

## RESEARCH ARTICLE

# Combined heterogeneity in cell size and deformability promotes cancer invasiveness

Asadullah<sup>‡</sup>, Sandeep Kumar<sup>\*,‡,§</sup>, Neha Saxena, Madhurima Sarkar, Amlan Barai and Shamik Sen<sup>§</sup>

## ABSTRACT

Phenotypic heterogeneity is increasingly acknowledged to confer several advantages to cancer progression and drug resistance. Here, we probe the collective importance of heterogeneity in cell size and deformability in breast cancer invasion. A computational model of invasion of a heterogeneous cell aggregate predicts that combined heterogeneity in cell size and deformability enhances invasiveness of the whole population, with maximum invasiveness at intermediate cell-cell adhesion. We then show that small cells of varying deformability, a subpopulation predicted to be enriched at the invasive front, exhibit considerable overlap with the biophysical properties of cancer stem cells (CSCs). In MDA-MB-231 cells, these include CD44<sup>hi</sup> CD24<sup>-</sup> mesenchymal CSCs, which are small and soft, and CD44<sup>hi</sup> CD24<sup>+</sup> hybrid CSCs, which exhibit a wide range of size and deformability. We validate our predictions by tracking the pattern of cell invasion from spheroids implanted in three-dimensional collagen gels, wherein we show temporal enrichment of CD44<sup>hi</sup> cells at the invasive front. Collectively, our results illustrate the advantages imparted by biophysical heterogeneity in enhancing cancer invasiveness.

This article has an associated First Person interview with the first author of the paper.

**KEY WORDS:** Biophysical heterogeneity, Breast cancer invasion, Cancer stem cells

## INTRODUCTION

Tumors are comprised of phenotypically distinct subpopulations, with heterogeneity associated with higher tumor growth and poor therapeutic efficiency (Lawson et al., 2018; Arozarena and Wellbrock, 2019). Several factors contribute to phenotypic heterogeneity, including genetic factors such as genomic instability (Burrell et al., 2013), non-genetic factors such as cellular plasticity (Brock et al., 2009), and by microenvironmental cues such as nutrient availability. In addition, during epithelial-mesenchymal transition (EMT), which plays a central role in cancer metastasis, epithelial cells rarely undergo a complete transition to a mesenchymal state, thereby giving rise to phenotypically distinct subpopulations (Pastushenko et al., 2018; Pastushenko and Blanpain, 2019). EMT also leads to the

generation of tumor-initiating cells, also referred to as cancer stem cells (CSCs) (Mani et al., 2008). Latest single-cell technologies have proven to be very useful in assessing and implicating molecular-level heterogeneity in tumor growth and relapse (Gan et al., 2018; Patel et al., 2014). The co-existence of multiple stable subpopulations within the same tumor may be indicative of cooperation between individual subpopulations (Tabassum and Polyak, 2015). Indeed, cooperation between phenotypically distinct subpopulations has been shown to play a beneficial role in tumor growth, invasion and drug resistance (Zhou et al., 2017). In the context of invasion, such cooperation between poorly invasive melanoma cells and proteolytically active melanoma cells drives collective invasion (Chapman et al., 2014). Similarly, in luminal breast cancer cells, lesser invasive CD44<sup>low</sup> follower cells invade collectively with CD44<sup>hi</sup> leader cells (Yang et al., 2019).

Cell size and cell deformability represent two key biophysical attributes of cells that have been shown to modulate cancer invasion and metastasis (Wolf et al., 2013; Kumar et al., 2018b). Intuitively, smaller cell size is expected to provide an invasion advantage not only during migration through the stroma, but also during transmigration through the endothelium. Studies have established small size as a key feature of stem cells as well as CSCs (De Paiva et al., 2006; Lia et al., 2015). In CSCs, the combination of small size and phenotypic plasticity has been shown to enable fast migration in aligned collagen matrices (Ray et al., 2017). In addition to smaller cell size, increased cell deformability has been shown to promote invasiveness in both cancer cell lines and patient-derived cells (Swaminathan et al., 2011; Xu et al., 2012; Harada et al., 2014), and is one of the key features of tumor-initiating CD44<sup>hi</sup> CD24<sup>low</sup> breast CSCs (Al-Hajj et al., 2003; Sheridan et al., 2006) and drug-resistant ovarian cancer cells (Kapoor et al., 2018), as well as cells migrating in non-proteolytic fashion (Das et al., 2019). Strikingly, studies have also demonstrated that more heterogeneous cellular populations possess a higher propensity for secondary metastasis (Cheung et al., 2016), suggesting that phenotypic heterogeneity may play a key role in mediating metastasis. Nevertheless, the extent of variation in cell size and cell deformability within these heterogeneous cell clusters, and their relative contribution to cancer invasiveness, remains obscure.

The relative importance of cell size and deformability in cancer invasiveness may also depend on the extent of cell-cell adhesion, which determines the mode of migration, i.e. single-cell migration versus collective cell migration. Whereas epithelial cells possessing high levels of E-cadherin exhibit collective migration, mesenchymal cells with low or no E-cadherin exhibit single-cell migration. EMT has long been considered as the key cellular transformation for cancer metastasis (Dongre and Weinberg, 2019). The process of EMT, originally thought to be a binary process, is increasingly understood to represent a spectrum of states with varying expression of epithelial/mesenchymal genes giving rise to the hybrid epithelial/mesenchymal phenotype (Jolly et al., 2018;

Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Mumbai 400076, India.

<sup>‡</sup>Present address: Lunenfeld Tanenbaum Research Institute, Toronto M5G1X5, Canada.

<sup>‡</sup>These authors contributed equally to this work

<sup>§</sup>Authors for correspondence (shamiks@iitb.ac.in, sandeep25789@gmail.com)

 A, 0000-0002-6610-3198; S.K., 0000-0001-9525-4836; A.B., 0000-0002-7510-1645; S.S., 0000-0002-8639-8659

Handling Editor: Andrew Ewald

Received 16 June 2020; Accepted 29 January 2021

Liao and Yang, 2020). This hybrid phenotype has been observed in circulating tumor cells (CTCs), which disseminate as tumor cell clusters with intact cell-cell adhesions (Yu et al., 2013; Fabisiewicz and Grzybowska, 2017). In addition to CTCs, the hybrid epithelial/mesenchymal phenotype has been observed in multiple types of cancers, with the hybrid subpopulation tending to localize at the leading edge of primary tumors (Puram et al., 2017). However, the extent to which the biophysical properties of this subpopulation enables its enrichment at the invasive front has not been adequately studied.

In this paper, using MCF-7 and MDA-MB-231 human breast cancer cell lines, which vary in the extent of cell-cell adhesion, we first document the extent of heterogeneity in cell size, deformability and cell motility. We then probe the importance of this biophysical heterogeneity on cancer invasiveness using a computational model of cancer invasion. Our results suggest that combined heterogeneity in cell size and deformability enhances the random scattering of a tumor-mimicking cell cluster independent of the extent of cell-cell adhesion, with comparable invasiveness of subpopulations of different sizes and deformabilities. Although intermixing of different subpopulations is observed even during directed cell migration, our results predict smaller-sized cells of varying deformability as one prominent subpopulation that gets enriched at the invasive front. The phenotypic characteristics of this subpopulation exhibit considerable overlap with the properties of CD44<sup>hi</sup>-expressing CSCs in MDA-MB-231 cells. Finally, using spheroid invasion experiments, we demonstrate the temporal enrichment of CSCs at the invasive front.

## RESULTS

### Biophysical heterogeneity of breast cancer cell lines

Phenotypic heterogeneity was characterized by measuring variability in cell-cell adhesion, cell size, cell speed and cell deformability of MCF-7 and MDA-MB-231 breast cancer cells. In comparison to strong E-cadherin localization at cell-cell junctions in MCF-7 cells, E-cadherin levels and peripheral localization were substantially reduced in MDA-MB-231 cells, which are more mesenchymal in nature (Fig. 1A,B). Though the two cell lines exhibited overlapping size distributions, the average size of MDA-MB-231 cells was nearly twice that of MCF-7 cells (Fig. 1Ci). The tracking of the size of individual cells over a 24 h period revealed the presence of cells that exhibited a relatively unchanged size, as well as cells that exhibited a large variation in size (Fig. 1Cii,Ciii; Movie 1). The distribution of maximum change in cell size (with respect to cell size at  $t=0$  h) revealed significantly lesser variation in the size of MDA-MB-231 cells compared to MCF-7 cells (Fig. 1D). MDA-MB-231 cells were nearly twice as motile compared to that of MCF-7 cells (Fig. 1E). However, no correlation between cell size and motility was observed in both the cell types, with the values of the correlation coefficients ( $r_{MCF}$  and  $r_{MDA}$ ) being close to zero (Fig. 1F). Further quantification of cell stiffness revealed that MDA-MB-231 cells were marginally softer than MCF7 cells (Fig. 1G). Together, these results suggest that highly invasive MDA-MB-231 cells tend to be larger, softer and more migratory compared to non-invasive MCF-7 cells.

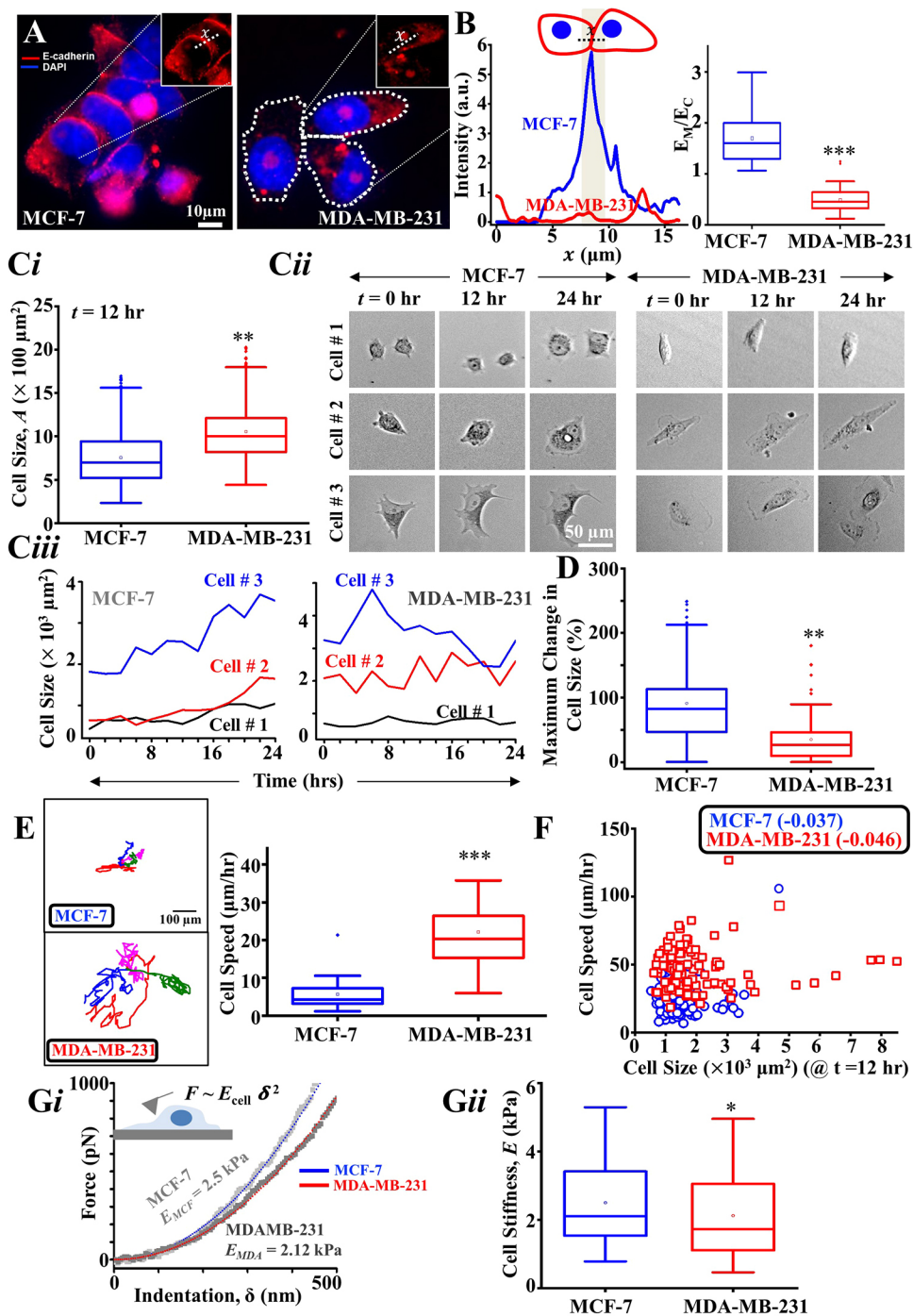
To check whether the observed functional-level heterogeneity is also present at the transcript level, publicly available single-cell RNA-seq datasets from breast cancer samples (GSE75688) (Chung et al., 2017) were analyzed by performing single sample enrichment analysis for stemness, EMT and cell biophysical characteristics (refer to Materials and Methods). Although publicly available genesets were used for calculating stemness and EMT scores,

characterization of cell mechanical properties using gene set enrichment analysis (GSEA) was performed using a manually defined BiophysicalGeneset comprising of regulators of cytoskeletal dynamics, including actin, vimentin, actin crosslinking proteins, myosin and Rho GTPases and their downstream effectors (refer to Materials and Methods). Additionally, to gain insights into intratumor and intertumor heterogeneity, enrichment scores were studied for samples from individual patients, as well as for the entire dataset. This analysis indeed revealed a wide heterogeneity in stemness, EMT and BiophysicalGeneset scores from within tumor from the same patient, as well as within the entire dataset (Fig. S1A-D). Next, to study the correlation between biophysical characteristics and stemness/EMT, we compared the enrichment score of BiophysicalGeneset with that of stemness and EMT scores for individual cells. Interestingly, this comparison showed a strong co-enrichment of BiophysicalGeneset and stemness, as well as BiophysicalGeneset and EMT scores (Fig. S1E,F). Though previous studies have shown a strong correlation between stemness and EMT (Chung et al., 2017; Mani et al., 2008), the reported relationship with biophysical characteristics is novel and was not clear from previous studies. In line with the heterogeneity observed in human breast cancer samples, a wide heterogeneity in expression levels of genes regulating cellular biophysical properties was observed in MCF7 and MDA-MB-231 cell lines (Fig. S2) (Jonasson et al., 2019; Birts et al., 2020). Comparatively, less heterogeneity was observed in EMT and stemness markers in cell line datasets. Collectively, these results suggest that breast cancer cells exhibit heterogeneity at the phenotypical level, as well as in gene expression patterns.

