Morphodynamical cell-state description via live-cell imaging trajectory embedding

Authors: Jeremy Copperman[1,*], Sean M. Gross[1], Young Hwan Chang[1,2], Laura M. Heiser[1,2,*] and Daniel M. Zuckerman[1,*]

Affiliations:

[1] Department of Biomedical Engineering, Oregon Health and Science University, Portland OR 97239, U.S.A.

[2] Knight Cancer Institute, Oregon Health and Science University, Portland OR 97239, U.S.A

* corresponding authors (copperma@ohsu.edu, heiserl@ohsu.edu, zuckermd@ohsu.edu)

Abstract

Time-lapse imaging provides powerful insight into the dynamical response of cells to perturbation, but the quantitative analysis of morphological changes over time is a challenge. Here, we exploit the concept of "morphodynamical trajectories" to analyze cellular behavior using morphological feature trajectory histories, rather than the common practice of examining morphological feature time courses in the space of single-timepoint (snapshot) morphological features. Our morphodynamical trajectory embedding analysis yielded quantitative and descriptive models of future time points based on the extended history information of MCF10A mammary epithelial cells treated with a panel of ligands. The trajectory analysis constructs a shared morphodynamical cell-state landscape, where the response of MCF10A cells induced by various extracellular signals is characterized by ligand-specific regulation of state transitions. Additionally, we show that including trajectories in single-cell morphological analysis enables (i) systematic characterization of cell state trajectories, and (ii) better separation of phenotypes and more descriptive models of ligand-induced differences as compared to snapshot-based analysis. This morphodynamical trajectory embedding is broadly applicable for the quantitative analysis of cell responses via live-cell imaging across many biological and biomedical applications.

Introduction

Cells can take on many different phenotypic states, which are modulated by extracellular signals; however, the quantitative understanding of complex morphological states remains a challenging problem. Single timepoint measurements provide some information about cell state but do not capture how these responses evolve over time. Live-cell imaging has long been used to characterize dynamical changes in single-cell morphology, or cellular morphodynamics[1,2]. Recent advances in live-cell imaging technologies has allowed for unprecedented resolution into the behavior and interactions of cellular populations with single-cell resolution[3–6]. To date, most analyses of live-cell image data have been primarily based upon a static classification of cell morphology, and do not directly classify the rich dynamic landscape of cell morphology trajectories[7–9].

Here we developed a generalizable morphodynamical trajectory embedding method that can be used to analyze live-cell imaging datasets composed of unlabeled phase-contrast microscopy images as well as multiplexed imaging with molecular reporters. Quantitative analysis of cell morphology in live-cell imaging typically involves the characterization of static cell images, with extracted single-cell time courses analyzed in the feature space of the snapshot (single-timepoint) cellular morphology[8–12]. Analysis methods directly based upon trajectory features have been used for classification of mitosis and apoptosis[13], and monitoring cell-signaling responses through fluorescent reporters[14]. We show here that by mapping the *trajectory space* of cells, rather than examining single-cell time courses in spaces built from snapshots, we can increase the information extracted from live-cell imaging experiments and improve the quantitative description of cellular response to extracellular signals.

Live-cell imaging provides temporal information not available from other single-cell and omics measures. Single-cell RNA sequencing (scRNA-seq) offers the ability to assay thousands of molecular read-outs across thousands or hundreds of thousands of cells, and a common data analysis procedure is the extraction of continuous low-dimensional cell-state spaces, or cell-state manifolds, from high-dimensional molecular data[15]. Because sequencing is a destructive readout, single-cell trajectories in this space can only be inferred indirectly through population time-series modeling[16–19] or pseudo-time approaches[20,21], while “RNA velocity” considers the complementary models of transcript degradation or splicing pathways[22,23] to map cell-state dynamics. Though unlabeled imaging is restricted to the observation of cellular morphology with limited information

about molecular state or information regarding the basis for dynamic changes, live-cell imaging is unique in that cell trajectories can be assayed for extended periods of time.

Live-cell imaging based cell-state analysis has been applied in a variety of contexts, with cell states defined in a variety of ways to capture different aspects of biology. In the context of high-content screening, live-cell imaging has been used to develop gene-level functional annotation, including RNAi gene knockout screens and drug screening[7]. This approach was extended to study how single-cell trajectories evolve in an assumed cell-state space consisting of 8 pre-determined cell-cycle states[8]. Gordonov et al. developed an unsupervised approach to characterize live-cell state, analyze cell shape space, and obtain models of cell response including three distinct cell states[10]. This workflow of live-cell imaging, segmentation, featurization, and tracking has been used to describe cell-state as a continuum[11] and a cell trajectory-based description of an epithelial-mesenchymal transition (EMT)[12] in the space of single-timepoint snapshot features. Notably, trajectory information, including combined motility and morphological features computed as averages over single-cell trajectories, have been used to classify cell state[24]. Our morphodynamical trajectory embedding approach differs in that we directly analyze morphological feature trajectories, rather than average morphological or cellular feature averages over time.

From the physical point of view, trajectories are the natural space from which to classify a system out of equilibrium, such as a living cell[25,26]. However, this realization is not useful without the capability to measure the trajectory space of a system of interest. Floris Takens' seminal trajectory embedding theorem[27] proves that in a deterministic, possibly chaotic dynamical system, there exists a 1:1 correspondence between the full dynamical phase space and that formed by concatenating incomplete observations of the system across time—the "trajectory embedding" space, also referred to as delay-embedding. To characterize cell morphodynamics, we take single-cell morphological features (snapshots) and concatenate them across time (trajectories), to form morphodynamical tensors. For N_f features and n_τ trajectory timepoints the morphodynamical tensor of a cell is N_f x n_τ. In stochastic systems or systems of sufficient complexity as a biological cell, a 1:1 correspondence to a deterministic dynamical map is not necessarily achievable, but trajectory embedding can still lead to an improved characterization of the dynamical behavior of such a complex, inherently stochastic system[28–30]. Trajectory embedding methodology has been applied in fields as diverse as weather prediction[31], economics[32], and molecular dynamics[33–35], but to our knowledge has not been applied in the context of classifying cell state in live-cell imaging assays.

We develop the morphodynamical trajectory embedding method in a dataset of MCF10A mammary epithelial cells in 2D culture perturbed by a set of six ligands (EGF, HGF, OSM, BMP2+EGF, IFNG+EGF, TGFB+EGF) spanning major extracellular signaling pathways and inducing distinct cellular responses. These ligands target canonical AKT/ERK, JAK/STAT, and SMAD signaling pathways and are implicated in regulating cell proliferation, differentiation state, and motility. The live-cell imaging is part of a broader data collection effort through the Library of Integrated Network-Based Cellular Signatures (LINCS) consortium[36,37] MCF10A project[38] where characterization of the molecular and phenotypic response to these ligand perturbations is explored. Molecular responses, as measured through bulk cell measurements (RNAseq, protein expression levels via RPPA, and chromatin state ATACseq) and cyclic immunofluorescence (cycIF) indicate that cellular responses under the different ligand treatments activate canonical cell signaling pathways leading to a unique cell response in each ligand condition[38]. Our live-cell imaging cell-trajectory-based analysis was developed to characterize the morphodynamical changes accompanying these molecular responses.

Results

We applied our trajectory embedding analysis to systematically characterize cell state captured via live-cell imaging of MCF10A mammary epithelial cells treated with a panel of ligands that drive distinct phenotypic responses: PBS (no ligand control), EGF, HGF, OSM, BMP2 + EGF, IFNG + EGF, and TGFB + EGF. Cells were observed via phase-contrast microscopy over 48 hours, with images collected every 30 min. An overview of the live-cell imaging trajectory embedding workflow is shown in Figure 1. The live-cell imaging assays we analyzed are part of the LINCS consortium[36,37] MCF10A project[38]. While lacking the molecular resolution of fluorescently labeled imaging, phase contrast microscopy has the benefit of minimally perturbing wild-type cell behavior, and the quantification of morphodynamical cell states demonstrates the biological information intrinsic to cellular trajectories in a broadly applicable and relatively simple imaging assay.

Comparing morphodynamical trajectories between the different ligand treatments requires the construction of a shared cell-state space, which we created by embedding the morphodynamical tensors of cells under all of the ligand treatment conditions together. Image sequences of embedded cell trajectory windows (snippets) mapped to the same location of the trajectory embedding space share morphodynamical features across treatments, see extended figure 2.

We find that in this shared morphodynamical trajectory space, ligand treatments alter the distribution of morphodynamical cell states, see Figure 2.

