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 is a powerful approach to gain insight into the dynamic responses of cells, but the quantitative analysis of morphological changes over time is a challenge. Here, we exploit the concept of "trajectory embedding" to analyze cellular behavior using morphological feature trajectory histories, i.e., multiple time points simultaneously, rather than the more common practice of examining morphological feature time courses in single timepoint (snapshot) morphological features. We apply this approach to analyze live-cell images of MCF10A mammary epithelial cells after treatment with a panel of microenvironmental perturbagens that strongly modulate cell motility, morphology, and cell cycle behavior. Our morphodynamical trajectory embedding analysis constructs a shared cell state landscape revealing ligand-specific regulation of cell state transitions and enabling quantitative and descriptive models of single-cell trajectories. Additionally, we show that incorporation of trajectories into single-cell morphological analysis enables (i) systematic characterization of cell state trajectories, (ii) better separation of phenotypes, and (iii) 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

In normal and diseased tissues, cells are continually exposed to a wide variety of extracellular stimuli, including growth factors and cytokines, that modulate morphological and phenotypic responses. However, quantitative assessment of complex morphological states remains a challenging problem. Single timepoint measurements provide some information about cell state but do not capture how 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 have 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 classification of cell morphology observed in individual time points 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 imaging with molecular reporters. Quantitative analysis of cell morphology in live-cell imaging typically involves the characterization of cellular morphology from static, single-timepoint images, extended over single-cell time courses[8–12]. Analysis methods directly based upon trajectory features, i.e., features aggregating multiple time points, have been used for classification of mitosis and apoptosis[13], and also for monitoring signaling responses via fluorescent reporters[14]. We show here that by mapping the multiple-time-point *trajectory space* of cells, rather than examining single-cell time courses built from snapshots, we can increase the information extracted from live-cell imaging experiments and improve the quantitative description of cellular responses.

Live-cell imaging provides temporal information not available from other single-cell and omics measures. Single-cell RNA sequencing (scRNA-seq) can assay thousands of molecular read-outs across thousands or hundreds of thousands of cells. 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], however, approaches such as “RNA velocity”[22,23] have been used to infer cell state dynamics. In contrast, live-cell imaging is unique in that cell responses can

be assayed for extended periods of time, even when unlabeled imaging is restricted to morphological features with limited information about molecular state or information regarding the basis for dynamic changes.

Cell responses to a perturbation can be quantified by grouping cells with shared phenotypes together and tracking the subpopulations of cells assigned to these cell states. Cell states can be discrete, such as assigning cell-cycle states G0-G1-G2-M, or continuous, such as a continuum of states spanning an epithelial-to-mesenchymal transition. Live-cell imaging analysis of cell states has been applied using many different experimental approaches and analysis strategies with cell states defined in a variety of ways to capture different biological aspects. Live-cell imaging has been used to develop gene-level functional annotation, including in RNAi gene knockout screens and drug screening[7], and 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 responses that included 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 to develop a cell trajectory-based description of an epithelial-mesenchymal transition (EMT)[12] in the space of single-timepoint snapshot features. Heryanto et al. utilized 3D shape descriptors to explore the relationship between 3D shape and cell motility[24]. Notably, trajectory information, including combined motility and morphological features computed as averages over single-cell trajectories, have been used to define and identify cell state[25]. The morphodynamical trajectory embedding approach described in this manuscript differs in that we directly analyze morphological feature trajectories, rather than average morphological or average cellular features over time. Preserving the feature trajectories ensures we maximally leverage the additional information contained in the time-ordered single-cell trajectory information.

Trajectories are the natural space from which to classify a system out of equilibrium, such as a living cell[26,27]. However, this realization is not useful without the capability to measure the trajectory space of the system of interest. Floris Takens' seminal trajectory embedding theorem[28] proves that in a deterministic dynamical system, there is a 1:1 correspondence between the space of the full dynamical system 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 to form morphodynamical feature trajectories, which we refer to as "trajectories." For N_f features and n_τ trajectory timepoints the trajectory 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[29–31]. Trajectory embedding methodology has been applied in fields as diverse as weather prediction[32], economics[33], and molecular dynamics[34–36], but to our knowledge has not been applied in the context of classifying cell state in live-cell imaging assays.

Here, we develop and apply the morphodynamical trajectory embedding method in a dataset of MCF10A mammary epithelial cells perturbed with a set of six ligands spanning major extracellular signaling pathways and inducing distinct cellular responses, including changes to cell proliferation, differentiation state, and motility[37]. The live-cell imaging data are part of a broader data collection effort through the Library of Integrated Network-Based Cellular Signatures (LINCS) consortium[38,39] MCF10A project[37] where the molecular and phenotypic responses to these ligand perturbations were explored. Molecular (RNAseq, protein expression levels via reverse phase protein array (RPPA), and chromatin state ATACseq) and cellular responses (cyclic immunofluorescence, cycIF) indicate changes in canonical cell signaling pathways and initiation of unique cellular responses in each ligand condition[37]. Our live-cell imaging cell-trajectory-based analysis was developed to characterize the morphodynamical changes associated with molecular responses.

Results

We applied our trajectory embedding analysis to systematically characterize cell state based on 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, as part of the LINCS MCF10A multi-omic profiling project[37]. Single-cells were segmented, featurized, and tracked through time as described in Methods. In our "trajectory embedding" approach, time-sequences of features (for the trajectory snippet length under consideration) were concatenated and used for UMAP[40] dimensionality reduction (to dimensionality d) as the basis for further analysis. The single-cell trajectories we analyze are the set of extracted single-cells and their tracks, or linkages between frames. An overview of the live-

cell imaging trajectory embedding workflow is shown in Figure 1. 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 analyzing all of the trajectories from the full set of treatments through the dimensionality reduction pipeline together. The single-cell trajectories we use for a trajectory embedding of length n_τ consist of all possible trajectory snippets of length n_τ in the full trajectory set; for example, a single cell which is tracked over 12 frames will have 5 possible trajectory snippets of length 8 in a sliding window manner (frames 1-8, 2-9, 3-10, 4-11, 5-12). Unless otherwise labeled, all available trajectory snippets over the 48-hour experiment are used. Snippets mapped to the same location in the reduced-dimensionality cell state space share qualitatively similar morphology across trajectory timepoints and across all 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.