### Combined heterogeneity in cell size and cell deformability enhances cancer invasiveness

To probe the importance of phenotypic heterogeneity in cancer invasion, we used our recently developed cellular Potts model (CPM)-based formalism, which allows us to account for multiple factors, including cell size and deformability, cell-cell adhesion, extracellular matrix (ECM) density and ECM remodeling. The simulation framework consists of three types of pixels representing the cell (pink), ECM fibers (blue lines) and the interstitial fluid (white) (Fig. 2A; see Materials and Methods for details) (Kumar et al., 2016b). At the start of simulations ( $t=0$ ), the cell aggregate is positioned at the center of the simulation grid. Although ECM fibers sterically hinder cell migration, the lateral invasion of cells is mediated by cell secreted matrix metalloproteinases (MMPs) which degrade the fibers and create space for migration. Simulations were performed for the case of a homogeneous cell aggregate and a heterogeneous cell aggregate. For the heterogeneous aggregate, size (or target cell area) and stiffness (or area constraint) of individual cells within the aggregate were assumed to follow Gaussian distributions (Fig. 2B). Furthermore, cells of varying sizes and stiffness were distributed such that there was no bias in their position within the cluster (Fig. S3A). Cell size and cell stiffness of the homogeneous cell aggregate were set to the mean cell size ( $\bar{A}$ ) and mean stiffness ( $\bar{\lambda}_a$ ) of the heterogeneous cell aggregate.

Spatiotemporal evolution of the system is based on random movement of individual pixels subject to transition probabilities using the Monte Carlo method, and is based on the total system energy (Eqn 1). One of the major factors contributing to the total energy of the system is dependent on the nature of adhesive interactions between the three different types of pixels in our simulations (cell-cell, cell-matrix and cell-fluid; first term in Eqn 1; Fig. S3B). Although cell-matrix ( $J_{cm}$ ) and cell-fluid ( $J_{cf}$ ) adhesion energies were held constant (i.e.  $J_{cm}=16$ ,  $J_{cf}=32$  in units of  $K_B T/L$ ),

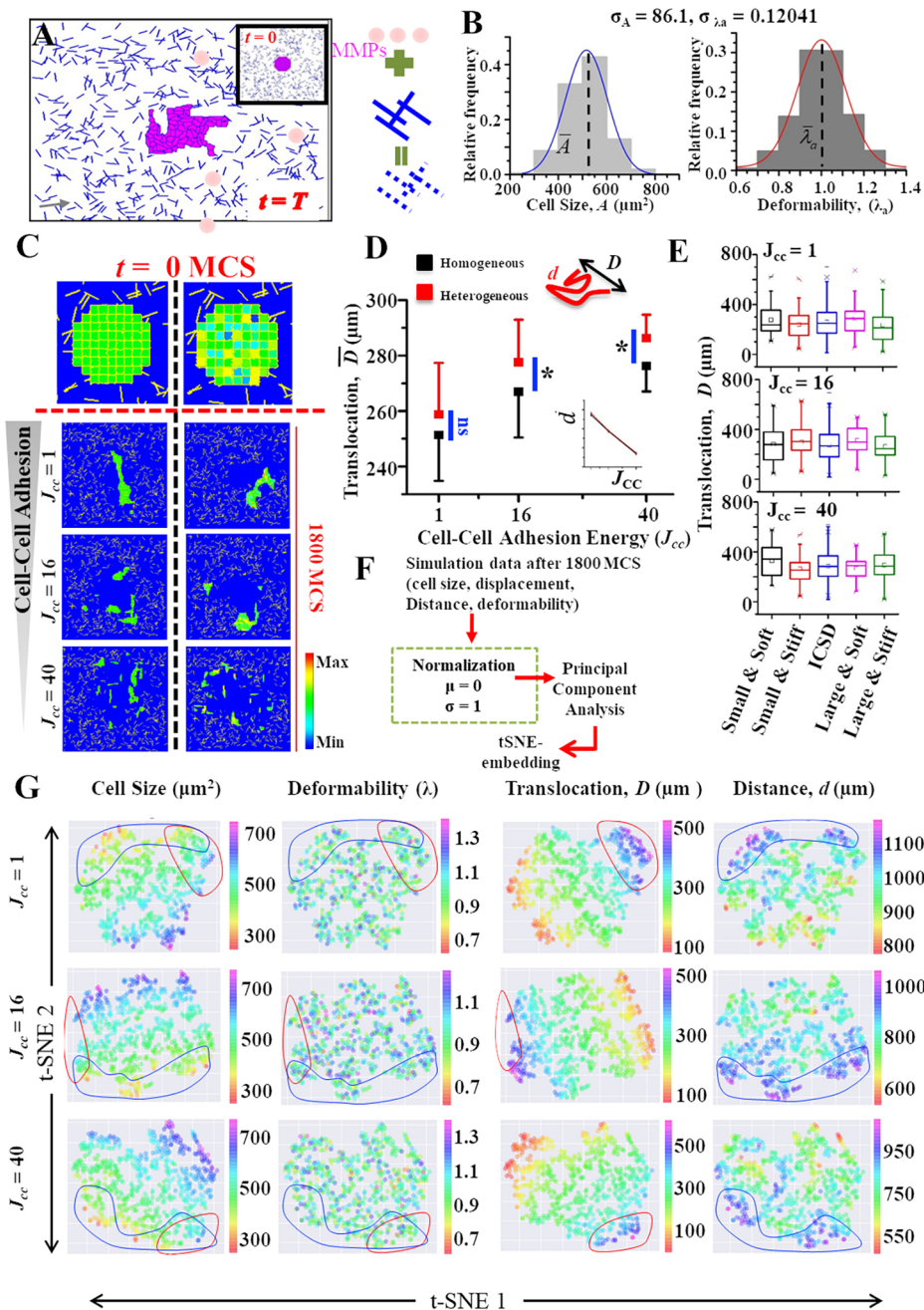


**Fig. 1. Phenotypic heterogeneity of breast cancer cells.** (A) Representative images of MCF-7 and MDA-MB-231 cells stained for E-cadherin. White dotted lines indicate cell boundaries of MDA-MB-231 cells. Insets show close ups of cell-cell junctions. (B) Line intensities of E-cadherin across cell-cell contacts in MCF-7 and MDA-MB-231 cells, and quantification of membrane to cytoplasmic E-cadherin levels (i.e.  $E_M/E_C$ ) ( $n > 40$  cells per condition pooled from  $N = 3$  independent experiments). (Ci) Distribution of cell size of MCF-7 and MDA-MB-231 cells ( $n > 200$  cells per condition pooled from  $N = 3$  independent experiments). (Cii, Ciii) Representative time-snapshots of three MCF-7 and MDA-MB-231 cells along with variation in their spread area over 24 h. (D) Distribution of maximum change in cell size over 24 h ( $n > 100$  cells per condition pooled from  $N = 3$  independent experiments). (E) Representative single cell trajectories of MCF-7 and MDA-MB-231 cells, and distribution of random cell motility ( $n > 100$  cells per condition pooled from  $N = 3$  independent experiments). (F) Plot of cell speed versus cell size in MCF-7 and MDA-MB-231 cells ( $n > 100$  cells per cell type pooled from  $N = 3$  independent experiments).  $r_{MCF}$  and  $r_{MDA}$  represent the values of Pearson correlation coefficient between cell size and cell speed in MCF-7 and MDA-MB-231 cells, respectively. (Gi) Schematic for probing cortical cell stiffness with atomic force microscope (AFM). Cells were indented with a soft pyramidal probe. Cortical stiffness was estimated by fitting the raw force-indentation curves with the Hertz equation. (Gii) Distribution of cell stiffness of MCF-7 and MDA-MB-231 cells ( $n > 100$  cells per condition pooled from  $N = 3$  independent experiments).  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$  (unpaired, two-tailed Student's  $t$ -test). In box-and-whisker plots, the box represents the 25–75th percentiles, and the median (line) and mean (small square) is indicated. The whiskers show  $1.5 \times$  of interquartile range (IQR). Outliers are indicated. a.u., arbitrary units.

cell-cell adhesion energies ( $J_{cc}$ ) were varied to mimic different extents of cell-cell adhesion to model the invasive behavior of both epithelial-like cells, such as MCF-7 cells, as well as mesenchymal-like cells, such as MDA-MB-231 cells. Whereas a choice of  $J_{cc} = 1$  led to collective migration mimicking the behavior of MCF-7 cells, a choice of  $J_{cc} = 40$  led to cell scattering mimicking the behavior of MDA-MB-231 cells (Fig. 2C; Movies 2–4). In comparison,  $J_{cc} = 16$  corresponds to cells with intermediate cell-cell adhesion in which both single cells and cell clusters (a cluster corresponds to a group of cells in contact with each other) were observed. The second term contributing to the total system energy is associated with deviations in the sizes of individual cells (Fig. S3C) from their target sizes (Fig. 2B).

The invasiveness of individual cells was quantified by measuring the total distance ( $d$ ) traveled by a given cell, as well as the net end-to-end distance ( $D$ ) traveled by the cell between its initial and final positions. Both  $d$  and  $D$  were measured based on the position of the cell centroid. Although no difference was observed in the population-averaged total distance ( $\bar{d}$ ) traveled by cells (Fig. 2D, inset), independent of  $J_{cc}$ , heterogeneity in cell size and deformability led to an increase in population-averaged net translocation ( $\bar{D}$ ) compared to the homogeneous case (Fig. 2D; Fig. S3D). Interestingly, the invasiveness of cells possessing heterogeneity only in size (i.e. stiffness of all cells  $= \bar{\lambda}_v$ ) or heterogeneity only in stiffness (i.e. size of all cells  $= \bar{A}$ ) were less invasive compared to cells possessing combined heterogeneity in cell size and deformability (Fig. S3E).





