A biophysical model of Epithelial-Mesenchymal Transition uncovers the frequency and size distribution of Circulating Tumor Cell clusters across cancer types

Federico Bocci[1], Mohit Kumar Jolly[2], and José Nelson Onuchic[1,*]
[1]Center for Theoretical Biological Physics, Rice University, Houston, TX, USA
[2]Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore, India

Keywords: metastasis, epithelial-mesenchymal transition, cell migration, hybrid epithelial/mesenchymal state, circulating tumor cell clusters

*Author for correspondence:
José Nelson Onuchic: jonuchic@rice.edu

Abstract

The gain of cellular motility via the epithelial-mesenchymal transition (EMT) is considered crucial in the metastatic cascade. Cells undergoing EMT to varying extents are launched into the bloodstream as single circulating tumor cells (CTCs) or multi-cellular clusters. The frequency and size distributions of these multi-cellular clusters has been recently measured, but the underlying mechanisms enabling these different modes of migration remain poorly understood. We present a biophysical model that couples the epithelial-mesenchymal phenotypic transition and cell migration to explain these different modes of cancer cell migration. With this reduced physical model, we identify a transition from individual migration to clustered cell migration that is regulated by the rate of EMT and the degree of cooperativity between cells during migration. This single cell to clustered migration transition can robustly recapitulate cluster size distributions observed experimentally across several cancer types, thus suggesting the existence of common features in the mechanisms of cell migration during metastasis. Furthermore, we identify three main mechanisms that can facilitate the formation and dissemination of large clusters: first, mechanisms that prevent a complete EMT and instead increase the population of hybrid Epithelial/Mesenchymal (E/M) cells; second, multiple intermediate E/M states that give rise to heterogeneous clusters formed by cells with different epithelial-mesenchymal traits; and third, non-cell-autonomous induction of EMT via cell-to-cell signaling that gives rise to spatial correlations among cells in a tissue. Overall, this biophysical model represents a first step toward bridging the gap between the molecular and biophysical understanding of EMT and various modes of cancer cell migration, and highlights that a complete EMT might not be required for metastasis.

Popular summary

The Epithelial-Mesenchymal Transition (EMT) has been identified as the first step that enables cancer metastases; through this process, cancer cells gain the motility necessary to migrate and invade. Cancer cells that undergo EMT can enter the circulatory system both as single cells or as multi-cellular clusters. While single cells are generally more frequent in human cancers, clusters are more prevalent in aggressive cancers that metastasize more. Although the molecular mechanisms of EMT are relatively conserved across cancers, how different cancers exhibit such tremendous variability in terms of cell migration remains unclear. We develop a biophysical model to investigate how EMT regulation at a single cell level can give rise to single cell and clustered cell migration. This model quantitatively reproduces size distributions of circulating tumor cell clusters reported in human circulation and mouse models, therefore identifying a unifying set of principles governing cell migration across different cancer types. Moreover, a model where cells only undergo a partial EMT to a hybrid epithelial/mesenchymal state can recapitulate different features observed in collective cancer cell migration, including the frequency of large clusters and flat distributions that cannot be captured by a model of complete EMT. Besides partial EMT, we propose additional mechanisms that can facilitate the formation of large tumor cell clusters, including multiple hybrid epithelial/mesenchymal cell states and signaling between cells that enables noncell autonomous EMT induction. Therefore, our general picture suggests universal traits in the migration of cancer cells and challenges the necessity of a complete EMT for cancer metastasis.

Introduction

Metastases are the leading cause of cancer-related deaths and still represent an insuperable clinical challenge (1). Typically, the epithelial cells in a primary tumor gain some degree of motility that enables migration through the adjacent tissues, invasion to breach the basement membrane and enter the bloodstream, and eventually colonize a distant organ to start a metastasis. Individually migrating cells - or circulating tumor cells (CTCs) - have been considered the primary drivers of metastasis. Recently, however, experimental evidence suggests that cells can also plunge into the bloodstream as multi-cellular clusters, which can more efficiently survive through the metastatic cascade and initiate a secondary tumor (2). In animal models, these clusters have been shown to initiate more than 95% of all metastases, despite accounting for as low as 3% of all CTC events (individual and clusters taken together) and typically only 5-8 cells large (3, 4).

The epithelial-mesenchymal transition (EMT) is commonly viewed as the mechanism that confers motility to epithelial cancer cells and enables migration (5, 6). EMT plays a crucial role during development and in physiological processes such as wound healing. During EMT, cells partially or completely lose E-cadherin-mediated adhesion junctions and apicobasal polarization while gaining motility and invasive traits (5). EMT is not a binary phenomenon; instead cells can display a spectrum of EMT states (7), including many hybrid epithelial/mesenchymal (E/M) cell states (8, 9). These hybrid E/M states can maintain some cell-cell adhesion typical of epithelial cells while acquiring the motility typical of mesenchymal cells, hence allowing collective migration leading to the formation of clusters of CTCs (7). So far, however, no quantitative model has been proposed to explain how a partial or complete EMT could potentially translate into the different frequencies and size distributions of CTC clusters observed in different tumors.

Here, we devise a biophysical model to investigate various modes of cancer cell migration enabled during a partial or complete EMT. In our model, cancer cells can undergo a complete EMT which allows solitary migration, but can also attain an intermediate, hybrid epithelial/mesenchymal (E/M) state that enables collective migration. These premises can explain the individual and clustered cell migration observed in different cancer types and robustly recapitulate the fraction and size distribution of CTC clusters experimentally observed in patients and mouse models, therefore offering a common framework for quantifying cell migration across different cancer types.

Furthermore, given the connection between hybrid epithelial/mesenchymal cell state(s), CTC clusters and poor prognosis (10), we seek mechanisms that increase the fraction of CTC clusters and allow the formation of larger clusters. First, a fast transition to a hybrid E/M state followed by a slow transition to a fully mesenchymal state can expand the fraction of hybrid E/M cells and facilitate the formation of large clusters. In the limiting case where cell can undergo only a partial EMT to a hybrid E/M state, but not a complete EMT, cells can still migrate in a solitary or collective fashion depending on the rate of EMT in the tumor tissue. Therefore, our model suggests that a partial EMT characterized by some conservation of an epithelial program might be sufficient to explain invasion in multiple cancers. Second, multiple intermediate hybrid E/M states facilitate the formation of large clusters and promote cell heterogeneity, mechanisms that can lead to higher plasticity, aggressiveness, and drug resistance (2). Third, cell-to-cell communication can propagate EMT induction signals and give rise to larger CTC clusters by enabling the induction of a hybrid E/M phenotype in a spatially coordinated way.

Results

A percolation model with escape to couple epithelial-mesenchymal transition and cell migration

To investigate how the epithelial-mesenchymal transition (EMT) can give rise to different modes of cell migration, we develop a biophysical framework that couples the dynamic at a single cell and multi-cell levels. To model EMT, we consider a scenario where cells can assume one of three possible states: epithelial (E), hybrid epithelial/mesenchymal (H) and mesenchymal (M). In this setup, E cells have maximal adhesion with their neighboring cells and cannot migrate, while M cells do not have any adhesion with neighbors and can migrate solitarily. H cells have both adhesion and migration traits. Cells can switch from E to H state in partial EMT transitions and from H to M state in full EMT transitions with rate k_{EMT} (Fig. 1A).

Second, to model cell migration, H and M cells can escape from the lattice and are then replaced by new E cells (Fig. 1B). This replacement represents the turnover of cancer cells, owing to their enhanced proliferation potential, that may also reach the periphery of the tumor. In fact, cells at the periphery of a tumor can be more exposed to EMT-inducing signals whereas cells in the interior exhibit more epithelial traits (11, 12).

M cells can only escape as single cells with a basal rate k_{ESC}, while H cells can escape either as single cells or as multi-cellular clusters. When two or more H cells are in contact, the cluster can escape as a whole or break into smaller subclusters (Fig. 1B). Therefore, H cells can migrate as single cells if they are not part of a cluster or in the limit case of a subcluster of only one cell. In general, the escape rate of a subcluster composed by m H-cells enclosed within a larger cluster composed by s H-cells is:

$$K(m,s) = k_{ESC} m^c {p_E}^{n_E} {p_H}^{n_H} \quad (1)$$

In this formula, c is a migration cooperativity index that describes how preferable it is for the H cells that compose the cluster to migrate together. In the limit case where c=0, there is no preference for collective migration. The case of c=1 implies a purely additive motility (i.e. a cluster of 2 H cells has twice the motility of a single H cell), while c>1 implies cooperation between H cells during migration (i.e. the motility of a 2-cell cluster is more than twice the motility of a single cell). Moreover, the cells in the cluster need to break the adhesive bonds with their neighbors in order to escape and migrate. These bonds are broken with probabilities P_E (breaking of an E-H bond) and P_H (breaking of an H-H bond). P_E is always assumed to be smaller than P_H because the E-H bond retains more adhesion as compared to the H-H bond. n_E and n_H are the number of nearest neighbor cells of the cluster in an E state and an H state, respectively, and in general depend on the dimension and connectivity of the considered lattice (see Supplementary figure 1A for an example on a 2D square lattice). In the case of a single mesenchymal (M) cell, m=1 and there are no adhesion bonds, therefore eq. (1) naturally relaxes to the basal escape rate k_{ESC}.

The rates of EMT and escape (k_{EMT}, k_{ESC}) are the only timescales in the model. Therefore, we rescale time in units of $1/k_{ESC}$ so that the model is described by the dimensionless EMT rate $k = k_{EMT}/k_{ESC}$ while $k_{ESC} = 1$ in the new reference frame.

