Electronic Supplementary Material (ESI) for Integrative Biology.

Supplementary Information

Figures S1 – S6; Table S1; Videos S1 – S3

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

Simon Gordonov[a,b], Mun Kyung Hwang[b], Alan Wells[c], Frank B. Gertler[b,d], Douglas A. Lauffenburger[a,b,d], Mark Bathe[a]

[a] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

[b] The David H. Koch Institute for Integrative Cancer Research, Cambridge, MA, USA

[c] Department of Pathology, University of Pittsburgh, and Pittsburgh VA Health System, Pittsburgh, PA, USA

[d] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

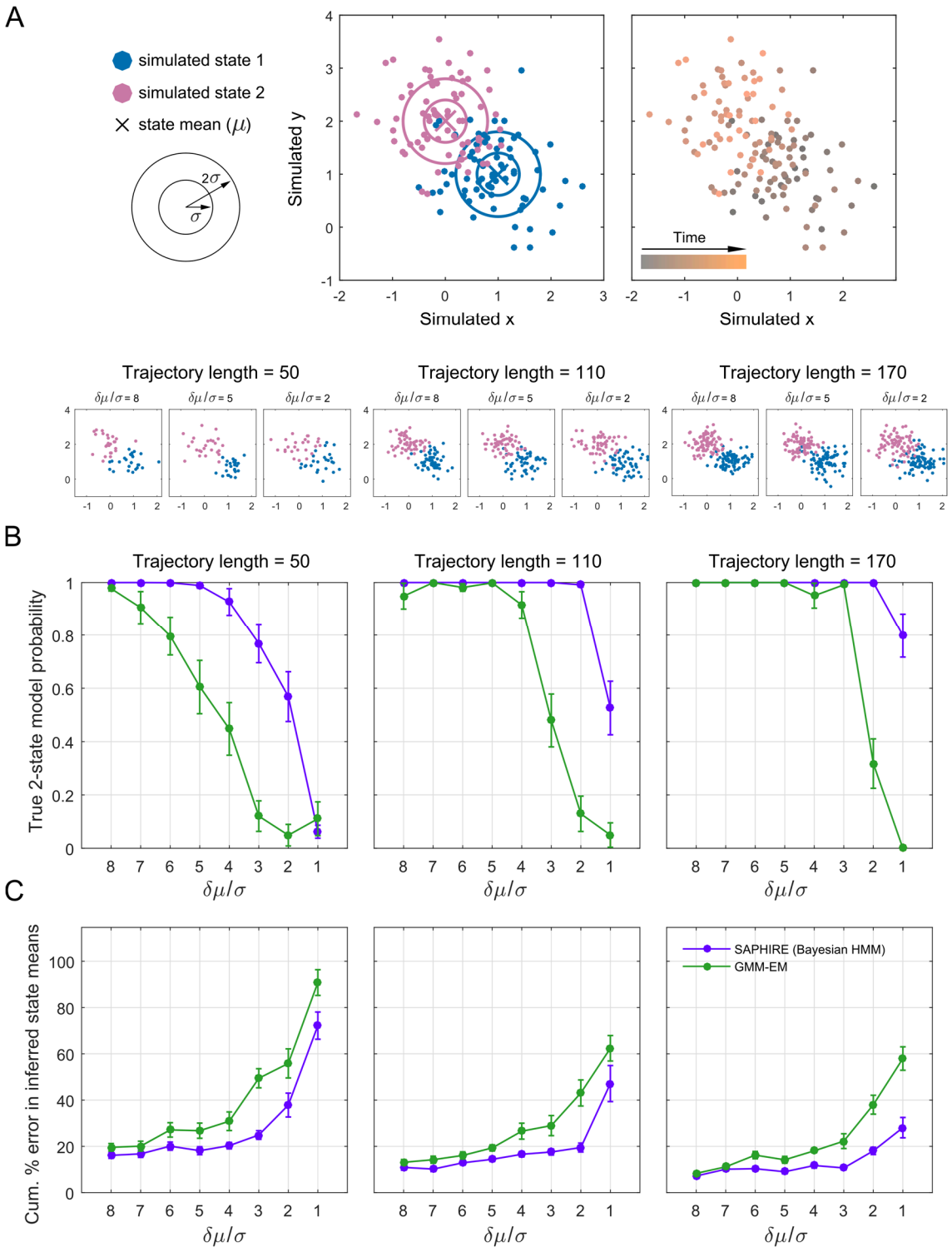
A
simulated state 1
simulated state 2
state mean (μ)
2σ
σ
Simulated y
Simulated x
Time
Trajectory length = 50
Trajectory length = 110
Trajectory length = 170
$\delta\mu/\sigma = 8$
$\delta\mu/\sigma = 5$
$\delta\mu/\sigma = 2$
B
True 2-state model probability
$\delta\mu/\sigma$
C
Cum. % error in inferred state means
SAPHIRE (Bayesian HMM)
GMM-EM