**Fig. 2. Effect of phenotypic heterogeneity on cancer invasiveness.** (A) Schematic of model. Cancer invasion was simulated by studying scattering of a cluster of 69 cells (pink pixels) positioned at the center of a two-dimensional space mimicking the ECM  $1 \times 1 \text{ mm}^2$  in size (shown in inset). The ECM consisted of 600 randomly positioned  $2 \mu\text{m}$  thick ECM fibers (blue lines)  $30\text{--}40 \mu\text{m}$  in length. Pixels that neither belong to a cell nor to an ECM fiber were considered as fluid pixels. ECM degradation is mediated by MMP molecules secreted by the cells, which diffuse into the surrounding ECM and degrade ECM fibers (see Materials and Methods for details). When an 'ECM pixel' is degraded, the pixel type is changed from 'ECM pixel' to 'fluid pixel'. (B) Cell phenotype was varied by varying cell size (A) and cell deformability (or area constraint,  $\lambda$ ), both of which were approximated as normal distributions. Sizes of individual cells within the cluster were assigned from a normal distribution with mean cell size  $\bar{A} = 517.5 \mu\text{m}^2$  and s.d.  $\sigma_A = 86.1 \mu\text{m}^2$ . Cell deformabilities were assigned from a normal distribution with mean deformability  $\lambda_a = 1 E/L^4$  and s.d.  $\sigma_\lambda = 0.12 E/L^4$ . (C) Time-dependent invasion of a cell cluster for three different values of cell-cell adhesion energy ( $J_{cc}=1, 16, 40$ ), with  $J_{cc}=1$  corresponding to collective cell invasion and  $J_{cc}=40$  corresponding to single cell invasion. Simulations were performed for homogeneous cell cluster, wherein each cell is  $517.5 \mu\text{m}^2$  in size and  $\lambda_a = 1 E/L^4$  in deformability, and for heterogeneous cell cluster, wherein cell size and deformability were drawn from the distribution in B. Each simulation was performed for 2010 MCS, with 22 independent sets of simulations performed per condition. (D) Cell invasion was quantified by measuring the total distance ( $d$ ) and the end-to-end distance ( $D$ ) traveled by a given cell in 1800 MCS. Quantification of population-averaged cell translocation ( $D$ ) for homogeneous and heterogeneous cell clusters for three different values of  $J_{cc}$ . Inset shows  $\bar{d}$  for homogeneous and heterogeneous cell clusters. Data are mean  $\pm$  s.e.m. (E) Quantification of translocation for different subpopulations of the heterogeneous cell cluster. Based on cell size and  $\lambda$ , cells were clustered into small and soft cell, small and stiff cells, cells with ICSD, large and soft cells, and large and stiff cells (see Materials and Methods for details). In box-and-whisker plots, the box represents the 25–75th percentiles, and the median (line) and mean (small square) is indicated. The whiskers show  $1.5 \times$  of interquartile range (IQR). Outliers are indicated. (F) Schematic of tSNE embedding of simulation data. (G) tSNE-based two-dimensional-embedding of simulation data based on cell size, deformability ( $\lambda$ ), net translocation ( $D$ ) and total distance traveled ( $d$ ). Red regions represent cells with high  $D$ . Blue regions represent cells with high  $d$ . \* $P < 0.05$ , ns, not significant (unpaired, two-tailed  $t$ -test).



To test whether higher  $\bar{D}$  of heterogeneous population is due to increased motility of one particular subpopulation, the population was divided into five subgroups depending upon their size and stiffness. These included cells of intermediate size and deformability (ICSD), small and soft cells, small and stiff cells, large and soft cells and large and stiff cells. ICSD cells correspond to cells possessing size and stiffness within 1 s.d. of the mean cell size and mean deformability of the population (see Materials and Methods). Whereas ‘small cells’/‘large cells’ correspond to cells smaller/larger than the size of ICSD cells, ‘soft cells’/‘stiff cells’ correspond to cells softer/stiffer than ICSD cells. In spite of considerable variation in the magnitude of translocation within each subpopulation, negligible differences were observed in the average translocation of the five subpopulations (Fig. 2E). As these observations were based on population averages, invasiveness behavior and its correlation with cell/size and deformability at a single-cell level remains obscure. Studying these aspects at the single-cell level becomes more important in this study due to intrapopulation heterogeneity. Therefore, to further probe the association between higher invasiveness and cell size/deformability at the single cell level, t-distributed stochastic neighbor embedding (tSNE)-based two-dimensional-embedding of all cells (from 22 independent simulations) was performed (Fig. 2F, Materials and Methods). Overlaying these two-dimensional projections with cell size,  $d$ ,  $D$  and  $\lambda$  for individual cells revealed that high translocation ( $D > 400 \mu\text{m}$ ) was exhibited by cells of varying sizes and deformabilities (Fig. 2G, red regions). Even though the majority of the cells that showed higher  $d$  were smaller in size (Fig. 2G, blue regions), no correlation was observed with  $D$ , suggesting that decrease in cell size may increase motility but not invasiveness. Collectively, these *in silico* studies predict that heterogeneity in cell size and deformability can promote cancer invasiveness independent of cell-cell adhesion.

### Chemotaxis is fastest at intermediate cell-cell adhesion

Though the above simulations revealed a synergistic effect of heterogeneity in cell size and deformability on invasiveness, these results may vary in the presence of a directional cue, which is relevant *in vivo*. Therefore, simulations were performed wherein directional migration of the heterogeneous cell aggregate was tracked in the presence of a stable linear chemical gradient (Fig. 3A). The efficiency of chemotaxis was modulated by varying the strength of chemotaxis ( $\mu$ ; see Materials and Methods). As effective chemotaxis was observed for  $\mu > 1000$  (Fig. S4), simulations were performed for  $\mu = 2000$  and  $\mu = 5000$ , with an increase in  $\mu$  leading to faster directed migration (Movies 5,6) (Kumar et al., 2018b). Under these conditions, cells were found to invade in a collective manner independent of  $J_{cc}$ , with negligible scattering observed even at  $J_{cc} = 40$  (Fig. 3B). Quantification of  $t^*$ , the time at which the first cell touches the right boundary of the lattice in a given simulation, was found to decrease with  $J_{cc}$  (Fig. 3C). Comparison of  $\bar{D}$  and  $\bar{d}$  at  $t_{min}$ , i.e., lowest  $t^*$  across all 22 simulations, revealed the highest translocation at intermediate cell-cell adhesion ( $J_{cc} = 16$ ) and highest persistence ( $\langle Dfd \rangle$ ) for  $J_{cc} = 16$  and  $J_{cc} = 40$  (Fig. 3D). In line with this, cellular level comparison for  $\mu = 5000$  revealed that cells that moved highest within  $t_{min}$  (i.e. highest  $d_{t_{min}}$ ), and also showed highest translocation, exhibited a wide variation in cell size and were of mixed deformabilities (Fig. 3E, red region). Additionally, at moderate and low cell-cell adhesion (i.e.  $J_{CC} \in \{16, 40\}$ ), there was a subpopulation that exhibited very high  $D$  but moderate  $d$  (blue region); a major fraction of cells in this subpopulation were large in size (blue arrows) and of mixed deformabilities. Simulations performed with lower chemotaxis strength (i.e.  $\mu = 2000$ ) showed a similar trend to that of  $\mu = 5000$ . Together, these analyses implicate intermediate cell-cell

adhesion in promoting invasion efficiency of cells in the presence of directional cues.

### Initial position and intermixing dictate enrichment of specific subpopulations at the invasive front

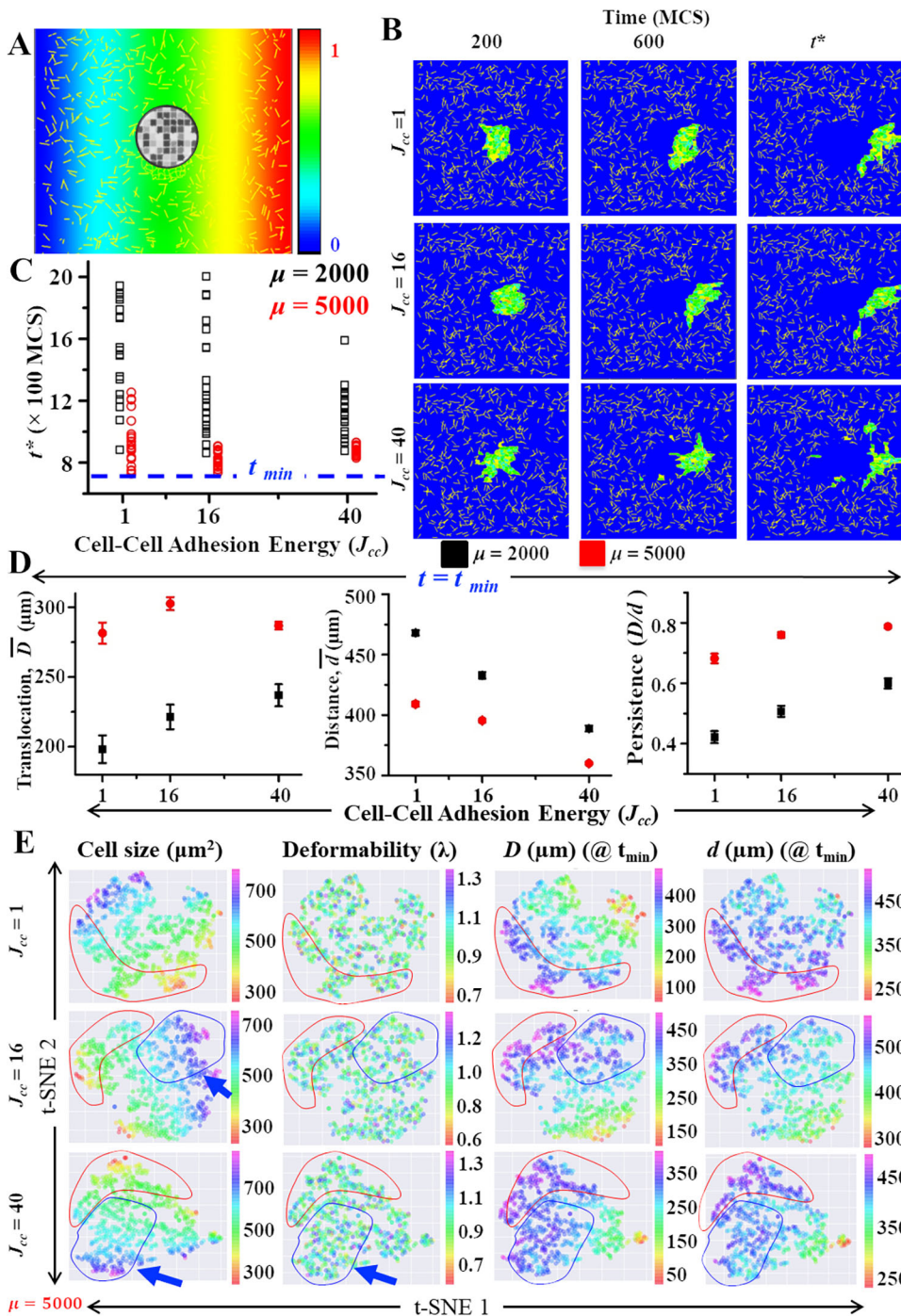
To further understand how heterogeneity in cell size and deformability imparts high invasiveness to invading populations, the distance from the right boundary at  $t = t^*$ , i.e.  $\delta$ , was compared across cells of varying sizes and deformabilities (Fig. 4A). For  $\mu = 5000$ , population averaged  $\delta$  or  $\bar{\delta}$  was comparable for all the three  $J_{cc}$  values. However, for  $\mu = 2000$ ,  $\bar{\delta}$  was lowest for  $J_{cc} = 1$ , suggesting that at moderate chemotactic strengths high cell-cell adhesion enables collective invasion of the whole population (Fig. 4B). As before, invasiveness of the population was highest, i.e.,  $\bar{\delta}$  was lowest, for the case of combined heterogeneity in cell size and deformability instead of heterogeneity in either cell size or in deformability (Fig. S5A).

To probe whether invading fronts were enriched in cells of particular sizes/deformabilities, the distribution of cell size and deformability was determined for cells that reached within  $50 \mu\text{m}$  from the right edge, i.e.,  $\delta \leq 50 \mu\text{m}$  (Fig. 4C). Although 40–50% of cells at the invasive front were of intermediate size and deformability, enrichment analysis, i.e. comparison of the proportion of cells of a given subpopulation at the invasive front in comparison to that in the entire population, provided interesting insight. Specifically, small cells of varying deformabilities, small and intermediate-sized soft cells, and intermediately stiff cells of varying cell sizes represent three heterogeneous subpopulations that were enriched at one or more values of  $J_{cc}$  (Fig. 4D).

To probe the possible mechanism(s) mediating enrichment, for all cells within the invasive front, the positions of cells at  $t = t^*$  were plotted as a function of their position at  $t = 50$  Monte Carlo Step (MCS) with different colors and shapes depicting different subpopulations (Fig. 4E,F). As the cells were positioned in a rectangular grid-like fashion at  $t = 0$  MCS (Fig. S3A), a time point of  $t = 50$  MCS was chosen to allow for some initial intermixing of the cells. For  $J_{cc} = 1$ , cells that ended up at the invasive front at  $t = t^*$ , were situated at distances  $\delta_i$  (distance of  $i^{\text{th}}$  cell from right boundary) ranging from 400 to 600  $\mu\text{m}$  at  $t = 50$  MCS. The distance 400  $\mu\text{m}$  corresponds to positions at the front of the cell aggregate closest to the right boundary, whereas 600  $\mu\text{m}$  corresponds to positions at the back of the cell aggregate farthest from the right boundary. In comparison, the initial distance ranged from 400 to 500  $\mu\text{m}$  for  $J_{cc} = 16$  and  $J_{cc} = 40$ . Consistent with these observations, the average number of neighbor changes of individual cells was highest in the case of  $J_{cc} = 1$  and lowest in the case of  $J_{cc} = 40$  (Fig. S5B). In combination, these results suggest that the initial position of cells and intermixing represent two mechanisms mediating enrichment, with the initial position playing a more dominant role at intermediate and low cell-cell adhesion, and intermixing playing a prominent role when cell-cell adhesions are intact. Interestingly, for  $J_{cc} = 1$ , where enrichment of cells from the rear of the cell aggregate was observed, the common feature of these cells was that they were either soft or of intermediate stiffness (Fig. 4F).