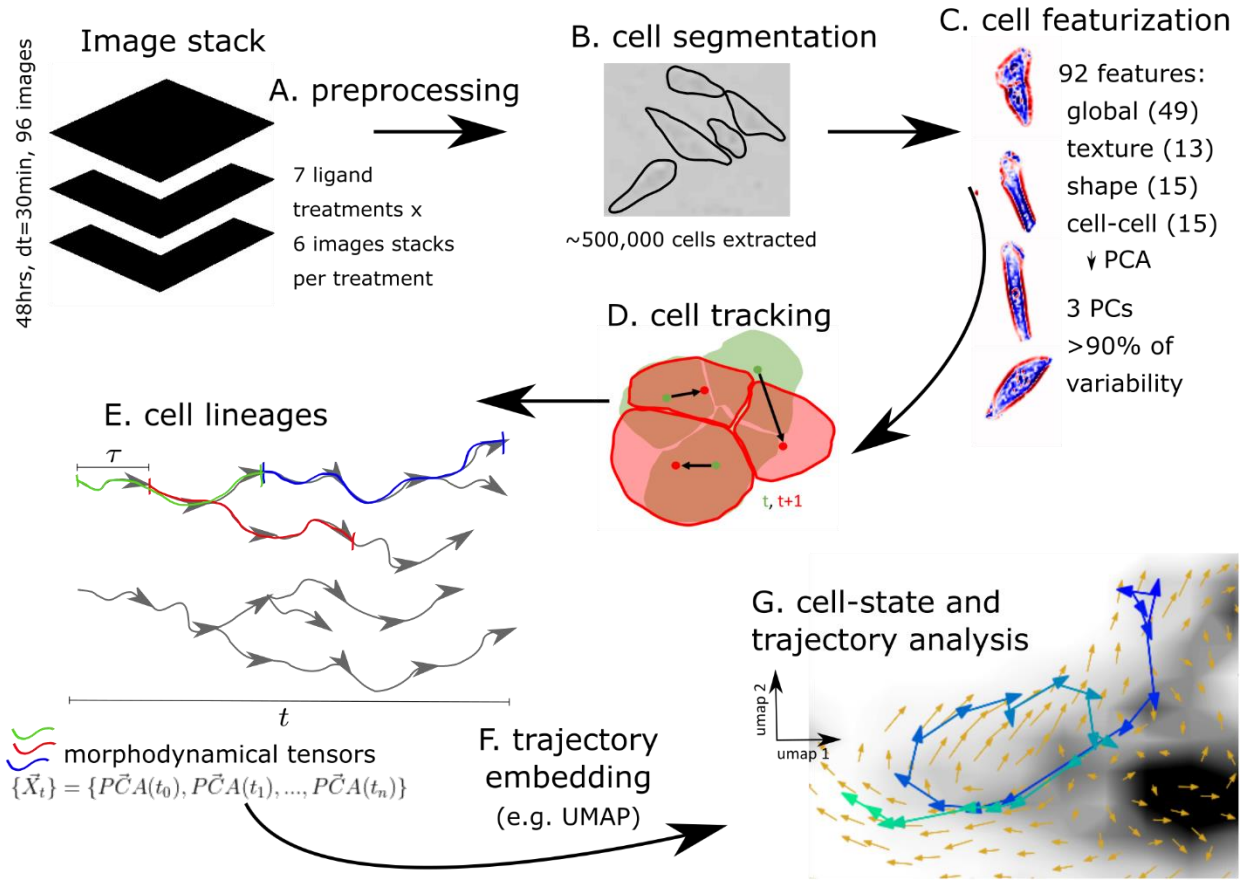


Figure 1: Live-cell imaging analysis and trajectory embedding pipeline.
Top left to bottom right: Data analysis pipeline starting from 48-hour image stacks to the morphodynamical trajectory analysis. Image processing steps include A. preprocessing (z-normalized phase-contrast pixel values colored red+ to blue-), B. cell segmentation, C. featurization, D. tracking (cell boundaries at t, t+30 min with cell centers connected by black arrows), E. extracting morphodynamical tensors (sliding window cell feature trajectory snippets) from cell lineages, F. trajectory embedding (UMAP), and G. cell-state and trajectory analysis.

Separation of unique and shared cell state under ligand perturbation

Ligand perturbation induces time-dependent morphologic and phenotypic differences in the cellular populations. Cells in the control condition (PBS, no ligand) did not proliferate, while cell populations grown in the other treatments display changes in proliferation and morphology as early as 6 hours (Figure 2A). For example, TGFB+EGF ligand treatment increased cell spacing and induced large lamellopodia, while OSM treatment induced tightly packed cell clusters; see Gross et al.[38] for further phenotypic profiling. We characterized the uniqueness of cell

morphodynamics under the different ligand treatments by quantifying the similarity between distributions of morphodynamical tensors between ligand conditions.

We found minimal ligand-specificity in the embedding space of morphological snapshots, with increased ligand specificity observed in the embedding space of morphological trajectories. At the snapshot level, which excludes trajectory information, Figure 2B (left) shows that cells occupy broad distributions in the embedding space, with distinct shifts in occupancy broadly separating OSM and TGFB+EGF from the rest of the ligand treatment conditions, consistent with qualitative visual phenotypic inspection. At a morphodynamical tensor trajectory length of 8 steps (3.5 hours), these broader relationships are preserved but the cell-state distributions in the embedded space become more condensed and display distinct peaks (Figure 2B right). The distinguishability of cell-state distributions between ligand treatments is reflected in a monotonic reduction in the shared area, or overlap, between cell-state probability distributions (Figure 2C) with increasing tensor trajectory length. The pairwise overlap decreased more rapidly than in a null model where the cell features were randomly scrambled within treatment (Figure 2C). Thus, the trajectory embedding leveraged information across timepoints leading to improved description of the ligand-specific morphodynamical responses.

