

## MIT Open Access Articles

*Time series modeling of live-cell shape dynamics for image-based phenotypic profiling*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Gordonov, Simon et al. "Time Series Modeling of Live-Cell Shape Dynamics for Image-Based Phenotypic Profiling." *Integr. Biol.* 8.1 (2016): 73–90.

**As Published:** <http://dx.doi.org/10.1039/c5ib00283d>

**Publisher:** Royal Society of Chemistry

**Persistent URL:** <http://hdl.handle.net/1721.1/106798>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike





# HHS Public Access

Author manuscript

*Integr Biol (Camb)*. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

*Integr Biol (Camb)*. 2016 January ; 8(1): 73–90. doi:10.1039/c5ib00283d.

## Time Series Modeling of Live-Cell Shape Dynamics for Image-based Phenotypic Profiling†

Simon Gordonov<sup>a,b</sup>, Mun Kyung Hwang<sup>b</sup>, Alan Wells<sup>c</sup>, Frank B. Gertler<sup>b,d</sup>, Douglas A. Lauffenburger<sup>a,b,d,\*</sup>, and Mark Bathe<sup>a,\*</sup>

<sup>a</sup> Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>b</sup> The David H. Koch Institute for Integrative Cancer Research, Cambridge, MA, USA

<sup>c</sup> Department of Pathology, University of Pittsburgh, and Pittsburgh VA Health System, Pittsburgh, PA, USA

<sup>d</sup> Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

### Abstract

Live-cell imaging can be used to capture spatio-temporal aspects of cellular responses that are not accessible to fixed-cell imaging. As the use of live-cell imaging continues to increase, new computational procedures are needed to characterize and classify the temporal dynamics of individual cells. For this purpose, here we present the general experimental-computational framework SAPHIRE (Stochastic Annotation of Phenotypic Individual-cell Responses) to characterize phenotypic cellular responses from time series imaging datasets. Hidden Markov modeling (HMM) is used to infer and annotate morphological state and state-switching properties from image-derived cell shape measurements. Time series modeling is performed on each cell individually, making the approach broadly useful for analyzing asynchronous cell populations. Two-color fluorescent cells simultaneously expressing actin and nuclear reporters enabled us to profile temporal changes in cell shape following pharmacological inhibition of cytoskeleton–regulatory signaling pathways. Results are compared with existing approaches conventionally applied to fixed-cell imaging datasets, and indicate that time series modeling captures heterogeneous dynamic cellular responses that can improve drug classification and offer additional important insight into mechanisms of drug action.

### Introduction

High-content imaging (HCI) is widely used to perform quantitative *in vitro* cell phenotyping in a broad range of applications from RNAi and drug screening to prediction of stem cell differentiation fates<sup>1–4</sup>. In contrast to population-level assays that measure concentrations and activities of molecular species pooled over heterogeneous cellular populations, HCI has

†Electronic supplementary information available: Fig. S1 - S6; Table S1; Videos S1 - S3

\* Correspondence: Douglas Lauffenburger (lauffen@mit.edu), 77 Massachusetts Avenue, Building 16, Room 343, Cambridge, MA 02139 USA, Tel. +1-617-252-1629 | Fax. +1-617-258-0204 Or Mark Bathe (mbathe@mit.edu), 77 Massachusetts Avenue, Building 16, Room 255, Cambridge, MA 02139 USA, Tel. +1-617-324-5685 | Fax. +1-617-324-7554.

the advantage of profiling cells *in situ* in a manner that captures both overall cellular morphology as well as sub-cellular features such as protein localization and their relative levels<sup>5,6</sup>. Shape is the most common property used to characterize cellular phenotype in part due to the ease of image-based quantification enabled by cytoskeletal staining and the importance of morphology in a wide variety of cellular processes. In practice, fixed-cell imaging is typically performed because it avoids large-scale handling of live cultures during imaging or generation of fluorescent reporter cell lines, and enables quantification of large numbers of cells at a single time point, increasing statistical power for comparing cellular phenotypes across experimental conditions<sup>7,8</sup>. Multivariate statistical modeling of fixed-cell image features has been effective in phenotype-based drug classification, providing important insight into signaling pathways involved in cellular morphogenesis<sup>9,10</sup>. Single-cell analysis using imaging has been particularly instrumental in identifying and deciphering cellular phenotypes in disease states<sup>11</sup>. User-defined shape categories coupled with supervised learning such as support vector machines, as well as unsupervised methods such as principal component analysis (PCA), have been used to generate quantitative profiles for comparing experimental perturbations and inferring spatial signaling mechanisms of shape regulation<sup>12–15</sup>.

However, fixed-cell assays, while relatively simple to perform through fluorescent staining and imaging, suffer from several important limitations. Principal among these is the loss of information regarding cellular dynamics in response to long-term or transient drug treatments. In addition, imaging artifacts may occur due to cell fixation and permeabilization, which may distort spatially resolved protein distributions<sup>16</sup>. For these reasons, live-cell imaging is increasingly being used to characterize cellular phenotypes, particularly in the subcellular analysis of cell shape dynamics and polarization. For example, computational tools for cell boundary tracking<sup>17–19</sup>, morphodynamics profiling<sup>20–23</sup>, measurement of fluorescent reporters<sup>24,25</sup>, and quantitative morphology and subcellular protein distribution analyses<sup>26</sup> in live cells have become an integral component of high-resolution analyses of cell shape and its regulation, particularly in the context of cell migration. In cell migration studies, live-cell shape and signaling analyses have been complemented by direct quantification of motility properties such as cell speed and persistence of motion to establish links between molecular mechanisms and migratory phenotypes<sup>27–32</sup>.

In these applications, the relative strengths of high-resolution, live-cell imaging versus fixed-cell HCI assays are apparent: the former captures rich, dynamic properties of single-cell behavior while the latter enables large-scale screening of hundreds to thousands of cells. In an effort to bridge this gap, several mathematical approaches have been developed to infer dynamic properties of cell populations from fixed-cell measurements in HCI studies. For example, ergodic rate analysis based on differential equation modeling has been used to infer transition rates through cell cycle stages from images of molecular reporters that define various mitotic phases in individual fixed cells<sup>33</sup>. Additionally, Bayesian network modeling of shape parameters coupled with RNAi knockdown of cytoskeleton-regulatory proteins has been used to infer shape state transitions of migratory cells and reveal underlying regulatory signaling modules<sup>34,35</sup>. However, these approaches assume quasi-steady-state of the cell population, assign cells into pre-defined phenotypic categories, and, in the case of Bayesian

networks, face difficulties in modeling repetitive processes such as motility cycle stages in migrating cells. Moreover, they are not directly applicable to the analysis of live cells over time to monitor individual cellular responses to drug perturbations.

To address these limitations, here we present a live-cell HCI framework that captures the dynamics of a large number of cells on the scale of a phenotypic screen. The approach combines high-content live imaging, image processing, multivariate data analysis, and probabilistic modeling to characterize cell shape dynamics in a drug-screening context. Inspired by existing methods for modeling cell cycle stages from time series images<sup>36,37</sup>, our framework employs hidden Markov modeling to describe shape dynamics as temporal sequences of morphological states that are observed as noisy, multivariate image data. Describing temporal trajectories of multivariate shape measurements as a finite set of states that is limited in number by Bayesian model selection that penalizes model complexity, results in an efficient description of time-dependent shape categories explored by a cell. This approach provides a means of modeling the shape dynamics of hundreds of individual cells, capturing temporal evolution in cell morphology directly from live images without assumptions of steady-state cell populations or predefined shape categories assumed by fixed-cell analyses. We show here in a proof-of-principle study that drug-response profiles derived from these models can be used for phenotypic drug comparisons and can reveal spatially distinct and pathway-specific roles that drug-targeted species play in modulating shape dynamics. Our computational framework is available as open source for the computational cell biology community to apply using MATLAB.

## Results

We developed a framework to characterize single-cell shape dynamics for large collections of cells from live-imaging assays (Fig. 1). Fluorescent cells treated with a variety of drug perturbations are imaged in a multi-well format followed by image processing and analysis to track and outline individual cells over time. Image-derived features, or descriptors, of cell shape are extracted to generate a data cube for the entire imaging screen, with the three dimensions corresponding to shape features, individual cells, and time points. PCA is used for dimensionality reduction along the shape feature axes to generate temporal “shape-space” trajectories of individual cells in principal component (PC) space. HMM is then used to annotate each cell shape trajectory using a small set of shape states that are determined using Bayesian model selection, revealing important time-dependent features of cell morphology while substantially reducing live-cell data size and complexity. Features extracted from temporal model annotations of hundreds of cells are then used for clustering and classifying perturbations. In the following sections we detail the steps of the framework and present a proof-of-principle study in profiling the heterogeneity in cell shape dynamics from a small-scale drug screen.

### Live imaging enables temporally-resolved readout of cell phenotypes

We first generated a triple negative breast cancer cell line, MDA-MB-231, with stable expression of two fluorescent reporters, LifeAct-eGFP and histone H2B-mCherry, for concurrent imaging of actin and nuclear dynamics, respectively. This enabled the

unambiguous identification, isolation, and tracking of individual cells using the segmented nuclei, while the actin reporter enabled quantitative characterization of cytoskeletal morphology. MDA-MB-231 cells were chosen for their mesenchymal-like properties and lack of epithelial-type intercellular junctions, making this cell line particularly amenable to studying phenotypic behavior on the single-cell level. The two-color fluorescent reporter cells were seeded in 96-well plate format and treated with different drugs that target distinct components of the actomyosin cytoskeleton-regulatory pathways including ROCK (Rho-associated protein kinase), myosin II, EGFR (epidermal growth factor receptor), calpain, MLCK (myosin light chain kinase), and MEK (mitogen-activated protein kinase kinase), to probe their effects on cell shape dynamics in two independent imaging experiments (Materials and Methods).

In one experiment, we included two controls in order to set a baseline for cell behavior under no drug treatment: one of growth media only and the other with 0.1% v/v DMSO in growth media to assess effects of DMSO alone since it was used to dissolve drug stock. We also included two different doses for the ROCK and MLCK inhibitors to assess drug concentration effects on cell shape dynamics characterized in our framework. In an additional experiment, we expanded the panel of drug treatments and used a different microscope, culture media, and cell passage number to assess these factors on reproducibility of experiments and subsequent analyses. Epifluorescence 2-D microscopy using a 10X/0.3NA air objective offered a suitable tradeoff between field-of-view and resolution, enabling us to maximize the number of cells captured with sufficient detail to characterize individual cell morphologies over approximately eighteen hours for each treatment condition.

Visual inspection of the compiled time series movies revealed asynchronous populations of dividing, apoptotic, and migratory cell phenotypes (Videos S1-3). For cells not undergoing division or death, we observed a broad range of shape changes, with some cells changing shape rapidly with pronounced cell body protrusions and retractions, while others exhibited less noticeable morphodynamic activity. Although the cells generally exhibited autonomous behavior that was largely independent of one another, we observed frequent contact and spatial overlapping of some cells in addition to entry and exit of cells into and out of the imaging fields of view. Collectively, the large quantity of imaging data and phenotypic cell behaviors rendered fully manual data parsing and analysis intractable, thereby motivating the need for an automated computational analysis pipeline for processing the time-series images.

