

Self-generated persistent random forces drive phase separation in growing tumors

Cite as: J. Chem. Phys. **153**, 201101 (2020); <https://doi.org/10.1063/5.0026590>

Submitted: 24 August 2020 . Accepted: 01 November 2020 . Published Online: 23 November 2020

 Sumit Sinha, and D. Thirumalai



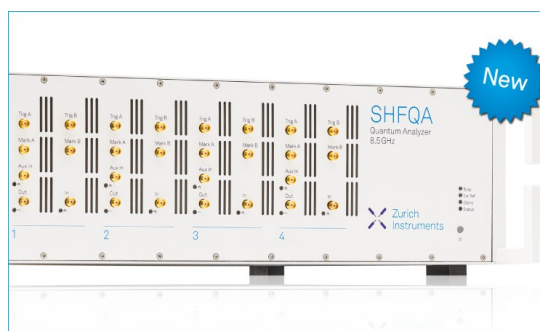
View Online



Export Citation



CrossMark



Your Qubits. Measured.

Meet the next generation of quantum analyzers

- Readout for up to 64 qubits
- Operation at up to 8.5 GHz, mixer-calibration-free
- Signal optimization with minimal latency

Find out more



Self-generated persistent random forces drive phase separation in growing tumors

Cite as: J. Chem. Phys. 153, 201101 (2020); doi: 10.1063/5.0026590

Submitted: 24 August 2020 • Accepted: 1 November 2020 •

Published Online: 23 November 2020



Sumit Sinha¹  and D. Thirumalai^{2,a)}

AFFILIATIONS

¹Department of Physics, University of Texas at Austin, Austin, Texas 78712, USA

²Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, USA

^{a)} Author to whom correspondence should be addressed: dave.thirumalai@gmail.com

ABSTRACT

A single solid tumor, composed of nearly identical cells, exhibits heterogeneous dynamics. Dynamics of cells in the core is glass-like, whereas those in the periphery undergoes diffusive or super-diffusive behavior. Quantification of heterogeneity using the mean square displacement or the self-intermediate scattering function, which involves averaging over the cell population, hides the complexity of the collective movement. Using the t-distributed stochastic neighbor embedding (t-SNE), a popular unsupervised machine learning dimensionality reduction technique, we show that the phase space structure of an evolving colony of cells, driven by cell division and apoptosis, partitions into nearly disjoint sets composed principally of the core and periphery cells. The non-equilibrium phase separation is driven by the differences in the persistence of self-generated active forces induced by cell division. Extensive heterogeneity revealed by t-SNE paves the way toward understanding the origins of intratumor heterogeneity using experimental imaging data.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0026590>

Intratumor heterogeneity (ITH), a pervasive phenomenon across cancers, is a major hurdle in developing effective treatment.^{1,2} ITH refers to the coexistence of genetically or phenotypically distinct cells within a single tumor.³ A source of ITH is genetic variations. Indeed, multi-region sequencing has revealed widespread genetic diversity within tumors.^{4–9} Stochastic variations due to differences in cancer microenvironment, which results in vastly different dynamics of cells in distinct regions of an evolving solid tumor, could also give rise to ITH. Evidence for the dynamically driven ITH has emerged recently from several imaging studies, which have mapped out the phenotypic properties (such as the shape and size) in three dimensional tumor spheroids.^{10–13} The growth of tumor spheroids is monitored by embedding them in a collagen matrix.^{10–14} Direct imaging reveals that the dynamics of cells in the tumor core differ dramatically compared to cells in the periphery,^{11–13,15} a clear signature of dynamical ITH, which we abbreviate as DITH. A characteristic of DITH is that the material properties of the cells are unaltered, implying that heterogeneity arises solely from microenvironment fluctuations. In this sense, DITH is reminiscent of dynamic heterogeneity in supercooled liquids that undergo glass transition.^{16,17}

Previously, we showed that the cell dynamics in a growing multicellular spheroid (MCS) is spatially heterogeneous,^{15,18} which

implies that cells in the core (periphery) exhibit sub-diffusive (super-diffusive) motion. These characteristics were first observed in imaging experiments tracking the displacement of cells moving in a collagen matrix and recently in other studies as well.^{12,13} However, characterizing the dynamics using conventional ensemble average measures, such as mean squared displacement or the self-intermediate scattering function, hides the rich dynamics, the cause of DITH.