In a generic lattice with arbitrary dimension and connectivity, the flux of escaping clusters of H cells per unit time is quantified by considering every cluster in the lattice, every possible subcluster that could arise from any of those clusters, and the escape rate of any subcluster given by eq. 1. This approach, however, becomes computationally too expensive when considering large lattices. Therefore, we consider a mean-field approximation for the cluster flux:

$$\Theta(\rho_H) = \sum_{s=1}^{+\infty} p_s(\rho_H) \sum_{m=1}^{s} n(m,s)\, K(m,s) \qquad (2)$$

In this expression, the first summation considers the probability to find clusters of a certain size s ($p_s(\rho_H)$) for a certain fraction of H cells in the lattice (ρ_H). The second summation considers the relative frequencies ($n(m,s)$) and escape rates ($K(m,s)$) of subclusters of size m<s that could arise from a given cluster of size s.

The analytical expression of the cluster size distribution $p_s(\rho_H)$, however, is not known in any dimension $D \geq 2$. Therefore, we consider cells arranged in an infinite one-dimensional chain. In this limit, $p_s(\rho_H) = {\rho_H}^s(1-\rho_H)^2$ from 1D percolation theory (13) and the relative frequencies of subclusters is $n(m,s) = 2(s-m+1)/[s(s+1)]$ (Appendix A). To make the model more realistic from a biological standpoint (i.e. tumor tissues are not 1D structures), we consider an "effective-2D" approximation. In the effective-2D setting, the infinite 1D chain is in contact with two replicas that have the same fractions of E, H and M cells (Fig. 1C). This assumption reflects the fact that, in a 2D or 3D tissue, larger clusters have more nearest neighbors, and so more bonds to break to migrate (Appendix B). Compared to a 2D lattice model, the effective-2D approximation underestimates the migration of large clusters for two main reasons. First, for the same fraction of H cells in the chain (ρ_H), the probability to form large clusters is larger in a 2D lattice (supplementary figure 1B). Second, clusters in the effective-2D model are linear chains. Therefore, their number of nearest neighbors, and so the number of bonds to break, is maximized as compared to a 2D lattice (supplementary figure 1C).

The temporal dynamics of E, H, and M cell fractions (ρ_E, ρ_H, ρ_M) are described by ordinary differential equations that consider EMT transitions, escape of H and M cells, and replacement by new E cells:

$$\frac{d\rho_E}{dt} = -k\,\rho_E + \Theta(\rho_H) + \rho_M \qquad (3a)$$

$$\frac{d\rho_H}{dt} = +k\,\rho_E - k\,\rho_H - \Theta(\rho_H) \qquad (3b)$$

$$\frac{d\rho_M}{dt} = +k\,\rho_H - \rho_M \qquad (3c)$$

Eqs. 3, along conservation of with cell fractions $\rho_E + \rho_H + \rho_M = 1$, provide the equilibrium cell fractions (${\rho_E}^{(eq)}$, ${\rho_H}^{(eq)}$, ${\rho_M}^{(eq)}$).

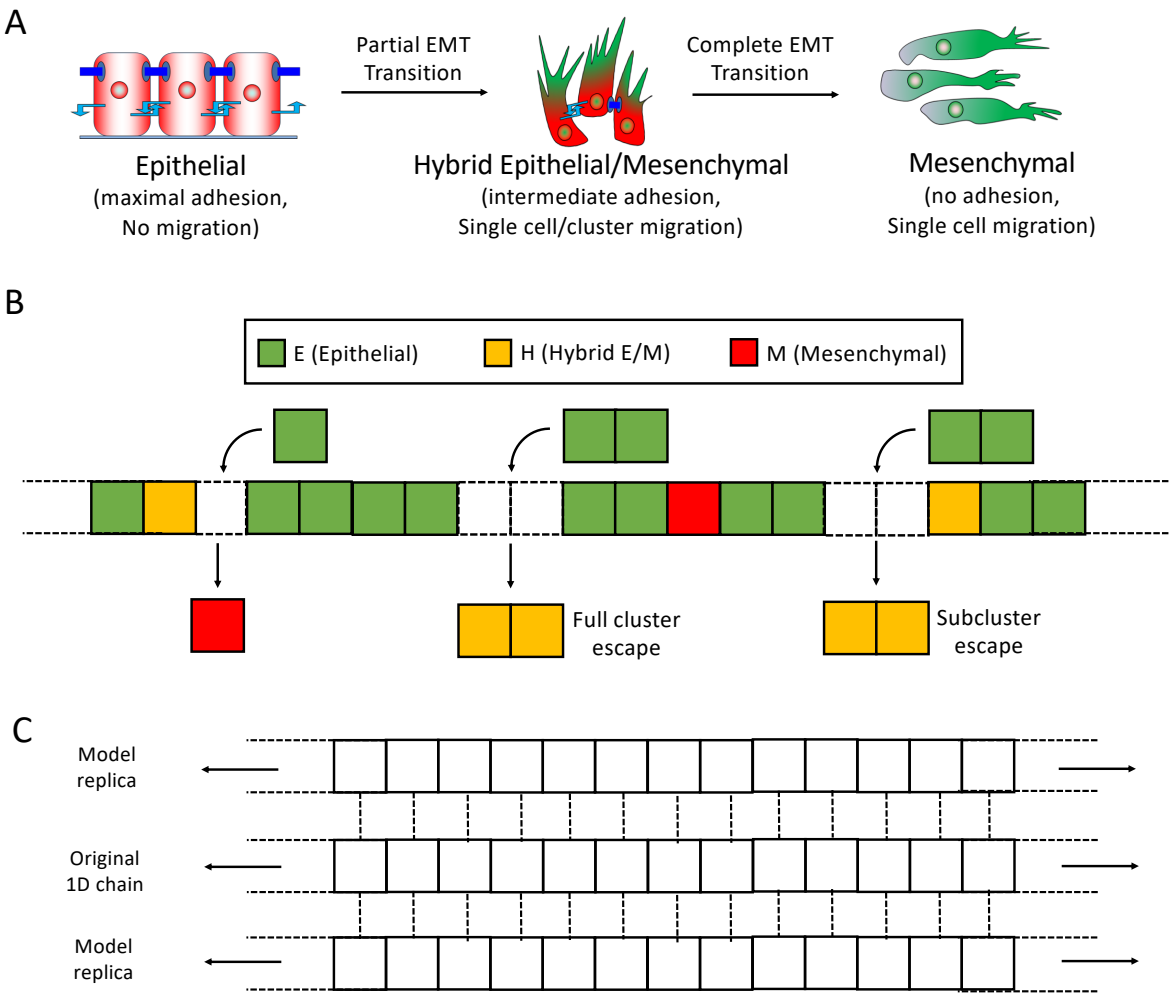

Figure 1. **A biophysical model to couple EMT phenotype transitions and cell migration. (A)** Epithelial (E) cells switch to a hybrid epithelial/mesenchymal (H) state in partial EMT transitions, and H cells switch to a Mesenchymal (M) state in complete EMT transitions. **(B)** When H and M cells migrate from the lattice, they are replaced by new epithelial cell that occupy the positions left in the lattice. When two or more H cells are neighbors, the entire cluster can migrate together (full cluster escape) or can break into smaller subcluster(s) that migrate(s) alone (subcluster escape). **(C)** In the effective-2D model, the original 1D chain (in the middle) is in contact with two replicas (top and bottom chains). The fractions of E, H, M cells in the replicas are equal to the fractions in the original chain.

The rate of EMT and cell cooperativity during migration give rise to a transition from single cell migration to cluster migration

First, we examine the steady state properties of the model as a function of the rate of EMT k. This parameter represents the ratio between the typical rate of EMT in a cancer cell and the typical rate of migration. Therefore, $k \ll 1$ implies that cell migration is faster than EMT, while $k \gg 1$ implies a fast EMT compared with cell migration.

In presence of a slow EMT ($k \ll 1$), the lattice is mostly composed by epithelial (E) cells. Conversely, a maximal fraction of H cells is observed when the typical timescales for EMT and migration are similar ($k \approx 1$). Finally, a very fast EMT ($k \gg 1$) results in a largely mesenchymal (M) lattice (Fig. 2A). As a result, cell migration is predominantly solitary at low values of k (migration of individual H cells) and at large values of k (migration of individual M cells). Clustered cell migration is observed, although at a small fraction, for intermediate levels of k that enable a high fraction of H cells in the lattice (Fig. 2B).

Increasing cell cooperativity during migration, however, gives rise to a switch from single cell migration to clustered cell migration. Specifically, the fraction of escape events that involve clusters of any size larger or equal than 2 cells (appendix C) increases as a combination of two factors. First, when the rate of EMT (k) is neither too high nor too low, ensuring a high fraction of hybrid H cells in the lattice; second, a high migration cooperativity (c) that promotes collective, rather than solitary, cell migration (Fig. 2C and supplementary figure 2A). Concurrently, parameters enabling a larger cluster escape fraction facilitate the formation of larger clusters, leading to an average cluster size of up to 5 cells (supplementary figure 2B). In fact, an intermediate rate of EMT (k) maximizes the fraction of H cells in the lattice, hence making the contact between two or more H cells more probable. Moreover, aggregates of H cells in the lattice tend to migrate together when the cooperativity (c) is high, hence forming larger clusters. These trends observed in cluster escape fraction and size are robust upon local and global variation of the model's parameters (supplementary figure 2C-D).

This variability in the frequency of CTC clusters is well-supported by multiple independent experiments that reported variable fractions of CTC clusters in circulation (3, 14–18). These CTC cluster fractions indicate that different cancer types are characterized by different migratory regimes, that in turn correspond to different combinations of EMT rate and cooperativity (k, c) in the model (supplementary figure 3). Potentially, different cancer subtypes or even patient-specific factors might further increase this variability.

Supporting the prediction of the effective-2D modeling framework, a similar single cell-cluster transition can be observed in both a 1D linear chain and in a square, two-dimensional lattice (supplementary figure 4).