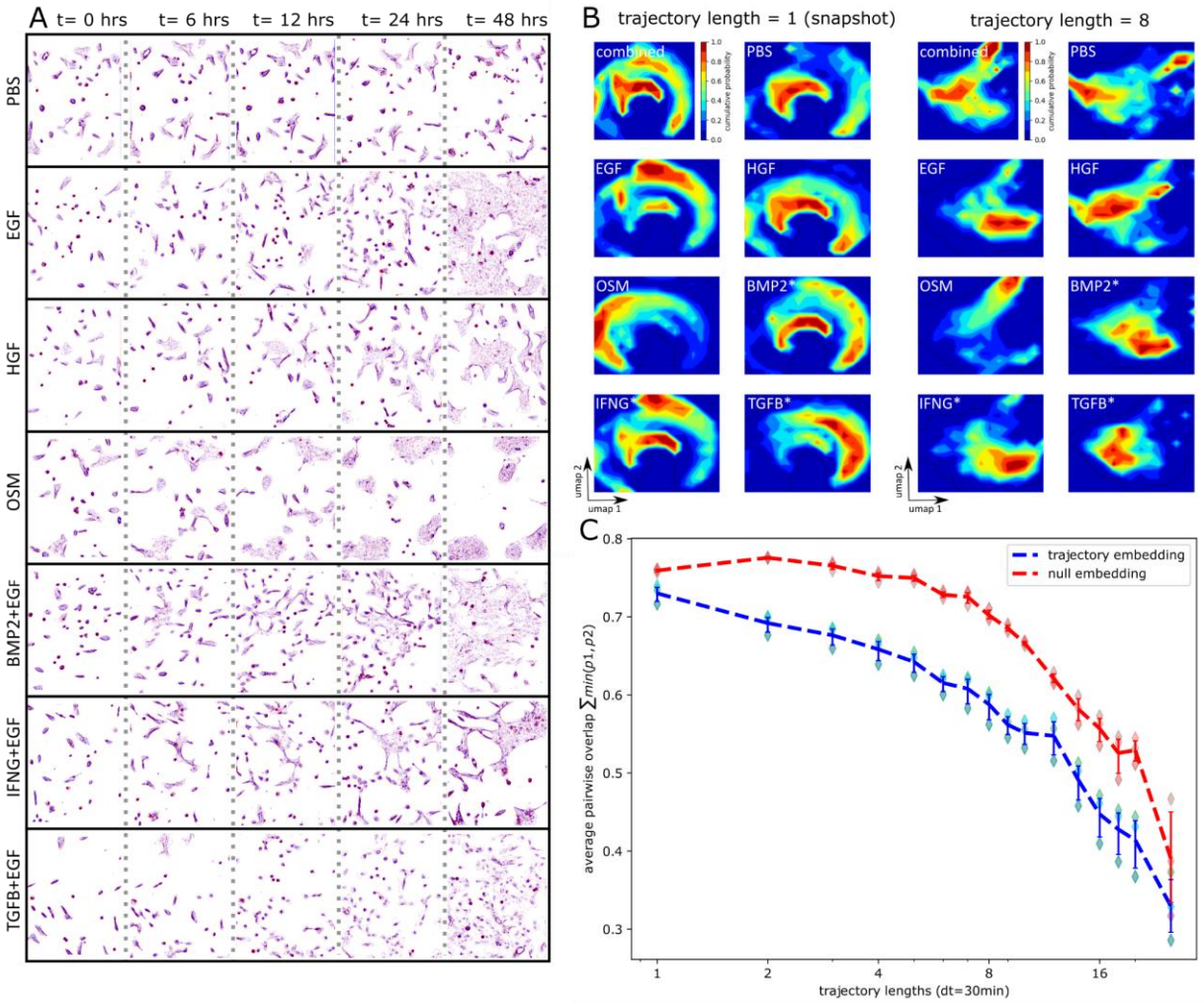


Figure 2: Trajectory embedding increases the distinguishability of cell states induced by ligand perturbation.

A) Representative background subtracted phase-contrast images (z-normalized phase-contrast pixel values colored red+ to blue-) at 0, 6, 12, 24, and 48 hrs. B) Cumulative distribution of cells in trajectory embedding space over the 48 hours of imaging. Left: snapshot space (trajectory embedding length = 1), right: trajectory embedding length =8. C) Average pairwise overlap (shared area under probability distributions) as a function of the morphodynamical tensor trajectory length used in the embedding, comparing trajectory embedding (blue dashed lines) and null model (red dashed lines), averaging over results obtained by dividing data into three sets (diamonds), with errorbars from a bootstrapped 95% confidence interval over the three data splits.

Improved cell-state description from morphodynamical trajectories

Cell states can be defined dynamically by looking for metastable regions of the morphodynamical trajectory embedding space. To compare cell states and pathways, we took trajectories that were longer than the morphodynamical tensor embedding length (3.5 hrs) and projected them onto the snippet-based landscape. The average of all cell-state trajectories passing through a local region in the landscape yields the cell-state "flow", which in a Markovian picture of a continuous stochastic process[39,40] is proportional to the effective force "pushing" a cell from one morphodynamical state to another. These cell-state force-fields are visualized in Figure 3. In the snapshot landscape, cell-state trajectories appear highly random and indicate little systematic variation between treatments (Figure 3, left column). In the trajectory embedding landscape, however, the cell-state force-field displayed treatment-specific convective flows (Figure 3, center column). These flows indicate stabilization in the unique regions of density peaks between treatments, providing direct evidence for the paradigm of metastable attractors in a landscape picture of cell state. Directly mapped trajectories in the embedded space are localized to these metastable cell states, and over the timescale of 10+ hours, cell-state changes reveal the transition pathways between metastable cell states (Figure 3, single-cell trajectories shown as blue to green lines, image sequences in the right column). Trajectories that appear random at the snapshot level unfold[41,42] and become systematic in the trajectory embedding space.

We measured how systematic and predictable the trajectories are by quantifying the randomness of the trajectories in the embedding space. A locality ratio $l = \sqrt{\langle (x(t+30min) - x(t))^2 \rangle} / \sqrt{\langle (x - \langle x \rangle)^2 \rangle}$ between the root-mean-square (RMS) displacement after one timestep (30 min) and the standard deviation in the displacement over the full population is a dimensionless measure of the predictability of a trajectory. In a completely random trajectory, this ratio is 1 since the variance in a single-timestep and the full population is identical. In a deterministic trajectory, this ratio is 0 because all trajectories emanating from the same point are identical and have no variance after a single-timestep. In a continuous, stochastic description of the trajectories in the embedded space[39,40], this locality ratio is proportional to the effective diffusion rate. Figure 4A shows that this locality ratio systematically decreases with trajectory embedding length in contrast with the null model, indicating that trajectories are increasingly systematic and predictable with increasing trajectory embedding length.

To determine the capability of the morphodynamical embeddings to characterize single-cell morphodynamical trajectories, we calculated the average log-likelihood of a validation set of

trajectories, given a model of trajectory likelihood trained with a training set of trajectories. The average log-likelihood is a direct measure of the descriptive capability of the embedding space and the predictability of the cell trajectories[43]. Here we used a Markovian transition matrix likelihood model, trained by counting transitions between k-means state centers on the landscape[44,45]. We utilized 100 k-means centers to discretize the morphodynamical embedding space, which provided sufficient state-centers to capture the observed patterns of cell-state "flow" while still retaining adequate sampling of state-state transitions. The average log-likelihood increases as a function of morphodynamical tensor embedding length and is higher than in a null model where cell features were randomly scrambled between treatments, shown in Figure 4B. We expect in general that greater trajectory embedding lengths will increase the descriptive capability of trajectory models, but only up to the point where adequate data quantity is obtained. With the cell segmentation and tracking challenges posed by the unlabeled bright field imaging data, longer trajectory embedding lengths indicate that less trajectory data was available, see extended data table 2. We chose a trajectory embedding length of 8 (3.5 hrs) for further analysis, because this resulted in increased trajectory likelihoods while retaining adequate sampling. Even partial trajectory information over a timescale of a few hours substantially improved the representation of cell state.

