Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mathematical Modeling and Transcriptome Profiling of Breast Cancer Cells During Tamoxifen Treatment Reveal Multiple Trajectories for Resistant Subtypes

Shigeyuki Magi (shigeyuki.magi@med.toho-u.ac.jp)
Toho University https://orcid.org/0000-0001-6062-5203
Sewon Ki
Masao Ukai
Atsuhiko Naito
Yutaka Suzuki
University of Tokyo
Mariko Okada

Article

Keywords: Cancer cells, Transcriptome Profiling, Resistant Subtypes, Trajectories

Posted Date: January 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-136924/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Cancer cells acquire drug resistance through following non-resistant, pre-resitant and resistant stages. Although the molecular mechanism of drug resistance is well investigated, the process of drug resistance acquisition remains largely unknown. Here, we elucidate the molecular mechanisms underlying the process of drug resistance acquisition by sequential analysis of gene expression patterns in tamoxifen-treated breast cancer cells. Single-cell RNA-sequencing of tamoxifen-treated cells revealed that tamoxifen-resistant cells can be subgrouped into two, one showing cancer stem cell-like metabolic regulation and the other showing high expression of genes encoding adhesion molecules and histone modifying-enzymes. Pseudotime analyses showed a cell transition trajectory to the two resistant subgroups that stems from shared pre-resistant state. An ordinary differential equation model based on the trajectory fitted well with the experimental results of cell growth. Based on the established model, it was predicted that inhibition of transition to both subtype would be required to prevent the appearance of tamoxifen resistance.

Introduction

Estrogen receptor (ER) is a hormone-dependent transcription factor that plays an important role in many physiological processes, including reproductive development, bone homeostasis, and cardiovascular remodeling. ER is also closely associated with breast cancer development[1,2]. Approximately 75% of all breast cancer cases are categorized into ER-positive luminal subtypes[3] and initially treated using an ER antagonist, tamoxifen (TAM). Unfortunately, around 40% of the TAM-responsive tumors progress to resistant and metastatic tumors after long-term treatment[4]. The molecular mechanisms by which those tumors exhibit TAM resistance have been well characterized and are shown to involve alterations in the estrogen–ER interaction-dependent gene expression[5], cholesterol pathway[6], and histone demethylase activity (which regulates cellular transcriptomic heterogeneity) [7] resulting in hyperactivation of alternative signaling pathways including ERK1/2[8], PI3K[9-11], and NF-κB signaling[12]. However, little is known about the process of TAM resistance acquisition, i.e., when or how the factors involved in resistance acquisition are altered, because most of the previous studies focused on the comparison between resistant and nonresistant states.

Previously, we analyzed the changes in gene expression of TAM-treated MCF-7 cell, a human ER-positive breast cancer cell line, by bulk RNA-sequencing (RNA-seq) and reported several molecular changes that preceded full acquisition of TAM resistance [15]. However, our analysis based on bulk RNA-seq might have overlooked the involvement of intracellular heterogeneity in the process of TAM resistance acquisition. In luminal breast cancer patients, single-cell transcriptome and epigenomic surveys revealed that non-genomic cell-to-cell variability generates phenotypic heterogeneity[14]. Therefore, understanding the process of TAM resistance acquisition at a single cell resolution may be important to fully understand the process of TAM resistance acquisition and develop strategies for preventing cancer recurrence.

In this study, we sequentially analyzed the transcriptomic changes in MCF-7 cells treated with TAM by single-cell RNA-seq. We report that tamoxifen-resistant cells can be subgrouped into two, one showing cancer stem cell-like metabolic regulation and the other showing high expression of genes encoding adhesion molecules and histone modifying-enzymes. Pseudotime analysis of single-cell RNA-seq data revealed a cell transition trajectory to the two resistant subpopulation that stems from shared pre-resistant state. Characteristic molecules for each subpopulation were: MYC, TATA-binding protein-associated factor 1 (TAF1), and promyelocytic leukemia (PML) for subtype 1; and BRD2, ERG, SMAD3 for subtype 2. An ordinary differential equation model based on the cell trajectory fitted well with the experimental results of cell growth and unveiled cellular processes that can be used as potential therapeutic targets.

Results

Time-series transcriptome profiles of MCF-7 cells during continuous TAM treatment

We first investigated the effect of continuous TAM treatment on human breast cancer MCF-7 cells, whose growth depends on ER signaling (**Figure 1a**). Treatment with 1 μM TAM inhibited cell growth (**Figure1b**) through decreasing the cells in S phase whereas increasing the cells in G1 phase (**Figure 1c** and **1d**), suggesting that TAM treament induced G1 arrest. The growth of cells was almost completely inhibited until week 5 but recoveered thereafter (**Figure 1b**). The cell cycle population of TAM-treated cells was also dysregulated until week 5 but became comparable to control cells after week 6 (**Figure 1c** and **1d**). These results showed the process by which cells survived and restored growth potential in the absence of ER signaling.

We next analyzed the bulk RNA-seq data of TAM-treated MCF-7 cells to identify the changes in gene regulatory network and compare with the changes in cellular phenotype. We previously performed time-course bulk RNA-seq analysis of TAM-treated MCF-7 cells and identified gene sets that play critical role at the “tipping” point of resistance acquisition [15]. In this study, we re-analyzed this dataset in order to focus on time-dependent changes in the expression of each gene, correcting for the effect of the difference in library preparation processes between samples (**Supplementary Figure 1**). A total of approximately 6,000 differentially expressed genes (DEGs) were identified between TAM-treated and control cells, of which approximately 3,000 were up-regulated (**Supplementary Figure 2**).

Gene expression in TAM-treated cells was normalized to control cells at each time point, and log2 fold-change (log2FC) values were calculated. The log2FC values of all genes at week 0 were set as a theoretical value of zero. We then obtained time-course patterns of log2FC values of 6,982 DEGs at 13 time points and analyzed by principal component analysis (PCA). The PC1 axis classified the expression profiles of all DEGs into two groups (**Figure 1f**). The separation in log2FC values was the greatest between weeks 4 and 5, which preceded the recovery of cell growth for 1 week to 2 weeks (**Supplementary Figure 3**). In addition, the separation in log2FC values between weeks 1 and 5 was greater than that between weeks 6 and 12, indicating that gene expression became more stable at the later stages.