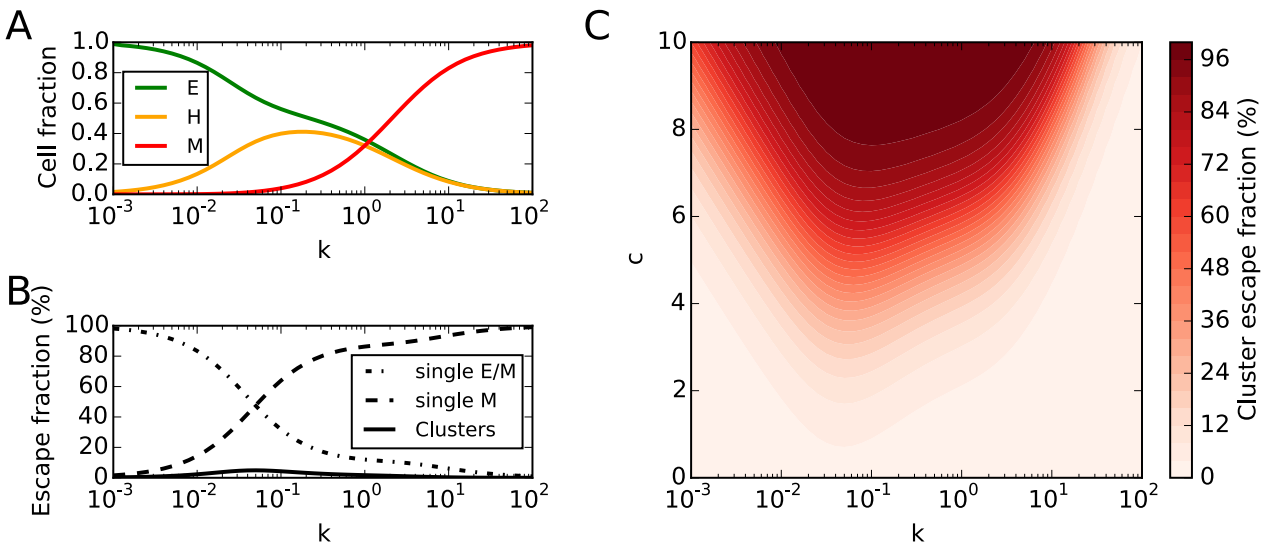

Figure 2. **Transition rate and migration cooperativity give rise to a switch from single cell migration to cluster migration. (A)** Fraction of epithelial (E, green), hybrid epithelial/mesenchymal (H, yellow) and mesenchymal (M, red) cells as a function of the rate of EMT (k). **(B)** Fractions of escape events due to clusters, single H cells and single M cells as a function of the rate of EMT (k). **(C)** Fraction of escape events due to clusters (cluster escape fraction) as a function of the rate of EMT (k, x-axis) and the migration cooperativity index (c, y-axis). All results are collected at steady state in the effective-2D model.

The single cell-cluster migration transition recapitulates cluster size distributions across several cancer types

We further investigate how the transition from individual migration to cluster-based migration is reflected in an increased probability to observe larger clusters by calculating the escape probability of clusters with different sizes (Appendix C).

First, for an intermediate rate of EMT (k) and low migration cooperativity (c), the cluster size distribution is steep and mostly dominated by single cells (i.e. no clusters). Increasing the cooperativity at a fixed EMT rate, however, modifies the shape of the distribution such that clustered cell migration becomes the primary migration mode (supplementary figure 5A). Additionally, for a high level of cooperativity, the cluster size distribution is steep for extreme values of k, while cluster cell migration is supported at intermediate levels of k (supplementary figure 5B).

Such variability can quantitatively capture cluster size distributions observed in different cancer types. First, data from melanoma patients acquired via microfluidic trap (19) shows a very steep distribution where large clusters are rare. This scenario is reproduced well by a model with low cooperativity (c=3) (Fig. 3A). Similarly, a cluster size distribution from prostate cancer (14) shares similar traits but slightly larger tail, and maps well on to a larger cooperativity index (c=4) (Fig. 3B). Microfluidic data of breast cancer patients (19) shows a more prominent role of large clusters, where approximately two-thirds of the observed clusters are composed by more than two cells. Such scenario is better recapitulated by a model with increased cooperativity (c=5.5) (Fig. 3C). For these three datasets, the frequency of single cells was not reported due to technical limitations. Our model predicts that single cell migration is the prevalent migration mode in these three cases, with a fraction of 92%, 84% and 55% of single CTCs, respectively (supplementary table 1). A mouse model of breast cancer shows a broad distribution with large clusters of 6-8 cells (Fig. 3D). While the model can qualitatively capture the transition toward a more clustered migration with a higher cooperativity (c=6), the probability to observe clusters remains relatively constant within a certain range of size in the experimental dataset (Fig. 3D). A further increase of cell cooperativity, however, gives rise to a different regime where small clusters become more probable than single cells (c=8.5, Fig. 3E). The onset of this single cell-cluster transition recapitulates well the size distribution measured in a mouse model of myeloma (20) that peaks for a cluster size of n=4 cells (Fig. 3E). In this case, single cells and small clusters of n=2 cells could not be measured experimentally, but our model predicts those to account approximately for only 3% and 8% of the total CTC fraction, while larger clusters of 3-5 cells account for about 68% of the total CTC fraction (supplementary table 1). CTC clusters detected in human glioblastoma (21) (Fig. 3F) are mapped onto an even higher cooperativity (c=9). Similar to the case of breast cancer (Fig. 3D), our model qualitatively captures the shape of the distribution but underestimates the frequency of large clusters composed by 7-8 or more cells in the case of glioblastoma (Fig. 3F). Finally, another mouse model of breast cancer (4) is characterized by a typical cluster size of 5-10 cells, and is largely devoid of individual CTCs and smaller CTC clusters. This size distribution can be mapped onto a very high cooperativity index (c=10) without increasing the rate of EMT (Fig. 3G). Thus, an increased value of the cooperativity parameter (c), without any further increase in rate of EMT, can account for the larger sized CTC clusters isolated from circulation. Indeed, at least some breast cancer cells have been proposed to invade collectively through a conserved epithelial program (22, 23).

The abovementioned datasets have been reproduced with a fixed EMT rate and increasing cooperativity. However, this is not the only combination in the two-dimensional parameter space of (c,k) that can reproduce these experimentally observed distributions. Indeed, several combinations of the EMT rate and cooperativity can quantitatively reproduce these distributions (supplementary figure 5C-I). In summary, a progressing tumor can access the cluster-based dissemination mode by (i) increasing the migration cooperativity at fixed EMT rate, (ii) increasing/decreasing its EMT rate from an extreme value and large enough cooperativity, or (iii) with a combination of these two pathways. (Fig. 3H).

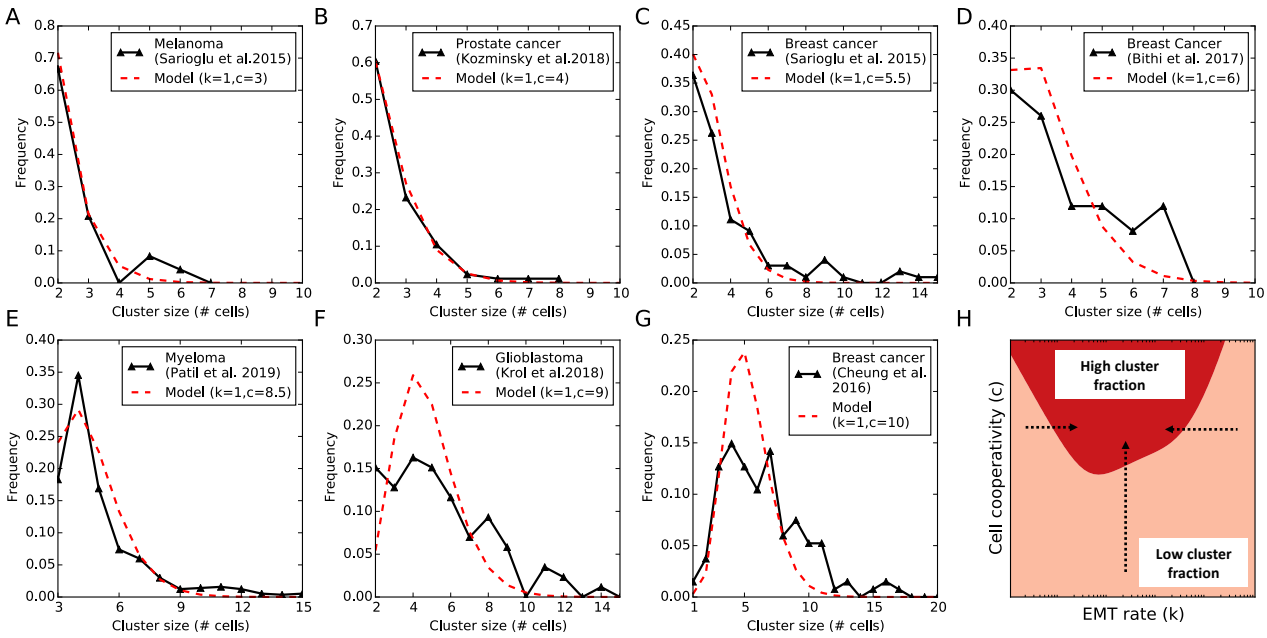

Figure 3. **The single cell-cluster migration transition recapitulates cluster size distributions across several cancer types.** Increasing the migration cooperativity index (c) at a fixed EMT rate (k=1) reproduces various experimental cluster size distributions: **(A)** Cluster size distribution observed in human melanoma (adapted from Sarioglu et al. (19)) and model's fit for a cooperativity index c=3. **(B)** Cluster size distribution in human prostate cancer (adapted from Kozminsky et al. (14)) and model's fit for c=4. **(C)** Cluster size distribution in human breast cancer (adapted from Sarioglu et al. (19)) and model's fit for c=5. **(D)** Cluster size distribution in a mouse model of breast cancer (adapted from Bithi et al. (24)) and model's fit for c=6. **(E)** Cluster size distribution in a mouse model of myeloma (adapted from Patil et al. (20)) and model's fit for c=8.5. **(F)** Cluster size distribution in human glioblastoma (adapted from Krol et al. (21)) and model's fit for c=9. **(G)** Cluster size distribution in a mouse model of breast cancer (adapted from Cheung et al. (4)) and model's fit for c=10. For datasets A-B-C-E-F, single cells were not measured; for dataset D, single cells and clusters of size n=2 cells were not measured. The original datasets were normalized to help comparison with the model. **(H)** Qualitative representation of the transition pathways from single cell migration to cluster cell migration in the (k, c) parameter space. This diagram depicts a 'binarized' version of Fig. 2C where the parameter space is differently colored depending on whether the fraction of clusters is smaller (orange) or larger (red) than 0.5.

