Extracellular Matrix regulates the morphodynamics of 3D migrating cancer cells

Christopher Z. Eddy,1 Helena Raposo,2 Ryan Wong,1 Fuxin Li,3 and Bo Sun1

1Department of Physics, Oregon State University, Corvallis OR, 97331
2Department of Chemical, Biological, and Environmental Engineering, Oregon State University, Corvallis OR, 97331
3School of Electrical Engineering and Computer Science, Oregon State University, Corvallis OR, 97331

Cell shape is linked to cell function. The significance of cell morphodynamics, namely the temporal fluctuation of cell shape, is much less understood. Here we study the morphodynamics of MDA-MB-231 cells in type I collagen extracellular matrix (ECM). We systematically vary ECM physical properties by tuning collagen concentrations, alignment, and gelation temperatures. We find that morphodynamics of 3D migrating cells are externally controlled by ECM mechanics and internally modulated by Rho-signaling. We employ machine learning to classify cell shape into four different morphological phenotypes, each corresponding to a distinct migration mode. As a result, we map cell morphodynamics at mesoscale into the temporal evolution of morphological phenotypes. We characterize the mesoscale dynamics including occurrence probability, dwell time and transition matrix at varying ECM conditions, which demonstrate the complex phenotype landscape and optimal pathways for phenotype transitions. In light of the mesoscale dynamics, we show that 3D cancer cell motility is a hidden Markov process whereby the step size distributions of cell migration are coupled with simultaneous cell morphodynamics. We also show that morphological phenotype transitions facilitate cancer cells to navigate non-uniform ECM such as traversing the interface between matrices of two distinct microstructures. In conclusion, we demonstrate that 3D migrating cancer cells exhibit rich morphodynamics that is regulated by ECM mechanics, Rho-signaling, and is closely related with cell motility. Our results pave the way to the functional understanding and mechanical programming of cell morphodynamics for normal and malignant cells.

INTRODUCTION

Shape defines the cell. In the 1677 book Micrographia, Robert Hooke showed sections within a herbaceous plant under a microscope. The shape of those sections resembles cells in a monastery, so he named the structures cells [1]. Many breakthroughs followed Hooke’s discovery, from the cell theory of Schwann and Schleiden, to the theory of tissue formation by Remak, Virchow and Kolliker, and the theory of cellularpathologie by Virchow, all of which are inspired by observations of cell shapes, or morphology in general [2, 3].

In our modern view cell shape is determined by cell function [4, 5]. A nerve cell has long branched protrusions for communication with other neurons; while the cuboidal shape of epithelial cells allow them to tile the surface of organs. Loss of characteristic shape, on the other hand, is associated with functional abnormality. Thus morphological characterization has been an important tool for diagnosis such as in red blood cell disease [6], neurological disease [7], and cancer [8, 9]. More recently, cell shape analysis is boosted by techniques from computer vision. As a result, it becomes possible to obtain high content information of cellular states from morphological data alone [9–12].

While most research focuses on the static cell morphology, the dynamic fluctuation of cell shape is much less understood. However, shape fluctuation – namely morphodynamics, is of central importance for dynamic cellular functions. The abnormal diffusion of small protrusions - microvilli - on the surface of a T cell allows the T cell to efficiently scan antigen-presenting surfaces [13]. For a migrating cancer cell, morphodynamics drives the motility of the cell in many ways similar to our body frame movements that enable swimming. In fact, just as there are different swimming styles, cancer cells have been observed to execute multiple programs – migration modes – during invasion in 3D tissue space [14]. Each mode has distinct signatures of morphology and morphodynamics, and are usually classified based on cell morphology as filopodial, lamellipodial, lobopodial, blebbing, and actin-enriched leading edge [15]. Cancer cell migration modes is controlled by intracellular signaling such as the Rock-myosin pathways [16, 17], and extracellular factors such as the elasticity, and degradability of the extracellular matrix (ECM) [15, 18]. The ability of a cancer cell to switch between migration modes is important for tumor prognosis. Many therapies, such as MMP inhibitors that target a particular mode of cell motility, fail to stop tumor metastasis largely because cells take other available migration programs [19, 20].

In this paper, we study the morphodynamics of MDA-MB-231 cells, a highly invasive human breast cancer cell line in 3D collagen matrices. We devise machine learning techniques to classify cell shapes into morphological phenotypes that correspond to known migration modes. This approach provides a mesoscale mapping of cell morphodynamics into transitions among morphological phenotypes. We find individual cells are capable of rapidly sampling multiple morphological phenotypes, implying spontaneous migration mode transitions. We find ECM mechanics coupled with cell mechanosensing pathways regulate the stability and transition rates between morphological phenotypes. We also find that such transi-
tions facilitate cancer cells to navigate ECM with inherent structural and mechanical heterogeneity.

![Image](image.png)

FIG. 1. Three-dimensional migration of MDA-MB-231 cells is accompanied with significant cell shape fluctuation. (A) A typical time lapse recording of 25 hours is projected onto a single image with colors representing time. (B) The real space (x-y plane), and shape fluctuations of 3 cells shown in (A). (C) The mean square displacement ($\sigma^2$) of selected cell geometric measures. Dots: experimental measurements. Solid lines: linear fit. Dashed lines: 95% prediction interval. Here the form factor is defined as perimeter$^2$/area. Curl is defined as the ratio between the major axis length and skeletonized contour length.

RESULTS

We find 3D migrating cancer cells demonstrate rapid shape fluctuations (Fig. 1(A-B)). In order to quantify the cell morphodynamics, we take time-lapse fluorescent images of MDA-MB-231 cells migrating in collagen matrices. The GFP-labeled cells typically stay within the focal depth of the objective lens (20X, NA 0.7) for 10-20 hours, while we obtain 2D cell images at a rate of 4 frames per hour (see SI Appendix section S1). After binarization and segmentation, we compute a total of twenty-one geometric measures which collectively quantify the shape of a cell (see SI Appendix section S2). These geometric measures characterize cell size (such as area and perimeter), deviation from circle (such as aspect ratio and form factor), surface topography (such as solidity), and backbone curvature (such as curl – the ratio between the major axis length and skeletonized contour length).

The morphodynamics of a cell manifests itself as a random walk in the geometric shape space concurrent with its motility in the 3D matrix (Fig. 1). However, unlike the real space motility that is slightly diffusive, we find cell morphodynamics is subdiffusive in the shape space (Fig. 1C and SI Appendix section S3). The subdiffusivity suggests physical barriers that are present both intrinsic to the cells and from the 3D ECM. Indeed, we find cells moving on 2D surface exhibit faster shape fluctuations than cells embedded in 3D ECM. Still, on flat surfaces cells show subdiffusive random walks in shape space and superdiffusive walks in real space (SI Appendix section S3).

![Image](image.png)

FIG. 2. Develop a supervised machine learning model to classify cells into morphological phenotypes corresponding to different migration modes. (A) MDA-MB-231 cells in 3D collagen matrices exhibit multiple morphological phenotypes that are characteristic of four distinct migration modes: actin-enriched leading edge (AE), small blebbing (BB), filopodial (FP), and lamellipodial (LA). Scale bars are 20 μm. (B) The cell images are quantified using a total of 21 geometric measures such as area, solidity, and aspect ratio. With 3800 manually labeled single cell images we have trained a supported vector machine (SVM) to calculate probability scores (Val) for a cell to belong to each morphological phenotypes (classes). We assign a cell to the class C with the maximum score ($Max(C, Val)$), if this maximum score is greater than a threshold of 60% ($Max(C, Val) > 0.6$). We consider a cell to be at an intermediate state if none of the four classes have a score higher than 0.6. In (B) a sample cell image is classified as a lamellipodial cell (LA), because LA class has a score of greater than 0.6. (C) To better visualize the high dimensional geometric measures, we apply t-SNE method to generate a 3D projection of the cell shape spaces. 25,000 unseen data set is presented here. Different morphological phenotypes are well separated. AE (yellow): actin-enriched leading edge. BB (green): small blebbing. FP (magenta): filopodial. LA (blue): lamellipodial.

