Spherical Manifolds Capture Drug-Induced Changes in Tumor Cell Cycle Behavior

Olivia Wen[1,2], Samuel C. Wolff[2,3], Wayne Stallaert[6], Didong Li[4], Jeremy E. Purvis[2,3,†] & Tarek M. Zikry[2,4,5,†]

[1]Department of Biology, [2]Computational Medicine Program, [3]Department of Genetics, [4]Department of Biostatistics, [5]School of Data Science and Society, University of North Carolina at Chapel Hill, NC
[6]Department of Computational and Systems Biology, University of Pittsburgh, PA
[†]Corresponding Authors: jeremy_purvis@med.unc.edu, tarek@unc.edu

CDK4/6 inhibitors such as palbociclib block cell cycle progression and improve outcomes for many ER+/HER2- breast cancer patients. Unfortunately, many patients are initially resistant to the drug or develop resistance over time in part due to heterogeneity among individual tumor cells. To better understand these mechanisms of resistance, we used multiplex, single-cell imaging to profile cell cycle proteins in ER+ breast tumor cells under increasing palbociclib concentrations. We then applied spherical principal component analysis (SPCA), a dimensionality reduction method that leverages the inherently cyclical nature of the high-dimensional imaging data, to look for changes in cell cycle behavior in resistant cells. SPCA characterizes data as a hypersphere and provides a framework for visualizing and quantifying differences in cell cycles across treatment-induced perturbations. The hypersphere representations revealed shifts in the mean cell state and population heterogeneity. SPCA validated expected trends of CDK4/6 inhibitor response such as decreased expression of proliferation markers (Ki67, pRB), but also revealed potential mechanisms of resistance including increased expression of cyclin D1 and CDK2. Understanding the molecular mechanisms that allow treated tumor cells to evade arrest is critical for identifying targets of future therapies. Ultimately, we seek to further SPCA as a tool of precision medicine, targeting treatments by individual tumors, and extending this computational framework to interpret other cyclical biological processes represented by high-dimensional data.

Keywords: Manifold learning; Dimensionality reduction; ER+/HER2- Cancer.

1. Introduction

Despite promising results of CDK4/6 inhibitors for treating ER+/HER2- breast cancer, 10-20% of patients show initial drug resistance, and all patients develop resistance over time.[1] Resistance is thought to arise from the heterogeneity of molecular states in individual tumor cells, and one potential source of this cell-to-cell heterogeneity is the cell cycle. In recent years, single-cell studies have revealed that the cell cycle can show remarkable flexibility.[2] For example, individual tumor cells may progress through cell cycle phases with variable durations, or show altered expression levels of core cell cycle regulators.[3–5] The ability of cells to upregulate or downregulate certain protein signaling pathways is referred to as cell cycle plasticity.

Prior studies of ER+/HER2- cells —both in a cell culture model and from a primary tumor sample— have demonstrated how molecular differences allow individual tumor cells to evade CDK4/6 inhibitor therapy through alternative cell cycle paths.[6] Therefore, to mechanistically understand how resistance develops, it is important to develop robust analytical methods for characterizing the underlying manifolds along which cell cycle trajectories proceed.

Commonly, to more easily detect trends in increasingly high-dimensional single-cell data, dimensionality reduction methods are used. Dimensionality reduction techniques transform high-dimensional data into a low-dimensional space such that the most valuable information, or original structure, of the data is preserved. Manifold learning, a nonlinear approach to dimensionality reduction, is often applied to high-dimensional data as a tool for visualization, data exploration, and statistical analysis. In this study, we collected and analyzed single-cell data of T47D, a model human ER+/HER2- breast cancer cell line,[7] that we introduced to varying doses of palbociclib, one of three FDA-approved CDK4/6 inhibitors (Fig. 1). Complex molecular signatures were obtained for each cell by performing iterative indirect immunofluorescence imaging (4i)[8] using 20 cellular features relevant to proliferation. The result is a high-dimensional dataset that we can broadly interpret as a representation of the cell cycle. We seek to reduce the high-dimensional representation and visualize it in a more interpretable space. The selection of an appropriate method for characterizing cell cycle data is a necessary and crucial step for both biological and statistical interpretations. However, it is difficult to assess the performance of these methods as no fixed statistics exist to directly compare the effectiveness of one method over another. Thus, we must biologically interpret the results using known cell cycle markers and trends.

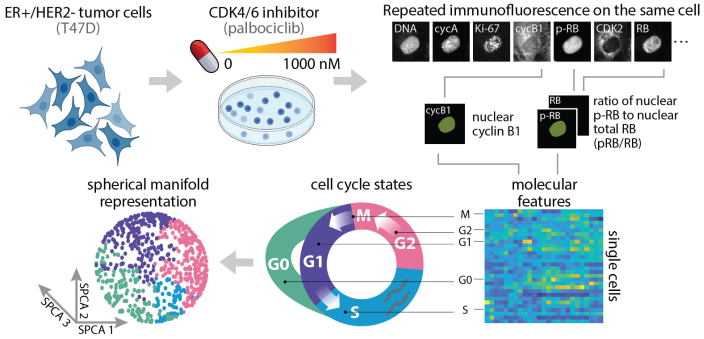

Fig. 1. **Pipeline for generating cell cycle manifold from single-cell images**. T47D tumor cells were treated with increasing concentrations of palbociclib. 4i was then performed using a panel of cell cycle-specific markers, resulting in a tabular dataset after raw image processing and segmentation. Because individual cells are not synchronized, individual cells span a range of cell cycle states. SPCA was then applied to estimate a hypersphere manifold representation of the cell cycle in a lower dimensional space. Each dot represents an individual tumor cell in a specific cell cycle state.

Here, we present spherical principal component analysis (SPCA) as an effective tool for modeling the underlying cell cycle structure of single-cell ER+/HER2- breast cancer cell line data (Fig. 1).[a] By assuming the data lie on a reduced spherical space, SPCA helps preserve gradual cell state transitions and cell-to-cell heterogeneity. Using trajectory inference approaches, we demonstrate how SPCA captures cyclical patterns of cell cycle regulators not

[a]All code, additional experimental details, and full supplemental feature plots can be found at `https://github.com/purvislab/SingleCell_HyperSphere`. Data are available at `https://doi.org/10.5281/zenodo.13621367`.

found in potential of heat-diffusion for affinity-based transition embedding (PHATE) or principal component analysis (PCA) models. Structural differences in spherical manifolds across treatment conditions also point to driving factors of CDK4/6 inhibitor response, which can help identify downstream clinical targets involved in treatment-resistant pathways.