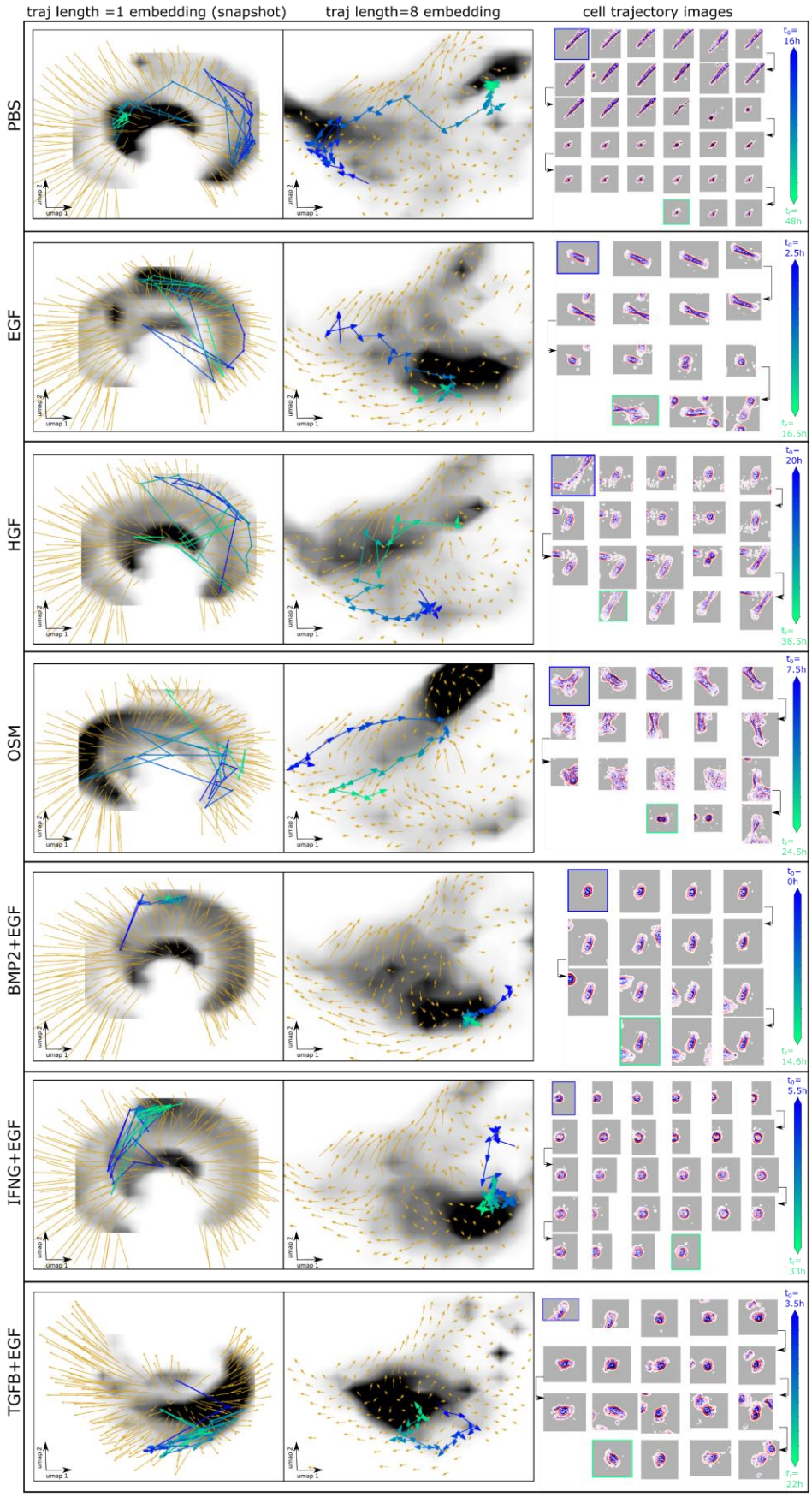


Figure 3: Trajectory embedding enables the determination of metastable cell states and pathways across ligand treatments. Left and middle column: average displacement proportional to the effective force (orange arrows), and cell density (grayscale) for snapshot embedding (left: snapshot, trajectory length = 1, right: trajectory length = 8). Representative cell morphodynamical trajectories (blue to green line with arrows showing the direction of motion in the embedding space) from t_0 to t_f determined by the available cell tracks. Right column: Cell images every hour along the extracted trajectory.

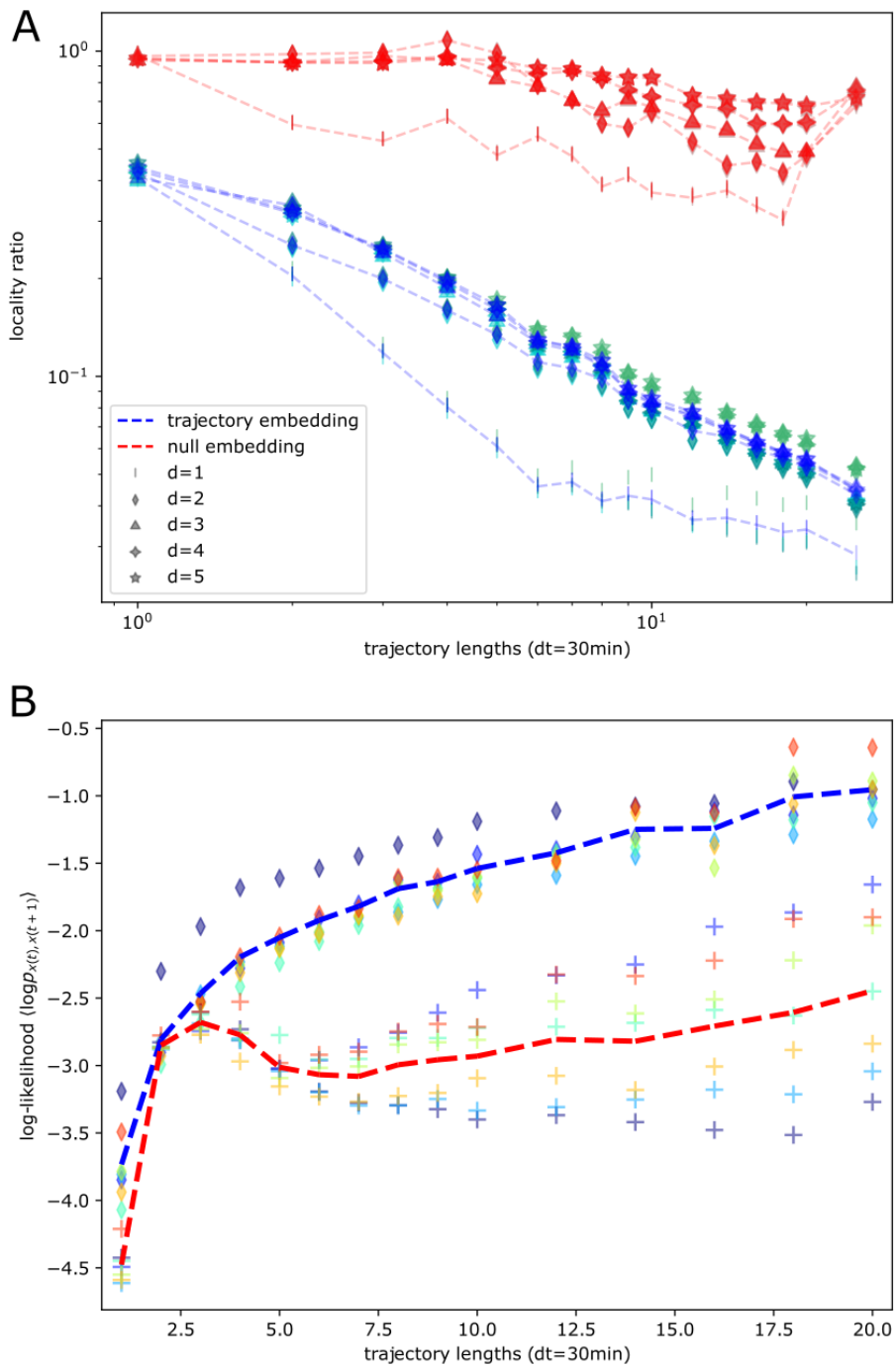


Figure 4: Trajectory embedding increases the predictability of cell trajectories.
A) Ratio between the single-step (dt=30min) and full rms displacement in the trajectory embedding space as a function of the trajectory tensor length, null model with randomly scrambled features within treatments (reds) and trajectory embedding (blues), and UMAP embeddings with d=1 (lines), d=2 (diamonds), d=3 (triangles), d=4 (squares), and d=5 (stars). Three replicates are shown per embedding (sea green, turquoise, teal).
B) Average log-likelihood per trajectory step from the validation set cell trajectories, as a function of the trajectory length, averages for the trajectory embedding (blue dashed line) and for the null model (red dashed line), from UMAP d=2 embeddings. Individual treatments (PBS, EGF, HGF, OSM, BMP2+EGF, IFNG+EGF, TGFB+EGF) for the null model (crosses) and for the trajectory embedding (diamonds).

Morphodynamical transitions precede cell-cluster formation

The direction of the cell-state flow in the trajectory embedding space indicates that a group of cells coming together and forming an attached multi-cellular cluster proceeds via a morphodynamical intermediate that resembles a mesenchymal-like morphological state. This flow was consistently observed in the clockwise (from right to left, then up) flow present in the cell-state "force-fields" (Figure 4), cell-state transition networks (Figure 5A), and the time-dependent cell-state distributions (Figure 5C). This flow from isolated cells into multi-cellular clusters is observed in all of the ligand conditions despite morphological differences in multicellular clusters between treatments.

We identified metastable states spanning the cell-state landscape from density peaks arising in individual treatments (Figure 5A), which we grouped into 6 macrostates. We first utilized an unsupervised kinetic clustering approach[46] which separated three major metastable basins consistent with the regions of the landscape enriched in the EGF, TGFB+EGF, and OSM conditions (respectively lower right, lower left, and upper regions, Figures 2B and 5C). We manually refined these regions to distinguish between clustered cell states (states 4, 5, and 6) differentially occupied between ligand treatment, and assigned biologically meaningful labels to these macrostates based upon observed cell morphology. Cells were mapped into these macrostates by first finding the closest metastable state (Figure 5 A.-S.) in the embedding space, and then assigning the macrostate label accordingly. Ligand treatments were distinctly characterized by the transition network between the states (Figure 5A), indicating ligand-specific regulation of cell morphodynamics. In general, cell-state distributions were more similar across ligand treatments at early times, and became more condensed and distinct at later times, reflected in the time-dependent morphodynamical cell-state distribution (Figure 5C). Separated cells (states 1 and 2) are kinetically separated from cell clusters (states 4 and 5) and as cells grow to confluence, cell cluster formation proceeds via cells displaying extended cytoskeletal features and increased cell border contrast (Figure 5B).