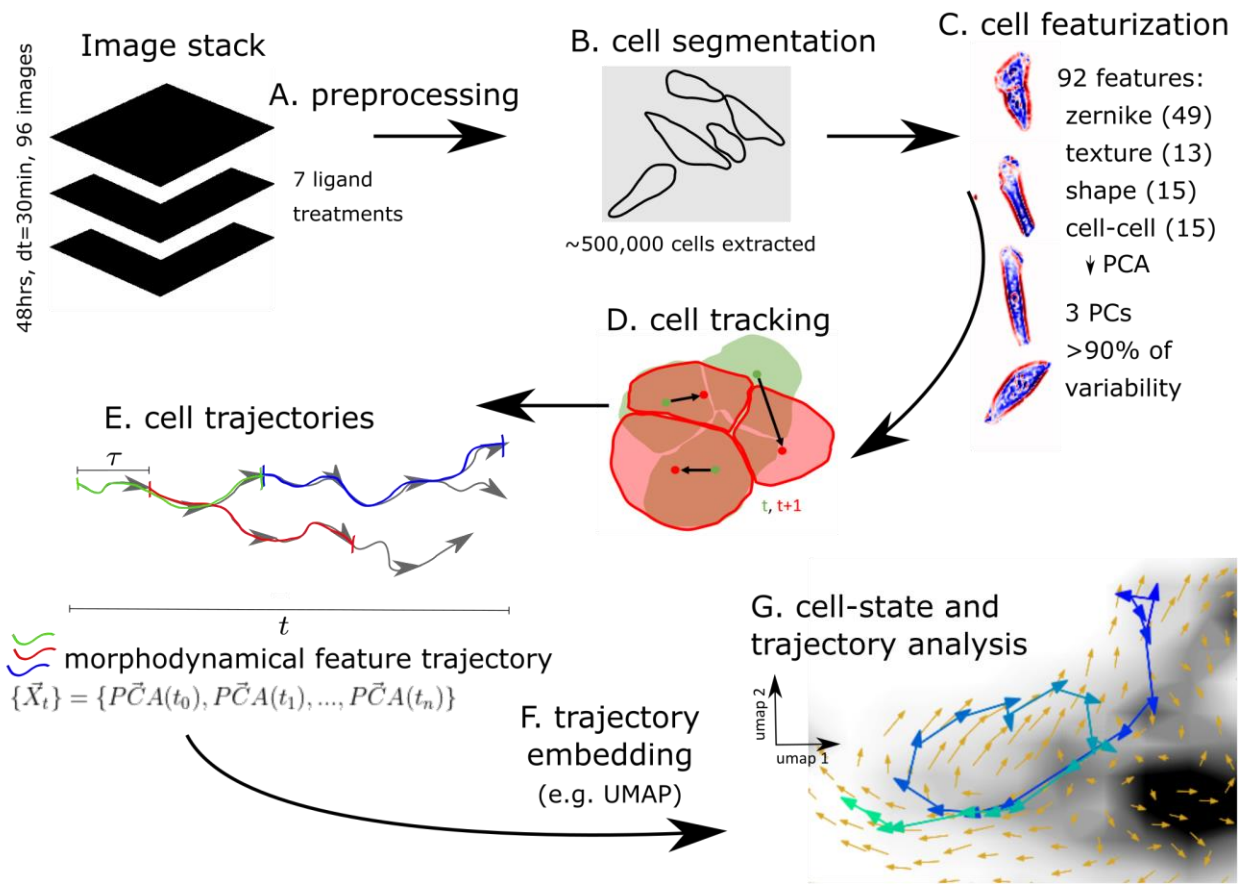


Figure 1: Live-cell imaging analysis and trajectory embedding pipeline.
The data analysis pipeline starts from 48-hour image stacks and proceeds to the morphodynamical trajectory analysis. Top left to bottom right: Image processing steps include A. preprocessing, B. cell segmentation, C. featurization (z-normalized phase-contrast pixel values colored red positive to blue negative), D. tracking (cell boundaries at t, t+30 min with cell centers connected by black arrows), E. extracting morphodynamical trajectories as sliding window cell feature trajectory snippets from cell linkages (3 possible trajectory snippets of length 3 shown in green, red, and blue), F. trajectory embedding (UMAP), and G. cell state and trajectory analysis.

Separation of unique and shared cell state under ligand perturbation

Ligand perturbation can induce time-dependent morphologic and phenotypic changes, including altered cycling rate, motility, and cytoskeletal features. 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.[37] for further phenotypic profiling. These striking changes motivate our interest in quantitative analysis of morphodynamical trajectories.

We characterized changes to cell morphodynamics under the different ligand treatments by quantifying the similarity between distributions of morphodynamical trajectories between ligand conditions. We found similar ligand-specific distributions in the embedding space of morphological snapshots, with increased ligand-specific uniqueness observed in the embedding space of morphological trajectories. At the snapshot level, which excludes trajectory information, Figure 2B (left) shows that over the course of the experiment cells occupy broad distributions in the embedding space, with distinct shifts in occupancy separating OSM and TGFB+EGF from the rest of the ligand treatment conditions, which is consistent with the very different morphologies observed under these two treatments (Figure 2A). At a 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 uniqueness 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 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.

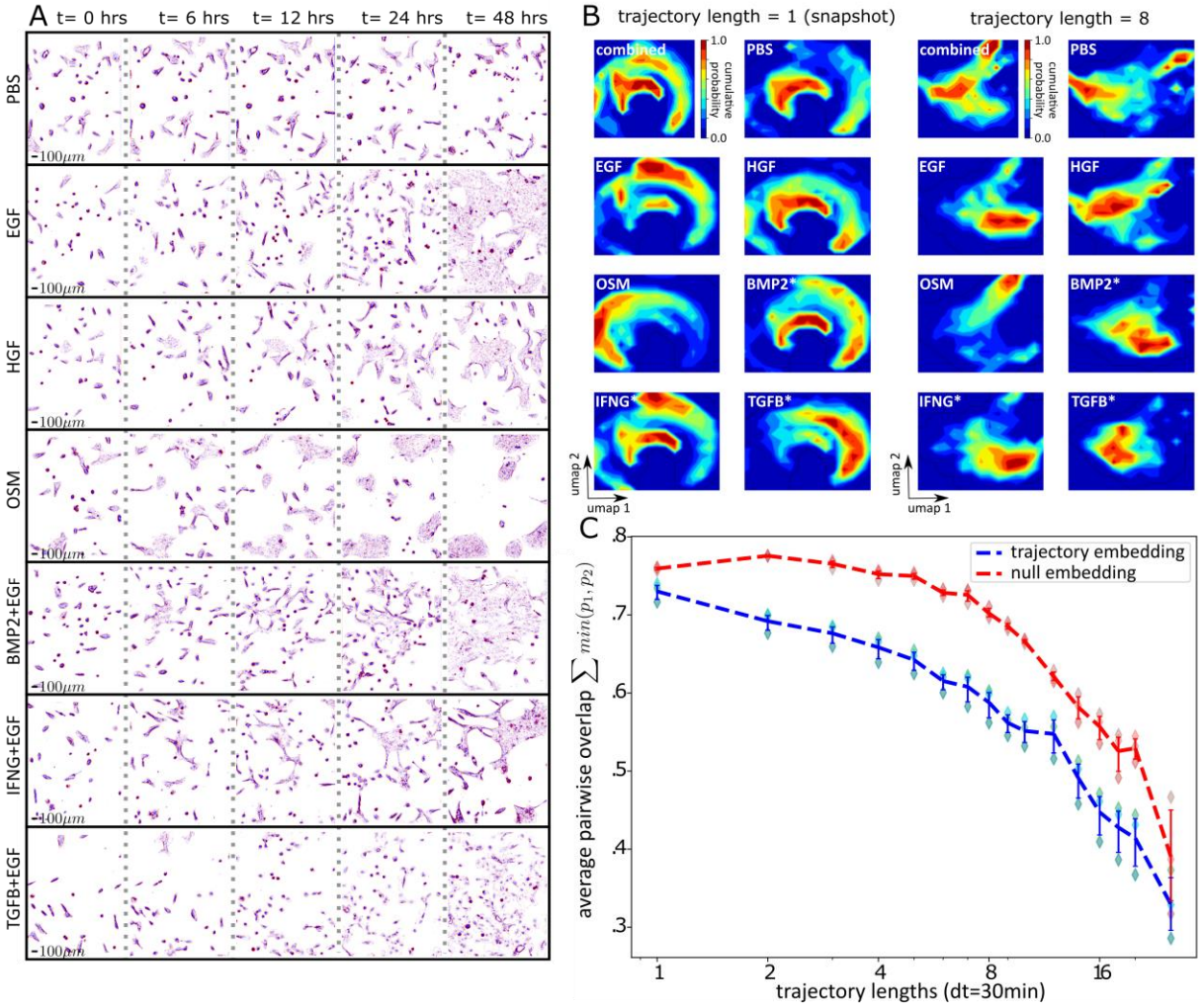


Figure 2: Trajectory embedding increases the distinguishability of cell states induced by ligand perturbation.

A) Representative background subtracted phase-contrast images (size 1.6mm x 1.9mm with z-normalized phase-contrast pixel values colored red (positive phase contrast) to blue (negative phase contrast) in the set of ligand conditions (top to bottom) at 0, 6, 12, 24, and 48 hrs (left to right). B. Distributions 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 over all treatment pairs (shared area under probability distributions) as a function of the morphodynamical feature trajectory length used in the embedding ($\log_2$ x-axis scale), comparing trajectory embedding (blue dashed lines) and null model with randomly scrambled time labels within treatments (red dashed lines), averaging over results obtained by dividing data into three sets by field of view (diamonds), with error bars from a bootstrapped 95% confidence interval over the three data splits. A trajectory length of 1 corresponds to a snapshot description.

Improved cell state description from morphodynamical trajectories

Cell states can be defined by identifying metastable regions of the morphodynamical trajectory embedding space where trajectories remain localized for extended time periods. To compare cell states and transitions, we extracted the single-cell dynamics in the embedding space, i.e., in the space of trajectory snippets defined in a sliding window. If there are *T* timepoints in a full single-cell trajectory, then there are $T\text{-}n_\tau + 1$ snippets of length n_τ, each of which is a point in the embedding space. Together, these points trace out a trajectory in the embedding space with $T\text{-}n_\tau$ transitions between snippets (e.g., from the snippet consisting of frames 1-8 to the snippet consisting of frames 2-9). We used the dynamical information about snippet-to-snippet transitions to calculate dynamics in the morphodynamical cell state space. 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[41,42] 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 displays 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. Individual single-cell trajectories in the embedded space can stay partly localized to these metastable cell states for extended time periods. 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 less complete snapshot level unfold[43,44] and become systematic in the trajectory embedding space.

The usefulness of a single-cell trajectory depends upon how well it characterizes cell state transitions and transition dynamics. We measured how systematic and predictable the trajectories are by quantifying the randomness of the trajectories in the embedding space. We defined the predictability of a trajectory as 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 in the embedding space after one timestep (30 min) and the standard deviation in the displacement over the full population. Angled brackets $\langle \cdots \rangle$ indicate averages over all trajectories and timepoints t. In a completely random trajectory, this ratio is 1 because the variance in a single timestep and the full population is identical. In a

deterministic trajectory, all trajectories emanating from the same point are identical and have no variance after a single timestep--the only contribution to the ratio is the relative average displacement in the time interval. In a continuous, stochastic description of the trajectories in the embedded space[41,42], this locality ratio is related 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 less random and more predictable with increasing trajectory embedding length.