2. Related Work

Manifold Estimation

Manifold learning is often a necessary step of high-dimensional data analysis. One branch of manifold learning techniques is manifold estimation. Manifold estimation approaches identify a low-dimensional embedding that preserves local and global structures without imposing assumptions about the structure of the data. PHATE is an example of a manifold estimation approach that has been found to be successful in producing clean, denoised models of biological data and preserving continuous trajectories.[9] PHATE captures local and global relationships by computing neighborhood relationships between cells, performing diffusion using local affinities, and projecting diffusion distances to create a two- or three-dimensional embedding. These embeddings can be used for hypothesis generation and visual comparison of cell cycle progressions.[6,10] However, because PHATE and other manifold estimation approaches do not assume an underlying structure, no statistical inferences can be made on these embeddings or between manifolds produced by different datasets.

Manifold Approximation

Other techniques, called manifold approximation, assume data to lie on an underlying structure. Thus, fitted values and error metrics can be computed. The most commonly used method for manifold approximation[11] is PCA.[12] PCA identifies features that are responsible for the most variance and projects the linearly transformed data onto a subspace of fewer dimensions. PCA has a long history of use in the biological field and requires low computational power, but is sensitive to noise, making it suboptimal for use with heterogeneous data such as single-cell data.[13,14]

Other manifold approximation methods assume data to lie on a more complex surface. SPCA, a variant of PCA, is one such method that assumes data lie on a sphere in a lower dimensional space.[15] For a dataset X reduced using SPCA, a spherical manifold M is parameterized by a radius r, a center c, and an affine subspace V. First, the subspace V where the optimal sphere lies is estimated from the input dataset X. A loss function is minimized to identify an optimal sphere by reducing the number of points that lie outside or inside the surface of the sphere. The optimal center and radius are estimated from the minimization of the loss function. From the parameters c (center), r (radius), and V (subspace), a projection for X onto the sphere is defined. SPCA has previously been applied to cell cycle data of retinal pigmented epithelial (RPE) cells and has been found to fit the data better than other methods.[10,14] However, a deep dive of the exact cell cycle trends was not explored, nor were SPCA manifolds of different datasets, such as cells of different treatment conditions, compared.

Trajectory Inference

Trajectory inference methods can be applied to high-dimensional, single-cell datasets to quantify the progression of dynamic cellular processes.[16] Revelio, a method that leverages PCA,

revealed single-cell transcriptomic data to follow a 2D circular trajectory.[17] However, this method seeks to remove cell cycle effects whereas we aim to study them and the response of proteomic states to forms of perturbation. Revelio also orders cells according to gene markers of cell state transitions and is suboptimal when applied to cells that follow different dynamics.[18]

One robust method for capturing dynamic cellular processes and representing noisy, single-cell data is Slingshot, a curve-based trajectory inference method.[19] Slingshot identifies single or multiple branched trajectories using two main steps: (1) constructing a minimum-spanning tree between clusters of data to identify a global lineage structure and (2) fitting smooth principal curves to each lineage. Orthogonal projections of each data point onto the curve assign a pseudotime, representing cell cycle progression, for each cell.

Slingshot also allows for various levels of supervision and flexibility in the choice of upstream data analysis methods. At a minimum, Slingshot requires data that has been clustered and reduced, a list of cluster labels, and specification of the dimensionality reduction method performed. Additional supervision can be achieved by specifying a start cluster, an end cluster, or the number of lineages to infer. Previously, Slingshot has been found to identify smooth cell cycle trends in PHATE embeddings.[6] The flexibility of Slingshot and its success with cell cycle data makes it an ideal method for comparing cell cycle paths inferred using different manifold learning approaches.

3. Methods

Experimental Details

T47D ER+/HER2- breast cancer cells were obtained from the ATCC (catalog number HTB-133) and maintained at 37°C with 5% CO_2 in RPMI-1640 media supplemented with 10% fetal bovine serum (FBS). Cells were plated on a glass 96-well plate coated with poly-L lysine at 25,000 cells per well. Cells were allowed to adhere for 24 hours at 37°C with 5% CO_2 in RPMI-1640 media with 10% FBS. After 24 hours, media and non-adherent cells were removed. RPMI-1640 media with 10% FBS was added containing vehicle, or palbociclib at 0, 1, 10, 100, or 1,000 nM. Cells were incubated at 37°C with 5% CO_2. After 24 hours of treatment, cells were fixed with PFA, and iterative indirect immunofluorescence imaging (4i) was performed as described below. Single-cell proteomic measurements for samples were obtained using 4i by adapting the protocol previously described in Refs. 6,8. Following image and data preprocessing, cell cycle phases were annotated using a three component Gaussian Mixture Model (sklearn v0.24.1) on the log-transformed measurements of DNA content, cyclin A, and cyclin B1, as these features were previously shown to minimally represent the cell cycle.[20] The full and close-up 4i images used for this study can be seen in Fig. S1, S2[a].

Manifold Approximation with SPCA

SPCA[15] was implemented in Python to identify the c_i, r_i, and V_i of the sphere that characterizes cells from each treatment condition i. Three dimensions were chosen to aid in visual comparison, but other methods, such as the identification of an elbow plot,[15] exist to identify the optimal lower dimension of a dataset. To identify a shared subspace V_G for comparison of all treatment conditions on a uniform scale, we applied SPCA using the complete dataset.

20-feature cell signatures from each treatment condition i were projected onto spheres with center c_i and radius r_i in subspace V_G. Orientations of plots were selected based on the best visual separation of phases or treatment conditions.

Manifold Comparisons with PHATE and SPCA

We visually and statistically compared the performance of the spherical manifolds approximated by SPCA to three-dimensional manifolds produced by PHATE and PCA. PHATE[9] was performed in Python (phate v1.0.11) on the complete dataset of all treatment conditions. A k-nearest neighbor graph was constructed to create the three-dimensional PHATE structure using the following hyperparameters: $n\_components = 4$, n_jobs = -1, $knn = 200$, and $t = 12$. The hyperparameters were tuned according to hyperparameters selected for previous PHATE models of cell cycle data.[6,20,21]

Python was also used to perform PCA[12] (scikit-learn v1.3.2). PCA was run using the complete dataset such that all treatment conditions can be evaluated in the same space. To produce a three-dimensional visualization, $n\_components$, the number of features to extract in the reduced dataset, was set to three.

Cell Cycle Trajectory Inference Using Slingshot

To assess the recapitulation of temporal trends, we applied trajectory inference to infer cell cycle paths. Slingshot was performed in R (slingshot v2.6.0) using each of the manifold learning approaches (PHATE, PCA, and SPCA) as the upstream dimensionality reduction method. Slingshot trajectories were inferred through cells from each of the treatment conditions. We provided cell cycle phase annotations (G0, G1, S, G2/M) as cluster labels and specified G0 as the start cluster. Pseudotimes were normalized to a scale of 0 to 1 to allow for the comparison of lineages on a uniform scale. To identify feature expression trends over pseudotime, locally estimated scatterplot smoothing (LOESS)[22] curves were fit using Python (v2.1.2).