Figure S1

Figure S1. Simulations demonstrating differences in model selection and parameter estimation of Bayesian HMM in SAPHIRE versus GMM under varying degrees of state discriminability.

(A) Two bivariate Gaussian states (blue and pink) with equal variances along x and y (left panel) were used to create temporal trajectories with a single state transition. In practice, the states are unknown and must be inferred from a time series cellular trajectory of coordinates (right panel). State circles represent one and two true standard deviations from the mean (circle centers) and points are random samples drawn from the states. Simulated trajectories of different lengths (different number of samples drawn from the states) and varying resolvability of the states (how well separated the states are, $\delta\mu / \sigma$) are shown. **(B)** Comparison between Bayesian HMM and GMM ability to infer the correct 2-state model (versus a 1-state, or 3-state model) as a function of how well separated the two states are ($\delta\mu / \sigma$) for different trajectory lengths. Here, $\delta\mu$ is the Euclidean distance between the means (centers) of the two states and σ is the standard deviation of each state, set to be the same for the two states in the simulations. Error bars represent +/- standard error of the mean for 20 state-drawn samples for each $\delta\mu / \sigma$. Inclusion of temporal information in the true 2-state trajectories enables the Bayesian HMM in SAPHIRE to infer the correct 2-state model with higher probability (purple curve) compared to GMM with expectation maximization (green curve), which does not take temporal information of the trajectory into account. For both the Bayesian HMM and GMM inference methods, longer trajectories and larger separation of underlying states improved inference of the correct 2-state model. **(C)** The percent error in inferred state means for the Bayesian HMM and GMM. Cumulative percent error in the state means was calculated as $100\sum_{s}\left(|\mu_{s_a,x}-\mu_{s_i,x}|+|\mu_{s_a,\mathrm{y}}-\mu_{s_i,\mathrm{y}}|\right)\Big/\delta\mu$ where s_a is the actual (true) state (two states in these simulations), $|\cdot|$ denotes absolute value, and s_i is the inferred state closest to the true state s_a, for all 2-state inferred models, regardless of whether they are the most probable model or not.