After quantitatively demonstrating the shape fluctuations of migrating cancer cells, we next investigate cell morphodynamics at a mesoscale that allows us to gain insights on cell migration modes. This is possible because different migration modes are associated with dis-
tinct characteristic cell morphologies (Fig. 2A) [14, 15]. Using 3800 manually labeled single cell images, we have trained machine classifiers that classify cell morphology into four morphological phenotypes based on and named after their corresponding migration modes.

We consider four morphological phenotypes including two amoeboidal ones: actin-enriched leading edge (or AE in short) and small blebbing (BB); as well as two mesenchymal ones: filopodial (FP) and lamellipodial (LA). Of note, another migration mode, namely lobopodial or nuclear piston mode, has not been observed in our experiments which is consistent with previous reports [21]. Once the classifier is trained, morphological phenotypes are determined automatically from a cell image if a particular phenotype receives more than 60% probability score (Fig. 2B). For a small fraction of cells (∼10%), none of the four phenotypes receive more than 60% probability score, we consider these cells to be in an intermediate state.

We have trained two classifiers (see SI Appendix S4). The first one is based on support vector machines (SVM [22, 23]) involving 21 geometric measures of binarized cell images. The second one is based on Random-Forest model using the same geometric properties [24]. The two classifiers agree with each other well on test data sets (90% overlapping). The SVM classifier particularly has a higher success rate of classifying unseen data (92%). In the following we mainly report the results from SVM algorithm. The validity of the SVM classifier is also evident from the color-coded t-SNE embedding of unseen data (Fig. 2C, 25,000 data points), as data points belong to different classes (morphological phenotypes) form separable clusters in the embedding space.

By applying the SVM classifier to time lapse recordings of 3D migrating MDA-MB-231 cells we find cells spontaneously make transitions among different morphological phenotypes. Fig. 3A shows snapshots of a typical cell. The cell switches directly from filopodial (F) to lamellipodial (L) mode, then to small blebbing (B) mode via intermediate state (I). Therefore using machine learning technique we map cell morphodynamics into transitions between morphological phenotypes, or their associated migration modes.

In order to understand the mechanisms underlying cell morphological phenotype transitions, we examine the effects of manipulating Rho-signaling, which is a master regulator that determines the mechanical state of a cell. To this end, we apply Y27632 [25], a Rho-inhibitor; and CN03 [26], a Rho-activator to MDA-MB-231 cells cultured in collagen ECM (see SI appendix S5). Y27632 reduces actomyosin contractility, promoting transitions from blebbing to mesenchymal phenotypes [27]. On the other hand, CN03 elevates myosin II activity, leading to retraction of filopodia to rounded cell shapes (Fig. 3B). These results are consistent with previous reports on the molecular control of cell migration modes by Rho-signaling [15].

While previous studies focus on the end points of manipulating Rho-signaling, morphodynamic analysis offers insights to the transition paths between migration modes. In particular, we take advantage of a modified t-SNE algorithm [28], which projects a cell image in the embedding space defined by the training sets (Fig. 2C, SI appendix S4). This approach allows us to map the continuous shape change of a cell as a trajectory in the (dimensional-reduced) mesoscale morphodynamic space. The trajectories start immediately after introducing Y27632 or CN03, and ends after 4 hours of incubating with the drugs. The forward time directions are shown as thick curves with arrows as guide to the eyes. Two representative trajectories (one with circular symbols, and another one with triangular symbols) per each treatment are shown as colored symbols connected by black lines, where color represents the instantaneous phenotype. Unconnected light-colored dots show training sets which are the same as in Fig. 2C. Note to better visualize the 3D trajectories coordinates have been rotated with respect to Fig. 2C.

FIG. 3. Rho-signaling internally controls mesoscale morphodynamics of 3D cultured MDA-MB-231 cells. (A) A sample time series of morphological phenotype. Insets: three snapshots showing the GFP-labeled cell morphology. Abbreviations: F – filopodial, L – lamellipodial, I – intermediate state, B: small blebbing. (B) Representative morphological changes under treatment of CN03 and Y27632, which upregulates and downregulates Rho A respectively. (C-D) Characteristic morphodynamic trajectories of cells in the t-SNE embedded shape space. The trajectories start immediately after introducing Y27632 or CN03, and ends after 4 hours of incubating with the drugs. The forward time directions are shown as thick curves with arrows as guide to the eyes. Two representative trajectories (one with circular symbols, and another one with triangular symbols) per each treatment are shown as colored symbols connected by black lines, where color represents the instantaneous phenotype. Unconnected light-colored dots show training sets which are the same as in Fig. 2C. Note to better visualize the 3D trajectories coordinates have been rotated with respect to Fig. 2C.
when treated with Y27632 blebbing cells turn to filopodial or lamellipodial via strongly converging trajectories most of which first visiting AE states (see also SI appendix S5). Fig. 3C shows two representative trajectories. AE state exhibits weak cell-ECM adhesions and F-actin rich protrusions [30, 31]. Our results suggest AE states mediate Rho-signaling controlled transition from amoeboidal to mesenchymal motility. On the other hand, CN03 treatment causes the majority of mesenchymal cells to switch to blebbing modes. However, without going through AE states, CN03 leads to strongly fluctuating and diverging trajectories from multiple cells (Fig. 3D shows two representative trajectories, see also SI appendix S5).

We next investigate the external control of cell morphodynamics. In particular, we focus on the role of ECM physical properties in regulating cell morphological phenotype transitions. In order to control the microstructure of collagen matrices we employ three methods as shown in Fig. 4A-D (see also SI appendix S6). First, increasing gelation temperature from room temperature (RT) to 37 °C, while keeping collagen density at [col] = 1.5 mg/mL significantly reduce fiber length and pore size (Fig. 4B). Second, increasing collagen density to [col] = 3.0 mg/mL while keeping gelation temperature at room temperature moderately reduces pore size, preserves clear fibrous structure, and increases stiffness (Fig. 4C). Finally, keeping gelation at room temperature and [col] = 1.5 mg/mL while generating an unidirectional flow field during gelation leads to aligned collagen fibers in the ECM. This method creates strong anisotropy in the ECM.

We find the occurrence probability (or population fraction) of different morphological phenotypes are remarkably different at different ECM conditions. As shown in Fig. 4E, increasing gelation temperature does not affect the probability of AE and LA cells. However, the homogeneous matrix microstructure at 37 °C significantly reduces the fraction of FP cells from 43% to 25%, while increases fraction of BB cells from 15% to 25%. Compared with increasing gelation temperature, doubling collagen concentration leads to less dramatic changes of the ECM microstructure. Correspondingly, only moderate changes of phenotype probabilities are observed. On the other hand, when matrix anisotropy is increased by aligning ECM fibers, we find significant shift of cells from amoeboid phenotypes to filopodial mode. Taken together, these results show that ECM heterogeneity and anisotropy determine the probability of different morphological phenotypes.

We have also examined the stability of each morphological phenotypes by measuring the average dwell time – duration of a cell to stay continuously in a morphological phenotype before transition to another (SI appendix S7). As shown in Fig. 4F, in all three cases manipulating ECM physical properties moderately increase the dwell time of all four morphological phenotypes. Therefore the changes in the phenotype probability can not be explained by the phenotype stability alone, and in some cases move in opposite trend from the dwell times observed.