4. Results

Recapitulating the Cell Cycle

From the 20-feature proteomic signatures and cell cycle phase labels (G0, G1, S, G2/M) of 64,502 T47D cells, we generated tabular datasets of cells from each treatment condition (n_0=10,366, n_1=10,675, n_{10}=13,051, n_{100}=15,688, n_{1000}=14,722). Each row describes a cell's unique molecular state, thus providing a complete representation of the cell cycle altogether. To identify a lower dimensional manifold that preserves the cyclical nature of the cell cycle, we performed SPCA[15] for each palbociclib dose resulting in five three-dimensional hyperspheres characterized by unique centers and radii projected onto a shared global reduced space.

We expect neighborhood relationships to be preserved such that cells in similar states, and thus with similar molecular signatures, are located near each other on a lower-dimensional manifold. Similarly, cells with different proteomic profiles are located far apart in 20-dimensional space and should remain further away on a three-dimensional manifold. To assess the ability of SPCA to capture differences in cell states, we visualized the distribution of cell cycle phases across the SPCA hyperspheres (Fig. 2A). For each of the treatment conditions, we obtained a spherical manifold that successfully captured differences between phases and the canonical progression of cells through the cell cycle, from G0, G1, S, to G2/M. Across

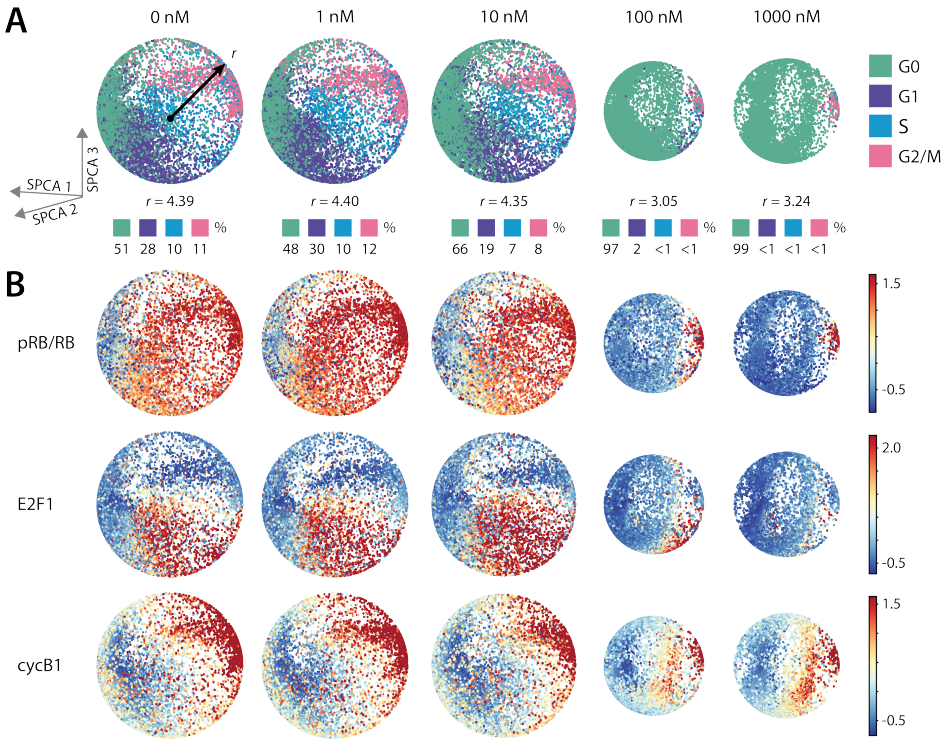

Fig. 2. **SPCA captures shifts in cell cycle phases and regulators across treatment conditions**. Data points from each treatment condition were projected onto three-dimensional hyperspheres identified by SPCA in a shared subspace. Points are colored according to (A) their cell cycle phase label or (B) normalized expression level of pRB/RB, E2F1, or cycB1.

all conditions, cells belonging to the same cell cycle phase were located near each other and within distinct regions along the surface of the spheres. We observed the cell cycle phases to be evenly distributed and occupy the same regions in the 0 nM, 1 nM, and 10 nM projections. At 100 nM and 1,000 nM, we saw an increase in the proportion of G0 cells concentrated mainly in the western hemisphere and along the vertical center axis, notably in the direction of cells in proliferative cell states. The small proportion of proliferative (G1, S, G2/M) cells was visible in a small region on the eastern hemisphere of the two manifolds. Additionally, the hyperspheres representative of cells treated with 100 nM and 1,000 nM had smaller radius sizes compared to the hyperspheres of lower palbociclib doses. Thus, we observed a delineation between the lower ($\leq$10 nM) and higher ($\geq$100 nM) treatment conditions. For all figures, plots for all features are available in the supplement[a].

We next investigated the ability of SPCA to capture more gradual cell-to-cell transitions by inspecting changes in the expression of each of the 20 cell cycle regulators (Fig. 2B). The resulting plots visually recapitulated known trends in protein expression levels for every feature. A high ratio of pRB to RB (pRB/RB) is needed to transition past the restriction point in late G1 to S phase. RB, or retinoblastoma protein, is hypophosphorylated by CDK4/6 and cyclin D1 complexes and hyperphosphorylated by cyclin E-CDK2.[23] Therefore, we expected cells in G0 and early G1 to have relatively lower pRB/RB values. In the G0 and G1 regions of the $\leq$10 nM palbociclib-treated cells, we observed an increasing gradient of pRB/RB values.

The cells with the highest pRB/RB ratios aligned with cells in the G2/M regions (Fig. 2A, B). Known trends were also observed for E2F1 and cyclin B1.[24,25] The highest values of E2F1 were located in S and nearby G1 regions while G2/M and bordering S phase cells expressed the highest values of cyclin B1. The 100 nM and 1,000 nM hyperspheres revealed more nuanced results. Compared to the lower treatment conditions, the G0 cells in the two highest treatment conditions expressed the lowest amounts of pRB/RB and E2F1, visible by the contrast in the color intensities of the G0 regions. Although the expression of pRB/RB and E2F1 decreased to a more extreme state, cyclin B1 followed a different trend. A subset of G0 cells demonstrated low expression of cyclin B1 while another group, most notably under the 1,000 nM palbociclib dose, had higher levels of cyclin B1 nearing those characteristic of proliferative G2/M cells.