To determine the capability of the morphodynamical embeddings to characterize single-cell morphodynamical trajectories, we first used a subset of the data to train a model of trajectory likelihood, then calculated the average log-likelihood of a held-out test set of trajectories. The average log-likelihood is a direct measure of the predictability of the cell trajectories[45]. Here we used a Markovian transition matrix likelihood model, trained by counting transitions between Voronoi states defined by k-means cluster centers on the landscape[46,47]. We utilized 100 k-means centers to discretize the morphodynamical embedding space, which provided sufficient Voronoi centers to capture the observed patterns of cell state flow between metastable cell states while still retaining adequate sampling of state-state transitions. The average log-likelihood increases as a function of morphodynamical trajectory 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, reconstructing complete trajectories is challenging (extended data table 2). For snippet lengths longer than 8 frames, we did not observe an increase in ligand-specific trajectories relative to the null model (Figure 2C), which likely is related to the decrease in the number of extracted trajectories longer than 8 frames (extended data table 2). Thus we chose a trajectory embedding length of 8 (3.5 hrs) for further analysis. The decreased overlap between ligand-specific distributions (Figure 2C), decreased locality ratio (Figure 4A), and increased trajectory likelihood (Figure 4B) indicate that even partial trajectory information over a timescale of a few hours substantially improved the representation of cell state.

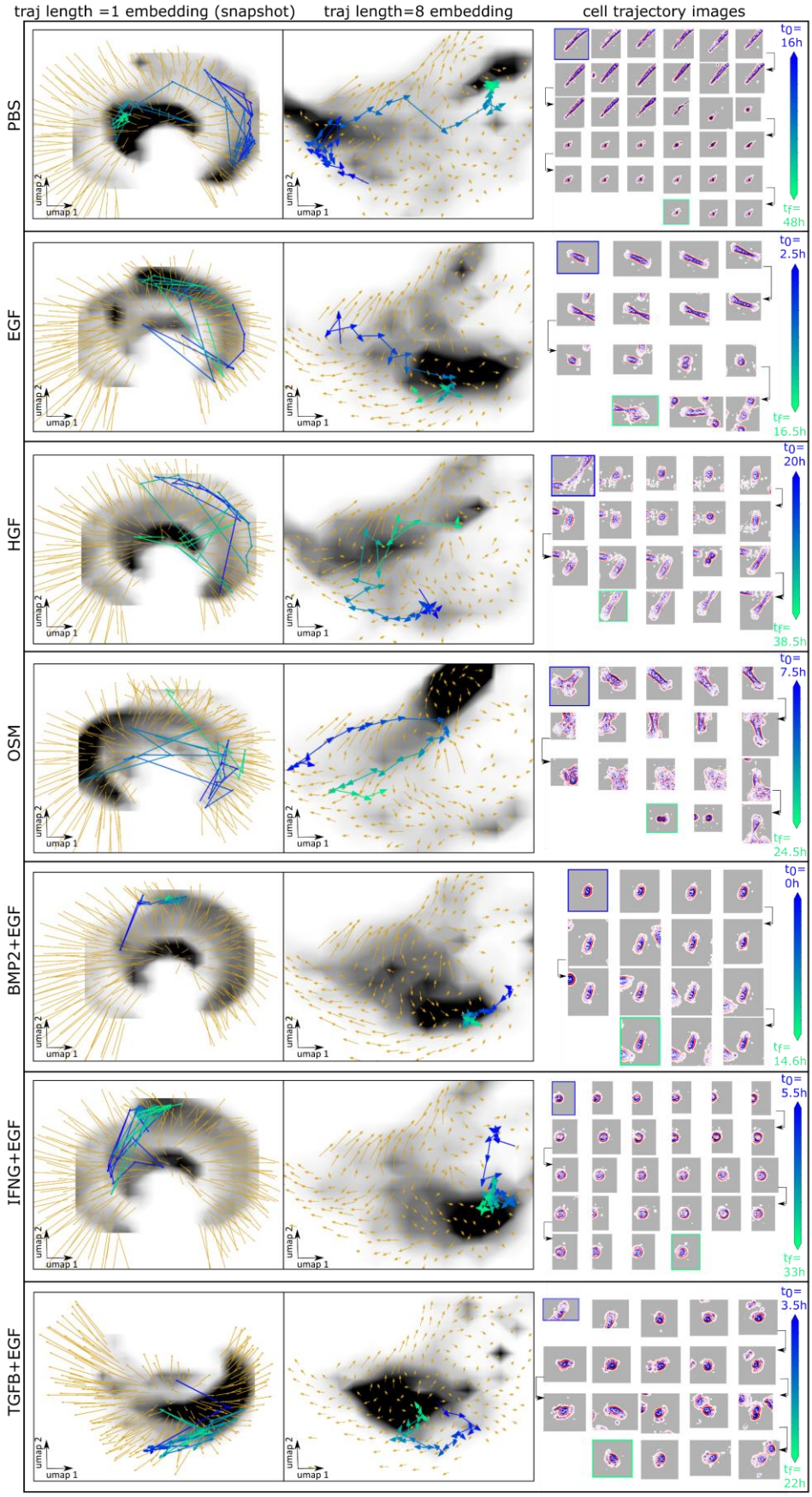


Figure 3: Trajectory embedding enables the determination of metastable cell states and pathways across ligand treatments. Left and middle column: average displacement proportional to the effective force (orange arrows), and cell density (grayscale) for snapshot embedding (left: snapshot, trajectory snippet length = 1, right: trajectory snippet length = 8). Representative cell morphodynamical trajectories capturing cell state transitions (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, with the tracked cell centered in the image frame and neighboring cells moving in and out of view, except in the IFNG+EGF images where the tracked cell is temporarily clipped at the edge of the microscope field of view.