Can we infer DITH directly from the cell trajectories in an evolving tumor? Computer simulations of physical models for evolving cells and, more importantly, direct imaging can be used to generate the needed trajectories. Here, we show that the t-distributed stochastic neighbor embedding (t-SNE), a popular unsupervised machine learning technique for analyzing big data, is ideally suited to answer the question posed above. The t-SNE method is among the best dimensionality reduction technique,^{19–22} allowing us to visualize the emergent heterogeneous dynamics *without any inherent bias* in the trajectory analysis. It has been extensively used in various areas ranging from genomics^{23,24} and neuroscience²⁵ to condensed matter physics.^{26–29}

We performed t-SNE on data generated using simulations of an expanding tumor spheroid model.^{15,18,30,31} The results revealed

massive dynamical heterogeneity that depends on the radial distance from the tumor center, which accords well with the conclusions in recent experiments.^{11–13} t-SNE also resolves the dynamical phase space structure of cells in the core and periphery. Division of the dynamical phase space structure primarily into two disjoint sets is a consequence of differences in the persistence of self-generated active force (SGAF), which is dynamically generated in our model due to an imbalance in the cell division and apoptosis rates. The cells in the periphery experience highly persistent forces that are predominantly pointed in the radially outward direction.

t-SNE: For completeness, we provide a brief description of the t-SNE method.^{19–22} Let us consider a n dimensional vector, $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, which, in our case, is the time dependent cell displacement or forces experienced by the cells. Here, n is the cell identity. The components of $\mathbf{x}_i$ s are $\{x_{i,1}, x_{i,2}, \dots, x_{i,D}\}$ with $D \gg 1$. For our case, $D = 1080$ (explained later). The t-SNE projects $\mathbf{x}_i$ s onto a low dimensional (usually 2 or 3) space $\mathbf{y}_i$ s while being faithful to the information content in the high dimensional space.

To determine the $\mathbf{y}_i$ s, a joint probability p_{ij} in high dimensional space, measuring the likelihood that points $\mathbf{x}_i$ and $\mathbf{x}_j$ are close to each other, is constructed. Following the standard practice, we take $p_{ij} = \frac{p_{ji} + p_{ji}}{2n}$, where the conditional probabilities $p_{ji} \propto \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_i^2})$, where $\|\cdot\|$ is a measure of distance, $\sum_{i,j} p_{ij} = 1$, and p_{ii} is set to zero. The variance σ_i^2 is chosen such that the perplexity ($\mathcal{P}_i$) of the distribution is given by

$$\mathcal{P}_i = 2^{-\sum_{j \neq i} p_{ji} \log_2 p_{ji}}. \quad (1)$$

The perplexity is independent of i ($\mathcal{P}_i = \mathcal{P}$). The maximum perplexity can be $(n - 1)$, resulting in $\sigma_i = \infty$, which would lead to a uniform distribution (i.e., $p_{ji} = \frac{1}{n-1}$). The perplexity value, which can be interpreted as the number of effective neighbors, influences the outcome of t-SNE.³²

The joint probability q_{ij} , measuring the likelihood that points $\mathbf{y}_i$ and $\mathbf{y}_j$ are in proximity, is t-distributed (i.e., $q_{ij} \propto [1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2]^{-1}$) with $q_{ii} = 0$ and $\sum_{i,j} q_{ij} = 1$. To compute the $\mathbf{y}_i$ s, we use the Kullback-Leibler divergence ($\mathcal{L} = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$) as a loss function ($\mathcal{L}$) to minimize the difference between p_{ij} and q_{ij} . We determine $\mathbf{y}_i$ s by minimizing $\mathcal{L}$ using a gradient method. The gradient minimization is numerically implemented using the updating scheme,

$$\mathbf{y}_i[t] = \mathbf{y}_i[t-1] + \eta \frac{\partial \mathcal{L}}{\partial \mathbf{y}_i} + \alpha(t)(\mathbf{y}_i[t-1] - \mathbf{y}_i[t-2]). \quad (2)$$

In Eq. (2), η is the learning rate and $\alpha(t)$ is the momentum term that is included to speed up the optimization. The $\mathbf{y}_i[0]$ s are sampled from a normal distribution of mean 0 and variance 0.0001. In order to quantify the accuracy of tSNE plots, we made sure that $\mathcal{L} \rightarrow 0$.

The parameters in the t-SNE algorithm are $\mathcal{P}$, η , the momentum (α), and the number of iterations. We used $\eta = 200$, $\alpha(t) = 0.5$ for $t \leq 250$, and $\alpha = 0.8$ for $t \geq 250$. We performed 2000 iterations. The perplexity is varied depending on the situation. We projected three large datasets onto two dimensions with coordinates tSNE1 and tSNE2.