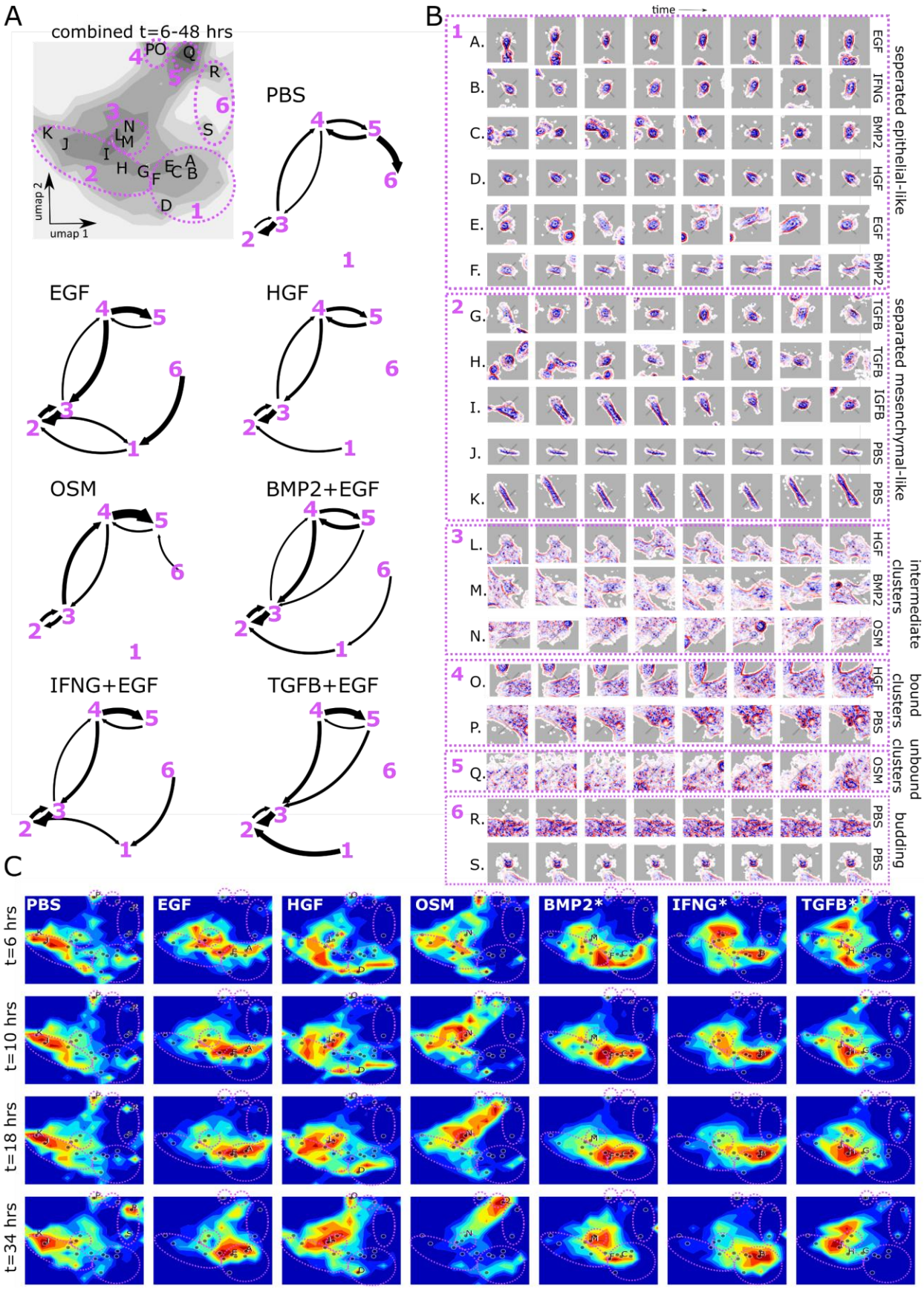

A
combined t=6-48 hrs
umap 1
umap 2
PBS
EGF
HGF
OSM
BMP2+EGF
IFNG+EGF
TGFB+EGF
B
time
separated epithelial-like
separated mesenchymal-like
intermediate clusters
bound clusters
unbound clusters
budding
C
t=6 hrs
t=10 hrs
t=18 hrs
t=34 hrs
PBS
EGF
HGF
OSM
BMP2*
IFNG*
TGFB*

Figure 5: Trajectory embedding resolves pathway of cell cluster formation via mesenchymal-like intermediate.
A) Upper left: Cumulative distribution of all cells under all treatments (grayscale) with labeled individual density peaks (A-.-S.), and overall qualitative cell-state network (mauve dashed circles) with numeric labels (1-6). Cell-state transition networks with arrow weight proportional to conditional transition probability, with transition probabilities <3% not drawn. B) Representative cell image trajectory snippets (embedding length = 8 = 3.5 hrs) extracted at density peak locations (A.-S.), with labeled treatment conditions and state definitions (right of images) and density peak locations chosen from time-dependent distributions (right panel C, labels A-S). C) Time-dependent cumulative distributions (rainbow) with local cell-state flow (local average trajectory displacement, white arrows) from 12-hr windowed averages (or maximum allowed window average). *(+EGF)

Discussion and Conclusions

In vivo, cells continually modulate their phenotypic state in response to the local microenvironment[36]. During development, cells must precisely control cell states[47], responses to extracellular signals, and cell motility, while the loss of cell state control typifies disease states such as cancers[48]. These dynamic cell behaviors are observed via live-cell imaging but it can be difficult to quantify the relationships between diverse cellular phenotypes that are dynamically interchanging. Our morphodynamical trajectory embedding provides a method to quantitate dynamic morphologic behaviors.

Trajectory embedding leverages the unique capability of live-cell imaging to follow single-cells in time, and constructs a space suited to study cell states and their dynamical relationships. Cell state plasticity has been identified as a mechanism of therapeutic resistance, and a therapeutic target in the context of advanced cancers[49–55]. Cell state has long been defined as a metastable attractor in a high-dimensional landscape, a paradigm famously depicted in Waddington's epigenetic landscape[56–58]. Live-cell trajectory embedding brings the cell-state landscape paradigm from theoretical biology to direct application, where cell-states and the pathways of cell-state changes can be resolved, validated, and potentially leveraged for actionable control strategies[17,59–64].

Our morphodynamical trajectory embedding procedure quantifies the space of morphological trajectories directly, leading to an improved description of dynamical cell-state changes compared to using only morphological snapshots. We applied the methodology in the context of unlabeled phase-contrast imaging of cell cultures which approach confluence[65,66], without paired ground

truth labeled images[67], suggesting some robustness to cell segmentation and tracking errors. The morphodynamical trajectory embedding method is general and can be applied to any live-cell imaging modality where cells can be characterized and tracked through time, and we expect morphodynamical trajectory embedding will increase the information extracted from live-cell imaging with multichannel molecular labeling as well.

Biological interpretation and validation of the morphodynamical cell states extracted here is required. Our observation that cell cluster formation is preceded by a mesenchymal-like shift in cell state aligns with the maturation of transverse arc stress fibers as a precursor to stable cell-cell junctions observed by Rajakylä et al,[68] but data integration with fixed, molecularly-labeled multiplexed imaging, as well as single-cell transcriptomics and proteomics, are critical to develop the applicability and utility of the information obtained from live-cell imaging alone. Manifold-based or mutual-information approaches have had some success with single-cell data integration[69–71], and may be useful in this regard to integrate live-cell imaging trajectory embeddings with molecularly resolved data, a critical data analysis goal needed to provide insight into the biological relevance of morphodynamical cell states.

We intend to explore independent approaches to determine the molecular mechanisms underlying morphodynamical cell states, with the broader goal of addressing the dynamical nature of cell state change, cell-state regulation via extracellular signals and cell-cell signals, and dysregulation in the context of human disease and cancer. With the trajectory embedding method we present here, we can now study the emergence of metastable attractors and the regulation of dynamic cell-state changes, directly confirmed via live-cell trajectories.

Methods

Live-cell imaging of MCF10A cells Cells were cultured with 7 different ligand treatments (PBS, EGF, HGF, OSM, BMP2 + EGF, IFNG + EGF, TGFB +EGF), imaged every 30 minutes for 48 hours via bright-field phase contrast with an Incucyte microscope (1020x1280, 1.49 μm/pixel). 6 image stacks were collected for each ligand treatment. Experimental protocols can be found in detail at the publicly available Synapse database[38].

Image preprocessing Foreground (cells) and background pixel classification was performed using manually trained random forest classifiers using the ilastik software[72]. Images were z-

normalized (mean subtracted and normalized by standard deviation) and background pixel values were set to a value of 0. In cell images, these z-normalized pixel values are shown from red to blue (positive to negative).

