Supporting Information for

Cell cycle plasticity underlies fractional resistance to palbociclib in ER+/HER2− breast tumor cells

Tarek M. Zikry[a,b,1], Samuel C. Wolff[a,c,1], Jolene S. Ranek[a,c], Harris M. Davis[a,c], Ander Naugle[a,c], Namit Luthra[a,c], Austin A. Whitman[d], Katarzyna M. Kedziora[e], Wayne Stallaert[f], Michael R. Kosorok[b], Philip M. Spanheimer[d,g,2], Jeremy E. Purvis[a,c,d,2]

[a] Computational Medicine Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA.

[b] Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, 27599, USA.

[c] Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA.

[d] Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA.

[e] Center for Biologic Imaging, Department of Cell Biology, University of Pittsburg, Pittsburgh, PA, 15620, USA.

[f] Department of Computational and Systems Biology, University of Pittsburg, Pittsburgh, PA. 15620, USA.

[g] Department of Surgery, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA.

[1] These authors contributed equally to this work

2 Corresponding authors:

Philip Spanheimer, MD
Division of Surgical Oncology, CB# 7123
UNC Department of Surgery
Chapel Hill, NC 27599-7213
philip_spanheimer@med.unc.edu

Jeremy Purvis, PhD
Mary Ellen Jones Building 11018C, CB#7488
116 Manning Drive
Chapel Hill, NC 27599-7488
jeremy_purvis@med.unc.edu

Primary Antibody	Species	Vendor/Cat#	Dilution
CDK2	Goat	R&D Systems/AF4654	1:100
CDK4	Rabbit	Abcam/ ab108355	1:400
CDK6	Rabbit	Abcam/ab124821	1:250
Cdt1	Rabbit	Cell Signaling Technology (CST)/8064	1:200
Cyclin A	Mouse	Santa Cruz/sc-271682	1:50
Cyclin B1	Goat	R&D Systems/AF6000	1:100
Cyclin D1	Mouse	Santa Cruz/sc-20044	1:100
Cyclin E	Mouse	Santa Cruz/sc-247	1:50
E2F1	Mouse	Santa Cruz/sc-251	1:100
Estrogen Receptor (ER)	Rabbit	Abcam/ab32063	1:200
Ki67	Rabbit	Abcam/ab15580	1:800
p16	Rabbit	Abcam/ab108349	1:400
p19 (CDKN2D)	Rabbit	Invitrogen/PA5-83665	1:200
p21	Goat	R&D Systems/AF1047	1:200
PanCK	Mouse	CST/4545	1:200
Phospho-Retinoblastoma (pRB)	Rabbit	CST/8516	1:1000
Progesterone Receptor (PR)	Mouse	Thermo Fisher/MA5-12658	1:400
Retinoblastoma (RB)	Mouse	CST/9309	1:500

Table S1. List of primary antibodies utilized in the 4i rounds of imaging along with host species, vendor information and dilution utilized in the experiment.

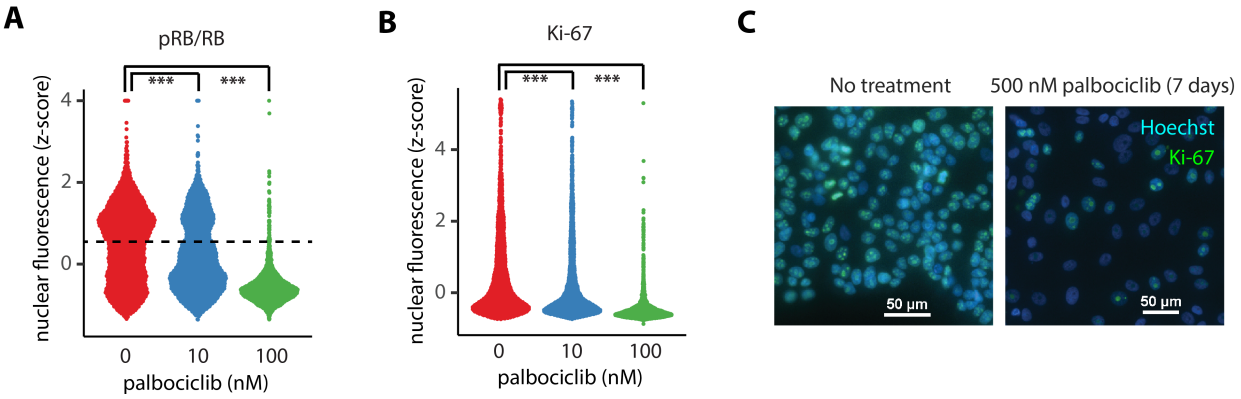
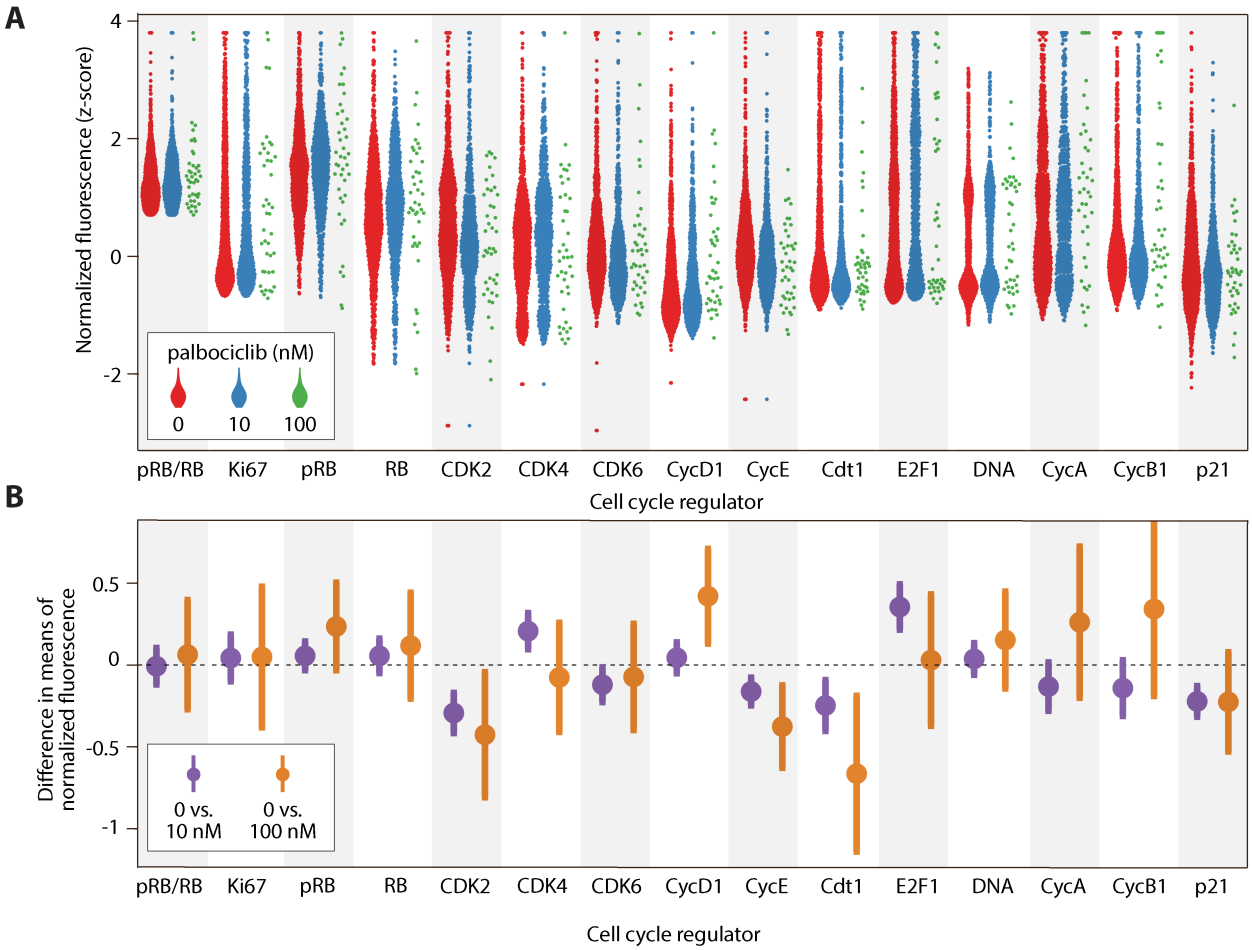


Figure S1. Fractional resistance in a biological replicate and under long-term treatment with palbociclib. A. Distribution of pRB/RB in T47D cells under 0, 10, or 100 nM palbociclib. The dotted line marks the statistically determined threshold for demarcating cells as proliferating despite palbociclib treatment. Cells above the dotted line were used for characterizing proliferative cells. **B.** Distribution of nuclear Ki-67 levels at 0, 10, or 100 nM palbociclib. *** indicates a *P*-value < 0.001 using a two-sided Kolmogorov-Smirnov test between untreated and treated cells. **C.** The ER+/HER2- cell line MCF7 was treated with 500 nM palbociclib for 7 days. Despite a reduction in growth rate, a significant fraction of cells (~12%) retained pRB/RB and KI-67 expression.