To reveal further details of morphological phenotype dynamics, we have computed the phenotype transition matrix: rates r that characterize the probability of transitions per hour between any two phenotypes (Fig. 4G-J) (see also SI appendix S7). While the rates vary dramatically for different ECM conditions (arrows in Fig. 4G-J), we notice several remarkable common features. First, direct transitions along FP - BB path rarely happens (r < 0.03 hr⁻¹). Instead, amoeboidal - mesenchymal transitions are primarily mediated by LA and AE states, presumably by turning cell-ECM adhesion on and off. On the other hand, transitions within the amoeboidal (AE - BB) and mesenchymal (FP - LA) modes are frequent, and the rates can go up to 1 per hour. Finally, while the morphological phenotype transitions are intrinsically non-equilibrium processes, probability fluxes between states are generally very small (SI appendix S7). This means that an approximate detailed balance exists among morphological phenotypes. In comparison with other nonequilibrium stationary processes at mesoscale [32], we speculate morphological phenotype transitions are not gated by active processes such as ATP consumption.

The transition rates also offer insights to understand the ECM-dependence of the fraction of cells in each morphological phenotype (Fig. 4E). For instance, as gelation temperature increases from RT to 37 °C, rates from AE to FP decreases by 52 percent, and rates from BB to AE decreases by 22 percent (Fig. 4G and Fig. 4I). As a result, we observe more blebbing cells and less filopodial cells in collagen matrices prepared at 37 °C. This is consistent with the mechanical mechanism of blebbing formation [33, 34]. Blebs form when actomyosin contractility exceeds the binding between cortical actin and cell membrane. A blebbing cell turns to AE when actin polymerization causes sharp protrusion on the membrane. Our results suggest that as collagen ECM loses structure heterogeneity, actin polymerization is less effective to drive the transition from BB to AE states.

Conversely, as ECM becomes more anisotropic (Fig. 4G and Fig. 4J), the transition rate from LA to FP increases as much as 27 percent, while rate from AE to BB decrease by 44 percent. Together, these altered rates lead to a significant fraction of blebbing cells turning to filopodial as shown in Fig. 4E. Filopodial protrusions consist of elongated F-actin bundles supported by elevated actin polymerization and cross-linking by Ena/VASP proteins [35]. Our results suggest that the mechanical barrier separating filopodia and blebbing protrusions is too high for actomyosin contractility to overcome directly. Instead, a blebbing cell turning into a filopodial one has to first
transform into AE or LA states. Because the morphological phenotype of a cell is linked to its 3D migration mode, we next investigate if the invasion potential of MDA-MB-231 cells depends on the mesoscale morphodynamics. Due to the short dwell times for each morphological phenotype, we only consider two coarse-grained classes of morphologies: mesenchymal (ME), which consists of FP and LA states; and amoeboidal (AM), which consists of AE and BB states. In particular, we measure for short time scales the step size distributions and for longer time scales the mean square displacement of the cells in randomly aligned collagen matrices geled at room temperature (Fig. 5).

Interestingly we find the steps are better described by a log-normal, rather than Gaussian distribution (SI appendix S8) due to frequent large steps. Fig. 5A show the mean and variance of the fitting parameters. It is clear that the steps in physical space are coupled with the corresponding mesoscale dynamics. For cells that dwell in the amoeboidal class, both mean and variance of the steps are the smallest. Correspondingly, the mean square displacement of amoeboid cells have a small slope, corresponding to an effective diffusivity of \(8 \mu m^2/\text{hour}\) (for each spatial dimension, Fig. 5B). On the other hand, cells make larger steps when dwell in the mesenchymal class, and the effective diffusivity increases by three-fold to \(26 \mu m^2/\text{hour}\). Importantly, when cell migration is simultaneously coupled with phenotype transitions, the class-switching steps have distinct statistical distributions (Fig. 5A).

Our analysis shows that not only it is important to distinguish different morphological phenotypes in studying the motility of cancer cells, but also one may need to take into account of phenotype transitions. For instance, without accounting for the class-switching events, the weighted average of mean square displacements from mesenchymal and amoeboidal cells underestimates the observed cell motility by 15% (Fig. 5B, the weighted average MSD curve corresponds to a diffusivity of approximately \(20 \mu m^2/\text{hour}\), as compared with \(24 \mu m^2/\text{hour}\) for full cell trajectories). Since phenotype transitions occur rapidly at single cell level regardless of ECM concentration, porosity and rigidity, we conclude that mesoscale dynamics contributes to determine the invasive potential of cancer cells.

During metastasis a migrating cancer cell must navigate ECM of distinct mechanical properties. Therefore we next investigate how cell morphodynamics facilitate cell traverse interfaces and adapt to ECM of distinct mechanics. To this end, we create collagen matrices consist of two integrated layers (SI appendix S9). The RT layer is prepared at room temperature that shows a porous fibrous network, and the \(37 ^\circ C\) layer is prepared at \(37 ^\circ C\) showing a much more homogeneous structure (Fig. 6A). Without additional cues MDA-MB-231 cells randomly navigate the ECM, occasionally traverse the interface to experience a sudden change of ECM physical properties. Over the course of 24 hours we do not observe durotaxis.

Consistent with the corresponding results in uniform
ECM, we find the likelihood of observing a filopodial cell is significantly higher in the ECM layer prepared at room temperature, while for blebbing cells the probability is higher in the gel layer prepared at 37 °C (Fig. 6B). The dwell times of phenotypes, on the other hand, follow the same trend of occurrence probabilities but change rather moderately (Fig. 6C).

The shift of phenotype homeostasis once again can be understood from the phenotype transition matrices. To simplify the discussion in Fig. 6D we plot the top four matrix elements of the transition matrices calculated from cells in each of the two layers. In the RT layer, filopodial cell population is enriched by frequent LA-AE exchange and LA to FP transition. As cells cross the interface, LA to FP becomes less likely, and the AE-BB pathway is steered to favor blebbing states.

To further quantify the effect of the interface in modulating cell morphodynamics, we calculate spatial frequencies of dwell events (Fig. 6E) and AE-originating transition events (Fig. 6F). We define a coordinate system where the y-axis is along the interface passing x=0 (Fig. 6A). This allows us to combine data from multiple repeating experiments where cell locations are seeded randomly. After aligning the coordinates, we define the spatial frequency of transition (or dwell) events from state $i$ to state $j$ as $R(i \rightarrow j, x)$, which satisfies

$$P(i \rightarrow j, x) = R(i \rightarrow j, x)M(x)$$

Here $P(i \rightarrow j, x)$ is the probability density of observing event $i \rightarrow j$ per unit time (1 hour), and $M(x)$ is the cell density (along x-axis). We use a 1-D Gaussian kernel to estimate $P(i \rightarrow j, x)$ and $M(x)$ (SI appendix S9). As a result, the spatial frequency $R(i \rightarrow j, x)$ represents the likelihood of a cell to undergo a specific type of transition over unit time (1 hour) as a function of distance to the interface.

The spatial frequency of dwell events clearly show that while FP state is increasingly stable into the RT layer, LA and BB states are more stable in the 37 °C layer. AE state, on the other hand, is most stable at the interface (Fig. 6E). Therefore AE state plays a special role in mediating the cell adaptation across distinct ECM layers. Indeed, we find a gradual shift of favorable AE-originating transitions as distance to the interface varies. The frequency of AE to LA events, the main amoeboidal to mesenchymal path, peaks in the RT layer. AE to BB events, which is mainly responsible of enriching blebbing cells, has peak frequency in the 37 °C layer. Taken together, we find morphological phenotype transitions and the associated migration mode switching are integral parts of cancer cell invasion and adaptation to complex ECM.

**DISCUSSION**

In this paper, we report the morphodynamics of MDA-MB-231 cells in type I collagen ECM as a model system of metastatic cancer cells migrating in 3D tissue. MDA-MB-231 cells rapidly change their geometry, exhibiting a subdiffusive random walk in the shape space. This occurs simultaneously with their superdiffusive walks in the real space (Fig. 1).