We then investigated the relationships between gene expression patterns and gene functions. Cluster analysis of the z-score of log2FC values revealed six groups of genes with distinct expression patterns: cluster A, rapidly decreasing expression; cluster B, initially up-regulated and recovered at week 5; cluster C, rapidly increasing before cell growth rate recovery; cluster D, a gradual increase in expression concomitant with growth rate recovery; cluster E, initially up-regulated and then down-regulated; and cluster F, gradually decreasing (**Figure 1f**). The enrichment analysis of each group was carried out using the Reactome pathway (**Figure 1g**) and KEGG (**Figure 1h**) databases. Cluster A genes showed a rapidly decreasing expression pattern compared with cluster F genes. Cluster A was enriched in genes related to both receptor tyrosine kinase signaling and fatty acid metabolism. On the contrary, clusters C and D consisted of genes showing consistent up-regulation. Genes in cluster C responded early to TAM than genes in cluster D, and some of these cluster C genes were related to TGFβ signaling or extracellular matrix–receptor interaction, such as *TGFB2*, *SMAD3*, and *MMP9*, whereas others encoded collagen or laminin proteins. This implies that the reorganization of the gene network regulating epithelial–mesenchymal transition or extracellular matrix secretion, both of which contribute to cancer malignancy, occurred before cell growth recovery in the TAM-treated condition. Genes related to ribosomes were also enriched in cluster C, indicating that ribosomal biogenesis may be up-regulated during continuous TAM treatment. On the other hand, cluster D, in which gene expression level increased gradually, was enriched in genes functioning in the thyroid hormone signaling pathway, HIF1 pathway and glycolytic process, and downstream signaling of RAS, among others. These data show that a Warburg-like effect co-occurs with TAM resistance acquisition. Genes in cluster B showed a dynamic expression pattern characterized by a transient decline from week 1 to week 4, followed by recovery to the basal level. This trend was similar to the growth rate pattern observed in the TAM-treated condition. Cluster B contained numerous genes involved in cell cycle regulation, such as *CCND1* and *E2F1*, and DNA replication, such as *RAD51*. Genes in cluster E were up-regulated only when cell growth was effectively inhibited by TAM; this pattern was opposite to that observed in cluster B. Cluster E was enriched in genes involved in interferon signaling, FoxO signaling, and autophagy. The interferon and FoxO signaling pathways exhibit anti-survival functions in cancer cells exposed to anti-cancer agents[16], whereas autophagy contributes to cell survival under normal conditions[17], suggesting that genes in cluster E reflect both antitumor as well as adaptation mechanisms triggered by the TAM treatment. Cluster F was the largest dynamic cluster containing 2,088 genes and was characterized by a consistent decrease in gene expression. Enrichment analysis showed that cluster F contained genes related to the estrogen signaling pathway, suggesting that TAM treatment inhibits the estrogen-dependent gene expression mechanism, and TAM resistance observed in our experiment may be supported by an estrogen/ER-independent mechanism.

Single-cell RNA-seq analysis of MCF-7 cells under continuous TAM treatment

Because cell-to-cell heterogeneity of phenotypic features is a key mechanism of drug resistance [18], we investigated TAM-induced changes in gene expression profiles at a single-cell level. On the basis of the results of the cell growth assay and bulk RNA-seq data, we focused on four time points: week 0 (before starting TAM treatment), week 3 (beginning of complete cell growth inhibition), week 6 (end of complete

cell growth inhibition), and week 9 (at the acquisition of TAM resistance) (**Figure 2a**). RNA-seq analysis of 1,108 single cells yielded 577 high-quality single-cell gene expressions (**Figure 2a**; see the **Materials and Methods section**). The Pearson correlation coefficient of 11,413 gene expression values between individual cells decreased at week 3 and then gradually increased at weeks 6 and 9 (**Figure 2b**). This changing pattern of correlation coefficient suggests the selection of cells that can survive the TAM treatment and subsequently transition into multiple stable states.

To visualize cell-to-cell diversity in detail, we conducted uniform manifold approximation and projection (UMAP), one of the standard methods of dimensional reduction. Before drawing the UMAP plot, we calculated the probability score of cell cycle progression in each cell using Seurat 3 software[19] to correct for the bias caused by the difference in the cell cycle stage. PI staining showed the same trend in bulk cell populations (**Figure 1c**), indicating that the percentage of cells in the S phase and G2/M phase decrease only when cells cannot grow (**Figure 1b** and **2c**). All cells were mapped on a three-dimensional UMAP plot and projected in two dimensions (**Fig. 2d** and **Supplementary Figure 4**). Single-cell data were roughly divided into two groups: week 0 and the others (weeks 1–9). Cells were widely distributed in space at weeks 3 and 6 but were localized in two separate regions at week 9. These cells could be clustered into six subpopulations in the UMAP plot (**Figure 2e**). Cells in subgroups 1 and 6 belonged to the week 0 group, and these cells were almost diminished at week 3. By contrast, cells in subgroups 2 and 3 newly emerged at week 3. Finally, these cells were converted into two groups: one containing subgroup 4, and the other containing subgroup 5.

Cluster-specific gene modules and their functions

We first investigated marker genes in each subgroup. The top five genes showing the highest specificity scores were selected in each subgroup (**Supplementary Figure 5a**). Subgroups 1 and 6 were the major groups at week 0, and marker genes in these subgroups included typical ER pathway target genes such as *AREG*[20] and *GREB1*[21]. This result clearly showed that transcription activity of the ER was down-regulated in the other subgroups. On the other hand, marker genes in other subgroups, especially subgroup 2, were rather nonspecific. Interestingly, almost all marker genes of subgroups 4 and 5 also showed high expression in subgroup 3, suggesting that the pre-resistant subgroup 3 could potentially mature into distinct resistant subgroups by rewiring the genetic network.

Next, we analyzed the genetic modules specifically expressed in each subgroup or each week (**Figure 2f** and **Supplementary Figure 5b**). Subgroups 1 and 6 contained highly expressed gene modules 1, 2, and 12, which function on estrogen-dependent gene expression, unfolded protein response, and amino acid and nucleotide metabolism. On the other hand, gene expression module 5 was relatively low. Subgroups 2 and 3 were the major subpopulations in weeks 3 and 6. Both these subgroups showed high expression levels of genes in module 5, some of which are involved in interferon signaling, TGFβ signaling, and tight junctions. These enriched terms were quite similar to the early responsive cluster C in the bulk RNA-seq experiment (**Figure 1g** and **h**). Subgroup 4, whose population was increased at week 9, showed high expression levels of genes in modules 4, 6, and 7, as shown in the heatmap. These genes encoded cell

adhesion-related molecules such as integrin β4 (*ITGB4*), laminin β2 (*LAMB2*), and zyxin (*ZYX*), and some genes were involved in ROCK activation mechanisms. These modules also include several terms related to signal transduction, such as the VEGF pathway and thyroid hormone signaling. In addition, some chromatin remodeling enzymes and lysine-specific histone demethylases were included in module 6. These results indicate that TAM-resistant cells in subgroup 4 showed higher activities of cell adhesion and migration, with an altered signaling pathway and epigenetic status. Subgroup 5, which represented another major population during week 9, contained highly expressed genes in modules 8–11. This result indicates that genes related to innate immune responses, oxidative phosphorylation, and translation are highly expressed in the cells in subgroup 5. In addition, module 11 contained genes related to carbon metabolism, especially the glycolysis/glycogenesis pathway, suggesting that cells in subgroup 5 exhibit unique metabolic adaptation to TAM-induced stress. On the basis of the aforementioned results, we uncovered that TAM-resistant ER-positive breast cancer cells obtained from the same parental cell line could be divided into two types, one of which acquired the re-wired metabolic network (subgroup 5) and another acquired the migratory phenotype with high expression levels of adhesion molecules and RTK signaling-related genes (subgroup 4).