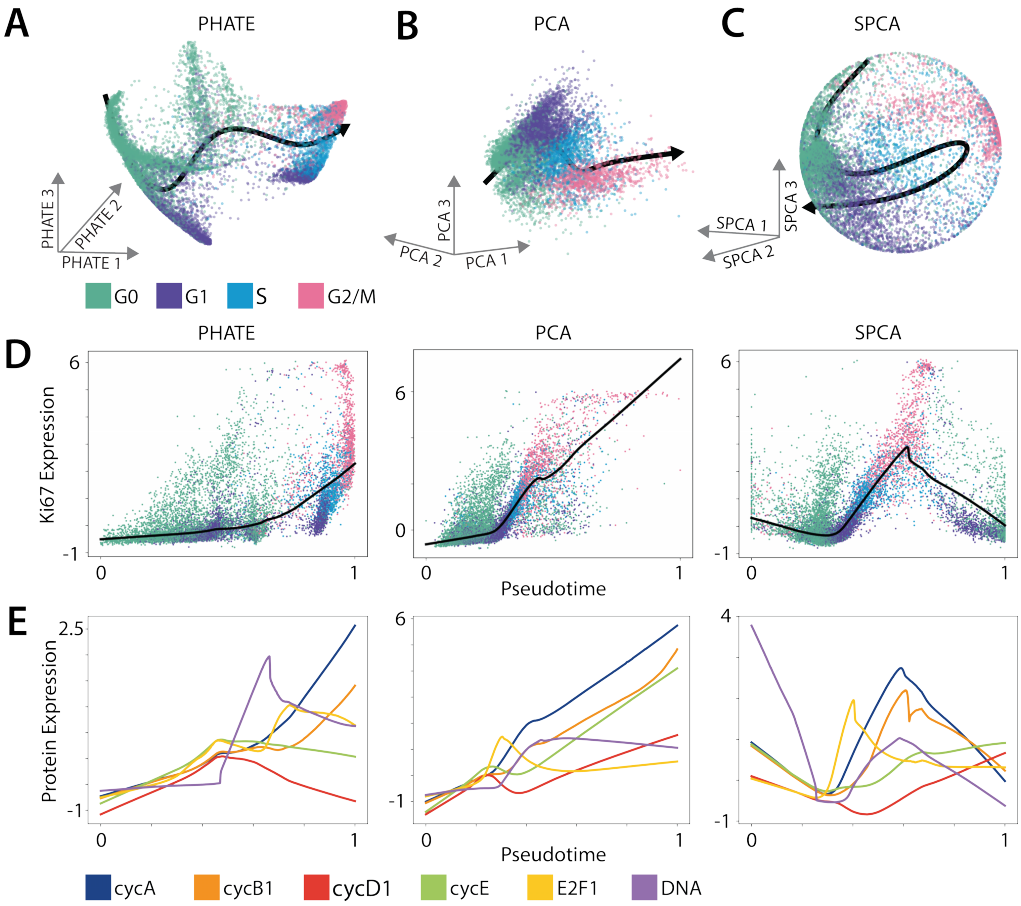

Fig. 3. **SPCA recapitulates cyclical protein level trends**. (A) PHATE, (B) PCA, and (C) SPCA were performed on untreated cells (0 nM palbociclib). Data points from each manifold learning method were plotted in three dimensions and colored according to their cell cycle phase annotation. Trajectories identified by Slingshot (black line) were overlaid onto their respective plots. (D) Ki67 expression of each cell was plotted according to the cell's normalized Slingshot pseudotime. A LOESS curve (black line) was fit through the points for each method. (E) LOESS curves fit through points plotted according to Slingshot pseudotime and median levels of core cell cycle regulators (cyclin A, cyclin B1, cyclin D1, cyclin E1, E2F1, and DNA content) were overlaid for PHATE, PCA, and SPCA.

Comparing Cell Cycle Structures from PHATE, PCA, and SPCA

SPCA successfully captured the overall structure of the cell cycle and known proteomic trends. To validate the effectiveness of SPCA as a representative tool for modeling the cell cycle, we compared SPCA to two other manifold learning methods, PHATE and PCA. Unlike SPCA, PHATE does not allow for the projection of multiple datasets to a shared space. PCA does have this capability but some uninterpretable alignment of principal component spaces is required. Due to these limitations, both PHATE and PCA were performed on the entire dataset such that all treatment conditions could be compared on a uniform scale. First, we examined the distribution of cell cycle phases in untreated cells using all three methods (Fig. 3A-C). Overall, we found cells belonging to the same cell cycle phase to be concentrated in the same region. However, the separation of phases varied. The least visual separation of phases was observed in the structure for PCA. In the projections produced by PHATE and SPCA, we saw greater separation between phase regions. We also observed a separation within phases in the PHATE manifold, specifically in G0 and G1, showing a discontinuous progression of phases. G0 cells occupied three main arms in one region of the PHATE structure while G1 cells were clustered in one of two regions on opposite sides of the manifold. One G1 cluster was located along an arm of the structure shared with G0 cells and the other group bordered the S phase region. Upon visual inspection, all manifolds suggested a canonical ordering of cell cycle phases. To assess how well these structures represented temporal trends, we performed Slingshot,[19] a trajectory inference method, to infer cell cycle paths. For each method, Slingshot identified a single trajectory through the canonical ordering of cell cycle phases - G0, G1, S, and G2/M - when provided a starting phase of G0. However, while the trajectories identified using PHATE and PCA proceeded in one direction from G0 to G2/M, the trajectory found from SPCA returned to the G0 and G1 regions, indicating a cyclic pattern (Fig. 3D).

Using the normalized pseudotime assigned to each cell, we next examined how expression levels of each feature fluctuated throughout the identified cell cycle paths (Fig. 3D). Ki67 is a key proliferative marker that accumulates over the course of the cell cycle reaching a peak in G2 and M.[26,27] Temporal orderings of cells identified for PHATE, PCA, and SPCA all followed an increasing trend of Ki67 expression (Fig. 3B). While each method ordered cells from G0, G1, S, to G2/M, cells in the SPCA pseudotime ordering returned to a state of G0 or G1 following the G2/M phase. Similarly, Ki67 expression decreased to a level consistent with that of the initial G0 cells. Cells were also more evenly distributed over pseudotime time for SPCA compared to PHATE, which had a separation between arrested and proliferating cells, and PCA, which had a majority of cells concentrated in the first half of the trajectory. Thus, SPCA successfully captured the gradient of protein accumulation we observed for key cell cycle regulators (Fig. 2, S3)[a] whereas PHATE and PCA identified less continuous trends.