The biological significance of the morphodynamics is further demonstrated by classifying cell shapes into morphological phenotypes corresponding to different migration programs (Fig. 2). This allows us to study cell morphodynamics at the mesoscale, in terms of morphological phenotype transitions. Utilizing machine learning and visualization techniques, we show that cell morphodynamics is regulated by Rho-signaling (Fig. 3), which is a molecular control hub of cell mechanosensing and force generation. It has been shown previously that Rho/Rac signaling regulates the shift between mesenchymal and amoeboidal motility [16, 18]. Our analysis further reveals that instead of favoring a particular mode of motility, perturbations of Rho signaling alter the migration
mode transition rates. In particular, down regulating Rho leads to overall amoeboidal-to-mesenchymal transition that routes through AE and LA states. Up regulation of Rho, on the other hand, leads to strongly fluctuating morphodynamics that enriches blebbing cells. The irreversibility of up and down regulating Rho signaling results suggest a complex phenotype landscape that controls 3D cancer cell motility.

We study morphological phenotype transitions in ECM of distinct physical properties and find ECM microstructure modulates the probabilities, dwell times, and transition rates of morphological phenotypes. Collagen matrices with homogeneous structure, as those prepared at higher temperature, enrich the population of blebbing cells. By comparing the transition matrices, we find the enrichment of blebbing cells is directly related with the reduced transition rate from BB to AE state, and also indirectly contributed by the mesenchymal-to-amoeboidal transition through LA and AE states. Similarly, collagen matrices with structural anisotropy enrich the population of filopodial cells. The enrichment is directly attributed to an increased LA to FP rate, and indirectly contributed by the amoeboidal-to-mesenchymal transition mediated by LA and AE states. These results show that it is possible to execute external control of cell morphodynamics (and the corresponding 3D migration modes) through ECM mechanics. Importantly, taking into account of the phenotype transitions allows us to better predict the outcome of manipulating cell migration mode through ECM physical properties [14, 36].

In light of the rapid phenotype transitions exhibited by individual cells, 3D cancer cell motility may be considered as a hidden Markov process where each phenotype is associated with characteristic step size distributions (Fig. 5). Specifically, we find steps that occur simultaneously with a phenotype transition have distinct sizes compared with steps that occur while cells dwell in a particular morphological phenotype. This makes morphodynamics a crucial factor in determining the invasive potential of cancer cells. To our knowledge, this aspect has been so far largely overlooked in the literature.

In the lens of a hidden Markov process, morphodynamics may facilitate cancer invasion because phenotype transitions allow cancer cells to search for and commit to a more effective migration program [37]. Using a ECM model consisting of two mechanically distinct layers, we show the cells gradually adjust their morphodynamics as they approach and cross the layer interface (Fig. 6). Therefore morphological phenotype transitions may be essential in cancer cell metastasis by enabling the cells to navigate non-uniform ECM.

In summary, we demonstrate the morphodynamics of 3D migrating cancer cells as a powerful tool to inspect the internal state and microenvironment of the cells. Investigated at mesoscale, the morphodynamics imply that 3D cancer cell migration is inherently plastic [15]. The plasticity is controlled by the mode transition matrices, rather than a deterministic decision tree [14]. In order to further exploit the information provided by the cell shape fluctuations, future research is needed to decode morphodynamics as a rich body language of cells, and to control morphodynamics as a route of mechanical programming of cell phenotype.

---

**FIG. 6.** Morphological phenotype transition facilitates cell migration in heterogeneous ECM. (A) Time-lapse projection of 3D migrating MDA-MB-231 cells navigating engineered heterogeneous ECM. The ECM contains two adjacent layers that are prepared at room temperature (RT) and 37°C respectively. A confocal reflection image shows the ECM structure next to the interface (dashed line, y-axis). Scale bar: 50 μm. (B) Fraction of cells of each morphological phenotype in both sides of the interface. (C) Dwell time of cells of each morphological phenotype in both sides of the interface. (D) Phenotype transition rates in both sides of the interface. (E) Spatial frequency of dwell events. (F) Spatial frequency of AE to AE, AE to LA and AE to BB events. See main text for the definition of spatial frequency R(i → j, x). In (E-F) dashed lines indicate the interface (x=0) separating the 37°C gel (left) and RT gel (right). A total of 3,800 hours of single cell recordings from three independent experiments have been used to calculate the results in (B-F).
MATERIALS AND METHODS

See SI Appendix S1-S9 for details of 3D cell culture, microscopy, pharmacological treatments, and data analysis.

We thank Prof. Michelle Digman and Prof. Steve Pressé for helpful discussions. The funding for this research results from a Scialog Program sponsored jointly by Research Corporation for Science Advancement and the Gordon and Betty Moore Foundation through a grant to Oregon State University by the Gordon and Betty Moore Foundation. Part of this research was conducted at the Northwest Nanotechnology Infrastructure, a National Nanotechnology Coordinated Infrastructure site at Oregon State University which is supported in part by the National Science Foundation (grant NNCI-1542101) and Oregon State University. C. Eddy and B. Sun are supported by DOE award W81XWH-20-1-0444 (BC190068). B. Sun is also supported by the National Institute of General Medical Sciences award 1R35GM138179.