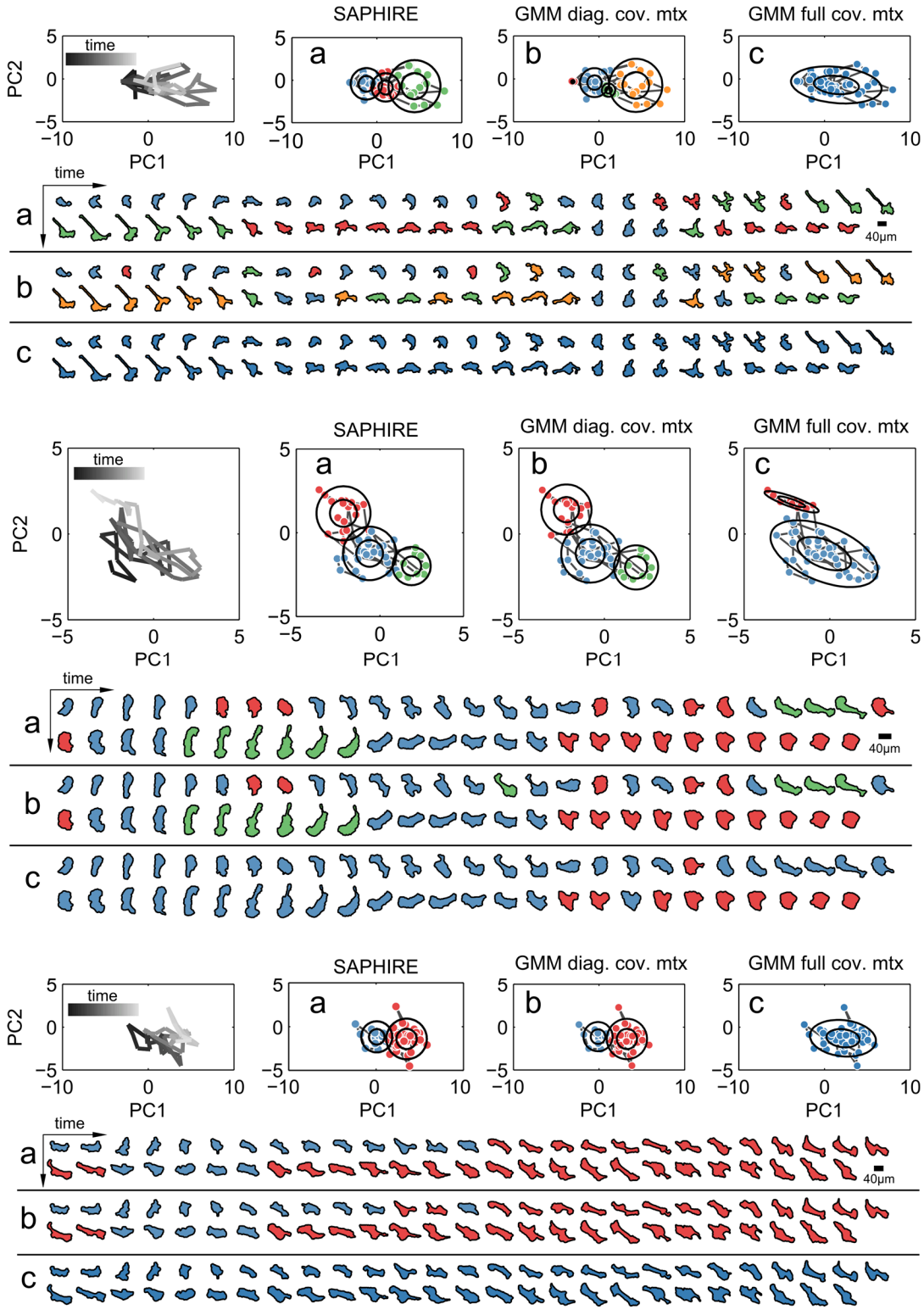
SAPHIRE
GMM diag. cov. mtx
GMM full cov. mtx
time
PC1
PC2
a
b
c
40μm

Figure S2

Figure S2. Gaussian mixture modeling (GMM) with full covariance matrix specification leads to state under-fitting and undesirable grouping of diverse morphologies into the same state.

Three examples of individual cell shape trajectories from the expanded drug panel imaging experiment modelled with SAPHIRE to derive annotated underlying shape state sequences from PCA shape-space trajectories **(a)**. The same cell shapes over time are classified into groups using GMM with a diagonal, equal-variance constraint for the covariance matrix **(b)** or full covariance matrix **(c)**, with BIC used for the GMMs to find the most probable shape state model for each cell trajectory individually, independent of other cells. The full covariance matrix GMM lumps cell with heterogeneous morphologies (e.g., rounder, elongated, branched) into similar groups, showing that Gaussian states with diagonal, equal variances (circles as opposed to ellipses) better resolve and describe the underlying morphological states of MDA-MB-231 cells in shape-space.

