Cell shape noise strength regulates shape dynamics during EMT-associated cell spreading

Wolfram Pönisch[1,†], Iskra Yanakieva[1,†], Guillaume Salbreux[2,*], Ewa K. Paluch[1,*]

[1] Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom.

[2] Department of Genetics & Evolution, University of Geneva, Geneva, Switzerland

† These authors contributed equally.

* Corresponding authors: guillaume.salbreux@unige.ch (G.S.), ekp25@cam.ac.uk (E.K.P)

ABSTRACT

Cellular shape is intimately linked to cell function and state, and transitions between cell states are tightly coupled to shape changes. Yet, shape has been largely overlooked in state transitions studies. Here, we combine morphometric analysis with theoretical modeling and molecular perturbations to interrogate cell shape dynamics during epithelial-to-mesenchymal transition (EMT). Using stochastic inference, we extract the morphogenetic landscape underlying EMT. We show that within this landscape, EMT-associated cell spreading reflects a transition between shape attractors. Strikingly, we observe a peak in cell shape noise strength concomitant with spreading, and show that higher shape noise accelerates transitions between shape attractors. Our analysis framework will be widely applicable to quantitative cell shape investigations in physiology and disease. Together, our results identify a key role for cellular stochasticity as a regulator of shape change rates, and highlight that shape dynamics yield rich phenotypic information that enhance our understanding of cellular states.

INTRODUCTION

Most cell types in an organism display functionally adapted, well-defined shapes, and cell shape deregulation is a harbinger of disease, from developmental disorders to cancer[1–3], a feature utilized in histological diagnostics. As such, precise control of cell shape is a key aspect of cell function, and morphological changes accompany most cellular state and fate transitions during embryonic development, tissue homeostasis and disease progression.

Cell state transitions have been extensively studied, yet most investigations focus on rewiring of underlying of gene regulatory networks (GRNs). Large scale -omics analyses, combined with advanced bioinformatics pipelines have established highly quantitative frameworks for the definition of cell states based on differences in patterns of gene expression[4]. These patterns are often described using the metaphor of the Waddington epigenetic landscape where GRN configurations are described through landscapes with peaks and valleys, and valleys represent attractor GRN states towards which cells are drawn. Mathematically, this concept of state attractors has provided a powerful framework, yielding many insights into state determination and the mechanisms of cell state change at the gene expression level[5–7].

In contrast, changes in cell morphology associated with state transitions have received much less attention. This is partly because cell shape changes ultimately result from changes in cell mechanics and in forces exerted on the cell surface. While mechanical changes can be triggered by GRN rewiring, they are mediated by physical reorganization of cytoskeletal architecture, of the plasma membrane, and other mechanically active cellular structures[8–10]. Understanding how these mechanical properties relate to GRN configurations is challenging

and remains very poorly understood. As a result, morphological changes associated with cell state transitions have been largely overlooked.

Another obstacle in investigating morphological transitions is the technical challenge of quantifying cell shape, particularly in live cells. Historically, shape changes have mostly been monitored using hand-picked descriptors of interest, such as aspect ratio or projected area to describe cell spreading. While such measures can be informative[11,12], they provide only limited information on overall cell morphology. In recent years, enhanced capabilities of imaging and segmentation have been enabling more systematic approaches to cell shape investigations based on morphometrics. In these approaches, cell shape is quantified through a high-dimensional set of features, either intuitive[13–18] or using mathematical descriptors[19–26]. Using dimensionality reduction, shape features are then mapped into a low-dimensional morphospace, which quantitatively describes cell shapes. These new developments provide a mathematical framework for the quantitative investigation of cellular shape changes.

Recent studies have used morphospace analysis to investigate epithelial-to-mesenchymal transition (EMT)[27–29], a phenotypic state change of key importance for development, wound healing and disease[30,31] . During EMT, cells transition from compact epithelial morphologies to spread cell shapes and increased cell motility (Figure 1A). These shape changes enable cells of epithelial origin to migrate away from their original position, a key function of EMT[31]. As such, the extent of cell shape change has been used as a single cell dynamic readout of EMT progression[32,28]. Furthermore, combined analysis of 2D cell shape trajectories and intracellular distributions of mesenchymal markers, suggests that EMT is driven by a transition between attractors in this composite feature space[28]. To what extent attractors

can be identified when considering cell shape alone, and how cells dynamically transition between such attractors has, to our knowledge, not been investigated. As a result, the mechanisms controlling the dynamics of shape change in EMT remain poorly understood.

Here, we investigated 3D cell shape dynamics during EMT and asked whether the shape change can be described as a transition between shape attractors in a morphogenetic landscape. We used Madin-Darby canine kidney (MDCK) epithelial cells, widely used in studies of EMT[33,34], and implemented morphospace analysis to investigate the stochastic dynamics of complex 3D cell shape change. We then used stochastic inference to extract the morphogenetic landscape and the strength of cell shape noise underlying morphospace shape trajectories. We found that EMT-associated cell spreading can be mathematically accounted for by a change in shape attractor. Strikingly, we also observed a strong increase in cell shape noise at the time of the shape transition. Theoretical analysis and perturbation experiments strongly suggest that cell shape noise and fluctuations accelerate cell spreading. Together, our study shows that EMT-associated cell spreading can be mathematically described as a transition between cell shape attractors, and that enhanced cell shape noise facilitates this transition, thus playing a key role in controlling cell shape change.

RESULTS

Generating a morphospace for the analysis of EMT-associated cell shape dynamics

We investigated 3D cell shape changes displayed by MDCK cells during EMT using live imaging and cell morphometric analysis. Cells were cultured on soft polyacrylamide gels to promote cuboidal cell morphologies in the epithelial state[35] (Figure 1B). To visualize cell shape changes, we used mosaic labeling by mixing MDCK cells stably expressing Lifeact-GFP to mark F-actin with unlabeled MDCK cells (Figures 1B and 1C, Movie S1). This mosaic F-actin labeling, where most Lifeact-GFP-expressing cells are surrounded by unlabeled neighbors, enabled semi-automated 3D cell shape segmentation. Nuclei were labeled in all cells using the live stain SiR-DNA (Figure 1B and 1C, Movie S1) to visualize cell colonies. We induced EMT by adding Hepatocyte Growth Factor (HGF)[34] and acquired 3D image stacks for 10 to 22 h after EMT induction at 10 min time intervals. While control cells remained cuboidal throughout the imaging window, cells undergoing EMT displayed dramatic shape changes of the whole cell body as they spread on the substrate (Figure 1C, Movie S1).

To quantitatively characterize the 3D shape dynamics of MDCK cells during EMT, we developed a morphometric analysis pipeline (Figure 1D; Methods). Cells were segmented with Imaris (Bitplane/Oxford Instruments), yielding timecourses of 3D triangulated meshes (Figure 1D(i)). We first aligned the long cell axes by rotating cell meshes in the plane of the substrate, so that a cell's orientation did not impact its position in morphospace (Figure 1D(ii)). We then mapped the cell surface mesh points onto a sphere and performed a spherical harmonics (SPHARMs) decomposition to extract a list of SPHARMs coefficients that represent the cell shape, providing a robust, automated quantification of cell morphology

during EMT (Figure 1D(iii), S1; Methods). We determined that considering SPHARMs functions up to a degree of 23, corresponding to 1728 coefficients, was sufficient to reconstruct cell shapes in our dataset with a high degree of accuracy (Figure S1). To facilitate the visualization and interpretation of this high-dimensional dataset, we then performed dimensionality reduction using principal component analysis (PCA; Figure 1D(iv); Methods), enabling the representation of cell shapes as points in a low-dimensional morphospace (Figure 1D(v)). We applied PCA on a dataset of cells undergoing EMT at 0 h and 18 h after EMT induction (0 h and 18 h EMT cells thereafter), as these time points reflect the extreme stages of the shape trajectories, corresponding to epithelial and mesenchymal morphologies, respectively (Figure 1C). This determined the principal components (PCs) of MDCK cell shape variation during EMT, which we used as axes of a morphospace to mathematically represent and analyze cell shapes in the rest of the study.

Morphometric analysis of EMT-associated cell shape change

Next, we investigated the PCs of cell shape variation. We first checked how much of the variance of our dataset of 0 h and 18 h EMT cell shapes was accounted for by increasing numbers of principal components. We found that PC1 and PC2 accounted for 46.3% and 10.2% of the cell shape variance, respectively, and that 22 PCs were required to account for more than 90% of the total variance of the dataset (Figure 1E). While part of the variance in our cell shape dataset reflects EMT-associated cell shape change, some of the variance corresponds to cell shape variability between cells at any given timepoint. To determine which of the first 22 PCs most strongly reflect EMT-associated cell shape changes, we

compared the standard deviations of each PC values for EMT cells at 0 h and 18 h to the distance d_i between the PC means at 0 h and 18 h (Figure 1F):

$$d_i = \left|\langle \mathrm{PC}i \rangle_{\mathrm{EMT,0h}} - \langle \mathrm{PC}i \rangle_{\mathrm{EMT,18h}}\right| \tag{1}$$

where $i = 1 \ldots 22$ (Figure 1F). We found that the distance exceeded the standard deviations, $d_i > \sigma_{i,\mathrm{EMT,0h}}$ and $d_i > \sigma_{i,\mathrm{EMT,18h}}$, only for PC1. For subsequent PCs, cell shape variability pre- and post-EMT exceeds the average change in PC value over the course of EMT. This indicates that PC1 reflects most of the EMT-associated shape variation in our dataset. Visual analysis indicates that an increase in PC1 corresponds to increased projected area and decreased cell height, reflecting cell spreading, while an increase in PC2 reflects asymmetric elongation of the cell (Figure 1G, Movie S2, Movie S3).

We then analyzed control and EMT cells in the two-dimensional PC1-PC2 morphospace by plotting the locations of control and EMT cells at 0 h and 18 h (Figures 2A and S2A-S2D). We found that control cells were distributed around PC1 ≈ -10 and PC2 ≈ 0 and their distribution did not considerably change after 18 h of imaging. In contrast, while the distribution of EMT cells overlapped with that of control cells at 0 h (Figure S2C), 18 h EMT cells displayed a clear shift along PC1 and were distributed around PC1 ≈ 10 and PC2 ≈ 0 (Figure S2D). Consistent with our analysis of PC distance and standard deviation (Figure 1F), the mean morphospace position of the cells mostly changed along PC1. The width of the distributions of cell positions in morphospace increased both along PC1 and PC2, suggesting an increase in overall cell shape variability during EMT (Figures 2A and S2A-S2D). Together, our morphometric analysis yields a high-fidelity representation of the 3D shapes of cells

undergoing EMT. Our analysis further identified PC1 as a one-dimensional cell shape descriptor that best accounts for cell shape change during EMT.

Stochastic inference analysis of morphospace trajectories

We then investigated the temporal dynamics of EMT-associated cell shape changes. First, we characterized ensemble cell shape dynamics by averaging the complete set of SPHARMs coefficients of control and EMT cells over the imaging timecourse and reconstructing the resulting morphologies, yielding ensemble-averaged cell shapes (Figure 2B). While the reconstructed average control cell did not display major shape change, the average EMT cell exhibited spreading between 0 h and 9-12 h and maintained a spread morphology after 12 h. To further quantify spreading dynamics, we analyzed the average PC1 coordinate of control and EMT cells over time (Figure 2C). Consistent with ensemble average cell shapes (Figure 2B), while PC1 of control cells increased weakly with time, PC1 of EMT cells steeply increased between 0 h and 9-12 h, and then saturated. Cell projected area and cell height, which we identified as associated with changes in PC1 (Figure 1G), displayed a similar behavior, saturating around 10 h post EMT induction (Figures S2E and S2F). In contrast, between 0 h and 6-9 h, PC2 displayed no difference in average dynamic behavior between control cells and cells undergoing EMT up to 6 h and showed a weak difference after 6-9 h (Figures S2G), while PC3 displayed only very minor differences over the total timecourse (Figures S2H). The dynamics of PC2 and PC3 further support our conclusion that PC1 captures most of the morphological changes associated with EMT.

We then asked if individual cell shape dynamics during EMT could be accounted for through a morphogenetic landscape formalism (Figure 3A). Visual analysis of morphospace cell shape trajectories (Figures 3B, S2I and S2J) suggested that between 3 h and 12 h, coinciding with the time of cell spreading (Figures 2B and 2C), individual cells displayed high variation in their morphospace position, suggesting enhanced cell shape fluctuations. To further explore the these fluctuations, we asked if the dynamics of cell shape during EMT could be described as a stochastic process driven by deterministic and random forces. We considered a stochastic process described through the following Langevin equation:

$$\underbrace{\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{PC1}(t)}_{\text{Rate of change of PC1}} = \underbrace{\Phi(\mathrm{PC1},t)}_{\text{Deterministic force}} + \underbrace{\sqrt{2D(\mathrm{PC1},t)}\Gamma(t)}_{\text{Stochastic force}} \tag{2a}$$

written in the Itô convention[36,37], and where $\Gamma(t)$ is a Gaussian white noise fulfilling:

$$\langle\Gamma(t)\rangle = 0 \quad \text{and} \quad \langle\Gamma(t)\Gamma(t')\rangle = \delta(t-t') \tag{2b}$$

In these equations, $\Phi(\mathrm{PC1},t)$ represents the deterministic morphospace force originating from the underlying morphogenetic landscape (Figure 3A); this force drives cell shape towards cell shape attractors. The term $\sqrt{2D(\mathrm{PC1},t)}\Gamma(t)$ represents the stochastic dynamics of cell shape changes, where the diffusion coefficient $D(\mathrm{PC1},t)$ reflects the strength of cell shape noise, which originates from fast processes that are not resolved spatially or temporally but effectively lead to forces driving undirected cell shape changes. In this framework, cell shape fluctuations result from the local structure of the morphogenetic landscape and from the strength of cell shape noise. Of note, the morphospace forces do not have the dimension of mechanical forces but units of [1/h], but they reflect the action of mechanical forces that drive cell shape change.

To infer the deterministic force Φ and the diffusion coefficient D, we adapted a method known as *stochastic (force) inference*[38] (Methods) and expanded it to allow for a time dependence of the deterministic and stochastic force fields. Indeed, in the most generic description, both the deterministic force Φ and the diffusion coefficient D can depend on both cell shape PC1 and time t. Stochastic force inference analysis of the one-dimensional PC1 trajectories yielded the deterministic force fields Φ (Figures 3C and 3D) and diffusion coefficients D (Figures 3E and 3F) for control and EMT cells. We then tested whether the inferred dynamics provide an accurate representation of the stochastic dynamics of the EMT-associated cell shape changes by numerically reconstructing PC1 trajectories from the inferred $\Phi(\mathrm{PC1}, t)$ and $D(\mathrm{PC1}, t)$. We found that trajectories looked visually similar between our experimental observations and *in-silico* reconstructed trajectories, and their stochastic dynamics exhibited excellent agreement (Figures S3A-S3F; Methods). This suggests that our stochastic analysis pipeline reliably extracts the deterministic force field Φ and diffusion coefficient D from cell shape trajectories.

Stochastic inference reveals the morphogenetic landscape and cell shape noise dynamics underlying EMT

We then analyzed the deterministic force Φ for control and EMT cells (Figure 3C and D). The magnitude of Φ was generally larger for EMT cells than for control cells. For control cells, Φ was positive in the first 6 hours (Figure 3C). After that, we could identify a stable attractor point around $\mathrm{PC1} \approx 0$ (characterized by $\Phi = 0$ and $\frac{\mathrm{d}\Phi}{\mathrm{dPC1}} < 0$), which we interpret as a shape attractor associated with the epithelial cell state. However, the magnitude of the

deterministic force Φ of control cells was small (average mean absolute rate of change per hour of $0.45\ \mathrm{h}^{-1}$), consistent with PC1 remaining mostly unchanged in control cell shape trajectories (Figure S3A). In contrast, for EMT cells, a stable attractor appears to emerge as early as 2-3 h after EMT induction, with the deterministic force pushing cell morphologies towards a stable point $\mathrm{PC1} \approx 10$ (characterised by $\Phi = 0$ and $\frac{\mathrm{d}\Phi}{\mathrm{dPC1}} < 0$ and mean absolute rate of change of PC1 of $3\ \mathrm{h}^{-1}$) (Figure 3D). Hence, $\mathrm{PC1} \approx 10$ can be interpreted as a shape attractor associated with the mesenchymal state. After 12 h, the magnitude of Φ becomes smaller and the deterministic force magnitude is no longer sufficient to drive any major shape changes (mean absolute rate of change of PC1 per hour of $0.39\ \mathrm{h}^{-1}$).