### Comparison of biophysical properties of simulation predicted invasive sub-population with CSCs

The above analysis has helped us identify small cells of varying deformabilities, soft cells of varying cell sizes, and intermediately stiff cells of varying cell sizes, as three heterogeneous subpopulations that are enriched at the invasive front depending on  $J_{cc}$ . To further probe which of these subpopulations is maximally invasive, simulations were performed using heterogeneous populations ( $A$ ,  $B$



**Fig. 3. Effect of phenotypic heterogeneity on chemotactic migration.** (A) Schematic of chemotaxis. Directed cell migration is driven by the presence of a stable chemotactic gradient with maximum chemokine concentration at the right edge of the two-dimensional ECM. (B) Representative time snaps of the migration of cancer cells under the influence of the chemotactic field for different values of  $J_{cc}$ .  $t^*$  corresponds to the time when the first cell reached the right boundary in a given simulation. (C) Quantification of  $t^*$  for different values of  $J_{cc}$  for two different values of chemotactic strength ( $\mu$ ), a measure of the ability of cancer cells to sense the chemokine and exhibit chemotaxis.  $t_{min}$  corresponds to the lowest value of  $t^*$  across all simulations. (D) Quantification (mean  $\pm$  s.e.m.) of ( $D, d$ ) and persistence ( $D/d$ ) at  $t_{min}$  for different values of  $J_{cc}$  and different values of  $\mu$ . (E) tSNE-based two-dimensional-embedding of simulation data based on cell size,  $\lambda$ ,  $D$  (at  $t_{min}$ ), and  $d$  (at  $t_{min}$ ). Red regions represent cells with high  $D$  and  $d$ . Blue regions represent cells with high  $D$  but moderate  $d$ . Blue arrows highlight large cells within the blue regions.

and C), wherein cell size/deformability of these populations were drawn from uniform distributions covering the same range of cell size/deformability as that of the above three subpopulations. In addition, a population  $D$  was generated by combining populations  $A$ ,  $B$  and  $C$  (Fig. 5A). Although  $t^*$  values were comparable for the four populations, for all values of  $J_{cc}$ , lowest  $\bar{\delta}$  was observed for population  $A$ , i.e. small cells of varying deformabilities (Fig. 5B; Movie 7). However, within this population, no dependence of  $\delta$  on deformability was observed (Fig. 5C).

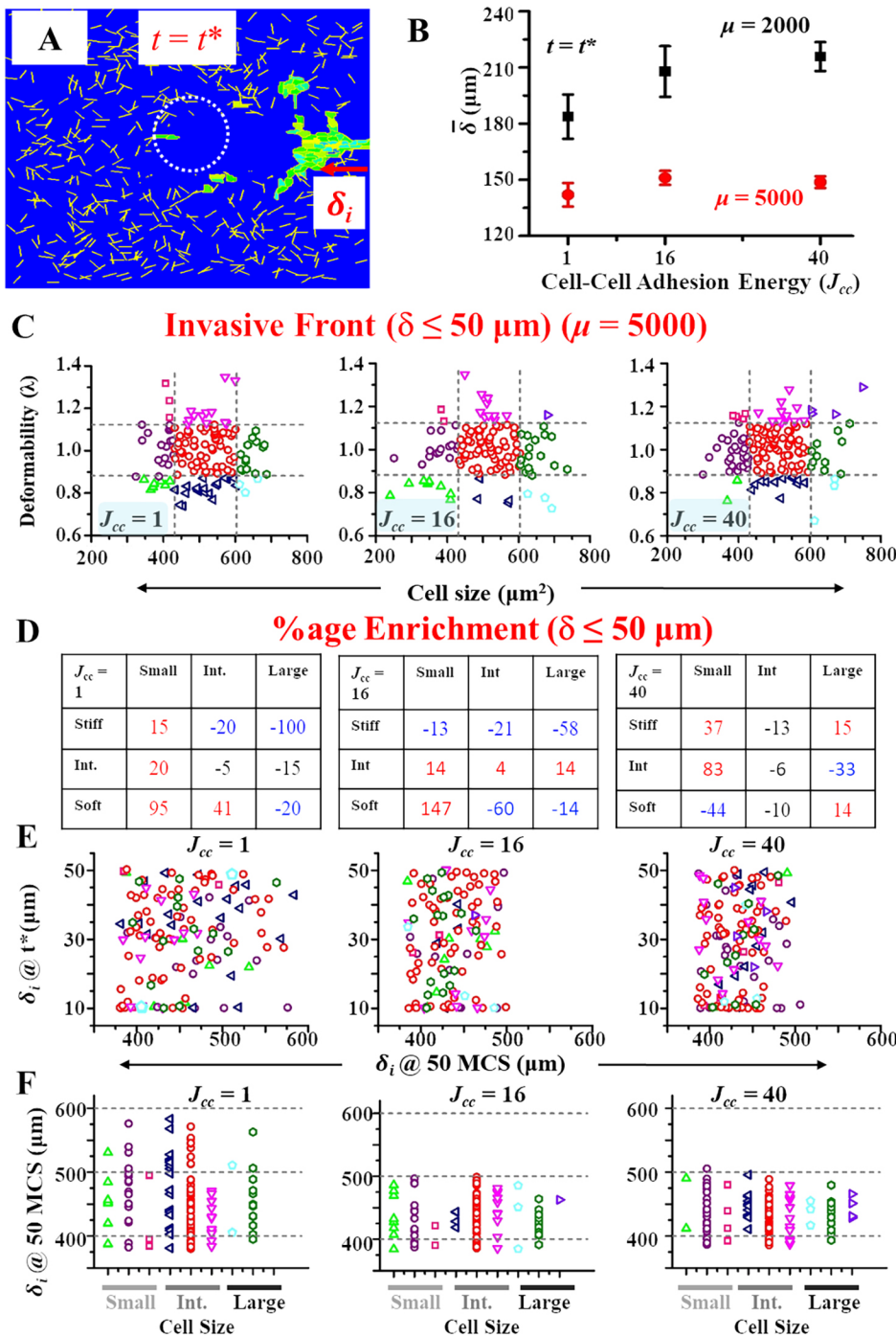
To compare and contrast the biophysical properties of population  $A$  with that of CSCs, a biophysical characterization of CD44<sup>hi</sup> CD24<sup>-</sup> mesenchymal CSCs (herein referred to as mCSCs) and CD44<sup>hi</sup> CD24<sup>+</sup> hybrid CSCs (herein referred to as hCSCs) was

undertaken (Fig. 5D,E; Fig. S6A). MDA-MB-231 cells were found to contain  $\sim 60\%$  mCSCs and  $\sim 6\%$  hCSCs (Fig. S6A). Although mCSCs were smaller and softer compared to the parental population, hCSCs exhibited size and stiffness values comparable to the parental population (Fig. 5F,G). In addition to highlighting differences between the biophysical properties of mCSCs and hCSCs, our results are indicative of an overlap in the biophysical properties of the maximally invasive population  $A$  predicted by our simulations with that of CSCs.

### CSCs are enriched at the invasive front in a temporal manner

To experimentally determine the spatial position of CSCs during invasion, MDA-MB-231 spheroids were implanted in three-





**Fig. 4. Spatial mapping of phenotypic heterogeneity at the invasive front.**

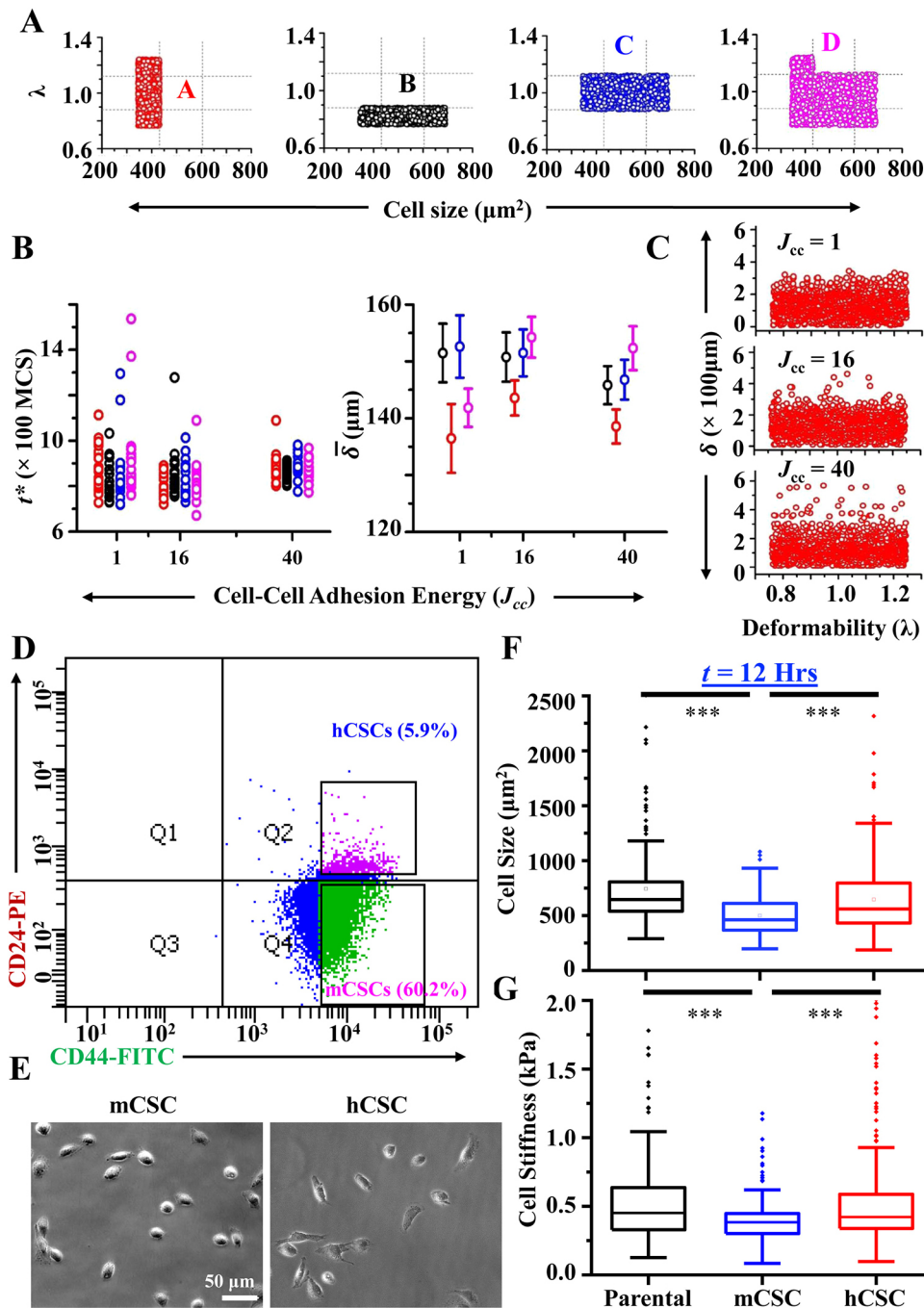
(A) Spatial distribution of phenotypically distinct subpopulations at  $t = t^*$  for  $J_{cc}=1$  and  $\mu=5000$ . The position of each cell was quantified by measuring the distance ( $\delta$ ) from the right edge of the two-dimensional ECM. (B) Plot (mean  $\pm$  s.e.m.) of  $\bar{\delta}$  for the entire population for different values of  $J_{cc}$ . (C) Plot of cell size versus cell deformability for cells at the invasive front (i.e.  $\delta \leq 50 \mu\text{m}$  at  $t = t^*$ ). Grey dotted lines drawn at  $(A \pm \sigma_A) / (\lambda \pm \sigma_\lambda)$  intervals allowed us to segregate cells into nine different groups based on size and deformability. (D) Percentage enrichment of cells of each subpopulation at the invasive front relative to the whole population. (E) For cells at the invasive front, a plot of the final position of cells (i.e.  $\delta_i$  at  $t^*$ ) versus their initial position (i.e.  $\delta_i$  at  $t=50$  MCS) for different values of  $J_{cc}$ . (F) Initial position of cells (i.e.  $\delta_i$  at  $t=50$  MCS) for the nine different groups of cells segregated on the basis of size and deformability. For ease of visualization, each cell was color coded based on its size and deformability (green top-pointed triangles, small and soft cells; purple circles, small cells with intermediate deformability; pink squares, small and stiff cells; dark blue left-pointed triangles, soft cells of intermediate size; red circles, intermediate stiff cells of intermediate size; pink bottom-pointed triangles, stiff cells of intermediate size; light blue pentagons, large and soft cells; green hexagons, large cells of intermediate deformability; purple right-pointed triangles, large and stiff cells).

dimensional collagen gels and the spatiotemporal positioning of CSCs was assessed. Although the proportion of mCSCs within the spheroids was nearly identical to that in MDA-MB-231 cells cultured in two-dimensional collagen-coated substrates, the proportion of hCSCs was higher (Fig. S6B,C). The temporal pattern of cell scattering was quantified 1, 3 and 5 days after implantation (day 0) (Fig. 6A). Spheroids were implanted in 1.5 mg/ml collagen gels with an average pore size of  $\sim 2 \mu\text{m}$  (Fig. 6B). To probe the initial spatial distribution of CSCs and non-CSCs (nCSCs) within the spheroid, cells were stained for CD44, F-actin and DAPI (Fig. 6C). Quantification of integrated CD44 and F-actin intensity as a function of radial position from the spheroid center revealed a uniform distribution, suggesting that

CSCs (i.e. CD44<sup>hi</sup> cells) were randomly positioned within the spheroid (Fig. 6D). Temporal tracking of spheroids revealed an increase in spheroid size, as well as an outward scattering of cells from the spheroid core (Fig. 6E).