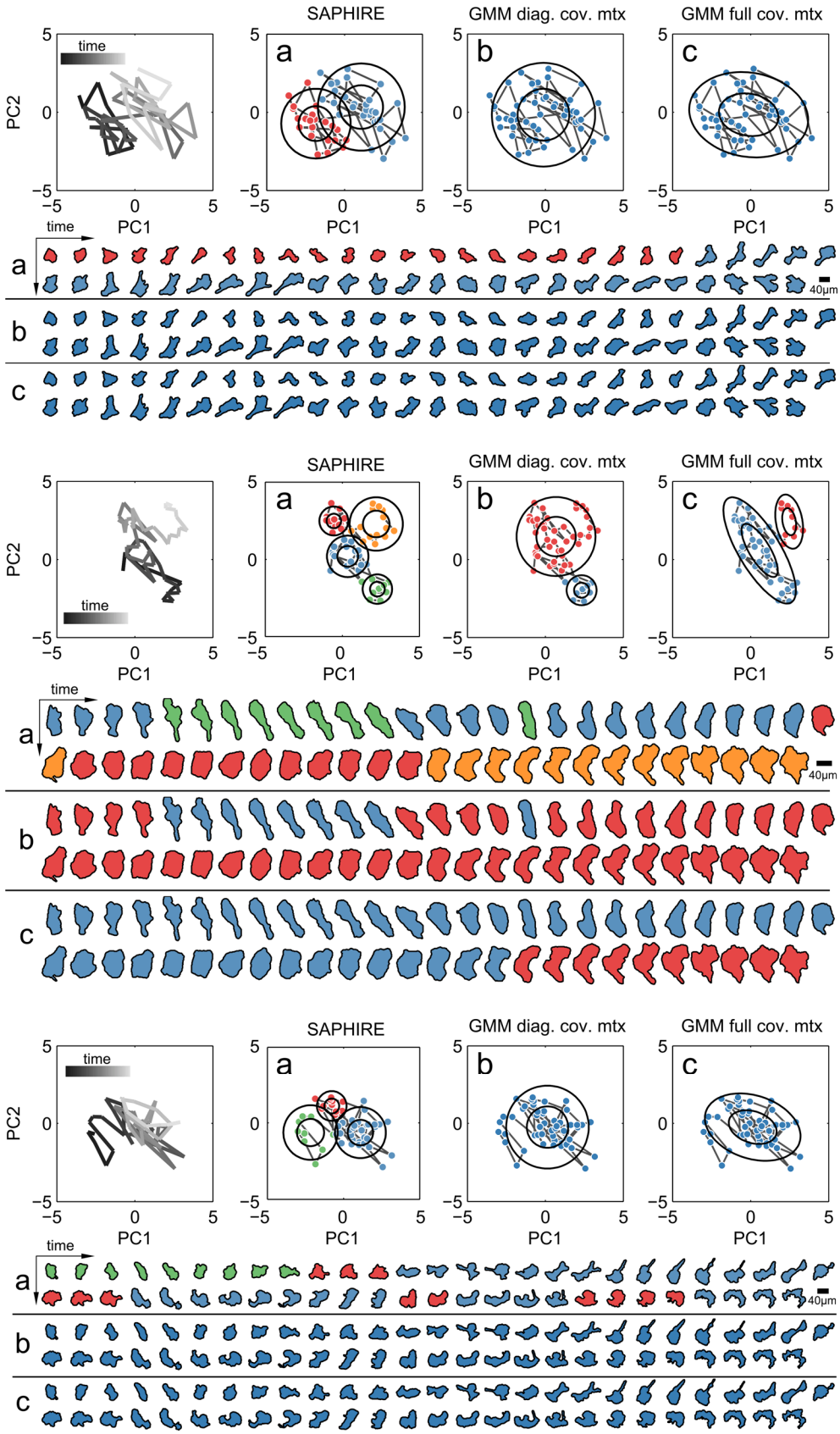
SAPHIRE
GMM diag. cov. mtx
GMM full cov. mtx
time
PC1
PC2
a
b
c
40μm