### **Automated processing with quality control ensures accurate parsing of image data**

A priority in formulating our probabilistic modeling approach was to ensure that any data input into the model accurately captures cell shape properties. We therefore developed an image processing pipeline to segment and track cells with as many automated, user-free steps as possible to increase throughput and minimize user subjectivity (Materials and Methods; Fig. 2A). Automated image parsing, cell segmentation, and tracking were first performed for all acquired image time series (see examples in Videos S1-3). A major challenge, however, involved the treatment of variations in image illumination between

fields and within fields over time, as well as variability in fluorescent reporter levels between cells, leading to occasional image processing errors such as over- or under-segmentation of cells (Fig. 2B). We therefore developed a GUI-based quality control (QC) module that enabled manual validation and modification of segmented regions and cell tracks. Here, user input is desirable because cell boundaries can be defined unambiguously by the human eye while automated segmentation is notoriously sensitive to image properties. This module was used to discern and correct artifacts introduced during image acquisition and automated processing steps. The QC module also enabled us to identify and label dividing and dying cells, which were removed from subsequent analyses (Fig. 2C).

### Drug treatments diversify morphologies of MDA-MB-231 cells

The automated image processing and QC pipeline resulted in a collection of accurately segmented and tracked cell trajectories. We next extracted a set of eighteen morphological features from segmentation masks of each cell and time point (Table S1). Following z-score normalization of each feature across cells, we applied PCA to the data in order to capture and visualize the variability in shape between all cell images from the imaging experiment with the expanded panel of drug treatments (Fig. 3A). Here, we call the basis of the projected data onto the first two PCs the “shape-space.” The first two PCs captured over 80% of the shape variability in the observations. In order to assess how PCA distributes and orients morphological properties of cells in shape-space, we selected random time snapshots from four different cells in seven regions of PC space and visualized the overlaid actin and nuclei images (Fig. 3B). This visualization revealed that morphologies of MDA-MB-231 cells across all seven drug treatments vary from large and spread (Fig. 3B, panel d), to round cells with pronounced cortical actin at the cell periphery (Fig. 3B, panels e and f), as well as polarized cells with varying degrees of elongation (Fig. 3B, panels g, a, b), and branched morphologies (Fig. 3B, panel c).

To visualize how morphologies vary in different directions and radial distances away from the data mean in shape-space, we created a polar representation of the two PCs. The polar shape-space was subdivided into twelve equal angular bins, with each bin subdivided further into quartiles. Fig. 3C shows the representative cell shapes in each radial quartile of the angular bins as well as the contributions of the original eighteen shape features to the values of the two PCs. The highest variance in shape, along the first PC, predominantly captures cell elongation and branching, while the second PC captures cell spreading area and roundness. Features such as major axis length, geodesic diameter, and maximum Feret length capture similar length properties of cells and therefore have similar PC coefficients and directions in shape-space. These features are anti-correlated with solidity, circularity, and extent, which point in the opposite direction for both PCs, showing that branched and elongated morphologies have smaller solidities and circularities, as expected. As captured mainly by the second PC, cells with longer minor axes and larger areas have generally lower eccentricities and smaller ratios of major to minor axis lengths.

Furthermore, we found that cells are densely packed around the data mean at the origin in shape-space, with the point density dissipating radially (Fig. 3A). This suggests that although MDA-MB-231 morphologies are visually distinct, shape properties of the

population as a whole vary on a continuum, with no clearly distinguishable groups. Interestingly, however, when we plotted cell shape trajectories one at a time we found increased grouping and clustering of points. We applied  $k$ -means clustering to individual cells and to the cell population as a whole for  $k = 2$  to  $k = 10$  clusters, and computed average silhouette values in order to quantify the degree to which points form well separated, or cohesive, groups (Fig. 3D). As may visually be discerned in Fig. 3A, example cell trajectory  $i$  forms 2 groups while cell trajectory  $j$  forms 3 groups, which is confirmed by the maximum silhouette value that occurs for  $k = 2$  and  $k = 3$  clusters, respectively. On average, points from individual cell trajectories had higher cluster cohesion than did random samples with the same number of points as the trajectories or the entire cell population as a whole.

### Probabilistic modeling of morphological states reveals heterogeneity in cell shape dynamics

Having observed that individual cell trajectories form more cohesive clusters than the population as a whole, we next sought to develop a principled and reproducible means of modeling morphological dynamics on a single-cell basis. We reasoned that the higher cluster cohesiveness within individual trajectories signifies the presence of underlying “states” that the cell explores in shape-space over time. Although Gaussian mixture modeling (GMM) is a highly useful approach for unsupervised, model-based data clustering that would eliminate subjective, user-defined delineation of states<sup>37</sup>, it typically ignores the temporal nature of data. Including temporal information during model inference is reasonable *a priori* because GMM inherently assumes that observations are independent<sup>38</sup>, which is not necessarily satisfied for sequential time series measurements from the same cell, whose shape may be highly correlated in time. We therefore applied an HMM framework in which temporal dependencies are directly incorporated during model inference, using a modified approach developed for annotating modes of single-particle motion in live cells<sup>39</sup>. Within this HMM framework, the “hidden” underlying states correspond to cell shape states that produce observable emissions that are associated with the measurable cell shapes computed in PC shape-space (Materials and Methods). Additionally, Bayesian model selection enables us to penalize models with greater numbers of underlying shape states to fit the set of time series points satisfying Occam’s razor or the Principle of Parsimony<sup>40,41</sup>.

To compare our approach with GMMs that ignore dependencies in time series data, we simulated shape-space trajectories from underlying hidden Gaussian states to explore differences with HMM-based annotation using SAPHIRE. Specifically, we sought to compare state identification capability of the two methods, with SAPHIRE using Bayesian model selection and GMM using the Bayesian information criterion to penalize model complexity. Our two-state model simulations revealed that SAPHIRE is more likely to choose the model with the correct number of states and has better accuracy in inferring locations of state means compared with GMM due to SAPHIRE’s incorporation of temporal dependencies in the data during the state inference steps (Fig. S1).

Next, we applied SAPHIRE to experimental observations of cells in the two-dimensional PC shape-space to infer a time series model for each cell individually. The inferred model specifies the most probable number of hidden shape states, which is unknown *a priori*, as



well as the state parameters. Each state is a symmetric, bivariate Gaussian distribution in shape-space with two variables corresponding to the two PCs. The inferred parameters for each Gaussian distribution are the standard deviation and the mean, which capture temporal shape variation and average morphology of the cell within the state, respectively. The model also specifies a state transition probability matrix that describes the probability of transitions between shape states. Each cell is annotated in time with the most probable, or maximum likelihood, shape state sequence, enabling us to determine the most likely state that the cell exists in at each point in time. We first compared shape state inference of our approach with that of commonly-used GMM under different model constraints. Allowing for elliptically-shaped Gaussian states in the GMM led to undesirable grouping of cell shapes with visually-dissimilar morphologies into the same state (Fig. S2), and, similar to the results of the numerical simulations (Fig. S1), SAPHIRE was better able to capture distinct morphological phases of cells that gradually move between distinct regions of shape-space (Fig. S3). In certain cases of cells moving between two visually distinguishable underlying states, the state annotations of cell trajectories were similar for SAPHIRE and GMM (Fig. S4). Moreover, from all modeled cell trajectories we found that a larger fraction of cells exists in two to three underlying states, while fewer cells explore either a single state or four or more states (Fig. S5).

Despite some similarities in the numbers of states explored by cells, the inferred parameters of transition dynamics and state annotations varied considerably between cells. These variations not only existed for cells across different drugs, but also across cells within a given treatment and for cells with the same number of shape states. Fig. 4 shows the inferred shape states for two example cells treated with the same drug having the same number of inferred states, but with notable heterogeneity in state parameters and transition dynamics. Some cells exhibited rapid back-and-forth switching between states corresponding to periodicity in actin protrusions and retractions, resulting in shapes that are elongated, rounded, or those in between (Fig. 4A). Other cells exhibited individual instances of state transitions, progressively changing shape in a given direction in shape-space, such as going from larger and more spread to smaller and rounder morphologies (Fig. 4B). The maximum likelihood state annotations capture different phases in the morphodynamic history of a cell, with state transition parameters derived from the state sequence providing useful insights into its dynamic behavior. For instance, the model for the cell in Fig. 4A reveals that the cell is more likely to stay in state 1 ( $p = 0.19$ ) or state 2 ( $p = 0.29$ ) than transition to other states, and that direct transitions between elongated (state 1) and a rounded (state 3) morphology without going through an intermediate shape (state 2) is not likely ( $p = 0$ ).

### Shape state annotations serve as phenotypic readouts of drug action

Our probabilistic modeling framework produced annotated sequences of shape dynamics for all individual cell trajectories from the inhibitor screen, with each trajectory comprised of morphological states evolving in PC shape-space over time. We next explored how shape dynamics compare between cells treated with the distinct compounds. We first generated phenotypic signatures from the annotated state sequences for each cell. These signatures capture where in PC shape-space states are located, how long a cell spends in its inferred



states, and how frequently and in what directions in PC space a cell makes shape transitions (Fig. 5A).

We next assessed how the distribution of cellular shape states in polar PC space was affected by the experimental treatment conditions (Fig. 5B). Cells in the DMSO and growth media controls were positioned closely and were fairly evenly distributed around the mean of the shape-space data without exhibiting biased morphologies in any particular direction compared with the morphologies induced by drugs. The ROCK and myosin II inhibitors pushed cells predominantly towards highly elongated and branched morphologies with longer dwell times in these states. On the other hand, MLCK inhibition had the opposite effect, biasing cells toward smaller and rounder morphologies along the negative first PC axis. In the experiment with the expanded panel of drugs, both MEK inhibitors tested led to a broader variety of states, predominantly either rounder, or more elongated, shapes, although no strong biases of shape state location or state dwell time were observed upon MEK inhibition (Fig. S6A). Similarly, EGFR inhibition led to a broader and relatively more uniform distribution of morphological states around shape-space, albeit with a noticeable shift towards less elongated morphologies, whereas calpain inhibition biased cells into more elongated morphologies. Further, doubling the dose of the myosin II inhibitor shifted cells slightly towards more branched morphologies while diminishing elongated states, whereas an increase in dose of the MLCK inhibitor appeared to accentuate the bias towards smaller and more rounded shapes (Fig. 5B). For either drug, however, the overall distributions of cellular shape states and state dwell times were similar across both doses tested, as well as between the two imaging experiments.

Live-cell analysis also enables the determination of whether drugs differentially affect the trajectories that cells take through shape-space when they transition between distinct morphological states. To our surprise, state transition directions were similar across the drugs tested, with cells moving between two relatively narrow angular ranges, 120 to 180 degrees or 300 to 360 degrees in shape-space (Fig. 5C; Fig. S6B) that correspond to decreased elongation with increased roundness versus increased elongation and decreased roundness, respectively (Fig. 3C). This finding suggests that most of the dynamics in shape that MDA-MB-231 cells undergo are along this morphological axis of increasing or decreasing elongation, regardless of drug treatment, which may be indicative of cytoskeletal protrusion and retraction cycles, while transitions toward larger, spread morphologies, for example, are quite rare for this cell type. Despite the similarity in the directions of state transitions taken by cells across drugs, ROCK and myosin II inhibition led to larger magnitudes and slightly broader distributions in state transitions, suggesting that these drugs induce more pronounced variations along the roundness-elongation axis of cell shape, likely by reducing transcellular actomyosin tension.