We next analyzed the diffusion coefficient D, characterizing the strength of the noise acting on cell shape, for control and EMT cells (Figures 3E and 3F). We found that EMT cells displayed an overall larger diffusion coefficient D than control cells and that for EMT cells, D increased with increasing PC1. Furthermore, we observed that D exhibited non-monotonous temporal behavior for EMT cells: D first increased, reached a maximum around 6 h, corresponding to the mid-timepoint of cell spreading (Figure 2C), and then decreased. This observation is in agreement with the high fluctuations in morphospace position observed in the EMT shape trajectories between 3 h and 12 h (Figures 3B, S2I and S2J). Finally, we note that multiplicative noise, in the Itô interpretation used here, can shift the peak in steady-state probability distribution away from the deterministic attractor. We find however that this effect here is quantitatively small (Supplemental Information, FigureS4). Taken together, our stochastic dynamics analysis of shape trajectories shows that EMT is associated with changes in the morphogenetic landscape underlying cell shape and is accompanied by a peak in cell shape noise strength around the time of spreading.

A minimal stochastic model recapitulates cell shape attractors associated with EMT

To simplify the comparison of dynamic cell shape trajectories between different conditions, we next considered a minimal stochastic model of EMT, governed by the following equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{PC1}(t) = -\frac{\mathrm{d}}{\mathrm{dPC1}}V_m(\mathrm{PC1}) + \sqrt{2D_m(t)}\Gamma(t) \tag{3}$$

where we introduced the morphospace potential V_m:

$$V_m(\mathrm{PC1}) = -\int \Phi_m(\mathrm{PC1})\ \mathrm{dPC1} \tag{4}$$

derived from a deterministic force field $\Phi_m(\mathrm{PC1})$. Equation 3 is equivalent to Equation 2, with the simplifications that the deterministic force is no longer time-dependent, but only depends on PC1, and that the diffusion coefficient $D_m(t)$ only depends on time and not on PC1. We found that the morphospace potential V_m inferred with this minimal model, displayed a single minimum for both control cells and cells undergoing EMT (Figure 3G). The minima locations, reflecting stable shape attractors, were located at $\mathrm{PC1} \approx 0$ and $\mathrm{PC1} \approx 10$ for control and EMT cells, respectively, in agreement with the shape attractors inferred with Equation 2 (Figure 3C and 3D). The inferred temporal dynamics of the diffusion coefficient D_m further showed that the strength of cell shape noise dramatically increased during EMT and peaked around 6 h (Figure 3H). Finally, we checked that the peak in D_m is not just a feature of PC1 but could also be observed when considering shape trajectories in a high-dimensional space consisting of the first 22 principal components (Figure S3G; Methods). Our analysis shows that even a minimal stochastic model recapitulates the main features of

EMT morphodynamics (Figure 3C-F), namely the existence of stable shape attractors and a peak in the strength of cell noise. We thus used this minimal model to investigate the function of cell shape noise in the rest of the study.

A theoretical model shows that a transient peak in noise facilitates transitions between attractors

We next asked if the increased shape noise strength we observe during EMT cell spreading could facilitate the cell shape transition. To this aim, we built a simple mathematical model to analyze the dynamics of a state descriptor ρ, starting at an initial level ρ_{start} (which in our experiments would correspond to the epithelial shape attractor) and evolving towards a potential minimum ρ_{m} (corresponding to the mesenchymal shape attractor) with different levels of noise. System dynamics were described by Equation 3, using a simple quadratic potential $V_s = \frac{k_s \rho^2}{2}$, corresponding to a landscape with a single minimum at $\rho_{\mathrm{m}} = 0$ (Figure 4A), and a time-dependent noise strength:

$$D_{\mathrm{s}}(t) = D_{\mathrm{b}} + \Delta D \exp\left[-\left(\frac{t - t_D}{\tau}\right)^2\right] \tag{5}$$

which describes a noise peak with a background noise strength D_{b}, a peak height ΔD, a peak time t_D, and a peak width characterized by the time τ. We considered three cases (Figure 4B): and a noise peak ($D_{\mathrm{b}} = 3\ \mathrm{h}^{-1},\ \Delta D = 13\ \mathrm{h}^{-1}$, qualitatively recapitulating the noise peak in Figure 3H of EMT cells), a constant low noise ($D_{\mathrm{b}} = 3\mathrm{h}^{-1},\ \Delta D = 0$), and a constant high noise ($D_{\mathrm{b}} = 16\ \mathrm{h}^{-1},\ \Delta D = 0$). The model parameters and mathematical solution are detailed in the Methods and Supplemental Information. We then asked how many

trajectories reached the potential minimum ρ_{m} after a given time t. We found that a higher proportion of trajectories with a constant high noise or with a noise peak passed the potential minimum compared to trajectories with constant low noise (Figure 4C). In our experiments, cells display a spectrum of mesenchymal shapes following EMT (Figure 2A), we thus analyzed transition dynamics towards a mesenchymal region corresponding to $\rho > -\rho_{dist}$ with a threshold ρ_{dist} (Figure 4A). We simulated trajectories for each condition and quantified the mean first passage time (MFPT) of trajectories to reach the mesenchymal region (Figure 4D). We found that across all thresholds, trajectories with high constant noise and with a noise peak reached the mesenchymal region earlier than cells with low constant noise. However, high noise could also drive trajectories away from the minimum once they have passed it. Hence, we speculated that the experimentally observed decrease in noise on longer timescales could help trajectories remain robustly near the new potential minimum. To test this idea in our model, we calculated the percentage of trajectories of ρ within the mesenchymal region, $P_{\mathrm{mesenchymal}}$, as a function of time (Figure 4E). We found that initially, more trajectories reached the mesenchymal region for the constant high noise and noise peak conditions, consistent with high noise facilitating the shape change. However, for later times, more trajectories remained in the mesenchymal region for the constant low noise condition compared to the other two. For later timepoints, trajectories with constant high noise were consistently least represented in the mesenchymal region. For the noise peak condition, the proportion of trajectories in the mesenchymal region steadily increased, converging towards the low noise case over time since the diffusion coefficient D_{s} in the noise peak case converged towards the diffusion coefficient of the low noise case $D_{\mathrm{s}} = 3\ \mathrm{h}^{-1}$. Taken together, our mathematical model suggests that while high cell shape noise can accelerate a transition towards a new cell shape attractor, a transient peak in cell shape

noise strength may make the transition both faster and more robust by keeping the trajectories in the mesenchymal region after the transition.

Trajectories with higher cell shape noise strength display faster cell spreading during EMT

We then asked how cell shape noise strength affected cell shape transition dynamics in individual EMT trajectories. As the time of shape transition can be difficult to pinpoint in the noisy cell shape trajectories (Figure 3B), we developed a mathematical framework to define the efficiency of spreading probabilistically. We assumed that cells undergoing EMT displayed epithelial shapes at 0 h and mesenchymal shapes at 18 h post HGF addition and used our shape data to train a support vector machine classifier for binary classification of epithelial and mesenchymal shapes. The underlying dataset used for training were the PC1 and PC2 coordinates of EMT cells at 0 h and 18 h after EMT induction (Figure 4F). We included PC2 to increase the accuracy of the predictor, as we observed a weak difference in the ensemble dynamics of PC2 between EMT and control cells (Fig. S2G). We then used this classifier to derive the probability map of mesenchymal shapes, which assigns a mesenchymal shape score to every location in the PC1-PC2 morphospace (Figure 4F) and tracked how this score changed for individual cells undergoing EMT (Figure 4G). Finally, we used the resulting mesenchymal shape score timecourse of individual cells to define cell transition time as the time to reach a threshold mesenchymal shape score and quantified the percentage of cells that pass it between 0 h and 12 h (Figure 4H), the time window when most of the cell spreading occurs (Figure 2C). As expected, we found that a larger fraction of EMT cells reached any given threshold mesenchymal shape score compared to control cells

(Figure 4H). This confirmed that our data-driven method allows for robust quantification of the efficiency of EMT-associated cell spreading through the timing and extent of the cell shape change.

We then asked how cell spreading efficiency related to cell shape noise strength for individual cells. We computed an average diffusion coefficient D_t for each cell morphospace trajectory (Methods). Consistent with our stochastic inference analysis (Figure 3H), we found that control cells had lower D_t than EMT cells (Figure 4I). We next grouped EMT cell shape trajectories into three equally sized populations, displaying low ($D_t < 4.8\ \mathrm{h}^{-1}$), moderate ($4.8\ \mathrm{h}^{-1} \leq D_t \leq 10.8\ \mathrm{h}^{-1}$), or large ($D_t > 10.8\ \mathrm{h}^{-1}$) cell shape noise strength. Interestingly, we found that cells with strong and moderate noise strength were more likely to cross any mesenchymal shape score (Figure 4J) than cells with weak noise strength. Moreover, the larger D_t, the earlier cells reached mesenchymal shapes across all mesenchymal shape score thresholds (Figure 4K). Together, these observations show that higher cell shape noise correlates with faster and more extensive cell spreading during EMT.

Experimentally reducing cell shape noise delays EMT-associated cell spreading

Finally, we asked how reducing cell shape noise experimentally affects cell shape transition dynamics. To this end, we tested if perturbing the mechanical balance at the cell surface might affect cell shape noise during EMT. Cell spreading requires an increase in actin protrusivity and a reduction in plasma membrane tension to accommodate increased protrusion formation[39–41]. We thus sought to perturb actin and membrane dynamics, and assessed effects on cell spreading dynamics and cell shape noise strength. We first interfered

with actin protrusivity by inhibiting the activity of the small GTP-ase Rac1, a key regulator of protrusive actin polymerization in EMT[33,42]. Upon Rac1 inhibition with the small molecule inhibitor NSC23766[43], cells displayed considerably reduced spreading 18 h after EMT induction compared to untreated EMT cells (Figures 5A and 5A', Movie S4). Morphospace analysis further demonstrated that cells with reduced Rac1 activity generally displayed less spread cell shapes characterized by lower PC1 (Figures 5B, 5C and S5A). Stochastic inference showed that Rac1 inhibition in EMT cells shifted the shape attractor to lower PC1 values, and dramatically decreased cell shape noise strength D_{m} (Figure 5D). We then analyzed cell shape spreading efficiency. We computed the mesenchymal shape score timecourses (Figure S5B) and found that Rac1 inhibition drastically reduced the percentage of cells that reached any mesenchymal shape threshold after 12 h (Figure 5E), consistent with the overall reduction in cell spreading (Figures 5B and 5C). Interestingly, the cells that displayed some spreading reached their spread shape significantly later than untreated cells across all mesenchymal shape scores (Figures 5E' and S5B).

We then treated cells with methyl-β-cyclodextrin (MbCD), a compound that depletes cells of cholesterol and has been shown to increase membrane tension in various cell types[44,45,41], including MDCK cells[46]. While some MbCD-treated cells displayed reduced spreading in EMT, many cells displayed spreading visually comparable to untreated cells 18 h after EMT induction (Figure 5A'', Movie S4). Morphospace analysis demonstrated that MbCD-treated cells displayed similar values of PC1 and reduced variability in PC2 compared to untreated cells 18 h after EMT induction (Figure 5F). The ensemble EMT spreading dynamics were also similar in MbCD-treated and untreated cells (Figures 5G and 5SA). Stochastic inference further showed that the morphospace potential of MbCD-treated EMT cells was comparable

to that of untreated EMT cells, with a minimum around PC1=10 (Figure 5H). Strikingly, MbCD treatment led to a strong reduction in cell shape noise strength D_{m} (Figure 5H). We then analyzed individual cell spreading dynamics during EMT in presence of MbCD treatment. We found that the percentage of cells acquiring mesenchymal shapes after 12 h was only weakly reduced compared to untreated cells (Figure 5I), in agreement with the comparable morphospace potentials of MbCD-treated and untreated EMT cells (Figure 5H). However, MbCD-treated cells reached the mesenchymal threshold shape later than non-treated cells across all mesenchymal cell shape scores (Figure 5I', MbCD treatment leads to a ~20% increase in time cells take to reach the threshold mesenchymal shape). These findings agree with the theoretical prediction that higher cell shape noise strength could accelerate the transition from a shape attractor to another (Figure 4D). Together, our theoretical model and perturbation experiments strongly suggest that the peak in cell shape noise observed during EMT plays a functional role, accelerating EMT-associated cell spreading.

DISCUSSION

In this study, we investigated cell shape dynamics during EMT-associated cell spreading. We employed a data-driven approach to quantify the morphodynamics of 3D cell shape changes of MDCK cells undergoing EMT. Most studies of morphogenesis use intuitive shape measures (e.g., volume, solidity, aspect ratios) as quantitative features describing cell morphology[15,16,18]. However, such features are often subjectively chosen and can thus be biased. Furthermore, they do not constitute unique descriptors, and thus cannot be used to directly reconstruct the original cell shapes. To avoid these limitations, we used spherical harmonics decomposition, a mathematical framework that has been recently attracting increasing attention for cell shape representation[19,22–26]. For unbiased and interpretable dimensional reduction, we employed PCA, as it represents low-dimensional features as linear combinations of the high-dimensional spherical harmonics descriptors. The linearity of the PCA mapping allows for defining distances between distinct shapes, and for the reconstruction of theoretical shapes corresponding to each position of the PCA morphospace (see grey shapes in e.g. Figure 2A). Together, this analysis enabled the representation of complex cellular shape and shape dynamics of individual cells in a low-dimensional morphospace in an objective and interpretable way.

We then used mathematical modeling to analyze the stochastic dynamics of EMT cell shape trajectories in morphospace. Our stochastic model shows that EMT-associated shape changes can be described using the mathematical framework of attractors. How EMT-associated signaling pathways regulate the epithelial and mesenchymal shape attractors we identify here will be an interesting avenue for future studies. It will also be interesting to investigate to what extent mechanical constraints on the organization of cellular

components controlling cell surface mechanics contribute to determining cell shape attractors. Recent work suggests that actin networks can only display a limited number of configurations, which could reflect constraints on the wiring of upstream signaling pathways[13], but could also reflect mechanical and structural constraints on the configurations an actin network can display[47,48]. It is tempting to speculate that such molecular and mechanical constraints could drive the emergence of attractors in actin networks organization, which could in turn control attractors in cell shape.

Strikingly, we found that the transition between shape attractors is associated with a peak in cell shape noise strength, and our model and perturbation experiments strongly suggest that noise functionally facilitates cell shape change during EMT. Noise has been shown to be a driver of transitions between attractors in quantum optics, chemical reaction kinetics and population dynamics studies[49,50]. In GRN studies, noise stemming from fluctuations in the expression and interactions of genes and proteins[51–53] has been shown to potentially affect transitions, as noise can distort the underlying epigenetic landscape, making it an essential regulator of cell state[54]. Though cell shape noise has been less explored, it has been proposed to play a functional role in driving chemotactic response in neuronal growth cones[55] and cell sorting during embryonic development[56]. Building on our findings, it will be exciting to investigate to what extent cell shape noise also contributes to other types of cellular shape transitions, and how shape noise is regulated.

The coupling between cell shape and the underlying mechanics, and cell state, has recently been attracting increasing attention[41,57,17]. The mathematical framework we developed here for investigating the stochastic dynamics of cell shape changes could be expanded to include other cellular features, such as the levels and distributions of state markers. Such multi-

dimensional analysis will constitute a powerful investigation tool that could help establishing a more holistic definition of what constitutes a cell state[58].