To probe for the possible enrichment of CSCs at the invasive front, the spheroid core of day 5 spheroids were divided into seven zones, and the average integrated CD44 intensity was computed for each zone (Fig. 6F). Zones 0-2 were chosen such that they spanned the area occupied by day 0 spheroids. Plots of integrated CD44, intensity normalized with respect to the average CD44 intensity in the innermost zone (i.e. zone 0) at day 1, day 3 and day 5, revealed time-dependent enrichment of CD44<sup>hi</sup> cells at the outermost zone (Fig. 6G). A threefold enrichment in the outermost zone (i.e. zone 4)





**Fig. 5. Small cells of varying deformabilities may correspond to CSCs.**

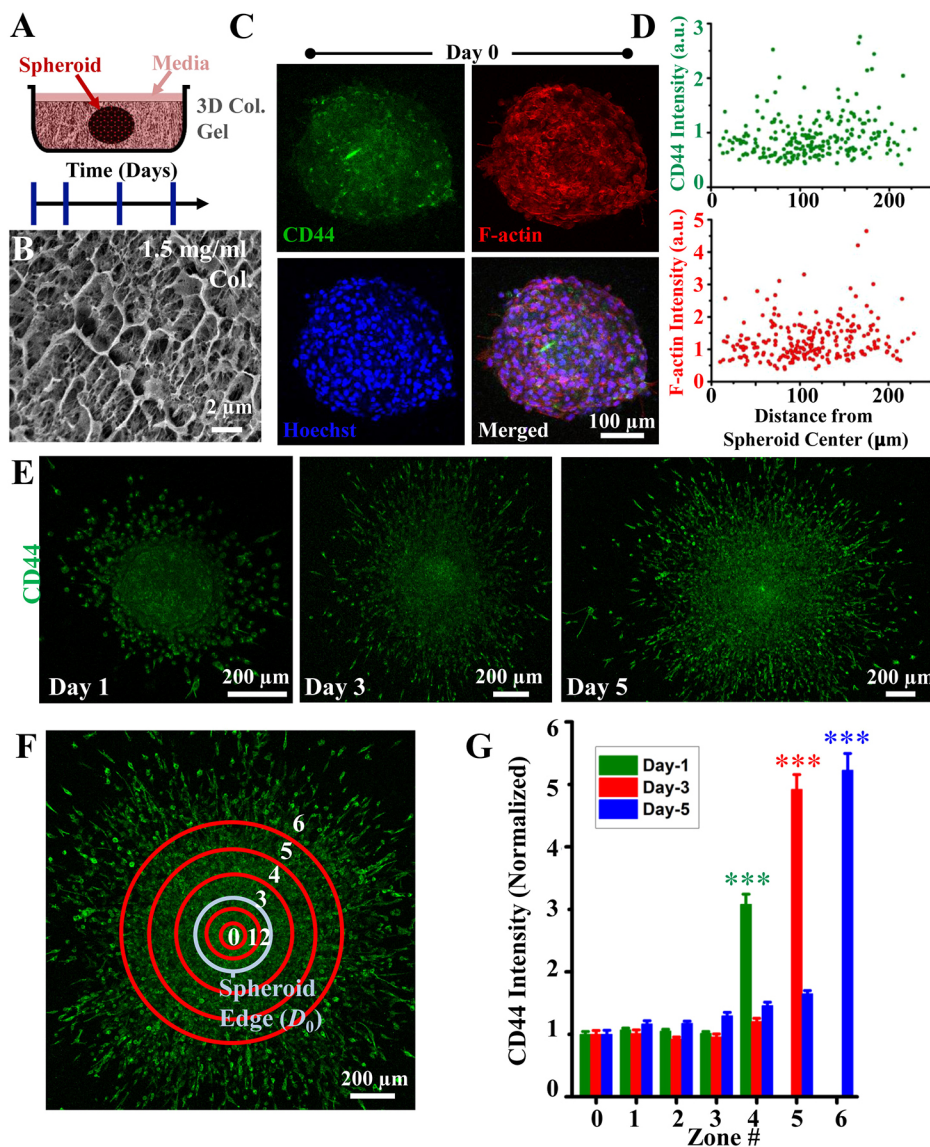
(A) Distribution of cell size and deformability for four distinct populations: small cells of varying deformabilities (A), soft cells of varying cell sizes (B), intermediately stiff cells of varying cell sizes (C), and a combination of populations A, B and C, i.e. population D. (B) Quantification (mean  $\pm$  s.e.m.) of  $t^*$  and  $\delta$  for populations A to D. (C) Dependence of  $\delta$  on deformability for population A. (D) FACS expression profile of CD44 and CD24 in MDA-MB-231 cells. Q1, Q2, Q3 and Q4 represent quadrant gating of FACS. Inset boxes correspond to gating of FACS to get pure populations of mCSCs and hCSCs. (E) Representative phase contrast images of CD44<sup>hi</sup> CD24<sup>-</sup> mCSCs and CD44<sup>hi</sup> CD24<sup>+</sup> hCSCs. Scale bar: 50  $\mu\text{m}$ . (F) Cell size distribution of parental, mCSCs and hCSCs after 12 h culture on collagen-coated glass coverslips ( $n > 190$  cells per condition pooled from  $N = 3$  independent experiments). (G) Cell stiffness distribution of parental, mCSCs and hCSCs after 12 h culture on collagen-coated glass coverslips ( $n = 180$  cells per condition pooled from  $N = 3$  independent experiments). \*\*\* $P < 0.001$  (one-way ANOVA/Fisher test). In box-and-whisker plots, the box represents the 25–75th percentiles, and the median (line) and mean (small square) is indicated. The whiskers show 1.5 $\times$  of interquartile range (IQR). Outliers are indicated.

at day 1 increased to fivefold enrichment at day 3 in zone 5, and remained unchanged thereafter at day 5 in the outermost zone 6.

To probe the mechanism underlying the enrichment of CD44<sup>hi</sup> cells at the invasive front, interconversion experiments were performed wherein mCSCs, hCSCs and nCSCs (CD44<sup>-low</sup> cells) were isolated from parental MDA-MB-231 cells by fluorescence-activated cell sorting (FACS) and individually expanded up to a duration of 48 h (Fig. 7A). FACS analysis of the individually grown cultures allowed us to determine the rates of interconversion in these three subpopulations. Although both nCSCs and hCSCs converted to mCSCs, mCSCs mostly exhibited symmetric division (Fig. 7B). Collectively, these results suggest that enrichment of CD44<sup>hi</sup> cells at the invasive front may be partly associated with transdifferentiation of nCSCs into mCSCs.

## DISCUSSION

In this paper, we began by mapping biomechanical heterogeneity in breast cancer cell lines by measuring cell size and cell deformability of two cell types, which vary in the extent of cell-cell adhesion. Although both cell types exhibit overlapping size and deformability distributions, MDA-MB-231 cells are on an average larger in size and more motile compared to MCF-7 cells. Our simulations predict that cancer invasiveness is maximum when heterogeneity in both cell size and cell deformability is present. Although a directional cue may drive collective invasion of the entire cell aggregate independent of the extent of cell-cell adhesion, invasiveness is most efficient at intermediate cell-cell adhesion. Analysis of subpopulations present at the invasive front suggests that small cells of varying deformabilities represent the most invasive subpopulation and may correspond to a



**Fig. 6. Temporal enrichment of CSCs at the invasive front.** (A) Schematic of the spheroid invasion assay. Single spheroids are implanted into three-dimensional collagen gels and their invasion tracked over a period of 6 days. (B) Representative cryo FEG-SEM image of 1.5 mg/ml collagen gel. (C) Representative image of a spheroid stained for CD44, phalloidin (for visualizing the F-actin cytoskeleton) and Hoechst 33342 for imaging nuclei. (D) Radial distribution of CD44 and F-actin intensity in cells within the spheroid at day 0. (E) Representative images of spheroids at days 1, 3 and 5, visualized by CD44 staining. (F) Segregation of invasive area of day 5 spheroids into seven zones. Zones 0, 1 and 2 were contained within the area encompassing day 0 spheroids. (G) Temporal zone-wise distribution of integrated CD44 intensity normalized with respect to CD44 intensity of the innermost zone (i.e. zone 0) (mean±s.e.m.;  $n > 40$  cells per zone per time point pooled from  $N = 2$  spheroids). \*\*\* $P < 0.001$  (one-way ANOVA/Fisher test for parametric data and Mann–Whitney test for non-parametric data). a.u., arbitrary units.

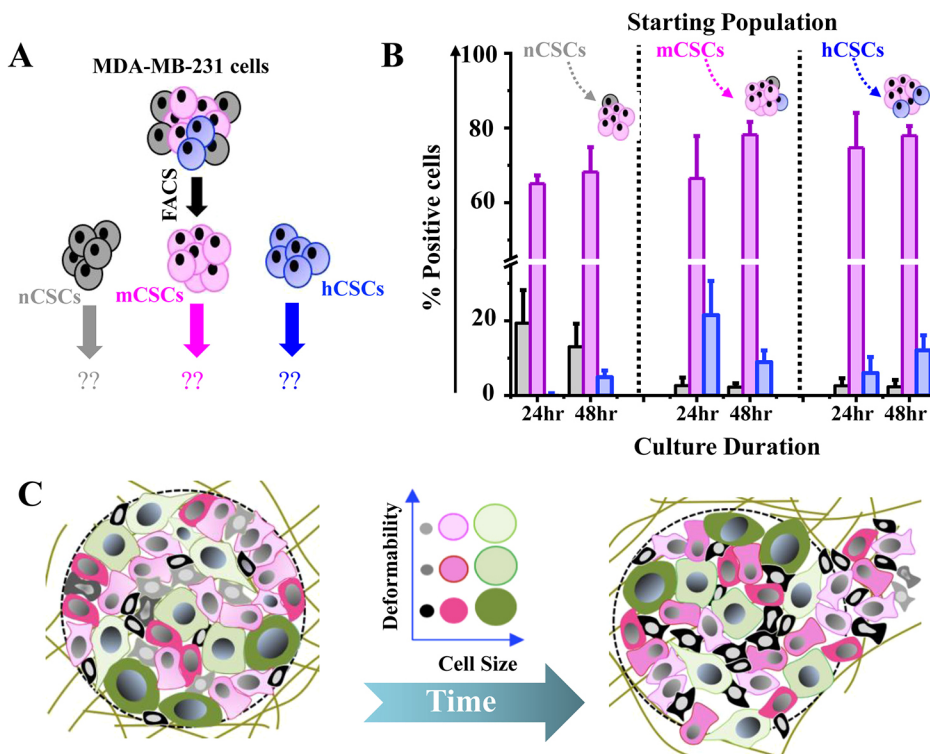
subpopulation of  $CD44^{hi}$  CSCs. In addition to highlighting the potential role of combined heterogeneity in cell size and deformability in hastening cancer invasion, our results suggest biophysical heterogeneity drives invasiveness (Fig. 7C). Although our simulations were performed keeping the extent of cell-cell adhesion fixed for a given simulation, variable extent of cell-cell adhesion, which is likely to be more relevant *in vivo*, is expected to generate a combination of single cells and cell clusters at the invasive front.