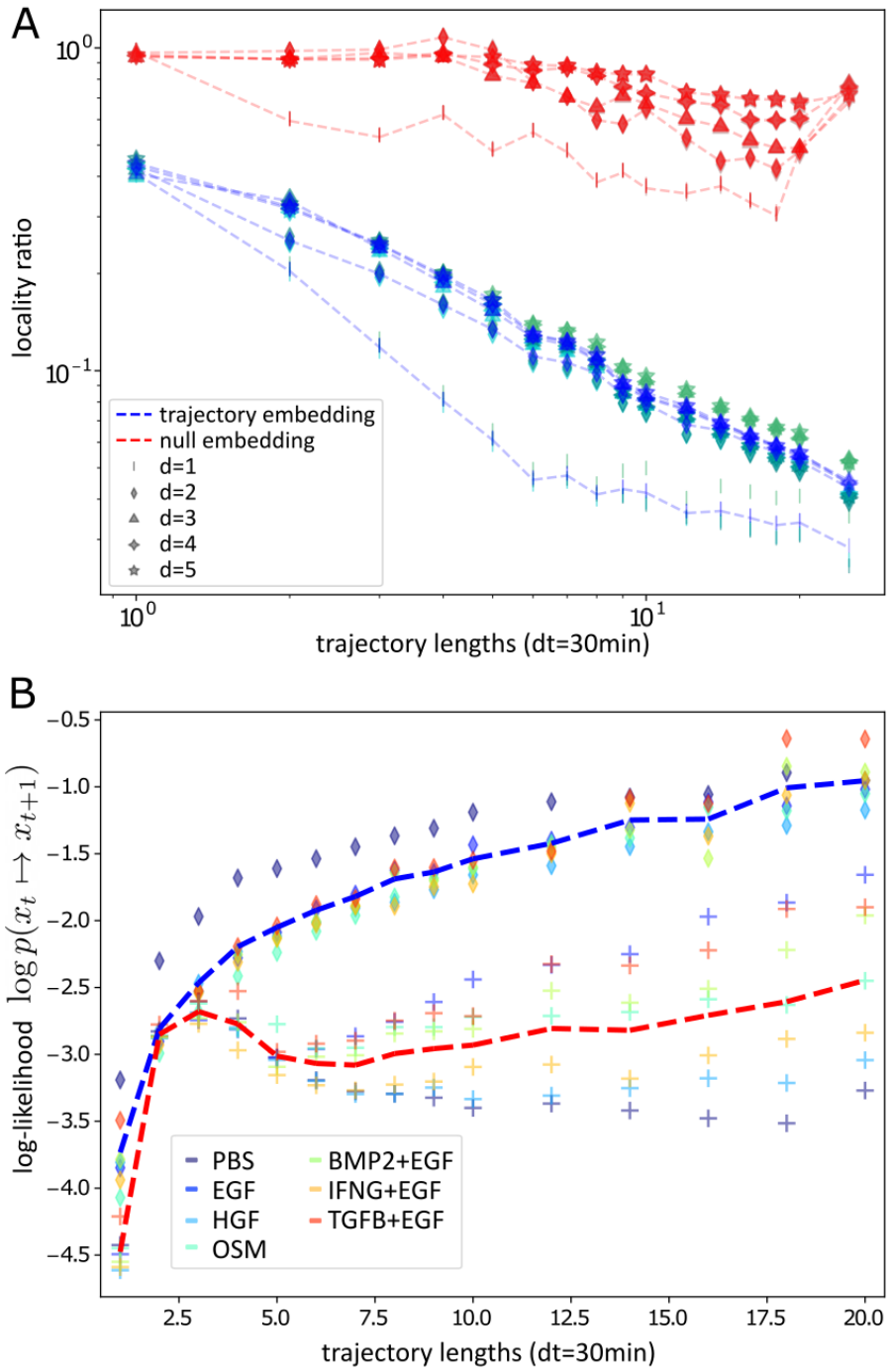


Figure 4: Trajectory embedding increases the predictability of cell trajectories.
A) Ratio between the single-step (dt=30min) and full RMS displacement in the trajectory embedding space as a function of the trajectory length, null model with randomly scrambled features within treatments (reds) and trajectory embedding (blues), and UMAP embeddings with d=1 (lines), d=2 (diamonds), d=3 (triangles), d=4 (squares), and d=5 (stars), with d the number of UMAP components. 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 (colors) for the null model (crosses) and for the trajectory embedding (diamonds).

Morphodynamical transitions precede cell cluster formation

Ligand treatments displayed characteristic transitions between cell 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 (extended figure 3).

Cell states in the embedding space were identified using a combination of quantitative methods refined by visual inspection Our primary conclusions, importantly, do not depend on fine details of the states employed for analysis. We identified fine-grained metastable states spanning the cell state landscape from density peaks arising in individual treatments (Figure 5A, 5B A.-S.), which we collected into 6 macrostates, or groupings, and assigned qualitatively descriptive labels based upon observed cell morphology. These labels are intended to be descriptive for ease of interpretation and suggest future areas of study to more deeply explore the biological underpinnings of our findings. “Separated epithelial-like” cell states are rounder, active cells enriched in EGF condition, while “separated mesenchymal-like” cell states display more extended cytoskeletal features such as lamellopodia and are enriched in the TGFB condition. The “intermediate clusters” cell state divides the cells which are mostly separated from the cell states where cells are tightly clustered and attached together. These multicellular clustered or attached cell states we divided into “bound clusters” which display a thick border around attached cells that span multiple cells, “unbound clusters” which lack this border and whose outer cells have extended cytoskeletal features, and “budding” cells which consist of semi-round, “crinkly” cells attached to the outer border of a cell cluster.

The direction of the cell state flow in the trajectory embedding space indicates that a group of cells coming together and forming an attached multicellular cluster proceeds via macrostate 2, “separated mesenchymal-like” cells with extended cytoskeletal features. This flow was consistently observed in the cell state “force-fields” (Figure 4), cell state transition networks (Figure 5A), and the time-dependent cell state distributions (Figure 5C), and in all of the ligand conditions despite morphological differences in multicellular clusters between treatments.

Cell cluster formation was most strongly associated with OSM treatment,[37] consistent with figure 5A showing that in the OSM treatment the transition probability from the “intermediate cluster” state 3 into the multicellular clustered state was higher than the transition probability out.

Separated cells (states 1 and 2) are separated from cell clusters (states 4 and 5) and cell cluster formation proceeds through states 2→3→4 via cells displaying extended cytoskeletal features and increased cell border contrast (Figure 5B).

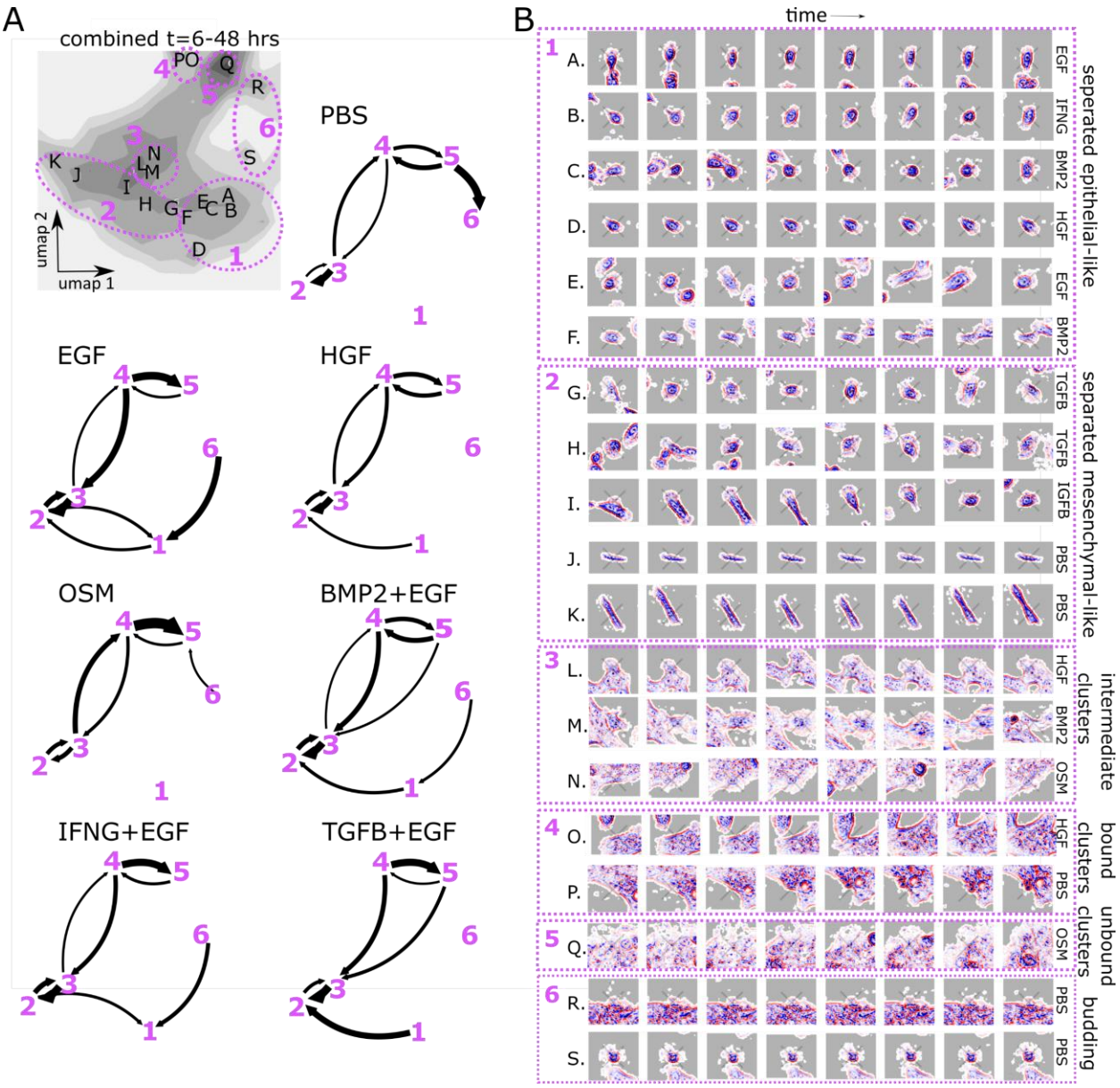


Figure 5: Trajectory embedding resolves pathway of cell cluster formation via mesenchymal-like intermediate.
A) Upper left: Cumulative distribution of all cells under all treatments (grayscale) with labeled fine-grained density peaks (A-S), and overall qualitative cell state network of macrostates (mauve dashed circles) with numeric labels (1-6). Cell state transition networks with arrow weight proportional to conditional transition probability, not overall transition flows, with transition probabilities <3% not drawn. B) Representative trajectory snippets (embedding length = 8 = 3.5 hrs) extracted at density peak locations (A.-S.) with the treatment condition of the representative trajectory snippet, grouped by macrostate with morphologically descriptive macrostate labels (right of images).

Discussion and Conclusions

In vivo, cells continually modulate their phenotypic state in response to the local microenvironment[38]. During development, cells must precisely control cell states[48], responses to extracellular signals, and cell motility, while the loss of cell state control is associated with various diseases, including cancer[49]. Dynamic cell behaviors are observed via live-cell imaging but quantifying the relationships between diverse cellular phenotypes that may be dynamically interchanging has been challenging. 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 coordinate space based on time-aggregated "hyper features" better suited to study cell states and their dynamical relationships as compared to standard featurization based on single timepoints.