A model of only partial, and not complete EMT, can explain both single cell and cluster migration

The biophysical model of EMT with 3 cell states (E, H, M) could not quantitatively capture (i) broad distributions where clusters of different sizes have similar frequencies (see Fig. 3D, 3F) and (ii) the presence of large CTC clusters of more than 8-10 cells in tumors characterized by collective migration (see Fig. 3F-G). Different molecular mechanisms, however, can interfere with the multiple steps of EMT, stabilize hybrid epithelial/mesenchymal phenotype(s) and give rise to more CTC clusters (25–28) by either (i) facilitating the transition from E to H, or (ii) preventing the transition from H to M. To address this scenario, we separately consider a rate of partial EMT that regulates the transition from epithelial (E) to hybrid (H) and a rate of complete EMT that regulates the transition from hybrid (H) to mesenchymal (M) (Appendix D).

The interplay of timescales for partial and complete EMT regulates the composition of the tumor tissue. As expected, a fast partial EMT, coupled with a slow complete EMT, generates a high fraction of hybrid E/M cells in the lattice (Fig. 4A). A slow partial EMT, however, results in an epithelial lattice, while the tissue is mostly mesenchymal when both the processes are fast (supplementary figure 6A-B). Therefore, an increased rate of complete EMT ($k_{Complete\ EMT}$) does not favor a large fraction of cells disseminating as clusters (supplementary figure 6C). Indeed, a high expression of 'phenotypic stability factors' that can stabilize hybrid epithelial/mesenchymal phenotype(s) and increase the mean residence times in these phenotypes by preventing them from undergoing a complete EMT (i.e. reduce $k_{Complete\ EMT}$ effectively) have been correlated with higher aggressiveness and worse patient survival in multiple cancer datasets (26–29), corroborating the evidence of clusters being the primary 'bad actors' of metastasis.

In the limit where the rate of progression to a complete EMT state is very slow ($k_{Complete\ EMT} \ll k_{Partial\ EMT}$), the model is restricted to the epithelial state and the hybrid state (Fig. 4B). In this 2-state limit, the model reaches a fraction of H cells very close to 1 for larger enough $k_{Partial\ EMT}$ (Fig. 4C).
In this situation, the lattice undergoes percolation of H cells, and the level of cooperation during migration distinguishes between a solitary or a collective migration. Without cooperativity, only single cells and small clusters are observed, while a higher cooperativity gives rise to larger clusters (Fig. 4D). Strikingly, this 2-state model reproduces effectively the experimental CTC cluster size distributions characterized by a more collective migration. Compared to the 3-state model considered earlier (Fig. 3), the 2-state model can correctly captures the broad size distributions observed in breast cancer and glioblastoma (21, 24) (Fig. 4E-F) and the collective invasion with large clusters >10 cells reported in breast cancer (4) (Fig. 4G). This picture is consistent with the fact that, at least in breast cancer, a conserved epithelial program, rather than a complete EMT, is required for collective invasion (22).

These results suggest that a partial EMT to a hybrid epithelial/mesenchymal state can be sufficient to explain the different regimes of solitary and collective migration observed across different cancers, and therefore that single cell migration via a complete EMT might not be essential for metastasis.

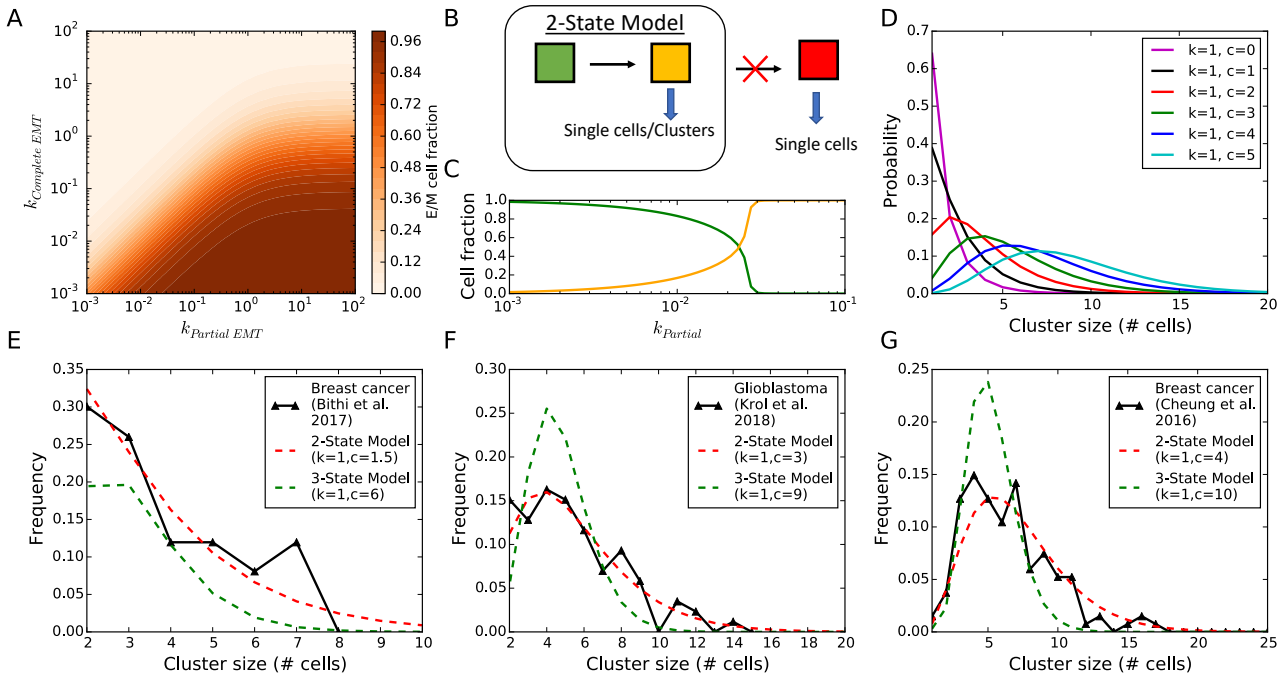

Figure 4. **The percolation limit of hybrid H cells gives rise to collective migration. (A)** Steady state cell fraction of hybrid H cells as a function of the rates of partial EMT and complete EMT in the case of linear cooperativity (c=1). **(B)** In the limit where the rate of complete EMT goes to zero, only the E and H states are available in the model. **(C)** Steady state fractions of E and H cells in the 2-State model as a function of the rate of partial EMT. **(D)** Cluster size distribution in the 2-state model for increasing values of migration cooperativity (c) at fixed rate of partial EMT (here denotes as k). **(E-F-G)** Experimental cluster size distribution human breast cancer (E – adapted from Bithi et al. (24)), Glioblastoma (F – adapted from Krol et al. (21)), and a breast cancer mouse model (G – adapted from Cheung et al. (4)) and best fit as predicted by the 2-state model for increasing values of cooperation (c=1.5, 3, 4) at fixed k (transition rate for the 3-state model, rate of partial EMT for the 2-state model). The best fit is find by minimizing the square root distance between the experimental dataset and the distribution predicted by the model while varying the parameter c.

Multiple intermediate hybrid states facilitate cluster dissemination and heterogeneity

Recent experiments suggest that multiple hybrid states with mixed epithelial and mesenchymal traits can exist with distinct epigenetic landscapes and morphological features (30–32).

Therefore, we generalize the model to an arbitrary number n of intermediate hybrid states. It is assumed that an epithelial cell transitions through the n intermediates in an ordered fashion to become mesenchymal. Furthermore, the intermediate states are orderly ranked, and every transition ($H_i \rightarrow H_{i+1}$) makes the cell less epithelial and more mesenchymal, i.e. cells in more advanced intermediates are increasingly less adhesive and have an increasingly higher probability to break the bonds with their neighbors. A different scenario, not considered here, would envision multiple parallel pathways from E to M that pass through different intermediate states. In the model, cells in different intermediate states can migrate together as clusters (Appendix E).

Increasing the number of intermediate states results in a larger cluster escape fraction and makes more probable, by orders of magnitude, to observe rare events of large clusters (Fig. 5A-B). Moreover, multiple intermediate states enable cluster-based dissemination with a higher fraction of CTC clusters at a much lower cooperativity index, compared with the case of only one intermediate state (supplementary figure 7A).

Moreover, multiple intermediate states result in a more heterogeneous cluster composition. Specifically, in a case with three intermediate states (H_1, H_2, H_3), clusters are composed by cells that are more epithelial-like (E-like or H_1), epithelial/mesenchymal-like (E/M or H_2), and mesenchymal-like (M-like or H_3). This heterogeneity has been well-reported in both clusters cultivated *in vitro* and circulating tumor clusters *in vivo*, where staining revealed that (i) only certain cells in the cluster express epithelial or mesenchymal markers and (ii) certain cells robustly express both epithelial and mesenchymal markers (3, 33–35).

Similarly, multiple intermediate states give rise to a spectrum of single migrating cells with different EMT traits (Fig. 5D). If EMT is slower than the cell migration ($k_{EMT} \ll 1$), most cells migrate as soon as they transition to an E-like state. Therefore, in this case, most of the migrating cells lean toward the epithelial end of the EMT spectrum (Fig. 5D, left region). Conversely, for a faster EMT ($k_{EMT} \sim 1$), most migrating cells lean toward the mesenchymal end of the EMT spectrum (Fig. 5D, right region). The heterogeneity of cluster composition and single cell states is conserved even for higher levels of migration cooperativity. While the fraction of single migrating cells decreases for higher c, single migrating cells are still observed in E-like, E/M, and M-like states. (supplementary figure 7B-C). This heterogeneity in cell migration has been well reported experimentally, and could be informative of the tumor progression and response to treatments. We compared the model's predictions with the CTC cell fractions measured in the bloodstream of breast cancer patients (36) (Appendix F). Notably, patients that responded positively to treatment maps onto smaller EMT rate (k) in the model after treatment, while patients that underwent tumor progression maps onto a larger EMT rate after treatment (supplementary figure 8).