Our results predict that heterogeneity in cell size alone or deformability alone is not sufficient to positively influence cancer invasiveness; instead, heterogeneity in both cell size and deformability is required. Combined heterogeneity may promote invasiveness by allowing cells to continuously reposition themselves within the cell cluster. Such internal reorganization is particularly evident during collective migration with intact cell-cell adhesions (i.e.  $J_{cc} = 1$ ), wherein cells cover the largest distances to reach the boundary and also undergo maximum neighbor changes. This type of re-arrangement enabled the enrichment of cells from the tumor interior, with these cells being softer than the bulk population. In contrast, at moderate and low cell-cell adhesion, as cells are more prone to scattering, the scope for re-arrangement may be lesser. Under

these conditions, initial position of cells is likely to play a more important role in determining their presence at the invasive front. Highest population invasiveness observed for  $J_{cc} = 16$  may be attributed to the synergistic effects of intermixing and initial positioning. Our observation of fastest chemotaxis observed at intermediate cell-cell adhesion is consistent with the notion that groups of cells exhibit better gradient-sensing via collective guidance through the regulation of contact inhibition of locomotion (CIL) (Camley et al., 2016).

In skin and mammary primary tumors, screening of a large panel of cell surface markers identified different tumour subpopulations exhibiting different extents of EMT, with complete EMT requiring cells to transit through multiple intermediate hybrid states (Pastushenko et al., 2018). The hybrid epithelial/mesenchymal state is associated with higher tumor initiating potential, increased metastasis and chemoresistance (Grosse-Wilde et al., 2015; Biddle et al., 2016; Navas et al., 2020). In breast cancer cells, although  $CD44^{hi} CD24^{-}$  CSCs are mesenchymal in nature,  $ALDH^{+}$  CSCs, as well as  $CD44^{hi} CD24^{+}$  CSCs, exhibit a hybrid epithelial/mesenchymal phenotype (Liu et al., 2014; Grosse-Wilde et al., 2015). In our simulations,  $J_{cc} = 16$  may correspond to this hybrid epithelial/mesenchymal state in which both single cells and cell clusters





**Fig. 7. Interconversion between CSCs and proposed model of invasion by a phenotypically heterogeneous tumor.** (A) Experimental setup for probing interconversion between different CSC subpopulations. nCSCs, mCSCs and hCSCs were isolated from MDA-MB-231 cells using FACS and then expanded separately. After 24 and 48 h, the individual cell fractions were re-analyzed using FACS to assess the relative proportions of CSCs in each cell fraction. (B) Relative proportion of nCSCs, mCSCs and hCSCs in cell fractions obtained by expanding nCSCs, mCSCs and hCSCs for 24 and 48 h ( $N=3$  independent experiments; data are mean $\pm$ s.e.m.) Insets show pictorial representations of the compositions of the cell fractions after 48 h. (C) Proposed model of invasion by a tumor comprising cells of varying size and deformability. The relative proportions of phenotypically distinct subpopulations at the invasive front is dependent on the extent of cell-cell adhesion.

(i.e. groups of cells with intact cell-cell contacts) were observed. Cells of intermediate deformability, predicted by our simulations to be enriched at the invasive front for  $J_{cc}=16$ , may correspond to a subset of the  $CD44^{hi} CD24^{+}$  hCSCs. Although nCSCs gave rise to predominantly mCSCs in our experiments, conversion of nCSCs to a  $CD44^{hi} CD24^{+}$  drug tolerant state in the presence of the chemotherapeutic drug docetaxel highlights the importance of stress-induced phenotypic plasticity (Goldman et al., 2015). The extent to which phenotypic plasticity is influenced by physical stresses felt by the cell nucleus during three-dimensional migration remains to be explored (Mukherjee et al., 2020).

A correlation between smaller cell size and stemness has been speculated, both in normal and diseased contexts (Lia et al., 2015). Consistent with this notion, compared to larger PC3 pancreatic cancer cells, smaller-sized cells were found to possess more CSC-like characteristics, including clonogenic and tumorigenic properties (Li et al., 2008). Similarly, spheres generated by culturing A431 epidermoid cancer cells under non-adherent conditions were enriched with podoplanin $^{+}$  small cells with increased tumorigenic potential (Bortolomai et al., 2010). In SUM149 inflammatory breast cancer cells, ALDH $^{+}$  CSCs have been shown to be softer compared to the parental population (Chen et al., 2019). Here, we show that in MDA-MB-231 cells, the two CSC subpopulations differ considerably in their size and deformability. Although  $CD44^{hi} CD24^{-}$  mCSCs were significantly smaller and softer than parental cells, similar to the above studies,  $CD44^{hi} CD24^{+}$  hCSCs exhibited a wide variation in their size and stiffness. Interconversion experiments suggest that although both nCSCs ( $CD44^{low}$  cells) and hCSCs convert primarily to mCSCs, mCSCs exhibit symmetric division. This may partly explain the enrichment of  $CD44^{hi}$  cells at the invasive front. As mCSCs represent the majority fraction of  $CD44^{hi}$  cells, it is likely that the majority of the  $CD44^{hi}$  cells at the spheroid invasive front are mCSCs. In line with this, a computational study predicted that EMT-inducing factors, such as TGF $\beta$ , secreted by stromal cells can give rise to the spatial segregation of different CSC subsets with mCSCs localized at the

migration front (Bocci et al., 2019b). However, whether or not such spatial segregation of CSCs also occurs in cell types with moderate or strong cell-cell adhesion remains to be explored.

Using computational modeling, we have previously shown that, in addition to the proportion of CSCs, the extent of their spatial scattering also influences tumor growth (Kumar et al., 2016a). In our spheroid invasion assays, CSCs and nCSCs were randomly localized at day 0. However, within 24 h, enrichment of  $CD44^{hi}$  cells in the outermost zone of the spheroids was clearly observed, suggesting that CSCs serve as leader cells. This is in line with previous reports in primary human breast cancers wherein  $CD44^{hi} CD24^{-}$  mesenchymal cells localized at the tumor invasive front, and ALDH $^{+}$  cells were found in the tumor interior (Liu et al., 2014). Although similar localization of  $CD44^{hi}$  cells at the invasive front was observed during the collective invasion of luminal breast cancer cells, these cells were found to exhibit a hybrid epithelial/mesenchymal gene signature (Yang et al., 2019). Although Notch signaling is known to induce EMT via its ligands Delta and Jagged, EMT induction through Jagged gives rise to clusters of hybrid epithelial/mesenchymal cells (Boareto et al., 2016). The proportion of these hybrid epithelial/mesenchymal cells and the size of the cell clusters has been predicted to depend on the rate at which cells undergo EMT and cell-cell adhesion (Bocci et al., 2019a). This model can successfully reproduce the typical cluster size distributions observed in CTCs isolated from patient blood.

In conclusion, in this study we have established the advantage of phenotypic heterogeneity in breast cancer invasion. Future experiments taking cell cycle stage into account can help us assess the contribution of cell cycle heterogeneity to our measurements of biophysical heterogeneity. Our multiscale stochastic computational model recapitulating experimental observations predicts a potential role of combined heterogeneity in cell size and deformability in promoting cancer invasion, with gradual enrichment of small cells at the invasive front, which may experimentally correspond to CSCs.



Future studies focused on probing interactions between the different subpopulations may provide us with strategies to inhibit heterogeneity-driven cancer invasion.

## MATERIALS AND METHODS

### Experimental methods

MCF-7 and MDA-MB-231 breast cancer cell lines were obtained from National Center for Cell Science (Pune, India) and cultured in high glucose Dulbecco's modified Eagle medium (DMEM, Invitrogen, AL007A) containing 10% fetal bovine serum (FBS, HiMedia, RM9952), maintained at 37°C at 5% CO<sub>2</sub> humidified atmosphere, and passaged using 0.25% trypsin-EDTA (HiMedia, TCL099). For experiments, cells were cultured for 24 h on glass coverslips coated with rat tail collagen I (Sigma-Aldrich, C3867) at a concentration of 25 µg/ml. For E-cadherin staining, cells were plated sparsely on collagen-coated glass coverslips. After 72 h, cells were fixed, permeabilized with 0.5% Triton-X 100 (Sisco Research Laboratories, 64518) and stained with anti-mouse E-cadherin antibody (Cell Signal Technology, 14472, 1:400) overnight at 4°C. Next day, after washing, cells were incubated with goat-anti-mouse IgG Alexa Fluor 555 (Invitrogen, A21422, 1:1000) for 2 h at room temperature. After staining nuclei with Hoechst 33342 (Thermo Fisher Scientific, H3570), images were captured using an Olympus IX81 inverted microscope at 40× magnification. Quantification of E-cadherin intensity was performed using Fiji Image J software by tracking the line intensities across the cell edges and quantifying the ratio of membrane to cytoplasmic E-cadherin intensity.

For cell size measurement, after 24 h of culture on collagen-coated glass coverslips, cells were imaged at 10× magnification using an inverted microscope (Olympus IX71). Cell area was quantified using ImageJ software by manually outlining cells. For tracking temporal evolution of cell size after 24 h culture, cells were imaged every 10 mins over a duration of 24 h (Zeiss Spinning Disk Microscope). The same movies were analyzed for measuring random cell motility based on the positions of cell centroids determined by manually outlining the periphery of individual cells in ImageJ. Cell cortical stiffness measurements were obtained by indenting cells with soft pyramidal probes of nominal spring constant 160 pN/nm (22 kHz, Olympus, TR800PB), and fitting the first 500 nm of indentation data with the Hertz model. For quantification of cell size, cell motility and cell cortical stiffness, experiments were repeated three times and a comparable number of measurements from independent experiments were pooled.

For flow cytometric analysis, MDA-MB-231 cells were stained with fluorescence-labeled antibodies CD44-FITC (BD biosciences, 555478) and CD24-PE (BD biosciences, 555428) as per the manufacturer's protocol. In brief, after trypsinization, MDA-MB-231 cells were incubated with CD44-FITC and CD24-PE for 45 min. After labeling, cells were washed with FACS buffer (PBS with 1% FBS) and cell sorting was performed in FACS Aria Fusion (Becton Dickinson), using FACS Diva software (Becton Dickinson). Suitable negative isotype controls were used to rule out the background fluorescence. Sorted cells were collected in complete medium (DMEM high glucose plus 10% FBS plus Antibiotic-Antimycotic solution (HiMedia, A002A) to perform further experiments. The percentage of each positive population was determined using quadrant or region statistics. Data were analyzed using FACS Diva Software (Beckton Dickinson).

For spheroid invasion experiments, MDA-MB-231 cells were trypsinized and resuspended in medium. Spheroids were generated using the hanging drop method by suspending 4000 cells in 15 µl medium supplemented with 6.25 µg/ml rat tail collagen I (Corning, 354249). The drops were then incubated at 37°C in 5% CO<sub>2</sub> for 48 h to generate spheroids. After isolation, spheroids were seeded on a bed of 1.5 mg/ml collagen gels, and then incubated with three-dimensional collagen solution at 37°C for gelation to occur. After 30–45 mins, wells were flooded with medium. Cells were fixed and stained for nuclei, F-actin (Thermo Fisher Scientific, A22282) and CD44 (Novus, NBP1-47386) at day 0 (2 h after spheroids were embedded in collagen gels), day 1, day 3 and day 5. For quantification of CD44 intensity, images were captured using an LSM microscope (Zeiss) at 10× magnification under identical gain and exposure settings for all the conditions. Captured images were processed and analyzed using Fiji

ImageJ. Raw integrated intensities of single cells were determined by manually outlining individual cells using the selection tool in Fiji ImageJ. For obtaining the normalized zone-wise distribution of CD44 intensity, the average integrated CD44 intensity of each zone, determined using the raw intensities of cells within that zone, was divided by the average integrated CD44 intensity of the innermost zone. This work was approved by the biosafety committee of the institute.