Trajectory embedding analysis enables quantification of cell state dysregulation, including cell state plasticity which has been identified as a mechanism of disease progression and therapeutic resistance in advanced cancers[50–56]. Cell state has been defined as an attractor in a high-dimensional landscape, a paradigm famously depicted in Waddington's epigenetic landscape[57–59]. We observe that cells dynamically transition between morphodynamical cell states and that the transition frequency is strongly modulated by ligand treatment. These dynamic cell state relationships can provide a framework for understanding cell-cell heterogeneity and heterogenous cell responses to perturbation. Live-cell trajectory embedding brings the cell state landscape

paradigm from theoretical biology to direct application, where cell states and the transitions between them can be resolved, validated, and potentially leveraged for actionable control strategies[17,60–65].

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 chose a broad cell feature set but many other cell featurizations have been developed, including other novel cell shape descriptors[66,67] and machine learning-based approaches[68]. Trajectory embedding, in principle, has the capability to map dynamical information from any reasonable featurization towards a more complete description of cell state.

The morphodynamical trajectory embedding method can be applied to any live-cell imaging modality where cells can be characterized and tracked through time, and in particular, we expect that this method will be especially powerful for analysis of live-cell approaches that incorporate genetically-encoded, fluorescently-labeled reporters, both *in vitro*[69–71] and *in vivo*[72–74]. Practical limitations will always come into play, nevertheless. For example, we applied our method to unlabeled phase-contrast imaging of cell cultures that approach confluence[75,76], without paired ground truth labeled images[77], requiring an analysis approach with some robustness to cell segmentation and tracking errors. Trajectory data quantity and quality will generally pose a constraint on trajectory length used in the morphodynamical trajectory embedding analysis. Even when many very long trajectories are available, analysis based upon shorter trajectory lengths might more directly capture processes and relationships of specific interest, such as the connection between specific morphological features and cell motility. General procedures to identify the optimal trajectory length for analysis need to be developed, building on the preliminary approaches employed here.

Biological interpretation and validation of the morphodynamical cell states extracted here will be important, and our findings help to motivate specific hypotheses that could be explored in future studies. Identification of the molecular programs associated with particular cell states and cell state transitions would provide insight into how these processes are mediated in normal tissues and how they may go awry in diseases. 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,[78] but live-cell imaging

coupled with deeper molecular profiling data—such as multiplexed imaging[79–82] and single-cell transcriptomics[83,84]-- are needed in order 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[85–87], and may enable integration of 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.

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 Data used in this study were recently described by Gross, et al.[37] In brief, cells were plated at ~50,000 cells/ml (0.5ml per well) on collagen-coated 24-well plates, and cultured with 7 different ligand treatments (PBS, EGF 2ug/ml, HGF 8ug/ml, OSM 2ug/ml, BMP2 + EGF 2ug/ml, IFNG + EGF 2ug/ml, TGFB +EGF 2ug/ml), after an 8 hour attachment period in growth media containing EGF and a 12 hour period in imaging media lacking EGF. Wells were imaged every 30 minutes for 48 hours via bright-field phase contrast with an Incucyte microscope (1020x1280, 1.49 μm/pixel), with the initial frame coinciding with the addition of the ligands and fresh imaging media. 6 image stacks were collected for each ligand treatment. Experimental protocols can be found in detail at the publicly available Synapse database[88].

Image preprocessing Foreground (cells) and background pixel classification was performed using manually trained random forest classifiers using the ilastik v1.3.3 software[89]. 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[90] v0.6.5 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 allowing the Cellpose algorithm to

determine a new cell size at each round until no more new cells were found (pixels of previously segmented cells set at each round to the background value of 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 Three classes of features were used to characterize individual cells: (i) texture features, (ii) shape features, and (iii) features characterizing adjacent cells. As preliminary steps, 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); then the long axis was defined by the non-mass-weighted moment of inertia of the cell mask and aligned along a reference axis. (i) Two types of internal cell features were calculated. Zernike moments (49 features) were used to characterize the overall spatial phase contrast signal and Haralick texture features (13 features) were used to characterize the phase contrast texture; these were calculated in the Mahotas[91] image analysis package. The sum average Haralick texture feature was discarded due to normalization concerns. (ii) Shape features (15 features) were calculated as the absolute value of the frequency coefficients of the Fourier transform of the distance to the boundary as a function of the radial angle around cell center[92], with the sum of shape features normalized to 1. (iii) The 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 (no neighboring cell), and value 1 if in contact with the cell foreground mask. The absolute values of the Fourier transform coefficient of this indicator as a function of radial angle around the 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. Additional information regarding the cell featurization can be found in extended figure 4A.

After computing the raw features as described, 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. More PCA components can be retained with only a small computational cost in the dimensionality reduction step (here UMAP) which typically scales

linearly with the data vector length. Additional information regarding the cell feature PCA reduction can be found in extended figure 4B.

Cell tracking Segmented cells were tracked between live-cell imaging frames to extract the set of single-cell trajectories. Image stacks were first registered translationally without allowing rotational or general affine transformation using the pystackreg implementation of subpixel registration[93]. Cell centers were recorded as the equally weighted center of mass of the single-cell masks. Cells were tracked between frames by first separating each contiguous cluster of cells as defined by connected sets of the foreground/background cell mask. If cell clusters occupied less than 10,000 pixels (typical cell very roughly 30x30 pixels), cells were simply tracked by minimum distance with a cutoff of 45 pixels. Tracking cells by minimum distance refers to linking a cell at frame t to the cell at frame t+1 which has the minimum distance between cell centers. 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. Two or more cells may have the same parent cell as the closest cell in the previous frame. This can be the result of a cell division event, or can be the result of a missed track or missed cell segmentation. We do not separate cell division events into their own category, but simply use the cell tracks, or linkages between frames, to identify the unique cell trajectory history. Where trajectories are split as in a cell division, trajectories are not truncated and begun anew, but rather each daughter is tracked backward in time leading to two trajectories which are treated separately, see cell image sequences from HGF and TGFB+EGF conditions in Figure 3 for examples. That is, daughter cells will have overlapping or shared history from the parent cell.

Morphodynamical trajectory embedding High-dimensional time-sequences of features are used as the input to dimensionality reduction algorithms as the basis for morphodynamical cell state analysis. Extracted cell linkages were constructed from available cell tracks. These cell linkages are not complete, that is incomplete segmentation/tracking means that some cells cannot be traced all the way back to an initially plated progenitor cell. The available cell history is the unique backwards trace of any extracted single-cell through time. 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 time points) with cell history equal to or longer than a given length, the trajectory

snippet set, see extended figure 5 for a graphical description. Note there is a large amount of duplication in this sliding window division of available cell trajectories. Cell trajectories that are longer than the trajectory snippet length used in the trajectory embedding allow for the determination of cell states and pathways. From the linkages, 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 is shown in 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 feature trajectory of length n_τ (with N features, the feature trajectory for each cell is N x n_τ). These morphodynamical feature trajectories (spanning all image stacks and all treatments) are flattened into vectors and then embedded using UMAP[94] into a space of dimension d. Changing UMAP embedding hyperparameters will alter fine details regarding the trajectory embedding landscape, but we find our overall results to be robust with very little change in the measured overlap between ligand populations or trajectory likelihood as shown in extended figure 6. 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[95] 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 We employ several measures motivated by standard concepts of stochastic physics. 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 and the angle brackets $\langle \dots \rangle$ indicate averages over all trajectories and time points. For completely random trajectories, this ratio is 1, and tends to the relative average displacement for a continuous, deterministic dynamic. If cells are obeying stochastic Markovian dynamics characterized by a diffusion equation, then the

average displacement $\langle \Delta\vec{x} \rangle$ is proportional to the effective force $\vec{F}(\vec{x})$ via $\langle \Delta\vec{x} \rangle = \gamma^{-1} \vec{F}(\vec{x})\Delta t$ with γ the friction.

Cell state clustering and prediction The cell state dynamics were characterized by building a fine-grained discretized transition matrix model in the continuous morphodynamical trajectory embedding space. The embedded space was binned using k-means clustering with $k = 100$ clusters. In the discrete space, a transition matrix between bins was estimated from the observed transition counts C_{ij} from a microbin i to microbin j as $T_{ij} = C_{ij}/C_i$ and $C_i = \sum_j C_{ij}$. This transition matrix is commonly referred to as a "Markov Model" broadly used in the analysis of molecular systems.[96] 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 To assess the quality of the description of the cell state dynamics, we adopted a self-consistent measure of how likely a test set of single-cell trajectories were within a transition matrix model trained from a separate training set of trajectories. Data were split into a training set (5/6 images stacks per treatment) and a validation set (1/6 image stacks per treatment). The training set was used to train a transition matrix likelihood model, and the average log-likelihood per trajectory step was calculated from the test 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 To capture the broad relationships defined in the continuous trajectory embedding space we defined a small set of discrete cell "states". We first picked out all metastable locations on the landscape. We identified these metastable locations by local maxima in the population density, see Figure 5 and extended figure 3. These fine-grained metastable cell states spanning the trajectory embedding landscape (19 total) were picked via density peaks in the individual treatments. To group these fine-grained cell states into macrostate groupings, we first utilized an unsupervised kinetic clustering approach[97] 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 then manually refined these regions to distinguish between clustered cell states (states 4, 5, and 6) differentially occupied between ligand treatment. These

6 qualitative cell macrostates were used to define cell macrostate transition networks (transition matrix, see **Cell state clustering and prediction** and Figure 5A). Cells were mapped into these macrostates by first finding the closest fine-grained metastable state in the embedding space, and then assigning the macrostate label accordingly. The morphodynamical trajectory embedding space indicates a continuum of cell states, thus intermediate metastable states as indicated by density peaks which are nearest-neighbors but assigned to different macrostates, such as E and F assigned to the epithelial-like macrostate and G and H assigned to the mesenchymal-like macrostate, may be very similar and not strictly distinct. 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. Error bars and 95% confidence intervals are estimated from the data splits by Bayesian bootstrapping[98].