Figure S3

Figure S3. Probabilistic time series modeling using SAPHIRE better resolves cell shape states compared to GMM when cells progressively explore shape-space over time.
Three PCA shape-space cell trajectories are shown, with annotation with the most likely shape state model and state sequence using SAPHIRE **(a)**, and GMM using diagonal, equal-variance constraint on the covariance matrix **(b)** or full covariance matrix **(c)** using BIC to select the most likely model for the GMMs. The cells shown generally move through shape-space continuously in particular directions over time (e.g., left to right for the first cell in the upper panel), with SAPHIRE able to capture these states and state transitions, whereas the GMM is unable to resolve them. This is consistent with numerical simulations (Fig. S1) showing that time series information taken into account by SAPHIRE, which is neglected by the GMM, is better able to resolve, model, and annotate the underlying temporal shape state behavior of individual cells. The inability to properly resolve and capture shape states by GMM leads to under-fitting of the number of states and improper grouping of morphologically dissimilar shapes into the same state.

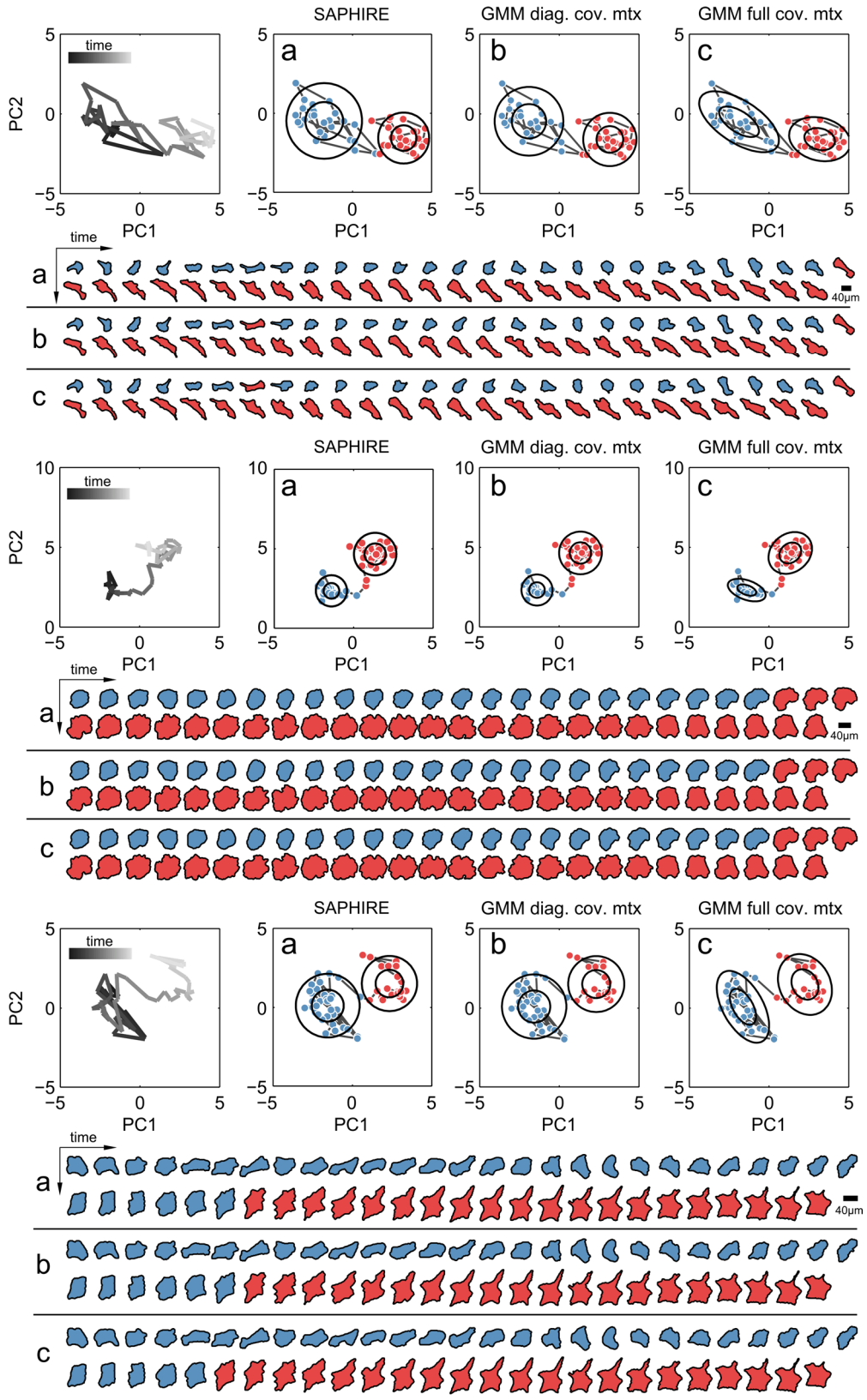
SAPHIRE
GMM diag. cov. mtx
GMM full cov. mtx
time
PC1
PC2
a
b
c
40μm