Finally, we found that modulating the strength of cell shape noise can influence the rate of EMT-associated cell spreading without altering the morphogenetic landscape and, thus, contribute to the robustness and timing of the process. Mechanisms to adjust the timing of cell spreading could be functionally essential during EMT in physiological processes like wound healing and development. Delayed wound closure increases the risk of infections[59]. In development, cell shape and fate changes need to be coordinated in time and space. For instance, in zebrafish gastrulation the timing of gastrulating cell specification is essential for coordinating cell protrusivity and internalization capacity while preserving tissue patterning[60]. In both contexts the exact timing of cell shape change may be key for organism survival. Our work suggests the intriguing possibility that cells could tune the levels of cell shape noise to flexibly adjust the timing of cellular shape transitions.

FIGURES

Figure 1

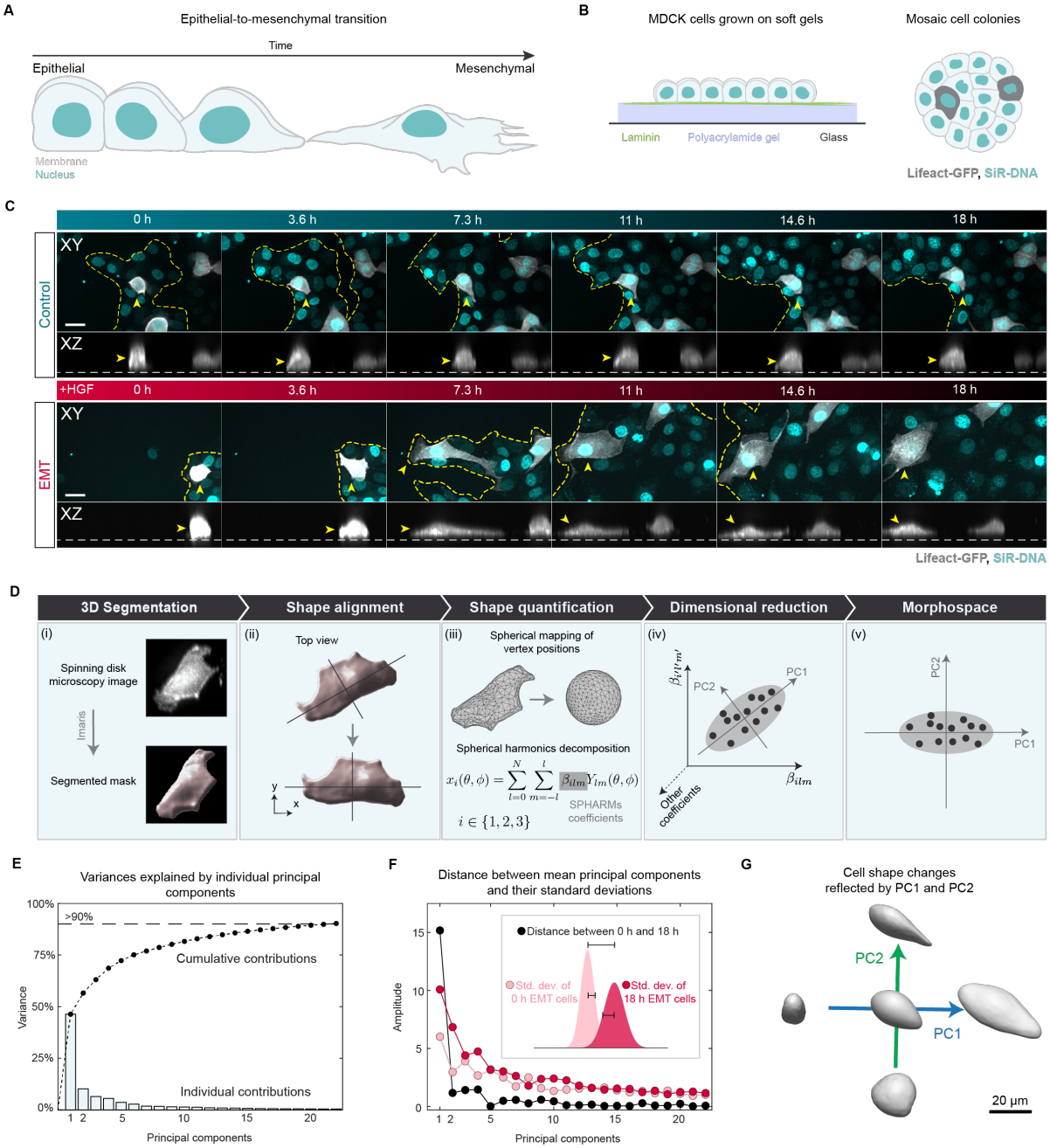

Figure 1: Morphometric analysis of cell shape during EMT

(A) Schematic of EMT-associated cell shape changes.

(B) Schematic of the experimental approach. MDCK cells were cultured on soft polyacrylamide gels coated with laminin to promote cuboidal epithelial morphologies prior

to EMT induction. Mosaic cell labeling was achieved by mixing MDCK cells stably expressing Lifeact-GFP (gray) with unlabeled MDCK cells (light blue). All nuclei were labelled with SiR-DNA (cyan).

(C) Representative time series of control and EMT cells, imaged with spinning disk confocal microscope. XY – top view, XZ – side view in maximum intensity projections. Labels: Lifeact-GFP (F-actin, gray) and SiR-DNA labels (nuclei, cyan). Yellow dashed line: colony outline, arrowhead: representative Lifeact-GFP cell. Scale bar: 20 µm.

(D) Morphometric analysis pipeline for quantitative cell shape description. (i) Segmentation of 3D cell shapes from experimental data. (ii) Alignment of cell shape so that the long axis of all cells is oriented along the x-axis. (iii) Decomposition of cell shape into SPHARMs coefficients β_{ilm}. (iv) Dimensional reduction of high-dimensional data set of SPHARM coefficients β_{ilm} using principal component analysis. (v) Low-dimensional mapping of cell shape in a morphospace. See also Methods for details on the analysis pipeline.

(E) Individual and cumulative percentage of the total variance explained by each principal component.

(F) Comparison of the distance d_i (Equation 1) between the average $\mathrm{PC}i$ at 0 h and at 18 h in EMT cells and of their standard deviations $\sigma_{i,\mathrm{EMT,0h}}$ and $\sigma_{i,\mathrm{EMT,18h}}$ for $i = 1, \ldots, 22$.

(G) Reconstruction of cell shape changes reflected by changes in PC1 and PC2, see Methods for details of the reconstruction.

Figure 2

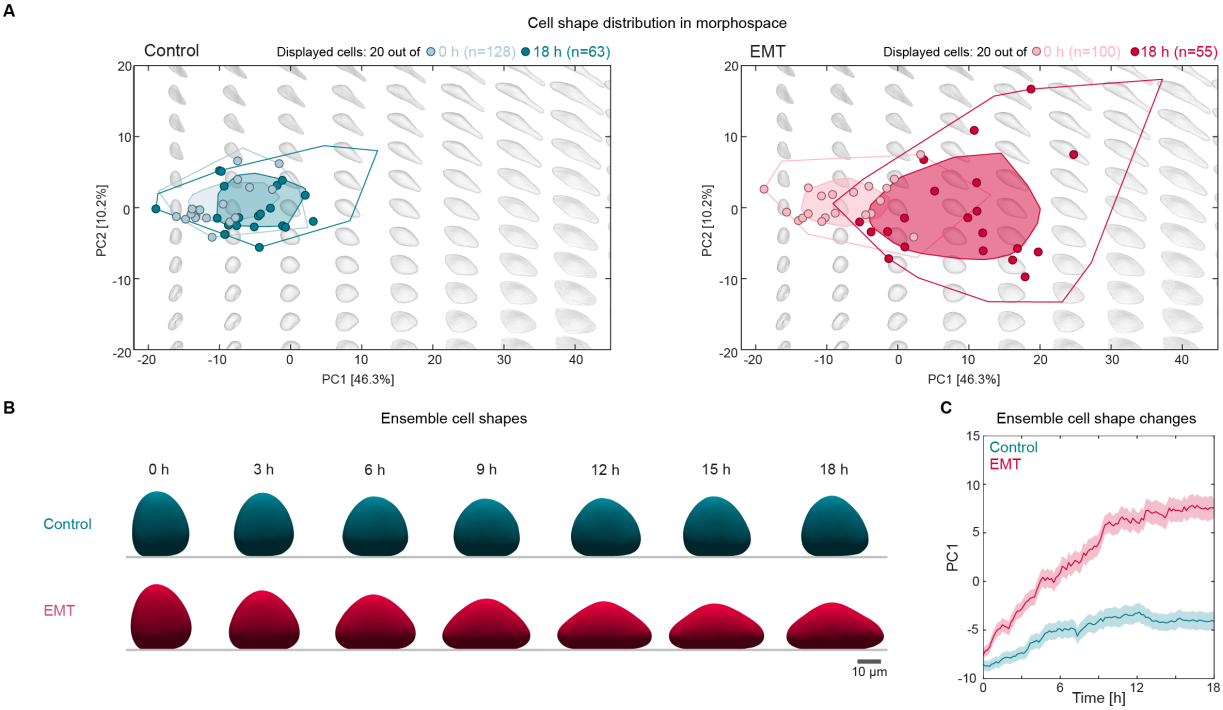

Figure 2: Ensemble cell shape changes during EMT

(A) Distribution of 0 h and 18 h control and EMT cells in the PC1-PC2 morphospace. For each condition, we show the morphospace positions of 20 randomly chosen cells (see Figure S2A and S2B for all datapoints). Background gray shapes: reconstructed theoretical cell shapes for different combinations of PC1 and PC2. Datapoints: individual cell shapes. Polygons: bagplots[61], where the inner polygon (bag) is constructed based on the Tukey depth and contains at most 50% of the points and the outer polygon (fence) corresponds to the convex hull of all datapoints.

(B) Mean shape dynamics of control and EMT cells. The mean shapes were generated by averaging the SPHARMs coefficients $\beta_{ilm}(t)$ of all cells at the given times and reconstructing the shape associated with the average SPHARM coefficient $\langle\beta_{ilm}(t)\rangle$ (Methods).

(C) Ensemble dynamics of PC1 of control (N=8, n=102) and EMT cells (N=12, n=145). Plots: mean ± SEM.

Figure 3

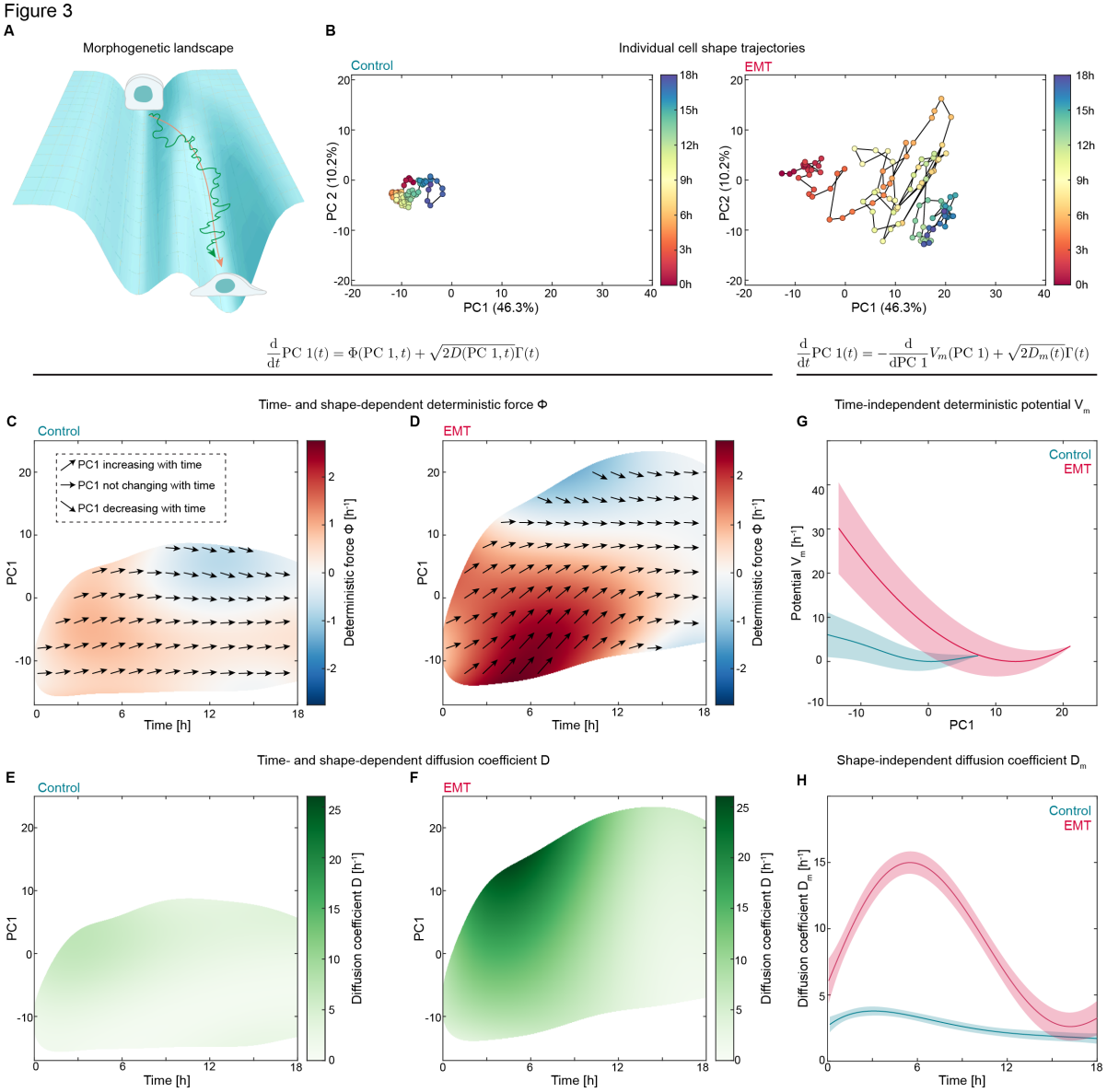

Figure 3: Stochastic inference analysis of morphospace trajectories reveals cell shape attractors and a peak in cell shape noise strength during EMT

(A) Illustration of a morphogenetic landscape. The landscape dictates changes in cell shape, with minima corresponding to local attractors towards which cell shapes converge. The deterministic forces in Equation 2 mediate a persistent trend along the morphogenetic landscape towards attractors (orange trajectory), while contributions due to the stochastic forces in Equation 2 mediate a jittery movement (green trajectory).

(B) Individual cell shape trajectory of a control and an EMT cell mapped into the PC1-PC2

morphospace. See Figure S2I and S2J for more trajectories.

(C, D) Deterministic force field $\Phi(\mathrm{PC1}, t)$ (Equation 2) inferred from morphospace trajectories of control and EMT cells. The arrows indicate how PC1 changes with time.

(E, F) Diffusion coefficient $D(\mathrm{PC1}, t)$ (Eq. 2) inferred from morphospace trajectories of control and EMT cells. Larger values correspond to larger cell shape noise strength.

(G) Potential $V_m(\mathrm{PC1})$ of the minimal stochastic model (Equation 3) inferred from morphospace trajectories of control and EMT cells. A minimum in the potential corresponds to a shape attractor. The curves are the best fit and the errorbars indicate the 95% prediction intervals.

(H) Diffusion coefficient $D_m(t)$ of the minimal stochastic model (Equation 3) inferred from morphospace trajectories of control and EMT cells. The curves are the best fit and the errorbars indicate the 95% prediction intervals.

