

Collective and Individual Migration following the Epithelial-Mesenchymal Transition

Ian Y. Wong, Sarah Javaid, Elisabeth A. Wong, Sinem Perk, Daniel A. Haber, Mehmet Toner and Daniel Irimia

	<u>Page #</u>
Note SN1: Approximating Cell Migration Dynamics using Lifetime Averaged Metrics	2
Note SN2: Principal Component Analysis of Collective and Individual Migration	4
Note SN3: Principal Component Analysis of Rsk Inhibitors	8
Note SN4: Derivation of Binary Alloy Solidification Model	9
Note SN5: Methods and Materials	12
Table S1: Comparison of Migration Phenotype and Biomarker Expression	28
Table S2: Loading, Covariance Matrix and Confidence Ellipses for Figure S12	30
Table S3: Loading, Covariance Matrix and Confidence Ellipses for Figure S26	48
Figure S1: Collective and Individual Migration Trajectories	18
Figure S2: Representative Migration Trajectories with Nearest Neighbor Dynamics	19
Figure S3: Distribution of Collectively or Individually Migrating Subpopulations	20
Figure S4: Distribution of Timescales for Transitions from Collective to Individual Migration	21
Figure S5: Induction of EMT Morphology and Biomarker Expression through Snail	22
Figure S6: MCF-10A: Scatterplot Matrix of Migration Metrics	23
Figure S7: MDA-MB-231: Scatterplot Matrix of Migration Metrics	24
Figure S8: MCF-10A Snail: Scatterplot Matrix of Migration Metrics	25
Figure S9: Comparison of Migration Metrics: MCF-10A, MCF-10A Snail, MDA-MB-231	26
Figure S10: Comparison of Migration Velocities with Final Position relative to Front	27
Figure S11: Comparison of Phenotypic Profiling with Immunostaining	28
Figure S12: Principal Component Analysis: MCF-10A, MCF-10A Snail, MDA-MB-231	29
Figure S13: Immunostaining: MCF-10A, MCF-10A Snail with Different Pillar Spacings	31
Figure S14: Comparison of Migration Metrics: MCF-10A Snail in Different Pillar Spacings	32
Figure S15: Immunostaining of Wound Healing Assay	33
Figure S16: Comparison of Migration: 2D Wound Healing Assays with Different ECM	36
Figure S17: Comparison of Migration: Isolated Cells in 2D with Different ECM	37
Figure S18: Immunostaining: MCF-10A, MCF-10A Snail with Inhibited Proliferation	38
Figure S19: Comparison of Migration Metrics: MCF-10A Snail with Inhibited Proliferation	39
Figure S20: Proliferation Dynamics with Hoechst Staining	40
Figure S21: Immunostaining at Different Plating Densities	41
Figure S22: DMSO Model Fit	43
Figure S23: BID1870 Model Fit	44
Figure S24: U0126 Model Fit	45
Figure S25: FMKMEA Model Fit	46
Figure S26: Principal Component Analysis: Rsk Inhibitors	47
Video S1: Collective Migration of a Purely Epithelial Population (MCF-10A)	
Video S2: Collective and Individual Migration of an Induced EMT Population (MCF-10A Snail)	
Video S3: Migration of an Induced EMT Population (MCF-10A Snail) with U0126	
Video S4: Migration of an Induced EMT Population (MCF-10A Snail) with FMKMEA	

Note SN1. Approximating Cell Migration Dynamics using Lifetime Averaged Metrics

Quantitative metrics were computed for each cell trajectory using MATLAB code. Only cell trajectories longer than 12 time frames (4 hrs) were considered. The coordinates and time frames associated with the start and end of each trajectory were stored for subsequent analysis.

Nearest Neighbors: The number of nearest neighbors was computed for each segmented object in each time frame by counting the other segmented objects within a 28 μm search radius (one lattice spacing). This number was then averaged over the lifetime of each cell trajectory. As expected from a square lattice, cells could have at most 8 nearest neighbors.