### Computational methods

For studying the impact of phenotypic heterogeneity on cancer invasion, we adapted our recently developed CPM-based two-dimensional model of cancer invasion, taking cell size and cell deformability into account (Kumar et al., 2016b). In our model, a cell aggregate (comprised of 69 cells) is situated at the center of a two-dimensional ECM lattice (1 mm×1 mm size) comprised of randomly positioned ECM fibers, with each ECM fiber being 2 µm in thickness and 30–40 µm in length. In simulations taking phenotypic heterogeneity into account, size and area constraint ( $\lambda_a$ ) of individual cells were assigned values from normal distributions as depicted in Fig. 2B. Although cell sizes were drawn from a normal distribution with mean size ( $\bar{A} = 517.5 \mu\text{m}^2$  and s.d.  $\sigma_A = 86.1 \mu\text{m}^2$ , cell deformabilities (i.e.  $\lambda_a$ 's) were drawn from a normal distribution with  $\bar{\lambda} = 1 E/L^4$  and s.d.  $\sigma_\lambda = 0.12 E/L^4$ . For homogeneous cell clusters, cell size and cell deformability of all the cells were set to the means of the above distributions, i.e.  $A=517.5 \mu\text{m}^2$  and  $\lambda=1E/L^4$ . Based on cell size and deformability, for initial analysis, the heterogeneous population was divided into five groups, namely small and soft cells ( $A < (\bar{A} - \sigma_A)$ ,  $\lambda_a < (\bar{\lambda}_a - \sigma_\lambda)$ ), small and stiff cells ( $A < (\bar{A} - \sigma_A)$ ,  $\lambda_a > (\bar{\lambda}_a + \sigma_\lambda)$ ), cells with intermediate size and deformability ( $(\bar{A} - \sigma_A) \leq A \leq (\bar{A} + \sigma_A)$ ,  $(\bar{\lambda}_a - \sigma_\lambda) \leq \lambda_a \leq (\bar{\lambda}_a + \sigma_\lambda)$ ), large and soft cells ( $A > (\bar{A} + \sigma_A)$ ,  $\lambda_a < (\bar{\lambda}_a - \sigma_\lambda)$ ), and large and stiff cells ( $A > (\bar{A} + \sigma_A)$ ,  $\lambda_a > (\bar{\lambda}_a + \sigma_\lambda)$ ). Apart from ECM fibers and cells, the remaining pixels in the lattice were considered as 'fluid' pixels mimicking the interstitial fluid. The different extent of adhesion between the different types of pixels was accounted for by using distinct adhesion energies ( $J_{xy}$ ) between entities  $x$  and  $y$ . Although cell-matrix ( $J_{cm}$ ), cell-fluid ( $J_{cf}$ ), matrix-fluid ( $J_{mf}$ ) and fluid-fluid ( $J_{ff}$ ) adhesion energies were held constant ( $J_{cm}=16 k_B T/L$ ,  $J_{cf}=32 k_B T/L$ ,  $J_{mf}=J_{ff}=35 k_B T/L$ ), cell-cell adhesion energies ( $J_{cc}$ ) were varied to mimic different extents of cell-cell adhesion, with  $J_{cc}=1 k_B T/L$  mimicking collective cell migration and  $J_{cc}=40 k_B T/L$  mimicking single cell migration.

Spatiotemporal evolution of the simulation lattice was based on random movement of individual pixels subject to transition probabilities based on the Monte Carlo method. Each simulation step consisted of randomly choosing two neighboring pixels, with one designated as the source pixel and the other as the target pixel. An attempt to update the lattice was made only when both the source and the target pixels represented either a 'cell pixel' or a 'fluid pixel' (i.e. empty space), and the source and the target pixels did not belong to the same cell. The proposed change is accepted with probability  $p=1$  if  $\Delta E < 0$  and  $p = e^{-\Delta E/k_B T_m}$  otherwise, where  $\Delta E$  represents a change in system energy due to the proposed change, and  $k_B$  represents the Boltzmann constant.  $T_m$  represents noise in the system and was set to 0.01  $T$ . An attempt to update the lattice was made only when both the source and the target pixels represented either a 'cell pixel' or a 'fluid pixel', with 'matrix pixels' not participating in the random Monte Carlo updates. The total system energy is given by the expression:

$$E_{Total} = \sum_{\forall i,j \text{ and } \sigma(i) \neq \sigma(j)} J_{\tau(\sigma(i)), \tau(\sigma(j))} + \sum_{\forall \sigma} \lambda_a (a(\sigma) - a_0)^2 + \sum_{\forall \sigma} \lambda_p (p(\sigma) - p_0)^2 + w(\sigma) - \mu \times [v(target) - v(source)]. \quad (1)$$

In the above expression,  $\sigma(i)$  represents the ID of pixel  $i$  and  $\tau(\sigma)$  represents the type of cell. The first term in the energy expression (i.e.  $J_{\tau_1, \tau_2}$ ) represents the boundary energy per unit length between cells of type  $\tau_1$  and  $\tau_2$ , and is indicative of the adhesion energy between two cells. The second and third terms represent the energy associated with changes in size and perimeter of cells from their preferred area ( $a_0$ )/perimeter ( $p_0$ ). Although the area constraint ( $\lambda_a$ ) represents the bulk stiffness or inverse compressibility of a given cell, the perimeter constraint ( $\lambda_p$ ) is indicative of line tension. In our

simulations, whereas  $\lambda_a$  of individual cells were assigned values as detailed above, a constant value of  $\lambda_p=0.5k_B T/L^2$  was chosen for all cells as performed previously (Kumar et al., 2016b). For a cell of target size  $a_0$ , the target perimeter ( $p_0$ ) was set to  $2\sqrt{\pi a_0}$ . The fourth energy term [ $w(\sigma)$ ] is associated with the active motility of a cell due to its inherent polarization, and given by the expression  $w(\sigma) = -\mu_0 \hat{p}$ , where  $\mu_0$  represents the strength of motility and  $\hat{p}$  represents the polarity vector, as described previously (Kabla, 2012). Although  $\mu_0=50k_B T/L$  was kept constant across all simulations,  $\hat{p}$  at time  $t$  was defined as the average of the previous ten displacement vectors, as performed in our previous work (Kumar et al., 2016b). The last term in the energy expression was included to model chemotaxis of cells in the direction of the chemoattractant gradient (Kumar et al., 2018b), with  $\mu$  representing the effective chemical potential, and  $v(\text{target})$  and  $v(\text{source})$  representing the concentrations of chemoattractant at target and source pixel, respectively. This term was used in the energy expression only for simulating cancer invasion in the presence of a chemotactic field. Parametric studies were performed for  $\mu=\{200, 500, 1000, 2000, 5000\}$ . For  $\mu \geq 2000$ , few cells reached the right boundary (i.e. location of highest chemokine concentration) within the maximum simulation duration of 2010 MCS for all values of  $J_{cc}$ . Thus, robust chemotaxis was observed for  $\mu \geq 2000$ .

Cancer cells are known to invade by MMP-mediated degradation of the surrounding ECM depending on the physicochemical properties of the ECM, including ligand density and stiffness (Das et al., 2013, 2017). Soluble MMPs activated by membrane-bound MT1-MMP then diffuse into the extracellular space and degrade the surrounding ECM (Kumar et al., 2018a). To incorporate this effect, we have integrated our CPM model with reaction-diffusion MMP dynamics. To mimic the ECM density-dependent MMP secretion profile observed in breast cancer cells, in our model, the number of MMP molecules secreted per unit time by a cell was assumed to be  $0.05 \text{ sec}^{-1}$  at the site of cell-fiber contact. Diffusion and degradation of soluble MMPs were incorporated in our model using the reaction diffusion formalism given

by the equation  $\frac{\partial \text{MMP}(\mathbf{x}, t)}{\partial t} = D \cdot \nabla^2 \text{MMP}(\mathbf{x}, t) - \delta \times \text{MMP}(\mathbf{x}, t)$ , where  $\text{MMP}(\mathbf{x}, t)$  represents the concentration of MMP molecules at point  $\mathbf{x}$  and time  $t$ ,  $D$  represents the diffusion coefficient, and  $\delta$  the degradation rate of soluble MMPs.  $D$  and  $\delta$  were chosen to be  $1.0 \times 10^{-9} \text{ cm}^2 \text{ sec}^{-1}$  and  $0.002 \text{ sec}^{-1}$ , respectively.

#### Simulation implementation, visualization and data analysis

The complete simulation framework was implemented using the open source package CompuCell3D (CC3D) (Swat et al., 2012). For visualization, \*.vtk files were generated from the CC3D simulations and visualized in CC3D itself. For quantifying migration trajectories, the cell centroid was tracked and logged to CSV files. These files were then processed in R using custom written scripts to extract different invasiveness metrics. The CC3D code used to implement the complete model is available at <https://github.com/sandeep13712/senlab-phenotypicheterogeneity>.

#### Quantification of heterogeneity from single-cell RNA-seq datasets

The breast cancer dataset was downloaded from the GEO repository (GSE75688) (Chung et al., 2017). Non-tumor and bulk samples were removed from the dataset. Then, control RNA readings were removed, and multiple readings for the same gene were added up to get a dataset with unique gene names. An obtained preprocessed dataset was then used for GSEA using the single sample GSEA method (Barbie et al., 2009). Stemness and EMT enrichment analyses were performed using MALTA\_CURATED\_STEMNESS\_MARKERS and HALLMARK\_EPITHELIAL\_MESENCHYMAL\_TRANSITION gene sets downloaded from the Molecular Signatures Database (Subramanian et al., 2005; Liberzon et al., 2011; Malta et al., 2018). For calculating the enrichment of biophysical characteristics, a custom gene set BiophysicalGeneset comprising MYH9, MYH10, MYO10, MYO5B, MYO5C, RAC1, RAC2, ACTB, ACTG1, ACTN1, ACTN4, VIM, ROCK1, ROCK2 and CDC42 was manually curated and used. For tSNE visualization, data were first log transformed and then the first ten principal components were used for computing the two-dimensional-tSNE projection. To quantify biophysical heterogeneity in the MCF7 and MDA-MB-231 cell lines from the

GSE115869 and GSE124989 datasets (Birts et al., 2020; Jonasson et al., 2019), downloaded datasets were log-transformed after adding 1 to all transcript counts and then the distributions of genes relevant for different aspects of cellular mechanics were quantified. Analysis was performed using Python scripts [available in github repository (<https://github.com/sandeep13712/senlab-phenotypicheterogeneity>)].

#### Acknowledgements

We thank IIT Bombay-Industrial Research and Consultancy Centre for providing Bio-AFM, FACS and confocal microscopy facilities.

#### Competing interests

The authors declare no competing or financial interests.

#### Author contributions

Conceptualization: A., S.K., S.S.; Methodology: A., S.K., N.S., M.S., A.B., S.S.; Software: A., S.K.; Formal analysis: A., S.K., N.S., M.S., A.B.; Investigation: A., S.K., N.S., M.S., A.B., S.S.; Resources: S.S.; Data curation: A., S.K., N.S., M.S., A.B.; Writing - original draft: A., S.K., S.S.; Supervision: S.S.; Project administration: S.S.; Funding acquisition: S.S.

#### Funding

This work was supported by Science and Engineering Research Board (SERB), Department of Science and Technology, Ministry of Science and Technology, Government of India (EMR/2016/005454). N.S. was supported by a fellowship from the Department of Science and Technology, Government of India.

#### Supplementary information

Supplementary information available online at <https://jcs.biologists.org/lookup/doi/10.1242/jcs.250225.supplemental>

#### Peer review history

The peer review history is available at <https://jcs.biologists.org/lookup/doi/10.1242/jcs.250225.reviewer-comments.pdf>

#### References

- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. and Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100**, 3983-3988. doi:10.1073/pnas.0530291100
- Arozarena, I. and Wellbrock, C. (2019). Phenotype plasticity as enabler of melanoma progression and therapy resistance. *Nat. Rev. Cancer* **19**, 377-391. doi:10.1038/s41568-019-0154-4
- Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C. et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-112. doi:10.1038/nature08460
- Biddle, A., Gammon, L., Liang, X., Costea, D. E. and Mackenzie, I. C. (2016). Phenotypic plasticity determines cancer stem cell therapeutic resistance in oral squamous cell carcinoma. *eBiomedicine* **4**, 138-145. doi:10.1016/j.ebiom.2016.01.007
- Birts, C. N., Banerjee, A., Darley, M., Dunlop, C. R., Nelson, S., Nijjar, S. K., Parker, R., West, J., Tavassoli, A., Rose-Zerilli, M. J. J. et al. (2020). p53 is regulated by aerobic glycolysis in cancer cells by the CtBP family of NADH-dependent transcriptional regulators. *Sci. Signal* **13**, aau9529. doi:10.1126/scisignal.aau9529
- Boareto, M., Jolly, M. K., Goldman, A., Pietila, M., Mani, S. A., Sengupta, S., Ben-Jacob, E., Devi, G. R., Levine, H. and Onuchic, J. N. (2016). Notch-Jagged signalling can give rise to clusters of cells exhibiting a hybrid epithelial/mesenchymal phenotype. *J. R. Soc. Interface* **13**, 20151106. doi:10.1098/rsif.2015.1106
- Bocci, F., Jolly, M. K. and Onuchic, J. N. (2019a). A biophysical model uncovers the size distribution of migrating cell clusters across cancer types. *Cancer Res.* **79**, 5527-5535. doi:10.1158/0008-5472.CAN-19-1726
- Bocci, F., Gearhart-Serna, L., Boareto, M., Ribeiro, M., Ben-Jacob, E., Devi, G. R., Levine, H., Onuchic, J. N. and Jolly, M. K. (2019b). Toward understanding cancer stem cell heterogeneity in the tumor microenvironment. *Proc. Natl. Acad. Sci. USA* **116**, 148-157. doi:10.1073/pnas.1815345116
- Bortolomai, I., Canevari, S., Facetti, I., De Cecco, L., Castellano, G., Zacchetti, A., Alison, M. R. and Miotti, S. (2010). Tumor initiating cells: development and critical characterization of a model derived from the A431 carcinoma cell line forming spheres in suspension. *Cell Cycle* **9**, 1194-1206. doi:10.4161/cc.9.6.11108
- Brock, A., Chang, H. and Huang, S. (2009). Non-genetic heterogeneity—A mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genetics* **9**, 265-273. doi:10.1038/nrg2556