We next asked how overall feature trends followed known accumulation patterns of key cell cycle regulators, specifically cyclins, E2F1, and DNA (Fig. 3E).[24,25,28] Expression of these cell cycle markers follows a cyclical pattern and aligns with key molecular events. Only cyclin A and cyclin B1 trends for PHATE and PCA as well as E2F1 trends for PCA aligned with expected points of accumulation, whereas all trends, except for cyclin E, identified using SPCA followed known expression patterns. For PHATE and PCA, the majority of feature trends

followed strictly increasing patterns. Cyclin D1 for PHATE and E2F1 for PCA experienced a peak in expression and a decrease, returning near initial expression levels. DNA content trends for the two methods, and cyclin E for PHATE peaked and revealed a more subtle decrease. All features for SPCA demonstrated a cyclic pattern such that final expression levels nearly matched initial levels, except for DNA content which had a lower final expression than the G0 cells identified to be at the beginning of the cell cycle path.

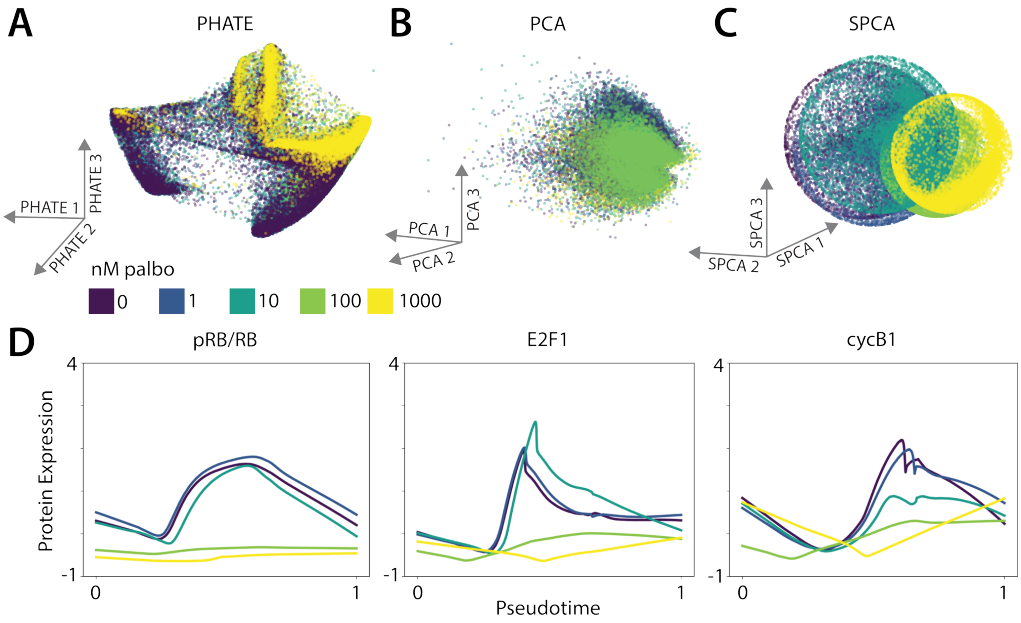

Fig. 4. **SPCA captures dose-dependent shifts in cell cycle manifold**. (A) PHATE and (B) PCA were performed using 20-feature single-cell signatures from all treatment conditions. (C) SPCA was performed for individual treatment conditions and the data points were projected onto their respective hyperspheres in a shared space identified by performing SPCA using all cells. Points are colored according to palbociclib dose for each individual cell. (D) LOESS curves were fit through points plotted according to Slingshot pseudotime generated using SPCA and median protein expression levels (pRB/RB, E2F1, cycB1) across five treatment conditions.

In response to palbociclib treatment, a greater proportion of cells become arrested (Fig. 2A). Therefore, we expect cells treated with different doses of palbociclib to reflect differences in the makeup of cell states and behaviors. PCA showed minimal delineation between treatment conditions (Fig. 4B) whereas PHATE and SPCA structures (Fig. 4A, C) captured differences between $\leq$10 nM and $\geq$100 nM palbociclib-treated cells. Cells belonging to the 100 nM and 1,000 nM treatment conditions concentrated along the arms of G0 cells of lower treatment conditions in the PHATE structure (Fig. 3A). Interestingly, cells of higher treatment conditions did not concentrate in the areas occupied by G0 cells in spheres of $\leq$10 nM doses. Instead, in addition to shrinking in size, SPCA spheres for 100 nM and 1,000 nM migrated in the direction of proliferative cell states. When we compared feature expression trends over pseudotime across treatment conditions for each method, SPCA more accurately captured cell cycle trends (Fig. 2B) and characterized behaviors expected of a dose response (Fig. 4D).

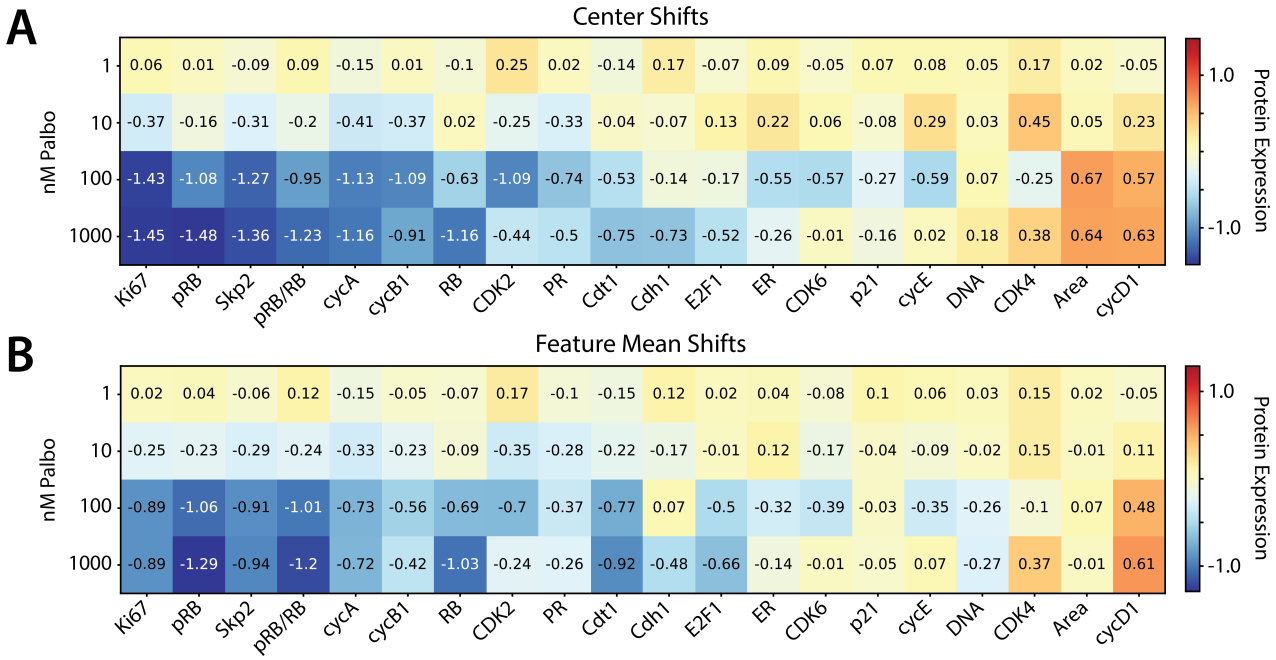

Fig. 5. **Shifts in centers of SPCA hyperspheres reveal changes in cell cycle regulation across treatment conditions**. (A) Shifts in the centers and (B) mean protein abundance of three-dimensional hyperspheres identified by SPCA for each treatment condition were calculated from each dose response to the untreated condition. Results from each pairwise comparison are represented in each row of the heatmaps.