Velocity: The velocity v_i ($i = x, y$) was calculated from the position x_i at time τ and $(\tau-4)$, respectively as: $v_i = [x_i(\tau) - x_i(\tau-4)]/4$. Since images were taken every 15 min, this corresponded to the average velocity every hour, which was chosen to reduce motions due to nuclei shape changes rather than cell migration. This velocity every hour was then averaged over the lifetime of the cell trajectory. Note that cells were restricted to move on a square lattice, so that lifetime averaged velocities were computed as $v_r = |v_x| + |v_y|$, rather than $v_r = \sqrt{v_x^2 + v_y^2}$.

Straightness Index: The straightness index (directional persistence) was computed as the ratio of the distance between initial and final positions for each cell, divided by the integrated distance traveled. $SI = \frac{x_i(t_{end}) - x_i(t_{start})}{\sum_{t_{start}+1}^{t_{end}} (x_i(\tau) - x_i(\tau-1))}$.

Sinuosity: Following the definition proposed by Benhamou (J. Theor Biol 229, 209, 2004), a supplemental measure of tortuosity called sinuosity was used, based on discretized movements. First, the distribution of turn angles ϕ between all time frames was used to compute a standard deviation σ . Second, the square root of the average distance p traveled at each time step was determined. The sinuosity takes the ratio of these two values $S = \sigma / \sqrt{p}$

To confirm that these lifetime-averaged metrics were representative of the overall cell migration dynamics, approximately 200 trajectories of induced EMT cells (MCF-10A Snail) were analyzed within a region of interest. The individually migrating cells start at various locations relative to the collectively migrating front (Figure S1AB). Those cells starting behind the collectively migrating front may initially migrate collectively with many neighbors but then break away, migrating individually for the rest of the trajectory. In contrast, cells that start ahead of the collectively migrating front may primarily migrate individually. It can also be observed that individually migrating cells start at various times throughout the experiment (Figure S1C).

To assess these behaviors more quantitatively, the migration dynamics of representative individual migration trajectories were replotted with the number of nearest neighbors at each time point (Figure S2). The ten trajectories in Figure S2 were classified as individual based on lifetime averaged nearest neighbors, final position in Y, average velocity and straightness (Note SN2). These individually migrating cells started at various times throughout the experiment and some display transient periods of collective migration, as characterized by large numbers of neighbors for extended periods (\sim hours), before finally breaking away and migrating individually ($N \sim 0$). Cells may undergo the opposite transition as well, from individual migration back to collective migration (Figure S2B,F. Light green trajectory, Figure S2D,G: purple trajectory). It should be noted that these trajectories are characterized as mostly individual migration (with zero neighbors) as well as transient periods of collective migration (with several neighbors), which would likely lifetime average to $N > 1$. To account for these dynamic fluctuations, individually migrating cells were classified by a cutoff of lifetime averaged nearest neighbors $N \sim 2.5$, as well as the final position, average velocity, etc.

To examine the relative frequencies of these transient dynamics within the population, each track was plotted in terms of the initial and final neighbors, as well as the lifetime averaged neighbors (Figure S3). These were manually verified to be representative of the cell dynamics. This analysis reveals four distinct subpopulations:

- 1) 28% of cells *always migrate collectively* with many neighbors (average $N > 2.5$).
 - 2) 53% of cells *mostly migrate collectively with some individual migration* (average $N > 2.5$).
- Approximately 30 cells had a lifetime average $N < 2.5$, but were scored as collective based on their final position behind the collective front, slow speed and tortuous trajectories.

- 3) 11% of cells *mostly migrate individually with some collective migration* (average $N < 2.5$).
- 4) 8% of cells *always migrate individually* with zero neighbors (average $N < 1$).

It should be noted that this representation captures the magnitude of these transient motions but omits the transition timescales from collective to individual migration (or vice versa) (Figure S4). In this analysis, the subpopulations that always migrate individually or collectively are represented with a transition time of 0. The median transition time for collective to individual migration is ~ 4 h or longer, while the median transition time for individual to collective migration is ~ 2 h. These sustained transition times are likely to occur from collective migration, not simply cells moving in close proximity (i.e. streaming migration).