(Figure S2 continued on next page)

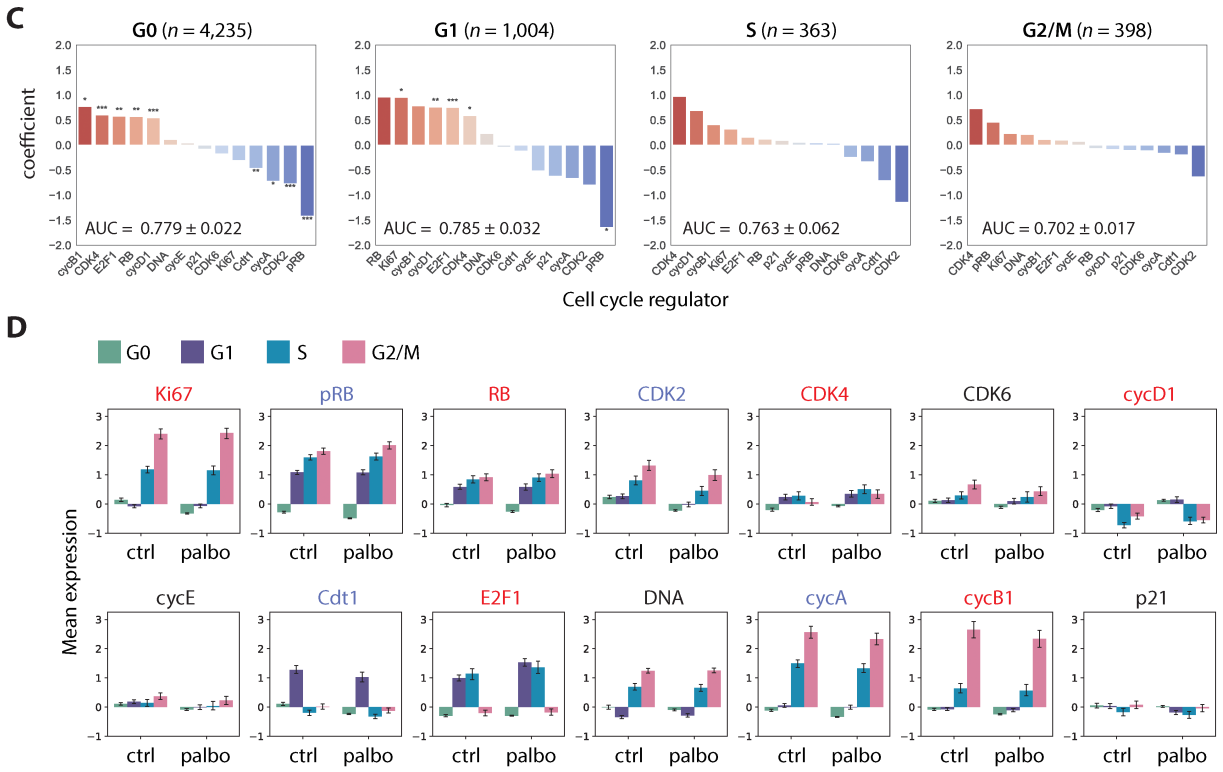
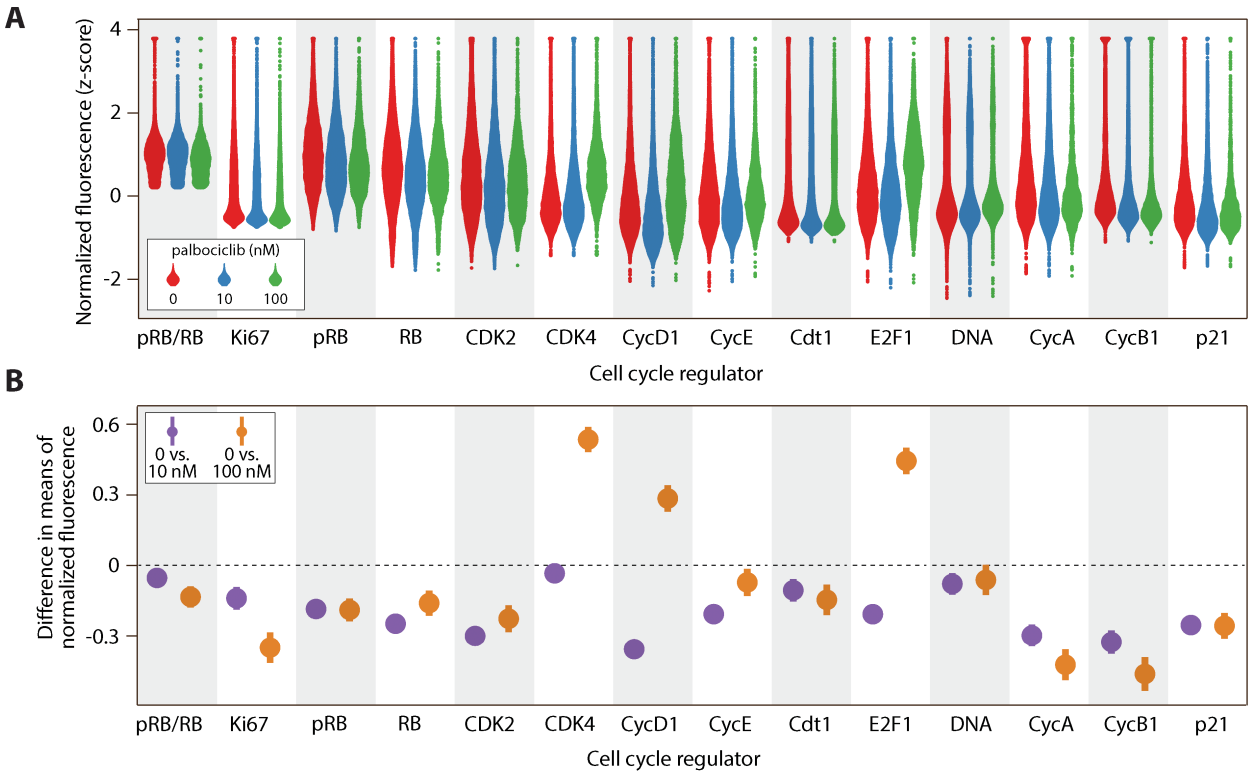


Figure S2. Shifts in expression of cell cycle proteins among fractionally resistant tumor cells for T47D biological replicate. A. Beginning with a downsampled dataset containing 2,000 cells from each condition (see **MATERIALS AND METHODS**), we identified the proliferating cells using the pRB/RB threshold defined in **Figure S1A**. Proliferating single-cell distributions of cell cycle regulators at 0 nM, 10 nM, or 100 nM palbociclib, showing an expected reduction in the number of cells at higher drug concentrations. **B.** 95% confidence intervals (CI) of proliferating cells for the differences in mean expression (normalized z-scores) between either untreated cells and 10 nM palbociclib (*purple*); or untreated and 100 nM palbociclib (*orange*). Confidence intervals overlapping with the dashed line at 0 indicate a lack of statistical significance. Confidence intervals are wider when comparing 0 vs. 100 nM due to lower sample sizes at the highest dose of palbociclib. **C.** Logistic regression on all 6,000 cells predicting the odds that a given cell is either untreated (0 nM) or treated (10 nM or 100 nM) based on expression of its cell cycle regulators. Cell-to-cell increases in regulators shown in red (e.g., CDK4), or decreases in regulators shown in blue (e.g., CDK2), increase the odds of association with treated (10 nM or 100 nM) versus untreated cells. A separate regression was performed for each phase (G0, G1, S, G2/M), where the last three are considered proliferating (high pRB/RB). This analysis was performed on all cells, including non-proliferating (G0) cells. Significance: *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$. **D.** Z-score normalized expression levels of cell cycle features, stratified by cell cycle phase for all 6,000 cells. Bar height is the mean expression for untreated (ctrl) or palbociclib-treated (10 nM and 100 nM) cells. Error bars represent confidence intervals.