Trajectory analysis of TAM resistance

To confirm the cell transition trajectory into two different types of resistant subgroups, we conducted pseudotime analysis (**Figures 3a–c**). The pseudotime of each cell calculated from the gene expression data was correlated with the sampling time after starting the continuous TAM treatment (**Figure 3d**). The pseudotime of cells in subgroup 4 was higher than that of cells in subgroup 5, suggesting that cells in subgroup 4, showing high expression of epigenetic modulators, are more divergent from parental cells than cells in subgroup 5 (**Figure 3e**). To estimate important molecules involved in the emergence of subgroups 4 and 5, we analyzed DEGs along with an estimated cell trajectory. A total of 273 and 79 genes were detected as highly expressed genes in subgroups 4 and 5, respectively (**Figure 3f** and **g**). Then, we analyzed the upstream regulators of these genes using the ChIP-Atlas database[22] (**Figure 3h** and **3i**). Upstream analysis of the trajectory to subgroup 4 (**Figure 3f** and **3h**) showed that only 4 of the top 10 factors represented ChIP-seq data obtained from MCF-7 cells, and most of the others were obtained from the triple-negative breast cancer (TNBC) cell line (**Figure 3h**, shown in green). These results also suggest that most of upregulated genes in subgroup 4 are controlled by bromodomain-containing proteins, BRD4 and BRD2, which recognize acetylated histones and act as super enhancers [23,24]. In addition, our results also suggest the possible involvement of oncogenic transcription factors SMAD3 and ERG in the trajectory to subgroup 4. Genes encoding these molecules were up-regulated before the cells acquired TAM resistance (**Figure 3j**, top and middle), as shown in bulk RNA-seq data (**Figure 1**). This indicates that subgroup 4 genes have different epigenetic status, which is clearly distinct from that of parent MCF-7 cells; this result was consistent with the enrichment analysis of specific gene modules (**Figure 2f**).

In subgroup 5, 7 of the top 10 factors were obtained from ER-positive breast cancer or normal cells (**Figure 3i**), suggesting that subgroup 5 retained the transcriptional network of parental MCF-7 cells. ChIP-

seq analysis using anti-WDR5 antibody showed the best q-value and fold enrichment score. WDR5 has been associated with histone methylation[25]. Although *WDR5* was not up-regulated in bulk RNA-seq data (**Figure 3j**, bottom), other histone methylases might regulate genes related to subgroup 5. Other up-regulated candidates estimated from ChIP-Atlas database included *MYC*, *TAF1*, *PML*, and BRD4. Among these genes, *MYC*, *TAF1*, and *PML* were up-regulated before week 4, as shown by bulk RNA-seq analysis (**Figure 3h**, bottom). These data indicate that MYC, TAF1, and PML1 may contribute to one of the emerging TAM-resistant subpopulations. Taken together, our analysis revealed key molecular candidates that drive two different TAM-resistant subgroups.

Mathematical modeling of the TAM resistance acquisition process

Finally, we constructed a mathematical model that reproduces the process of acquiring TAM resistance, based on cell trajectories obtained by pseudotime analysis, to estimate the core processes that contribute to the growth and differentiation of resistant cell populations (**Figure 4a**). This model consists of four major cell subpopulations: cells initially sensitive to TAM (X_S, clusters 1 and 6 in **Figure 3e**), pre-resistant cells (X_P, clusters 2 and 3), resistant cells showing high expression of carbon metabolism-related genes (X_{R1}, cluster 5), and resistant cells with highly adhesive and invasive phenotypes (X_{R2}, cluster 4). We hypothesized that TAM induces the death of X_S and X_P cells but not that of resistant cell populations and sigmoidally promotes cell state transition from X_S to X_P and from X_P to X_{R1} or X_{R2} in response to a period of TAM treatment. This hypothesis assumes that cell state transition is caused by the accumulation of genetic or epigenetic changes induced by continuous TAM treatment. We obtained 20 model parameter sets, which could reproduce two experimental time-course datasets, total cell growth rate in the presence of TAM (**Figure 1b**) and rate of cell subpopulation (**Figure 3e**), simultaneously (**Fig. 4b and c and Supplementary Figure 6a**) using the BioMASS computational framework[26].

Comparing each parameter set, we found two remarkable features of the well-fitting parameter sets. First, the growth rate of subpopulation X_{R2} (rate constant of parameter v_{12}) was significantly greater than that of X_{R1} (v_9) (**Supplementary Figure 6b**). This finding is consistent with the result that the subpopulation of X_{R2} expressed some cell division-related genes (**Figure 2f**). Second, a parameter determining the steepness of sigmoidal change in v_{10} during the time course was greater than the steepness of sigmoidal change in v_7 (**Supplementary Figure 6c**). This implies that the cell transition speed from X_P to X_{R2} showed switch-like alteration triggered by the accumulation of epigenetic alterations. This finding is substantiated by the results of single-cell RNA-seq analysis, which showed that the genetic feature of X_{R2} displayed high expression levels of chromatin-modifying enzymes (**Figure 2f**), and pseudotime analysis, in which X_{R2} was the most differentiated subtype compared with other subtypes (**Figure 3e**). These results indicate the potential of a constructed model and obtained parameter sets in reproducing not only the growth rate and ratio of cell populations during the time course but also the biological features of subpopulations.

On the basis of the model and parameter sets, we then performed sensitivity analysis of each single reaction to examine whether a change in each single reaction affects the size of the resistant cell

population. We focused on the integral growth rate after week 3, which determines the total increase in cell number after the acquisition of TAM resistance, and calculated the sensitivities of each reaction on that (**Figure 4d**). The results indicated that v_{12}, growth velocity of X_{R2}, was the most critical phenomenon for the increase in resistant cell numbers. On the contrary, neither the reverse transition from resistant cell types to X_P (or from X_P to X_S) nor cell death caused by TAM could be a negative regulator of the increase in resistant cell numbers alone. Thus, we showed that the increase in TAM-resistant cells is strongly controlled by the growth rate of X_{R2}. Finally, we examined the synergistic inhibitory effect of two biological events in the model, cell growth and forward state transition to two different subtypes, on TAM-resistant cell growth (**Figure 4e**). Combined inhibition of cell growth rate of both resistant subtypes caused synergistic regression (growth rate <1) at broad inhibitory ranges than the other conditions (**Figure 4e**, bottom right). However, combination inhibition including at least one transition to the resistant subpopulation induced potent regression when the cell transition was completely reduced. This result indicates that inhibition of cell state transition by, for example, epigenetic inhibitors, has the potential to be more effective than targeting the growth of the resistant subpopulation itself.

Discussion

In this study, we analyzed transcriptional changes in MCF-7 cells during continuous TAM treatment using both bulk and single-cell RNA-seq. The results of bulk RNA-seq analysis revealed several time-course patterns of gene expression during the continuous TAM treatment. A subset of genes, including clusters B and E, showed low or high expression immediately before acquiring the growing ability in the presence of TAM, respectively. It is reasonable to speculate that the recovery of gene expression levels in cluster B is accompanied by the recovery of growth rate because these genes included positive cell cycle regulators. The expression levels of these genes may be regulated by E2F families, suggesting that the growth of TAM-resistant cells also depends on the CDK4–E2F cell cycle machinery, supporting the effect of the CDK4/6 inhibitor on ER cells[27]. Combined with the expression pattern of ESR1, our results implied that the expression levels of E2F gene families are maintained by estrogen–ESR1-dependent signaling in the absence of TAM; however, this was superseded by other signaling pathways, such as central carbon metabolism-related HIF1 machinery, in TAM-resistant cells (**Figure 1h**).