Moreover, readouts of shape state locations and state transitions can be used to group treatments based on similarities in induced phenotype dynamics. We derived a dynamic shape state “profile” for each treatment from the shape state and state transition histograms (Fig. 5B and 5C; Fig. S6A and S6B) and hierarchically clustered the treatments using profile similarities for each imaging experiment separately (Fig. 6A and 6B). As anticipated based on the preceding results, the two MEK inhibitors clustered closest together, as did the

DMSO and growth media controls. These results serve as internal controls to validate the HMM annotation and phenotypic drug comparison procedure proposed here. The MEK inhibitors were also found to have dynamic shape state profiles more similar to those of the EGFR and calpain inhibitors than to the MLCK, ROCK, and myosin II inhibitors. Myosin II inhibition with Blebbistatin produced phenotypes most similar to those with ROCK inhibition. Perhaps surprisingly, MLCK inhibition, which is known to alter myosin II activity, induced shape dynamics more similar to those under MEK and EGFR inhibition than under myosin II and ROCK inhibition. Overall, the computational analyses from these imaging experiments demonstrate that temporal dynamics of individual cells can be combined into quantitative profiles that serve as useful readouts for phenotypic drug comparison and for inferring shape-regulatory roles of targeted signaling molecules.

### **State-space temporal modeling of cell morphology improves drug classification over existing image-based profiling methods**

SAPHIRE generates phenotypic profiles of experimental treatments from single-cell models of HMM-annotated morphological state trajectories in shape-space. This approach differs from existing fixed-cell HCI profiling methods, which, instead of modeling cellular properties of a given cell over time, characterize cellular properties for different cells in a population at a given time point. Therefore, we sought to assess the benefit that live-cell temporal modeling on a per-cell basis has in comparison with fixed-cell approaches<sup>10,42–44</sup> in classifying treatment conditions in our screen (Fig. 7). We chose four profiling methods that have previously been compared amongst one another and have shown high drug classification accuracies in a large-scale drug screen study (see<sup>15</sup> and Materials and Methods). To make a fair comparison with live-cell analyses, we used live-cell population data from our inhibitor screen at five different time points post-treatment for the fixed-cell analysis methods in order to mimic a fixed-cell time course experiment. We additionally profiled the dynamics of individual cells but without the HMM that is used in SAPHIRE to assess the value of the HMM state-space shape representation for classifying treatments.

For the three treatments tested that were amenable to the classification analysis approach as in<sup>15</sup>, of the four fixed-cell methods the “Factor Analysis + Means” profiles performed best, correctly classifying 5 of 6 treatments (Fig. 7A and 7B). This approach has also previously been shown to have good classification performance in a large-scale screen of dozens of drugs from different mechanistic classes<sup>15,44</sup>. Profiles generated using Gaussian mixtures<sup>43</sup> and the K-S statistic<sup>10</sup> correctly classified 4 of 6 treatments, whereas the simplest “Means”<sup>42</sup> approach only classified 2 of 6 treatments correctly. Treatment profiles generated using SAPHIRE resulted in correct classification of all 6 treatments, demonstrating improvement over the fixed-cell profiling methods tested. Moreover, the HMM annotation of morphological states using SAPHIRE appears to be critical for improving correct treatment classification, as profiles derived from single-cell temporal dynamics without the HMM misclassified 2 of the 6 treatments.

To further understand which properties of the state-space treatment profiles generated using SAPHIRE improve treatment classification, we separately considered the use of features that capture state transitions versus features that only consider properties of the states themselves

(Fig. 7C). Treatment classification confusion matrices using these state features separately revealed that state transitions correctly classify all six treatments, whereas exclusion of temporal transition information leads to misclassification of a treatment, yielding a classification performance similar to the fixed-cell “Factor Analysis + Means” profiling approach.

Collectively, these results demonstrate that all of the methods implemented, with the exception of the “Means” approach, correctly classify the MLCK and myosin II inhibitors, but fail to differentiate between cells treated with MEK inhibitor versus control, or those treated with ROCK versus myosin II inhibitors. Thus, as shown in Fig. 7D, when the shape distributions of two treatments differ, both state-space modeling of individual cells over time and existing fixed-cell profiling methods that measure features of different cells at multiple time points, can correctly classify and resolve treatment effects on cell shape. When shape distributions of two treatments are similar, however, state-space modeling of single-cell temporal transitions within the population distributions can improve discriminability and classification of treatments compared with existing methods that only capture shape properties of distinct cells within a population. Therefore, annotation of cell morphologies with a state-space representation using HMM, and in particular capturing state transition dynamics on a per-cell basis, can improve the accuracy of classifying treatment conditions in an HCI experiment.

## Discussion

Here, we presented a computational framework that applies probabilistic time series modeling to characterize shape dynamics of individual cells under the action of drug perturbations assayed using live-cell imaging. Modeling temporal dynamics of cell shape using HMM condenses complex multivariate imaging data into simpler sequences of morphological states that evolve over time. Advantages of our approach over existing methods that are designed for fixed-cell imaging applications were explored. Quantitative features extracted from HMM state sequences using our approach serve as temporal signatures for comparing morphodynamic behaviors between cells, which are not available from fixed-cell analysis procedures. Temporal signatures from multiple identically treated cells are combined to serve as phenotypic profiles for clustering and classifying experimental perturbations.

It has been proposed that cell populations can assume either discrete or continuous morphological landscapes, principally determined by cell type, extracellular environment, types of perturbations, and the particular phenotype captured<sup>45</sup>. For example, in cases when particular morphologies are reached under effect of genetic perturbations, cells may exhibit more stable steady-state morphologies leading to discrete shape states<sup>13</sup>. Our results suggest, however, that shorter-term responses to pharmacological perturbation of highly asynchronous cell populations fall on a more continuous, finely graded morphological spectrum of shapes that conceals more cohesive and well separated shape state clusters within temporal trajectories of individual cells (Fig. 3). Consequently, two different approaches to cell profiling, one using commonly-applied fixed-cell analyses of populations of distinct cells and the other using temporal modeling of individual live cells presented

here, may collectively serve as a powerful combined approach to reveal idiosyncrasies in phenotypic effects of different molecular perturbations (Fig. 8).

In particular, within the context of image-based phenotypic profiling, our results illustrate that state-space temporal modeling of cell shape using HMM can improve the resolution of cellular classification in response to distinct drug treatments. This result, shown in Fig. 7D, highlights an important practical outcome of this work in the context of HCI. Existing phenotypic profiling methods correctly classified some treatments and misclassified others in the actomyosin cytoskeleton-focused drug screen. Our analyses demonstrate that modeling morphological transitions over time on an individual cell basis using HMM can provide useful live-cell phenotypic information for discriminating between treatment effects. Some of these effects are not captured by simply modeling cell shape distribution properties of different cells in a population, even if the same population is profiled over time. This suggests that the framework presented offers additional information beyond fixed-cell profiling in image-based classification using morphological data. An interesting application of temporal modeling would be to complement fixed-cell assays that utilize immunofluorescence staining in order to augment phenotypic profiling and better resolve underlying molecular differences driving distinct cellular drug responses assayed using live imaging.

In addition to investigating the utility of temporal shape modeling in HCI applications, we characterized differential cell shape dynamics induced by small-molecule inhibitors in order to infer relationships between actomyosin signaling mechanisms and shape phenotypes. Similarities in shape dynamics between DMSO and non-DMSO controls, between the two MEK inhibitors, and under different doses of the same drugs, validates the morphological states and state transitions as fundamentally reflecting target-specific effects. Similar dynamic responses in cell shape induced by EGFR, MEK, and calpain inhibition suggest that these species function along a common signaling axis of cell shape regulation in MDA-MB-231 breast cancer cells (Fig. 6B). This result is in concordance with previous molecular mechanistic studies performed in fibroblasts, which showed that m-calpain activity is directly regulated by Erk by altering its spatial localization and association with PI(4,5)P2 at the plasma membrane<sup>46</sup>. This process can be mediated by EGFR signaling through PLC- $\gamma$ , which depletes PI(4,5)P2 at the leading edge resulting in localization and activation of m-calpain at the trailing end that plays an important role in migratory cell polarization. The increased elongation phenotype induced by calpain inhibition revealed from our analysis (Fig. S6A) reflects the role of m-calpain in adhesion remodeling, the inhibition of which leads to trailing end retraction defects<sup>47,48</sup>.

Moreover, our observation that ROCK and myosin II inhibitors biased cells toward branched and elongated morphologies is consistent with reported effects of increased actomyosin tension promoting cell polarization and limiting the formation of new protrusions<sup>49–52</sup>. The increased elongation and branching dynamics we observed reflect the impairment in actomyosin contractility and loss of cell polarity upon ROCK and myosin II inhibition, leading to multiple competing protrusions extending outwards around the cell periphery. Given that MLCK and ROCK are both known to induce myosin II phosphorylation and activation, we may have expected their inhibition to promote similar shape dynamics.

Instead, the surprisingly dissimilar and near-opposite morphologies that their inhibition produced (Fig. 5B) suggests that MLCK and ROCK differentially affect cytoskeletal dynamics; this notion is supported by previous studies which showed that MLCK and ROCK have spatially distinct roles in regulating myosin II activity, and whose activities have opposite effects on the number of lamellipodial protrusions, cell elongation, and polarization<sup>50,53</sup>. The MDA-MB-231 cell shapes induced by MLCK inhibition in our work are reflective of single-front motile cell morphologies for which formation of additional lamellipodial protrusions is limited (Video S3), as has been suggested to also occur in keratocytes at different stages of development<sup>53</sup>. The differential effects of ROCK and MLCK inhibition on shape dynamics that we identified through our analyses (Fig. 8) may have consequential effects on overall cell polarization and migratory behavior in cancer that warrants further investigation in future work.

More generally, the plasticity and variability in shape revealed in our analyses reflects the highly dynamic cytoskeletal changes characteristic of protrusion and retraction events, such as those of migratory cells undergoing directional changes or motility cycles. In light of this, further exploration of interest would be to characterize and understand what cellular functions the morphological states captured by HMM annotation may represent. Qualitative visual inspection suggests that cells in rounder states have higher amounts of cortical actin, whereas those in more elongated states are polarized and may therefore be migratory (Fig. 2B, Videos S1-3). Future studies will focus on characterizing morphological state properties in more depth by quantifying cytoskeletal organization in addition to whole-cell shape, incorporating signaling status from live or fixed-cell fluorescent reporters in post-capture analysis<sup>24,54</sup>, and combining model-annotated state-space dynamics with motility measurements to establish temporal relationships between morphology and migratory behavior<sup>27,55,56</sup>.

Notwithstanding, the current work demonstrates that phenotypic information derived from imaging-based models of temporal shape dynamics reflects underlying signaling pathway activities that regulate cell shape, which enables hypothesis generation for more in-depth, mechanistic follow-up studies. *In vitro* profiling of phenotype dynamics is also relevant in the context of early-phase drug discovery. In that context, time series modeling of live-cell phenotypes can complement existing fixed-cell imaging methods as well as genetic and biochemical approaches for understanding cellular drug responses, particularly with the advent of genome editing approaches such as CRISPR/Cas9<sup>57-61</sup>. The collective implementation and application of these approaches will become an increasingly important component of integrative drug profiling and holistic understanding of cell function.