Figure 4

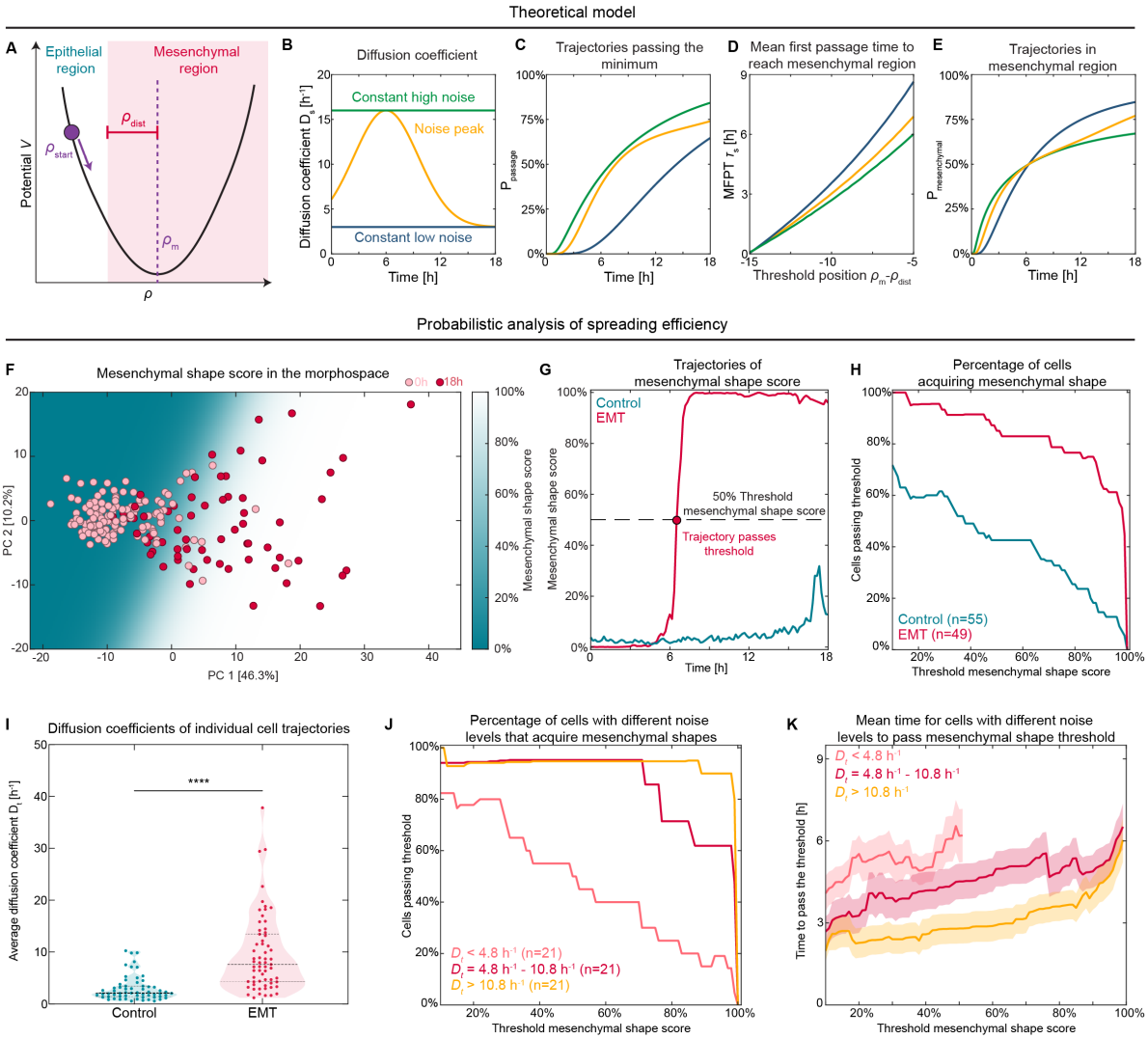

Figure 4: A peak in cell shape noise strength may increase the efficiency of EMT-associated cell spreading

(A) Illustration of stochastic dynamics model for a state descriptor ρ in a monostable potential $V_s = k_s\rho^2/2$, starting at the initial position $\rho_{\text{start}} < 0$; $\rho_{\text{m}} = 0$ is the potential minimum. The threshold $\rho_{\text{dist}} > 0$ separates the epithelial and mesenchymal regions.

(B) Illustration of the three cases of the diffusion coefficient $D_s(t)$ considered in the model (see Equation 5 and Methods).

(C) Percentage of trajectories, P_{passage}, that passed the potential minimum after a time t. (D) Mean first passage time τ_s of trajectories passing the mesenchymal threshold as a function

of the threshold position $\rho_{\mathrm{m}} - \rho_{\mathrm{dist}}$.

(E) Percentage of trajectories, $P_{\mathrm{mesenchymal}}$, that are located within the mesenchymal region after the time t. In (C-E), blue line: “constant low noise” case, green line: “constant high noise”, yellow line: “noise peak” case.

(F) Probability map of mesenchymal shapes in the PC1-PC2 morphospace. Color gradient: mesenchymal shape score of each morphospace position. Pink and red datapoints: 0 h and 18 h EMT cells, respectively (same dataset as in Figure S2A and S2B).

(G) “Mesenchymal shape score” of individual representative control and EMT cells as a function of time.

(H) Percentage of control (N=8, n=55) and EMT (N=12, n=49) cells that pass a given mesenchymal shape score threshold (x-axis) within 18 h.

(I) Distribution of average diffusion coefficients D_t of control (N=8, n=61) and EMT (N=12, n=63) cells, (P<0.0001, Mann-Whitney test), dashed line: median, dotted lines: quartiles.

(J) Percentage of EMT cells with weak cell shape noise strength ($D_t < 4.8\ \mathrm{h}^{-1}$), moderate cell shape noise strength ($4.8\ \mathrm{h}^{-1} \leq D_t \leq 10.8\ \mathrm{h}^{-1}$), and large cell shape noise strength ($D_t > 10.8\ \mathrm{h}^{-1}$) that reaches given mesenchymal shape score thresholds (x-axis) within 12 h. n=21 for each category.

(K) Mean time taken by EMT cells to reach a given mesenchymal shape score threshold (x-axis). Mean values of the times are represented as mean ± SEM.

Figure 5

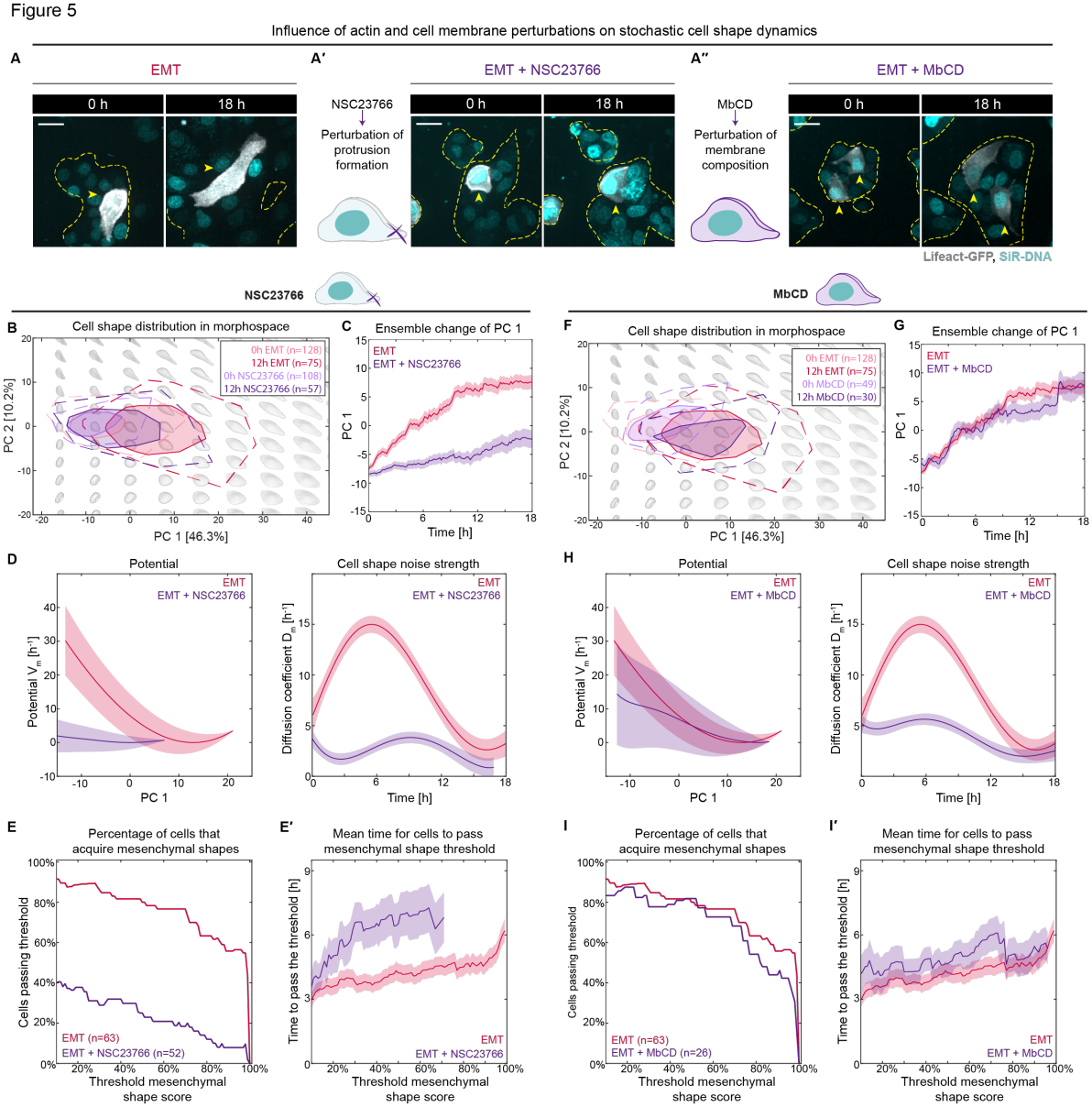

Figure 5: Reducing cell shape noise strength interferes with EMT-associated cell spreading

(A, A', A'') Representative images at 0 h and 18 h after EMT induction of unperturbed EMT cells (A) and cells treated with 100 µM Rac1 inhibitor NSC23766 (A') or 1 mM MbCD (A''). Labels: Lifeact-GFP (F-actin, gray) and SiR-DNA (nuclei, cyan). Yellow dashed lines: colony outlines, arrowheads: representative Lifeact-GFP cells. Scale bars: 20 µm.

(B) Bagplot representation of the distribution of untreated and NSC23766-treated EMT cells in the morphospace at 0 h and 12 h. Background gray shapes: reconstructed theoretical cell

shapes for different combinations of PC1 and PC2.

(C) Ensemble dynamics of PC1 of untreated and NSC23766-treated EMT cells. Mean ± SEM.

(D) Potential $V_m(\text{PC1})$ and diffusion coefficient $D_m(t)$ of untreated and NSC23766-treated EMT cells inferred from their PC1 morphospace trajectories. $V_m(\text{PC1})$ and $D_m(t)$ curves are the best fit and error bars indicate the 95% prediction intervals.

(E) Percentage of untreated or NSC23766-treated EMT cells that acquire mesenchymal morphologies within 12 h as a function of the mesenchymal shape score threshold.

(E') Mean time taken by untreated or NSC23766-treated EMT cells to reach a given mesenchymal shape score threshold. Mean ± SEM.

(F) Bagplot representation of the distribution of untreated and MbCD-treated EMT cells in the morphospace at 0 h and 12 h. Background gray shapes: reconstructed theoretical cell shapes for different combinations of PC1 and PC2.

(G) Ensemble dynamics of PC1 of untreated and MbCD-treated EMT cells. Mean ± SEM.

(H) Potential $V_m(\text{PC1})$ and diffusion coefficient $D_m(t)$ of untreated and MbCD-treated EMT cells inferred from their PC1 morphospace trajectories. $V_m(\text{PC1})$ and $D_m(t)$ curves are the best fit and error bars indicate the 95%

(I) Percentage of untreated or MbCD-treated EMT cells that acquire mesenchymal morphologies within 12 h as a function of the mesenchymal shape score threshold.

(I') Mean time taken by untreated or MbCD-treated EMT cells to reach a given mesenchymal shape score threshold. Mean ± SEM.

AUTHOR CONTRIBUTIONS

E.K.P., G.S., I.Y. and W.P. developed the project idea; I.Y. developed and performed the wet lab experiments; I.Y. and W.P. conducted the spinning disk microscope imaging; I.Y. and W.P. conducted the segmentation; W.P. was responsible for the data analysis and theoretical modelling; E.K.P., G.S., I.Y. and W.P. prepared the manuscript; E.K.P. and G.S. supervised the project, discussed the data and results.

ACKNOWLEDGEMENTS

We would like to thank Margherita Battistara, Kevin Chalut, Joe Howard, Laura Machesky Xiaoyi Ouyang, Ana Raffaelli, Aki Stubb, Marta Urbanska for feedback on the manuscript, Nir Gov and Ruther Hermsen for helpful discussions about the data analysis and theoretical modelling. We thank Lakshmi Balasubramaniam for help with gel protocol optimization. We also thank Windie Höfs and Guillaume Charras for providing the MDCK cell lines and input on cell culture conditions. We thank the Cambridge Advanced Imaging Centre for assistance with microscopy.

We also thank the Paluch lab (in particular Margherita Battistara, Will Foster, Fiona Morgan, Tasmin Sarkany, Belle Sow, Aki Stubb, Baptiste Vauléon, Liz Williams, Alex Winkel) and the Salbreux lab (in particular Silvia Grigolon, Diana Khoromskaia, Quentin Vagne, Max Kerr Winter) for discussions and feedback throughout the project.

E.K.P. acknowledges funding from the European Research Council (Consolidator Grant 820188-NanoMechShape), Medical Research Council UK (MRC Programme Award MC_UU_12018/5), and the Philip Leverhulme Trust (Leverhulme Prize in Biological Sciences

to E.K.P.). W.P. acknowledges financial support from the Herchel Smith Fund (Herchel Smith Postdoctoral Fellowship).

SUPPLEMENTAL INFORMATION

Figures S1-S5

Document S1. Model and theoretical calculations.

Movie S1. Times series of cell shape dynamics in control and EMT cells.

Movie S2. Movie of what shape change PC1 corresponds to.

Movie S3. Movie of what shape change PC2 corresponds to.

Movie S4. Times series of cell shape dynamics in untreated EMT cells and EMT cells with perturbed actin or membrane dynamics.

DATA AVAILABILITY

All data and Matlab scripts are available upon request.

METHODS

Cell culture

MDCK-II cell lines (wild type (WT) and stable lines expressing Lifeact-GFP or H2B-iRFP) were a gift from Professor Guillaume Charras, London Centre for Nanotechnology, United Kingdom. MDCK cells were cultured at 37°C in a humidified atmosphere of 5% CO_2. Cells were passaged every 3-4 days using standard cell culture protocols and disposed of after no more than 30 passages. The culture medium was composed of high-glucose Dulbecco's Modified Eagle Medium, GlutaMAX (Thermo Fisher Scientific, #31966047) supplemented with 10% FBS (Thermo Fisher Scientific, #10500-064) and 1% penicillin-streptomycin

(Thermo Fisher Scientific, #P4333) and 25 mM HEPES (Thermo Fisher Scientific, #15630056). Cells were tested for mycoplasma monthly and free of contamination.

Hydrogel fabrication

Polycrylamide hydrogels were fabricated according to the procedure described by Labouesse et al.[62] with minor changes. Briefly, coverslips and cell culture chambers were silanized with 2% 3-(Trimethoxysilyl)propyl methacrylate (Merck, #M6514) and 1% acetic acid diluted in ethanol and air-dried in an oven at 60°C. Hydrogel solutions were prepared to produce gels of 3.5 kPa stiffness (48 µl 40% Acrylamide (Merck, #A4058) 36 µl 2% Bisacrylamide (Merck, #M1533); 20 µl 2 M 2-(prop-2-enoylamino)hexanoic acid (AK Scientific, #6845AL), 389.5 µl MiliQ water, 1.5 µl TEMED (Merck, #T22500), 5 µl 10% APS (Merck, #A3678) and polymerized between a support coverglass (cell culture chamber or a functionalized coverslip) and a top coverslip for 25-35 minutes. To ensure that cell morphology was not affected by the underlying glass substrate, gel mix quantity was adjusted depending on coverslip diameter to produce gels with thickness above 100 µm. After polymerization, gels were soaked in PBS overnight. The top coverslips were removed on the following day. Hydrogels were then activated by 30 min treatment in 0.2 M EDAC (Merck, #03450) and 0.5 M NHS (Acros Organics, #157272500) in MES buffer, pH 6.1. They were coated with 25 µg/ml Laminin (Sigma-Aldrich, #L2020) diluted in HEPES buffer (50 mM pH 8.5) overnight at 4°C. After washes, gels were blocked with 0.5 M Ethanolamine (Merck, #411000) in HEPES buffer. Gels were stored in basal medium at 4°C for no more than two weeks.