Position data: We collected the time traces of ≈ 5000 cells [$\mathbf{r}(t) = x(t), y(t), z(t)$] between time interval $T_{w1} = \tau \leq t \leq 11\tau$, where $\tau = 15$ h represents the cell cycle time (see Refs. 15 and 18

for details). The cell positions were recorded every 500 s. The sampling rate was chosen to roughly mimic the frame rate (one per 14 min) of microscopy measurements in experiments.^{11,13} The trajectory obtained from simulations was divided into 1080 ($\frac{T_{w1}}{500\text{ s}}$) time windows. Each time window can be thought of as a dimension. Therefore, the trajectories of each cell resides in 1080 dimensions. In each time window t_i , a cell is displaced by $|\delta x(t_i)| = |x(t_{i+1}) - x(t_i)|$ along the x-coordinate and similarly along y and z coordinates. Here, $|\cdot|$ represents the absolute sign. Thus, for each cell, we have 1080 ($t_i, |\delta x_i|$) pairs. One can choose any of the three (i.e., x, y, or z) coordinates. However, for our purpose, we used the x coordinate.

Before applying the t-SNE, for each cell, 1080 $|\delta x_i|$ s were sorted from the smallest to the largest value. Sorting is an important aspect of using tSNE on cell trajectories, which relies on the following argument. Imagine two cells (A and B) in the same region (region here means the distance from the tumor center) of the tumor. One expects that their dynamics would be similar. However, a naive application of tSNE method might lead to the opposite result. The reason is that the conditional probabilities in tSNE are based on the Euclidean distance $\|\delta x_A - \delta x_B\| = \sum_{i=1}^D (\delta x_{A,i} - \delta x_{B,i})^2$. If the data are not sorted, cells in the same region may not have small $\|\delta x_A - \delta x_B\|$, which is avoided if the data are curated.

Force data: We also used time traces of forces [$\mathbf{F} = (F_x, F_y, F_z)$] on individual cells (≈ 5000 cells) in the t-SNE analysis. Forces, with F_x, F_y and F_z , were recorded every ≈ 10 min between time interval $T_{w1} = \tau \leq t \leq 11\tau$.

Interpenetration data: This dataset contains the interpenetration distances for ≈ 5000 cells that were present in the simulation for time interval T_{w1} . The interpenetration of the i th cell at time t_k is given by $h_i(t_k) = \frac{1}{NN(i, t_k)} \sum_{j=1}^{NN(i, t_k)} h_{ij}(t_k)$, where $h_{ij}(t_k) = \max\{0, R_i(t_k) + R_j(t_k) - |\mathbf{r}_i(t_k) - \mathbf{r}_j(t_k)|\}$. Here, $R_j(t_k)$ ($\mathbf{r}_j(t_k)$) is the radius (position) of the j th cell at time t_k . $NN(i, t_k)$ is the number of nearest neighbors of the i th cell at time t_k . For each cell, we have 1080 (t_k, h_k) pairs.

Results: Figure 1(a) shows the clustering obtained when the trajectories of cells [1080 ($t_i, |\delta x_i|$)] are projected onto tSNE1 and tSNE2 ($\mathcal{P} = 100$). In Fig. 1, each dot represents a single cell and is colored depending on its distance from the center of the tumor, R_c . It should be emphasized that in performing t-SNE, we did not use the information of the cell distance from the tumor center. The colors aid in visualizing the cells. The results in Fig. 1(a) show that there is a pattern in the way the dots are arranged. Majority of the red dots (cells farthest from the tumor core) are at one end, with blue dots at the other end (cells closer to the core). In the other words, there is a dynamic phase separation, which we show below is a consequence of cell division and apoptosis. The partitioning into two disjoint patterns in Fig. 1(a) implies that the dynamics of the cells is dependent on their distance from the center of the tumor, as noted in experiments.^{11,12} However, the boundary between the two regions (roughly high and low density) is not sharp. The t-SNE method, based on machine learning, is able to delineate massive heterogeneity in a single tumor with identical cells, which is hidden in observables, such as the mean squared displacement or self-intermediate scattering function.^{15,18} Thus, unbiased analyses of the cell trajectories are required to shed light on the origin of DITH in solid tumors.^{16,33,34}

In a recent experiment,¹² the solid tumor was divided into two core and periphery regions. It was shown that the tumor