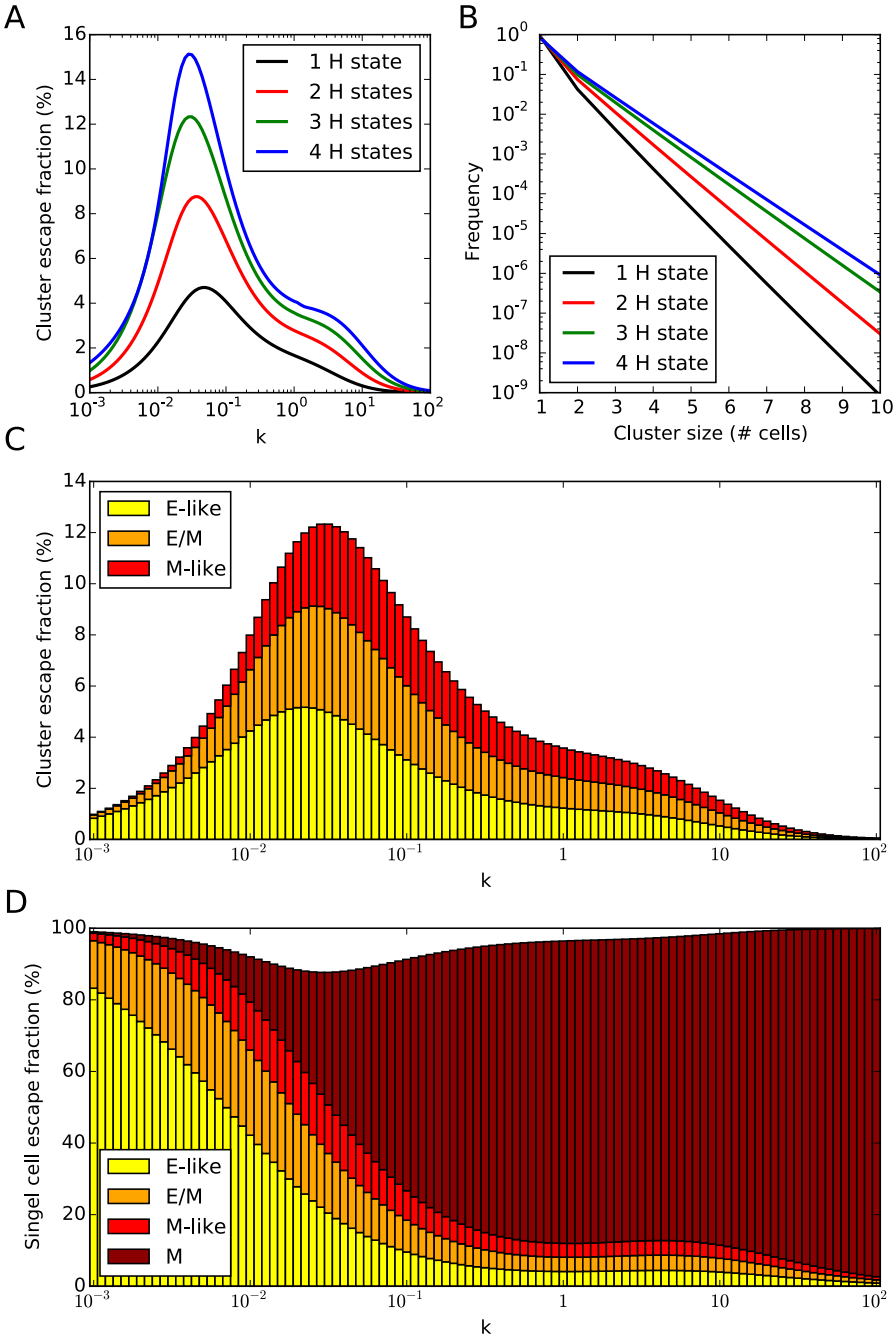

Figure 5. **Multiple intermediate hybrid E/M states facilitate cluster dissemination and heterogeneity. (A)** Cluster escape fraction as a function of the EMT rate (k) and **(B)** cluster size distribution for models with an increasing number of intermediate hybrid E/M states in the effective-2D model. **(C)** Cluster escape fraction as a function of the EMT rate (k) in the model with 3 intermediate states. Color bars show the average cluster composition in terms of fractions of E-like, E/M and M-like cells that compose the clusters. **(D)** Single cell escape fraction as a function of the EMT rate (k) in the model with 3 intermediate states. Color bars show the fraction of single cell migration due to E-like, E/M, M-like and Mesenchymal cells. All results were collected at steady state and with linear cooperativity (c=1).

Cell-to-cell EMT induction creates larger clusters and gives rise to spatial correlations among cells

EMT is not necessarily a cell autonomous process. Cell-to-cell signaling can additionally regulate EMT through multiple mechanisms, including contact-based ligand-receptor interactions and secretion of exosomes (5).

Therefore, we consider a scenario where cells in a partial EMT state can, in turn, induce EMT in neighboring cells. We assume, however, that this contact-based EMT induction requires adhesion between neighboring cells. For this reason, M cells cannot induce EMT in neighbors because, although they have an active EMT program, they have no residual cell adhesion. Similarly, an E cell will receive a stronger EMT induction from its neighbor H cells as compared to another H cell, because the adhesion of an E-H bond is greater. Therefore, in addition to the cell autonomous EMT rate (k), we introduce a noncell autonomous EMT term that depends on a basal cell-to-cell signaling rate ($k_{Cell-Cell}$) and the number of neighboring H cells that a cell has (Appendix G).

Increasing the noncell autonomous EMT rate ($k_{Cell-Cell}$) generates a higher fraction of escaping clusters (Fig. 6A) and larger clusters (supplementary figure 9). Interestingly, if either k or $k_{Cell-Cell}$ are larger than 1 (the typical timescale for escape in the model), the fraction of clusters decreases because the fast EMT rate gives rise to a high fraction of M cells that migrate in a solitary fashion.

Furthermore, cell-to-cell EMT induction gives rise to spatial correlations among H cells. This correlation can be quantified by the average number of H neighbors that a H cell has (n_H). In a linear chain, each cell has two nearest neighbors that are hybrid (H) with a probability equal to the fraction of H cells in the lattice (ρ_H). Therefore, on average, each H cell is expected to have $n_H = 2\rho_H$ hybrid H nearest neighbors in the absence of correlations. In the case of strong cell-to-cell EMT induction, however, n_H becomes significantly larger than $2\rho_H$ because H cells tend to be in contact rather than stochastically scattered, and the correlation index $(n_H - 2\rho_H)/2\rho_H$ becomes positive (Fig. 6B).

A direct comparison shows a slight anti-correlation when $k_{Cell-Cell} < k$ because H-H cell contacts can be lost via cell migration. The correlation is maximal when $k_{Cell-Cell}$ is similar to the escape rate ($k_{ESC} = 1$) and decreases when $k_{Cell-Cell}$ is significantly larger (Fig. 6C). Confirming the notion that intermediate levels of $k_{Cell-Cell}$ maximize the correlation among H cells, the probability distribution of cluster escape as a function of cluster size has a larger tail for $k_{Cell-Cell} = 1$ and is steeper for extremal values of $k_{Cell-Cell}$ (Fig. 6D).

These results show that cell-to-cell signaling is crucial to enable the formation of large clusters of H cells from a single EMT event by propagating EMT from neighbor to neighbor (Fig. 6E).