Cell segmentation Single-cells were segmented from the preprocessed images using the cytoplasm model of the Cellpose[73] software, a deep learning approach trained broadly across cell types and imaging modalities. The Cellpose algorithm requires estimating the size of the cells before segmentation; due to the variability in sizes and shapes of the MCF10A cells, segmentation was performed iteratively over multiple rounds until no more new cells were found (pixel values of segmented cells set to 0). Image preprocessing and segmentation scripts can be found on the github repository, see data and code availability. The unlabeled, bright-field phase-contrast imaging used here leads to image analysis challenges, particularly for cell segmentation. It is difficult to judge the quality of many extracted cell segmentations when local cell density is high, see extended data table 1 for manual validation.

Cell featurization Segmented cells were extracted, and mask-centered into zero-padded equal sized arrays larger than the linear dimension of the biggest cell (in each treatment). Principal components of each cell were aligned, and then single-cell features were calculated. Zernike moments (49 features) and Haralick texture features (13 features) were calculated in the Mahotas[74] image analysis package. The sum average Haralick texture feature was discarded due to normalization concerns. Shape features (15 features) were calculated as the absolute value of the fourier transform of the distance to the boundary as a function of the radial angle around cell center[75], with the set of shape features normalized to 1. Cell environment was featurized in a similar fashion, where an indicator function with value 0 if the cell boundary was in contact with the background mask, and value 1 if in contact with the cell foreground mask. Absolute value of the Fourier transform of this indicator as a function of radial angle around cell-center then featurized the local cell environment (15 features), with the sum of cell environment features normalized to 1. Note the first component of the cell environment features is practically the fraction of the cell boundary in cell-cell contact. The high-dimensional cell feature space was dimensionally reduced using principal component analysis (PCA), retaining the largest 3 eigen-components of the feature covariance matrix (spanning all treatments and image stacks) which captured >90% of the variability.

Cell tracking Image stacks were first registered translationally using the pystackreg implementation of subpixel registration[76]. Cells were tracked between frames by first separating each contiguous cluster of cells. If cell clusters were less than 10,000 pixels (typical cell very roughly 30x30 pixels), they were simply tracked by minimum distance with a cutoff of 45 pixels. For larger cell clusters, clusters were first tracked by minimum distance with a cutoff of 300 pixels. Tracked cell clusters were each individually registered rotationally and translationally (again using pystackreg), and individual cells in the clusters were tracked between frames by maximum overlap with a cutoff of 10 pixels.

Morphodynamical tensor embedding All partial cell lineages were contructed from available cell tracks. These cell lineages are not complete, that is incomplete segmentation/tracking means that cells cannot be traced all the way back to an initially plated progenitor cell. The unique backwards trace of any extracted single-cell through time is the available cell history. Of the total 476,855 cells extracted here; 137,845 could be traced back only one step, while 36,919 could be traced back 10 steps, see extended data table 2. We consider the set of all cells (from possibly different experimental timepoints) with cell history up to a given length, the trajectory snippet set. Note there is a large amount of duplication in this sliding window division of available cell trajectories. Cell trajectories which are longer than the trajectory snippet length used in the trajectory embedding allow for the determination of cell-states and pathways. From the lineages, all cell trajectory snippets of length n_τ (all possible cell histories with length n_τ) were extracted in a sliding window manner. The number of available trajectory snippets for each treatment are shown extended data table 2. The PCs of each cell in the trajectory snippet were then concatenated together (e.g. for 2 PCs $\{\overrightarrow{X^{n_\tau}}\} = \{PCA_1(t_0), PCA_2(t_0), \dots, PCA_1(t_1), PCA_2(t_1), \dots, PCA_1(t_{n_\tau}), PCA_2(t_{n_\tau})\}$ to form the trajectory snippet supervector, which we define as the morphodynamical tensor of length n_τ (with N features, the tensor for each cell is N x n_τ). These morphodynamical tensors (spanning all image stacks and all treatments) are flattened into vectors and then embedded using UMAP[77] into a space of dimension d. The trajectory embedding analysis allows for the robust and systematic characterization of cell state trajectories even in this challenging data analysis regime with many missing and partially segmented cells.

Overlap coefficient To compare the similarity of two probability distributions over a shared space, we use the overlap coefficient[78] defined by the sum of the minimum value of two probability

distributions. The overlap is 0 for completely distinct non-overlapping distributions, and 1 for identical distributions.

Stochastic dynamics: locality ratio, cell-state force-fields A measure of the randomness of motion in stochastic dynamics is the effective diffusion rate defined at a timescale τ by $D \equiv < \Delta x^2 >_\tau / \tau$ with here τ the time between frames of 30 minutes. We characterize how random trajectories are by the ratio of the single-step rms (root-mean-square) displacements to the total RMS displacement, a locality ratio $l = \sqrt{\langle (x(t+30min) - x(t))^2 \rangle} / \sqrt{\langle (x - \langle x \rangle)^2 \rangle}$ where the mean-squared displacements are summed over dimension. For completely random trajectories, this ratio is 1, and tends to zero for a continuous, deterministic dynamic. If cells are obeying a stochastic Markovian dynamics characterized by a diffusion equation, then the average displacement is proportional to the effective force via $< \Delta \vec{x} >= \hat{\gamma}^{-1} \vec{F}(\vec{x}) \Delta t$ with γ the friction tensor.

Cell state clustering and prediction The embedded space was binned into "microbins" using k-means clustering with $k = 100$ clusters. Qualitative results are similar as long as there are enough clusters to capture features of the continuous embedding space. In this discrete space, a transition matrix between bins was estimated from transition counts as $T = C_{ij}/C_i$ and $C_i = \sum_j C_{ij}$. Note that this transition matrix does not share a steady-state distribution with the cell populations, as cell birth and death states are not included. This transition matrix was used as a (highly simplified) model of the single-step trajectory likelihood.

Trajectory likelihood Data was split into a test set (5/6 images stacks per treatment) and a validation set (1/6 image stacks per treatment). The test set was used to train a transition matrix likelihood model, and the average log-likelihood per trajectory step was calculated from the validation set trajectories using the transition matrix as $< L >= \frac{1}{N_2} \sum_{x0=i, x1=j} \log(T_{ij})$ with N_2 the set of all 2-step trajectories in the validation set (initial point $x0$ mapped to bin i and next point $x1$ mapped to bin j.

Cell metastable state extraction and grouping Cell states spanning the trajectory embedding landscape (19 total) were picked via density peaks in the individual treatments. These were grouped via proximity into 6 qualitative cell macrostates, and used to define cell macrostate transition networks (transition matrix, see **Cell state clustering and prediction**). Cell state

names are descriptive for ease of interpretation but not based upon validated biological interpretation.

Statistical analysis. Overlap and locality ratio results were validated by calculating over 3 replicates of the data, split into 3 groups composed of 2 image stacks for each treatment. Means over the replicates, and the individual replicate data points are plotted to allow visual estimation of the robustness of the analysis. Errorbars and 95% confidence intervals are estimated from the data splits by Bayesian bootstrapping[79].

Acknowledgements

We thank Ian McLean and Mark Dane for project guidance and assistance accessing LINCS data, David Aristoff and Gideon Simpson for mathematically oriented discussion insight, and John Russo and Luke Ternes for input regarding computational implementation. J.C. is supported by the Damon Runyon Cancer Research Foundation Quantitative Biology Fellowship. Y.H.C is supported in part by the National Cancer Institute (U54CA209988, U2CCA233280, U01 CA224012). D.M.Z. acknowledges support from the OHSU Center for Spatial Systems Biomedicine. These studies were supported in part by NIH research grants U54-CA209988 and U54-HG008100, and the Anna Fuller Foundation.

Competing Interest

The authors declare no competing interests.

Data and Code Availability

All codes and scripts to perform the analysis in this work can be found at the project github repository.

https://github.com/jcopperm/celltraj

LINCS MCF10A Molecular Deep Dive data is available in some formats from the synapse database[38] and additional data is available upon request.

References

1. Shao, D., Rappel, W. J. & Levine, H. Computational model for cell morphodynamics. *Phys. Rev. Lett.* **105**, 108104 (2010).
2. Ma, X., Dagliyan, O., Hahn, K. M. & Danuser, G. Profiling cellular morphodynamics by spatiotemporal spectrum decomposition. *PLoS Comput. Biol.* **14**, e1006321 (2018).