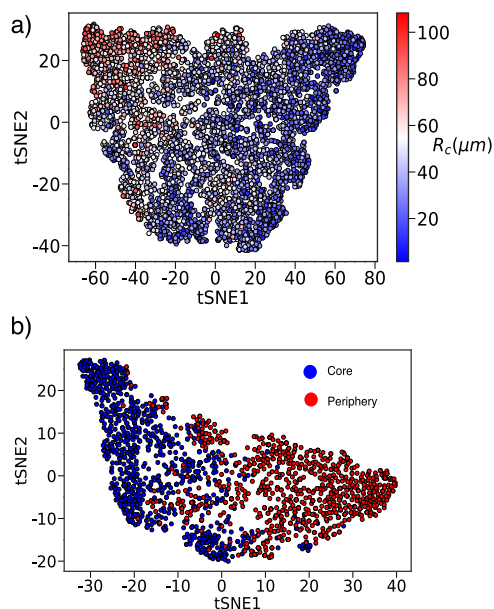


FIG. 1. Heterogeneity is radially dependent on an evolving tumor spheroid. (a) Cell trajectories (~ 5000) projected onto tSNE1 and tSNE2. Each dot represents a cell. The cell label depends on the radial distance from the center of the tumor (R_c). The color gradient is suggestive of extensive dynamical heterogeneity. The cells in blue represents those closer to the center, whereas cells in red are farther away from the center (see the R_c scale on the right). (b) Cell trajectories sampled from the core (blue dots) and periphery (red dots). There are 1580 (946) core (periphery) cells. Projection of cell trajectories onto t-SNE coordinates shows that the cells in the two regions have a resolvable dynamical phase space structure, by which we mean that the data cluster according to the motilities of the cells. This is easily seen from the separation of the red and the blue dots.

core (periphery) exhibits sub-diffusive (super-diffusive) dynamics, implying that the cells explore distinct non-overlapping regions of phase space dynamically. In order to assess if t-SNE separates the dynamics in the two regions, we collected position data of core ($R_c < 30 \mu\text{m}$) and periphery ($R_c > 60 \mu\text{m}$). We applied the SNE algorithm on this mixed dataset. Figure 1(b) shows the t-SNE clustering of cells belonging to core (blue dots) and periphery (red dots; $\mathcal{P} = 100$). It is clear from Fig. 1(b) that the red and blue cells are approximately phase separated. The distinct dynamical phase space explored by the cells in the core and periphery, as illustrated by Fig. 1(b), sheds light on the non-equilibrium phase separation between the tumor core and periphery.¹²

Phase separation of cells into the core and periphery is a consequence of self-generated active forces (SGAFs) arising from cell division. The SGAF is spatially dependent, leading to distinct cell motility in the core and periphery. In order to understand the dynamical phase space structure of cells predicted by t-SNE, we probed the nature of forces exerted on the cells. We first calculated force–force persistence of the i th cell, $FF_i(t) = \frac{\mathbf{F}_i(t+\delta t) \cdot \mathbf{F}_i(t)}{|\mathbf{F}_i(t+\delta t)| |\mathbf{F}_i(t)|}$, where $\mathbf{F}_i(t)$ ($|\mathbf{F}_i(t)|$) is the force (force magnitude) on the i th cell at time t and $\delta t = 0.05 \tau$ or 40 min. $FF_i(t)$ is a measure of force persistence an i th cell experiences and takes on values $[-1, 1]$. If $FF(t) = 1$ ($FF(t) = -1$), a cell experiences force in the same (opposite) direction at time t and

$t + \delta t$. We calculated $FF_i(t)$, with $\tau \leq t \leq 11\tau$, for all the cells that belong to the core and periphery and performed t-SNE analysis. The t-SNE projection of the $FF(t)$ data in Fig. 2(a) reveals contrasting force persistence for cells in the core and periphery. $FF(t)$ in the two regions partition into two disjoint sets, which is vividly illustrated in Fig. 2(a). The distinct behavior of force persistence is indicative of the super-diffusive and sub-diffusive behavior of cells in the periphery and core, respectively.^{11,12} The contrasting force behavior in the core and periphery is intrinsically related to spatial propensity for cell division. The increased stress (due to jamming) in the core