Acknowledgments

We thank Ian McLean and Mark Dane for project guidance and assistance accessing LINCS data, David Aristoff and Gideon Simpson for mathematical insights, 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 DRQ-09-20. 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 and NSF grant MCB 2119837. 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 in the project github repository.

https://github.com/jcopperm/celltraj

LINCS MCF10A Molecular Deep Dive data is available in some formats from the synapse database[88] 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 (1979)* **346**, 1257998 (2014).
4. Stephens, D. J. & Allan, V. J. Light microscopy techniques for live cell imaging. *Science* vol. 300 82–86 Preprint at https://doi.org/10.1126/science.1082160 (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 Preprint at https://doi.org/10.1146/annurev-physiol-022516-034055 (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 Preprint at https://doi.org/10.1016/j.molcel.2015.02.033 (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 Preprint at https://doi.org/10.1016/j.ymeth.2017.02.007 (2017).
10. Gordonov, S. *et al.* Time series modeling of live-cell shape dynamics for image-based phenotypic profiling. *Integrative Biology* **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 Preprint at https://doi.org/10.1016/j.coisb.2017.12.008 (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. Heryanto, Y. D. *et al.* Integrated analysis of cell shape and movement in moving frame. *Biol Open* **10**, (2020).
25. Wu, Z. *et al.* DynaMorph: Learning morphodynamic states of human cells with live imaging and sc-RNAseq. *bioRxiv* 2020.07.20.213074 Preprint at https://doi.org/10.1101/2020.07.20.213074 (2020).
26. Jaynes, E. T. Macroscopic Prediction. in 254–269 (Springer, Berlin, Heidelberg, 1985). doi:10.1007/978-3-642-70795-7_18.
27. 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).
28. Takens, F. Detecting strange attractors in turbulence. in 366–381 (Springer, Berlin, Heidelberg, 1981). doi:10.1007/bfb0091924.
29. Muldoon, M. R., Broomhead, D. S., Huke, J. P. & Hegger, R. Delay embedding in the presence of dynamical noise. *Dynamics and Stability of Systems* **13**, 175–186 (1998).
30. Ragwitz, M. & Kantz, H. Markov models from data by simple nonlinear time series predictors in delay embedding spaces. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **65**, 12 (2002).
31. 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).
32. Alexander, R., Zhao, Z., Székely, E. & Giannakis, D. Kernel analog forecasting of tropical intraseasonal oscillations. *J Atmos Sci* **74**, 1321–1342 (2017).
33. 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.
34. 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.
35. Wang, J. & Ferguson, A. L. Recovery of Protein Folding Funnels from Single-Molecule Time Series by Delay Embeddings and Manifold Learning. *Journal of Physical Chemistry B* **122**, 11931–11952 (2018).
36. Topel, M. & Ferguson, A. L. Reconstruction of Protein Structures from Single-Molecule Time Series. (2020).

37. Gross, S. M. *et al.* A multi-omic analysis of MCF10A cells provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses. *Commun Biol* **5**, (2022).
38. 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 Preprint at https://doi.org/10.1016/j.cels.2017.11.001 (2018).
39. 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).
40. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv* (2018).
41. Gardiner, C. Stochastic Methods: A Handbook for the Natural and Social Sciences. Preprint at (2009).
42. 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).
43. Kennel, M. B., Brown, R. & Abarbanel, H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys Rev A (Coll Park)* **45**, 3403–3411 (1992).
44. 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).
45. Kellogg, E. H., Lange, O. F. & Baker, D. Evaluation and optimization of discrete state models of protein folding. *Journal of Physical Chemistry B* (2012) doi:10.1021/jp3044303.
46. Noé, F. & Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr Opin Struct Biol* **18**, 154–162 (2008).
47. 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).
48. 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).
49. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
50. Chaffer, C. L., San Juan, B. P., Lim, E. & Weinberg, R. A. EMT, cell plasticity and metastasis. *Cancer and Metastasis Reviews* **35**, 645–654 (2016).
51. Singh, M., Yelle, N., Venugopal, C. & Singh, S. K. EMT: Mechanisms and therapeutic implications. *Pharmacol Ther* **182**, 80–94 (2018).
52. 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).
53. 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).
54. Risom, T. *et al.* Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. *Nat Commun* **9**, 3815 (2018).

55. 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).
56. Karacosta, L. G. *et al.* Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. *Nat Commun* **10**, 1–15 (2019).
57. 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.
58. Matsushita, Y. & Kaneko, K. Homeorhesis in Waddington's Landscape by Epigenetic Feedback Regulation. (2019).
59. Banerji, C. R. S. *et al.* Cellular network entropy as the energy potential in Waddington's differentiation landscape. *Sci Rep* **3**, 3039 (2013).
60. 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.
61. Gray, J. W. & Pallavicini, M. G. Ara-C scheduling: Theoretical and experimental considerations. *Med Pediatr Oncol* **10**, 93–108 (1982).
62. Swierniak, A., Polanski, A. & Kimmel, M. Optimal control problems arising in cell-cycle-specific cancer chemotherapy. *Cell Prolif* **29**, 117–139 (1996).
63. 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).
64. 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.
65. Cunningham, J. *et al.* Optimal control to reach eco-evolutionary stability in metastatic castrate-resistant prostate cancer. *PLoS One* **15**, e0243386 (2020).
66. Osada, R., Funkhouser, T., Chazelle, B. & Dobkin, D. Shape distributions. *ACM Transactions on Graphics (TOG)* **21**, 807–832 (2002).
67. Chuang, G. C. H. & Kuo, C. C. J. Wavelet descriptor of planar curves: Theory and applications. *IEEE Transactions on Image Processing* **5**, (1996).
68. Ternes, L. *et al.* A multi-encoder variational autoencoder controls multiple transformational features in single-cell image analysis. *Communications Biology 2022 5:1* **5**, 1–10 (2022).
69. C. Greenwald, E., Mehta, S. & Zhang, J. Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. *Chem Rev* **118**, 11707–11794 (2018).
70. Linghu, C. *et al.* Spatial Multiplexing of Fluorescent Reporters for Imaging Signaling Network Dynamics. *Cell* **183**, 1682-1698.e24 (2020).
71. Yang, J. M. *et al.* Deciphering cell signaling networks with massively multiplexed biosensor barcoding. *Cell* **184**, 6193-6206.e14 (2021).
72. Helmchen, F. & Denk, W. Deep tissue two-photon microscopy. *Nature Methods 2005 2:12* **2**, 932–940 (2005).
73. Alexander, S., Weigelin, B., Winkler, F. & Friedl, P. Preclinical intravital microscopy of the tumour-stroma interface: invasion, metastasis, and therapy response. *Curr Opin Cell Biol* **25**, 659–671 (2013).

74. Bullen, A., Friedman, R. S. & Krummel, M. F. Two-photon imaging of the immune system: A custom technology platform for high-speed, multicolor tissue imaging of immune responses. *Curr Top Microbiol Immunol* **334**, 1–29 (2009).
75. Vicar, T. *et al.* Cell segmentation methods for label-free contrast microscopy: Review and comprehensive comparison. *BMC Bioinformatics* **20**, 360 (2019).
76. Tscherepanow, M., Zöllner, F. & Kummert, F. *Automatic Segmentation of Unstained Living Cells in Bright-Field Microscope Images*.
77. 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).
78. 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).
79. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods 2014 11:4* **11**, 417–422 (2014).
80. Lin, J. R., Fallahi-Sichani, M. & Sorger, P. K. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nature Communications 2015 6:1* **6**, 1–7 (2015).
81. Tsujikawa, T. *et al.* Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Rep* **19**, 203–217 (2017).
82. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).
83. Angerer, P. *et al.* Single cells make big data: New challenges and opportunities in transcriptomics. *Curr Opin Syst Biol* **4**, 85–91 (2017).
84. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols 2018 13:4* **13**, 599–604 (2018).
85. Yang, K. D. & Uhler, C. Multi-Domain Translation by Learning Uncoupled Autoencoders. (2019).
86. Cao, K., Hong, Y. & Wan, L. Manifold alignment for heterogeneous single-cell multi-omics data integration using Pamona. *bioRxiv* 2020.11.03.366146 Preprint at https://doi.org/10.1101/2020.11.03.366146 (2020).
87. 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).
88. Gross, S. M. *et al.* A multi-omic analysis of MCF10A cells provides a resource for integrative assessment of ligand-mediated molecular and phenotypic responses. *Commun Biol* **5**, (2022).
89. 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.
90. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18**, 100–106 (2021).
91. Coelho, L. P. Mahotas: Open source software for scriptable computer vision. *J Open Res Softw* **1**, e3 (2013).