3. Chen, B. C. *et al.* Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science (80-.).* **346**, 1257998 (2014).
4. Stephens, D. J. & Allan, V. J. Light microscopy techniques for live cell imaging. *Science* vol. 300 82–86 (2003).
5. Specht, E. A., Braselmann, E. & Palmer, A. E. A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annual Review of Physiology* vol. 79 93–117 (2017).
6. Liu, Z., Lavis, L. D. & Betzig, E. Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level. *Molecular Cell* vol. 58 644–659 (2015).
7. Neumann, B. *et al.* High-throughput RNAi screening by time-lapse imaging of live human cells. *Nat. Methods* **3**, 385–390 (2006).
8. Held, M. *et al.* CellCognition: Time-resolved phenotype annotation in high-throughput live cell imaging. *Nat. Methods* **7**, 747–754 (2010).
9. Nketia, T. A., Sailem, H., Rohde, G., Machiraju, R. & Rittscher, J. Analysis of live cell images: Methods, tools and opportunities. *Methods* vol. 115 65–79 (2017).
10. Gordonov, S. *et al.* Time series modeling of live-cell shape dynamics for image-based phenotypic profiling. *Integr. Biol.* **8**, 73–90 (2016).
11. Chang, A. Y. & Marshall, W. F. Dynamics of living cells in a cytomorphological state space. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21556–21562 (2019).
12. Wang, W. *et al.* Live-cell imaging and analysis reveal cell phenotypic transition dynamics inherently missing in snapshot data. *Sci. Adv.* **6**, eaba9319 (2020).
13. Huh, S., Ker, D. F. E., Su, H. & Kanade, T. Apoptosis detection for adherent cell populations in time-lapse phase-contrast microscopy images. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 7510 LNCS 331–339 (Springer Verlag, 2012).
14. Jacques, M., Dobrzyński, M., Gagliardi, P. A., Sznitman, R. & Pertz, O. CODEX, a neural network approach to explore signaling dynamics landscapes. *Mol. Syst. Biol.* **17**, e10026 (2021).
15. Moon, K. R. *et al.* Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Current Opinion in Systems Biology* vol. 7 36–46 (2018).
16. Chapman, M. P. *et al.* A model of phenotypic state dynamics initiates a promising approach to control heterogeneous malignant cell populations. in *2016 IEEE 55th Conference on Decision and Control (CDC)* 2481–2487 (IEEE, 2016). doi:10.1109/CDC.2016.7798634.

17. Chapman, M. P. *et al.* Modeling differentiation-state transitions linked to therapeutic escape in triple-negative breast cancer. *PLOS Comput. Biol.* **15**, e1006840 (2019).
18. Karacosta, L. G. *et al.* Mapping Lung Cancer Epithelial-Mesenchymal Transition States and Trajectories with Single-Cell Resolution. *bioRxiv* 570341 (2019) doi:10.1101/570341.
19. Cho, H. & Rockne, R. C. Mathematical modeling with single-cell sequencing data. *bioRxiv* 710640 (2019) doi:10.1101/710640.
20. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
21. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. *Cell* **157**, 714–725 (2014).
22. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
23. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
24. Wu, Z. *et al.* DynaMorph: Learning morphodynamic states of human cells with live imaging and sc-RNAseq. *bioRxiv* 2020.07.20.213074 (2020) doi:10.1101/2020.07.20.213074.
25. Jaynes, E. T. Macroscopic Prediction. in 254–269 (Springer, Berlin, Heidelberg, 1985). doi:10.1007/978-3-642-70795-7_18.
26. Pressé, S., Ghosh, K., Lee, J. & Dill, K. A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **85**, 1115–1141 (2013).
27. Takens, F. Detecting strange attractors in turbulence. in 366–381 (Springer, Berlin, Heidelberg, 1981). doi:10.1007/bfb0091924.
28. Muldoon, M. R., Broomhead, D. S., Huke, J. P. & Hegger, R. Delay embedding in the presence of dynamical noise. *Dyn. Stab. Syst.* **13**, 175–186 (1998).
29. Ragwitz, M. & Kantz, H. Markov models from data by simple nonlinear time series predictors in delay embedding spaces. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **65**, 12 (2002).
30. Stark, J., Broomhead, D. S., Davies, M. E. & Huke, J. Delay embeddings for forced systems. II. stochastic forcing. *J. Nonlinear Sci.* **13**, 519–577 (2003).
31. Alexander, R., Zhao, Z., Székely, E. & Giannakis, D. Kernel analog forecasting of tropical intraseasonal oscillations. *J. Atmos. Sci.* **74**, 1321–1342 (2017).
32. Yiwen, Y., Guizhong, L. & Zongping, Z. Stock market trend prediction based on neural networks, multiresolution analysis and dynamical reconstruction. in *IEEE/IAFE Conference on Computational Intelligence for Financial Engineering, Proceedings (CIFEr)*

155–157 (IEEE, 2000). doi:10.1109/cifer.2000.844615.

33. Ferguson, A. L., Panagiotopoulos, A. Z., Kevrekidis, I. G. & Debenedetti, P. G. Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach. (2011) doi:10.1016/j.cplett.2011.04.066.
34. Wang, J. & Ferguson, A. L. Recovery of Protein Folding Funnels from Single-Molecule Time Series by Delay Embeddings and Manifold Learning. *J. Phys. Chem. B* **122**, 11931–11952 (2018).
35. Topel, M. & Ferguson, A. L. Reconstruction of Protein Structures from Single-Molecule Time Series. (2020).
36. Keenan, A. B. *et al.* The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. *Cell Systems* vol. 6 13–24 (2018).
37. Niepel, M. *et al.* A Multi-center Study on the Reproducibility of Drug-Response Assays in Mammalian Cell Lines. *Cell Syst.* **9**, 35-48.e5 (2019).
38. Gross, S. M. *et al.* A LINCS microenvironment perturbation resource for integrative assessment of ligand- mediated molecular and phenotypic responses. *bioRxiv* 2021.08.06.455429 (2021) doi:10.1101/2021.08.06.455429.
39. Gardiner, C. Stochastic Methods: A Handbook for the Natural and Social Sciences. (2009).
40. Risken, H. & Frank, T. *The Fokker-Planck Equation: Methods of Solutions and Applications (Springer Series in Synergetics)*. *The Fokker-Planck Equation: Methods of Solutions and Applications (Springer Series in Synergetics)* (1996).
41. Kennel, M. B., Brown, R. & Abarbanel, H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403–3411 (1992).
42. Abarbanel, H. D. I. & Kennel, M. B. Local false nearest neighbors and dynamical dimensions from observed chaotic data. *Phys. Rev. E* **47**, 3057–3068 (1993).
43. Kellogg, E. H., Lange, O. F. & Baker, D. Evaluation and optimization of discrete state models of protein folding. *J. Phys. Chem. B* (2012) doi:10.1021/jp3044303.
44. Noé, F. & Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **18**, 154–162 (2008).
45. Schwantes, C. R. & Pande, V. S. Improvements in Markov State Model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **9**, 2000–2009 (2013).

46. Reuter, B., Weber, M., Fackeldey, K., Röblitz, S. & Garcia, M. E. Generalized Markov State Modeling Method for Nonequilibrium Biomolecular Dynamics: Exemplified on Amyloid β Conformational Dynamics Driven by an Oscillating Electric Field. *J. Chem. Theory Comput.* **14**, 3579–3594 (2018).
47. Ankam, S., Teo, B. K., Kukumberg, M. & Yim, E. K. High throughput screening to investigate the interaction of stem cells with their extracellular microenvironment. *http://dx.doi.org/10.4161/org.25425* **9**, 128–142 (2013).
48. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
49. Chaffer, C. L., San Juan, B. P., Lim, E. & Weinberg, R. A. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev.* **35**, 645–654 (2016).
50. Singh, M., Yelle, N., Venugopal, C. & Singh, S. K. EMT: Mechanisms and therapeutic implications. *Pharmacol. Ther.* **182**, 80–94 (2018).
51. Lin, C. H., Jokela, T., Gray, J. & LaBarge, M. A. Combinatorial Microenvironments Impose a Continuum of Cellular Responses to a Single Pathway-Targeted Anti-cancer Compound. *Cell Rep.* **21**, 533–545 (2017).
52. Watson, S. S. *et al.* Microenvironment-Mediated Mechanisms of Resistance to HER2 Inhibitors Differ between HER2+ Breast Cancer Subtypes. *Cell Syst.* **6**, 329-342.e6 (2018).
53. Risom, T. *et al.* Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. *Nat. Commun.* **9**, 3815 (2018).
54. Loret, N., Denys, H., Tummers, P. & Berx, G. The Role of Epithelial-to-Mesenchymal Plasticity in Ovarian Cancer Progression and Therapy Resistance. *Cancers (Basel).* **11**, 838 (2019).
55. Karacosta, L. G. *et al.* Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. *Nat. Commun.* **10**, 1–15 (2019).
56. Waddington, C. H. *The strategy of the genes: A discussion of some aspects of theoretical biology*. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology* (Allen, 1957). doi:10.4324/9781315765471.
57. Matsushita, Y. & Kaneko, K. Homeorhesis in Waddington's Landscape by Epigenetic Feedback Regulation. (2019).
58. Banerji, C. R. S. *et al.* Cellular network entropy as the energy potential in Waddington's differentiation landscape. *Sci. Rep.* **3**, 3039 (2013).
59. Bather, J. A., Fleming, W. H. & Rishel, R. W. Deterministic and Stochastic Optimal

Control. *J. R. Stat. Soc. Ser. A* (1976) doi:10.2307/2344363.