## Materials and Methods

### Software and image data availability

The files necessary to run SAPHIRE are provided as supplemental software. A demo with sample cell image data, instructions, and scripts is provided for using the trajectory editing GUI tool, time series modeling, phenotypic profile computation for groups of cells, and visualization of SAPHIRE outputs.

## Generation of fluorescent reporter cells

Cells were generated with fluorescent reporters for both actin to provide for cell shape and histones to label nuclei. A pBABE-HistoneH2B-mCherry retroviral plasmid was a gift from Dr. Iain Cheeseman (MIT Whitehead Institute). LifeAct-eGFP was inserted between the XhoI-EcoRI sites in pMSCV-puro vector using standard molecular biology techniques. Replication-incompetent virus was purified from HEK-293T cells using standard protocols. Supernatants containing LifeAct-eGFP or H2B-mCherry packaged virus were harvested at 48–72 h after transfection, and passed through 0.45 $\mu$ m filter prior to use for transduction of target cells (Pall Corp., Cortland, NY). MDA-MB-231 cells were transduced with LifeAct-eGFP and H2B-mCherry filtered viral supernatants containing 8 $\mu$ g/mL polybrene (Millipore). Selection and propagation of transduced MDA-MB-231 cells was performed by culture in 1 $\mu$ g/mL Puromycin (Sigma) in complete DMEM media. To make the stable fluorescent cell population more uniform in the expression of the two reporters, cells were sorted for double-positive intensities in the 80-90 percentile of the population using a MoFlo3 flow cytometer (Beckman Coulter, Inc.).

## Cell culture, live-cell imaging, and drug perturbations

The triple negative breast cancer cell line MDA-MB-231 (ATCC) stably expressing LifeAct-eGFP and histone H2B-mCherry was used in all experiments. Cells were cultured in high-glucose Duplecco's Modified Eagle Medium (DMEM) (Life Technologies) supplemented with 10% HyClone fetal bovine serum (Thermo Scientific), 1% penicillin/streptomycin (Gibco), and 1% GlutaMAX (Life Technologies) at 37°C and 5% CO<sub>2</sub>. For drug perturbation imaging experiments, Nunclon Delta 96-well optical bottom plates (Thermo Scientific) were coated with 5 $\mu$ g/cm<sup>2</sup> pH-neutralized, acid-extracted, nonpepsin digested collagen I (BD Biosciences) for 1 hour at 37°C and 5% CO<sub>2</sub>. All wells were then washed twice with PBS and once with culture media prior to cell seeding. Following plate coating with collagen, cells were seeded at a density of 1000 cells/cm<sup>2</sup> in culture medium and incubated at 37°C and 5% CO<sub>2</sub> for 24 hours. Following the 24-hour incubation, the culture media was replaced with drug-containing imaging culture media and the plates were immediately transferred to the microscope for live imaging. The time delay between drug addition and start of image acquisition was approximately 30 minutes to 1 hour.

Two imaging experiments were performed: for the experiment with the expanded panel of drugs cells were imaged on an IncuCyte ZOOM system incubated at 37°C and 5% CO<sub>2</sub> under a standard scanning protocol (Essen Bioscience) in drug-containing DMEM growth media (see above). Cells were imaged over approximately 18 hours at 20-minute intervals, with three wells per drug condition and four fields of view per well, producing an image data set of 84 fields (4452 time series image frames), from which 293 individual cell trajectories, with an average of 53 time frames, were obtained for further analyses. For the other experiment, which contained the experimental controls, cells were imaged in Leibovitz's L-15 media (Life Technologies) supplemented with the same additives as the DMEM culture media above, on a Nikon Eclipse Ti microscope equipped with an Andor Zyla sCMOS camera. Cells were imaged over approximately 18 hours at 8-minute intervals, with three wells per drug condition and four fields of view per well, producing an image data set of 96 fields (16128 time series image frames), from which 435 individual cell trajectories, with an



average of 100 time frames, were obtained for further analyses. All inhibitor stocks were dissolved in dimethyl sulfoxide (DMSO) (Sigma), other than Y-27632, which was dissolved in high-purity water. All final concentrations of inhibitors in culture media used for imaging were at 0.1% v/v DMSO or lower. In both experiments cells were imaged using a 0.30 NA Nikon Plan Fluor 10X air objective. The inhibitors used in this work and their vendor sources were as follows: AZD6244 (MEK inhibitor; Selleck Chem.), ML-7 (MLCK inhibitor; Enzo Life Sciences), Blebbistatin(+/-) (non-muscle myosin II inhibitor; Enzo Life Sciences); Gefitinib (EGFR inhibitor; LC Labs), PD0325901 (MEK inhibitor; LC Labs), Y-27632 (ROCK inhibitor; Enzo Life Sciences), PD150606 (Calpain inhibitor; EMD Millipore).

### Image processing

Time series stacks were exported as monochrome images from the red and green channel for LifeAct-eGFP (actin) and histone H2B-mCherry (nuclei) reporters, respectively. All images were batch processed using a custom pipeline written in MATLAB (Mathworks, Inc.). First, time series images were drift-corrected with the StackReg plugin in ImageJ using the nuclear reporter channel relative to the first time frame and cropped to maintain identical field of view regions in the nuclei and actin reporter image stacks for each field individually. Nuclei in each image were then segmented using point-source detection<sup>62</sup>. User-assisted cell body segmentation from the actin reporter channel was performed for the first frame of each time series stack. Otsu intensity and Canny edge detection thresholds were set by the user for the first frame of a time series for each field of view, with the cell masks generated by combining the binary Otsu foreground and Canny edge pixels. The thresholds were modified through time for each field of view automatically for the second frame onward using a parameter gradient search that minimized the sum of the difference in area, perimeter, and solidity of foreground objects between a previous and subsequent frame. For each connected foreground object, nuclei centers were used to form holes in the binary object mask that enabled detection of touching cells by computing the Euler number for the object. Segmented cell body regions without spatial overlap with segmented nuclei were removed. Segmented nuclei were tracked over time using the IDL tracking method implementation<sup>63</sup>. To avoid bias of cell-cell interaction effects on morphology, touching cells were automatically identified as two or more segmented nuclei within a segmented cell body region in a given frame, and subsequently flagged for removal. A graphical user interface was developed to correct cell body segmentation inaccuracies and to label dividing or dying cells (Fig. 2). Continuous cell trajectories with no division, death, or intercellular spatial interactions were retained for both imaging experiments and subsequent analyses.

### Cell shape quantification and shape-space definition

Eighteen whole-cell shape features were computed from the binary masks of each cell at every time point in both imaging experiments (Table S1). Here, we call a temporal snapshot of a cell a “cell object”. For each imaging experiment separately, the raw shape features from all cell objects were combined into an  $n \times m$  matrix  $A$ , where  $n$  is the total number of cell objects and  $m$  is the number of shape features. The matrix was z-score normalized across cell objects for each feature (along columns). For the imaging experiment where the expanded panel of drugs was tested, we performed dimensionality reduction using PCA onto

a two-dimensional basis from the covariance matrix of  $A$ , making the PCs linear combinations of the shape features with points in PC space corresponding to cell objects. The PCA coefficients derived from cells imaged in the expanded drug panel experiment were then used to project the shape features from the other experiment (with drug doses and controls) onto the same two-dimensional PC basis. Reducing data dimensionality using PCA dampens the effects of features in the model that contribute little to shape variability across cells and removes correlations between features used for subsequent modeling steps. With this projection, the shape trajectory of a cell is converted from a  $T \times m$  matrix, where  $T$  is the number of time points, to a  $T \times 2$  matrix, where the two columns correspond to the two linearly independent PCs. The  $T \times 2$  matrices for each cell represent bivariate temporal trajectories in the two-dimensional PC space that we define as the “shape-space”. The bivariate temporal trajectories are then used for subsequent shape state identification and time series modeling. GMM using expectation maximization and Bayesian information criterion (BIC) for model selection were used for GMM analyses of simulated and experimental data and implemented using the *mclust* package in R.

### Probabilistic time series modeling

A bivariate temporal trajectory for an individual cell whose shape dynamics we wish to model is comprised of  $T$  time points each with an  $(x_t, y_t)$  coordinate at each time point  $t \in \{1, \dots, T\}$  in two-dimensional PC shape-space. We model the coordinates as emissions  $e_t = (x_t, y_t)$  from  $K$  number of “hidden” shape states,  $\{s_i\}_{i=1}^K$ ,  $K = \{1, 2, 3, \dots\}$ . We represent each shape state as a symmetric, bivariate Gaussian distribution with mean  $\mu_j = (\mu_{x,j}, \mu_{y,j})$  and standard deviation along both PC coordinates,  $\sigma_j = \sigma_{x,j} = \sigma_{y,j}$ . The covariance matrix of  $x$  and  $y$  is diagonal, and we write the probability of the point  $e_t$  in shape-space coming from shape state  $s_j$  as:

$$p_{s_i}(e_t) = \frac{1}{2\pi\sigma_i^2} \exp\left(-\frac{(x_t - \mu_{x,i})^2 + (y_t - \mu_{y,i})^2}{2\sigma_i^2}\right)$$

We next consider the set of points  $e_t$  for the entire cell trajectory ( $e_t \forall t \in \{1, \dots, T\}$ ), resulting in a  $T \times 2$  matrix,  $\mathbf{e}$ . Because the number of hidden shape states, and the parameters  $\mu_j$  and  $\sigma_j$  for each state  $s_j$  that lead to the emissions  $\mathbf{e}$  are unknown, we test models  $M_k$  with different numbers of hidden states,  $k \in K$ , and select the model that best fits the cell trajectory shape-space data,  $\mathbf{e}$ . Bayesian model selection is used to evaluate how well each model fits the data in order to select the best model, inherently penalizing increased model complexity (i.e., increasing number of states),

$$P(M_k|\mathbf{e}) = \frac{P(\mathbf{e}|M_k)P(M_k)}{P(\mathbf{e})} \propto P(\mathbf{e}|M_k)$$

with the proportionality holding since we consider the prior probabilities of all models  $M_k$  to be equal. Importantly, points in  $\mathbf{e}$  are not independent as they represent a temporal evolution in shape of the same cell. A hidden Markov model in the framework is used to incorporate the temporal dependencies in  $\mathbf{e}$  and infer parameters that describe the dynamic properties

(state transitions). Therefore, for each model,  $M_k$ , its full set of parameters  $\theta_k$  that must be inferred from the data are

$$\theta_k = \left[ \{\mu_{x,i}, \mu_{y,i}, \sigma_i\}_{i=1}^K, \{\pi_i\}_{i=1}^K, \{\phi_{ij}\}_{i,j=1}^K \right]$$

where for each shape state (bivariate Gaussian distribution)  $\mu_i$  and  $\sigma_i$  correspond to the state mean and standard deviation in shape-space, respectively,  $\phi_{ij}$  is the probability of transitioning from state  $s_i$  to state  $s_j$  within the state transition probability matrix,  $\phi$ , and  $\pi_i$  is the probability of the cell starting in state  $s_i$  at the first time point. Models are compared amongst each other independent of a particular realization, or values, of the parameters, as well as of the possible hidden shape state sequences. Therefore, the likelihood  $P(\mathbf{e} | M_k)$  is marginalized over all the parameters  $\theta_k$  and hidden state sequences  $s_k = \{s_t\}$  for  $t \in \{1, \dots, T\}$  to obtain the total marginalized likelihood of the model  $M_k$ ,

$$P(\mathbf{e} | M_k) = \int \sum_{s_k} \left( \pi_{s_1} \prod_{t=2}^T \phi_{s_{t-1}, s_t} \prod_{t=1}^T p_{s_t}(\mathbf{e}_t) \right) P(\theta_k | M_k) d\theta_k$$

The summation over states in brackets,  $[\cdot]$ , is the probability,  $P(\mathbf{e} | M_k, \theta_k)$  of the observed temporal sequence of coordinates in shape-space of the cell trajectory,  $\mathbf{e}$ , conditioned on a particular model with  $k$  hidden shape states,  $M_k$ , and its parameters,  $\theta_k$ . Summation over the hidden state sequences is performed using the forward algorithm. Metropolis Markov Chain Monte Carlo (MCMC) with importance sampling is used to sample parameter space in order to integrate the marginalized likelihood, with the prior probability of the parameters given the model,  $P(\theta_k | M_k)$ , taken as constant<sup>39</sup>. The resulting MCMC integration yields the total marginalized likelihood of each model,  $P(\mathbf{e} | M_k)$ , and its maximum likelihood (ML) parameters,  $\hat{\theta}_k$ . The model with the highest marginalized likelihood is chosen to describe the shape dynamics of the cell, with the most likely hidden shape state sequence calculated by the Viterbi algorithm using the ML parameters.