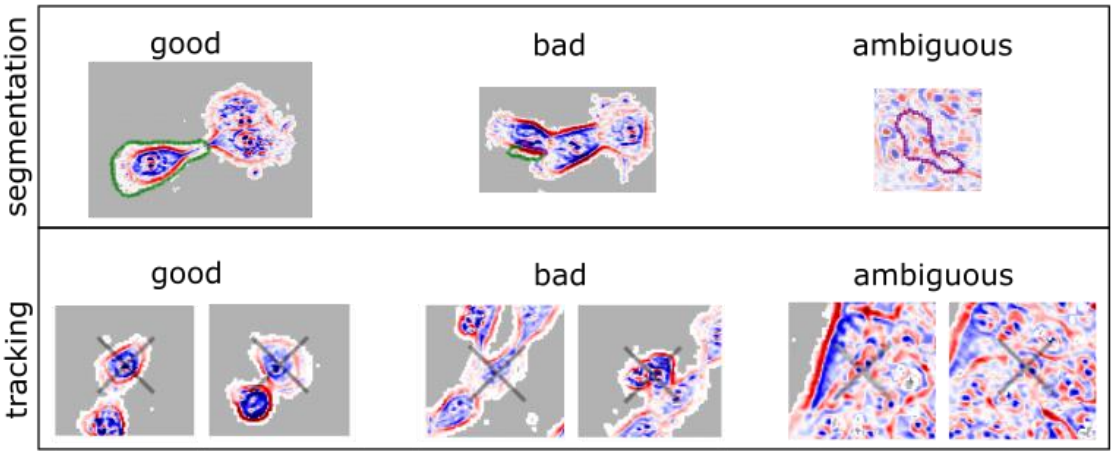
92. Alizadeh, E., Xu, W., Castle, J., Foss, J. & Prasad, A. TISMorph: A tool to quantify texture, irregularity and spreading of single cells. *PLoS One* **14**, e0217346 (2019).
93. Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Transactions on Image Processing* **7**, 27–41 (1998).
94. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37**, 38–47 (2019).
95. 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).
96. Noé, F. & Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr Opin Struct Biol* **18**, 154–162 (2008).
97. 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).
98. 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%

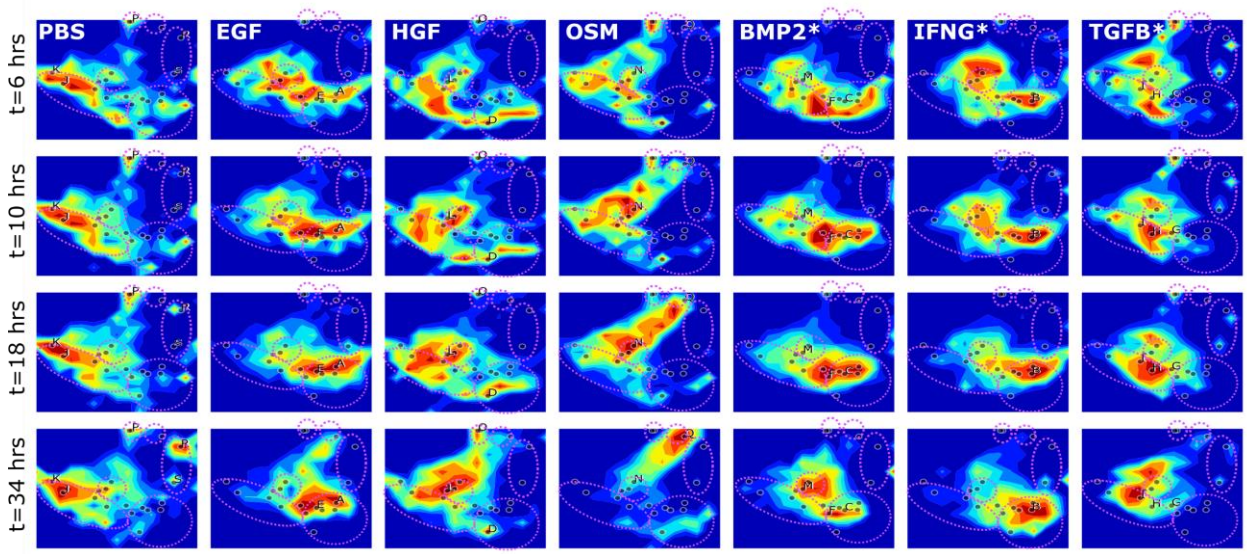
Extended Figure 1. Segmentation and tracking manual validation. Examples of good, bad, and ambiguous qualitative validation categories for segmentation and tracking.



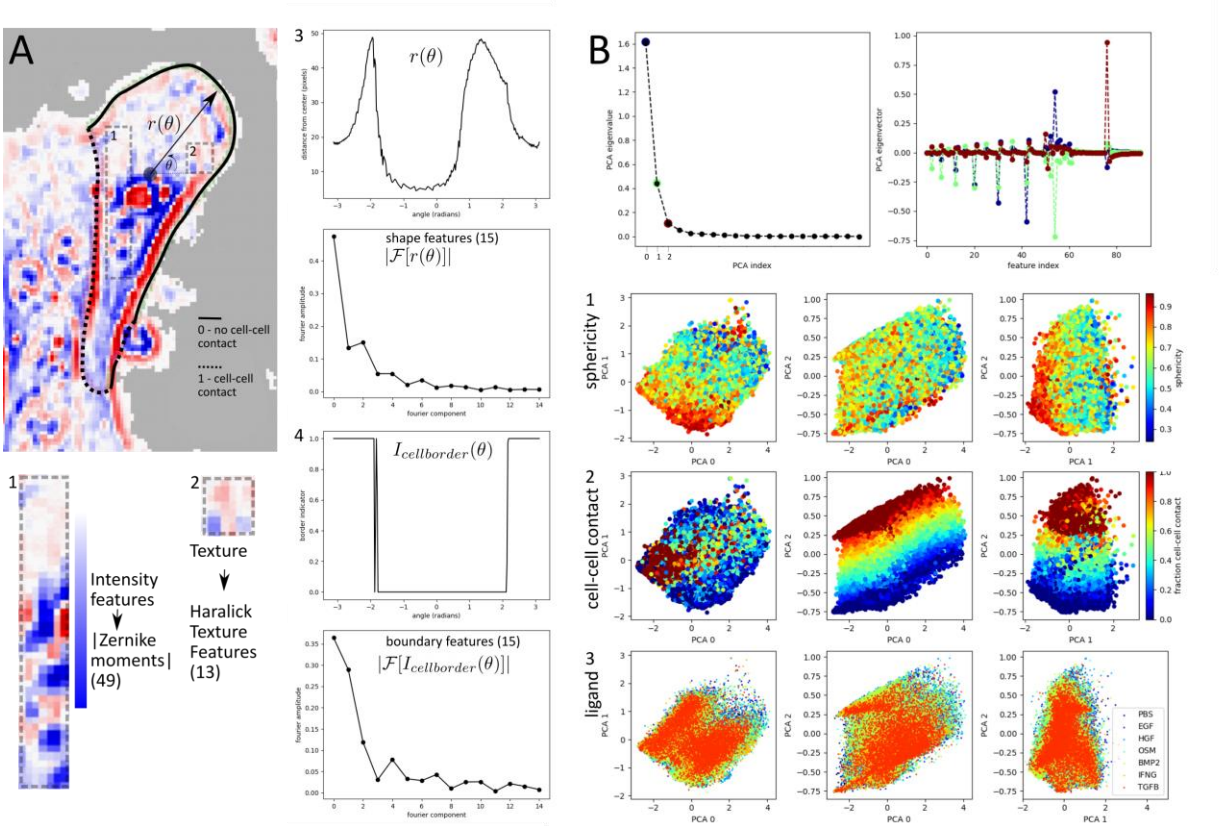
Extended Data Table 2. Number of extracted trajectory snippets with increasing snippet length. *(+EGF)

