoA GTPase and various transcription factors in the RhoA pathway.* J Cell Physiol, 2018. **233**(9): p. 6381-6392.

67. Ries, C., *Cytokine functions of TIMP-1.* Cell Mol Life Sci, 2014. **71**(4): p. 659-72.

68. Friedmann-Morvinski, D., et al., *Targeting NF-kappaB in glioblastoma: A therapeutic approach.* Sci Adv, 2016. **2**(1): p. e1501292.

69. Tannous, B.A. and C.E. Badr, *A TNF-NF-kappaB-STAT3 loop triggers resistance of glioma-stem-like cells to Smac mimetics while sensitizing to EZH2 inhibitors.* Cell Death Dis, 2019. **10**(4): p. 268.

70. Emuss, V., et al., *KSHV manipulates Notch signaling by DLL4 and JAG1 to alter cell cycle genes in lymphatic endothelia.* PLoS Pathog, 2009. **5**(10): p. e1000616.

71. Yamamoto, M., et al., *NF-kappaB non-cell-autonomously regulates cancer stem cell populations in the basal-like breast cancer subtype.* Nat Commun, 2013. **4**: p. 2299.

72. Hossain, F., et al., *Notch Signaling Regulates Mitochondrial Metabolism and NF-kappaB Activity in Triple-Negative Breast Cancer Cells via IKKalpha-Dependent Non-canonical Pathways.* Front Oncol, 2018. **8**: p. 575.

73. Dittmer, A. and J. Dittmer, *A CAF-Fueled TIMP-1/CD63/ITGB1/STAT3 Feedback Loop Promotes Migration and Growth of Breast Cancer Cells.* Cancers (Basel), 2022. **14**(20).

74. Ando, T., et al., *Tissue inhibitor of metalloproteinase-1 promotes cell proliferation through YAP/TAZ activation in cancer.* Oncogene, 2018. **37**(2): p. 263-270.

75. Justo, B.L. and M.G. Jasiulionis, *Characteristics of TIMP1, CD63, and beta1-Integrin and the Functional Impact of Their Interaction in Cancer.* Int J Mol Sci, 2021. **22**(17).

76. Wang, S.F., et al., *Growth differentiation factor 15 induces cisplatin resistance through upregulation of xCT expression and glutathione synthesis in gastric cancer.* Cancer Sci, 2023. **114**(8): p. 3301-3317.

77. Kim, M. and D.J. Kim, *GFRA1: A Novel Molecular Target for the Prevention of Osteosarcoma Chemoresistance.* Int J Mol Sci, 2018. **19**(4).

78. Nagarajan, S., et al., *ARID1A influences HDAC1/BRD4 activity, intrinsic proliferative capacity and breast cancer treatment response.* Nat Genet, 2020. **52**(2): p. 187-197.

79. Seachrist, D.D., L.J. Anstine, and R.A. Keri, *FOXA1: A Pioneer of Nuclear Receptor Action in Breast Cancer.* Cancers (Basel), 2021. **13**(20).

80. Zheng, Z.Z., et al., *Super-enhancer-controlled positive feedback loop BRD4/ERalpha-RET-ERalpha promotes ERalpha-positive breast cancer.* Nucleic Acids Res, 2022. **50**(18): p. 10230-10248.

81. Zhuang, H.H., et al., *Superenhancers as master gene regulators and novel therapeutic targets in brain tumors.* Exp Mol Med, 2023. **55**(2): p. 290-303.

82. Johansson, J., et al., *TGF-beta1-Induced Epithelial-Mesenchymal Transition Promotes Monocyte/Macrophage Properties in Breast Cancer Cells.* Front Oncol, 2015. **5**: p. 3.

83. Siaw, J.T., et al., *Loss of RET Promotes Mesenchymal Identity in Neuroblastoma Cells.* Cancers (Basel), 2021. **13**(8).

84. Zhou, B., et al., *Notch signaling pathway: architecture, disease, and therapeutics.* Signal Transduct Target Ther, 2022. **7**(1): p. 95.

85. Bai, J.W., et al., *Notch Signaling Pathway and Endocrine Resistance in Breast Cancer.* Front Pharmacol, 2020. **11**: p. 924.