### SAPHIRE model-derived phenotypic profiles for drug comparisons

Annotated shape state sequences of individual trajectories from identically treated cells were combined to generate a phenotypic profile for that treatment condition. A phenotypic profile is a 48-element numerical vector composed of four types of histograms that capture shape dynamics: state radial distances and state dwell times, collectively called “state location features”, as well as state transition magnitudes and state transition dwell times, collectively called “state transition features”. These features are derived as follows (see illustration in Fig. 5A).

For each inferred Gaussian state  $s_i$  in shape-space from a temporally annotated sequence, we convert the coordinates of the state mean value in the Cartesian PCA axes  $(\mu_{x,s_i}, \mu_{y,s_i})$  to polar coordinates  $(\mu_{r,s_i}, \mu_{\theta,s_i})$ . As a result,  $\mathbf{r}_{s_i}$  becomes the location vector of the state in shape-space and  $|\mathbf{r}_{s_i}|$  is the radial distance of the state from the origin, where  $|\cdot|$  denotes vector magnitude.  $\theta$  is the angle  $\mathbf{r}_{s_i}$  makes in the counterclockwise direction with the positive PC1 axis for state  $s_i$ . For each state  $s_i$  the state radial distance is described by  $|\mathbf{r}_{s_i}|$ , or

equivalently  $\sqrt{\mu_{x,s_i}^2 + \mu_{y,s_i}^2}$ . The normalized state dwell time for  $s_j$  is calculated as  $\sum_{t=1}^T \delta_{t,s_j} / T$ , where  $\delta_{t,s_j} = 1$  if the state is  $s_j$  in the model-annotated state sequence at time  $t$ , and 0 otherwise, with  $T$  equal to the number of elements in the sequence (i.e. number of time frames in the cell trajectory). For a given cell trajectory, we derive an angular histogram with twelve bins on the range of 0 to 360 degrees, with the first bin 0 to 30 degrees, second bin 30 to 60 degrees, and so on until the twelfth bin that ranges from 330 to 360 degrees. For all states  $s_j \in \mathcal{S}$  in the cell sequence, the state radial distance value of each bin is computed as  $\langle |r_{s_i}| \rangle$  where  $\langle \cdot \rangle$  denotes the mean, for all states  $s_j$  with  $\mu_{\theta,s_j}$  falling within the angular range of the bin, to generate a histogram of state radial distances. A state radial distance profile for a treatment condition is then computed as the average within each bin in the twelve-bin state radial distance histograms for all cells in that treatment condition. The state dwell time profile of the treatment condition is similarly computed except averaging normalized state dwell time values for states of each cell instead of the state radial distances for twelve angular bins.

We additionally derive state transition profiles for each treatment condition. Consider for a given cell a state transition from  $s_i$  to  $s_j$ . The state transition direction  $\theta_{s_{ji}}$  for  $s_j$  from state  $s_i$  is the angle that the vector  $r_{s_j} - r_{s_i}$  makes in the counterclockwise direction with the positive PC1 axis in shape-space. The state transition magnitude for  $s_j$  from state  $s_i$  is computed as  $|r_{s_j} - r_{s_i}|$  scaled by the normalized state transition dwell time. The normalized state transition dwell time for  $s_j$  from state  $s_i$  is calculated as the number of time points a cell spends in state  $s_j$  after transitioning to it from state  $s_i$ , divided by the sequence length,  $T$ . The state transition dwell time value of each of the twelve angular bins is the mean of the normalized state transition dwell times of all states whose means  $(\mu_{r,s_j}, \mu_{\theta,s_j})$  are in the bin. For transition directions  $\theta_{s_{ji}}$  for all  $i \neq j$  and a given cell, the state transition magnitude value of each of the twelve angular bins is computed as  $\langle |r_{s_j} - r_{s_i}| \rangle$  for all  $\theta_{s_{ji}}$  falling within the angular range of each bin, to produce a histogram of state transition magnitudes. A state transition magnitude profile for a treatment condition is then calculated as the average within each bin in the twelve-bin state transition histograms for all cells in that treatment condition. A state transition dwell time profile is similarly computed except averaging normalized state transition dwell times across cells for the twelve angular bins instead of the transition magnitudes.

To quantify similarities in effects of treatment conditions on cell shape dynamics, clustering using average linkage of Euclidean distances between phenotypic profiles was performed for treatment pairs to generate cluster dendrograms. Internal nodes were re-sorted into optimal leaf order without dividing the clusters or changing overall tree connectivity so that treatments with similar profiles are next to each other by maximizing the sum of similarities  $(\max_{ij} (d_{ij}) - d_{ij})$ , where  $i$  is a leaf adjacent to leaf  $j$  between adjacent tree leaves.

Permutation testing was used to assess significance of Euclidean distances (similarities) between all pairs of treatments as follows. Dynamic phenotype profiles (48-element vectors) based on state locations and state transitions and corresponding dwell times (see above) of

individual cells within any two treatment conditions being compared were randomly reassigned into the two treatments while preserving the number of cells and their signatures. Signature values were averaged between the randomly-assigned cells for a given treatment to generate a treatment profile. A permutation test p-value for similarity between two treatment profiles was quantified by comparing the actual Euclidean distance between the treatment profiles relative to the null distribution of distances between profiles that were obtained by repeating the cell assignment randomizations and distance calculations 10,000 times for each pair of treatments.

### Drug classification and comparison with existing methods

To compare between various image-based profiling methods for classifying treatment conditions, we followed an analysis procedure similar to the one described previously<sup>15</sup>. Four existing state-of-the-art methods for generating treatment profiles from fixed-cell measurements were implemented for comparison with the temporal modeling framework presented in this work: “Means”<sup>42</sup>, “K-S Statistic”<sup>10</sup>, “Factor Analysis + Means”<sup>44</sup>, and “Gaussian Mixture”<sup>43</sup>, all of which were previously compared in an HCI drug classification performance study<sup>15</sup>. For each method, all cellular image snapshots were extracted from the time series movies for all treatment conditions at 1, 4, 8, 12, and 16 hours following treatment addition, as if a researcher would make fixed-cell imaging measurements at these time points in an HCI experiment.

Briefly, the “Means” approach averages each of the 18 cell shape features for each treatment condition for the five time points to generate a phenotypic profile for a treatment. For the “K-S Statistic”, the cumulative distribution function (*cdf*) is compared between the treatment and DMSO control cells for each shape feature separately. A signed Kolmogorov-Smirnov (K-S) statistic is then computed for each treatment, which is equal to the maximum distance between the two *cdfs* and set to positive if the treatment *cdf* is above the control *cdf* and negative otherwise. A “K-S Statistic” profile for a treatment is a vector of concatenated signed K-S statistics for each shape feature. For the “Factor Analysis + Means” method we first performed factor analysis on cells from all treatments at the five time points, selecting the number of factors using the Kaiser criterion, and then computed the average value of the scores of each factor for the cells in each treatment condition to generate a profile for that treatment. Finally, for the “Gaussian Mixture” method, a GMM with no covariance matrix constraints was fit to cell observations from all treatments at each time point separately using all shape features as variables, testing models with 2 to 30 mixtures, and selecting the best-fitting model using BIC. For each treatment, the posterior probability of each cell under that treatment belonging to each of the mixtures was computed and averaged across cells for each mixture to generate a profile at a given time point. The averaged posterior probabilities of all mixtures and time points were combined to generate a phenotypic profile for each treatment.

Treatment classification accuracy was assessed for each of the profiling methods described above, as well as using the profiles derived from the single-cell temporal modeling using SAPHIRE. To assess the value of the HMM in treatment classification, treatment profiles extracted from shape-space analyses of cellular dynamics without the HMM annotations

were generated identically as for SAPHIRE but by treating each time point in a cell trajectory as a separate “state” (i.e. no modeling of PCA coordinates from underlying hidden Gaussian states). Treatment condition classification accuracy was assessed by computing pair-wise Euclidean distances between phenotypic profiles of treatments using each of the profiling approaches separately. A treatment was designated as being correctly classified if its phenotypic profile was closest, in terms of distance, to that of the same treatment at a different dose, compared to the profiles from other treatments <sup>15</sup>.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We would like to acknowledge Shannon Hughes for guidance on experimental elements of this work, and Nilah Monnier, Syuan-Ming Guo, Zachary Barry, and Russell McConnell for helpful discussions. We are grateful to the assistance of staff from the Imaging and Flow Cytometry cores of the Swanson Biotechnology Center at the MIT Koch Institute for Cancer Research. This work was supported by NIGMS grant GM69668 (A.W., D.A.L.), NSF PoLS PHY 1305537 (M.B.), and the Ludwig Postdoctoral Fellowship (M.K.H.).