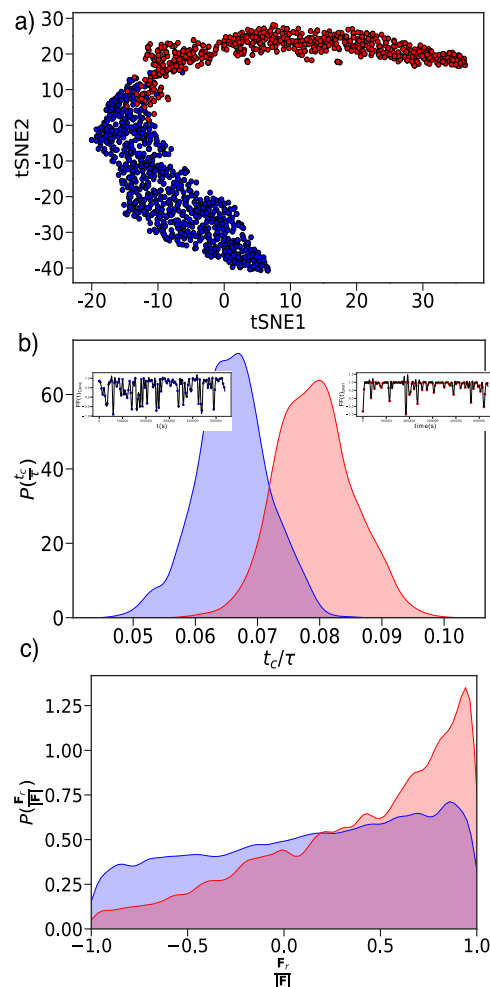


FIG. 2. Self-generated force due to cell division and apoptosis. (a) t-SNE projection of $FF(t)$ for cells in the core (blue) and periphery (red). The cells in the two regions cluster into distinct regions. (b) Distribution of t_c/τ obtained using Gaussian process regression for cells in the core (blue) and periphery (red). Cells in the periphery have higher persistence times (mean value is $t_c/\tau = 0.08$) compared to cells in the core (average is $t_c/\tau = 0.06$). Inset in the left (right) shows the $FF(t)$ fit using GPR for cells in the core (periphery). The blue (red) dots correspond to cells in the core (periphery) and the black solid line is the GPR fit. (c) Probability distribution of $F_r/|\mathbf{F}|$ for cells in the core (blue) and periphery (red). Cells in the periphery experience force predominantly in the radial direction.

suppresses cell division, whereas cells on the periphery can readily divide.^{35,36} This imbalance in cell division in the two regions leads to contrasting force persistence.

We calculated the $\frac{t_c}{\tau}$ distribution to quantify how long SGAF is persistent in the two regions. In order to extract $\frac{t_c}{\tau}$, we fit $FF_i(t)$ using Gaussian process regression (GPR) with the standard radial-basis function (RBF) kernel [$k(t, t') \sim e^{-\frac{(t-t')^2}{2\sigma^2}}$]. Figure 2(b) shows that the distribution of $\frac{t_c}{\tau}$ for cells in the core and periphery is resolvable. Furthermore, we find that the mean persistence time in the periphery ($\frac{t_c}{\tau} = 0.08$) is greater than that in the core ($\frac{t_c}{\tau} = 0.06$). Increased persistence of forces in the periphery results in greater directed movement. Insets in Fig. 2(b) show the FF(t) fit using GPR for one cell in the core and periphery. Experiments have noted that the cells in the periphery move predominantly radially outward.¹¹ Therefore, we calculated the radial force $\frac{F_{ir}}{|F_i|}$ exerted on the cells in the two regions. Here, $F_{i,r} = F_i \cdot \hat{r}_i$, where $\hat{r}_i$ is the radial unit vector, $\hat{r}_i = \frac{\mathbf{r}_i - \mathbf{r}_{com}}{|\mathbf{r}_i - \mathbf{r}_{com}|}$, and $\mathbf{r}_{com}$ is the center of mass of the tumor. If $\frac{F_{ir}}{|F_i|} = 1$, the force is radially directed outward, whereas $\frac{F_{ir}}{|F_i|} = -1$ implies inwardly directed force. The probability distribution of $\frac{F_{ir}}{|F_i|}$ for cells in the core is predominantly skewed more toward unity [Fig. 2(c)] than the core cells. The radially outward force explains the invasive characteristics of the cells at the tumor boundary.¹¹ These forces originate solely due to the imbalance in cell division in the core and peripheral region of the tumor. Cell division as a source of active stress has been reported before,³⁷ but the emergence of the highly persistent nature of forces adds new insights into the physics of tumor expansion.

Armed with the unbiased identification of the cells in the core and the periphery, we set out to find out if these features are manifested in other characteristics as well. Experiments have established that the core cells are tightly packed or jammed and have smaller volume compared to cells in the periphery.¹³ Therefore, we expected that the inter-cellular distance can be used to differentiate between the cells in the two regions. We recorded the interpenetration (h_{ij}) distances (h_{ij} s) for cells in the time window $T_{w1} = \tau_{min} < t < 11\tau_{min}$. Figure 3 shows the result of t-SNE clustering based on h_{ij} data for cells in the core and periphery ($P = 50$). To our surprise, the t-SNE