(Figure S3 continued on next page)

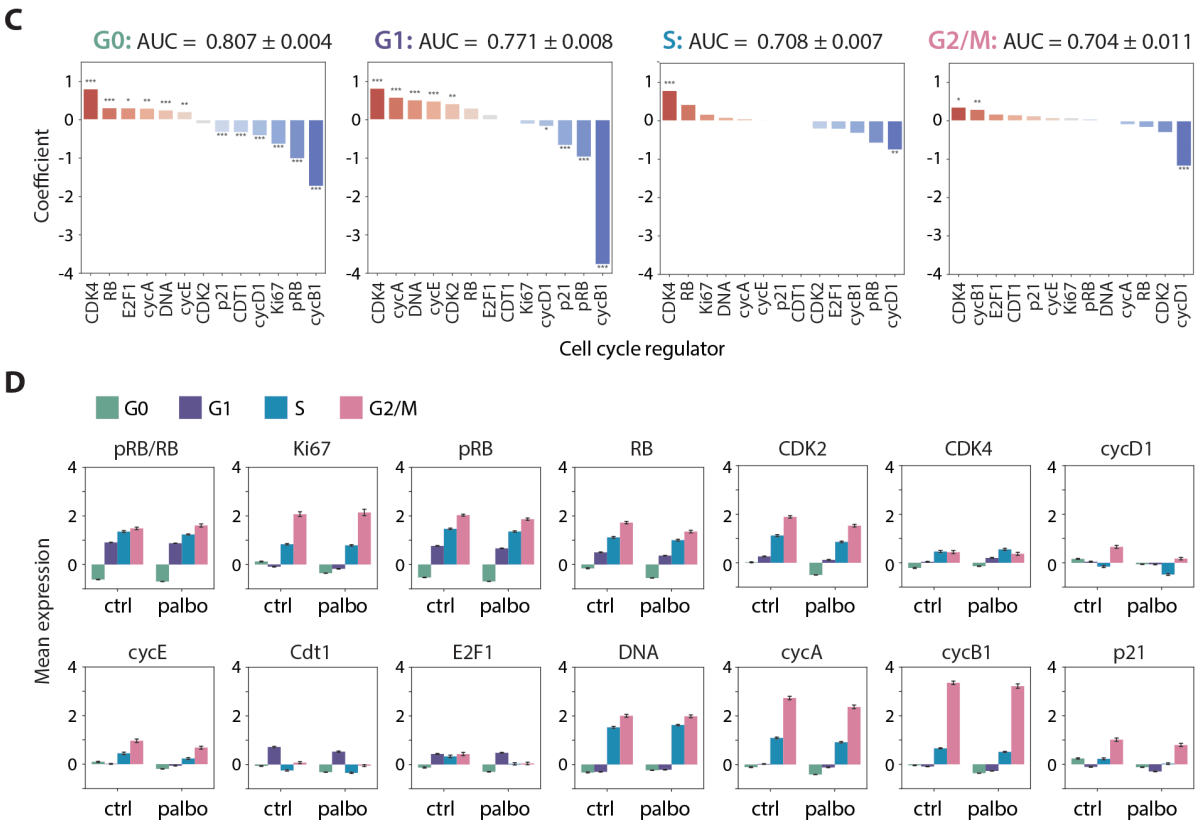


Figure S3. Shifts in expression of cell cycle proteins among fractionally resistant tumor cells for T47D biological replicate . A. Beginning with a downsampled dataset containing 20,000 cells from each condition (see **MATERIALS AND METHODS**), we identified the proliferating cells using a pRB/RB threshold. Proliferating single-cell distributions of cell cycle regulators at 0 nM, 10 nM, or 100 nM palbociclib, showing an expected reduction in the number of cells at higher drug concentrations. **B.** 95% confidence intervals (CI) of proliferating cells for the differences in mean expression (normalized z-scores) between either untreated cells and 10 nM palbociclib (*purple*); or untreated and 100 nM palbociclib (*orange*). Confidence intervals overlapping with the dashed line at 0 indicate a lack of statistical significance. Confidence intervals are wider when comparing 0 vs. 100 nM due to lower sample sizes at the highest dose of palbociclib. **C.** Logistic regression on all 60,000 cells predicting the odds that a given cell is either untreated (0 nM) or treated (10 nM or 100 nM) based on expression of its cell cycle regulators. Cell-to-cell increases in regulators shown in red (e.g., CDK4), or decreases in regulators shown in blue (e.g., cycD1), increase the odds of association with treated (10 nM or 100 nM) versus untreated cells. A separate regression was performed for each phase (G0, G1, S, G2/M), where the last three are considered proliferating (high pRB/RB). This analysis was performed on all cells, including non-proliferating (G0) cells. Significance: *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$. **D.** Z-score normalized expression levels of cell cycle features, stratified by cell cycle phase for all 60,000 cells. Bar height is the mean expression for untreated (ctrl) or palbociclib-treated (10 nM and 100 nM) cells. Error bars represent confidence intervals.

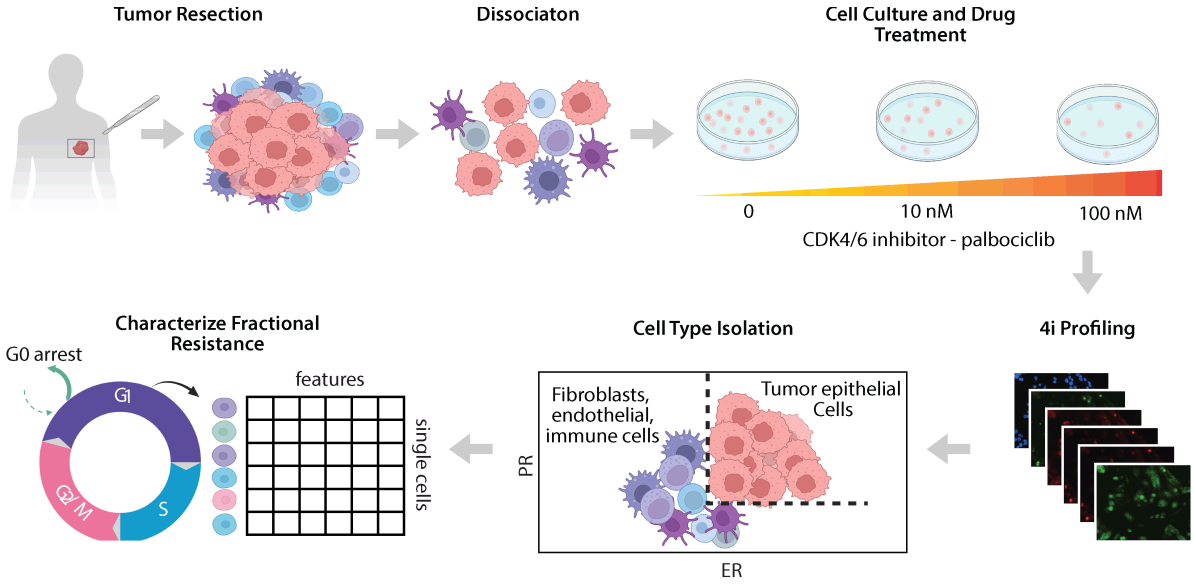


Figure S4. Experimental and analysis pipeline for characterizing fractional resistance in primary tumor cells. Detailed methods may be found in the **MATERIALS AND METHODS** section of the main text.