## Abbreviations

<b>BIC</b>	Bayesian information criterion
<b>EGFR</b>	epidermal growth factor receptor
<b>GMM</b>	Gaussian mixture modeling
<b>HCI</b>	high-content imaging
<b>HMM</b>	hidden Markov modeling
<b>PC</b>	principal component
<b>PCA</b>	principal component analysis
<b>MEK</b>	mitogen-activated protein kinase kinase
<b>MLCK</b>	myosin light chain kinase
<b>ROCK</b>	Rho-associated protein kinase
<b>SAPHIRE</b>	Stochastic Annotation of Phenotypic Individual-cell Responses

## References

1. Fuchs F, Pau G, Kranz D, Sklyar O, Budjan C, Steinbrink S, Horn T, Pedal A, Huber W, Boutros M. *Mol. Syst. Biol.* 2010; 6:370. [PubMed: 20531400]
2. Loo L-H, Wu LF, Altschuler SJ. *Nat. Methods.* 2007; 4:445–453. [PubMed: 17401369]
3. Treiser MD, Yang EH, Gordonov S, Cohen DM, Androulakis IP, Kohn J, Chen CS, Moghe PV. *Proc. Natl. Acad. Sci. U. S. A.* 2010; 107:610–615. [PubMed: 20080726]
4. Vega S, Dhaliwal A, Arvind V, Patel PJ, Beijer NRM, De Boer J, Murthy NS, Kohn J, Moghe PV. *Integr. Biol.* 2015; 7:435–446.
5. Glory E, Murphy RF. *Dev. Cell.* 2007; 12:7–16. [PubMed: 17199037]



6. Mazumder A, Pesudo LQ, McRee S, Bathe M, Samson LD. *Nucleic Acids Res.* 2013; 41:9310–9324. [PubMed: 23935119]
7. Wawer MJ, Li K, Gustafsdottir SM, Ljosa V, Bodycombe NE, Marton M. a, Sokolnicki KL, Bray M-A, Kemp MM, Winchester E, Taylor B, Grant GB, Hon CS-Y, Duvall JR, Wilson JA, Bittker J. a, Dan ik V, Narayan R, Subramanian A, Winckler W, Golub TR, Carpenter AE, Shamji AF, Schreiber SL, Clemons P. a. *Proc. Natl. Acad. Sci. U. S. A.* 2014; 111:10911–10916. [PubMed: 25024206]
8. Ng AYJ, Rajapakse JC, Welsch ROYE, Matsudaira PT, Horodincu V, Evans JG. *J. Biomol. Screen.* 2010; 15:858–868. [PubMed: 20525958]
9. Bakal C, Aach J, Church G, Perrimon N. *Science.* 2007; 316:1753–6. [PubMed: 17588932]
10. Perlman ZE, Slack MD, Feng Y, Mitchison TJ, Wu LF, Altschuler SJ. *Science.* 2004; 306:1194–1198. [PubMed: 15539606]
11. Candia J, Maunu R, Driscoll M, Biancotto A, Dagur P, McCoy JP, Sen HN, Wei L, Maritan A, Cao K, Nussenblatt RB, Banavar JR, Losert W. *PLoS Comput. Biol.* 2013; 9:e1003215. [PubMed: 24039568]
12. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, Castoreno AB, Eggert US, Root DE, Golland P, Sabatini DM. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:1826–31. [PubMed: 19188593]
13. Yin Z, Sadok A, Sailem H, McCarthy A, Xia X, Li F, Garcia MA, Evans L, Barr AR, Perrimon N, Marshall CJ, Wong STC, Bakal C. *Nat. Cell Biol.* 2013; 15:1–13. [PubMed: 23263367]
14. Ku C-J, Wang Y, Weiner OD, Altschuler SJ, Wu LF. *Cell.* 2012; 149:1073–83. [PubMed: 22632971]
15. Ljosa V, Caie PD, Ter Horst R, Sokolnicki KL, Jenkins EL, Daya S, Roberts ME, Jones TR, Singh S, Genovesio A, Clemons P. a, Carragher NO, Carpenter AE. *J. Biomol. Screen.* 2013; 18:1321–9. [PubMed: 24045582]
16. Schnell U, Dijk F, Sjollem KA, Giepmans BNG. *Nat. Methods.* 2012; 9:152–8. [PubMed: 22290187]
17. Bosgraaf L, Van Haastert PJM. *Cell Adhes. Migr.* 2010; 4:46–55.
18. Tsygankov D, Bilancia CG, Vitriol E. a, Hahn KM, Peifer M, Elston TC. *J. Cell Biol.* 2014; 204:443–460. [PubMed: 24493591]
19. Barry DJ, Durkin CH, Abella JV, Way M. *J. Cell Biol.* 2015; 209:163–180. [PubMed: 25847537]
20. Machacek M, Danuser G. *Biophys. J.* 2006; 90:1439–52. [PubMed: 16326902]
21. Tsukada Y, Aoki K, Nakamura T, Sakumura Y, Matsuda M, Ishii S. *PLoS Comput. Biol.* 2008; 4:e1000223. [PubMed: 19008941]
22. Driscoll MK, Mccann C, Kopace R, Homan T, Fourkas JT, Parent C, Losert W. *PLoS Comput. Biol.* 2012; 8:e1002392. [PubMed: 22438794]
23. Veronika M, Welsch R, Ng A, Matsudaira P, Rajapakse JC. *BMC Bioinformatics.* 2011; 12:S19.
24. Welch CM, Elliott H, Danuser G, Hahn KM. *Nat. Rev. Mol. Cell Biol.* 2011; 12:749–56. [PubMed: 22016058]
25. Driscoll MK, Losert W, Jacobson K, Kapustina M. *Cytoskeleton.* 2015; 281:1–27.
26. Tsygankov, D.; Chu, P-H.; Chen, H.; Elston, TC.; Hahn, KM. *Methods in Cell Biology.* 1st. Waters, J.; Wittman, T., editors. Vol. 123. Elsevier; Amsterdam: 2014. p. 409-427.
27. Keren K, Pincus Z, Allen GM, Barnhart EL, Marriott G, Mogilner A, Theriot JA. *Nature.* 2008; 453:475–80. [PubMed: 18497816]
28. Bear JE, Haugh JM. *Curr. Opin. Cell Biol.* 2014; 30C:74–82.
29. Liu X, Welf ES, Haugh JM. *J. R. Soc. Interface.* 2015; 12:20141412. [PubMed: 25904526]
30. Weiger MC, Ahmed S, Welf ES, Haugh JM. *Biophys. J.* 2010; 98:67–75. [PubMed: 20085720]
31. Weiger MC, Wang C-C, Krajcovic M, Melvin AT, Rhoden JJ, Haugh JM. *J. Cell Sci.* 2009; 122:313–323. [PubMed: 19126672]
32. Henry SJ, Crocker C, Hammer DA. *Integr. Biol.* 2014; 6:348–356.
33. Kafri R, Levy J, Ginzberg MB, Oh S, Lahav G, Kirschner MW. *Nature.* 2013; 494:480–3. [PubMed: 23446419]

34. Sailem H, Bousgouni V, Cooper S, Bakal C. *Open Biol.* 2014; 4:130132. [PubMed: 24451547]
35. Sero JE, Sailem HZ, Ardy RC, Almuttaqi H, Zhang T, Bakal C. *Mol. Syst. Biol.* 2015; 11:790.
36. Held M, Schmitz M. H. a, Fischer B, Walter T, Neumann B, Olma MH, Peter M, Ellenberg J, Gerlich DW. *Nat. Methods.* 2010; 7:747–54. [PubMed: 20693996]
37. Zhong Q, Busetto AG, Fededa JP, Buhmann JM, Gerlich DW. *Nat. Methods.* 2012; 9:711–713. [PubMed: 22635062]
38. Fraley C, Raftery AE. *J. Am. Stat. Assoc.* 2002; 97:611–631.
39. Monnier N, Barry Z, Park HY, Su K, Katz Z, English BP, Dey A, Pan K, Cheeseman IM, Singer RH, Bathe M. *Nat. Methods.* 2015; 12:1–7. [PubMed: 25699311]
40. Posada D, Buckley T. *Syst. Biol.* 2004; 53:793–808. [PubMed: 15545256]
41. Raftery AE. *Sociol. Methodol.* 1995; 25:111–163.
42. Tanaka M, Bateman R, Rauh D, Vaisberg E, Ramachandani S, Zhang C, Hansen KC, Burlingame AL, Trautman JK, Shokat KM, Adams CL. *PLoS Biol.* 2005; 3:e128. [PubMed: 15799708]
43. Slack MD, Martinez ED, Wu LF, Altschuler SJ. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:19306–19311. [PubMed: 19052231]
44. Young DW, Bender A, Hoyt J, McWhinnie E, Chirn G-W, Tao CY, Tallarico J. a, Labow M, Jenkins JL, Mitchison TJ, Feng Y. *Nat. Chem. Biol.* 2008; 4:59–68. [PubMed: 18066055]
45. Yin Z, Sailem H, Sero J, Ardy R, Wong STC, Bakal C. *Bioessays.* 2014; 36:1195–203. [PubMed: 25220035]
46. Leloup L, Shao H, Bae YH, Deasy B, Stolz D, Roy P, Wells A. *J. Biol. Chem.* 2010; 285:33549–66. [PubMed: 20729206]
47. Glading A, Lauffenburger D. a, Wells A. *Trends Cell Biol.* 2002; 12:46–54. [PubMed: 11854009]
48. Franco SJ, Huttenlocher A. *J. Cell Sci.* 2005; 118:3829–38. [PubMed: 16129881]
49. Elliott H, Fischer RS, Myers K. a, Desai R. a, Gao L, Chen CS, Adelstein RS, Waterman CM, Danuser G. *Nat. Cell Biol.* 2015; 17:137–147. [PubMed: 25621949]
50. Totsukawa G, Wu Y, Sasaki Y, Hartshorne DJ, Yamakita Y, Yamashiro S, Matsumura F. *J. Cell Biol.* 2004; 164:427–439. [PubMed: 14757754]
51. Vicente-Manzanares M, Zareno J, Whitmore L, Choi CK, Horwitz AF. *J. Cell Biol.* 2007; 176:573–80. [PubMed: 17312025]
52. Vicente-Manzanares M, Koach M. a, Whitmore L, Lamers ML, Horwitz AF. *J. Cell Biol.* 2008; 183:543–54. [PubMed: 18955554]
53. Lou SS, Diz-Munoz a, Weiner OD, Fletcher D. a, Theriot J. a. *J. Cell Biol.* 2015; 209:275–288. [PubMed: 25918227]
54. Komatsu N, Aoki K, Yamada M, Yukinaga H, Fujita Y, Kamioka Y, Matsuda M. *Mol. Biol. Cell.* 2011; 22:4647–4656. [PubMed: 21976697]
55. Lock JG, Mamaghani MJ, Shafiqat-Abbasi H, Gong X, Tyrcha J, Strömblad S. *PLoS One.* 2014; 9:e90593. [PubMed: 24587399]
56. Schoenauer Sebag, a.; Plancade, S.; Raulet-Tomkiewicz, C.; Barouki, R.; Vert, J-P.; Walter, T. *Bioinformatics.* 2015; 31:i320–i328. [PubMed: 26072499]
57. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson J. a, Barthorpe S, Lutz SR, Kogera F, Lawrence K, McLaren-Douglas A, Mitropoulos X, Mironenko T, Thi H, Richardson L, Zhou W, Jewitt F, Zhang T, O'Brien P, Boisvert JL, Price S, Hur W, Yang W, Deng X, Butler A, Choi HG, Chang JW, Baselga J, Stamenkovic I, Engelman J. a, Sharma SV, Delattre O, Saez-Rodriguez J, Gray NS, Settleman J, Futreal PA, Haber D. a, Stratton MR, Ramaswamy S, McDermott U, Benes CH. *Nature.* 2012; 483:570–575. [PubMed: 22460902]
58. Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, Saez-Rodriguez J. *PLoS One.* 2013; 8:e61318. [PubMed: 23646105]
59. Niepel M, Hafner M, Pace E. a, Chung M, Chai DH, Zhou L, Schoeberl B, Sorger PK. *Sci. Signal.* 2013; 6:ra84. [PubMed: 24065145]
60. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. *Science.* 2013; 339:819–23. [PubMed: 23287718]

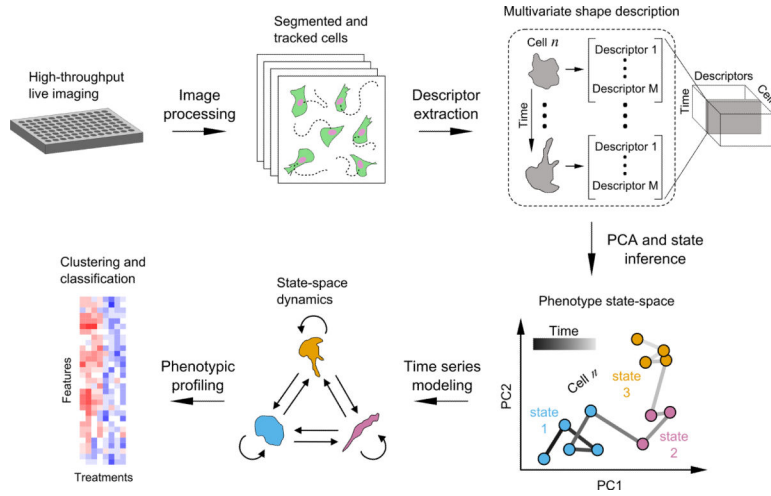
61. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. *Science*. 2013; 339:823–6. [PubMed: 23287722]
62. Aguet F, Antonescu CN, Mettlen M, Schmid SL, Danuser G. *Dev. Cell*. 2013; 26:279–291. [PubMed: 23891661]
63. Crocker JC, Grier DG. *J. Colloid Interface Sci*. 1996; 179:298–310.