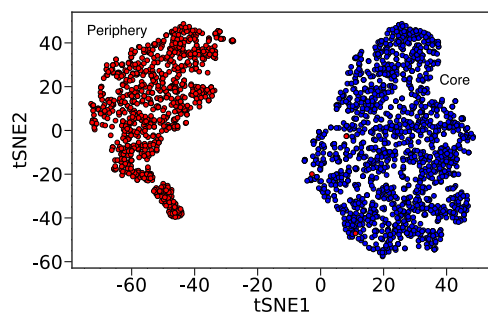


FIG. 3. Core and periphery cells have resolvable interpenetration. Interpenetration, h_{ij} (defined in the text), based classification for core and periphery cells using t-SNE. Cells in the core (periphery) are represented using blue (red) dots. The t-SNE algorithm resolves the interpenetration data for cells in the core and periphery.

algorithm clustered the cells remarkably well with clear phase separation between the cells in the core and the periphery. These results imply that the density in the two regions differ greatly because the interior cells are jammed, whereas the motility of the cells near the periphery is high. The emergence of the radially dependent density, with jamming in the interior, is consistent with the experiments, which show that pressure is higher in the core than at the periphery.³⁵

In order to provide a geometrical interpretation of the SGAF-driven phase separation, we followed Merkel and Manning³⁸ who predicted that $S = \frac{A}{V^{2/3}}$ could serve as an order parameter for the rigidity transition in 3D confluent tissues. The variables A and V are, respectively, the surface area and volume of the cell. The rigidity transition occurs at $S = 5.41$ in three dimensions (see Ref. 39 for results in two dimensions). Because the core (periphery) is solid-like (fluid-like), we expected that the shape parameter would reveal the observed differences in the motilities within a *single* tumor. We calculated the Voronoi volume and the area of the cells in at time $\approx 11\tau$. We excluded the cells at the boundary as their Voronoi volume is not defined. Figure 4 shows the distributions of the shape parameter distribution in the interior and the periphery. Remarkably, the distribution for cells in the core is narrow with a peak at 5.41, close to the predicted³⁸ solid to fluid transition value for confluent tissues. However, the packing of cells in the interior in our simulations is qualitatively different and does not reach confluency. In contrast, the distribution in the boundary broadly peaked with a mean of around 5.6. The inset in Fig. 4 shows that the cells in boundary have a bigger Voronoi volume as compared to cells in the core. We should emphasize that the variations in the shape parameter are observed in a tumor in which the low motile and high motile cells are simultaneously present. It appears that the shape parameter is good predictor of the transition from a jammed to a motile (super diffusive) state even in a continuously growing tumor whose dynamics is determined by cell division and apoptosis.

We used the unsupervised clustering technique (t-SNE)^{19,21,22} to elucidate the extent of heterogeneity in an evolving solid tumor consisting of nearly identical cells. The unbiased t-SNE analysis of

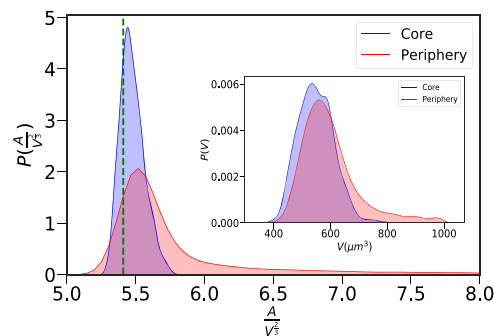


FIG. 4. Distribution of the shape parameter in the core and the periphery. Blue curve shows the distribution for the cells in the core and red indicates the cells in the periphery. The green line demarcates the solid to fluid boundary at $\frac{A}{V^{2/3}} = 5.41$. The inset shows the distributions of the Voronoi volume of cells in the core and the periphery.

the simulation data shows unambiguously that the dynamical behavior of cells in a growing tumor spheroid depends on the distance from the tumor core. The gradual change in the dynamical behavior from a jammed state in the tumor interior to highly motile (super-diffusive) behavior at the periphery is due to the generation of self-generated persistent forces that arises dynamically due to the inequality between cell division and apoptosis rates. The t-SNE method resolves the dynamical phase space structure of identical cells, revealing a plausible mechanism for non-equilibrium phase separation.¹² Our results, establishing dynamic heterogeneity in a single tumor consisting of nearly identical cells, imply that average properties in non-equilibrium systems may have little physical meaning.⁴⁰