On the other hand, the significance in cluster E is rather difficult to interpret. Some groups have previously reported that interferon regulatory factor-1 (IRF1) is critical for TAM-mediated apoptosis, and its related pathway is also up-regulated in TAM-treated cells[28]. IRF1 was shown to induce apoptosis in breast cancer cells [29]. However, another group showed that interferon-responsive genes are up-regulated in both TAM-resistant and radioresistant MCF-7 cells and contribute to the cross resistant[30]. These previous reports imply that the bilateral function of interferon signaling may accelerate the adaptation of cancer cells to the TAM-treated condition by increasing cell-to-cell variability (**Figure 2b**). Non-genetic cell-to-cell variability, believed as a major contributor to the production of outlier cells, which can adapt to severe conditions[31,32], could play an important role in the acquisition of TAM resistance under our experimental conditions because few genes are mutated at the time when genes in cluster E were up-regulated[15]

(**Figure 1f**). Previous studies show contradictory results on the relationship between chemosensitivity and FoxO-autophagy signaling. It has been reported that 4-hydroxytamoxifen induces autophagic cell death[33,34]; however, another group reported that inhibition of autophagy restored the responsiveness for anti-estrogen therapy. In our single-cell RNA-seq analysis, cells in subgroup 4, which showed high expression levels of autophagy-related genes, were broadly distributed on the cell landscape (**Supplementary Figure 2b**). Our results suggest the possibility that the modulation of autophagy and interferon signaling in the early phase of endocrine therapy prevents the transition of cells to resistant types.

Several studies showed that TAM is localized to mitochondria and endoplasmic reticulum and shows non-genomic toxicity by inhibiting the electron transport chain complexes[35,36]. Some results in our transcriptomic analysis can be explained by such estrogen-independent mode of action of TAM. The overrepresentation of genes related to translation in cluster C and that of genes involved in the detoxification of reactive oxygen species, which are mainly produced in mitochondria, in cluster D (**Figure 1g** and **1h**) can be interpreted as a protective response for the dysfunction of these organelles. High expression levels of ribosomal and mitochondrial genes (modules 8 and 10) were also detected in TAM-resistant subgroup 5 by single-cell RNA-seq analysis (**Figure 2f**). Another phenomenon related to mitochondrial dysfunction was the up-regulation of glycolytic pathway enzymes induced by the HIF1 signaling pathway (**Figure 1h**), which was coincident with the growth ability of cells in the presence of TAM (**Figure 1b**). We detected the overexpression of genes encoding glycolytic and gluconeogenetic enzymes in subgroups 3 and 5. Indeed, MYC, which drives a gene expression of ribosomal proteins[37], hexokinase 2, and lactate dehydrogenase[38] was increased during the time course, and was predicted as one of the main regulators of resistant subgroup 5 (**Figure 3i** and **j**). These results suggest that subgroup 5 genes overcome the non-genomic toxicity of TAM by up-regulating the ribosomal and mitochondrial functions via HIF1 or MYC activity.

Our single-cell RNA-seq analysis revealed the existence of two different resistant subpopulations and the role of important molecules in the emergence of each resistant subpopulation. Subgroup 5 was predicted to be initiated by the activity of TAF1 and PML1 molecules, in addition to MYC (**Figure 3i**). Interestingly, the second bromodomain-specific inhibitor of TAF1 represses *MYC* expression, and its effect is synergistic to the BRD4 inhibitor[39]. On the basis of the results of this study and previous studies, we infer that the differentiation of pre-resistant cells to resistant subgroup 5 requires TAF1 and BRD4 activity for up-regulating *MYC* gene expression. PML is believed to possess tumor-suppressing activity by controlling the cell cycle and apoptosis[40]; however, recent studies revealed that PML is overexpressed and promotes metastasis, especially in TNBC[41,42]. Although the overexpression of PML in luminal types is uncommon, silencing PML functions elicits not only growth suppression in TNBC[43] but also oncosphere formation, a readout of self-renewal potential, in PML-overexpressing luminal type breast cancers[44]. In addition, PML overexpression in MCF-10A cells promotes fatty acid oxidation and ATP production via the tricarboxylic acid cycle[42]. Taken together, these results suggest that TAM-resistant cells in subgroup 5 are similar to proliferative cancer stem cells, which exhibit self-renewal potential and rely on both oxidative

phosphorylation and glycolytic metabolism [45]. These results also suggest that the pharmacological PML inhibitor, arsenic trioxide, could prevent the transformation of pre-resistant cells into resistant cells. Another resistant subgroup (subgroup 4) showed a mesenchymal phenotype, with high expression levels of prostate cancer-related genes and chromatin-modifying enzymes. Some of the TAM-resistant specimens showed the overexpression of androgen receptor (AR), and exogenously AR-overexpressed MCF-7 cells resistant to TAM-induced growth inhibition[46]. In addition, one of the histone demethylases, KDM5B, was reported to modulate resistance to endocrine therapies by increasing transcriptional heterogeneity [7]. These studies support our predictions that one of resistant subpopulation acquired new TAM-resistant features by androgen receptor signaling and histone-modifying enzymes. These results also raise the possibility that inhibitors of AR and KDM5 prevent cell transformation to this subgroup.

In summary, our time-series single-cell sampling and multidimensional data analysis highlighted that the acquisition of drug resistance relies on heterogeneity and emphasized the importance of multiple molecules in phenotype transitions. We propose that dual inhibition of key molecules involved in each cell transition trajectory can be an effective strategy for preventing endocrine therapy resistance (**Figure 4e**). To the best of our knowledge, this is the first study to reproduce the characteristics of more than one TAM-resistant cell populations by creating a mathematical model of cell transitions based on cell trajectories obtained by single-cell analysis. Although several issues such as the financial cost of treatment need to be overcome, the combination of single-cell RNA-seq analysis of clinical cancer samples with the mathematical modeling approach can contribute to the design of a personalized treatment strategy for each patient in the future.

Materials And Methods

Cell culture

Human breast adenocarcinoma MCF-7 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and antibiotics, as described previously[15].

Cell growth assay

Approximately 1 x 10^6 MCF-7 cells were seeded in a 100-mm dish containing 10 mL of culture medium supplemented with or without 1 μM TAM. After a week, cells were detached and collected with trypsinization, and the concentration of the cell suspension was measured using a hematocytometer. The cell growth rate per week was calculated by dividing 1 x 10^6 with the total number of cells in each cell suspension.

Cell cycle analysis by flow cytometry

MCF-7 cells were trypsinized, washed with phosphate-buffered saline (PBS), and fixed with 80% ethanol. Subsequently, the fixed cells were washed with PBS and stained with PI staining solution (BD bioscience, CA, U.S.A.) for 15 min. The PI-stained cells were subjected to flow cytometry using the FACSCanto II Flow

Cytometer (BD bioscience), and the number of cells at each cell cycle stage was analyzed using the FlowJo 7.6.5 software.

Bulk RNA-seq analysis of TAM-resistant cells

Bulk RNA-seq data of MCF-7 cells have been published previously[15]. Briefly, RNA was extracted from MCF-7 cells treated with or without 1 μM TAM using QIAshredder (QIAGEN, Netherlands) and RNeasy Mini Kit (QIAGEN), which was then used for RNA-seq analysis. Different sequencing methods were used, which resulted in either 100-bp paired-end reads or 36-bp single-end reads (**Supplementary Figure 1**). To remove the influence of different sequencing methods, we used only the first 36 bp of the first single-end read of the paired-end data. After removing adaptor sequences and checking the sequence quality using Trim Galore[47], the reads were aligned to the human reference genome (version GRCh38), and the read number was counted by featureCounts[48] without multi-mapping and multi-overlapping. The expression level of each gene was quantified as transcripts per million (TPM). TPM data of each sample were used for PCA to analyze the variability and reproducibility of the data. Differential gene expression analyses were performed using DESeq2[22]. A total of 6,982 genes, whose expression levels were altered at least three time points, were selected as DEGs.