Author Manuscript

Author Manuscript

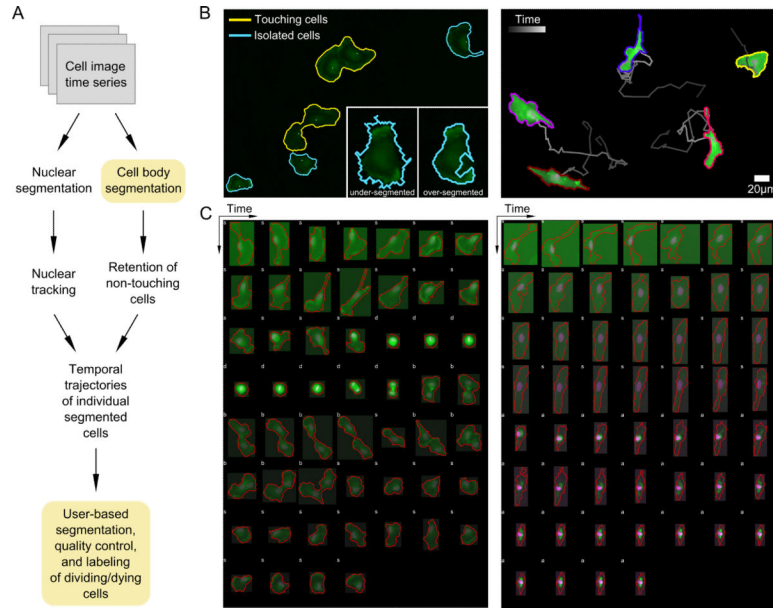
Author Manuscript

Author Manuscript



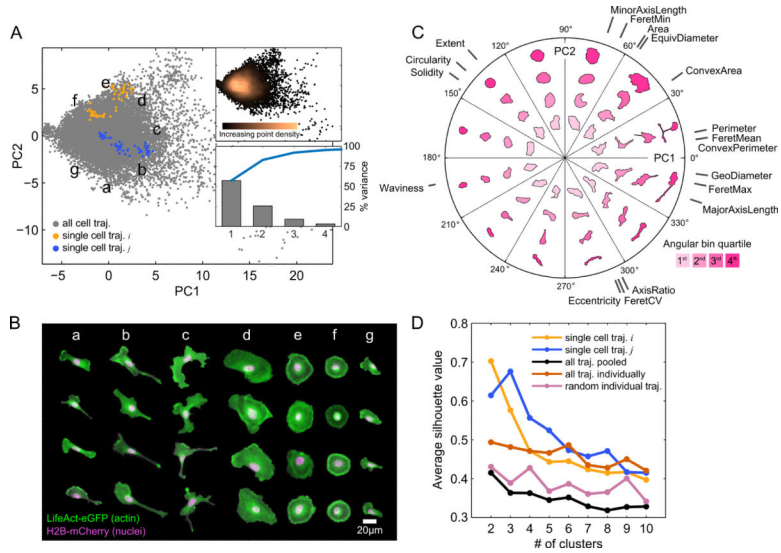
**Figure 1. Schematic of SAPHIRE (acronym) for live imaging-based modeling of cellular phenotypes**

Live fluorescent cells treated with drugs or other perturbations are imaged in multi-well format on a microscope with an incubation system. Temporal image acquisition is followed by image processing and analysis to segment and track individual cells over time. Cell shape descriptors are extracted to generate a data “cube” with dimensions corresponding to [shape features]  $\times$  [individual cells]  $\times$  [time-points] for the entire imaging screen. Principal component analysis is applied to generate the temporal “shape-space” trajectories of individual cells. For each cell trajectory, probabilistic time series modeling is applied to infer the most likely underlying model of phenotypic states (e.g. blue, pink, and yellow) and its parameters, which are used to annotate a temporal sequence of states and state transitions. Model-derived temporal features of single-cell dynamics from multiple cells are then used to generate phenotypic profiles for drug clustering and classification.



**Figure 2. Processing and analysis of two-color fluorescent cell movies for segmentation, tracking, phenotype annotation, and quality control**

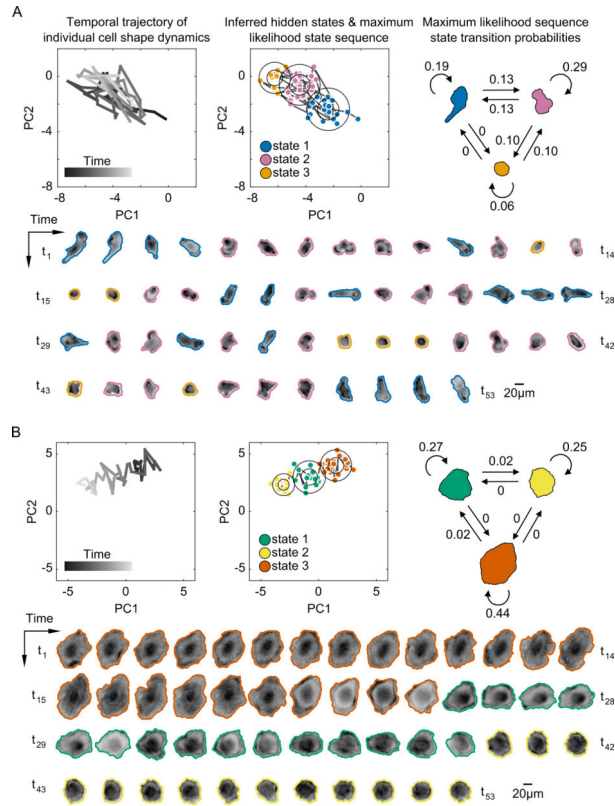
(A) Schematic of the image time series processing workflow used to generate tracked cell outlines in multi-well imaging experiments. Boxes in yellow indicate steps with user input. (B) Left, an example image time series field of view showing touching (yellow) and isolated (non-touching) cells (cyan) automatically identified in the processing pipeline. Inset shows automatic thresholds that may lead to under- or over-segmentation that are subsequently adjusted by the user with a GUI tool. Right, example of individual cells segmented and tracked using the pipeline in (A) from time series imaging over approximately sixteen hours. (C) Quality control and phenotype labeling of processed cell image time series. Two example cell trajectories imaged for eighteen hours at 20-minute intervals are shown, one undergoing cell division (left), characterized by cell rounding in mitosis followed by cytokinesis, and another undergoing death (right), with progressive cytoskeletal shrinking, non-uniformity in actin structure (green), and disruption of nuclear morphology (pink). Isolated cells are automatically labeled ‘s’ while cells touching others or the image boundary are labeled ‘b’. Dividing or dying cells are interactively labeled ‘d’ and ‘a’, respectively.



**Figure 3. Visualization and clustering of live-cell shape features reveals a morphologically diverse shape-space following drug treatments**

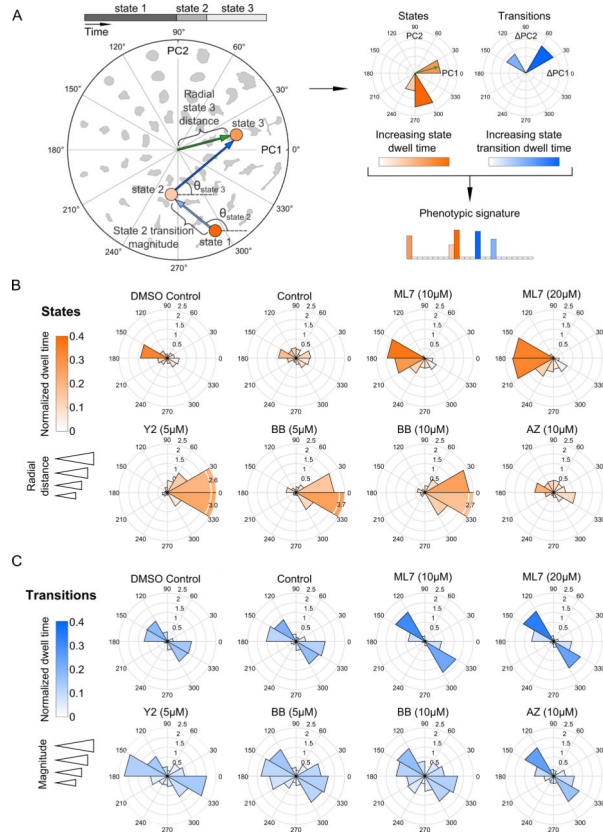
(A) PCA of approximately 20,000 temporal snapshots (gray points) from live imaging of 293 cells treated with seven drugs (inhibitors of: EGFR, Calpain, MEK (2), Myosin II, ROCK, and MLCK). Two PCs explain over 80% of the variability in the original shape features (upper right panel). Inset shows a continuous shape-space with highest point density around the mean of the data for the cell population, while single-cell trajectories form more well separated clusters in time. (B) Visualization of 28 live-cell snapshots randomly chosen in different regions (a-g) of shape-space in (A). Under the influence of the seven drugs, shapes vary from large and spread, to round with cortical actin, to varying degrees of elongation, to branched morphologies. (C) Polar coordinate PC shape-space visualization. Shapes are shown in different angular bins and radial distances from the mean of the data, which is the origin of the two PC axes in (A). Cell shapes are colored based on the radial distance quartiles within each angular bin. The orientations of the original shape features that contribute to each PC are shown. Lengths of gray lines for each feature correspond to the relative magnitudes of their PC coefficients. (D) K-means clustering of morphologies in shape-space on individual cell trajectories (examples, yellow and blue) reveals clusters that are tighter and better separated on average for all imaged trajectories (orange) compared with clustering of cells pooled together (black) or of random single-cell trajectories of the same lengths (pink).