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

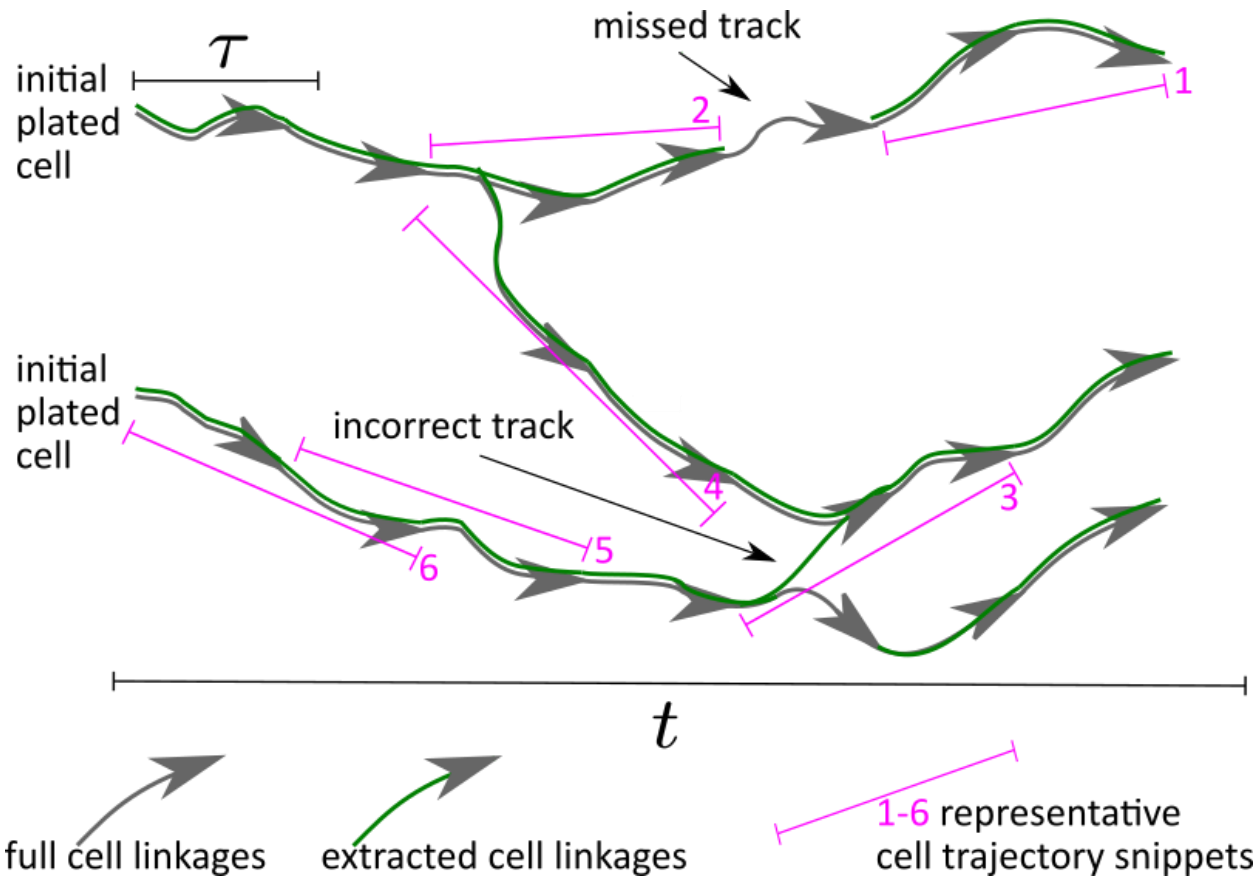
Extended Figure 2: Trajectory embedding constructs a common space to evaluate unique and shared cell morphodynamics. Top left: outline of the combined density distribution in the trajectory embedding (snippet length = 8) space (gray), with locations of the density peaks in individual treatments marked 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.



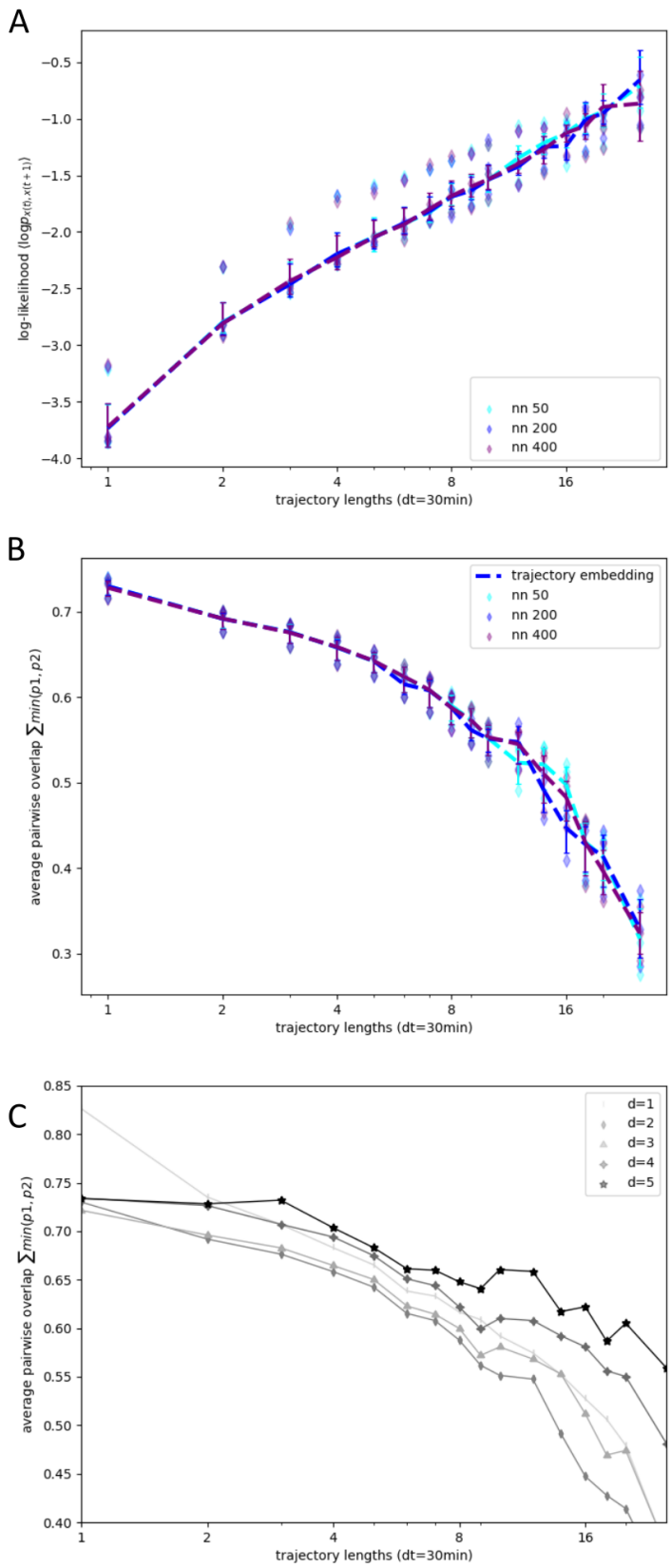
Extended Figure 3:
Time-dependent cumulative distributions (rainbow) from 12-hr windowed averages (or maximum allowed window average), and density peak locations selected as fine-grained metastable states (labels A-S).
*(+EGF)



Extended Figure 4:
Cell features and PCA reduction. A) Segmented single-cell with cell border in contact with another cell (dashed line), and not in contact (solid line), and cell feature description 1. Phase contrast overall features described by Zernike moment absolute values. 2. Texture described by Haralick texture features. 3. Cell shape featurized by the absolute value of the Fourier coefficients of the distance to the cell center as a function of the angle θ. 4. Local cell environment described by the absolute value of the Fourier transform of $I(\theta)$ indicating cell-cell contact. B) PCA feature reduction. PCA eigenvalues (top left) and eigenvectors for the top 3 components explaining >90% of the feature variance (top right). PCA landscape colored by 1. Approximate sphericity given by the first Fourier component of $r(\theta)$, approximate fraction of the cell boundary in contact with another cell given by the first Fourier component of the cell-cell border indicator $I(\theta)$, and 3. by ligand condition.



Extended Figure 5:
Graphical illustration of the full set of cell linkages connecting 2 initially plated cells to the cells at the final timepoint (gray arrows). The extracted cells and linkages from the cell segmentation and tracking steps as green arrows indicating the available partial set of cells and linkages used in the data analysis with errors and missing cells, and some possible trajectory snippets (yellow highlights) extracted in a sliding window manner along the extracted linkages.



Extended Figure 6:
A) Trajectory log-likelihood with trajectory length at half the value of the UMAP n_neighbors parameter (nn=50, cyan) and twice the value of the n_neighbors parameter (nn=400, purple), and the value used in Figures 2-5 (nn=200). B) Average pairwise overlap between ligand population distributions at half the value of the UMAP n_neighbors parameter (nn=50, cyan) and twice the value of the n_neighbors parameter (nn=400, purple), and the value used in Figures 2-5 (nn=200). C) Average pairwise overlap between ligand population distributions with UMAP embedding dimension (d=1-5, light-gray to black).