SPCA Elucidates Mechanisms of CDK4/6 Inhibitor Resistance

To identify which specific factors were driving shifts in cell cycles across treatment conditions, evident by visual observations of feature expression differences (Fig. 2B) and the shift in positions of the 100 nM and 1,000 nM SPCA structures from the lower treatment conditions (Fig. 4C), we compared centers of the spheres. We quantified center shifts by subtracting the 20-feature center for 0 nM from the centers of each treatment condition. There was a clear distinction in protein levels between cells treated with lower (1 nM and 10 nM) and higher (100 nM and 1,000 nM) doses of palbociclib (Fig. 5A). Notably, we found a more significant depletion of proteins including Ki67, pRB, Skp2, cyclin A, cyclin B1, and RB, and enrichment of CDK4, cell area, and cyclin D1 in higher treatment conditions. Overall, the same trends were identified by comparing differences in feature means (Fig. 5B). However, the differences between the treatment groups were not as substantial, specifically for Ki67, Skp2, cyclin A, cyclin B1, and cell area which showed almost no change in mean expression across treatment conditions. The greatest depletion was found in pRB expression while cyclin D1 accumulation was the highest among all cell cycle regulators according to mean expression shifts from 100 nM and 1,000 nM to untreated cells. CDK4 expression was the second most elevated protein according to mean expression. Although CDK4 was also enriched according to center shifts between 1,000 nM and 0 nM treatment groups, CDK4 expression peaked in the 10 nM center shift as opposed to in the 1,000 nM mean expression shift. A similar pattern was found for cyclin E which was also elevated in the 10 nM to 0 nM comparison of centers, suggesting

an increase in cyclin E expression in 10 nM palbociclib-treated cells, while mean expression values revealed the opposite.

Because a majority of cells under 100 nM and 1,000 nM palbociclib treatment were arrested in G0 (Fig. 2A) and Slingshot identified a cyclical return to G0 cells in SPCA's cell cycle trajectory (Fig. 3C, D), we wanted to determine if there were differences between these groups of G0 cells. G0 cells were partitioned according to the median pseudotime of cells in a treatment condition. We will refer to the group of G0 cells with a pseudotime less than the median pseudotime value as 'early G0' and the remaining G0 cells bordering G2/M phase as 'late G0'. When we compared the proteomic signatures of early and late G0 cells, we observed notable differences between the groups (Fig. 6). For ≤100 nM doses, we observed higher expression of cell cycle regulators Ki67, Skp2, RB, CDK2, PR, Cdt1, Cdh1, ER, CDK6, p21, and cyclin E, and a decrease in cell area in late G0 cells. The greatest contrast was observed between early and late G0 cells treated with ≤10 nM palbociclib. An opposite trend was observed for the 1,000 nM early G0 cells which had higher expression of cell cycle markers including cyclin D1, CDK4, and cyclin E compared to the late G0 cells for 1,000 nM and other early G0 cells.

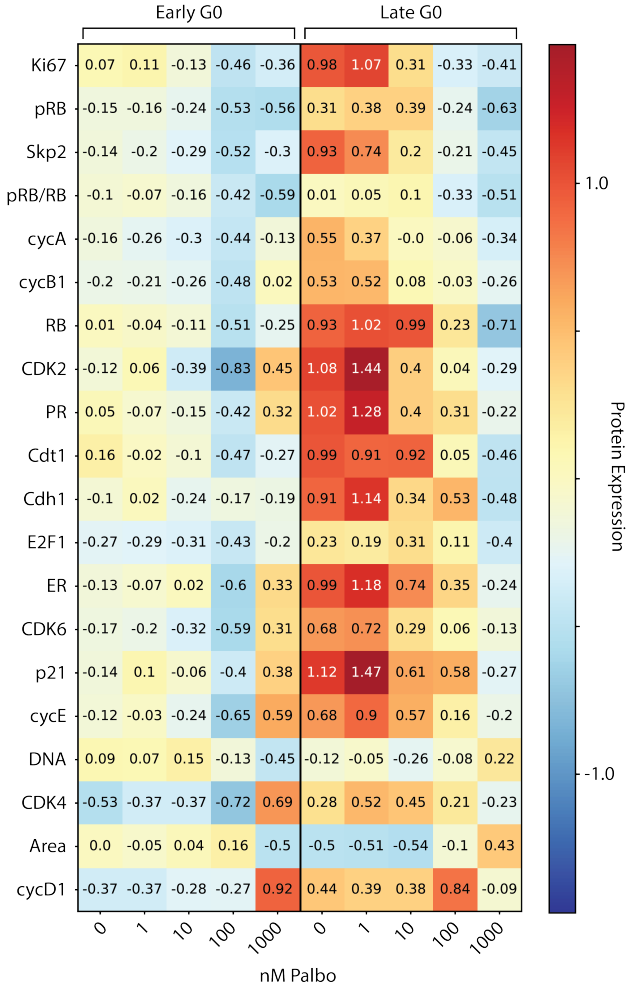

Fig. 6. **SPCA and Slingshot identify differences in G0 cells**. G0 cells were separated according to median normalized pseudotime. Mean protein levels for each cell cycle feature are represented in each row of the heatmap.

5. Discussion

We validated SPCA as a tool for characterizing cell cycle plasticity of breast tumor cells in response to palbociclib treatment. SPCA recapitulated the underlying cyclical structure of multiplex, single-cell breast tumor data and enabled direct visual and quantitative comparisons across treatment conditions. SPCA captures heterogeneity of molecular states by preserving fundamental differences between stages of the cell cycle, shown by the delineation of each cell cycle phase, while revealing gradual transitions in protein expression patterns. In addition to the continuous progression of cell states, the even distribution of phases and cells across the spherical manifolds suggests that the structure is representative of cell cycle data. Other

methods such as PHATE and PCA do not allow for as much flexibility in quantitative analysis or comparison of treatment conditions compared to SPCA. These methods failed to capture the cyclical nature of phase progression and protein expression.