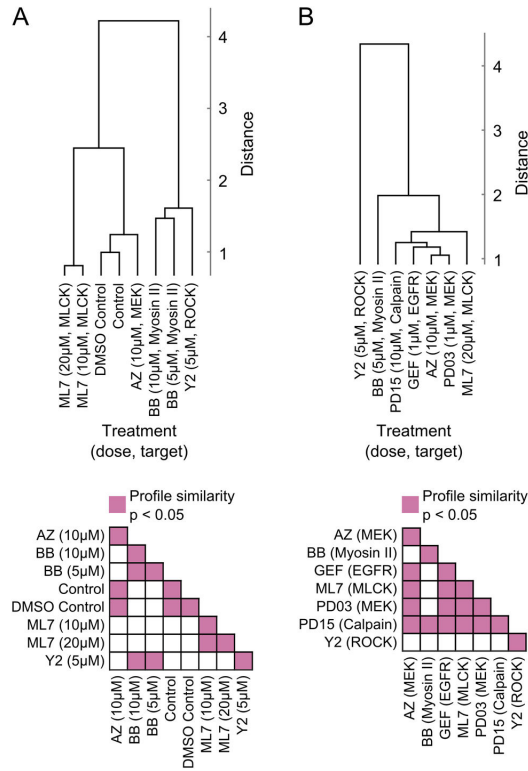
**Figure 4. Probabilistic modeling of temporal shape-space trajectories captures heterogeneous dynamic transitions in cell morphology**

(A) A cell exhibiting repetitive back-and-forth switching between three inferred morphological states of increasing elongation of the cell body from yellow to pink to blue. Top left, the temporal trajectory of the cell through PC shape-space. Top center, Gaussian shape states inferred by Bayesian HMM in SAPHIRE with circles corresponding to one and two standard deviations from the mean (circle center), respectively. Trajectory time points are colored based on the maximum likelihood (ML) hidden state inferred from the model. Top right, a diagram of the dynamic state transitions derived from the ML state sequence with numbers next to arrows corresponding to transition frequencies between, or within, states. Live-cell actin reporter images are outlined with the cell body mask boundaries and colored according to their ML states. (B) A cell with switch-like transitions between three morphological states, changing shape continuously from larger and spread to smaller and round. Shape modeling of this cell trajectory is performed independently of the cell in (A).

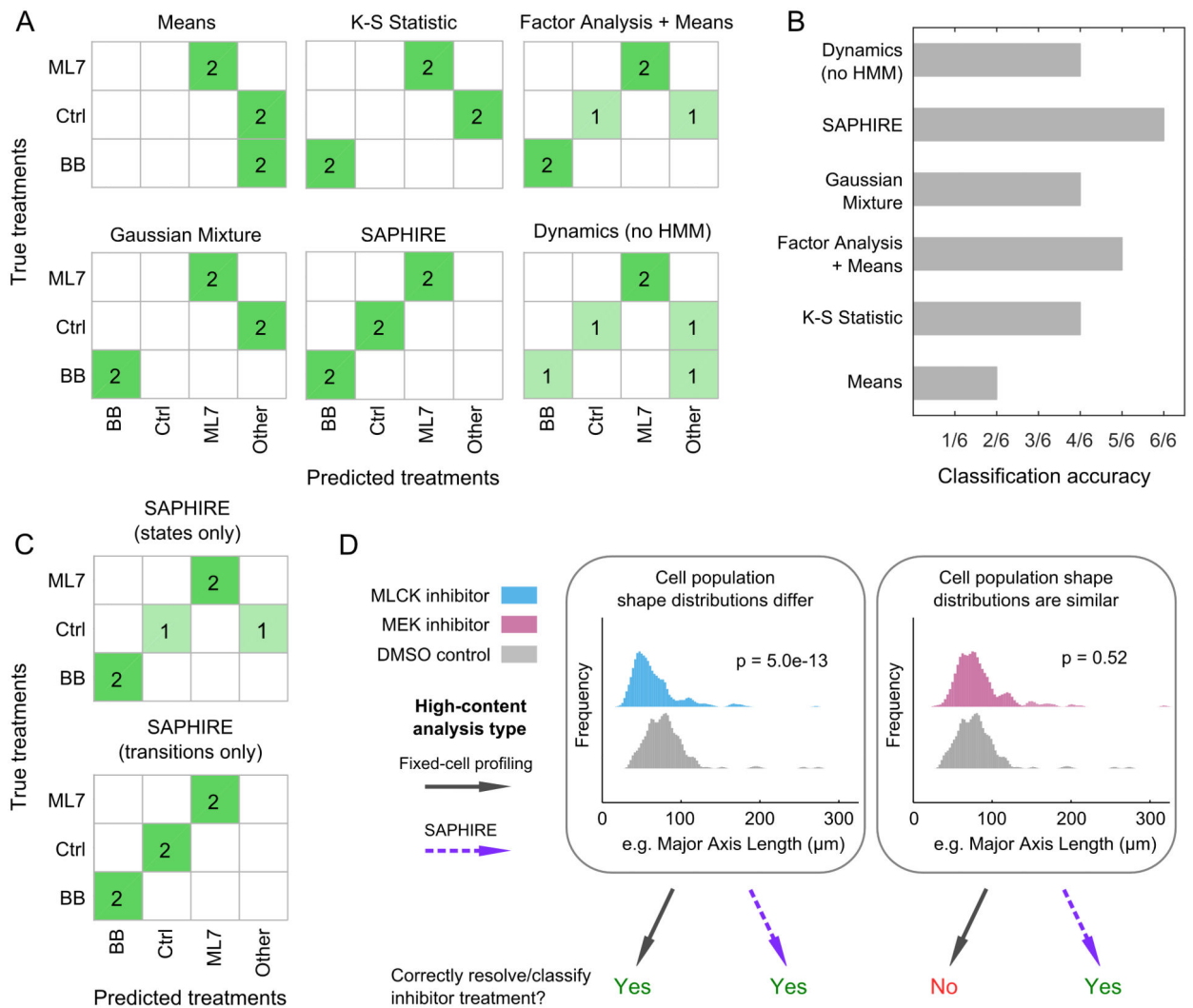


**Figure 5. Dynamic features of model-annotated shape state sequences from multiple cells enable phenotypic comparisons between experimental treatment conditions**

(A) Traversal of a hypothetical cell through polar shape-space showing shape state and state transition features for deriving a phenotypic signature of a cell trajectory. Circles are locations of inferred state means, blue arrows are state transitions, and the green arrow is an example of a radial distance of a particular state from the shape-space center. State transitions capture direction of cellular shape changes over time regardless of state location, whereas state locations capture the particular morphological properties of a cell in a given region of shape-space where the state resides. States and transitions are used to generate a phenotypic signature for each cell individually to enable comparisons between cells. (B) Polar distributions of combined states from all single-cell trajectory models for treatments tested in the imaging screen. Rose plot petals correspond to directional bins in polar shape-space as in (A), but represent average responses of all cells in a given treatment. Longer petals signify radial distances that are further from the shape-space origin. Petal color depth relates to state dwell time of cells in a given slice of polar shape-space, normalized to total trajectory length. (C) Effects of drugs on directions of cellular transitions in polar shape-space. Rose plot petals are average state transition directions of cells in a given treatment. Longer petals signify larger transition magnitudes between states, meaning that state means are farther apart. Petal color depth relates to dwell times of the target states normalized to total trajectory length.



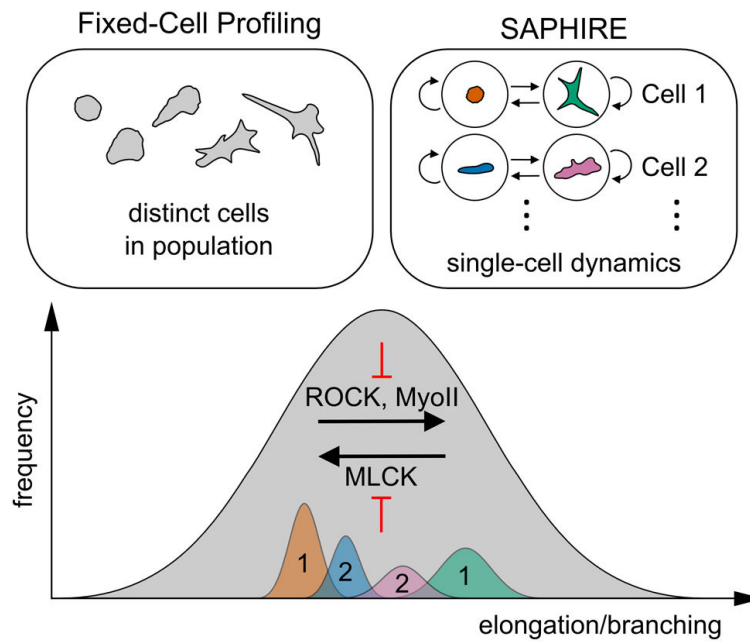
**Figure 6. Comparisons of drug effects on morphological states and state transition dynamics** (A) Top, hierarchical clustering of phenotypic profiles derived from single-cell models of shape dynamics under the effect of small-molecule inhibitors that target molecular species involved in regulating actomyosin organization. Increasing Euclidean distance indicates decreasing similarity between treatments that are grouped in the dendrogram using average linkage. Bottom, permutation tests indicate statistical significance of pairwise similarities between treatment profiles. (B) Clustering of model-derived phenotypic profiles, as in (A), of cell shape dynamics in response to treatment with an expanded panel of inhibitors.



**Figure 7. State-space modeling of cellular shape dynamics improves drug classification performance compared to existing image-based profiling methods**

(A) Confusion matrices showing nearest-neighbor classifications of individual treatments in the inhibitor screening experiment using existing fixed-cell drug profiling methods and the dynamic modeling approach presented in this work. The myosin II and MLCK inhibitor (BB and ML7, respectively) treatments are comprised of two different doses tested for each drug, “Ctrl” comprises the DMSO and non-DMSO controls, and “Other” comprises Y-27632 (ROCK inhibitor) and AZD6422 (MEK inhibitor) treatments that were only tested at a single dose and therefore not used for classification comparisons. (B) Overall treatment classification performances of the proposed and existing phenotypic profiling methods. Classification accuracy reflects the number of correctly predicted treatments out of six true treatment comparisons made in (A). (C) Confusion matrices showing treatment classification performance of SAPHIRE using either state features alone or temporal state transition features alone for generating profiles used in classification. (D) Utility of single-cell temporal modeling in phenotypic profiling depends in part on similarities of cellular feature distributions between treatments. Cellular distributions of a highly varying shape

property between cells, major axis length, are shown for pairwise treatment comparisons illustrating the benefits of single-cell temporal modeling in drug profiling and classification. Each distribution comprises individual cells measured at 1, 4, 8, 12 and 16 hours post-treatment. P-values correspond to Kolmogorov-Smirnov tests with the null hypothesis that the distributions of two treatments being compared are the same.



**Figure 8. Schematic of the conceptual difference between fixed-cell analyses versus SAPHIRE, and their combined insights into molecular mechanisms of cell shape regulation**

The distribution of morphologies of drug-treated MDA-MB-231 breast cancer cells falls on a continuum, illustrated in gray. The extent of cell elongation and branching is the dominantly varying morphological property of the breast cancer cell line analyzed in our drug screen (Fig. 3). Fixed-cell profiling captures the morphological landscape of distinct cells in the population at a given time point. In contrast, SAPHIRE captures the morphological evolution of individual cells over time. SAPHIRE uses an HMM framework to detect the presence of morphological states (e.g. orange and green states for one cell, blue and pink states for another) used to model the temporal dynamics of each cell independently within the population. Application of both of these approaches to live-cell shape profiling reveals that MLCK inhibition decreases, while ROCK and myosin II inhibition increases, cellular elongation and branching.