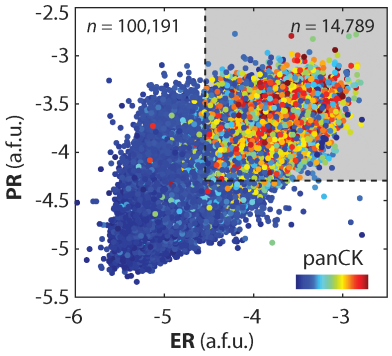


Figure S5. Image-based isolation of ER+/PR+ cells from resected tumor specimen. ~15% of cells were identified as ER+ and PR+ and subjected to downstream analysis.

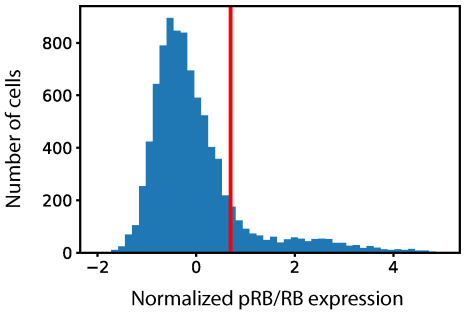
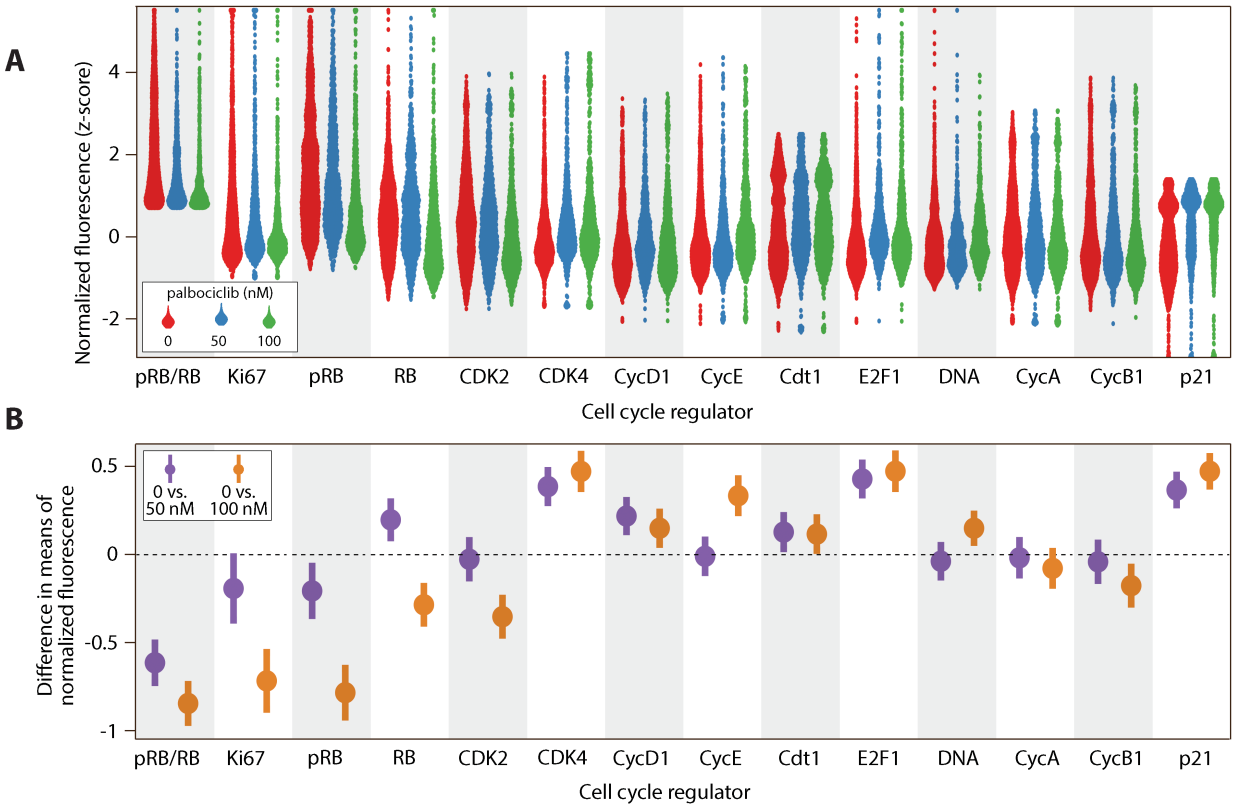


Figure S6. Distribution of pRB/RB in primary ER+/HER2- tumor cells. The red line marks the statistically determined threshold for demarcating cells as proliferating despite palbociclib treatment. Cells to the right of the line were used for downstream analyses.



(Figure S7 continued on next page)