Figure S4

Figure S4. Examples of cells with similar SAPHIRE and GMM shape state annotations. Cell trajectories in shape-space were annotated with the most likely shape state model and state sequence using Bayesian HMM **(a)**, and GMM using diagonal, equal-variance constraint on the covariance matrix **(b)** or full covariance matrix **(c)**. For the GMMs, the BIC was used to select the most likely state model. When a cell moves progressively through shape-space over time and resides in well-separated shape-space regions (early time in black and later time in gray/white in left panels), Bayesian HMM and GMM categorize morphologies using the same number of states, with state transitions (blue to red) found to occur at similar points in time.

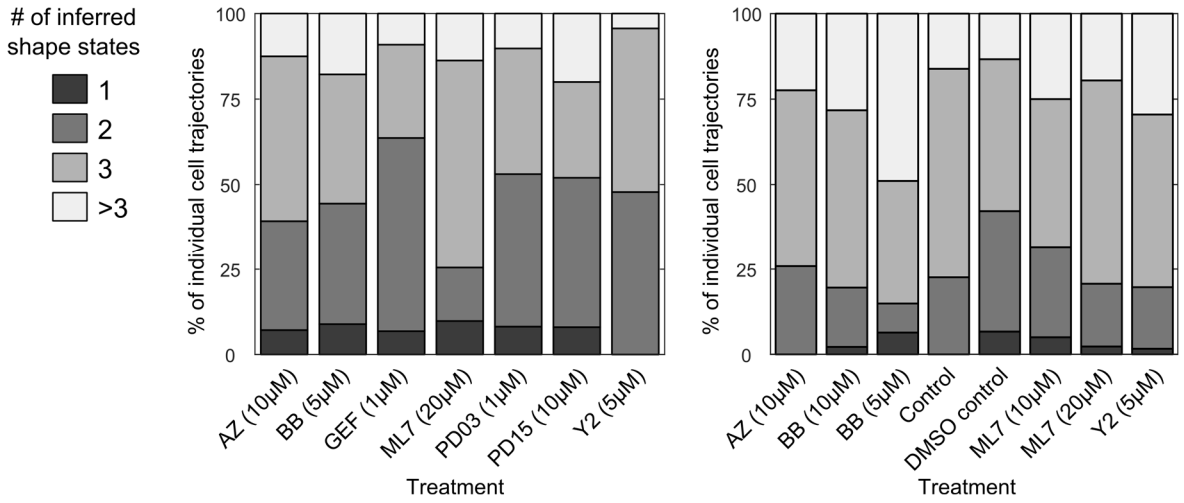

Figure S5

Figure S5. Effects of treatments on the number of morphological states explored by cells. Percentage of individual cells with given numbers of explored shape states as inferred from the SAPHIRE probabilistic models across different treatment conditions for the two imaging experiments. Modeling was applied separately for each temporal shape-space trajectory to infer the *a priori* unknown number of hidden shape states explored by each cell.