In summary, this analysis validates the use of lifetime averaged metrics. This approximation accounts for subpopulations that always migrate collectively or individually as well as subpopulations that display transient dynamics of individual or collective migration, respectively. Moreover, the scattering of individual cells from a collectively migrating front is shown to occur as a consequence of phenotypic plasticity in migration, since cells can transiently exhibit different dynamics. This rules out a pure sorting behavior where each subpopulation always migrates collectively or individually.

Note SN2: Principal Component Analysis of Collective and Individual Migration

Single cell analyses of six migration metrics: Final Y, velocity, sinuosity, straightness, neighbors, and starting Y, revealed that the EMT population (MCF-10A Snail) displayed heterogeneous collective or individual migration behaviors. In contrast, the purely epithelial population (MCF-10A) exhibited collective migration and the purely mesenchymal population (MDA-MB-231) displayed individual migration. Normalized migration metrics were classified into distinct phenotypes using a Gaussian mixture model by calling the MATLAB function *gmdistribution.fit*.

For an unbiased evaluation of the migration behaviors that define the heterogeneity of the Snail population, principal component analysis was used to reduce the dimensionality of the data into a new set of variables that better display the underlying variance (reviewed in Eriksson, L. *et*

al. *Multi- and Megavariate Data Analysis Part I: Basic Principles and Applications*. (Umetrics, 2006) and Jackson, J. E. *J Quality Technol* **12**, 201-213 (1980)).

Using custom MATLAB code, the various arrays for each migration metric was concatenated into an $N \times 6$ data array \mathbf{X} , where $N \sim 10,000$ represents the number of cells tracked over at least 6 separate experiments. Custom MATLAB code was used to normalize each lifetime averaged migration metric to a mean of 0 and a standard deviation of 1. The 6×6 covariance matrix \mathbf{C} was calculated using the MATLAB function *cov*. Note that elements of the covariance matrix C_{ij} represent the covariance between the respective metrics i, j . An examination of \mathbf{C} (Table S2A) reveals that correlations exist between various migration metrics, but the overall distribution of the data is not simply captured by any one metric or pair of metrics.

These migration metrics were then uncorrelated via projection onto a principal component space using the following procedure. The MATLAB function *svd* was used to apply singular value decomposition to \mathbf{C} , yielding a 6×6 matrix \mathbf{V} corresponding to the eigenvectors of \mathbf{C} (Table S2B) as well as 6 eigenvalues ordered from largest to smallest (Table S2C). Normalization of each eigenvalue by the sum of all the eigenvalues reveals the percentage of variance captured by each principal component (Table S2C). Principal component \mathbf{p}_1 is the direction of largest variance, accounting for 45% of the overall variance and the metric groupings described in the text. The variance decreases successively with the subsequent principal components, with minimal off-diagonal elements. Overall, 4 eigenvectors, corresponding to the largest 4 eigenvalues, are sufficient to capture $\sim 93\%$ of the variance of the data. Thus, the 6 migration metrics are well represented by projection onto a reduced dimensional space based on only $a = 4$ principal components. The truncated PCA loading \mathbf{P} is the 6×4 matrix kept from \mathbf{V} (Table S2B). The corresponding scores were given by $\mathbf{T} = \mathbf{XP}$, corresponding to the projection of the normalized data \mathbf{X} into principal component space \mathbf{T} . The primary result of this loading is to group final Y position, starting Y position, velocity and straightness index for individual migration as well as a second grouping of lifetime-averaged nearest neighbors and sinuosity for collective migration. (Figure S12A). This grouping arises primarily from the first principal component \mathbf{p}_1 , which captures the most variance in the underlying data and where the correlated differences are most apparent. We chose to plot this against the third principal component \mathbf{p}_3 , which displays the metrics in a more physically

intuitive way, particularly grouping velocity, straightness index and final Y. Nevertheless, qualitatively consistent trends occur when plotting the second principal component \mathcal{P}_2 instead.