[1] R. Hooke, Micrographia (The Royal Society, 1665).
[2] P. Mazzarello, Nat. cell biol., E13 (1999).
[3] E. Mayr, The Growth of the Biological Thought (Belknap, Cambridge, MA, 1982).
[4] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, Molecular biology of the cell (Garland Science, 2014).
[5] R. Singhvi, A. Kumar, G. P. Lopez, G. N. Stephanopoulos, D. I. Wang, G. M. Whitesides, and D. E. Ingber, Science 316, 3162–3162 (2007).
[6] M. Diez-Silva, M. Dao, J. Han, C.-T. Lim, and J. Suresh, MRS bulletin / Materials Research Society 35, 382 (2010).
[7] A. Serrano-Pozo, M. P. Frosch, E. Masliah, and B. T. Hyman, Cold Spring Harbor Perspectives in Medicine 6, a006189 (2011).
[8] Z. Yin, A. Sadok, H. Sailem, A. McCarthy, X. Xia, F. Li, M. A. Garcia, L. Evans, A. R. Barr, N. Perrimon, C. J. Marshall, S. T. C. Wong, and C. Bakal, Nat. Cell Biol. 15, 860 (2013).
[9] P.-H. Wu, J. M. Phillips, S. B. Khatau, W.-C. Chen, J. Stirman, S. Rossell, K. Tschudi, J. Van Patten, M. Wong, S. Gupta, A. S. Baras, J. T. Leek, A. Maitra, and D. Wirtz, Sci. Rep. 5 (2015), 10.1038/srep18437.
[10] C. Bakal, J. Aach, G. Church, and N. Perrimon, Science 316, 1753 (2007).
[11] V. K. Lam, T. C. Nguyen, B. M. Chung, G. Nehmetallah, and C. B. Raub, Cytometry Part A 93, 334 (2017).
[12] C. C. Caicedo1, S. Cooper, F. Heigwer, S. Warcha, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus, M. Wawer11, L. Paavolainen, M. D. Herrmann, M. Rohban, J. Hung, H. Hennig, J. Concannon, I. Smith, P. A. Clemons, S. Singh, P. Rees, P. Horvath, R. G. Linnington, and A. E. Carpenter, Nat. Methods 14, 849 (2017).
[13] E. Cai, K. Marchuk, P. Beemiller, C. Beppeler, M. G. Rubashkin, V. M. Weaver, A. Gérard, T.-L. Liu, B.-C. Chen, E. Betzig, F. Bartumeus, and M. F. Krummel, Science 356 (2017).
[14] C. D. Paul, P. Mistriotis, and K. Konstantopoulos, Nature Reviews Cancer 17, 131 (2017).
[15] R. J. Petrie and K. M. Yamada, J. Cell Science 125, 5017 (2012).
[16] V. Sanz-Moreno, G. Gadea, J. Ahn, H. Paterson, P. Marra, S. Pinner, E. Sahai, and C. J. Marshall, Cell 135, 510 (2008).
[17] S. Wilkinson, H. F. Paterson, and C. J. Marshall, Nat. Cell Biol. 7, 255 (2005).
[18] V. Sanz-Moreno and C. J. Marshall, Curr. Opin. Cell Biol. 22, 690 (2010).
[19] M. Pavlaki and S. Zucker, Cancer Metastasis Rev. 22, 177 (2003).
[20] K. Wolf, I. Mazo, H. Leung, K. Engelke, U. H. von Andrian, E. I. Deryugina, A. Y. Strongin, E. Bröcker, and P. Friedl, J. Cell Biol. 160, 267 (2003).
[21] R. J. Petrie, H. M. Harlin, L. T. Korsak, , and K. M. Yamada, J. Cell Biol. 216, 93 (2017).
[22] C. Cortes and V. N. Vapnik, Machine Learning 20 (1995), 10.1007/BF00994018.
[23] A. Ben-Hur, D. Horn, H. Siegelmann, and V. N. Vapnik, Journal of Machine Learning Research 2 (2001).
[24] L. Breiman, Machine learning 45, 5 (2001).
[25] M. Matsubara and M. J. Bissell, Oncotarget 7, 31602–31622 (2016).
[26] A. Daoud, U. Gopal, J. Kaur, and J. S. Isaacs, Oncotarget 8, 106807–106819 (2017).
[27] T. Yeung, P. C. Georges, L. A. Flanagan, B. Marg, M. Ortiz, M. Funaki, N. Zahir, W. Ming, V. Weaver, and P. A. Janmey, Cell Motil. Cytoskeleton 60, 24 (2005).
[28] L.J.P. van der Maaten, Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics JMLR W&CP 5, 384 (2009).
[29] G. J. Berman, D. M. Choi, W. Bialek, and J. W. Shaevitz, Journal of The Royal Society Interface 11, 20140672 (2014), https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2014.0672.
[30] K. Yoshida and T. Soldatli, J. Cell Science 119, 3833 (2006).
[31] J. B. Wyckoff, S. E. Pinner, S. Gschmeissner, J. S. Condeelis, and E. Sahai, Curr. Biol. 16, 1515 (2006).
[32] C. Battle, C. P. Broedersz, N. Fakhri, V. F. Geyer, J. Howard, C. F. Schmidt, and F. C. MacKintosh, Science 352, 604 (2016).
[33] J.-Y. Tinevez, U. Schütze, G. Salbreux, J. Roensch, J.-F. Joanny, and E. Paluch, Proceedings of the National Academy of Sciences 106, 18581 (2009).
[34] R. J. Petrie, N. Gavara, R. S. Chadwick, and K. M. Yamada, J. Cell Biol. 197, 439 (2012).
[35] M. R. Mejillano, S. Kojima, D. A. Applewhite, F. B. Gertler, T. M. Svitkina, and G. G. Borisy, Cell 118, 363 (2004).
[36] A. D. Doyle and K. M. Yamada, Exp. Cell Res. 343, 60 (2016).
[37] P. B. Gupta, C. M. Fillmore, G. Jiang, S. D. Shapira, K. Tao, C. Kuperwasser, and E. S. Lander, Cell 146, 633 (2011).
Supplementary information for Extracellular Matrix regulates the morphodynamics of 3D migrating cancer cells

Christopher Z. Eddy, Helena Raposo, Ryan Wong, Fuxin Li, Bo Sun

March 2, 2021
S1: Additional experimental details

3D cell culture

GFP-labeled MBA-MB-231 human breast cancer cells are purchased from GenTarget Inc. and are maintained according to the manufacturer’s instructions. To embed the cells in 3D collagen matrices, cells are suspended at very low density in neutralized collagen solutions. Generally, the suspension is then immediately transferred to glass bottom dishes (ibidi µ-dish) and incubated on either a warming plate set to 25°C, or in a tissue culture incubator (37°C, 5% CO2) for 30 minutes in order to solidify the matrix. For fiber alignment, a small magnetic iron shaving (<200 µm) is first placed onto the dish and then immersed in cell-collagen solution and placed onto the warming plate. We then drag the particle through the solution by an external magnet along a line for approximately 3 minutes while warming, and is then left to solidify (1). The cellularized ECM is then immersed with tissue culture medium and continuously incubated for 24 hours before imaging.

Microscopy and image analysis

Continuous imaging is done with a Leica TCS SPE confocal microscope equipped with a stage-top incubator (ibidi). Images are captured at a rate of 1 frame per 15 minutes. The raw images are gray scale with a resolution of 1024 x 1024 pixel². The voxel size has been calibrated to equal 0.538 µm. A single x-y plane is imaged every 10 µm in the z-dimension per experiment for up to 24 hours. Using custom Matlab scripts, cell images are maximum-projected onto a x-y plane and tracked over time (see section S2). The projected images are manually segmented, and screened to remove cells that are not entirely within the viewing window.

S2: Geometric characterization of cell images

Image processing

Following acquisition of fluorescent images, data regarding cell shape and position are obtained by processing and binarizing the time-lapse images using custom NIH ImageJ and Matlab scripts. First, fluorescence images are background subtracted using a rolling ball radius of 50 pixels (26.88 µm) and then log-transformed in order to make cell edges highly visible and so that less fluorescent-intense cells are also quantified. A manual threshold is then applied for each image. After, cells are manually segmented for each z-stack if applicable. Since consecutive z-stacks may have cell overlap, custom Matlab scripts are then used to determine if the same cell is in multiple z-stacks. After, we take a maximum projection (2D) of each cell. Geometrical measurements are then taken on binary objects using Matlab’s regionprops function, including area, perimeter, major axis length, minor axis length, solidity, eccentricity, convex area, extent, and equivalent diameter, convex perimeter, fiber length (skeletonized max length), maximum inscribed radius, and maximum bleb radius (maximum secondary circle). Additional measures of form factor ($\frac{\text{Area}}{\text{Perimeter}^2}$), aspect ratio ($\frac{\text{MajorAxisLength}}{\text{MinorAxisLength}}$), Convexity ($\frac{\text{Perimeter}}{\text{ConvexPerimeter}}$), Curl ($\frac{\text{MajorAxisLength}}{\text{FiberLength}}$), Perimeter Curl ($\frac{\text{Perimeter}}{\pi} (1 - \sqrt{1 - 4\pi \text{form factor}})$), Sphericity ($\frac{2\text{MaxInscribedRadius}}{\text{MajorAxisLength}}$), Incribed Area ($\frac{\text{MajorAxisLength}^2 \pi}{\text{MaxInscribedRadius}^2}$), and Bleb Ratio ($\frac{\text{MaxBlebRadius}}{\text{MaxInscribedRadius}}$) are subsequently calculated, totalling to 21 geometric measures. Collectively, this is shown schematically in figure 1.