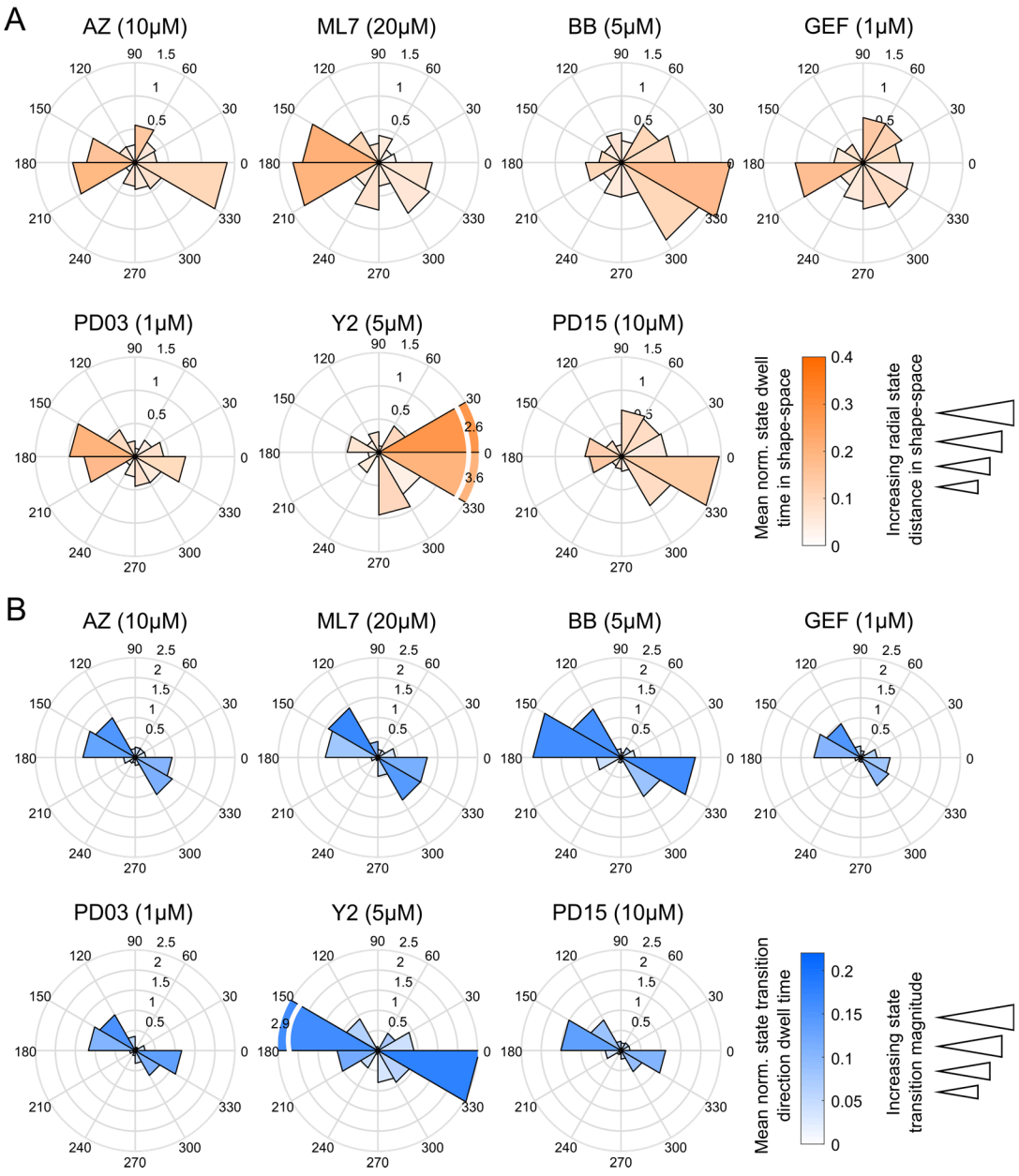

Figure S6

Figure S6. Phenotypic comparisons of an expanded panel of drugs that target actomyosin-regulatory proteins using model-annotated profiles of cell shape dynamics. **(A)** Effects of the expanded panel of drugs targeting various molecular species involved in regulating actomyosin dynamics on model-inferred cellular shape state locations in polar shape-space (as in Fig. 5B). **(B)** Effects of the expanded panel of drugs on cellular state transitions in shape-space (as in Fig. 5C).