Microscope image acquisition

Cells were seeded at subconfluent levels on 3.5 kPa hydrogels coated with with 25 µg/ml Mouse laminin (Merck, #L2020) in 2-well chambered coverglass (Nunc™ Lab-Tek™ II, #155379 or Ibidi µ-Slide, #80287) allowing control and EMT cells to be compared directly. Unlabeled MDCK cells were mixed with Lifeact-GFP MDCK cells (15:1 ratio) enabling the imaging of cell morphology of single labelled cells located in clusters of unlabeled cells. Nuclei in all cells were labelled using SiR-DNA (Spirochrome, SC007). 24 h after seeding, control samples were left untreated and EMT was induced in the rest by addition of 5 ng/ml HGF (R&D Systems, #294-HG)[34,63,64].

Time-lapse imaging using a spinning disk confocal microscope

Live-cell imaging was performed at 37 °C in CO_2-independent Leibovitz's L-15 medium, no phenol red (Thermo Fisher, # 21083027) supplemented with 10% FBS and 1% penicillin-streptomycin. Samples were grown on hydrogels in 2-well chambered coverglass allowing the control and HGF-treated cells to be imaged in parallel. Time-lapses were acquired using a 30x silicon oil immersion objective (UPLSAPO30XS, Olympus) at spinning disk imaging system with a Perkin Elmer MLS1 base laser engine, CSU-X1 Spinning Disk head, Olympus IX81 stand, equipped with a Hamamatsu Flash 4.0v2 sCMOS camera and controlled by Volocity software. Acquired z-stacks had a thickness of 60 µm and a step size of 0.75 µm, time interval was 10 minutes with a total duration of 10-22 h.

Chemical perturbations

Rac1 inhibitor NSC23766 was dissolved and stored as stock solution (10 mM) in DMSO at -20°C. Methyl-β-cyclodextrin (MbCD) was dissolved in cell culture medium and always prepared immediately before experiments. For multi-position live-cell imaging, imaging positions were identified and added to Volocity, inhibitor solution was added to the imaging medium at the spinning disk confocal microscope and the first 30 minutes of drug pre-treatment were acquired. After that the acquisition was stopped, 5 ng/ml HGF was added to the EMT samples and acquisition was continued for the following 10-22 h. All treatments were started after 24 h after cells were seeded on gels. Final concentrations used for the drug treatments were 100 µM Rac1 inhibitor NSC23766 (Tocris, #2161) and 1mM MbCD (Merck, #C4555).

Image analysis and dimensional reduction

Minimal image processing was used before image analysis consisting of image cropping in Fiji[65].

<u>Cell segmentation and tracking in 3D</u>

Semi-automatic segmentation and tracking were performed on 3D time lapses of single-labelled cells using the Surface tool in Imaris 9.1/10 (Bitplane/Oxford Instruments). The Lifeact-GFP (actin) channel was used for segmentation. The following parameters were used in the Surface Tool: Smoothing surface detail length 1.5 µm, background subtraction sphere diameter 8 µm, autoregressive motion tracking with 0 gap size. The faces and vertices of the triangulated surfaces were exported as *.wrl files.

Shape quantification

Pre-processing

The *.wrl files of the time-dependent shapes of individual cells were imported in MATLAB (versions R2022b-R2024a) with a custom script. At each time point, only the largest connected mesh was kept and hence, smaller connected meshes were removed, e.g. due to segmentation artefacts resulting from regions within the cell with low signal intensity.

Standard shape features

We computed standard shape features using a custom script in MATLAB (versions R2022b-R2024a). To calculate the volumetric center of mass $\vec{r}_{COM} = \{x_{\mathrm{COM}}, y_{\mathrm{COM}}, z_{\mathrm{COM}}\}$ of a cell, we used Gauss's theorem. Cell shape features such as principal axes lengths and directions, cell volume and cell surface area were calculated in a similar manner (Supplemental Information).

Spherical harmonics decomposition

We carried out the spherical harmonics decomposition in MATLAB (versions R2022b-R2024a). First, cells were aligned by calculating the eigenvectors and eigenvalues of the covariance matrix of the surface given by the

$$CV_{ij} = \begin{pmatrix} \langle \bar{x}^2 \rangle & \langle \bar{x}\bar{y} \rangle & 0 \\ \langle \bar{x}\bar{y} \rangle & \langle \bar{y}^2 \rangle & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \int \mathrm{d}S \begin{pmatrix} 0 \\ \bar{x}^2\bar{y} \\ 0 \end{pmatrix} \cdot \vec{n} & \int \mathrm{d}S \begin{pmatrix} 0 \\ 0 \\ \bar{x}\bar{y}\bar{z} \end{pmatrix} \cdot \vec{n} & 0 \\ \int \mathrm{d}S \begin{pmatrix} 0 \\ 0 \\ \bar{x}\bar{y}\bar{z} \end{pmatrix} \cdot \vec{n} & \int \mathrm{d}S \begin{pmatrix} \bar{x}\bar{y}^2 \\ 0 \\ 0 \end{pmatrix} \cdot \vec{n} & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \frac{1}{\int \mathrm{d}S \, \frac{1}{3} \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix} \cdot \vec{n}} \qquad (6)$$

with the centered cell surface points given by $\bar{x} = x - x_{\mathrm{COM}}$, $\bar{y} = y - y_{\mathrm{COM}}$ and $\bar{z} = z - z_{\mathrm{COM}}$. To numerically compute the covariance matrix of a triangulated cell surface mesh, we used Gauss's theorem (Supplemental Information). The z-direction was chosen perpendicular to the substrate and only the directions in the plane parallel to the substrate are considered, hence we set $CV_{i3} = 0$ and $CV_{3i} = 0$. Cells were rotated around the z-axis so that the long axis of the cells (corresponding to the eigenvector with the largest eigenvalue) was parallel to the x-direction. The initial direction of the cell alignment along the x-axis was chosen such that the third moment $\langle \bar{x}^3 \rangle$ of the mesh is positive. For later time points, the direction of alignment was chosen such that it was pointing in a similar direction as in the previous time point by ensuring that the scalar product of the eigenvectors at the current and previous time point is positive.

After alignment, the MATLAB function *reducepatch()* was used to reduce the number of faces to 1000 per cell, significantly speeding up the following analysis without sacrificing global shape information. For the spherical harmonics decomposition of triangulated shapes, we adapted tools previously developed to study brain structure morphologies and implemented in MATLAB[66,67] [https://brainimaging.waisman.wisc.edu/~chung/midus/]. First, the surface mesh was mapped onto a unit sphere utilizing an area-preserving surface flattening algorithm[66] [https://brainimaging.waisman.wisc.edu/~chung/midus/]. This algorithm maps the vertices of the original cell mesh to the surface of a sphere with the aim

of having the areas of the faces connecting groups of three vertices being conserved between the original mesh and the spherical mesh. While the algorithm allows for changes in the triangle shape (as long as the area is conserved), the spherical mesh has the same set of edges (connections between two neighboring vertices) as the original mesh, and hence conserves local information of the underlying cell shape. This ensures that similar shapes with different meshes will have similar mappings on the spherical surface. The new vertex positions on the sphere were then rescaled such that they are located on the surface of the unit sphere. The algorithm maps the positions $x(\theta,\phi)$, $y(\theta,\phi)$ and $z(\theta,\phi)$ of the vertices on the unit sphere surface with the polar angle ϕ and the azimuthal angle θ. These distributions on the surface of the unit sphere were then decomposed into spherical harmonics functions (SPHARMs)[67] [https://brainimaging.waisman.wisc.edu/~chung/midus/] so that

$$x_i(\theta,\phi) = \sum_{l=0}^{N} \sum_{m=-l}^{l} \beta_{ilm} Y_{lm}(\theta,\phi), \tag{7}$$

with the spherical harmonics functions $Y_{lm}(\theta,\phi)$ and the spherical harmonics coefficients $\beta_{1,lm}$, $\beta_{2,lm}$ and $\beta_{3,lm}$ which are unique descriptors of cell shape. We also introduced the coordinates x_1, x_2, x_3 which correspond to x, y, z. Here, N is the highest degree of spherical harmonics functions considered. The SPHARMs framework also permits us to reconstruct shapes from the coefficients β_{ilm} with $i \in \{1,2,3\}$ (Figure S1A).

To verify the quality of the spherical harmonics decomposition, we compared the distances between the triangulated mesh of the original surface and the mesh reconstructed after the spherical harmonics decomposition. For each vertex of an original mesh/reconstructed mesh, we computed the distance from the reconstructed triangulated mesh/original mesh, using the function *point2trimesh()* [Daniel Frisch (2024). point2trimesh( ) —

Distance Between Point and Triangulated Surface (https://www.mathworks.com/matlabcentral/fileexchange/52882-point2trimesh-distance-between-point-and-triangulated-surface), MATLAB Central File Exchange.]. We then calculated the maximum distance between a vertex and in the following, only considered those cell surfaces where the maximum error was below 2 µm. Moreover, we calculated the mean distances of all vertices of the original and reconstructed meshes and only consider cell surfaces that have a mean error below 0.15 µm.

To identify the minimal degree N of spherical harmonic functions (corresponding to $3(N+1)^2$ SPHARMs coefficients, Figure S1B) that represent the shape in the appropriate manner, we performed a spherical harmonics decomposition for values of $N = 1 \ldots 43$ of all cells (without drug treatments) after 0 h and 18 h. For all cells we calculated the maximum error and mean error and remove those cells that have too large errors (with the criteria described above) for the highest degree $N = 43$, for which we would usually expect to find a minimal error. For the remaining cells, we calculate the largest mean and maximum error as a function of N of all cells. We chose the decomposition degree N for which the largest mean error of any cell is below 0.15 µm and the largest maximum error of any cell is below 2 µm [Figure S1C and S1D]. We found that this is fulfilled for $N = 23$.

To reconstruct the average shapes of cells as a function of time, we calculated $\langle \beta_{j,lm} \rangle_{\text{cells}}$, and reconstructed the resulting cell shape.

<u>Dimensionality reduction</u>

For dimensional reduction of the high-dimensional shape features β_{ilm}, we utilised principal component analysis. To define the principal components, we chose the shapes features of

HGF-treated cells (without any additional drug treatment) at 0 h and 18 h. We computed their averaged cell shapes

$$\langle \beta_{ilm} \rangle = \frac{\langle \beta_{ilm}(0h) \rangle_{\text{cells}} + \langle \beta_{ilm}(18h) \rangle_{\text{cells}}}{2}, \tag{8}$$

and then defined the centered shape features of the cells as

$$\tilde{\beta}_{ilm}(t) = \beta_{ilm}(t) - \langle \beta_{ilm} \rangle. \tag{9}$$

We then used the MATLAB function *pca()* for the principal component analysis (PCA) of the centered shape features with the observation weights $\frac{1}{N_{0h}}$ for the cells at 0 h (with the number N_{0h} of cells used for the PCA at 0 h) and the weights $\frac{1}{N_{18h}}$ for the cells at 18 h (with the number N_{18h} of cells used for the PCA at 18 h). The observation weights compensated that at 0 h we have 128 cells that contribute to the PCA and at 18 h we have 63 cells that contributed to the PCA. This provided us with a mapping of the high-dimensional cell shape features into a low-dimensional morphospace. In this study, we take the principal components (PCs) to be unitless, such that principal components score has units of [1/µm]. We note that one could instead take the principal component scores of the PCA to be unitless, in which case the principal components (PCs) would be given in units of [µm]. Combining the principal component coefficients derived by the PCA and the shape features $\tilde{\beta}_{ilm}(t)$ of all cells, we calculated the principal component scores (called $\mathrm{PC}i$ with $i = 1,2, \ldots$) of each individual cell, corresponding to the coordinates in the low-dimensional morphospace. Using $\langle \beta_{ilm} \rangle$ as a reference, we reconstructed cell shapes for arbitrary values of the principal component scores, allowing us to reconstruct the cell morphology corresponding to any location in the morphospace. For this, we used MATLAB functions provided by previous studies[66,67] [https://brainimaging.waisman.wisc.edu/~chung/midus/].

To generate bagplots of the cell distribution in the morphospace (Figure 2A, 5B, 5F, S2C, S2D, S5D and S5E), we employed the LIBRA toolbox[61].

Stochastic inference and modelling

Numerical inference of deterministic drift and diffusion coefficient

Our stochastic inference approach was inspired by previous studies[38]. We assumed that the cell trajectories of the variable PC1 followed overdamped Langevin dynamics

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{PC1} = \Phi(\mathrm{PC1}, t) + \sqrt{2D(\mathrm{PC1}, t)}\Gamma(\mathrm{t}), \tag{10}$$

with a deterministic force field $\Phi(\rho, t)$, the diffusion coefficient $D(\mathrm{PC1}, t)$ and Gaussian white noise $\Gamma(\mathrm{t})$. Both $\Phi(\mathrm{PC1}, t)$ and $D(\mathrm{PC1}, t)$ can depend on PC1 and time t.

We derived the deterministic force field $\Phi(\mathrm{PC1}t)$ from the mean displacement

$$\Phi(\mathrm{PC1}, t) = \frac{\langle \mathrm{PC1}(t + \Delta t) - \mathrm{PC1}(t) \rangle}{\Delta t} \tag{11}$$

and the diffusion coefficient from the mean square displacement

$$D(\mathrm{PC1},t) = \frac{\langle \left(\mathrm{PC1}(t + \Delta t) - \mathrm{PC1}(t)\right)^2 \rangle}{2\Delta t} \tag{12}$$

(see Supplemental Information). To numerically infer the deterministic force $\Phi(\mathrm{PC1},t)$, we considered all relevant trajectories (distinguishing HGF and drug treatment, times between 0.5h before EMT induction until 23 h50 min after EMT induction) and considered the ensemble of all their possible displacements $(\mathrm{PC1}(t + \Delta t) - \mathrm{PC1}(t))/\Delta t$ for the time step $\Delta t = 10\ \mathrm{min}$ and fit polynomial surfaces up to the 5th order depending on time t and shape $\mathrm{PC1}(t)$ with the MATLAB function *fit()*. We weighted displacements depending on their time, to make sure that early time points, where we usually have more data, did not

overcontribute to the fitting. To identify the best polynomial fit, we computed the degree-of-freedom adjusted coefficient of determination R^2 for each fit (ranging from 0th order to 5th order) and chose the fit with the highest R^2. The same computation was done to derive the diffusion coefficient $D(\mathrm{PC1}, t)$ from the mean squared displacements of the trajectories $\left(\mathrm{PC1}(t+\Delta t) - \mathrm{PC1}(t)\right)^2 / 2\Delta t$. The fitting parameters used in Figures 3C, 3D, 3E and 3F are provided in the Supplemental Information.

When inferring the time-independent behavior of the deterministic force $\Phi(\mathrm{PC1})$, we only considered polynomials up to the 5th order that depended on PC1. When inferring the shape-independent behavior of the diffusion coefficient $D(\mathrm{t})$, we only considered polynomials up to the 5th order depending on time t.

For the time-independent deterministic force $\Phi(\mathrm{PC1})$, we derived the corresponding potential using

$$V(\mathrm{PC1}) = -\int \mathrm{dPC1}\ \Phi(\mathrm{PC1})\,. \tag{13}$$

For the deterministic force $\Phi(\mathrm{PC1}, t)$ and the diffusion coefficient $D(\mathrm{PC1}, t)$, we calculated the 95% prediction intervals (using the MATLAB function *predint()* with non-simultaneous bounds for the fitted curve) as a function on PC1 and t and estimated the fitting error $e(\mathrm{PC1}, t)$ as difference between the upper and lower prediction bounds. We only considered those regions with

$$e(\mathrm{PC1}, t) \leq 2\min[e(\mathrm{PC1}, t)]. \tag{14}$$

Changes in PC1 due to deterministic force Φ will translate into changes of the mean squared displacement and hence, can affect the inferred diffusion coefficient. The resulting contribution is given by (Supplemental Information)

$$D_{\Phi}(\mathrm{PC1}, t) = \frac{\Phi^2(\mathrm{PC1}, t)\Delta t}{2} \tag{15}$$

and we found that it has a negligible effect on the inferred diffusion coefficient (Figures S4 and S4B).