Furthermore, we note that the spheres characterizing cells in the higher treatment group reveal differences from lower treatment conditions that are not observed by PHATE or PCA. The decrease in radius size, paired with a shift in center location, and skewed distribution of phases suggests that cells experience a fundamental shift in their cell cycles at the 10 nM and 100 nM transition. The greater proportion of G0 cells and smaller radius size of the 100 nM compared to the 10 nM sphere indicate less heterogeneity in cell cycle states. The overall shifts in the positions of the hyperspheres and the migration of 100 nM and 1,000 nM palbociclib-treated G0 cells towards proliferative regions in lower treatment conditions also suggest that cells under higher dosage have different mean states and, thus, traverse alternative paths through the cell cycle. The dichotomy between $\leq$10 nM and $\geq$100 nM hyperspheres aligns with prior knowledge that the IC50 for palbociclib lies within this range.[29]

SPCA allows us to quantitatively assess this difference between low and high treatment conditions via a comparison of each hypersphere's 20-feature center coordinates. These structural characteristics of SPCA manifolds can reveal trends that cannot be realized by looking at feature expression alone. A decrease in the hypersphere radius size along dose increases (Fig. 2A), indicates a reduction in heterogeneity of cell states. Center shifts pointed to further depletion of cell markers (Ki67, Skp2, cyclin A, cyclin B1) compared to mean expression, but also an increase in cell area and DNA which were found to remain consistent (cell area) or be downregulated (DNA) according to mean expression under increasing palbociclib treatment. Conversely, Cdt1 was found to be one of the highest-ranking features with decreased mean expression in higher treatment groups, but this difference was not as prominent when examining center shifts. These differences highlighted by radius and center shifts may indicate which cell cycle regulators are most responsible for driving changes in cell cycle behavior, but future experiments will need to be done to validate this hypothesis.

Structural differences that allowed for the identification of cyclical cell cycle trajectories with SPCA, but not PHATE or PCA, are also worth further investigation. Differences in early and late G0 cells suggest greater heterogeneity of multiple molecular states within cells categorized as G0. These differences may suggest that improved methods of cell cycle phase annotation need to be performed and that our framework of using SPCA and Slingshot could be used as a tool for differentiating between cell states. For example, $\leq$10 nM late G0 cells with high expression of proliferative markers, but low cell area could suggest that these are new daughter cells. However, differences in early and late G0 cells may indicate true differences in G0 cells or CDK4/6 inhibitor resistance. Cells with low proliferation markers, such as $\leq$100 nM early and 1,000 nM late G0 cells, may indicate varying depths of quiescence.[21,30–32] Overexpression of cyclin D1 and elevated levels of CDK2, shown in late 1,000 nM G0 cells in Fig. 6 across doses, has previously been found to be a potential mechanism of CDK4/6 inhibitor resistance via formation of cyclin D1-CDK2 complexes.[6,33,34] Cyclin E overexpression and constitutive activation is another characteristic of breast tumor cell behavior and an indicator of CDK4/6 inhibitor response.[35–37] Thus, these cell profiles can be used to characterize cells

and identify potential mechanisms involved in treatment-resistant pathways.

In future work, this pipeline can be utilized for multi-modal precision medicine. Rather than estimating hyperspheres for each treatment dose to compare, we may estimate hyperspheres for individual tumors across a patient population. In this way, we can compare an individual's cancer progression and resistance, and find personalized biomarkers for clinical targeting. Furthermore, though we have demonstrated the use of these innovative computational and statistical techniques on a single-cell breast tumor dataset, this framework can be extended to other biological contexts. SPCA can be generalized to study not only disease responses along the cell cycle, but single-cell responses to other forms of perturbation as well, including stem cell differentiation pathways. Other cyclical biological processes such as circadian rhythm and weather patterns can also be studied, leveraging the inherent underlying structures of these data, although prior knowledge or assessment that the data is spherical, which was established for cell cycle data based on extensive study,[211417] is needed. This novel framework for modeling cyclical biological data can allow for the rapid identification and quantification of novel trends in responses to forms of perturbations to biological systems.

Acknowledgments D.L. was supported by NIH grants R01 AG079291, R56 LM013784, R01 HL149683, and UM1 TR004406. T.M.Z. was supported by NIH F31HL156464. This work was also supported by grants NSF-2242980 (J.E.P.), R01-GM138834 (J.E.P.), and R01-CA280482 (J.E.P.).

References

1. I. Gomes, C. Abreu, L. Costa and S. Casimiro, The evolving pathways of the efficacy of and resistance to cdk4/6 inhibitors in breast cancer, *Cancers* **15**, p. 4835 (2023).
2. S. J. Altschuler and L. F. Wu, Cellular heterogeneity: do differences make a difference?, *Cell* **141**, 559 (2010).
3. H. X. Chao, R. I. Fakhreddin, H. K. Shimerov, K. M. Kedziora, R. J. Kumar, J. Perez, J. C. Limas, G. D. Grant, J. G. Cook, G. P. Gupta *et al.*, Evidence that the human cell cycle is a series of uncoupled, memoryless phases, *Molecular systems biology* **15**, p. e8604 (2019).
4. C. Liu, Y. Konagaya, M. Chung, L. H. Daigh, Y. Fan, H. W. Yang, K. Terai, M. Matsuda and T. Meyer, Altered g1 signaling order and commitment point in cells proliferating without cdk4/6 activity, *Nature Communications* **11**, p. 5305 (2020).
5. H. W. Yang, S. D. Cappell, A. Jaimovich, C. Liu, M. Chung, L. H. Daigh, L. R. Pack, Y. Fan, S. Regot, M. Covert *et al.*, Stress-mediated exit to quiescence restricted by increasing persistence in cdk4/6 activation, *Elife* **9**, p. e44571 (2020).
6. T. M. Zikry, S. C. Wolff, J. S. Ranek, H. M. Davis, A. Naugle, N. Luthra, A. A. Whitman, K. M. Kedziora, W. Stallaert, M. R. Kosorok *et al.*, Cell cycle plasticity underlies fractional resistance to palbociclib in er+/her2- breast tumor cells, *Proceedings of the National Academy of Sciences* **121**, p. e2309261121 (2024).
7. S. Yu, T. Kim, K. H. Yoo and K. Kang, The t47d cell line is an ideal experimental model to elucidate the progesterone-specific effects of a luminal a subtype of breast cancer, *Biochemical and Biophysical Research Communications* **486**, 752 (2017).
8. G. Gut, M. D. Herrmann and L. Pelkmans, Multiplexed protein maps link subcellular organization to cellular states, *Science* **361**, p. eaar7042 (2018).
9. K. R. Moon, D. Van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. v. d. Elzen, M. J. Hirn, R. R. Coifman *et al.*, Visualizing structure and transitions in high-dimensional biological data, *Nature biotechnology* **37**, 1482 (2019).
10. W. Stallaert, K. M. Kedziora, C. D. Taylor, T. M. Zikry, J. S. Ranek, H. K. Sobon, S. R. Taylor, C. L. Young, J. G. Cook and J. E. Purvis, The structure of the human cell cycle, *Cell systems* **13**, 230 (2022).
11. I. T. Jolliffe, *Principal component analysis for special types of data* (Springer, 2002).
12. H. Hotelling, Analysis of a complex of statistical variables into principal components., *Journal of educational psychology* **24**, p. 417 (1933).
13. R. Xiang, W. Wang, L. Yang, S. Wang, C. Xu and X. Chen, A comparison for dimensionality reduction methods of single-cell rna-seq data, *Frontiers in genetics* **12**, p. 646936 (2021).
14. H. Luo, J. E. Purvis and D. Li, Spherical rotation dimension reduction with geometric loss functions, *Journal of Machine Learning Research* **25**, 1 (2024).
15. D. Li, M. Mukhopadhyay and D. B. Dunson, Efficient manifold approximation with spherelets, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **84**, 1129 (2022).
16. L. Deconinck, R. Cannoodt, W. Saelens, B. Deplancke and Y. Saeys, Recent advances in trajectory inference from single-cell omics data, *Current Opinion in Systems Biology* **27**, p. 100344 (2021).
17. D. Schwabe, S. Formichetti, J. P. Junker, M. Falcke and N. Rajewsky, The transcriptome dynamics of single cells during the cell cycle, *Molecular systems biology* **16**, p. e9946 (2020).
18. A. Riba, A. Oravecz, M. Durik, S. Jiménez, V. Alunni, M. Cerciat, M. Jung, C. Keime, W. M. Keyes and N. Molina, Cell cycle gene regulation dynamics revealed by rna velocity and deep-learning, *Nature communications* **13**, p. 2865 (2022).
19. K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom and S. Dudoit, Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics, *BMC genomics*