Enrichment analyses

Pathway enrichment analysis was carried out using the Targetmine platform. Redundant enrichment terms, shared by >70% of the genes of interest, were removed from the results, and the term with the lowest q-value was retained. The enrichment analysis of upstream transcriptional regulators was performed using the ChIP-Atlas database (https://chip-atlas.org) [19].

Single-cell RNA-seq analysis of TAM-resistant cells

Single cells were separated using the ICELL-8 system (Takara Bio, Shiga, Japan). MCF-7 cells treated with or without continuous TAM were trypsinized and collected following dilution with the culture medium. The cells were then washed twice with cold PBS and stained with Hechest33342 (5 μg/mL) and PI (1 μg/mL) for 15 min. After staining, the cells were diluted to a concentration of 20,000 cells/mL and loaded into the ICELL-8 single-cell system. Then, cDNA was prepared using 3'DE reagents (Takara Bio), according to the manufacturer's instructions, and subjected to 100-bp paired-end sequencing on the Illumina HiSeq 3000 platform (Illumina, CA, U.S.A.). Mapping of sequence reads to the human reference genome sequence and counting genes were carried out using the mappa and hanta software (Takara Bio). The gene count data of individual cells were cleaned using the Seurat 3 software[49]. A series of quality controls were implemented. First, any gene expressed in <5 cells at <5 counts per million was removed. Second, cells with <1,500 detected genes and >25% mitochondrial genes were filtered out. After filtering, the count data matrix consisting of 11,413 genes and 186, 189, 118, and 84 cells at weeks 0, 3, 6, and 9, respectively, was obtained. Next, any bias due to differences in the cell cycle stage was removed using the function *CellCycleScoring* and cell cycle gene set, and the effect of cell cycle phases on gene

expression data was regressed. The data were imported into the Monocle 3 software[26], and the data dimensions were reduced to three with UMAP. Then, cells were categorized into multiple classes. Gene module and pseudotime analyses were carried out using the Monocle 3 software, according to the developer's instructions (https://cole-trapnell-lab.github.io/monocle3/). The pseudotime of each cell was calculated on the basis of the relative distance from open circle #1 (set as pseudotime = 0) (**Figure 3c**).

Mathematical simulation

The process of TAM resistance acquisition was determined on the basis of the results of pseudotime analysis (**Figure 4a**). The mathematical model comprised 12 ordinary differential equations with 19 parameters. Details of the equations are summarized in **Supplementary Table 1**. Mathematical simulation and parameter search were performed using the BioMASS platform[26]. During the parameter search process, we attempted to minimize the weighted sum of squared percentage errors (*wSSPE*):

$$wSSPE = \sum_{i=1}^{n} \left(\frac{x_{sim,i} - x_{exp,i}}{x_{exp,i} + 1} \right)^2$$

where *wSSPE* is an objective function; *n* is the number of obtained data points to be fitted such as growth rate and rate of subpopulation at each time point and treatment; $x_{sim,i}$ and $x_{exp,i}$ are the i^{th} simulation and experimental data, respectively. Importantly, “weighted” SSPE (calculated by adding 1 to the denominator of objective function) was used instead of normal SSPE to achieve two purposes simultaneously: escaping division by zero and fitting the simulation results to two experimental datasets with different range limits.

Sensitivity analysis

The single parameter sensitivity of each reaction is defined as follows:

$$s_i(q(\boldsymbol{v}), v_i) = \frac{\partial \ln (q(\boldsymbol{v}))}{\partial \ln (v_i)} = \frac{\partial q(\boldsymbol{v})}{\partial v_i} \cdot \frac{v_i}{q(\boldsymbol{v})}$$

where v_i is the i^{th} reaction; $\boldsymbol{v}$ is a reaction vector ($\boldsymbol{v} = v_1, v_2, \ldots$); and $q(\boldsymbol{v})$ is a target function (e.g., the integral of growth rate in the current study). The sensitivity of each reaction was calculated with 1% increase in the reaction rate using the BioMASS platform[26].

Author contribution

All authors declare no competing interests.

References

1. Aranda, A. & Pascual, A. Nuclear hormone receptors and gene expression. *Physiol. Rev.*81, 1269–1304 (2001).
2. Burns, K. A. & Korach, K. S. Estrogen receptors and human disease: an update. *Arch. Toxicol.*86, 1491–1504 (2012).
3. Allred, D. C., Brown, P. & Medina, D. The origins of estrogen receptor alpha-positive and estrogen receptor alpha-negative human breast cancer. *Breast Cancer Res.*6, 240–245 (2004).
4. Ring, A. & Dowsett, M. Mechanisms of tamoxifen resistance. *Endocr. Relat. Cancer*11, 643–658 (2004).
5. Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M. & Schiff, R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nat. Rev. Clin. Oncol.*12, 573–583 (2015).
6. Hultsch, S. *et al.* Association of tamoxifen resistance and lipid reprogramming in breast cancer. *BMC Cancer*18, 850–14 (2018).
7. Hinohara, K. *et al.* KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance. *Cancer Cell*35, 330–332 (2019).
8. Gee, J. M., Robertson, J. F., Ellis, I. O. & Nicholson, R. I. Phosphorylation of ERK1/2 mitogen-activated protein kinase is associated with poor response to anti-hormonal therapy and decreased patient survival in clinical breast cancer. *Int. J. Cancer*95, 247–254 (2001).
9. Kirkegaard, T. *et al.* AKT activation predicts outcome in breast cancer patients treated with tamoxifen. *J. Pathol.*207, 139–146 (2005).
10. Shoman, N. *et al.* Reduced PTEN expression predicts relapse in patients with breast carcinoma treated by tamoxifen. *Mod. Pathol.*18, 250–259 (2005).
11. Tokunaga, E. *et al.* Activation of PI3K/Akt signaling and hormone resistance in breast cancer. *Breast Cancer*13, 137–144 (2006).
12. Zhou, Y., Eppenberger-Castori, S., Eppenberger, U. & Benz, C. C. The NFkappaB pathway and endocrine-resistant breast cancer. *Endocr. Relat. Cancer*12 Suppl 1, S37–46 (2005).
13. Liu, J., Dang, H. & Wang, X. W. The significance of intertumor and intratumor heterogeneity in liver cancer. *Exp. Mol. Med.*50, e416–e416 (2018).
14. Patten, D. K. *et al.* Enhancer mapping uncovers phenotypic heterogeneity and evolution in patients with luminal breast cancer. *Nat. Med.*24, 1469–1480 (2018).
15. Liu, R. *et al.* Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers. *J Mol Cell Biol*31, 166–664 (2018).
16. Lindner, D. J., Kolla, V., Kalvakolanu, D. V. & Borden, E. C. Tamoxifen enhances interferon-regulated gene expression in breast cancer cells. *Mol. Cell. Biochem.*167, 169–177 (1997).
17. Doherty, J. & Baehrecke, E. H. Life, death and autophagy. *Nat. Cell Biol.* 1–8 (2018). doi:10.1038/s41556-018-0201-5
18. Huang, S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metastasis Rev.*32, 423–448 (2013).

19. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell*177, 1888–1902.e21 (2019).
20. Ciarloni, L., Mallepell, S. & Brisken, C. Amphiregulin is an essential mediator of estrogen receptor alpha function in mammary gland development. *Proc. Natl. Acad. Sci. USA*104, 5455–5460 (2007).
21. Rae, J. M. *et al.* GREB 1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Res. Treat.*92, 141–149 (2005).
22. Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.*19, (2018).
23. Verdin, E. & Ott, M. *50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond.Nature reviews. Molecular cell biology*16, 258–264 (Nature Publishing Group, 2015).
24. Cheung, K. L. *et al.* Distinct Roles of Brd2 and Brd4 in Potentiating the Transcriptional Program for Th17 Cell Differentiation. *Mol. Cell*65, 1068–1080.e5 (2017).
25. Couture, J.-F., Collazo, E. & Trievel, R. C. Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat. Struct. Mol. Biol.*13, 698–703 (2006).
26. Imoto, H., Zhang, S. & Okada, M. A Computational Framework for Prediction and Analysis of Cancer Signaling Dynamics from RNA Sequencing Data-Application to the ErbB Receptor Signaling Pathway. *Cancers (Basel)*12, 2878 (2020).
27. Finn, R. S. *et al.* PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro. *Breast Cancer Res.*11, R77–13 (2009).
28. Fang, Q., Yao, S., Luo, G. & Zhang, X. Identification of differentially expressed genes in human breast cancer cells induced by 4-hydroxyltamoxifen and elucidation of their pathophysiological relevance and mechanisms. *Oncotarget*9, 2475–2501 (2018).
29. Stang, M. T. *et al.* Interferon regulatory factor-1-induced apoptosis mediated by a ligand-independent fas-associated death domain pathway in breast cancer cells. *Oncogene*26, 6420–6430 (2007).
30. Post, A. E. M. *et al.* Interferon-Stimulated Genes Are Involved in Cross-resistance to Radiotherapy in Tamoxifen-Resistant Breast Cancer. *Clin. Cancer Res.*24, 3397–3408 (2018).
31. Brock, A., Chang, H. & Huang, S. Non-genetic heterogeneity–a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.*10, 336–342 (2009).
32. Pisco, A. O. & Huang, S. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: 'What does not kill me strengthens me'. *Br. J. Cancer* (2015). doi:10.1038/bjc.2015.146
33. Bursch, W. *et al.* Active cell death induced by the anti-estrogens tamoxifen and ICI 164 384 in human mammary carcinoma cells (MCF-7) in culture: the role of autophagy. *Carcinogenesis*17, 1595–1607 (1996).
34. Kohli, L. *et al.* 4-Hydroxytamoxifen induces autophagic death through K-Ras degradation. *Cancer Res.*73, 4395–4405 (2013).
35. Theodossiou, T. A., Yannakopoulou, K., Aggelidou, C. & Hothersall, J. S. Tamoxifen subcellular localization; observation of cell-specific cytotoxicity enhancement by inhibition of mitochondrial ETC

complexes I and III. *Photochem. Photobiol.*88, 1016–1022 (2012).

36. Theodossiou, T. A., Wälchli, S., Olsen, C. E., Skarpen, E. & Berg, K. Deciphering the Nongenomic, Mitochondrial Toxicity of Tamoxifens As Determined by Cell Metabolism and Redox Activity. *ACS Chem. Biol.*11, 251–262 (2016).
37. Golomb, L., Volarevic, S. & Oren, M. p53 and ribosome biogenesis stress: the essentials. *FEBS Lett.*588, 2571–2579 (2014).
38. Li, B. & Simon, M. C. Molecular Pathways: Targeting MYC-induced metabolic reprogramming and oncogenic stress in cancer. *Clin. Cancer Res.*19, 5835–5841 (2013).
39. Sdelci, S. *et al.* Mapping the chemical chromatin reactivation landscape identifies BRD4-TAF1 cross-talk. *Nat. Chem. Biol.*12, 504–510 (2016).
40. Mazza, M. & Pelicci, P. G. Is PML a Tumor Suppressor? *Front. Oncol.*3, 174 (2013).
41. Ponente, M. *et al.* PML promotes metastasis of triple-negative breast cancer through transcriptional regulation of HIF1A target genes. *JCI Insight*2, e87380 (2017).
42. Carracedo, A. *et al.* A metabolic prosurvival role for PML in breast cancer. *J. Clin. Invest.*122, 3088–3100 (2012).
43. Arreal, L. *et al.* Targeting PML in triple negative breast cancer elicits growth suppression and senescence. *Cell Death Differ.*98, 10869–14 (2019).
44. Martín-Martín, N. *et al.* Stratification and therapeutic potential of PML in metastatic breast cancer. *Nat. Commun.*7, 12595–13 (2016).
45. Chae, Y. C. & Kim, J. H. Cancer stem cell metabolism: target for cancer therapy. *BMB Rep*51, 319–326 (2018).
46. De Amicis, F. *et al.* Androgen receptor overexpression induces tamoxifen resistance in human breast cancer cells. *Breast Cancer Res. Treat.*121, 1–11 (2010).
47. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*30, 923–930 (2014).
48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*15, 550 (2014).
49. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature*566, 496–502 (2019).