To further visualize this distribution of data, 95% confidence ellipses were calculated in principal component space (Table S2D). The center of the confidence ellipse was located at the respective centroids of each clustered subpopulation, while axes were defined by the elements of the matrix R_k : $R_k = \sqrt{\text{Cov}(T_k) c_{95,k} \frac{(N_k-1)^2}{N_k}}$, where k denotes the clustered collective or individual subpopulation, T_k is the projection of the normalized migration metrics of subpopulation k into principal component space, N_k are the number of cells comprising subpopulation k and $c_{95,k} = ((a/(N_k - a - 1))^*z)/(1 + (a/(N_k - a - 1))^*z)$, where $z = 2.39$ is the critical F value for $\alpha = 0.05$ (95% confidence) with degrees of freedom a and $(N - a - 1)$. Given this definition, the percentage of each subpopulation inside or outside the confidence ellipse could be calculated. It should be noted that the entire dataset has been mean-centered to zero, so the centroid of the migration metrics for the overall population is located at the origin. However, the centroids for the clustered collective or individual subpopulations are primarily offset from the origin along \mathcal{P}_1 . This occurs since \mathcal{P}_1 is the principal component that captures the most variance from the underlying dataset, so that the differences between the migration metrics of the two clustered subpopulations are most prominent along \mathcal{P}_1 . For this particular dataset, the differences between the collective and individual centroids along \mathcal{P}_3 are relatively small. In comparison, the collective and individual centroids for the Rsk inhibitor treated populations (Note SN3, Figure S26) are not only offset from each other along \mathcal{P}_1 due to the differences between collective and individual migration behavior, but additional offsets along \mathcal{P}_1 and \mathcal{P}_3 due to differences between control and Rsk inhibited migration.

Based on this analysis, the t_1/t_3 score plot reveals two clusters for cells that display either collective migration (green) or individual migration (red) (Figure S12B). The clusters determined by the Gaussian mixture model were overlaid with 95% confidence ellipses (F -distribution), indicating that a cell within that ellipse has 95% confidence of belonging to that particular cluster. These confidence ellipses correspond to a wider distribution of individual migration behaviors (\mathcal{P}_1 axis $a_{\text{ind}} = 4.12$) and a more compact distribution of collective migration behaviors (\mathcal{P}_1 axis $a_{\text{col}} = 2.48$).

The behaviors associated with purely epithelial (MCF-10A) were then projected onto the PCA model to quantify similarities or differences in migration behavior. Figure S12C shows the projection of the purely epithelial population (MCF-10A) onto the t_1/t_3 score space of the Snail model (cyan). The blue circle and green box correspond to the centroids of the purely epithelial population (MCF-10A) and the collectively invading subpopulation after EMT (MCF-10A Snail), respectively. The centroid of the purely epithelial population (MCF-10A) is closer to the centroid of collective migration in EMT (MCF-10A Snail) (distance ~ 1.16) compared to the centroid of individual migration in EMT (MCF-10A Snail) (distance ~ 1.85). Furthermore, 75% of the purely epithelial population (MCF-10A) were located within the 95% confidence ellipse for the collectively invading subpopulation after EMT (MCF-10A Snail). Altogether, this suggests that the statistical distributions of normalized collective migration behaviors are similar between purely epithelial population (MCF-10A) and the collectively migrating subpopulation after EMT (MCF-10A Snail), with some discrepancies due to differences in final Y.