19, 1 (2018).
20. J. S. Ranek, W. Stallaert, J. J. Milner, M. Redick, S. C. Wolff, A. S. Beltran, N. Stanley and J. E. Purvis, Delve: feature selection for preserving biological trajectories in single-cell data, *Nature Communications* **15**, p. 2765 (2024).
21. W. Stallaert, S. R. Taylor, K. M. Kedziora, C. D. Taylor, H. K. Sobon, C. L. Young, J. C. Limas, J. Varblow Holloway, M. S. Johnson, J. G. Cook *et al.*, The molecular architecture of cell cycle arrest, *Molecular Systems Biology* **18**, p. e11087 (2022).
22. W. G. Jacoby, Loess:: a nonparametric, graphical tool for depicting relationships between variables, *Electoral studies* **19**, 577 (2000).
23. S. Kim, A. Leong, M. Kim and H. W. Yang, Cdk4/6 initiates rb inactivation and cdk2 activity coordinates cell-cycle commitment and g1/s transition, *Scientific reports* **12**, p. 16810 (2022).
24. J. R. Nevins, The rb/e2f pathway and cancer, *Human molecular genetics* **10**, 699 (2001).
25. C. Norbury and P. Nurse, Animal cell cycles and their control, *Annual review of biochemistry* **61**, 441 (1992).
26. E. Endl and J. Gerdes, The ki-67 protein: fascinating forms and an unknown function, *Experimental cell research* **257**, 231 (2000).
27. M. Sobecki, K. Mrouj, J. Colinge, F. Gerbe, P. Jay, L. Krasinska, V. Dulic and D. Fisher, Cell-cycle regulation accounts for variability in ki-67 expression levels, *Cancer research* **77**, 2722 (2017).
28. T. Evans, E. T. Rosenthal, J. Youngblom, D. Distel and T. Hunt, Cyclin: a protein specified by maternal mrna in sea urchin eggs that is destroyed at each cleavage division, *Cell* **33**, 389 (1983).
29. J. Bollard, V. Miguela, M. R. De Galarreta, A. Venkatesh, C. B. Bian, M. P. Roberto, V. Tovar, D. Sia, P. Molina-Sánchez, C. B. Nguyen *et al.*, Palbociclib (pd-0332991), a selective cdk4/6 inhibitor, restricts tumour growth in preclinical models of hepatocellular carcinoma, *Gut* **66**, 1286 (2017).
30. J. S. Kwon, N. J. Everetts, X. Wang, W. Wang, K. Della Croce, J. Xing and G. Yao, Controlling depth of cellular quiescence by an rb-e2f network switch, *Cell reports* **20**, 3223 (2017).
31. J. M. Lemons, X.-J. Feng, B. D. Bennett, A. Legesse-Miller, E. L. Johnson, I. Raitman, E. A. Pollina, H. A. Rabitz, J. D. Rabinowitz and H. A. Coller, Quiescent fibroblasts exhibit high metabolic activity, *PLoS biology* **8**, p. e1000514 (2010).
32. K. J. Soprano, Wi-38 cell long-term quiescence model system: A valuable tool to study molecular events that regulate growth, *Journal of cellular biochemistry* **54**, 405 (1994).
33. M. T. Herrera-Abreu, M. Palafox, U. Asghar, M. A. Rivas, R. J. Cutts, I. Garcia-Murillas, A. Pearson, M. Guzman, O. Rodriguez, J. Grueso *et al.*, Early adaptation and acquired resistance to cdk4/6 inhibition in estrogen receptor–positive breast cancer, *Cancer research* **76**, 2301 (2016).
34. M. Palafox, L. Monserrat, M. Bellet, G. Villacampa, A. Gonzalez-Perez, M. Oliveira, F. Brasó-Maristany, N. Ibrahimi, S. Kannan, L. Mina *et al.*, High p16 expression and heterozygous rb1 loss are biomarkers for cdk4/6 inhibitor resistance in er+ breast cancer, *Nature communications* **13**, p. 5258 (2022).
35. J. Gray-Bablin, J. Zalvide, M. P. Fox, C. J. Knickerbocker, J. A. DeCaprio and K. Keyomarsi, Cyclin e, a redundant cyclin in breast cancer, *Proceedings of the National Academy of Sciences* **93**, 15215 (1996).
36. P. W. Hinds, S. Mittnacht, V. Dulic, A. Arnold, S. I. Reed and R. A. Weinberg, Regulation of retinoblastoma protein functions by ectopic expression of human cyclins, *Cell* **70**, 993 (1992).
37. C. E. Caldon, C. M. Sergio, J. Kang, A. Muthukaruppan, M. N. Boersma, A. Stone, J. Barraclough, C. S. Lee, M. A. Black, L. D. Miller *et al.*, Cyclin e2 overexpression is associated with endocrine resistance but not insensitivity to cdk2 inhibition in human breast cancer cells, *Molecular cancer therapeutics* **11**, 1488 (2012).