Figures

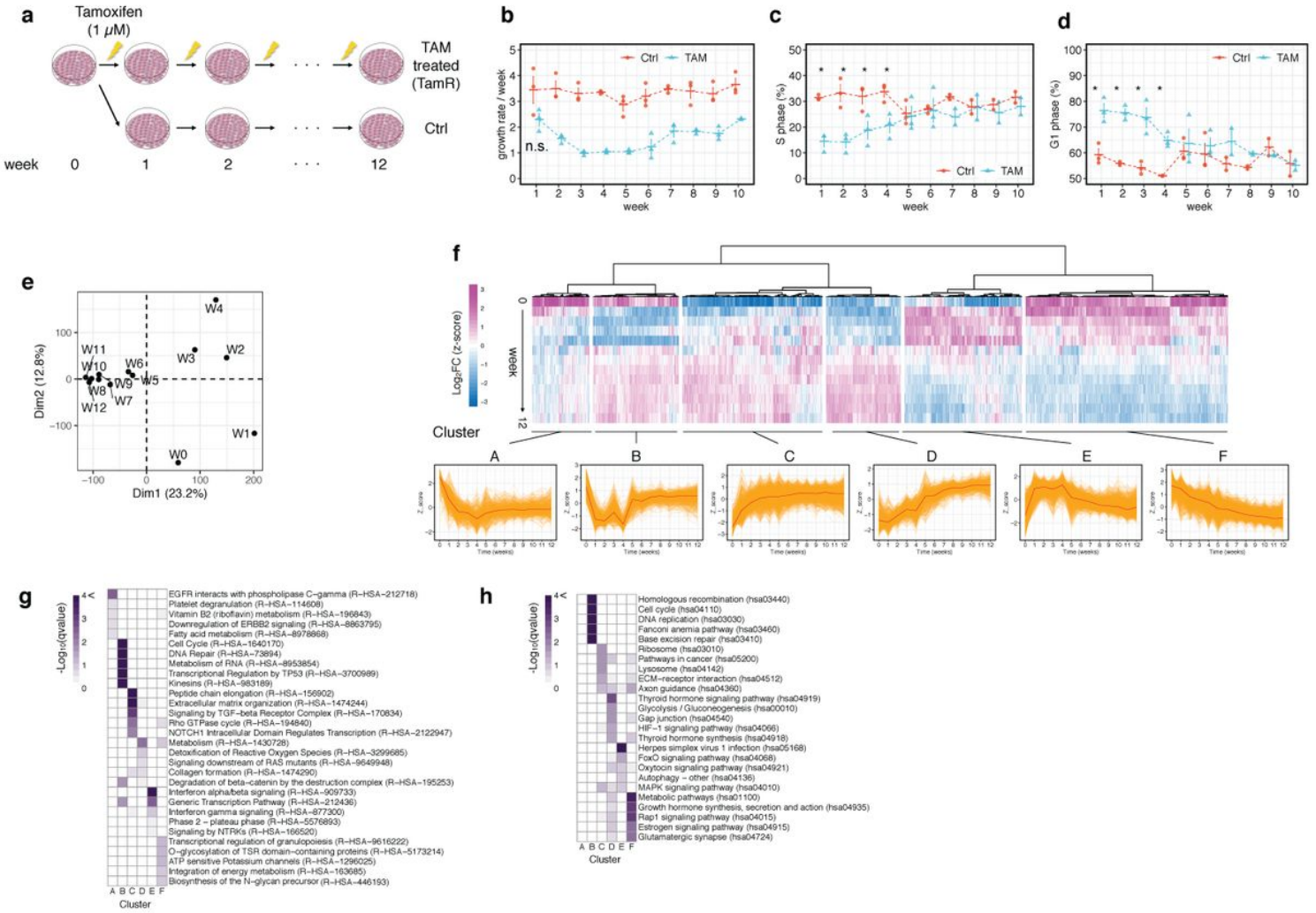

Figure 1

Figure 1

Functional analysis of gene expression patterns in human breast adenocarcinoma MCF-7 cells during the tamoxifen (TAM) resistance acquisition process. (a) Schematic overview of the experimental procedure. Each treatment was replicated five times. (b) Growth rate of TAM-treated and control (Ctrl) cells. Data represent mean ± standard error (SE). Asterisks indicate significant differences between Ctrl and TAM-treated cells ($p < 0.05$; two-tailed Welch's test). (c, d) Ratio of G1 (c) and S (d) phase in TAM-treated and Ctrl cells. Data represent mean ± SE. Asterisks indicate significant differences between Ctrl and TAM-

treated cells ($p < 0.05$; two-tailed Welch's test). (e) Principal component analysis (PCA) of log2 fold-change (log2FC) values. (f) Cluster analysis of gene expression patterns during the 9-week time course. The z-scores of log2FC in gene expression at each time point were classified into six clusters. The bottom line graphs showed individual (orange) or median (red) of gene expression patterns. (g, h) Heatmap of enrichment analysis data. Top five significant terms in the Reactome pathway database (g) and KEGG pathway database (h) in each cluster are presented.

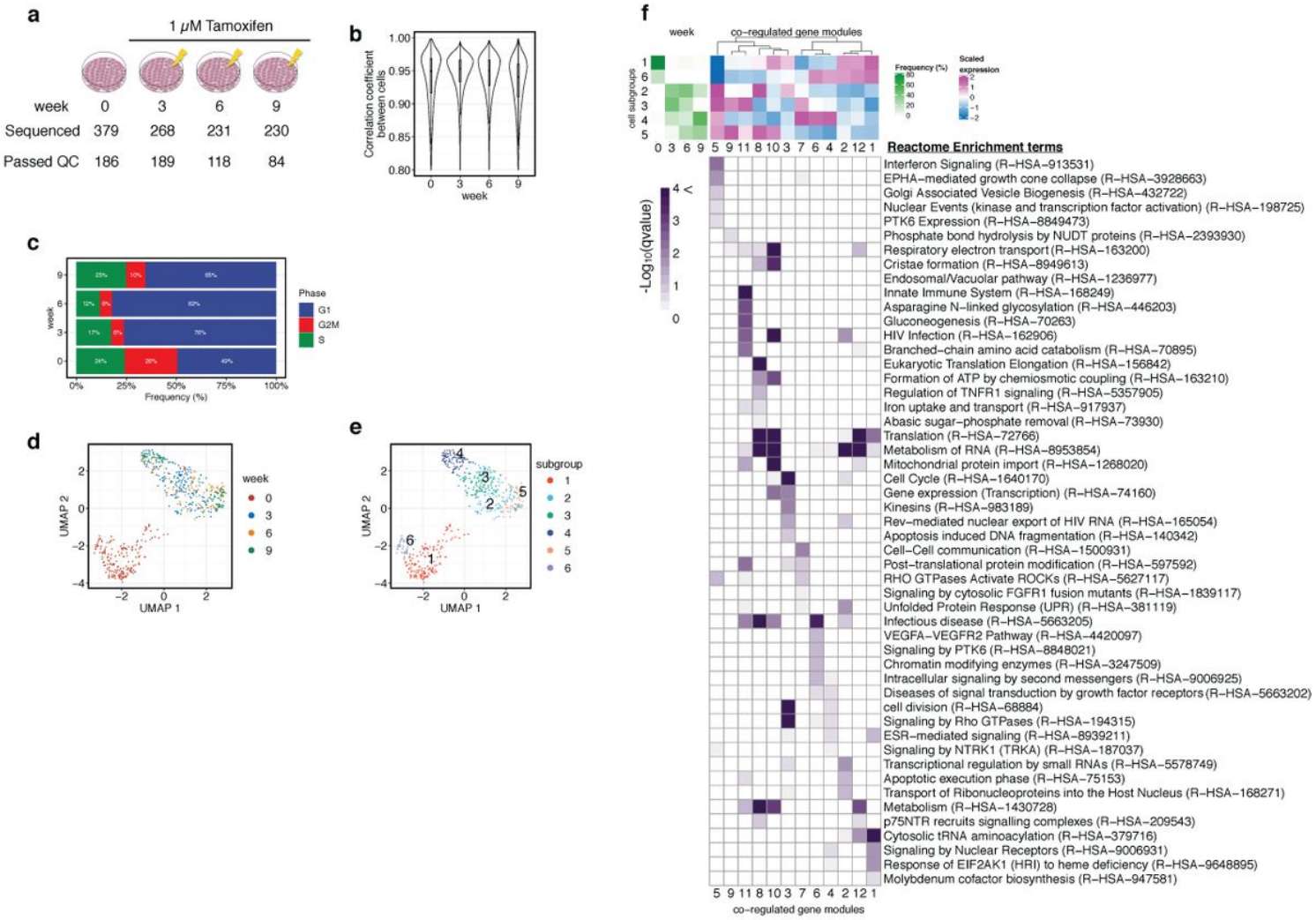

Figure 2

Figure 2

Single-cell RNA-seq analysis of TAM-resistant MCF-7 cells. (a) Schematic overview of the experimental procedure of single-cell RNA-seq. (b) Violin plot showing the distribution of the correlation coefficients of single-cell gene counting among cells at each time point. (c) Percentage of cells at different cell cycle stages at each time point. (d, e) Visualization of single-cell transcriptome data by UMAP. Single-cell data space was reduced to three dimensions, and the distribution of data was visualized using the first two dimensions. Cells were colored by week. estimated cell cycle stage in (d) and subgroups in (e). (f) Top, Clustering of single cell data presented in (d). Bottom, percentages of subgroups in each week. (g) Complex heatmap of cell subpopulation, co-regulated gene modules, and enriched functions. Top left heatmap presents the frequency of subgroups at each time point shown in (f). Top right heatmap presents the relative expression level of each gene module in each cluster. Bottom heatmap presents the enrichment terms in each gene module.