Tracking cell position

In order to track the real-space center of the cell, we use maximum inscribed circle (MIC) of the cell image. Compared to imaging the nucleus directly, which causes phototoxicity and is prone to poly(nucleus), MIC does not require additional probes. To further demonstrate the accuracy of MIC, we compared short videos of dual labeled MDA-MB-231 cells where the GFP channel labels the cytoplasm and RFP channel labels the cell nucleus (SYTO-64, ThermoFisher). We find that MIC agrees very well with direct nucleus staining when determining the cell position, as shown in figure 2. For most of the frames, the deviation is less than 10% of the cell long axis. The root mean squared deviation is approximately 3 microns.
**Fig. 1.** Schematic of image processing and measures taken from binary using custom Matlab and Python scripts (scale-bar = 10 μm).

| MEASURE          | Quantity          |
|------------------|-------------------|
| Area             | 490.81 μm²        |
| Major Axis Length| 65.47 μm          |
| Minor Axis Length | 12.53 μm         |
| Eccentricity     | 0.98              |
| Convex Area      | 768.30 μm²        |
| Equivalent Diameter | 25.00 μm    |
| Solidity         | 0.64              |
| Extent           | 0.31              |
| Perimeter        | 167.31 μm         |
| Convex Perimeter | 146.46 μm         |
| Fiber Length     | 71.22 μm          |
| Max Inscribed Radius | 5.48 μm    |
| Bleb Length      | 42.18 μm          |
| Aspect Ratio     | 5.23              |
| Form Factor      | 0.02              |
| Convexity        | 1.14              |
| Perimeter Curl   | 10.51 μm          |
| Curl             | 0.89              |
| Sphericity       | 0.17              |
| Inscribed Area   | 2455.99 μm        |
| Rel. Bleb Length | 7.69              |

**Fig. 2.** Positional data acquisition: Position measurements of a cell determined by the nucleus centroid stained with SYTO-64 (red) and max-inscribed circle center (green), and histogram of square deviations between the two positional measures. All scale-bars are 20 μm.

**S3: ECM dimension modulates real space and shape space dynamics**

The morphodynamics of a cell display a strongly sub-diffusive characterization in shape-space. Shown in Table 1, we quantify the fits of mean square displacement of measures and report the power over a ten hour lag period.
Table 1. Power \( (n) \) of anomalous diffusion quantified by fitting the mean square displacements of shape and position measures for cells embedded in a 1.5 mg/mL collagen matrix prepared at room temperature (3D) and cells plated on top of similarly prepared collagen matrices (2D).

| Measure                      | \( n \) (3D) | \( n \) (2D) |
|------------------------------|--------------|--------------|
| Area                         | 0.6261       | 0.9812       |
| Major Axis Length            | 0.5761       | 0.8781       |
| Minor Axis Length            | 0.4338       | 0.4702       |
| Eccentricity                 | 0.3638       | 0.3234       |
| Convex Area                  | 0.5738       | 0.9230       |
| Equivalent Diameter          | 0.5999       | 0.9246       |
| Solidity                     | 0.4493       | 0.5794       |
| Extent                       | 0.4965       | 0.5208       |
| Perimeter                    | 0.5635       | 0.8353       |
| Convex Perimeter             | 0.6043       | 0.9636       |
| Fiber Length                 | 0.5981       | 0.8931       |
| Maximum-Inscribed Radius     | 0.5059       | 0.6059       |
| Bleb Length                  | 0.5125       | 0.8510       |
| Aspect Ratio                 | 0.4240       | 0.6419       |
| Form Factor                  | 0.4642       | 0.5326       |
| Convexity                    | 0.2645       | 0.4477       |
| Perimeter Curl               | 0.4790       | 0.5681       |
| Curl                         | 0.2515       | 0.2775       |
| Sphericity                   | 0.5533       | 0.5016       |
| Inscribed Area               | 0.5139       | 0.8617       |
| Relative Bleb Length         | 0.4667       | 0.6052       |
| Real-Space Migration         | 1.2197       | 1.3881       |

S4: Morphological Phenotype Analysis

Details of machine-learning and SVM

In order to classify cells into particular migration mechanisms, we used support-vector machine (SVM) learning (2, 3). As a maximal-margin classifier, SVM was particularly attractive as the overlap between different phenotypes was unknown. Also important is that cells can display multiple phenotypes at once and thus a soft-margin classifier was vital. Lastly, given our small dimensional space and small training-sets, SVM was an optimum choice for classification purposes.

Labeled data were first acquired as described below. Images were binarized and then geometrical data was obtained on labeled cells for training. We performed parameter grid search for RBF, Linear, and polynomial kernel models, with 10-fold cross validation to determine best performance. A grid search determined that an RBF kernel with \( \gamma = 5.8^{-3} \) and \( C = 5000 \) yields an average training, validation, and test set accuracy of 93.4%, 86.0%, and 94.3%, respectively. However, this model was only nominally improved from the RBF model used with \( \gamma = 0.01 \) and \( C = 1000 \) (average training, validation, and test set accuracy of 93.0%, 85.6%, and 94.3%, respectively). A linear model also recorded strong performance with \( C = 193.2 \) with average training, validation, and test set accuracy of 90.8%, 86.0%, and 94.3%, respectively. Although easier to interpret, the linear model was found to not consistently match supervised classification. For this reason, along with the slight increase in performance, we proceeded with the RBF kernel SVM model. The optimized SVM model performs at 92.3%, 91.8%, and 94.3% on the train, validation, and held out test sets discussed in the Random-forest classifier.

Details of training, validation, and test sets

Training images were processed as previously described. At least 280 images of each class were acquired to train machine learning networks, including the use of an optimized small shearing augmentation (randomly sheared with magnitude < 0.4) and rotation on images as an over-sampling technique to make our models robust to slight perturbations and noise, as well as to expand training examples. To evaluate true performance in these networks, an unseen test set was prepared with 35 images. The training and test sets are available on Figshare (4).
Comparisons with Random-Forest classifier

To compare performance to other multi-classification models, we have trained a Random-Forest classifier, tuning hyperparameters of depth, number of estimators, and features [5]. The optimized model uses 200 estimators, with a maximum depth of 220 splits, and uses no feature subset selection, yielding a bagged ensemble of regression trees. We perform 10-fold cross validation on 80% of training data and use 20% for validation data, utilizing class weighting to account for the class imbalance during training. We also include our 35 image held out test set for evaluation. The mean accuracy score was 91.3 ± 1.7%, with average recall scores of 100 and 93.9% on train and validation scores, respectively, and 82.1% on the held out test set. SVM and Random-Forest classifiers generally agree very well in terms of predictions, although SVM performs notably better on the held-out test set. The Random-Forest classifier disagrees with the SVM classifier on 232/3012 training examples (7.7%), 41/754 validation examples (5.4%), and 3/35 test set examples (8.6%).

Details of parametric t-SNE embedding algorithm

To visualize a high-dimensional morphological trajectories in three-dimensional space, a t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm is prepared for dimensionality reduction utilizing the geometric characterization of cells without requiring labels [6]. Since the t-SNE algorithm is a non-parametric model, it utilizes all available data to separate data clusters well, and cannot classify new data points without retraining and often leads to different cluster shapes. To embed new data points or entire trajectories in a consistent morphological space, a parametric t-SNE model is prepared. This model uses an Adam optimizer set to minimize the Kullback-Leibler divergence loss, and is trained with 800 iterations with a batch size of 256 examples, and the tunable perplexity parameter set to 30.0. The training set consists of 1024 examples, with equal numbers of samples from each phenotype. The model is made up of three fully connected layers with subsequent ReLU activation after each layer, with an additional final fully connected layer calculating the three t-SNE components for each points. We then use the model to transform the high-dimensional trajectories of new cells into the three-dimensional space.

S5: Cell Chemical Treatment

To demonstrate the effects of the Rho-ROCK pathway on cellular morphology, we chemically induce activation and inhibition of pathway proteins. Y27632, a specific inhibitor of Rho-associated protein kinases as well as ROCK-II activity [7–9], was purchased from Sigma-Aldrich and diluted to a working concentration at 3 µg/mL (0.1% v/v DMSO) in serum-free growth medium. Similarly, Rho activator II (CN03), known to robustly increase the level of GTP bound RhoA [10, 11], was purchased from Cytoskeleton and diluted to a working concentration at 2 µg/mL (0.1% v/v DMSO) in serum-free growth medium. 1.5 mg/mL gels were prepared as discussed in the main text with cells at a low concentration and incubated in serum-free growth medium for 6 hours. Cells were then imaged with confocal microscopy for up to 4 hours. Then, serum-free growth medium is removed from the dish and replaced with fresh serum-free growth medium containing the prepared chemical. We then proceed to immediately image for up to another 20 hours.