60. Gray, J. W. & Pallavicini, M. G. Ara-C scheduling: Theoretical and experimental considerations. *Med. Pediatr. Oncol.* **10**, 93–108 (1982).
61. Swierniak, A., Polanski, A. & Kimmel, M. Optimal control problems arising in cell-cycle-specific cancer chemotherapy. *Cell Prolif.* **29**, 117–139 (1996).
62. Cunningham, J. J., Brown, J. S., Gatenby, R. A. & Staňková, K. Optimal control to develop therapeutic strategies for metastatic castrate resistant prostate cancer. *J. Theor. Biol.* **459**, 67–78 (2018).
63. Chapman, M. P., Mazumdar, E. V., Langer, E., Sears, R. & Tomlin, C. J. On the Analysis of Cyclic Drug Schedules for Cancer Treatment using Switched Dynamical Systems. in *2018 IEEE Conference on Decision and Control (CDC)* 3503–3509 (IEEE, 2018). doi:10.1109/CDC.2018.8619490.
64. Cunningham, J. *et al.* Optimal control to reach eco-evolutionary stability in metastatic castrate-resistant prostate cancer. *PLoS One* **15**, e0243386 (2020).
65. Vicar, T. *et al.* Cell segmentation methods for label-free contrast microscopy: Review and comprehensive comparison. *BMC Bioinformatics* **20**, 360 (2019).
66. Tscherepanow, M., Zöllner, F. & Kummert, F. *Automatic Segmentation of Unstained Living Cells in Bright-Field Microscope Images*.
67. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917–920 (2018).
68. Rajakylä, E. K. *et al.* Assembly of Peripheral Actomyosin Bundles in Epithelial Cells Is Dependent on the CaMKK2/AMPK Pathway. *Cell Rep.* **30**, 4266-4280.e4 (2020).
69. Yang, K. D. & Uhler, C. Multi-Domain Translation by Learning Uncoupled Autoencoders. (2019).
70. Cao, K., Hong, Y. & Wan, L. Manifold alignment for heterogeneous single-cell multi-omics data integration using Pamona. *bioRxiv* 2020.11.03.366146 (2020) doi:10.1101/2020.11.03.366146.
71. Schau, G., Burlingame, E. & Chang, Y. H. DISSECT: DISentangle SharablE ConTent for Multimodal Integration and Crosswise-mapping. in *Proceedings of the IEEE Conference on Decision and Control* vols 2020-December 5092–5097 (Institute of Electrical and Electronics Engineers Inc., 2020).
72. Sommer, C., Straehle, C., Kothe, U. & Hamprecht, F. A. Ilastik: Interactive learning and segmentation toolkit. in *Proceedings - International Symposium on Biomedical Imaging*

230–233 (2011). doi:10.1109/ISBI.2011.5872394.

73. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

74. Coelho, L. P. Mahotas: Open source software for scriptable computer vision. *J. Open Res. Softw.* **1**, e3 (2013).

75. Alizadeh, E., Xu, W., Castle, J., Foss, J. & Prasad, A. TISMorph: A tool to quantify texture, irregularity and spreading of single cells. *PLoS One* **14**, e0217346 (2019).

76. Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).

77. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–47 (2019).

78. Inman, H. F. & Bradley, E. L. The Overlapping Coefficient as a Measure of Agreement Between Probability Distributions and Point Estimation of the Overlap of two Normal Densities. *Commun. Stat. - Theory Methods* **18**, 3851–3874 (1989).

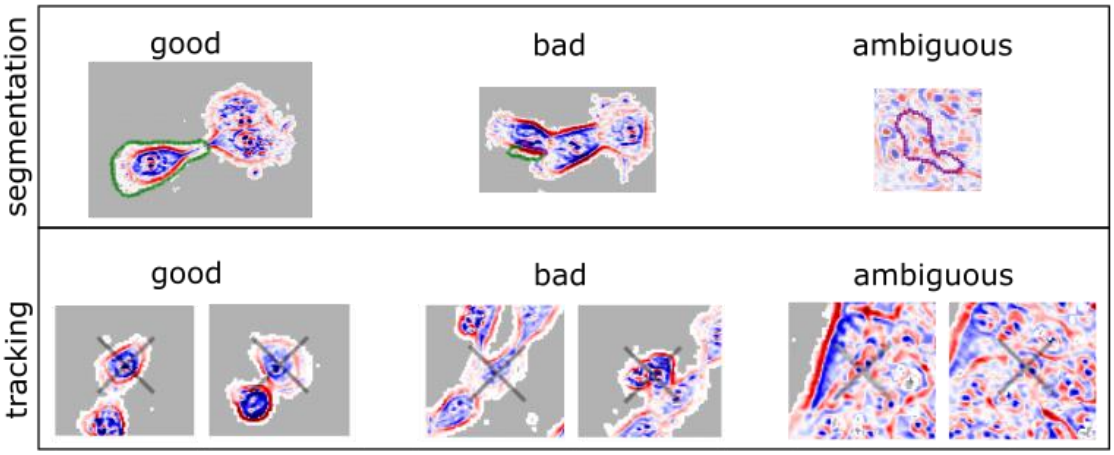
79. Mostofian, B. & Zuckerman, D. M. Statistical Uncertainty Analysis for Small-Sample, High Log-Variance Data: Cautions for Bootstrapping and Bayesian Bootstrapping. *J. Chem. Theory Comput.* **15**, 3499–3509 (2019).

Extended Data Tables and Figures

Extended Data Table 1. Segmentation and tracking manual validation. 100 cells per treatment were randomly selected, and evaluated by eye to qualitatively assess segmentation and tracking accuracy. Fraction segmented was estimated by the image area covered by segmented masks divided by the area selected as being occupied by cells from the ilastik random forest pixel classifier. *(+EGF)

ligand	total	EGF	HGF	OSM	IFNG*	BMP2*	TGFB*
% segmented	52%	53%	51%	46%	54%	52%	42%
% good seg	43%	52%	38%	33%	48%	53%	56%
% bad seg	26%	26%	24%	7%	25%	33%	34%
%ambiguous seg	31%	22%	38%	60%	27%	14%	10%
% tracked	60%	51%	56%	55%	60%	58%	79%
% good tracks	87%	95%	91%	42%	94%	96%	95%
% bad tracks	2%	5%	0%	3%	6%	0%	0%
%ambiguous tracks	11%	0%	9%	55%	0%	4%	5%

Extended Figure 1. Segmentation and tracking manual validation. Examples of good, bad, and ambiguous qualitative validation categories for segmentation and tracking.



Extended Data Table 2. Number of extracted trajectory snippets with increasing snippet length. *(+EGF)

Trajectory snippet length	PBS	EGF	HGF	OSM	BMP2*	IFNG*	TGFB*
1	49016	73093	69747	99933	63669	72389	49008
2	32251	35337	40640	58119	31558	39545	36736
4	21894	14066	22837	29783	14135	19501	16206
8	15342	4892	12572	13208	5772	9033	2936
16	9256	1416	5857	3861	1892	3603	884
32	4243	226	2161	924	409	1055	142
64	1002	16	347	139	0	141	0

Extended Figure 2: Trajectory embedding constructs a common space to evaluate unique and shared cell morphodynamics. Top left: outline of the combined density distribution in the trajectory embedding (snippet length = 8) space (gray), with locations of the density peaks in individual treatments marked with letters consistent with Figure 5 and exhibited cell trajectory snippets at locations marked with stars. Remaining boxes: cell trajectory snippet extracted at the marked location, but from the treatment labeled for each box. Time for each 8-step trajectory snippet shown runs from top to bottom.