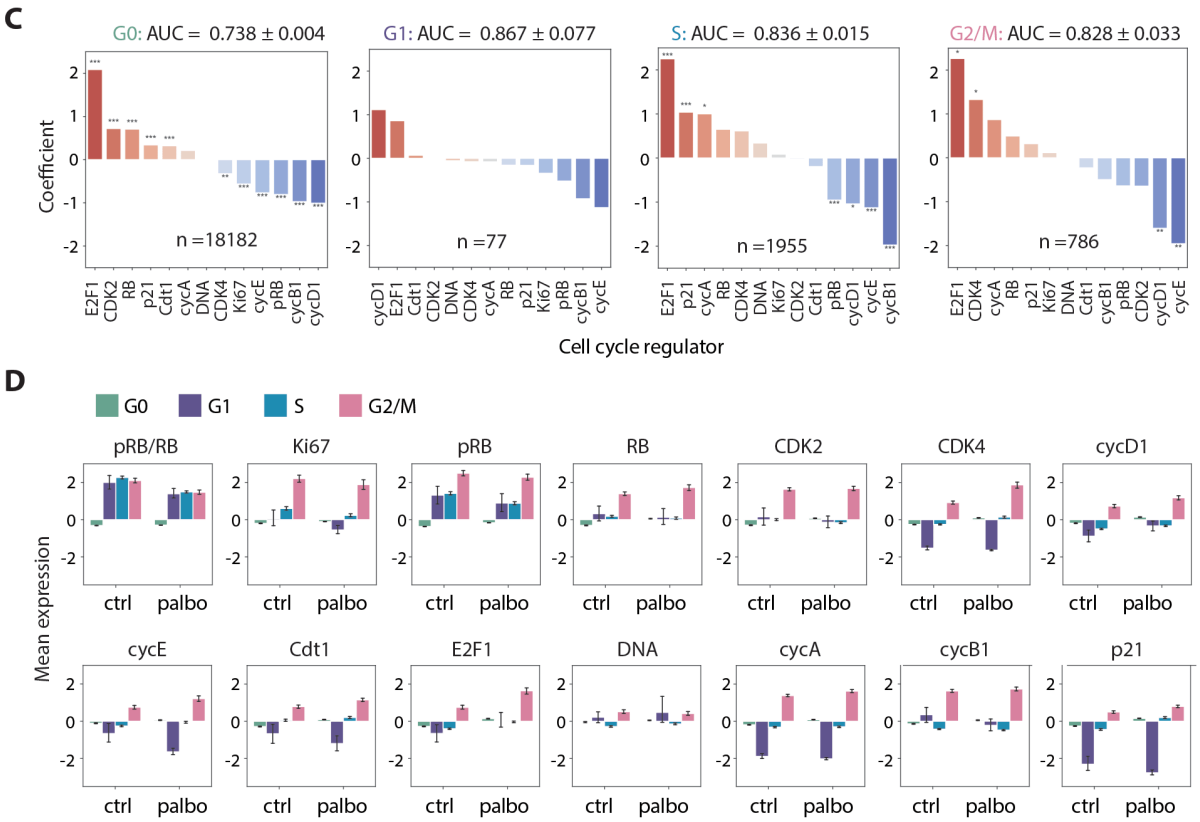


Figure S7. Palbociclib reveals shifts in expression of cell cycle proteins in primary tumor cells. A. Beginning with a downsampled dataset containing 2,000 cells from each condition (see **MATERIALS AND METHODS**), we identified the proliferating primary tumor cells using a pRB/RB threshold in an additional primary tumor sample from a different patient. Proliferating single-cell distributions of cell cycle regulators at 0 nM, 50 nM, or 100 nM palbociclib, showing an expected reduction in the number of cells at higher drug concentrations. **B.** 95% confidence intervals (CI) of proliferating cells for the differences in mean expression (normalized z-scores) between either untreated cells and 50 nM palbociclib (purple); or untreated and 100 nM palbociclib (orange). Confidence intervals overlapping with the dashed line at 0 indicate a lack of statistical significance. Confidence intervals are wider when comparing 0 vs. 100 nM due to lower sample sizes at the highest dose of palbociclib. **C.** Logistic regression on all 6,000 cells predicting the odds that a given cell is either untreated (0 nM) or treated (50 nM or 100 nM) based on expression of its cell cycle regulators. Cell-to-cell increases in regulators shown in red (e.g., E2F1), or decreases in regulators shown in blue (e.g., cycE), increase the odds of association with treated (50 nM or 100 nM) versus untreated cells. A separate regression was performed for each phase (G0, G1, S, G2/M), where the last three are considered proliferating (high pRB/RB). This analysis was performed on all cells, including non-proliferating (G0) cells. Significance: *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$. **D.** Z-score normalized expression levels of cell cycle features, stratified by cell cycle phase for all 6,000 cells. Bar height is the mean expression for untreated (ctrl) or palbociclib-treated (50 nM and 100 nM) cells. Error bars represent confidence intervals.

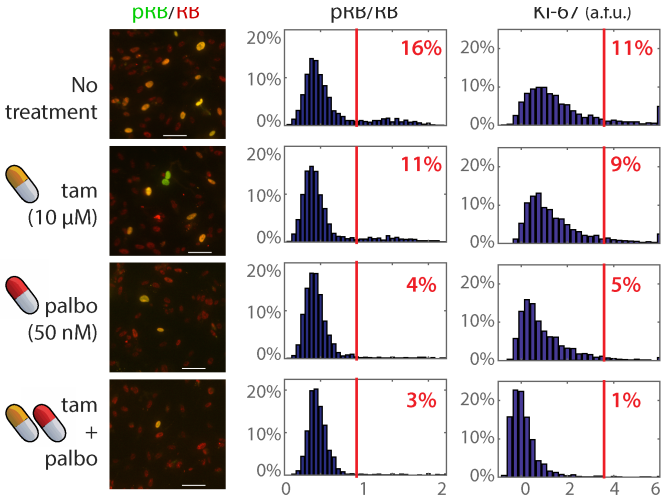
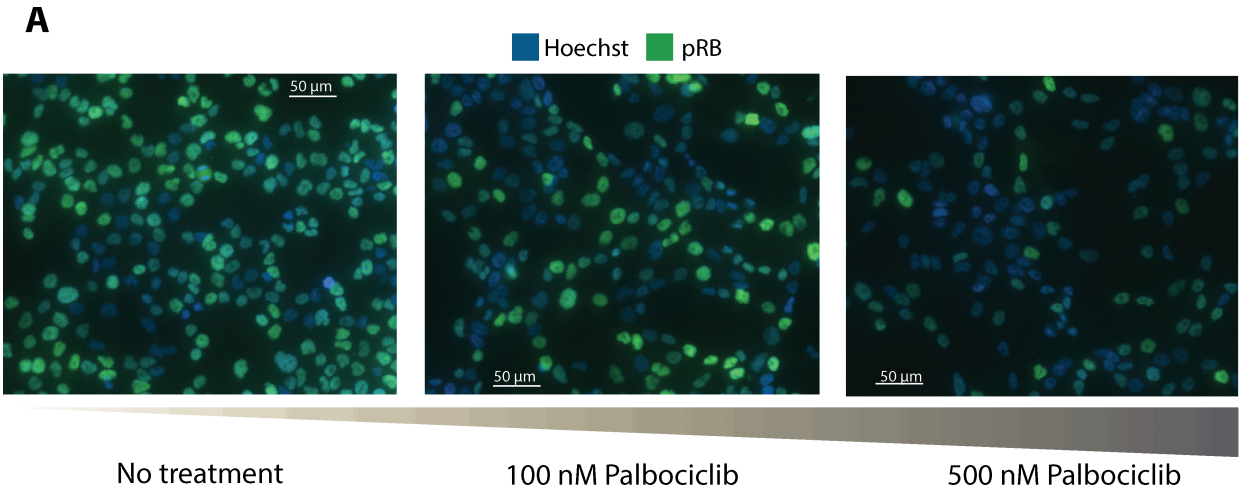


Figure S8. Fractional resistance in resected tumor cells under combination therapy. Primary ER+/PR+ tumor cells were treated with 10 μM tamoxifen and/or 50 nM palbociclib prior to 4i profiling. Y-axis denotes the percentage of cells at a given expression level.



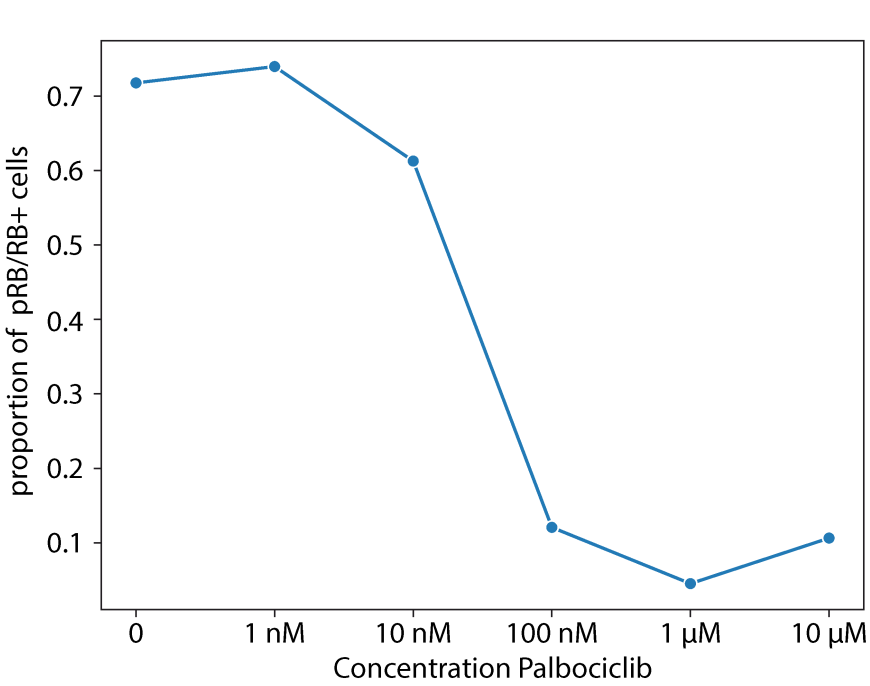


Figure S9. T47D single-cell proliferation in response to high concentrations of palbociclib A. Representative images of T47D cells at three different concentrations of palbociclib. pRB high cells, in bright green, show proliferating cells, and Hoechst staining in blue shows DNA content. Visual reductions in both DNA and pRB are visible at 100 nM and 500 nM concentrations of palbociclib, but proliferating "fractionally resistant" cells are still observable. **B.** Six-point palbociclib dose response curve for T47D cells. The proportion of pRB/RB+ cells is marked by first sketching to 10,000 cells per each treatment condition, and then identifying a higher modality of pRB/RB expression and calculating their proportion for each concentration (see **MATERIALS AND METHODS**). Proliferation of cells continues even with 10 µM of palbociclib, and not all fractionally resistant cells are eliminated.

DATA AND CODE AVAILABILITY

Preprocessed single-cell 4i datasets are publicly available in the Zenodo repository (1). Source code for image preprocessing, including cell segmentation, transformation, alignment, and quantification are publicly available in the GitHub repository (2). Source code for computational analyses, including functions for preprocessing, sketching, integration, trajectory inference, and other computational analyses as described in this manuscript are publicly available in the GitHub repository: (3).