Figure 4 shows additional trajectories of cells that responded to chemical treatment of CN03. Generally, cells treated with CN03 exhibit morphologies characteristic of cell contraction. Notably, cells do not produce protrusions following treatment and typically retract protrusions post-treatment. Visually, this is seen as a notable slide toward the blebbing migration mode (green) or toward lamellipodial cell-spreading (blue).

Conversely, figure 3 shows additional trajectories of cells that responded to chemical treatment of Y27632. Cells treated with Y27632 show characteristic production of protrusions and/or sustained pre-existing protrusions. In the figure, this is shown by most trajectories moving toward filopodial migration (magenta) or toward actin-enriched leading edge migration (yellow).
**S6: ECM Characterization**

**Gel Autocorrelation**

In order to quantify the density fluctuations of collagen ECM, we compute the autocorrelation functions of confocal reflection images of collagen gels. Reflection images are first background subtracted using a rolling ball with radius...
of 50 pixels (26.88 µm). Images were then log-transformed to make fibers highly quantifiable. Images are then mean subtracted, and the autocorrelation is calculated. Following, the autocorrelation is normalized and then smoothed using interpolation. The results are shown in Figure 5. The spatial uniformity indicates an appropriate distribution of randomly oriented gel fibers. Additionally, these show that the decay in the autocorrelation is slower for higher density gels, an indication of the fiber quantity, and much faster for higher temperature gel, caused by the much shorter fiber lengths and smaller pores (12). The anisotropy of decay in the final graph indicates the direction of alignment.

![Fig. 5. Autocorrelations shown for 25°C 1.5 mg/mL Randomly oriented gel (upper left), 25°C 3.0 mg/mL Randomly oriented gel (upper right), 37°C 1.5 mg/mL Randomly oriented gel (lower left), and 25°C 1.5 mg/mL Aligned oriented gel (lower right).](image)

**Gel Coherency**

In order to determine the degree of local and global alignment of collagen fibers, we take the pre-processed confocal reflection images and use OrientationJ with 9-14 circular ROIS per image (packed without overlap), with ROI sizes ranging from 145 to 450 pixel diameter. At least 3 images (one per experiment) were used to quantify coherency. Empirically, larger ROIs were reliable to quantify global alignment, while smaller ROIs were better used for local alignment measures. Smaller ROIs than 145 pixels were not used, as it was noticed the coherency measured by OrientationJ can be highly biased for large fibers such as in 25°C gels (145 pixels | 78 µm| is approximately twice the average fiber length | 41 µm| in a 25°C gel). The results are shown in figure 6 and follows previous literature (13).

![Fig. 6. Coherency measurements taken using custom Matlab scripts utilizing OrientationJ plug-in for ImageJ (NIH). Measurements are for 1.5 mg/mL collagen gels prepared at room temperature, where we have utilized the alignment protocol to compare against typical randomly-oriented collagen gels.](image)
Rheology

Strain sweep rheology measurements were performed on the varying density gels with a Discovery Hybrid Rheometer-3 (TA instruments) at a 1 Hz frequency in a parallel plate geometry. A standard peltier plate (TA instruments) allowed gels to be formed at 37°C. Young’s modulus is shown in figure 7. As shown, storage moduli in the linear regime for gels with 1.5, 3.0, and 4.5 mg/mL collagen were about 70, 180, and 250 Pa, respectively.

![Fig. 7. Strain-sweep rheology measurements of collagen fiber networks gelled at 37°C using DHR-3 Rheometer (TA Instruments) in a parallel plate geometry at a 1 Hz frequency.](image)

S7: Additional details of morphological phenotype dynamics

Dwell time definition

The dwell time of state i is determined by using

$$D_{i \rightarrow i} = \frac{1}{1 - r_{i \rightarrow i}}.$$  

where $r_{i \rightarrow j}$ is the transition rate from state i back to itself, and $D_{i \rightarrow i}$ is the corresponding dwell time. We found this definition of dwell time maximizes the usefulness of data and is more robust to experimental shortfalls that affect the naive method of simply counting the number of frames a cell remains in the same phenotype for.

Details of transition rate calculations

Assuming the probability distribution of cell morphological phenotypes follow a Boltzmann distribution, then the probability, $P_i$, given by

$$P_i = \frac{1}{1 + \sum_{j \neq i}^{N} \frac{r_{i \rightarrow j}}{r_{j \rightarrow i}}}$$

where $r_{i \rightarrow j}$ is the transition rate from state i to state j, and conversely $r_{j \rightarrow i}$ is the transition rate from state $j$ to state $i$. The transition matrix is thus calculated as probability per unit time, and hence each row in the transition matrix will sum to 1

$$(r_{i \rightarrow i} - 1) + \sum_{j}^{N} r_{i \rightarrow j} = 1.$$  

The transition rate $r_{i \rightarrow j}$ is simply calculated by counting the number of transitions from state i to state j, and then dividing by all observed transitions from state i. Using this method, we find that the transition rates are stable,
regardless of the length of trajectories in computational experiments.

Following SVM classification, phenotype dynamics could be properly drawn from data. Importantly, where maximum decisions by the SVM classifier do not exceed 60%, the classification is thus determined to an intermediate between two states. The intermediate state can be a chimera of two states, with most occurrences being as a cell transitions between morphological phenotypes. Because this state is not considered to be unique, we calculate the transition rate from state $i$ that passes through $N$ intermediate states prior to state $j$ as $1/N_{frames}$. We find by simulation that this method can most-accurately recover transition rates in comparison to methods using soft-max or ignoring intermediate states.

**Probability flux calculations**

To investigate broken detailed balance in morphological phenotype space, we report probability flux calculations shown in figure 8. The probability flux from state $i$ to state $j$ is given by

$$\Delta \phi_{i \rightarrow j} = \frac{N_{i \rightarrow j}}{\sum_k \sum_l N_{k \rightarrow l}}$$

where $N_{i \rightarrow j}$ is the number of transitions from state $i$ to state $j$, and both summations in the denominator are over all available states. Therefore, the net probability flux between states $i$ and $j$ is given by $\Delta \phi_{i \rightarrow j} - \Delta \phi_{j \rightarrow i}$. We find that the net flux between states are all less than 0.003 probability difference per hour for cells in any ECM condition we tested. To reveal if any small difference in probability flux may be significant, we also report the probability flux percent difference between states $i$ and $j$ is given by

$$\frac{\Delta \phi_{i \rightarrow j} - \Delta \phi_{j \rightarrow i}}{\Delta \phi_{i \rightarrow j} + \Delta \phi_{j \rightarrow i}}$$

We find that the maximum net probability flux percent difference for cells in any ECM condition evaluated is less than 9%.

![Figure 8](image_url)

**Fig. 8.** (A-D) Net probability flux calculations under varying ECM conditions. (E-H) Percent difference of net probability flux under varying ECM conditions. (A, E) collagen ECM prepared at room temperature (RT) and $[\text{col}] = 1.5$ mg/mL. (B, F) collagen ECM prepared at RT and $[\text{col}] = 3.0$ mg/mL. (C, G) collagen ECM prepared at 37 °C and $[\text{col}] = 1.5$ mg/mL. (D, H) collagen ECM prepared at RT and $[\text{col}] = 1.5$ mg/mL with flow-aligned collagen fibers.
**S8: Motility analysis of morphological phenotypes**

In order to determine the migrational persistence of migration modes, we report first the velocity autocorrelation observed for cells in collagen ECM prepared at room temperature (RT) and \([col] = 1.5 \text{ mg/mL}\). We find that the autocorrelation quickly decays to zero in a single time step (15 minutes) shown in Figure 9(A). We additionally checked the \(\cos(\theta)\) distribution between consecutive steps and found steps that the concurrent steps seem to be taken randomly in direction, as given by the U-shaped distribution of Figure 9(B).