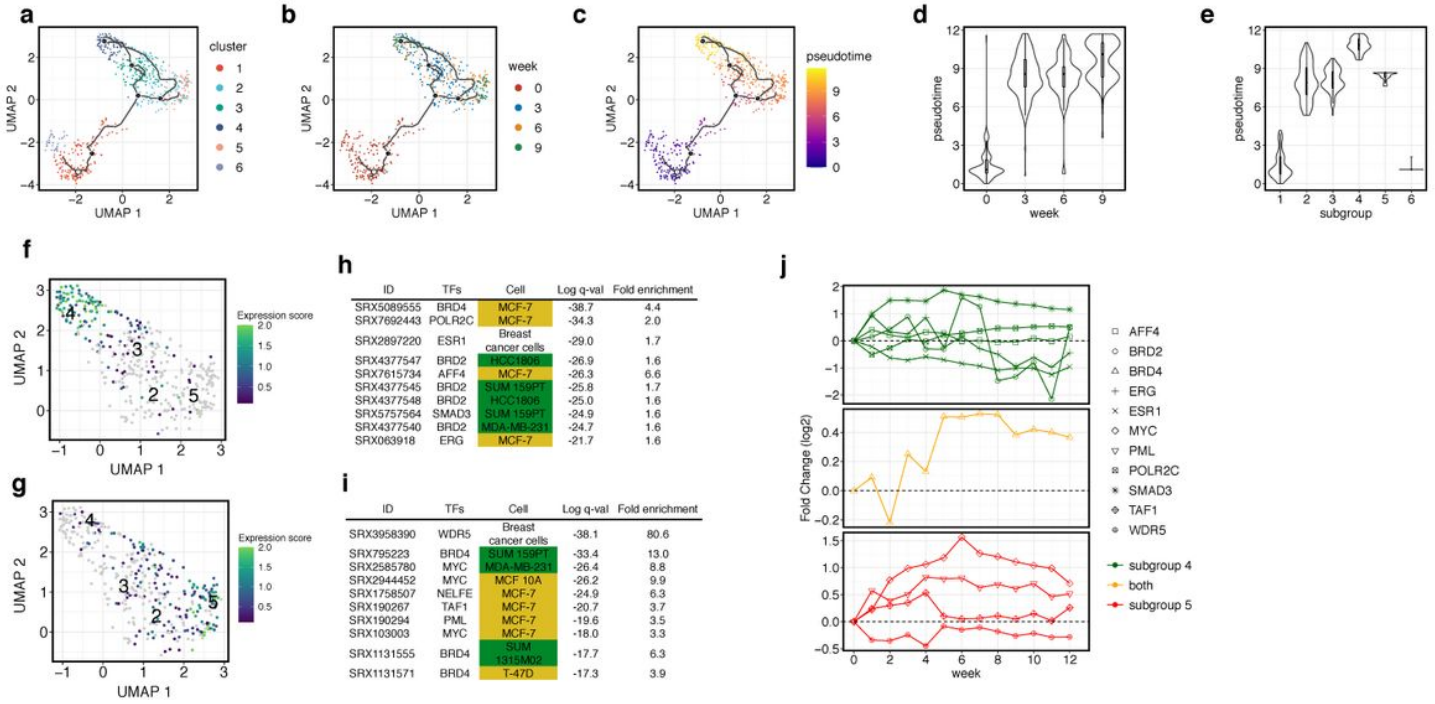

Figure 3

Figure 3

Trajectory analysis of the TAM resistance acquisition process. (a–c) Single-cell trajectory during the continuous TAM treatment. Graphs were colored by clusters (a), weeks (b), and pseudotime (c). (d, e) Violin plots showing the distribution of pseudotime in each week (d) and subgroup (e). (f, g) Expression score of up-regulated genes in subgroup 4 (f) and subgroup 5 (g), considering the cell trajectory. (h, i) Upstream factor analysis of up-regulated genes displayed in (f) and (g). Top 10 Q-value data are

presented. Color represents subtypes of breast cancers: yellow, MCF-7 cells; green, triple-negative breast cancer (TNBC). (j) Time-series bulk gene expression patterns listed in (h) and (i).

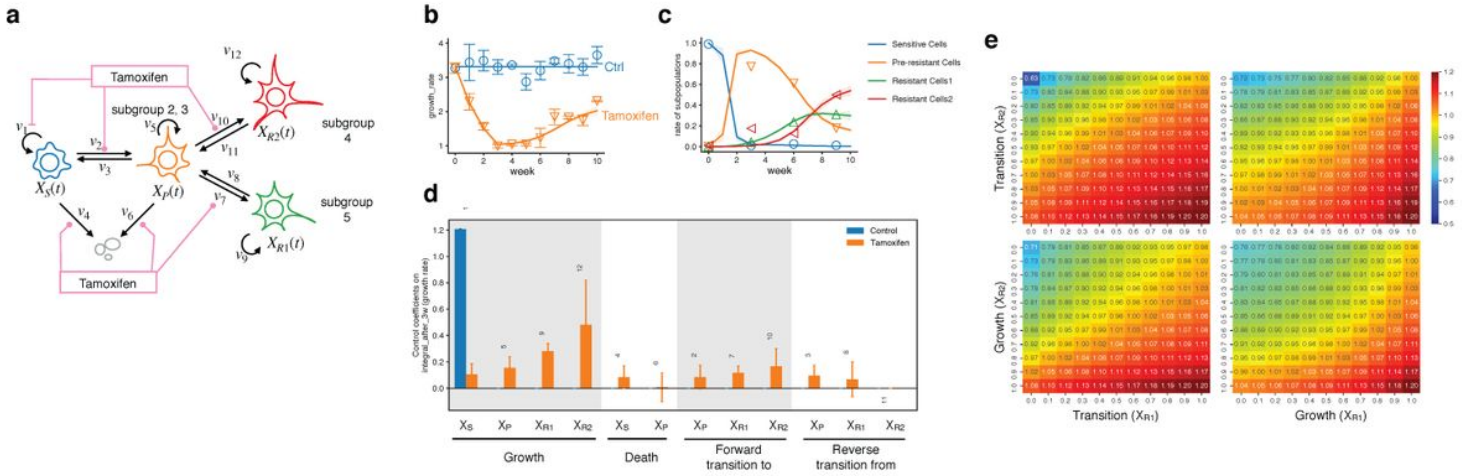

Figure 4

Figure 4

Ordinary differential equation-based model of the TAM resistance acquisition process. (a) Illustration of the model scheme. (b, c) Time-series analysis of changes in cell growth rate (b) and ratio of each subpopulation (c). Points: experimental data; error bars: standard deviation (SD) of experimental data; solid lines: averaged in silico simulation of 20 sets of parameters; shaded areas: SD of simulations. (d)

Results of sensitivity analysis of the integral of growth rate from the third week at each reaction; v1 ~ v12. Error bars represent the SD of simulations with 20 set parameters. (e) Heatmap of the mean growth rate of TAM-treated cells in week 3, with different inhibitory intensity of growth rate of XR1 or XR2 and cell transition to XR1 or XR2. The X- and Y-axes indicate the remaining reaction rate (i.e., 1.0 means 0% inhibition, and 0.0 means 100% inhibition). Data represent the mean of simulations with 20 set parameters.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS11109.xlsx
- 201221FigS.pdf