The authors thank Xin Li and Mauro Mugnai for valuable comments on the manuscript. This work was supported by grants from the National Science Foundation (Grant Nos. PHY 17-08128 and PHY-1522550). Additional support was provided by the Collier-Welch Reagents Chair (Grant No. F-0019).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- 1 G. H. Heppner, B. E. Miller *et al.*, "Tumor subpopulation interactions in neoplasms," *Biochim. Biophys. Acta (BBA)* **695**, 215–226 (1983).
- 2 N. McGranahan and C. Swanton, "Biological and therapeutic impact of intratumor heterogeneity in cancer evolution," *Cancer Cell* **27**, 15–26 (2015).
- 3 K. Hinohara and K. Polyak, "Intratumoral heterogeneity: More than just mutations," *Trends Cell Biol.* **29**, 569 (2019).
- 4 M. Gerlinger, A. J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey *et al.*, "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing," *New England J. Med.* **366**, 883–892 (2012).
- 5 M. Gerlinger, S. Horswell, J. Larkin, A. J. Rowan, M. P. Salm, I. Varela, R. Fisher, N. McGranahan, N. Matthews, C. R. Santos *et al.*, "Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing," *Nat. Genet.* **46**, 225 (2014).
- 6 K. Harbst, M. Lauss, H. Cirenajwis, K. Isaksson, F. Rosengren, T. Torngren, A. Kvist, M. C. Johansson, J. Vallon-Christersson, B. Baldetorp *et al.*, "Multi-region whole-exome sequencing uncovers the genetic evolution and mutational heterogeneity of early-stage metastatic melanoma," *Cancer Res.* **76**, 4765 (2016).
- 7 E. C. de Bruin, N. McGranahan, R. Mitter, M. Salm, D. C. Wedge, L. Yates, M. Jamal-Hanjani, S. Shafi, N. Murugaesu, A. J. Rowan *et al.*, "Spatial and temporal diversity in genomic instability processes defines lung cancer evolution," *Science* **346**, 251–256 (2014).
- 8 J. Zhang, J. Fujimoto, J. Zhang, D. C. Wedge, X. Song, J. Zhang, S. Seth, C.-W. Chow, Y. Cao, C. Gumbs *et al.*, "Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing," *Science* **346**, 256–259 (2014).
- 9 W. Cao, W. Wu, M. Yan, F. Tian, C. Ma, Q. Zhang, X. Li, P. Han, Z. Liu, J. Gu *et al.*, "Multiple region whole-exome sequencing reveals dramatically evolving intratumor genomic heterogeneity in esophageal squamous cell carcinoma," *Oncogenesis* **4**, e175 (2015).
- 10 R. Richards, D. Mason, R. Levy, R. Bearon, and V. See, "4D imaging and analysis of multicellular tumour spheroid cell migration and invasion," *bioRxiv:443648* (2018).
- 11 A. M. J. Valencia, P.-H. Wu, O. N. Yagurtcu, P. Rao, J. DiGiacomo, I. Godet, L. He, M.-H. Lee, D. Gilkes, S. X. Sun *et al.*, "Collective cancer cell invasion induced by coordinated contractile stresses," *Oncotarget* **6**, 43438 (2015).
- 12 W. Kang, J. Ferruzzi, C.-P. Spatarelu, Y. L. Han, Y. Sharma, S. Koehler, J. P. Butler, D. Roblyer, M. H. Zaman, M. Guo *et al.*, "Tumor invasion as non-equilibrium phase separation," *bioRxiv:2020.04.28.066845* (2020).
- 13 Y. L. Han, A. F. Pegoraro, H. Li, K. Li, Y. Yuan, G. Xu, Z. Gu, J. Sun, Y. Hao, S. K. Gupta *et al.*, "Cell swelling, softening and invasion in a three-dimensional breast cancer model," *Nat. Phys.* **16**, 101–108 (2019).
- 14 A. Palamidessi, C. Malinverno, E. Frittoli, S. Corallino, E. Barbieri, S. Sigismund, G. V. Beznoussenko, E. Martini, M. Garre, I. Ferrara, C. Tripodo, F. Ascione, E. A. Cavalcanti-Adam, Q. Li, P. P. Di Fiore, D. Parazzoli, F. Giavazzi, R. Cerbino, and G. Scita, "Unjamming overcomes kinetic and proliferation arrest in terminally differentiated cells and promotes collective motility of carcinoma," *Nat. Mater.* **18**, 1252–1263 (2019).
- 15 S. Sinha, A. N. Malmi-Kakkada, X. Li, H. S. Samanta, and D. Thirumalai, "Spatially heterogeneous dynamics of cells in a growing tumor spheroid: Comparison between theory and experiments," *Soft Matter* **16**, 5294–5304 (2020).
- 16 T. R. Kirkpatrick and D. Thirumalai, "Colloquium: Random first order transition theory concepts in biology and physics," *Rev. Mod. Phys.* **87**, 183 (2015).
- 17 L. Berthier and J.-L. Barrat, "Nonequilibrium dynamics and fluctuation-dissipation relation in a sheared fluid," *J. Chem. Phys.* **116**, 6228–6242 (2002).