![Figure 9](image.png)

**Fig. 9.** (A) Velocity autocorrelation and (B) binned directionality given by \(\cos(\theta)\) between consecutive velocity vectors for cell migration in collagen ECM prepared at room temperature and \([col] = 1.5 \text{ mg/mL}\). (Bin width = 0.1)

A soft-max classification strategy was used to determine step-sizes dominated by a single migration mode and to avoid confounding step sizes from classifications near the decision function boundary. To evaluate the motility of various migration mode transitions, step sizes were separated into components parallel and perpendicular to the direction of the previous step. Both directions had step-size distributions that were approximately Gaussian. Figure 10(A) shows the binned step size distribution in the persistent direction fit with a Gaussian distribution. We find that these fits yield mean step sizes close to zero and small variances, which are not concurrent with the mean-square displacements (MSD) of RT data shown in Figure 5B. Rather, we find more suitable fits are obtained by fitting the magnitude of the step sizes, which follow a log-normal distribution as shown in figure 10(B). These fits result in step distributions that yield similar MSD values to RT data. The parameters used to generate the log-normal fits are calculated from fitting the empirical cumulative distribution of step size magnitudes with a log-normal CDF, shown in figure 11.
Fig. 10. (A) Gaussian fitting (red) of binned step sizes of in the persistent direction (blue) [bin width = 1 µm] for various migration mode transitions. (B) Log-normal fitting (red) of binned step size magnitudes (blue) using the parameters shown in Figure 5A to generate the log-normal PDF [bin-width = 0.5 µm] for various migration mode transition.

Fig. 11. Empirical cumulative distribution (blue) and log-normal cumulative distribution function fit (red) of step sizes for transitions between amoeboid (AM) and mesenchymal (ME) modes.

S9: Interface calculations

Experiment Details

Interface experiment was done in triplet. Briefly, the outer gel was first made by gelling cells in collagen solution (1.5 mg/mL neutralized) at 25°C for 20 minutes on the DIGME stage (14), warmed at 25°C. After, the needle was gently removed and a new ice-cold collagen solution (1.5 mg/mL neutralized) containing cells was then dripped into the hole, gently swirled, and then placed into the incubator at 37°C for 15 minutes. After, 3 mL of growth medium was added on top and sat for 24 hours before imaging. Imaging was taken near the interface, imaging bright-field, fluorescence (green), and back-reflection confocal images. Using the back-reflection images, the interface was manually traced out through the z-stack. The distance to the interface from cell centroids in corresponding z-stacks were then measured away from the closest marked interface point using Matlab bwdist. The frequency vs distance is shown in Figure 12, indicating a large number of cells were images close to the interface.
Fig. 12. Frequency of cell locations away from interface (at 0). Negative is in 37°C gel, Positive indicates 25°C gel.

Details of interface analysis

Migration mode transitions were first determined by using continuous trajectories accounting properly for intermediate state classification, and then distances away from the interface were determined by the initial state location. A 1-D Gaussian kernel was used to acquire continuous local spatial probability densities of transitions (per hour), centered every 5.376 µm (10x the distance-to-pixel ratio) with a standard deviation of 26.88 µm. This yields the probability density (per hour) of observing a transition \( P(i \rightarrow j, x) \) at a location \( x \), as used in the main text. To mitigate the bias from non-uniform cell density, we then divide the prior probability by the spatial probability of observing a cell within the spatial window centered at \( x \), \( M(x) \). \( M(x) \) is calculated using the same Gaussian kernel.

[1] Jihan Kim, Yu Zheng, Amani A. Alobaidi, Hanqing Nan, Jianxiang Tian, Yang Jiao, and Bo Sun. Geometric dependence of 3d collective cancer invasion. *Biophysical Journal*, 118(5):1177 – 1182, 2020. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2020.01.008. URL http://www.sciencedirect.com/science/article/pii/S000634952030031X.

[2] Corinna Cortes and Vladimir N. Vapnik. Support-vector networks. *Machine Learning*, 20(3), 1995. doi: 10.1007/BF00994018.

[3] Asa Ben-Hur, David Horn, Hava T. Siegelmann, and Vladimir Vapnik. Support vector clustering. *J. Mach. Learn. Res.*, 2:125–137, March 2002. ISSN 1532-4435.

[4] Christopher Eddy. MDA-MB-231 Morphology Training and Test Sets. *FigShare*, 10 2020. doi: 10.6084/m9.figshare.13125086.

[5] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[6] Gordon J. Berman, Daniel M. Choi, William Bialek, and Joshua W. Shaevitz. Mapping the stereotyped behavior of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99):20140672, 2014. doi: 10.1098/rsif.2014.0672. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2014.0672.

[7] Renaud Poincloux, Olivier Collin, Floria Lizárraga, Maryse Romao, Marcel Debray, Matthieu Piel, and Philippe Chavrier. Contractility of the cell rear drives invasion of breast tumor cells in 3d matrigel. *Proceedings of the National Academy of Sciences*, 108(5), 2011. doi: 10.1073/pnas.1010396109.

[8] Simon Gordonov, Mun Kyung Hwang, Alan Wells, Frank B. Gertler, Douglas A. Lauffenburger, and Mark Bathe. Time series modeling of live-cell shape dynamics for image-based phenotypic profiling. *Integrative Biology*, 8(1), 2015. doi: 10.1039/c5ib00283d.

[9] Masahiro Matsubara and Mina J Bissell. Inhibitors of rho kinase (rock) signaling revert the malignant phenotype of breast cancer cells in 3d context. *Oncotarget*, 7(22), 2016. doi: 10.18632/oncotarget.9395.

[10] Meng-Horng Lee, Pei-Hsun Wu, Daniele Gilkes, Ivie Aifuwa, and Denis Wirtz. Normal mammary epithelial cells promote carcinoma basement membrane invasion by inducing microtubule-rich protrusions. *Oncotarget*, 6(32):32634–32645, 2015. ISSN 1949-2553. doi: 10.18632/oncotarget.4728.

[11] Abdelkader Daoud, Udhayakumar Gopal, Jasmine Kaur, and Jennifer S Isaacs. Molecular and functional crosstalk between extracellular hsp90 and ephrin a1 signaling. *Oncotarget*, 8(63), 2017. doi: 10.18632/oncotarget.22370.

[12] Christopher Allen Rucksack Jones, Long Liang, Daniel Lin, Yang Jiao, and Bo Sun. The spatial-temporal characteristics of type i collagen-based extracellular matrix. *Soft Matter*, 10(44):8855–8863, November 2014. ISSN 1744-683X. doi: 10.1039/c4sm01772b.

[13] T. D. Clemons, M. Bradshaw, P. Toshniwal, N. Chaudhari, A. W. Stevenson, J. Lynch, M. W. Fear, F. M. Wood, and K. Swaminathan Iyer. Coherency image analysis to quantify collagen architecture: implications in scar assessment. *RSC Adv.*, 8:9661–9669, 2018. doi: 10.1039/C7RA12693J. URL http://dx.doi.org/10.1039/C7RA12693J.

[14] Amani A. Alobaidi and Bo Sun. Probing three-dimensional collective cancer invasion with digme. *Cancer Convergence*, 1(1), 2017. doi: 10.1186/s41236-017-0004-9.