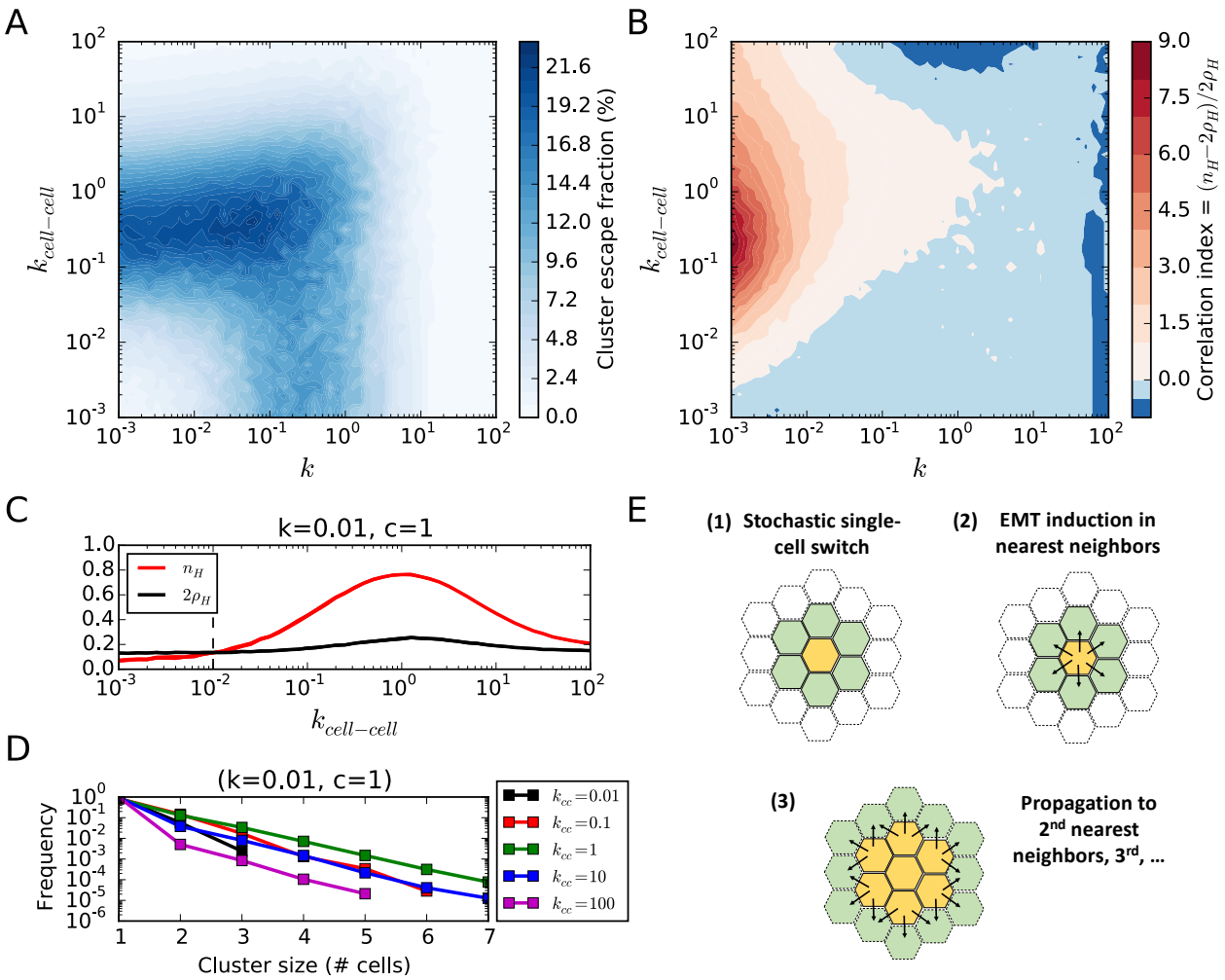

Figure 6. **Cell-to-cell EMT induction creates larger clusters and gives rise to correlations among hybrid E/M cells. (A)** Illustration of how cell-to-cell EMT induction can propagate and form large clusters of E/M cells. **(B)** Cluster escape fraction as a function of EMT rate (k) and cell-dependent EMT rate ($k_{Cell-cell}$) for c=1. **(C)** Degree of correlation between H cells in the lattice. Red indicates positive correlations between H cells (H cells tend to preferentially be in contact), while blue indicates anti-correlation. **(D)** n_H (red) and $2\rho_H$ (black) as a function of the cell-dependent EMT rate ($k_{Cell-cell}$) for k=0.01, c=1. **(E)** Cluster size distribution for different values of $k_{Cell-cell}$ and k=0.01, c=1. All these results were computed with a Gillespie simulation on a 100-cell 1D chain with the setting of the Effective-2D model.

Discussion

Are there common traits in the migration of cancer cells?

We propose a model that couples the epithelial-mesenchymal phenotypic transition at a single-cell level with cell migration. In the model, mesenchymal cells that have undergone a complete EMT can only migrate as single cells, while cells midway *en route* to EMT, or hybrid epithelial/mesenchymal (E/M), can either migrate as single cells or as multi-cellular clusters. The model predicts a transition from a migratory regime mostly dominated by single mesenchymal cells to a regime mostly dominated by hybrid E/M cell clusters. This transition depends on the rate at which cells undergo EMT and on the level of cooperativity that quantifies how probable it is for hybrid E/M cells to migrate together rather than as single cells. This simple physical model can quantitatively reproduce different cluster size distributions observed in the bloodstream of patients and/or in experimental mouse models. These experimental datasets exhibit a tremendous variability, ranging from cases where large clusters are highly improbable to cases of collective migration (4, 14, 19–21, 24), all of which can be captured by the model for varying parameter combinations.

Nonetheless, there are multiple limitations to our model. First, cancer cells in a tumor tissue can proliferate, hence introducing an additional modulation to cell patterning. Second, we do not consider the death of CTCs in circulation, which may depend on the cluster size (37). Third, additional mechanisms of aggregation can modulate the size of these clusters after detachment from the primary tumor, such as collisions in the bloodstream (38). Fourth, the CTC clusters may contain non-cancerous cells such as neutrophils (39). Finally, it is worth pointing out that the abovementioned datasets have been acquired in different systems and with different experimental techniques, which might have affected the accuracy in detection of CTC clusters. With these caveats being raised, these results support the idea that multiple aspects of cancer cell migration are highly conserved across cancer types, and microenvironmental, case-specific factors might set the conditions for a more individual or collective migration by regulating cancer cell heterogeneity in terms of EMT and migration, as effectively represented by the variation of the parameters in the model. Supporting this picture, it has been recently shown that multiple breast cancer cell lines with varying origin and morphological features can switch from solitary to collective migration by manipulating the physical properties of the extracellular medium *in vitro* (40, 41).

Is a complete EMT necessary for cancer metastases?

Cell invasion during the metastatic cascade can operate via multiple mechanisms, such as single cell invasion upon activation of a 'full EMT' program or collective invasion characterized by a basal-like phenotype that conserves E-cadherin contacts between cells (22, 42). EMT, at least in the cancer metastasis field, has been largely viewed as a binary process – no migration (no EMT) or individual migration (full EMT) – but recent evidence has suggested that EMT need not be an 'all-or-none' linear unidimensional process (43). The existence of one or more hybrid epithelial/mesenchymal states capable to invade while maintaining cell-cell adhesion has been proposed as a possible explanation for the different modes of invasion (43). Supporting this hypothesis, hybrid epithelial/mesenchymal phenotype(s) have been associated with stemness and tumor-initiation ability (31, 44–47), while inducing a complete EMT has been shown to even restrict tumor aggressiveness and metastatic potential in some cases (48–51). Here, we show that a model of only partial EMT that can be associated with a partially conserved epithelial program can still give rise to different regimes characterized by more individual or more collective cell migration. While a model of complete EMT could correctly recapitulate the size distribution of CTC clusters in contexts where the migration is mostly solitary, a model of only partial EMT could quantitatively reproduce cases of both solitary and collective migration. These observations suggest that, although a complete EMT can take place in certain types of cancer, a partial EMT might be sufficient to explain the different modes of invasion, and seems necessary for collective migration. Therefore, this model identifies hybrid E/M state(s) as a potent concept to unify qualitatively different modes of cancer cell migration. A similar variety of migratory modes is observed in the field of developmental biology, where the idea that a complete EMT and

collective migration represent extreme instances of a larger spectrum of behaviors is gaining traction as well (52).

What molecular mechanisms can maximize the metastatic potential of CTC clusters?
First, we show that multiple, instead of a single, hybrid E/M state give rise to heterogeneous clusters composed by cells with different states along the EMT spectrum. This heterogeneity might enable co-operation among different phenotypes in executing various aspects of metastasis, hence increasing the survival probability of clusters in circulation. For instance, completely mesenchymal CTCs are associated with a more favorable survival outcome in breast cancer (53), and completely mesenchymal cells lose their metastatic potential in squamous cell carcinoma (54). Moreover, cell-to-cell communication can regulate the patterning of cells with different EMT phenotypes, and EMT induction between neighboring cells can rapidly give rise to large E/M clusters. In this regard, molecules of the Jagged family that promote cell-to-cell signaling are considered promising therapeutic targets (55). The exchange of Jagged ligands that bind through the Notch receptor facilitates bi-directional cross-talk between cells, stabilizes hybrid E/M cell phenotype(s) and promotes resistance to therapies (55–57). Previous efforts to couple EMT with intercellular signaling have disregarded the loss of cell-cell adhesion leading to a reduced cell-to-cell signaling between mesenchymal cells (58). Also, we recently showed that JAG1, a Jagged ligand subtype, promotes the growth of tumor organoids in breast cancer cells *in vitro* (12). Overall, these observations suggest that Jagged can facilitate a 'plasticity window' that maximizes tumor aggressiveness in terms of EMT, proliferation and therapeutic resistance (45). Consistently, CTC clusters exhibit higher levels of JAG1 as compared to single CTCs in breast cancer (4, 23).

Conclusions

Overall, the emerging picture suggests similar mechanisms to acquire motility and invasive traits, which might be regulated differently in different cancers, hence leading to different modes of migration. Our model suggests that tumors that are more mesenchymal tend to disseminate more in the form of single CTCs and only rare clusters, while tumors where EMT is more tightly regulated and are more hybrid E/M (intended, generally, as either one or a spectrum of states with mixed epithelial-mesenchymal traits) may disseminate as clusters more efficiently (Fig. 7). Thus, counterintuitive to the long-standing paradigm, mechanisms that inhibit a complete EMT and instead stabilize hybrid E/M state(s) might bear a higher metastatic potential (27, 28, 59). Future investigations should therefore focus on identifying the molecular mechanisms that drive a progression to a mesenchymal state or maintain a partial EMT instead.

Cell migration appears ubiquitously in a multitude of physiological and pathological contexts, and it has been the subject of several theoretical and computational investigations. Most of these efforts, however, have been devoted toward understanding either the molecular mechanisms that enable such transition within a single cell (60, 61) or the different invasive modes that arise from the mechanical properties of cells (62, 63). This work offers a first step toward an integrated understanding of cell migration and its underlying signaling mechanisms. Future efforts will further investigate the coupling between the signaling and mechanical aspects of cancer cell migration *en route* to a better understanding of cancer metastasis.