SI MATERIALS AND METHODS

Primary human breast tumor cells

Under an Institutional Review Board (IRB) approved protocol, we obtained a tumor sample from a female patient with invasive lobular carcinoma that was positive for expression of the estrogen receptor (ER+) and negative for amplification of HER2 (HER2-). Written informed consent was obtained under a protocol approved by the Institutional Review Board at the University of North Carolina at Chapel Hill and included access to de-identified patient data which was obtained through an honest broker. The deidentified tumor specimen was obtained in the operating room suite within 15 minutes of resection. The sample was placed in DMEM/F12 (Gibco) media with 1% Penicillin-Streptomycin and transferred immediately to the laboratory on ice. The tumor specimen was sharply minced into 2-4 mm fragments. Enzymatic dissociation was performed using Gentle Collagenase/Hyaluronidase (Stemcell Technologies Inc. 07919) in DMEM/F12 supplemented with 5% BSA, Hydrocortisone (Stemcell Technologies Inc.), HEPES (Corning), and Glutamax (Gibco) for 16 hours at 37°C with cell agitation. The cells were gently centrifuged and washed twice with PBS supplemented with FBS and HEPES buffer. Cells were resuspended in ammonium chloride solution (Stemcell Technologies Inc. 07800) and incubated at 37°C with 5% CO2 to remove red blood cells. Cells were centrifuged and briefly trypsinized in warm 0.05% Trypsin-EDTA (Gibco) and DNase I. Cells were centrifuged and washed then resuspended in DMEM/F12 with 10% FBS. Cells were then strained using multiple rounds of sequential straining with 100 um and 40 um cell strainers to remove cell debris. folCells were counted using fresh trypan blue and the Countess cell counter (Life Technologies). Cells were plated on a glass 96-well plate coated with poly-L lysine at 100,000 cells per well. Cells were allowed to adhere for 48 hours at 37°C with 5% CO2 in DMEM/F12 media with 10% FBS. After 24 hours, media and non-adherent cells were removed. DMEM/F12 media with 10% FBS was

added containing vehicle, or palbociclib at 10 or 100 nM. Cells were incubated at 37°C with 5% CO2. After 24 hours of treatment, cells were fixed with PFA and 4i performed as described below.

FFPE slides sectioned at 4 microns were obtained from clinical pathology for the primary ER+/HER2- tumor. Immunohistochemistry (IHC) for Ki-67 antigen was performed using the Ki-67 Antibody (MIB-1, Dako) at 1:100 as we have previously described (4). A positive control was included. Ki-67 was scored according to the Ki-67 IHC MIB-1 pharmDx (Dako Omnis) Interpretation Manual for Breast Carcinoma. The Ki67 pharmDx score (%) was calculated as number of Ki-67 staining viable invasive (*in situ* disease was excluded) tumor cells divided by the total number of viable invasive tumor cells, multiplied by 100 for 2000 cells scored cells. The Ki67 staining for the primary tumor sample is shown in **Figure 1B**, which was scored as 14% Ki67+.

Iterative immunofluorescence

We followed the protocol of Gut *et al.* (5) with the following modifications. T47D (ATCC HTB-133) or primary cells from human tumors were fixed by adding 8% PFA (Thermo Scientific cat#28908) directly to the samples (1:1 v/v with media) for a final concentration of 4% PFA and incubated for 30 minutes at room temperature (RT). Samples were rinsed 3 times with PBS (pH=7.4)(200 µL/well for 96 well format) and incubated with 0.1% Triton X-100 (50 µL/well)(Fisher cat#BP151) for 15 minutes at RT to permeabilize the cells for immunofluorescence. Samples were then rinsed a single time with PBS and then incubated with Hoechst (Sigma cat#94403)(50 µL/well; 1:2500 dilution in PBS) for 15 minutes at RT to stain the DNA contained in the nucleus of the cells. Cells were rinsed once with PBS,100 µL/well of PBS was added to the wells, and cells were imaged. This ‘pre-stain’ is a key first step as it ensures that 1) the cells are well distributed in the well and 2) serves as a necessary quality control step to ensure that the cells are suitable for 4i. Samples deemed suitable for 4i were eluted, even though labeling with a primary antibody has not occurred. This is done as the elution process further opens the cells and permits optimal labeling.

Elution of samples was carried out by first rinsing the samples three times with water. Elution buffer (EB) was prepared fresh from a pre-mix stock (L-Glycine [0.5M](Sigma cat#50046), Urea [3M](Sigma cat#U4883) and Guanidinum Chloride [3M](Invitrogen cat#15502-016)) combined with TCEP-HCl [70 mM](Sigma #646547) and HCl (Fisher cat#SA49) to obtain a pH to 2.5. Samples were washed three times with EB (50 µL/well) for 10 minutes at RT with gentle

shaking. Of note, it is important not to exceed the number of washes or the duration of the washes as this may degrade the samples. Once elution was complete, the sample was rinsed one time with PBS prior to labeling with primary antibodies.

Labeling with primary antibodies first requires incubation with sBS (4i blocking solution) for 1 hour at room temperature (50 µL/well). The blocking solution was made up fresh and for every mL of solution one adds 14.6 mg Malemide [100 mM](Sigma cat#129585) and 5.35 mg NH_4Cl [100 mM](Sigma cat#A9434) to conventional blocking solution (cBS)(1% BSA (Sigma cat#A7906) in PBS). Once incubation with blocking solution was complete, samples were rinsed one time with PBS and primary antibodies (50 µL/well) were applied for an overnight incubation at 4°C with gentle rocking/shaking. It is important to note that the antibody solution was made in a conventional blocking solution at a dilution that is empirically determined and may contain several different antibodies. This does not present an issue as long as the antibodies have different species of origins (see **Table S1** for a list of primary antibodies used in this study). Alternatively, samples may be incubated with the primary antibody solution at room temperature for an hour or more, but labeling may not be as robust. Once the incubation with the primary antibody solution was complete, samples were rinsed one time with PBS, followed by three washes with PBS for 5 minutes each, followed by one final rinse in PBS. The final rinse with PBS was carefully aspirated off the sample to ensure all residual antibodies had been removed. Immediately following incubation with primary antibodies, fluorescent secondary antibodies specifically directed at the primary antibodies were applied. We used the Alexa series of secondary antibodies at a dilution of 1:500 in cBS along with Hoechst DNA stain at 1:2500. The secondary solutions (50 µL/well) were incubated for 1 hour at RT with gentle rocking and under conditions excluding light to prevent any photobleaching of the secondary fluorophores. Once this step was complete, cells were rinsed/washed in the same exact manner as the end of the primary antibody incubation step. During the wash step, fresh imaging buffer (IB) was prepared, which consists of N-acetylcysteine (NAC, Sigma cat#A7250) in water at a final concentration of 700 mM and pH of 7.4. We added 100 µL/well of IB to the samples and immediately imaged the cells.

Imaging was performed on a Nikon TiE inverted microscope utilizing a plan apo lambda 20X objective lens (NA = 0.75) with an Andor Zyla 4.2P sCMOS camera as a detector. NIS-Elements HCA (high content analysis) JOBS software was utilized in the acquisition of images as it permits the imaging of entire wells in a fast and automated fashion. Upon completion of imaging, samples were eluted per the protocol described above and the next round of labeling

and imaging was performed. It should be noted that every other round after elution, and before the next round of labeling, samples were imaged with the same exact experimental parameters with successful elution. This results in little to no fluorescent signal and ensures that the antibodies from the previous round have been successfully removed via the elution process and that no residual labeling is present to 'contaminate' the next round of imaging. This process of imaging and elution was repeated in an iterative manner to build a molecular profile for individual cells for each sample and treatment condition.

Image processing and cell property quantification

The image processing pipeline consisted of several steps to convert the raw images in the Elements nd2 format to a matrix of single cells with protein expression quantified in the nucleus, ring, and cytoplasm. The four primary steps were: 1) **cell segmentation** via the Cellpose algorithm (71) to define the nucleus for each round from the Hoechst staining, 2) cell segmented masks were **aligned** across all rounds of images, 3) **punchmasks** were manually drawn to exclude any debris (cellular or otherwise), and 4) **cell properties** were calculated from individual segmented nuclei for all the intensity channels. We followed the image preprocessing pipeline as described in the GitHub repository: (2).