Table S1

List and descriptions of image-derived cell shape features

Feature #	Feature Name	Feature description
1	Area	Number of pixels in the cell region.
2	Perimeter	Length of the cell region's boundary.
3	Equivalent Diameter	The diameter of a circle that has the same area as in the cell region.
4	Major Axis Length	Length of major axis of an ellipse that has the same normalized second central moments as the cell region.
5	Minor Axis Length	Length of the minor axis of an ellipse that has the same normalized second central moments as the cell region.
6	Eccentricity	The eccentricity of an ellipse that has the same second moments as the cell region.
7	Solidity	The ratio of the area of the cell region to the area of cell region's convex hull.
8	Extent	The ratio of the area of the cell region to the area of cell region's bounding box.
9	Convex Area	The area of the convex hull of the cell region.
10	Axis Ratio	The ratio of the major axis length to the minor axis length of the cell region.
11	Circularity	For the cell region, computed as: $(4*\pi*\text{Perimeter}) / (\text{Area}^2)$
12	Waviness	The ratio of the perimeter of the cell region's convex hull to the perimeter of the cell region.
13	Geodesic Diameter	The length of the longest geodesic path between all pairs of points on the boundary of the cell region. A geodesic path is the shortest path that connects two points on the cell region boundary that cannot traverse outside of the cell region.
14	Convex Perimeter	Length of the cell region's convex hull's boundary.
15	Feret Max	The maximum of the Feret lengths of the cell region over 180 directions sampled uniformly 0–360 degrees. Feret length is the measure of the cell region's size (length) along a specified direction, as would be measured with calipers.
16	Feret Min	The minimum of the Feret lengths of the cell region over 180 directions sampled uniformly 0–360 degrees.
17	Feret Mean	The mean of the Feret lengths of the cell region over 180 directions sampled uniformly 0–360 degrees.
18	Feret CV	The coefficient of variation (standard deviation divided by the mean) of Feret lengths of the cell region from 180 directions sampled uniformly 0–360 degrees.

Video S1. Segmented and tracked breast cancer cells following DMSO control treatment.
MDA-MB-231 cells stably expressing LifeAct-eGFP (green) actin reporter and Histone 2B-mCherry (pink) nuclear reporter were imaged for approximately 18 hours at 8-minute time intervals following treatment with bolus 0.1% v/v DMSO in growth media. Image acquisition began approximately 45 minutes post-treatment. Imaging was performed using time-lapse epifluorescence microscopy with a 10X/0.3NA air objective on a Nikon Eclipse Ti microscope incubated at 37°C. The video shows colored cell outlines corresponding to cell body masks along with temporal evolution of nuclear tracks (purple lines) using automated image processing prior to quality control/phenotype labeling with the SAPHIRE GUI tool (see Materials and Methods of the main text). Scale bar, 40μm.

Video S2. Segmented and tracked breast cancer cells following myosin II inhibition.
MDA-MB-231 cells stably expressing LifeAct-eGFP (green) actin reporter and Histone 2B-mCherry (pink) nuclear reporter following treatment with bolus 10μM Blebbistatin (myosin II inhibitor) in growth media. Image acquisition, processing, and video details are identical to those of Video S1. Scale bar, 40μm.

Video S3. Segmented and tracked breast cancer cells following MLCK inhibition.
MDA-MB-231 cells stably expressing LifeAct-eGFP (green) actin reporter and Histone 2B-mCherry (pink) nuclear reporter following treatment with bolus 10μM ML-7 (MLCK inhibitor) in growth media. Image acquisition, processing, and video details are identical to those of Video S1. Scale bar, 40μm.