- Burrell, R. A., McGranahan, N., Bartek, J. and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, **501**, 338–345. doi:10.1038/nature12625
- Camley, B. A., Zimmermann, F., Levine, H. and Rappel, W.-J. (2016). Emergent collective chemotaxis without single-cell gradient sensing. *Phys. Rev. Lett.* **116**, 098101. doi:10.1103/PhysRevLett.116.098101
- Chapman, A., del Ama, L. F., Ferguson, J., Kamarashev, J., Wellbrock, C. and Hurlstone, A. (2014). Heterogeneous tumor subpopulations cooperate to drive invasion. *Cell Rep.*, **8**, 688–695. doi:10.1016/j.celrep.2014.06.045
- Chen, W., Allen, S. G., Qian, W., Peng, Z., Han, S., Li, X., Sun, Y., Fournier, C., Bao, L., Lam, R. H. W. et al. (2019). Biophysical phenotyping and modulation of ALDH+ inflammatory breast cancer stem-like cells. *Small* **15**, 1802891. doi:10.1002/smll.201802891
- Cheung, K. J., Padmanaban, V., Silvestri, V., Schipper, K., Cohen, J. D., Fairchild, A. N., Gorin, M. A., Verdone, J. E., Pienta, K. J., Bader, J. S. et al. (2016). Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proc. Natl. Acad. Sci. USA* **113**, 854–863. doi:10.1073/pnas.1508541113
- Chung, W., Rum, H. H., Lee, H. O., Lee, K. M., Lee, H. B., Kim, K. T., Kim, S., Lee, J. E., Park, Y. H., Kan, Z. et al. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Comm.* **8**, 15081. doi:10.1038/ncomms15081
- Das, A., Kapoor, A., Mehta, G. D., Ghosh, S. K. and Sen, S. (2013). Extracellular matrix density regulates extracellular proteolysis via modulation of cellular contractility. *J. Carcinog. Mutagen.* **S13**, 003. doi:10.4172/2157-2518.S13-003
- Das, A., Monteiro, M., Barai, A., Kumar, S. and Sen, S. (2017). MMP proteolytic activity regulates cancer invasiveness by modulating integrins. *Sci. Rep.* **7**, 14219. doi:10.1038/s41598-017-14340-w
- Das, A., Barai, A., Monteiro, M., Kumar, S. and Sen, S. (2019). Nuclear softening is essential for protease-independent migration. *Matrix Biol.* **82**, 4–19. doi:10.1016/j.matbio.2019.01.001
- De Paiva, C. S., Pflugfelder, S. C. and Li, D. Q. (2006). Cell size correlates with phenotype and proliferative capacity in human corneal epithelial cells. *Stem Cells* **24**, 368–375. doi:10.1634/stemcells.2005-0148
- Dongre, A. and Weinberg, R. A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **20**, 69–84. doi:10.1038/s41580-018-0080-4
- Fabisiewicz, A. and Grzybowska, E. (2017). CTC clusters in cancer progression and metastasis. *Med. Oncol.* **34**, 12. doi:10.1007/s12032-016-0875-0
- Gan, Y., Li, N., Zou, G., Xin, Y. and Guan, J. (2018). Identification of cancer subtypes from single-cell RNA-seq data using a consensus clustering method. *BMC Med. Genomics* **15**, 813–824. doi:10.1186/s12920-018-0433-z
- Goldman, A., Majumder, B., Dhawan, A., Ravi, S., Goldman, D., Kohandel, M., Majumder, P. K. and Sengupta, S. (2015). Temporally sequenced anticancer drugs overcome adaptive resistance by targeting a vulnerable chemotherapy-induced phenotypic transition. *Nat. Comm.* **6**, 6139. doi:10.1038/ncomms7139
- Grosse-Wilde, A., Fouquier d'Hérouël, A., McIntosh, E., Ertaylan, G., Skupin, A., Kuestner, R. E., del Sol, A., Walters, K. A. and Huang, S. (2015). Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor survival. *PLoS ONE* **10**, e0126522. doi:10.1371/journal.pone.0126522
- Harada, T., Swift, J., Irianto, J., Shin, J.-W., Spinler, K. R., Athirasala, A., Diegmiller, R., Dingal, P. C., Dave, P., Ivanovska, I. L. et al. (2014). Nuclear lamin stiffness is a barrier to 3D migration, but softness can limit survival. *J. Cell Biol.* **204**, 669–682. doi:10.1083/jcb.201308029
- Jolly, M. K., Manib, S. A. and Levine, H. (2018). Hybrid epithelial/mesenchymal phenotype(s): The 'fittest' for metastasis? *BBA Rev. Cancer* **1870**, 151–157. doi:10.1016/j.bbcan.2018.07.001
- Jonasson, E., Ghannoum, S., Persson, E., Karlsson, J., Kroneis, T., Larsson, E., Landberg, G. and Stahlberg, A. (2019). Identification of breast cancer stem cell related genes using functional cellular assays combined with single-cell RNA sequencing in MDA-MB-231 cells. *Front. Genet.*, **10**, 500. doi:10.3389/fgene.2019.00500
- Kabla, A. J. (2012). Collective cell migration: leadership, invasion and segregation. *J. R. Soc. Interface*, **9**, 3268–3278. doi:10.1098/rsif.2012.0448
- Kapoor, A., Barai, A., Thakur, B., Das, A., Patwardhan, S. R., Monteiro, M., Gaikwad, S., Bukhari, A. B., Mogha, P., Majumder, A. et al. (2018). Soft drug-resistant ovarian cancer cells migrate via two distinct mechanisms utilizing myosin II-based contractility. *BBA Mol. Cell Res.* **1865**, 392–405. doi:10.1016/j.bbamcr.2017.11.012
- Kumar, S., Kulkarni, R. and Sen, S. (2016a). Cell motility and ECM proteolysis regulate tumor growth and tumor relapse by altering the fraction of cancer stem cells and their spatial scattering. *Phys. Biol.* **13**, 036001. doi:10.1088/1478-3975/13/3/036001
- Kumar, S., Kapoor, A., Desai, S., Inamdar, M. M. and Sen, S. (2016b). Proteolytic and non-proteolytic regulation of collective cell invasion: tuning by ECM density and organization. *Sci. Rep.* **6**, 19905. doi:10.1038/srep19905
- Kumar, S., Das, A. and Sen, S. (2018a). MMP secretion rate and inter-invasion spacing collectively govern cancer invasiveness. *Biophys. J.* **114**, 650–662. doi:10.1016/j.bpj.2017.11.3777
- Kumar, S., Das, A. and Sen, S. (2018b). Multicompartment cell-based modeling of confined migration: regulation by cell intrinsic and extrinsic factors. *Mol. Biol. Cell* **29**, 1599–1610. doi:10.1091/mbc.E17-05-0313
- Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N. and Werb, Z. (2018). Tumour heterogeneity and metastasis at single-cell resolution. *Nat. Cell Biol.* **20**, 1349–1360. doi:10.1038/s41556-018-0236-7
- Li, H., Chen, X., Calhoun-Davis, T., Claypool, K. and Tang, D. G. (2008). PC3 human prostate carcinoma cell holoclones contain self-renewing tumor-initiating cells. *Cancer Res.* **68**, 1820–1825. doi:10.1158/0008-5472.CAN-07-5878
- Lia, Q., Rycaja, K., Chena, X. and Tang, D. G. (2015). Cancer stem cells and cell size: a causal link? *Sem. Cancer Biol.* **35**, 191–199. doi:10.1016/j.semcancer.2015.07.002
- Liao, T.-T. and Yang, M.-H. (2020). Hybrid epithelial/mesenchymal state in cancer metastasis: clinical significance and regulatory mechanisms. *Cells* **9**, 623. doi:10.3390/cells9030623
- Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740. doi:10.1093/bioinformatics/btr260
- Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., Martin-Trevino, R., Shang, L., McDermott, S. P., Landis, M. D. et al. (2014). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Rep.* **2**, 78–91. doi:10.1016/j.stemcr.2013.11.009
- Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O. et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**, 338–354. doi:10.1016/j.cell.2018.03.034
- Mani, S. A., Guo, W., Liao, M.-J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shiptsin, M. et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715. doi:10.1016/j.cell.2008.03.027
- Mukherjee, A., Barai, A., Singh, R., Yan, W. and Sen, S. (2020). Nuclear plasticity increases susceptibility to damage during confined migration. *PLoS Comp. Biol.* **16**, e1008300. doi:10.1371/journal.pcbi.1008300
- Navas, T., Kinders, R. J., Lawrence, S. M., Ferry-Galow, K. V., Borgel, S., Hollingshead, M. G., Srivastava, A. K., Alcoser, S. Y., Makhlof, H. R., Chuaqui, R. et al. (2020). Clinical evolution of epithelial-mesenchymal transition in human carcinomas. *Cancer Res.* **80**, 304–318. doi:10.1158/0008-5472.CAN-18-3539
- Pastushenko, I. and Blanpain, C. (2019). EMT Transition states during tumor progression and metastasis. *Trends Cell Biol.* **29**, 212–226. doi:10.1016/j.tcb.2018.12.001
- Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S. et al. (2018). Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463–468. doi:10.1038/s41586-018-0040-3
- Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. A., Gillespie, S. M., Wakimoto, H., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L. et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401. doi:10.1126/science.1254257
- Puram, S. V., Tirosh, I., Park, A. S., Patel, A. P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C. L., Mroz, E. A., Emerick, K. S. et al. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624. doi:10.1016/j.cell.2017.10.044
- Ray, A., Slama, Z. M., Morford, R. K., Madde, S. A. and Provenzano, P. P. (2017). Enhanced directional migration of cancer stem cells in 3D aligned collagen matrices. *Biophys. J.* **112**, 1023–1036. doi:10.1016/j.bpj.2017.01.007
- Sheridan, C., Kishimoto, H., Fuchs, R. K., Mehrotra, S., Bhat-Nakshatri, P., Turner, C. H., Goulet, R., Jr, Badve, S. and Nakshatri, H. (2006). CD44<sup>+</sup>/CD24<sup>-</sup> breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. *Breast Cancer Res.* **8**, R59. doi:10.1186/bcr1610
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550. doi:10.1073/pnas.0506580102
- Swaminathan, V., Myhre, K., O'Brien, T. E., Berchuck, A., Blobe, G. C. and Superfine, R. (2011). Mechanical stiffness grades metastatic potential in patient tumor cells and in cancer cell lines. *Cancer Res.* **71**, 5075–5080. doi:10.1158/0008-5472.CAN-11-0247
- Swat, M. H., Thomas, G. L., Belmonte, J. M., Shirinifard, A., Hmeljak, D. and Glazier, J. A. (2012). Multi-scale modeling of tissues using CompuCell3D. *Methods Cell Biol.* **110**, 325. doi:10.1016/B978-0-12-388403-9.00013-8
- Tabassum, D. P. and Polyak, K. (2015). Tumorigenesis: it takes a village. *Nat. Rev. Cancer* **15**, 473–483. doi:10.1038/nrc3971
- Wolf, K., Lindert, M. T., Krause, M., Alexander, S., Riet, J. T., Willis, A. L., Hoffman, R. M., Fidler, C. G., Weiss, S. J. and Friedl, P. (2013). Physical limits of cell migration: control by ECM space and nuclear deformation and tuning by proteolysis and traction force. *J. Cell Biol.* **201**, 1069–1084. doi:10.1083/jcb.201210152



- Xu, W., Mezenцев, R., Kim, B., Wang, L., McDonald, J. and Sulchek, T.** (2012). Cell stiffness is a biomarker of the metastatic potential of ovarian cancer cells. *PLoS ONE* **7**, e46609. doi:10.1371/journal.pone.0046609
- Yang, C., Cao, M., Liu, Y., He, Y., Du, Y., Zhang, G. and Gao, F.** (2019). Inducible formation of leader cells driven by CD44 switching gives rise to collective invasion and metastases in luminal breast carcinomas. *Oncogene* **38**, 7113-7132. doi:10.1038/s41388-019-0899-y
- Yu, M., Bardia, A., Wittner, B. S., Stott, S. L., Smas, M. E., Ting, D. T., Isakoff, S. J., Ciciliano, J. C., Wells, M. N., Shah, A. M. et al.** (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**, 580-584. doi:10.1126/science.1228522
- Zhou, H., Neelakantan, D. and Ford, H. L.** (2017). Clonal cooperativity in heterogenous cancers. *Sem. Cancer Biol.* **64**, 79-89. doi:10.1016/j.semdb.2016.08.028