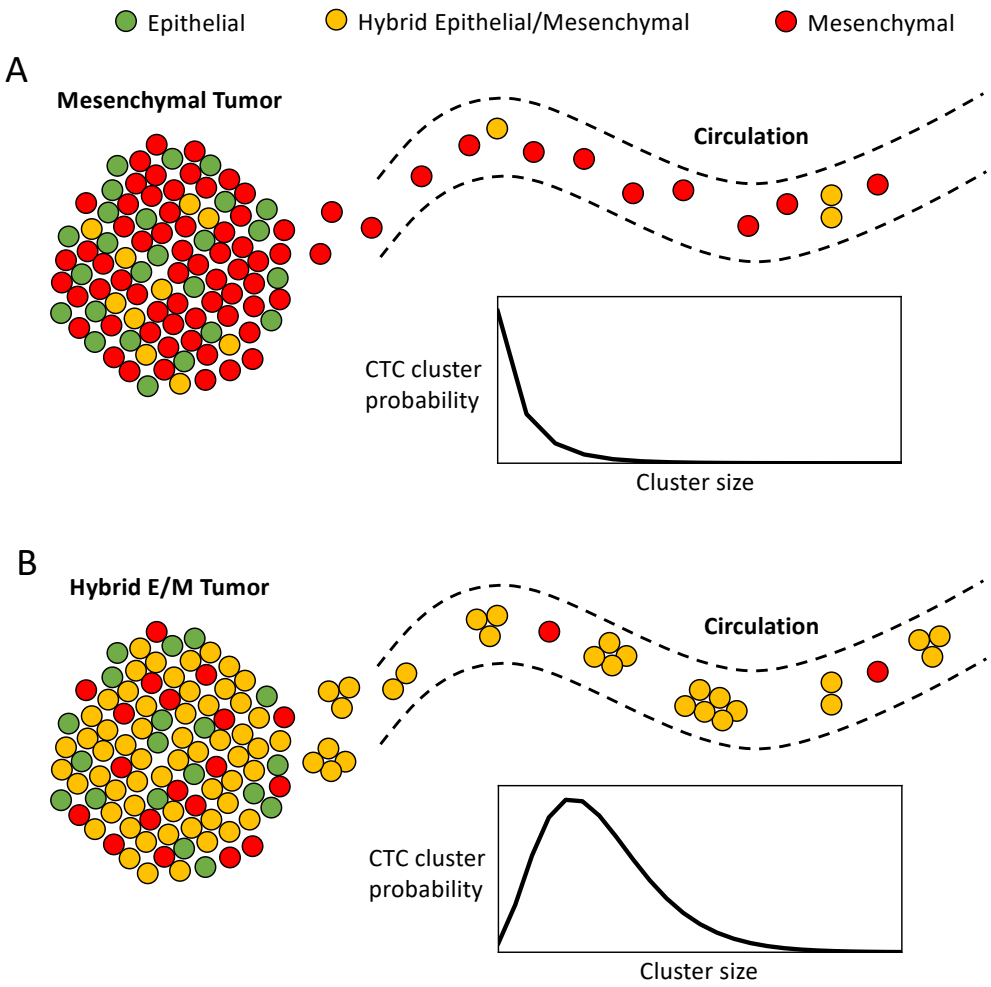

Figure 7. **Proposed mechanisms to enable different invasion modes in a tumor. (A)** Mesenchymal tumors where cells can undergo a complete EMT disseminate mostly single CTCs. **(B)** Tumors where EMT is more tightly regulated and a mesenchymal state is not accessible disseminate mostly CTC clusters composed of cells in one or more hybrid epithelial/mesenchymal states that conserve cell-cell adhesion but also exhibit motility.

Acknowledgements

The authors thank Prof. Herbert Levine (Northeastern University) for his valuable suggestions. The work in the Onuchic Lab was supported by the Center for Theoretical Biological Physics (NSF PHY-1427654). Federico Bocci was supported by the Marjory Meyer Hasselmann Fellowship. Mohit Kumar Jolly was supported by Ramanujan Fellowship provided by SERB, DST, Government of India (SB/S2/RJN-049/2018).

Author contribution

FB developed the model, produced the results and wrote the manuscript; MKJ and JNO supervised the research. All authors discussed the results and edited the manuscript.

Appendix

A – Frequency of subclusters in a one-dimensional chain

In a one-dimensional chain, a cluster of s cells can only be a linear chain. The number of possible subclusters of size m (m<s) that can be formed from the cluster of size s is:

$$N(m,s) = s - m + 1 \qquad (4)$$

The relative frequency of subclusters of a given size m equals the number of subclusters of size m divided by the total number of subclusters of any size that can be formed from the original cluster:

$$n(m,s) = \frac{N(m,s)}{\sum_{m=1}^{s} N(m,s)} = \frac{s-m+1}{\frac{s(s+1)}{2}} \qquad (5)$$

B – Number of nearest neighbors in mean field approximation

Here, n_E and n_H are the number of nearest neighbors of the cluster that are in an E and H state, respectively. In the one-dimensional chain, each cluster or subcluster has always 2 nearest neighbors. A cluster of size s has, on average, $2\rho_E/(\rho_E + \rho_M)$ nearest neighbors in an E state and $2\rho_M/(\rho_E + \rho_M)$ nearest neighbors in an M state. For a subcluster of size m, two scenarios are possible. Out of the $s - m + 1$ possible subclusters, $s - m - 1$ are completely enclosed within the main cluster, and their nearest neighbors are therefore 2 H cells. The remaining 2 subclusters of size m share 1 bond with an H cell within the cluster and another bond with either an E cell or an M cell with probabilities $\rho_E/(\rho_E + \rho_M)$, $\rho_M/(\rho_E + \rho_M)$, respectively.

In the effective-2D variation of the model, a cluster of size s has 2 nearest neighbors from the main lattice and s additional neighbors from each one of the two replicas, for a total of 2+2s nearest neighbors. Similarly, a subcluster of size m has 2+2m nearest neighbors. These additional neighbors are in an E, H or M state with probabilities ρ_E, ρ_H, ρ_M, respectively. In any dimension equal or larger than 2, the cluster size distribution is not known, and a simulation is required.

C – Equilibrium cluster size distribution

The probability distribution of escaped clusters as a function of cluster size, or cluster size distribution, can be reconstructed as follows. The total escape flux of clusters of fixed size s is:

$$F(s) = \sum_{s'>s} p_{s'}(\rho_H) \sum_{m=1}^{s} n(m,s')\, k(m,s')\delta_{ms} = \sum_{s'>s} p_{s'}(\rho_H) n(s,s')\, k(s,s') \qquad (6)$$

Where the condition $s' > s$ considers that a subcluster of size s can arise from clusters of size s' larger or equal than s, and the Kronecker's delta δ_{ms} ensures that only subclusters of size s are considered. Finally, to compute the cluster size distribution $f(s)$, one must consider that the escape flux of single cells (s=1) receives the additional contribution of M cells that can only migrate as single cells:

$$f(s) = \begin{cases} \dfrac{\rho_M + F(1)}{\rho_M + \Theta(\rho_H)}, & s = 1 \\ \dfrac{F(s)}{\rho_M + \Theta(\rho_H)}, & s > 1 \end{cases} \qquad (7)$$

D - Variation of EMT rates and 2-state model

The rates pf partial EMT and complete EMT are independent parameters, so that:

$$E \xrightarrow{k_{Partial\ EMT}} H \xrightarrow{k_{Complete\ EMT}} M$$

In the limit where $k_{Complete\ EMT} \sim 0$, the model has only the E and H states, with the following reactions:

$$E \xrightarrow{k_{Partial\ EMT}} H$$

$$H \xrightarrow{Escape} E$$

E - Multiple intermediate states

For a generic case of n IS, it is assumed that an E cell must transition through all ISs in an ordered fashion to become mesenchymal:

$$E \xrightarrow{k_{EMT}} H_1 \xrightarrow{k_{EMT}} H_2 \xrightarrow{k_{EMT}} \dots \xrightarrow{k_{EMT}} H_{n-1} \xrightarrow{k_{EMT}} H_n \xrightarrow{k_{EMT}} M$$

Cells in different IS can form clusters and escape together. Therefore, in a model with n IS, each IS *i* is characterized by a set of n probabilities: the probability of breaking a bond between an epithelial cell and a cell in the ith IS, plus all the probabilities of breaking bonds between a cell in the ith IS and a cell in any other IS. To estimate these probabilities, it is assumed that the cell states are ordered on an adhesion scale A, so that the E state has a maximal adhesion factor A_E=1 and the M state has a minimal adhesion factor A_M=0. All IS are placed in an ordered and linear fashion in (0,1), so that $A_{i-1}>A_i>A_{i+1}$. With this scale, a pairwise adhesion factor A_iA_j can be defined for two cells in IS *i* and *j*, respectively. This factor is bound in [0,1], and can be interpreted as the probability of maintaining the bond. The probability of breaking the bond can be thus defined as the complementary probability p_{ij}=1- A_iA_j. When applied to the simple case of one IS, this generalization replicates the values of p_E and p_H with A_E=1, A_H=0.5, A_M=0.

F - K score of patient datasets

In the considered datasets, CTCs were classified as E, E>M, E=M, M>E, M. These fractions were compared to the fractions of escaping single cells predicted by the model. Since the E state is not motile in the model, we considered a case with 4 intermediate hybrid states. The fractions of single migrating cells in the 4 intermediates and in the mesenchymal state were compared to the 5 CTC fractions from the dataset, hence defining a distance between the dataset and the model. The k score is the value of EMT rate (k) that minimizes the dataset between the experimental dataset and the model's prediction at a fixed cooperativity c=1.

G - Noncell autonomous EMT induction

For the case of cell-to-cell EMT induction, the rate of E-H transition for an E cell becomes:

$$k_{E-H} = k_{EMT} + k_{Cell-Cell} n_H \quad (8)$$

Where $k_{Cell-Cell}$ is a basal rate of EMT and n_H is the number of nearest neighbors in an H state. The rate of H-M transition for an H cell is:

$$k_{H-M} = k_{EMT} + A_H k_{Cell-Cell} n_H \quad (9)$$

where σ is a corrective factor that considers the reduced adhesion, and therefore ability to signal, among two H cells as compared to one E cell and one H cell. This corrective term can, in principle, assume values in [0,1], and was chosen as $\sigma = \frac{A_H}{A_E} = A_H = 0.5$. Notably, these rates now depend on the physical dimensions and topology of the lattice through n_H.