Data preprocessing

To compare tumor cells from a cell culture model of ER+ breast cancer (T47D) and tumor cells from a primary tumor sample resected from a patient, we performed a series of preprocessing steps. Following image preprocessing and cell property quantification, we computationally filtered cells within the primary tumor samples to retain only the tumor epithelial cells by gating cells according to the median expression of ER and PR (see **Figure S5**). Feature selection was then performed by selecting the intersection of core cell cycle regulators profiled in both datasets (P = 13). Lastly, T47D and primary tumor datasets were standardized independently by mean centering and scaling to unit variance. The abbreviated experimental pipelines for the T47D and primary tumor samples are shown in **Figure 1** and **Figure S4**, respectively.

Cell cycle annotations

Based on existing work annotating the cell cycle phases, the bimodal distribution of the ratio of phosphorylated to total RB levels (pRB/RB) was used to distinguish proliferative cells (G1/S/G2/M, high pRB/RB) from arrested cells (G0, low pRB/RB) (6–8). To agnostically set the pRB/RB threshold for both datasets without any underlying assumptions on the shape or spread of the distribution, we implemented a data normalization step outlined previously in Ref. (9),

based on the idea that if a distribution is bimodal, there will be a region of higher density on one side of the median as compared to the other. More precisely, given a sorted list of expression values, $x_{pRB/RB}$, we first computed the median of the distribution as m = median($x_{pRB/RB}$). We then folded the left side of the distribution, $x_{pRB/RB} < m$, over the right side of the median by $z_{pRB/RB}\,[x_{pRB/RB} < m] = 2*m - x_{pRB/RB}\,[x_{pRB/RB} < m]$, where z is the new one-sided distribution. Next, we computed a specified percentile, p, of this one-sided distribution and subtracted the median, denoting this difference as a, $z_{(pRB/RB,\,p)} - m = a$. The cutoff point of the second mode of the distribution (i.e., proliferative cells with high pRB/RB) was then defined according to the values of $x_{pRB/RB}$ that fell within the range $(m - a, m + a)$. More specifically, we denote $s = |(z \in (m - a, m + a))|$, where | is the cardinality of the set within the specified range. We define the point separating the modes of the distribution, c, as $c = s / n$, where n is the total number of cells in the distribution. We selected a percentile value of (p = 0.2) for T47D datasets (**Figures 1E, S1A**), (p = 0.7) for an additional T47D replicate (**Figure S2**), (p = 0.25) for the T47D triplicate (**Figure S3),** (p = 0.7) for the primary tumor samples (**Figures 3, S7)**, and (p = 0.7) for the primary tumor replicate (**Figure S6**) based on the distribution of pRB/RB expression values.

For the proliferative cells, as indicated above by high expression of pRB/RB, cell cycle phase annotations (G1, S, and G2/M) were subsequently determined by fitting a three component Gaussian Mixture Model to the log-transformed measurements of DNA content, cyclin A, and cyclin B1. Unsupervised clusters were annotated as follows: G1 (DNA content = 2C, low cyclin A), S (DNA content = 2-4C, medium cyclin A), and G2/M (DNA content = 4C, high cyclin A). The Gaussian Mixture Model was implemented using the sklearn 0.24.1 package in Python.

Sketching

To identify a limited subset of representative cells for each dataset and facilitate a direct comparison cell cycle states across samples and treatment conditions, we selected an equal number of cells (n = 2,000) from each treatment condition (untreated, 10 nM, and 100 nM palbociclib) within a dataset (T47D, primary tumor) using kernel herding sketching (10). 2,000 was chosen as it was near the maximum sample size in a respective treatment condition, with the primary tumor sample having 2,806 cells in the 100 nM condition. In the T47D triplicate sample (**Figures S3**), we sample 20,000 cells from each condition. Kernel herding sketching performs principled downsampling of the data and selects prototypical cells that are representative of the original distribution of cell type frequencies (e.g., cell cycle phases), while also ensuring rare cell types are sufficiently sampled. For each dataset, sketched cells from each condition were then vertically concatenated into a $N \times P$ matrix prior to downstream

analysis, where N is the number of sketched cells across three treatment conditions (N = 6,000 or 60,000) and P is the number of profiled proteomic imaging features (P = 14).

Confidence intervals

To identify shared and distinct mechanisms of resistance to palbociclib treatment, we examined the fractional arrest profiles of proliferating T47D and primary tumor cells as follows. For each dataset (T47D, primary tumor), we computed two sample *t*-tests assuming equal variance between each cell cycle effector in untreated proliferative cells and each treatment condition separately. More specifically, we computed 95% confidence intervals between proliferative untreated and 10 nM palbociclib cells, and 95% confidence intervals between proliferative untreated and 100 nM palbociclib cells.

Logistic regression

Logistic regression (11) is a supervised learning algorithm that can be used to predict the probability of a binary outcome (e.g., control, treated) based on a set of input features (e.g., proteomic imaging features). To ascertain changes in cell cycle regulators associated with palbociclib treatment, a logistic regression model was trained on the proteomic expression profiles of cells within each cell cycle phase (G0, G1, S, G2/M) for each dataset (T47D, primary tumor) to predict the treatment group of a tumor cell (control, treated). We join 10 nM and 100 nM as one treatment condition to be predicted due to low counts of proliferating cells when creating separate models by cell cycle phase. In this case, the control group consisted of untreated cells, whereas the treated group consisted of cells treated with either concentration of palbociclib (10 nM, 100 nM). For each dataset and phase, nested ten-fold cross-validation was performed using stratified random sampling to assign cells within a particular phase to either a training or a test set. Using a grid search, hyperparameters were tuned within each fold prior to training the model, and cells were classified as control or treated from the test data. Classification performance was subsequently assessed by computing the area under the receiver operator characteristic curve (AUC ROC). Logistic regression was implemented using the sklearn v0.24.1 package in Python.

Data integration with TRANSACT

TRANSACT (Tumor Response Assessment by Nonlinear Subspace Alignment of Cell lines and Tumors) (12) is a nonlinear data integration method that can be used to identify a shared subspace of preclinical cell lines and patient-derived samples. Briefly, TRANSACT merges datasets by performing kernel principal components analysis (13) on each individual dataset,

and then geometrically aligns these nonlinear principal components to extract principal vectors that represent similar nonlinear weighted combinations of expression profiles across data samples. A consensus data representation, corresponding to biological processes that are present within both preclinical cell lines and primary tumor samples, is then computed by optimizing the match between interpolated sets of principal vectors using geodesic flow (14). We performed data integration of T47D and primary tumor samples using TRANSACT to more robustly represent and compare cell cycle trajectories under palbociclib treatment. More specifically, we identified a shared latent space by first computing consensus features for T47D and primary tumor samples, and then projecting both datasets onto the consensus features. Here, the integrated dataset, $F^{T47D,Tum}$, consisted of 12,000 cells and 14 shared consensus features. Of note, integration was performed on the sketched datasets to ensure that the joint latent space was not overwhelmed by one data modality when performing downstream analyses, such as dimensionality reduction and trajectory inference. TRANSACT was implemented using the transact-dr v1.0.1 package in Python, where cell similarity was defined using a radial basis function with a scaling factor, $\gamma = 1/\sqrt{500}$.

PHATE dimensionality reduction

To visualize high dimensional single-cell 4i profiles of the cell cycle, we performed nonlinear dimensionality reduction with PHATE (Potential of Heat-diffusion for Affinity-based Trajectory Embedding) on the integrated dataset of T47D and primary tumor samples. PHATE (15) is a nonlinear dimensionality reduction method that effectively represents the geometry of complex continuous data structures and has been shown previously (16–18) to successfully recapitulate proliferative and arrest cell cycle trajectories. PHATE was implemented using the phate v1.0.7 package in Python by constructing a *k*-nearest neighbor graph (*k* = 150) according to pairwise Euclidean distances between all pairs of cells from the consensus feature space computed by TRANSACT, $F^{T47D,Tum}$.