Numerical generation of in silico trajectories of cell shape changes during EMT

To numerically generate stochastic trajectories following Langevin dynamics with a given force field $\Phi(\mathrm{PC1}, t)$ and diffusion coefficient $D(\mathrm{PC1}, t)$ as input, we used the numerical scheme[37]

$$\mathrm{PC1}(t + \Delta t) = \mathrm{PC1}(t) + \Phi(\mathrm{PC1}, t)\Delta t + \sqrt{2\Delta t D(\mathrm{PC1}, t)}w(\mathrm{t}) \qquad (16)$$

with the starting position $\mathrm{PC1}(0)$ and a Gaussian-distributed random variable $w(t)$ with zero mean and variance 1. We randomly sampled the starting positions of the experimental trajectories as starting position of the simulations and simulated 100 trajectories for each condition with a time step of 0.01 h and saved trajectories of the ρ-coordinate with a time interval of 0.17 h.

Verification of stochastic inference and *in silico* simulations

To compare numerically generated trajectories with our experimentally-derived trajectories, for both we simulated several in-silico trajectories (Figures S3A and S3B) and inferred the deterministic force field and diffusion coefficient from the in silico trajectories. We found good qualitative agreement with the fields inferred from the experimental data (see Figures S3C and S3D). For the inference, we ignored trajectories with shapes $\mathrm{PC1} < -30$ at 18 h, as these trajectories corresponded to physically invalid cell shapes and were a result from fitting errors in regions with low densities or no cell shapes during the stochastic inference.

Moreover, we compared the ensemble behavior for several trajectories $\langle \mathrm{PC1}(t) \rangle$ (Figure S3E, S3F) and found good qualitative agreement between the experiment and the simulations.

Total diffusion coefficient

To calculate the overall shape noise strength of not just $\mathrm{PC1}$, but the first 22 principal components combined (corresponding to 90% of the total variance of the data, Figure S3G), we defined the high-dimensional morphospace trajectory

$$\overrightarrow{\mathrm{PC}\imath}(t) = \{\mathrm{PC1}(t), \mathrm{PC2}(t), \dots, \mathrm{PC22}(t)\} \tag{17}$$

and inferred the diffusion coefficient by fitting polynomial surfaces to the mean squared displacement

$$D_{\mathrm{all}} = \frac{\left\langle \left(\overrightarrow{\mathrm{PC}\imath}(t + \Delta t) - \overrightarrow{\mathrm{PC}\imath}(t) \right)^2 \right\rangle}{2 \cdot 22\, \Delta t} \tag{18}$$

with $\Delta t = 10\ \mathrm{min}$.

Time to reach attractor minimum in the simplified model

We considered a stochastic process expressed by the Langevin equation[36] of ρ (Figure 4A), which is written as

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\frac{\mathrm{d}V_s}{\mathrm{d}\rho} + \sqrt{2D_s(t)}\Gamma(t), \tag{19}$$

with the initial position $\rho(0) = \rho_{\mathrm{start}} < 0$, the quadratic potential $V_s = k_s \rho^2/2$ (corresponding to a potential minimum $\rho_m = 0$) and the time-dependent diffusion coefficient (Figure 4B) which is defined as

$$D_s(t) = D_{\mathrm{b}} + \Delta D \exp\left[-\left(\frac{t - t_0}{\tau} \right)^2 \right]. \tag{20}$$

For the diffusion coefficient, we introduced the background noise strength $D_\mathrm{b} > 0$, the peak height $\Delta D \geq 0$, the peak time t_D and a time τ which describes the characteristic time of the peak. The probability of passing the potential minimum located at $\rho = 0$ at least once after time t was then given by

$$P_\mathrm{passage}(t) = 1 - \mathrm{erf}\left(-\frac{\rho_\mathrm{start}}{\sqrt{2\sigma_s^2(t)}}\exp(-k_s t)\right). \tag{21}$$

(Figure 4C; Supplemental Information). Here, we introduced the variance

$$\sigma_s^2(t) = \Delta D\tau \exp\big(k_s(-2t + 2t_0 + \tau^2 k_s)\big)\sqrt{\pi}\left[\mathrm{erf}\left(\frac{t - t_0}{\tau} - \tau k_s\right) + \mathrm{erf}\left(\frac{t_0}{\tau} + \tau k_s\right)\right] + \frac{2D_\mathrm{b}}{k_s}\exp(-k_s t)\sinh(k_s t)\,. \tag{22}$$

with the error function $\mathrm{erf}(x)$. The mean first passage time (MFPT) to reach the potential minimum then follows from

$$\tau_\mathrm{s} = \int_0^\infty \mathrm{d}t'\big[1 - P_\mathrm{passage}(t')\big]\,. \tag{23}$$

To calculate the mean first passage time to first time reach a region defined by $\rho \in [-\rho_\mathrm{dist}, \infty]$ (Figure 4A), we numerically generated trajectories with the starting location ρ_start and computed the time when the position of the simulated trajectories is $\rho \geq -\rho_\mathrm{dist}$ as a function of the boundary position $-\rho_\mathrm{dist}$ (Figure 4E). To compute the percentage of trajectories being in a region defined by $\rho \in [-\rho_\mathrm{dist}, \infty]$ (Figure 4A) at time t is given by

$$P_\mathrm{mesenchymal}(t) = \int_{-\rho_\mathrm{dist}}^\infty p(\rho, t)\,\mathrm{d}\rho = 1 - \frac{1}{2}\left[1 + \mathrm{erf}\left(-\frac{\rho_\mathrm{dist} + \exp(-k_s t)\,\rho_\mathrm{start}}{\sqrt{2\sigma_s^2(t)}}\right)\right]. \tag{23}$$

(Figure 4D and Supplemental Information), In Figure 4, we used the following parameters for the noise peak which correspond to our experimental observations: $k_s = 0.1\ \mathrm{h}^{-1}$, $\tau = 5\ \mathrm{h}$ and $t_0 = 6\ \mathrm{h}$. We consider three case: constant low noise ($D_\mathrm{b} = 3\ \mathrm{h}^{-1}$, $\Delta D = 0$), constant high noise ($D_\mathrm{b} = 16\ \mathrm{h}^{-1}$, $\Delta D = 0$) and noise peak ($D_\mathrm{b} = 3\ \mathrm{h}^{-1}$, $\Delta D = 13\ \mathrm{h}^{-1}$). Moreover

in Figure 4C and 4E we used the initial position $\rho_{\mathrm{start}} = -15$ and to define the mesenchymal region, we used $\rho_{\mathrm{dist}} = 8$ and tested that the presented results are robust when considering other combinations of ρ_{start} and ρ_{dist}. In Figure 4D, we used different values of ρ_{start} and investigated the role of the noise peak height ΔD for $D_{\mathrm{b}} = 3\ \mathrm{h}^{-1}$. In Figure 4E, we used the time step $\Delta t = 0.002\ \mathrm{h}$ and simulated 500000 trajectories for each condition. We employed the numerical scheme provided in Equation 16.

Estimation of average diffusion coefficient

We estimated the average diffusion coefficient (Figure 4I) of a single cell trajectory of PC 1 of the form $\{\mathrm{PC1}(0), \mathrm{PC1}(\Delta t), \mathrm{PC1}(2\Delta t), \mathrm{PC1}(3\Delta t), \mathrm{PC1}(4\Delta t), \dots\}$ with the time step $\Delta t = 10\ \mathrm{min}$ by calculating the mean squared displacement

$$D_t = \frac{\langle [\mathrm{PC1}(t + \Delta t) - \mathrm{PC1}(t)]^2 \rangle_{t=0\dots 12\mathrm{h}}}{2\Delta t}. \tag{24}$$

Here, we only considered the first $12\ \mathrm{h}$ of the trajectories after induction of EMT via HGF. We choose this time interval since the most significant cell shape changes of EMT are happening in the first $12\ \mathrm{h}$ and the cell shape noise strength is significantly reduced for later time points (see Figure 2I).

To define classes of weak, medium and strong average diffusion of trajectories, we considered the distribution of D_t of all cells treated with HGF (and no drug treatment) and considered the 33% and 67% percentiles as the thresholds of the classes.

First passage times of reaching mesenchymal shapes in the experiment

Support vector machine classifier

We trained a classifier for the identification of epithelial and mesenchymal shape classes with the PC1 and PC2 values of HGF treated cells at 0 h and 18 h. We assumed that at 0 h, all cells had epithelial shapes. Moreover, we assume that 18 h of HGF treatment were sufficient for all cells to undergo EMT and hence, all cells at 18 h were treated as having mesenchymal shapes. We then trained a support vector machine classifier in MATLAB with the function *fitcsvm()* with a first order polynomial kernel. For the training, we standardized the predictor data.

After training the classifier, we used the *predict()* function in MATLAB to estimate the mesenchymal probability of pairs of PC1 and PC2. We used this function when generating the probability map in Figure 4F and when generating trajectories of mesenchymal shape probability of individual cells (e.g., Figure 4G). This probability was called mesenchymal shape score.

Estimation of the mean first passage time

We considered only full 0 h-18 h trajectories (Figure 4H) and 0 h-12 h trajectories (Figure 4J, 4K, 5E and 5I) of the mesenchymal shape score of individual cells.

To estimate the probability of cells to reach mesenchymal shapes (Figures 4H, 4J, 5E and 5I), we defined a mesenchymal shape score threshold (>10% and <100%, Figure 4H) and only considered those trajectories that started with a score below the threshold (assumed to be cells with epithelial shapes). We then estimated the percentage of cells that passed this

threshold. We only presented the probabilities if we considered at least 10 trajectories for a given threshold.

To estimate the first passage time (Figure 4K, 5E and 5I), we considered all trajectories that started with epithelial shapes and passed the threshold and then computed their mean value and standard error (95%).

REFERENCES

1. Fujita, Y. (2011). Interface between normal and transformed epithelial cells: A road to a novel type of cancer prevention and treatment. Cancer Science *102*, 1749–1755. https://doi.org/10.1111/j.1349-7006.2011.02011.x.
2. Heisenberg, C.-P., and Bellaïche, Y. (2013). Forces in Tissue Morphogenesis and Patterning. Cell *153*, 948–962. https://doi.org/10.1016/j.cell.2013.05.008.
3. Bodor, D.L., Pönisch, W., Endres, R.G., and Paluch, E.K. (2020). Of Cell Shapes and Motion: The Physical Basis of Animal Cell Migration. Developmental Cell *52*, 550–562. https://doi.org/10.1016/j.devcel.2020.02.013.
4. Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nat Rev Mol Cell Biol *24*, 695–713. https://doi.org/10.1038/s41580-023-00615-w.
5. Huang, S. (2012). The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology? BioEssays *34*, 149–157. https://doi.org/10.1002/bies.201100031.
6. Moris, N., Pina, C., and Arias, A.M. (2016). Transition states and cell fate decisions in epigenetic landscapes. Nat Rev Genet *17*, 693–703. https://doi.org/10.1038/nrg.2016.98.
7. Teschendorff, A.E., and Feinberg, A.P. (2021). Statistical mechanics meets single-cell biology. Nat Rev Genet *22*, 459–476. https://doi.org/10.1038/s41576-021-00341-z.
8. Fletcher, D.A., and Mullins, R.D. (2010). Cell mechanics and the cytoskeleton. Nature *463*, 485–492. https://doi.org/10.1038/nature08908.
9. Clark, A.G., Wartlick, O., Salbreux, G., and Paluch, E.K. (2014). Stresses at the cell surface during animal cell morphogenesis. Current biology : CB *24*, R484-94. https://doi.org/10.1016/j.cub.2014.03.059.
10. McCreery, K.P., Stubb, A., Stephens, R., Fursova, N.A., Cook, A., Kruse, K., Michelbach, A., Biggs, L.C., Keikhosravi, A., Nykanen, S., et al. (2024). Mechano-osmotic signals control chromatin state and fate transitions in pluripotent stem cells. Preprint, https://doi.org/10.1101/2024.09.07.611779 https://doi.org/10.1101/2024.09.07.611779.
11. Chamaraux, F., Fache, S., Bruckert, F., and Fourcade, B. (2005). Kinetics of Cell Spreading. Phys. Rev. Lett. *94*, 158102. https://doi.org/10.1103/PhysRevLett.94.158102.
12. Cuvelier, D., Théry, M., Chu, Y.-S., Dufour, S., Thiéry, J.-P., Bornens, M., Nassoy, P., and Mahadevan, L. (2007). The Universal Dynamics of Cell Spreading. Current Biology *17*, 694–699. https://doi.org/10.1016/j.cub.2007.02.058.
13. Bryce, N.S., Failes, T.W., Stehn, J.R., Baker, K., Zahler, S., Arzhaeva, Y., Bischof, L., Lyons, C., Dedova, I., Arndt, G.M., et al. (2019). High-Content Imaging of Unbiased Chemical Perturbations Reveals that the Phenotypic Plasticity of the Actin Cytoskeleton Is Constrained. Cell Systems *9*, 496-507.e5. https://doi.org/10.1016/j.cels.2019.09.002.
14. Chang, A.Y., and Marshall, W.F. (2019). Dynamics of living cells in a cytomorphological state space. Proc. Natl. Acad. Sci. U.S.A. *116*, 21556–21562. https://doi.org/10.1073/pnas.1902849116.
15. Andrews, T.G.R., Pönisch, W., Paluch, E.K., Steventon, B.J., and Benito-Gutierrez, E. (2021). Single-cell morphometrics reveals ancestral principles of notochord development. Development *148*, dev199430. https://doi.org/10.1242/dev.199430.
16. Segal, D., Mazloom-Farsibaf, H., Chang, B.-J., Roudot, P., Rajendran, D., Daetwyler, S., Fiolka, R., Warren, M., Amatruda, J.F., and Danuser, G. (2022). In vivo 3D profiling of site-

specific human cancer cell morphotypes in zebrafish. Journal of Cell Biology *221*, e202109100. https://doi.org/10.1083/jcb.202109100.
17. Jain, A., Gut, G., Sanchis-Calleja, F., Okamoto, R., Streib, S., He, Z., Zenk, F., Santel, M., Seimiya, M., Holtackers, R., et al. (2023). Morphodynamics of human early brain organoid development. Preprint, https://doi.org/10.1101/2023.08.21.553827 https://doi.org/10.1101/2023.08.21.553827.
18. Dent, L.G., Curry, N., Sparks, H., Bousgouni, V., Maioli, V., Kumar, S., Munro, I., Butera, F., Jones, I., Arias-Garcia, M., et al. (2024). Environmentally dependent and independent control of 3D cell shape. Cell Reports, 114016. https://doi.org/10.1016/j.celrep.2024.114016.
19. Ruan, X., and Murphy, R.F. (2019). Evaluation of methods for generative modeling of cell and nuclear shape. Bioinformatics *35*, 2475–2485. https://doi.org/10.1093/bioinformatics/bty983.
20. Medyukhina, A., Blickensdorf, M., Cseresnyés, Z., Ruef, N., Stein, J.V., and Figge, M.T. (2020). Dynamic spherical harmonics approach for shape classification of migrating cells. Sci Rep *10*, 6072. https://doi.org/10.1038/s41598-020-62997-7.
21. Aspalter, I.M., Pönisch, W., Chalut, K.J., and Paluch, E.K. (2020). Mouse embryonic stem cells switch migratory behaviour during early differentiation. Preprint, https://doi.org/10.1101/2020.12.07.415307 https://doi.org/10.1101/2020.12.07.415307.
22. Dalmasso, G., Musy, M., Niksic, M., Robert-Moreno, A., Badía-Careaga, C., Sanz-Ezquerro, J.J., and Sharpe, J. (2022). 4D reconstruction of murine developmental trajectories using spherical harmonics. Developmental Cell *57*, 2140-2150.e5. https://doi.org/10.1016/j.devcel.2022.08.005.
23. Cavanagh, H., Kempe, D., Mazalo, J.K., Biro, M., and Endres, R.G. (2022). T cell morphodynamics reveal periodic shape oscillations in three-dimensional migration. J. R. Soc. Interface. *19*, 20220081. https://doi.org/10.1098/rsif.2022.0081.
24. Viana, M.P., Chen, J., Knijnenburg, T.A., Vasan, R., Yan, C., Arakaki, J.E., Bailey, M., Berry, B., Borensztejn, A., Brown, E.M., et al. (2023). Integrated intracellular organization and its variations in human iPS cells. Nature *613*, 345–354. https://doi.org/10.1038/s41586-022-05563-7.
25. Mazloom-Farsibaf, H., Zou, Q., Hsieh, R., Danuser, G., and Driscoll, M.K. (2023). Cellular harmonics for the morphology-invariant analysis of molecular organization at the cell surface. Nat Comput Sci *3*, 777–788. https://doi.org/10.1038/s43588-023-00512-4.
26. Hewitt, M.N., Cruz, I.A., Linbo, T.H., and Raible, D.W. (2024). Spherical harmonics analysis reveals cell shape-fate relationships in zebrafish lateral line neuromasts. Development *151*, dev202251. https://doi.org/10.1242/dev.202251.
27. Sailem, H., Bousgouni, V., Cooper, S., and Bakal, C. (2014). Cross-talk between Rho and Rac GTPases drives deterministic exploration of cellular shape space and morphological heterogeneity. Open Biol. *4*, 130132. https://doi.org/10.1098/rsob.130132.
28. Wang, W., Poe, D., Yang, Y., Hyatt, T., and Xing, J. (2022). Epithelial-to-mesenchymal transition proceeds through directional destabilization of multidimensional attractor. eLife *11*, e74866. https://doi.org/10.7554/eLife.74866.
29. Copperman, J., Gross, S.M., Chang, Y.H., Heiser, L.M., and Zuckerman, D.M. (2023). Morphodynamical cell state description via live-cell imaging trajectory embedding. Commun Biol *6*, 484. https://doi.org/10.1038/s42003-023-04837-8.
30. Nieto, M.A., Huang, R.Y.-J., Jackson, R.A., and Thiery, J.P. (2016). EMT: 2016. Cell *166*, 21–45. https://doi.org/10.1016/j.cell.2016.06.028.

31. Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G., et al. (2020). Guidelines and definitions for research on epithelial–mesenchymal transition. Nat Rev Mol Cell Biol *21*, 341–352. https://doi.org/10.1038/s41580-020-0237-9.
32. Leggett, S.E., Sim, J.Y., Rubins, J.E., Neronha, Z.J., Williams, E.K., and Wong, I.Y. (2016). Morphological single cell profiling of the epithelial–mesenchymal transition. Integr. Biol. *8*, 1133–1144. https://doi.org/10.1039/C6IB00139D.
33. Ridley, A.J., Comoglio, P.M., and Hall, A. (1995). Regulation of scatter factor/hepatocyte growth factor responses by Ras, Rac, and Rho in MDCK cells. Mol Cell Biol *15*, 1110–1122. https://doi.org/10.1128/MCB.15.2.1110.
34. Grotegut, S., von Schweinitz, D., Christofori, G., and Lehembre, F. (2006). Hepatocyte growth factor induces cell scattering through MAPK/Egr-1-mediated upregulation of Snail. EMBO J *25*, 3534–3545. https://doi.org/10.1038/sj.emboj.7601213.
35. Kaliman, S., Hubert, M., Wollnik, C., Nuić, L., Vurnek, D., Gehrer, S., Lovrić, J., Dudziak, D., Rehfeldt, F., and Smith, A.-S. (2021). Mechanical Regulation of Epithelial Tissue Homeostasis. Phys. Rev. X *11*, 031029. https://doi.org/10.1103/PhysRevX.11.031029.
36. Kampen, N.G. van (2007). Stochastic Processes in Physics and Chemistry Third Edition. N. G. van Kampen, ed. (Elsevier) https://doi.org/10.1016/B978-0-444-52965-7.X5000-4.
37. Risken, H. (1996). The Fokker-Planck equation: methods of solution and applications Second edition, 3rd printing. (Springer).
38. Frishman, A., and Ronceray, P. (2020). Learning Force Fields from Stochastic Trajectories. Phys. Rev. X *10*, 021009. https://doi.org/10.1103/PhysRevX.10.021009.
39. Raucher, D., and Sheetz, M.P. (2000). Cell Spreading and Lamellipodial Extension Rate Is Regulated by Membrane Tension. The Journal of Cell Biology *148*, 127–136. https://doi.org/10.1083/jcb.148.1.127.
40. Diz-Muñoz, A., Romanczuk, P., Yu, W., Bergert, M., Ivanovitch, K., Salbreux, G., Heisenberg, C.-P., and Paluch, E.K. (2016). Steering cell migration by alternating blebs and actin-rich protrusions. BMC Biol *14*, 74. https://doi.org/10.1186/s12915-016-0294-x.
41. De Belly, H., Stubb, A., Yanagida, A., Labouesse, C., Jones, P.H., Paluch, E.K., and Chalut, K.J. (2021). Membrane Tension Gates ERK-Mediated Regulation of Pluripotent Cell Fate. Cell Stem Cell *28*, 273-284.e6. https://doi.org/10.1016/j.stem.2020.10.018.
42. Yilmaz, M., and Christofori, G. (2009). EMT, the cytoskeleton, and cancer cell invasion. Cancer Metastasis Rev *28*, 15–33. https://doi.org/10.1007/s10555-008-9169-0.
43. Gao, Y., Dickerson, J.B., Guo, F., Zheng, J., and Zheng, Y. (2004). Rational design and characterization of a Rac GTPase-specific small molecule inhibitor. Proc. Natl. Acad. Sci. U.S.A. *101*, 7618–7623. https://doi.org/10.1073/pnas.0307512101.
44. Kilsdonk, E.P.C., Yancey, P.G., Stoudt, G.W., Bangerter, F.W., Johnson, W.J., Phillips, M.C., and Rothblat, G.H. (1995). Cellular Cholesterol Efflux Mediated by Cyclodextrins. Journal of Biological Chemistry *270*, 17250–17256. https://doi.org/10.1074/jbc.270.29.17250.
45. Biswas, A., Kashyap, P., Datta, S., Sengupta, T., and Sinha, B. (2019). Cholesterol Depletion by MβCD Enhances Cell Membrane Tension and Its Variations-Reducing Integrity. Biophysical Journal *116*, 1456–1468. https://doi.org/10.1016/j.bpj.2019.03.016.
46. Pietuch, A., Brückner, B.R., Fine, T., Mey, I., and Janshoff, A. (2013). Elastic properties of cells in the context of confluent cell monolayers: impact of tension and surface area regulation. Soft Matter *9*, 11490. https://doi.org/10.1039/c3sm51610e.

47. Truong Quang, B.A., Peters, R., Cassani, D.A.D., Chugh, P., Clark, A.G., Agnew, M., Charras, G., and Paluch, E.K. (2021). Extent of myosin penetration within the actin cortex regulates cell surface mechanics. Nat Commun *12*, 6511. https://doi.org/10.1038/s41467-021-26611-2.
48. Serres, M.P., Smith, M.B., Lavoie, G., Roux, P.P., and Paluch, E.K. (2023). Synergistic action of actin binding proteins regulate actin network organization and cell shape. Preprint, https://doi.org/10.1101/2023.09.15.553706 https://doi.org/10.1101/2023.09.15.553706.
49. Kramers, H.A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. Physica *7*, 284–304. https://doi.org/10.1016/S0031-8914(40)90098-2.
50. Horsthemke, W., and Lefever, R. (1984). Noise-induced Transitions: Theory and Applications in Physics, Chemistry, and Biology (Springer-Verlag).
51. Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. Nature *467*, 167–173. https://doi.org/10.1038/nature09326.
52. Kohar, V., and Lu, M. (2018). Role of noise and parametric variation in the dynamics of gene regulatory circuits. npj Syst Biol Appl *4*, 40. https://doi.org/10.1038/s41540-018-0076-x.
53. Pal, S., and Dhar, R. (2024). Living in a noisy world—origins of gene expression noise and its impact on cellular decision-making. FEBS Letters *598*, 1673–1691. https://doi.org/10.1002/1873-3468.14898.
54. Coomer, M.A., Ham, L., and Stumpf, M.P.H. (2022). Noise distorts the epigenetic landscape and shapes cell-fate decisions. Cell Systems *13*, 83-102.e6. https://doi.org/10.1016/j.cels.2021.09.002.
55. Betz, T., Koch, D., Lim, D., and Käs, J.A. (2009). Stochastic Actin Polymerization and Steady Retrograde Flow Determine Growth Cone Advancement. Biophysical Journal *96*, 5130–5138. https://doi.org/10.1016/j.bpj.2009.03.045.
56. Yanagida, A., Corujo-Simon, E., Revell, C.K., Sahu, P., Stirparo, G.G., Aspalter, I.M., Winkel, A.K., Peters, R., De Belly, H., Cassani, D.A.D., et al. (2022). Cell surface fluctuations regulate early embryonic lineage sorting. Cell *185*, 777-793.e20. https://doi.org/10.1016/j.cell.2022.01.022.
57. De Belly, H., Paluch, E.K., and Chalut, K.J. (2022). Interplay between mechanics and signalling in regulating cell fate. Nat Rev Mol Cell Biol *23*, 465–480. https://doi.org/10.1038/s41580-022-00472-z.
58. Rafelski, S.M., and Theriot, J.A. (2024). Establishing a conceptual framework for holistic cell states and state transitions. Cell *187*, 2633–2651. https://doi.org/10.1016/j.cell.2024.04.035.
59. Wilkinson, H.N., and Hardman, M.J. (2020). Wound healing: cellular mechanisms and pathological outcomes. Open Biol. *10*, 200223. https://doi.org/10.1098/rsob.200223.
60. Pinheiro, D., Kardos, R., Hannezo, É., and Heisenberg, C.-P. (2022). Morphogen gradient orchestrates pattern-preserving tissue morphogenesis via motility-driven unjamming. Nat. Phys. *18*, 1482–1493. https://doi.org/10.1038/s41567-022-01787-6.
61. Verboven, S., and Hubert, M. (2005). LIBRA: a MATLAB library for robust analysis. Chemometrics and Intelligent Laboratory Systems *75*, 127–136. https://doi.org/10.1016/j.chemolab.2004.06.003.
62. Labouesse, C., Tan, B.X., Agley, C.C., Hofer, M., Winkel, A.K., Stirparo, G.G., Stuart, H.T., Verstreken, C.M., Mulas, C., Mansfield, W., et al. (2021). StemBond hydrogels control the mechanical microenvironment for pluripotent stem cells. Nat Commun *12*, 6132. https://doi.org/10.1038/s41467-021-26236-5.

63. de Rooij, J., Kerstens, A., Danuser, G., Schwartz, M.A., and Waterman-Storer, C.M. (2005). Integrin-dependent actomyosin contraction regulates epithelial cell scattering. Journal of Cell Biology *171*, 153–164. https://doi.org/10.1083/jcb.200506152.
64. Farrell, J., Kelly, C., Rauch, J., Kida, K., García-Muñoz, A., Monsefi, N., Turriziani, B., Doherty, C., Mehta, J.P., Matallanas, D., et al. (2014). HGF Induces Epithelial-to-Mesenchymal Transition by Modulating the Mammalian Hippo/MST2 and ISG15 Pathways. J. Proteome Res. *13*, 2874–2886. https://doi.org/10.1021/pr5000285.
65. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nature methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.
66. Chung, M.K., and Taylor, J. (2004). Diffusion smoothing on brain surface via finite element method. In 2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (IEEE Cat No. 04EX821) (IEEE), pp. 432–435. https://doi.org/10.1109/ISBI.2004.1398567.
67. Chung, M.K., Dalton, K.M., Shen, L., Evans, A.C., and Davidson, R.J. (2007). Weighted Fourier Series Representation and Its Application to Quantifying the Amount of Gray Matter. IEEE Trans. Med. Imaging *26*, 566–581. https://doi.org/10.1109/TMI.2007.892519.

SUPPLEMENTARY FIGURES

Figure S1

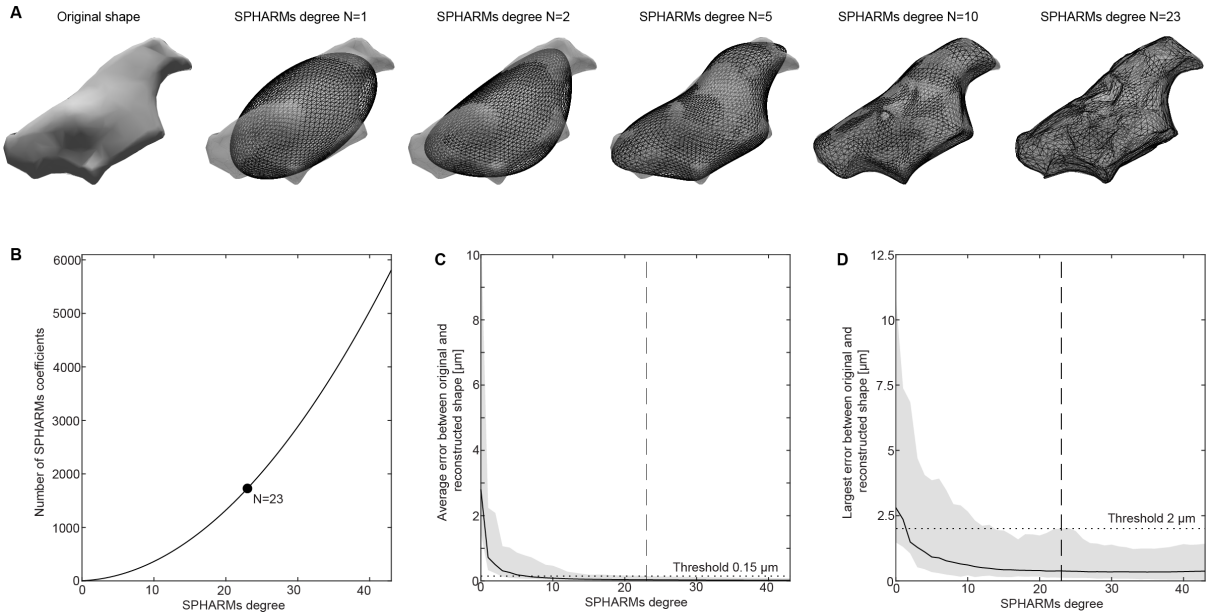

Figure S1: SPHARM reconstruction and identification of most suitable SPHARMs degree, related to Figure 1

(A) Reconstruction of cell shapes from the SPHARMs coefficients of different degrees. The grey cell shows the original cell shape, and the black meshes represent the reconstructed shapes.

(B) Number of SPHARMs coefficients as a function of SPHARMs degree N, given by $3(N+1)^2$. In this study, we used $N = 23$, corresponding to 1728 coefficients.

(C, D) Mean and largest error between original and reconstructed shapes for 0 h and 18 h control and EMT cells as a function of SPHARMs degree. We defined a threshold of 2 µm for the largest error and 0.15 µm for the mean error (dotted lines) as largest acceptable errors to identify the lowest possible SPHARMs degree. We found that this was fulfilled for $N = 23$ (black dashed lines). The error bars indicate the range of error values (lowest and highest maximum and average error of a cell).