- 18 A. N. Malmi-Kakkada, X. Li, H. S. Samanta, S. Sinha, and D. Thirumalai, "Cell growth rate dictates the onset of glass to fluidlike transition and long time superdiffusion in an evolving cell colony," *Phys. Rev. X* **8**, 021025 (2018).
- 19 L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.* **9**, 2579–2605 (2008); available at <http://jmlr.org/papers/v9/vandermaaten08a.html>
- 20 L. van der Maaten, "Learning a parametric embedding by preserving local structure," in *Artificial Intelligence and Statistics* (Artificial Intelligence and Statistics, 2009), pp. 384–391.
- 21 L. van der Maaten and G. Hinton, "Visualizing non-metric similarities in multiple maps," *Mach. Learn.* **87**, 33–55 (2012).
- 22 L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.* **15**, 3221–3245 (2014); available at <http://jmlr.org/papers/v15/vandermaaten14a.html>
- 23 D. Kobak and P. Berens, "The art of using t-SNE for single-cell transcriptomics," *Nat. Commun.* **10**, 5416 (2019).
- 24 G. C. Linderman, M. Rachh, J. G. Hoskins, S. Steinerberger, and Y. Kluger, "Fast interpolation-based t-SNE for improved visualization of single-cell RNA-SEQ data," *Nat. Methods* **16**, 243–245 (2019).
- 25 G. Dimitriadis, J. P. Neto, and A. R. Kampff, "t-SNE visualization of large-scale neural recordings," *Neural Comput.* **30**, 1750–1774 (2018).
- 26 J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," *Nat. Phys.* **13**, 431–434 (2017).
- 27 K. Ch'ng, N. Vazquez, and E. Khatami, "Unsupervised machine learning account of magnetic transitions in the Hubbard model," *Physical Review E* **97**, 013306 (2018).
- 28 A. G. Day, M. Bukov, P. Weinberg, P. Mehta, and D. Sels, "Glassy phase of optimal quantum control," *Phys. Rev. Lett.* **122**, 020601 (2019).
- 29 I. Tah, T. A. Sharp, A. J. Liu, and D. M. Sussman, "Quantifying the link between local structure and cellular rearrangements using information in models of biological tissues," *arXiv:2008.13702* (2020).
- 30 A. Malmi-Kakkada, X. Li, S. Sinha, and D. Thirumalai, "Dual role of cell-cell adhesion in tumor suppression and proliferation," *arXiv:1906.11292* (2019).
- 31 H. S. Samanta, S. Sinha, and D. Thirumalai, "Far from equilibrium dynamics of tracer particles embedded in a growing multicellular spheroid," *arXiv:2003.12941* (2020).
- 32 M. Wattenberg, F. Viégas, and I. Johnson, "How to use t-SNE effectively," *Distill* **1**, e2 (2016).
- 33 V. Almendro, A. Marusyk, and K. Polyak, "Cellular heterogeneity and molecular evolution in cancer," *Annu. Rev. Pathol. Med. Dis.* **8**, 277–302 (2013).
- 34 X. Li and D. Thirumalai, "Share, but unequally: A plausible mechanism for emergence and maintenance of intratumour heterogeneity," *J. R. Soc. Interface* **16**, 20180820 (2019).
- 35 M. E. Dolega, M. Delarue, F. Ingremeau, J. Prost, A. Delon, and G. Cappello, "Cell-like pressure sensors reveal increase of mechanical stress towards the core of multicellular spheroids under compression," *Nat. Commun.* **8**, 14056 (2017).

³⁶K. Alessandri, B. R. Sarangi, V. V. Gurchenkov, B. Sinha, T. R. Kießling, L. Fetler, F. Rico, S. Scheuring, C. Lamaze, A. Simon *et al.*, “Cellular capsules as a tool for multicellular spheroid production and for investigating the mechanics of tumor progression *in vitro*,” *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14843–14848 (2013).

³⁷A. Doostmohammadi, S. P. Thampi, T. B. Saw, C. T. Lim, B. Ladoux, and J. M. Yeomans, “Celebrating soft matter’s 10th anniversary: Cell division: A source of active stress in cellular monolayers,” *Soft Matter* **11**, 7328–7336 (2015).

³⁸M. Merkel and M. L. Manning, “A geometrically controlled rigidity transition in a model for confluent 3D tissues,” *New J. Phys.* **20**, 022002 (2018).

³⁹D. Bi, X. Yang, M. C. Marchetti, and M. L. Manning, “Motility-driven glass and jamming transitions in biological tissues,” *Phys. Rev. X* **6**, 021011 (2016).

⁴⁰S. J. Altschuler and L. F. Wu, “Cellular heterogeneity: Do differences make a difference?,” *Cell* **141**, 559–563 (2010).