References

1. Gupta GP, Massagué J (2006) Cancer metastasis: building a framework. *Cell* 127(4):679–695.
2. Cheung KJ, Ewald AJ (2016) A collective route to metastasis: Seeding by tumor cell clusters. *Science* 352(6282):167–169.
3. Aceto N, et al. (2014) Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 158(5):1110–1122.
4. Cheung KJ, et al. (2016) Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proc Natl Acad Sci* 113(7):E854--E863.
5. Nieto MA, Huang RY, Jackson RA, Thiery JP (2016) EMT: 2016. *Cell* 166(2016):21–45.
6. Brabletz T, Kalluri R, Nieto MA, Weinberg RA (2018) EMT in cancer. *Nat Rev Cancer* 18:128–134.
7. Jolly MK, et al. (2015) Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front Oncol* 5(155).
8. Sha Y, et al. (2018) Intermediate cell states in epithelial-to-mesenchymal transition. *Phys Biol* 16:0–21.
9. Pastushenko I, Blanpain C (2018) EMT Transition States during Tumor Progression and Metastasis. *Trends Cell Biol*:1–15.
10. Jolly MK, Mani SA, Levine H (2018) Hybrid epithelial/mesenchymal phenotype(s): the „fittest" for metastasis? *BBA - Rev Cancer* 1870(2):151–157.
11. Liu S, et al. (2014) Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* 2(1):78–91.
12. Bocci F, et al. (2019) Towards understaning Cancer Stem Cell heterogeneity in the tumor microenvironment. *Proc Natl Acad Sci* 116(1):148–157.
13. Saberi AA (2015) Recent advances in percolation theory and its applications. *Phys Rep* 578:1–32.
14. Kozminsky M, et al. (2018) Detection of CTC Clusters and a Dedifferentiated RNA-Expression Survival Signature in Prostate Cancer. *Adv Sci*:1801254.
15. Krebs MG, et al. (2012) Analysis of circulating tumor cells in patients with non-small cell lung cancer using epithelial marker-dependent and independent approaches. *J Thorac Oncol* 7(2):306–315.
16. Maddipati R, Stanger BZ (2015) Pancreatic cancer metastases harbor evidence of polyclonality. *Cancer Discov* 5(10):1086–1097.
17. Molnar B, Ladanyi A, Tanko L, Sreter L, Tulassay Z (2001) Circulating Tumor Cell Clusters in the Peripheral Blood of Colorectal Cancer Patients. *Clin Cancer Res* 7:4080–4085.
18. Kulasinghe A, et al. (2018) A Collective Route to Head and Neck Cancer Metastasis. *Sci Rep* 8(1):1–8.
19. Sarioglu AF, et al. (2015) A microfluidic device for label-free, physical capture of circulating tumor cell clusters. *Nat Methods* 12(7):685–691.
20. Patil R, et al. (2019) In Vivo Monitoring of Rare Circulating Tumor Cell and Cluster Dissemination in a Multiple Myeloma Xenograft Model. *bioRxiv*.
21. Krol I, et al. (2018) Detection of circulating tumour cell clusters in human glioblastoma. *Br J Cancer* 119:487–491.
22. Cheung KJ, Gabrielson E, Werb Z, Ewald AJ (2013) Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* 155(7):1639–1651.
23. Jolly MK, et al. (2017) Inflammatory breast cancer: a model for investigating cluster-based dissemination. *npj Breast Cancer* 3(1):21.
24. Bithi SS, Vanapalli SA (2017) Microfluidic cell isolation technology for drug testing of single tumor cells and their clusters. *Sci Rep* 7:41707.
25. Jia D, et al. (2015) OVOL guides the epithelial-hybrid-mesenchymal transition. *Oncotarget* 6(17):15436–48.
26. Jolly MK, et al. (2016) Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* 7(19):27067–84.
27. Bocci F, et al. (2017) Numb prevents a complete epithelial – mesenchymal transition by modulating Notch signalling. *J R Soc Interface* 14(136):20170512.
28. Bocci F, et al. (2018) NRF2 activates a partial Epithelial-Mesenchymal Transition and is maximally present in a hybrid Epithelial/Mesenchymal phenotype. *BiorXiv*.

29. Biswas K, Jolly M, Ghosh A (2019) Stability and mean residence times for hybrid E/M phenotypes. *Phys Biol* 16(2):025003.
30. Font-Clos F, Zapperi S, La Porta CAM (2018) Topography of epithelial–mesenchymal plasticity. *Proc Natl Acad Sci* 115(23):5902–5907.
31. Pastushenko I, et al. (2018) Identification of the tumour transition states occurring during EMT. *Nature* 556:463–468.
32. Jia D, et al. (2018) Testing the gene expression classification of the EMT spectrum. *Phys Biol* 16(2):025002.
33. May AN, Crawford BD, Nedelcu AM (2018) In Vitro Model-Systems to Understand the Biology and Clinical Significance of Circulating Tumor Cell Clusters. *Front Oncol* 8:63.
34. Lecharpentier A, et al. (2011) Detection of circulating tumour cells with a hybrid (epithelial/ mesenchymal) phenotype in patients with metastatic non-small cell lung cancer. *Br J Cancer* 105(9):1338–1341.
35. Hou JM, et al. (2011) Circulating tumor cells as a window on metastasis biology in lung cancer. *Am J Pathol* 178(3):989–996.
36. Yu M, et al. (2013) Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science (80-)* 339(6119):580–584.
37. Rostami P, et al. (2019) Novel approaches in cancer management with circulating tumor cell clusters. *J Sci Adv Mater Devices*.
38. Liu X, et al. (2019) Homophilic CD44 interactions mediate tumor cell aggregation and polyclonal metastasis in patient-derived breast cancer models. *Cancer Discov* 9(1):96–113.
39. Szczerba BM, et al. (2019) Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature*.
40. Guzman A, Ziperstein MJ, Kaufman LJ (2014) The effect of fibrillar matrix architecture on tumor cell invasion of physically challenging environments. *Biomaterials* (35):6954–6963.
41. Guzman A, Sánchez Alemany V, Nguyen Y, Zhang C, Kaufman L (2017) A novel 3D in vitro metastasis model elucidates differential invasive strategies during and after breaching basement membrane. *Biomaterials* (115):19–29.
42. Nguyen-Ngoca K-V, et al. (2012) ECM microenvironment regulates collective migration and local dissemination in normal and malignant mammary epithelium. *Proc Natl Acad Sci* 109(39):E2595-604.
43. Jolly MK, Ware KE, Gilja S, Somarelli JA, Levine H (2017) EMT and MET: necessary or permissive for metastasis? *Mol Oncol* 11(7):755–769.
44. Grosse-Wilde A, et al. (2015) Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor survival. *PLoS One* 10(5):e0126522.
45. Bocci F, Jolly MK, George JT, Levine H, Onuchic JN (2018) A mechanism-based computational model to capture the interconnections among epithelial-mesenchymal transition, cancer stem cells and Notch-Jagged signaling. *Oncotarget* 9(52):29906–29920.
46. Bocci F, Levine H, Onuchic JN, Jolly MK (2019) Deciphering the dynamics of Epithelial-Mesenchymal Transition and Cancer Stem Cells in tumor progression. *Curr Stem Cell reports*:1–11.
47. Bierie B, et al. (2017) Integrin-β4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. *Proc Natl Acad Sci* 114(12):E2337--E2346.
48. Celià-terrassa T, et al. (2012) Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *J Clin Invest* 122(5):1849–1868.
49. Ocaña OH, et al. (2012) Metastatic Colonization Requires the Repression of the Epithelial-Mesenchymal Transition Inducer Prrx1. *Cancer Cell* 22(6):709–724.
50. Tsai JH, Donaher JL, Murphy DA, Chau S, Yang J (2012) Spatiotemporal Regulation of Epithelial-Mesenchymal Transition Is Essential for Squamous Cell Carcinoma Metastasis. *Cancer Cell* 22(6):725–736.
51. Ruscetti M, Quach B, Dadashian EL, Mulholland DJ, Wu H (2015) Tracking and functional characterization of epithelial-mesenchymal transition and mesenchymal tumor cells during prostate cancer metastasis. *Cancer Res* 75(13):2749–2759.
52. Campbell K, Casanova J (2016) A common framework for EMT and collective cell migration.

Development 143(23):4291–4300.
53. Grosse-wilde A, et al. (2018) Loss of inter-cellular cooperation by complete epithelial- mesenchymal transition supports favorable outcomes in basal breast cancer patients. *Oncotarget* 9(28):20018–20033.
54. Tsuji T, et al. (2008) Epithelial-Mesenchymal Transition Induced by Growth Suppressor p12 CDK2-AP1 Promotes Tumor Cell Local Invasion but Suppresses Distant Colony Growth. *Cancer Res* 68(24):10377–10387.
55. Li D, Masiero M, Banham AH, Harris AL (2014) The notch ligand JAGGED1 as a target for anti-tumor therapy. *Front Oncol* 4:254.
56. Boareto M, et al. (2015) Jagged-Delta asymmetry in Notch signaling can give rise to a Sender/Receiver hybrid phenotype. *Proc Natl Acad Sci* 112(5):402–409.
57. Boareto M, et al. (2016) Notch-Jagged signalling can give rise to clusters of cells exhibiting a hybrid epithelial/mesenchymal phenotype. *J R Soc Interface* 13(118):20151106.
58. Shaya O, et al. (2017) Cell-cell contact area affects Notch signaling and Notch-dependent patterning. *Dev Cell* 40(5):505–511.
59. George JT, Jolly MK, Xu S, Somarelli JA, Levine H (2017) Survival Outcomes in Cancer Patients Predicted by a Partial EMT Gene Expression Scoring Metric. *Cancer Res* 77(22):6415–6428.
60. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E (2013) MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc Natl Acad Sci U S A* 110(45):18174–9.
61. Hong T, et al. (2015) An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput Biol* 11(11).
62. Kabla AJ (2012) Collective cell migration: Leadership, invasion and segregation. *J R Soc Interface* 9(77):3268–3278.
63. Yang X, et al. (2017) Correlating Cell Shape and Cellular Stress in Motile Confluent Tissues. *Proc Natl Acad Sci* 114(48):12663–12668.