Figure S2

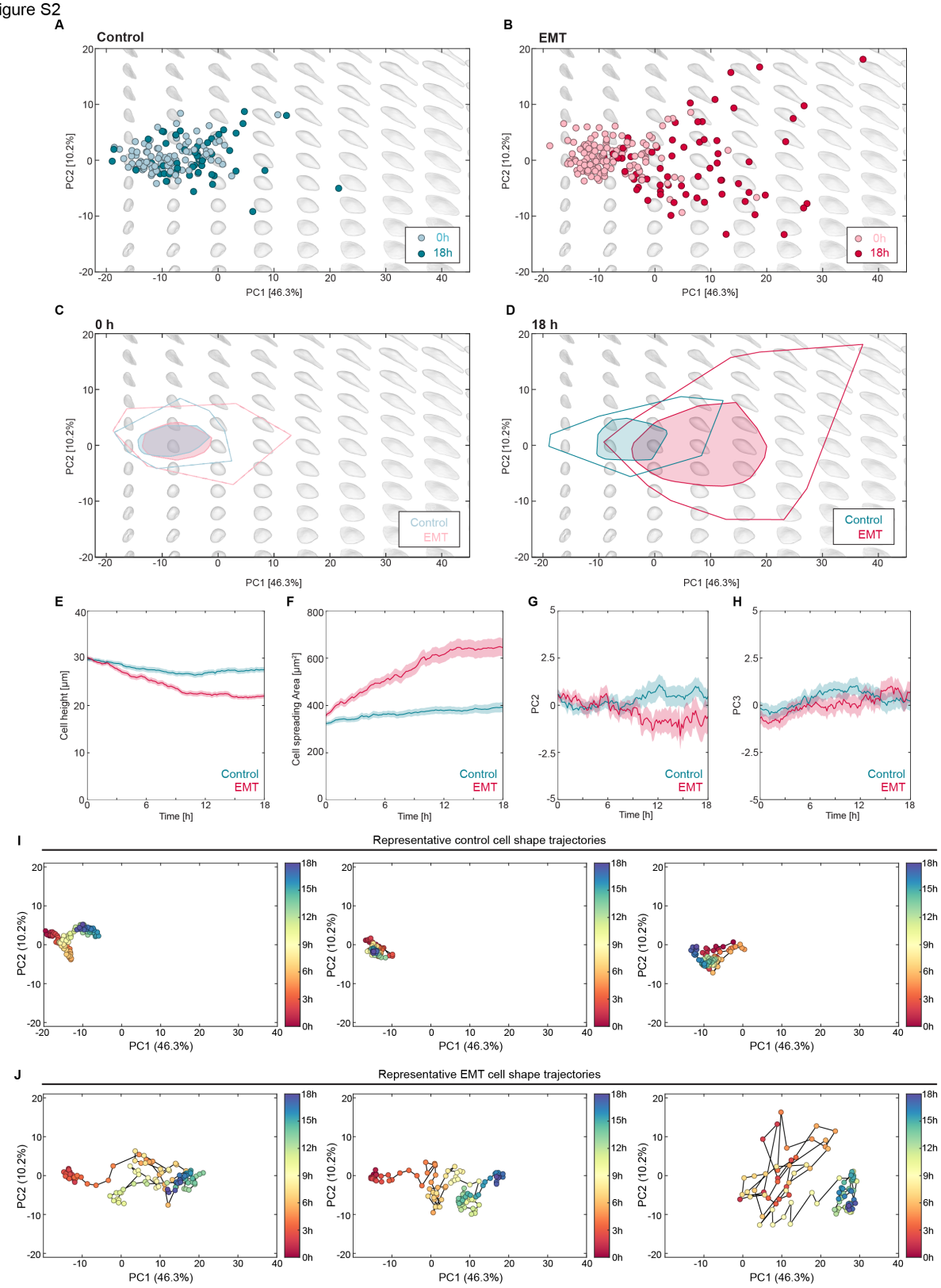

Figure S2: Distribution of control and EMT cells in the morphospace and cell shape dynamics during EMT, related to Figure 2 and 3

(A) Point cloud of all control cells in morphospace at 0 h (N=8, n=128) and 18 h (N=8, n=63).

(B) Point cloud of all EMT cells in the morphospace 0 h (N=12, n=100) and 18 h (N=12, n=55) after EMT induction.

(C) Bagplots of control and EMT cells in morphospace 0 h after EMT induction. Bagplots are a higher-dimensional expansion of 1d boxplots and used here to visualize the range of a point cloud distributions in the PC1-PC2 morphospace. See Methods for details on the bagplots. In the background of the bagplots, we provided cell shapes corresponding to the specific location in the morphospace.

(D) Bagplots of control and EMT cells in morphospace 18 h after EMT induction.

(A-D): Gray background shapes: reconstructed shapes corresponding to the specific location in the PC1-PC2 morphospace.

E,F) Ensemble dynamics of cell height (E) and cell projected area (F) of EMT cells (N=12, n=145) and control cells (N=8, n=102). Mean ± SEM.

(G,H) Ensemble dynamics of PC2 (G) and PC3 (H) of EMT cells (N=12, n=145) and control cells (N=8, n=102). Mean ± SEM.

(I) Representative cell shape trajectories of control cells mapped into the morphospace.

(J) Representative cell shape trajectories of an EMT cells mapped into the morphospace.

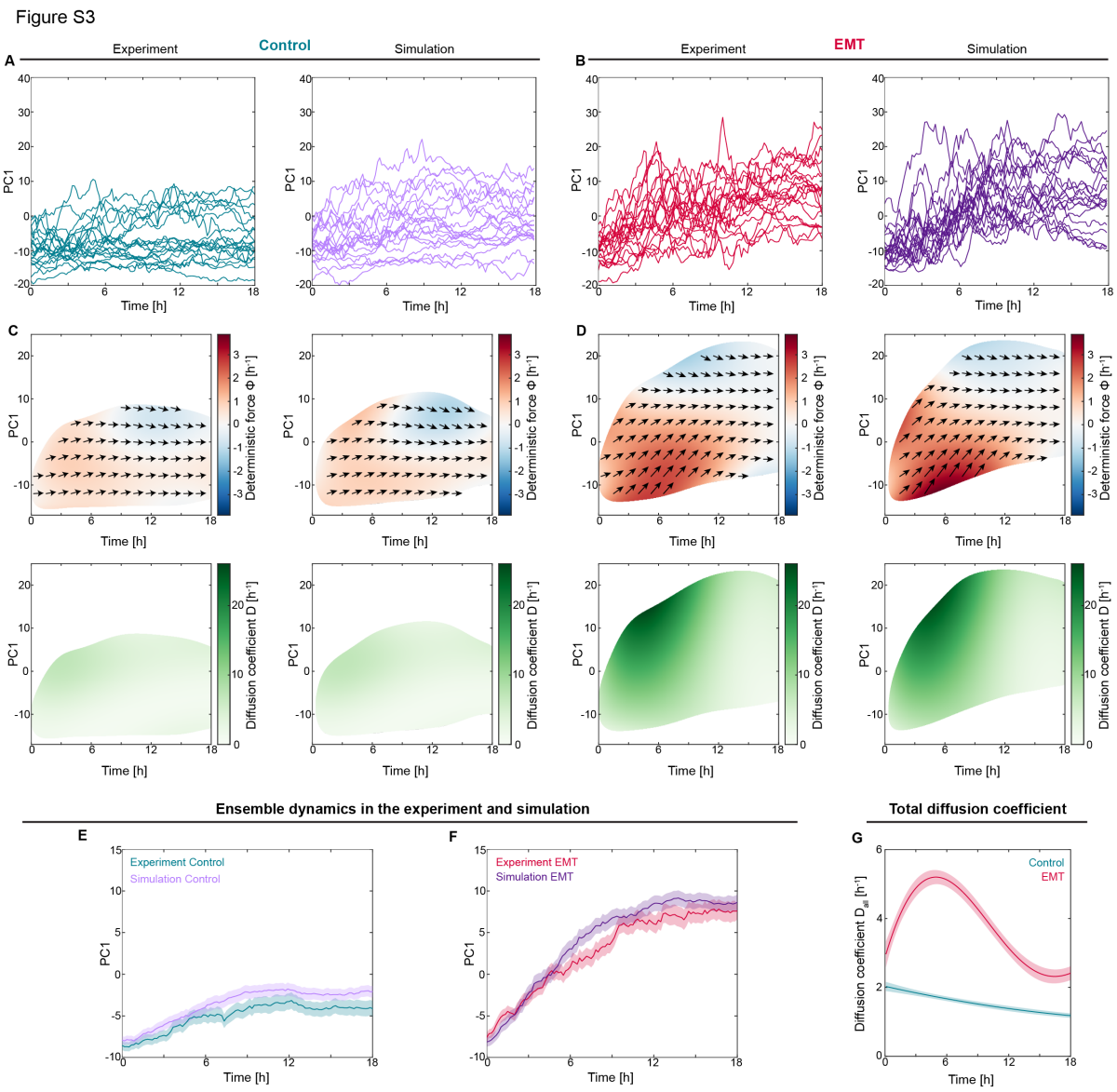

Figure S3: Verification of stochastic inference approach and noise peak, related to Figure 3

(A,B) Trajectories of ρ of experimental control (A) and EMT (B) cells and *in silico* trajectories for the inferred deterministic force field and diffusion coefficient (Figure 2F, 2G, 2H and 2I; Equation 2). Details of the generation of *in silico* trajectories is provided in Methods.

(C,D) Deterministic force field $\Phi(\rho, t)$ and diffusion coefficient $D(\rho, t)$ of Equation 2 inferred from experimental and simulated ρ data of control (C) and EMT (D) cells show good agreement between experiments and theory.

(E,F) Ensemble dynamics of experimental (N=12, n=145) and simulated PC1 trajectories (n=100) of control and EMT cells exhibit good qualitative agreement. Mean values of the shape features are represented as mean $\pm$ SEM.

(G) Total diffusion coefficient of the combination of the first 22 principal (Methods).

Figure S4

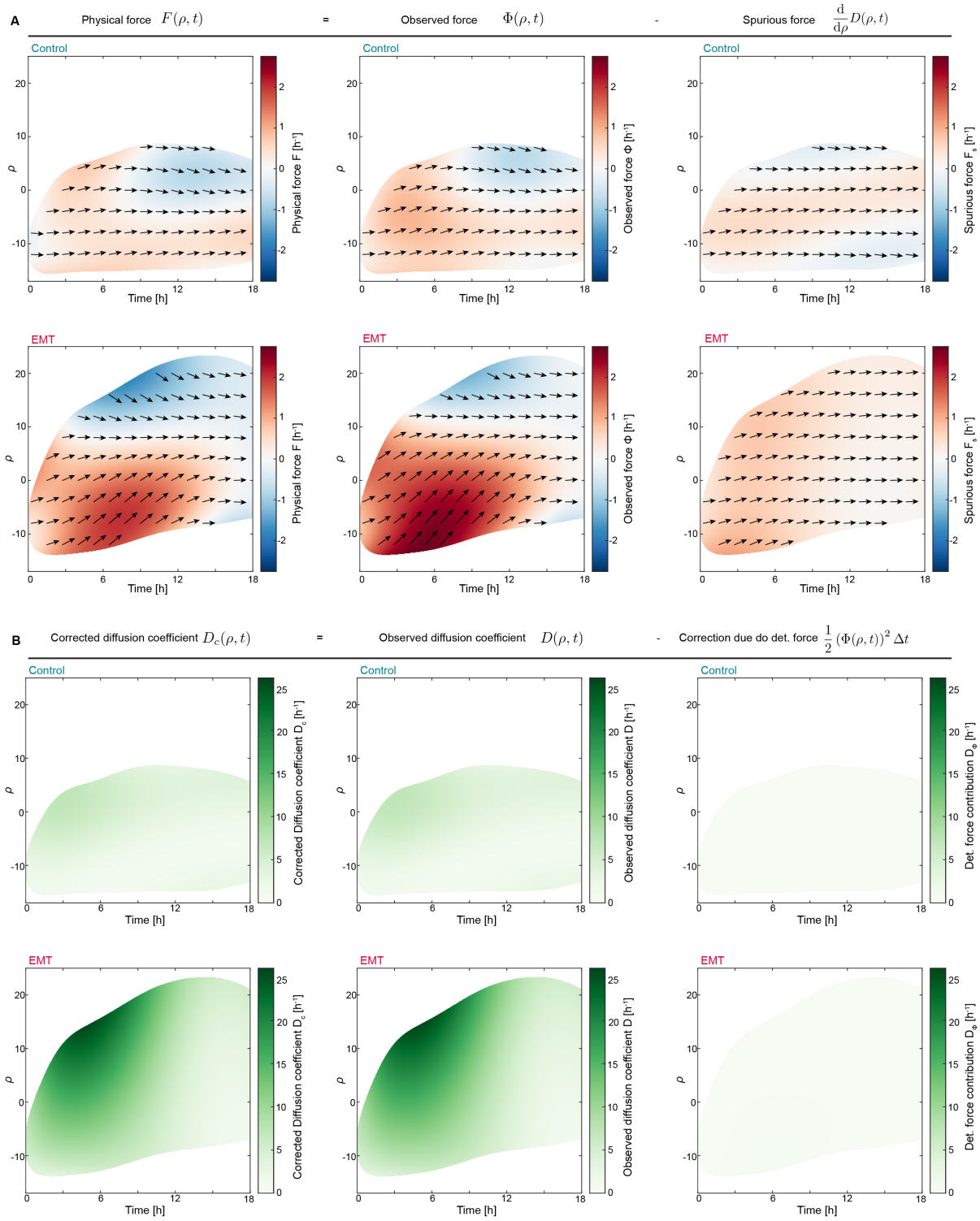

Figure S4: Contributions due to the spurious force and the mean squared displacement have no qualitative impact on the stochastic inference, related to Figure 3

(A) Contribution due to a spurious force $\mathrm{d}D(\rho,t)/\mathrm{d}\rho$ (Supplemental Information) to the

inferred deterministic force $\Phi(\rho, t)$ for control and EMT cells. We found that subtracting the spurious force from the inferred deterministic force field did not lead to any qualitative differences between $\Phi(\rho, t)$ and the corrected physical force $\mathrm{F}(\rho, t)$.

(B) Contributions of the deterministic force field $\Phi(\rho, t)$ to the diffusion coefficient $D(\rho, t)$ for control and EMT cells. The mean squared displacements mediated by the deterministic force field, given by $1/2\big(\Phi(\rho, t)\big)^2 \Delta t$ (for a time step $\Delta t = 10\ \mathrm{min}$; Methods; Equation 15) are so small that they can be neglected when calculating the diffusion coefficient D.

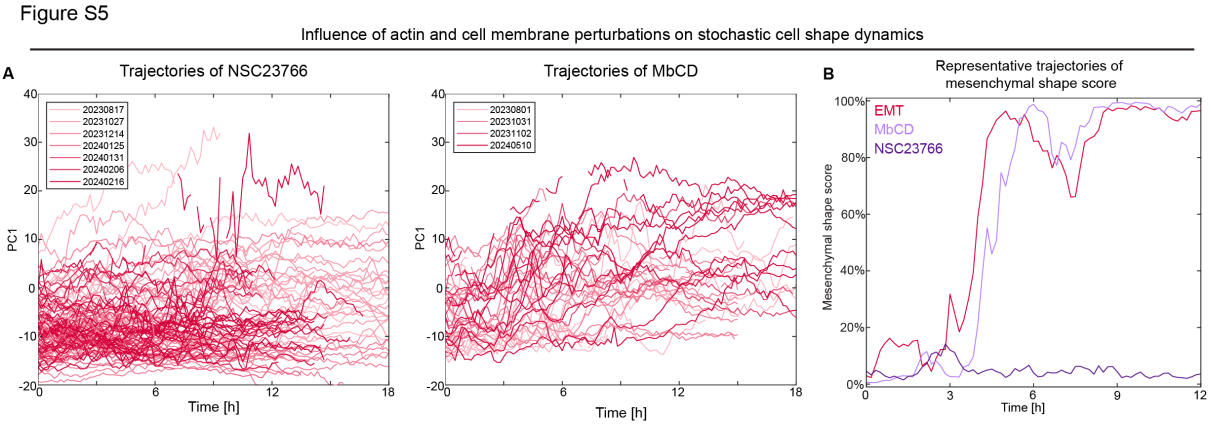

Figure S5: Effect of drug treatments on EMT-associated cell spreading

(A) PC1 dynamics of individual cells treated with 100 µM Rac1 inhibitor NSC23766 (left) and 1mM MbCD (right). Insets highlight independent experiments.

(B) Mesenchymal shape score as a function of time for of individual representative untreated cells (red line) and cells treated with 100 µM Rac1 inhibitor NSC23766 (dark purple line) and 1mM MbCD (light purple line).