Trajectory inference and alignment

To characterize trajectories through the cell cycle under palbociclib treatment, we performed trajectory inference using Slingshot (19) on each dataset (T47D, primary tumor) and treatment condition (untreated, 10 nM, and 100 nM palbociclib). This trajectory inference method was chosen as it was shown previously (20) to outperform alternative methods on inferring simple continuous or branched cellular trajectories. Slingshot was implemented using the slingshot v2.7.0 package in R by 1) fitting a minimum spanning tree through cluster centroids defined by cell cycle phase annotations, and then 2) estimating pseudotime by projecting cells onto the

principal curves fit through the PHATE embedding generated from the consensus feature space computed by TRANSACT. The *root* (starting) cluster was defined as the G0 phase. Across most inferred cellular trajectories, Slingshot identified the canonical ordering of cell cycle phases (G0 to G1 to S to G2/M). However, we note that in two scenarios (untreated T47D and 100 nM palbociclib primary tumor), Slingshot identified a minimum spanning tree spanning from G0 to G1 to G2/M phases for the primary tumor and G0 to S to G2/M for the T47D, respectively.

Given that trajectory inference was performed on cells from each treatment condition separately, we subsequently aligned the trajectories onto one common pseudotime axis using TrAGEDy to enable a direct comparison of continuous proteomic expression profiles across treatment conditions. TrAGEDy (Trajectory Alignment of Gene Expression Dynamics) (21) is a trajectory alignment method that can align cells from two independently generated trajectories and has been shown previously to enable robust comparisons of continuous expression trends across treatment conditions when aligning Slingshot trajectories from PHATE dimensionality reduced single-cell data. Methodologically, TrAGEDy first interpolates points at different regions of the trajectory to overcome any noise inherent to single-cell data. Next, the Spearman correlation is computed between the set of interpolated points along the two trajectories to define a trajectory similarity matrix. Lastly, TrAGEDy uses a dynamic time warping approach (22) with modifications to account differences in cell states in order to find the optimal alignment through the similarity matrix of interpolated points. This approach ensures that the original pseudotemporal ordering is preserved, while the distance between points across trajectories is minimized. For each dataset, we performed trajectory alignment with TrAGEDy by aligning the 10 nM and 100 nM trajectories to one another, followed by alignment to the untreated trajectory. TrAGEDy was implemented with 50 interpolating points using the R code provided in the GitHub repository at: https://github.com/No2Ross/TrAGEDy.

To visualize continuous feature expression trends, a generalized additive model (GAM) with a cubic spline basis function with shrinkage was fit for each feature as an outcome along the aligned pseudotime as sole covariate using the mgcv v1.8-42 package in R. Moreover, to identify an approximate transition point from arrest into proliferation, we computed the inflection point where approximately 50% of the cells were G0 and 50% of the cells were proliferative (non-G0) for each trajectory. To do so, we discretized the aligned pseudotime values into bins and then computed the ratio of G0/non-G0 cells for each bin. The transition point was defined as the aligned pseudotime value where this ratio was approximately one. For the untreated trajectories, we chose a smaller number of bins (n = 25) to find the inflection point due to the

larger number of proliferative cells, whereas for the treated trajectories, we chose a larger number of bins (n = 50). Of note, this transition point was excluded for the 100 nM palbociclib primary tumor trajectory due to the small sample size of proliferative cells.

CDK2 inhibitor co-treatment

T47D ER+/HER2- breast cancer cells were obtained from the ATCC (catalog number HTB-133). Cells were seeded on a poly-D-lysine coated glass-bottom plate at 20,000 cells per well. Cells were treated with a range of CDK2 inhibitor CVT-313 (Selleckchem Cat# S6537) concentrations (0, 0.01, 0.5, 1, 2, 5, 7.5, 10, and 20 μM) for each of CDK4/6 inhibitor palbociclib (Selleckchem Cat# PD-0332991) treatment condition (0, 10, 50 nM) for 24 hours. Cells were fixed at room temperature with 4% PFA. Following image preprocessing and cell property quantification, T47D single-cell data were standardized by mean centering and scaling to unit variance. Across all treatment conditions, the bimodal distribution of the ratio of phosphorylated to total RB (pRB/RB) was used to determine pRB/RB positive cells. Here, the threshold was computed with percentile value (p = 0.4) as previously described (See cell cycle annotations). For each palbociclib treatment condition (i.e., control, 10 nM-, 50 nM- palbociclib), fold changes in the proportion of pRB/RB positive cells under treatment with CVT-313 were computed by normalizing the proportion of pRB/RB positive cells by the average proportion of the first three doses of CVT-313 (0, 0.01, 0.1 μM). To assess the statistical significance of the proportion of pRB/RB positive cells, a two-sided Wilcoxon rank sum test was performed between control and 50 nM palbociclib treated cells using the ranksums function in scipy v1.10.1.

SUPPORTING REFERENCES

1. T. M. Zikry, *et al.*, Single-cell datasets for cell cycle plasticity underlies fractional resistance to palbociclib in ER+/HER2- breast tumor cells (2023) https:/doi.org/10.5281/ZENODO.10063003.

2. K. M. Kedziora, *4i_analysis* (Github) (January 18, 2024).

3. T. M. Zikry, *et al.*, *fractional_resistance: Characterizing fractional resistance in T47D and primary tumor samples* (Github) (May 16, 2023).

4. H. Kim, *et al.*, Tamoxifen response at single cell resolution in estrogen receptor-positive primary human breast tumors. *Clin. Cancer Res.* (2023) https:/doi.org/10.1158/1078-0432.CCR-23-1248.

5. G. Gut, M. D. Herrmann, L. Pelkmans, Multiplexed protein maps link subcellular organization to cellular states. *Science* **361** (2018).

6. A. B. Pardee, A restriction point for control of normal animal cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 1286–1290 (1974).

7. A. Zetterberg, O. Larsson, Kinetic analysis of regulatory events in G1 leading to proliferation or quiescence of Swiss 3T3 cells. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 5365–5369 (1985).

8. J. Moser, I. Miller, D. Carter, S. L. Spencer, Control of the Restriction Point by Rb and p21. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8219–E8227 (2018).

9. F. R. Hampel, E. M. Ronchetti, P. Rousseeuw, W. A. Stahel, *Robust statistics: the approach based on influence functions* (Wiley-Interscience; New York, 1986).

10. V. A. Baskaran, J. Ranek, S. Shan, N. Stanley, J. B. Oliva, Distribution-based sketching of single-cell samples in *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '22., (Association for Computing Machinery, 2022), pp. 1–10.

11. D. R. Cox, *Analysis of Binary Data* (Chapman and Hall, 1969).

12. S. M. C. Mourragui, *et al.*, Predicting patient response with models trained on cell lines and patient-derived xenografts by nonlinear transfer learning. *Proceedings of the National Academy of Sciences* **118**, e2106682118 (2021).

13. B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Comput.* **10**, 1299–1319 (1998).

14. B. Gong, K. Grauman, F. Sha, "Geodesic flow kernel and landmarks: Kernel methods for unsupervised domain adaptation" in *Domain Adaptation in Computer Vision Applications*, Advances in computer vision and pattern recognition., (Springer International Publishing, 2017), pp. 59–79.

15. K. R. Moon, *et al.*, Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).

16. W. Stallaert, *et al.*, The structure of the human cell cycle. *Cell Syst* **13**, 230-240.e3 (2022).

17. W. Stallaert, *et al.*, The molecular architecture of cell cycle arrest. *Mol. Syst. Biol.* **18**, e11087 (2022).

18. J. S. Ranek, W. Stallaert, J. Milner, N. Stanley, J. E. Purvis, Feature selection for preserving biological trajectories in single-cell data. *bioRxiv*, 2023.05.09.540043 (2023).

19. K. Street, *et al.*, Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

20. W. Saelens, R. Cannoodt, H. Todorov, Y. Saeys, A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).

21. R. F. Laidlaw, E. M. Briggs, K. R. Matthews, R. McCulloch, T. D. Otto, TrAGEDy: Trajectory Alignment of Gene Expression Dynamics. *bioRxiv*, 2022.12.21.521424 (2022).

22. A. Alpert, L. S. Moore, T. Dubovik, S. S. Shen-Orr, Alignment of single-cell trajectories to compare cellular expression dynamics. *Nat. Methods* **15**, 267–270 (2018).