Finally, the behaviors associated with purely mesenchymal (MDA-MB-231) were projected onto the PCA model. Figure S12D shows the projection of the purely mesenchymal population (magenta). The red diamond and magenta triangle correspond to the centroids of the individually invading subpopulation after EMT (MCF-10A Snail) and purely mesenchymal population (MDA-MB-231), respectively. The centroid of the purely mesenchymal population (MDA-MB-231) is closer to the centroid of individual migration in EMT (MCF-10A Snail) (distance ~ 0.64) compared to the centroid of collective migration in EMT (MCF-10A Snail) (distance ~ 2.37). Moreover, 99% of the purely mesenchymal cells (MDA-MB-231) were located within the 95% confidence ellipse for the individually invading subpopulation after EMT (MCF-10A Snail). Altogether, this suggests that the statistical distributions of normalized individual migration behaviors are similar for a purely mesenchymal population (MDA-MB-231) and the individually migrating subpopulation after EMT (MCF-10A Snail). Nevertheless, it should be noted that this analysis is based on lifetime-averaged migration metrics. This approximation is representative of single cell migration behaviors (Note SN1), but does not include the dynamic transitions between collective and individual migration in the induced EMT population (MCF-10A Snail). A more granular analysis incorporating details of single cell migration dynamics would reveal further differences in the purely mesenchymal population (MDA-MB-231), which always migrates individually, and the individually migrating

subpopulation after EMT (MCF-10A Snail), which includes cells that always migrate individually as well as cells that always migrate individually with some collective migration.

Note SN3: Principal Component Analysis of Rsk Inhibitor Panel

The migration behaviors of an induced EMT population (MCF-10A Snail) were analyzed in response to a panel of small molecule Rsk-inhibitors (BID-1870, U0126 and FMKMEA), relative to control (DMSO). The effectiveness was further quantified using principal component analysis, as described in Note S2. First, the control population was grouped into distinct subpopulations based on collective or individual migration behaviors. Figure S26A shows the previous loading plot with representative principal components (\mathbf{P}_1 / \mathbf{P}_3). This projection groups larger values of final Y, velocity, straightness and starting Y with individual migration, whereas larger values of sinuosity and nearest neighbors are associated with collective migration.

The data for the drug-treated conditions were then projected onto the Snail PCA model for direct comparison (Note SN2) and clustered into collectively or individually invading subpopulations using a Gaussian mixture model in MATLAB. The covariance of the truncated loading again shows that the principal components are sufficient to capture most of the variation even when migration is perturbed using Rsk inhibitors (Table S3A-F). Figure S26B shows the centroid for individually invading subpopulations as well as the corresponding 95% confidence ellipse (Table S3G), color-coded in red (DMSO), magenta (BID-1870), cyan (U0126) and blue (FMKMEA). The centroids for BID-1870 and U0126 are slightly displaced in t_1 from DMSO by 0.34 and 0.37, respectively. The largest displacement of 1.06 in t_1 from DMSO occurs for FMKMEA, corresponding to the strongest suppression of average velocity and distance invaded. The confidence ellipse for DMSO (red) has a \mathbf{P}_1 axis of 4.12, which is consistently narrower when Rsk inhibitors are applied. In particular, the confidence ellipses associated with U0126 and FMKMEA both have \mathbf{P}_1 axes of approximately 3.8, but BID1870 has an even smaller \mathbf{P}_1 axis of 3.29. This indicates that BID1870 results in the most compact statistical distribution, associated with decreased variation in migration behaviors.

Figure S26C shows the centroid for individually invading subpopulations as well as the corresponding 95% confidence ellipse (Table S3G), color-coded in green (DMSO), magenta (BID-1870), cyan (U0126) and blue (FMKMEA). The centroids for BID-1870 and U0126 are slightly displaced in t_1 from DMSO by -0.09 and 0.27, respectively. The largest displacement of

0.76 in t_1 from DMSO again occurs for FMKMEA, corresponding to the strongest suppression of average velocity and distance invaded. The confidence ellipse for DMSO (red) has a \mathbf{P}_1 axis of 2.49, compared to 2.41 for BID1870, 2.36 for U0126 and 2.18 for FMKMEA. For collective migration, FKMEA results in the most compact statistical distribution, associated with decreased variation in migration behaviors.

Finally, Figure S26D shows the centroid for the complete populations as well as the corresponding 95% confidence ellipse (Table S3G), color-coded in red/green (DMSO), magenta (BID-1870), cyan (U0126) and blue (FMKMEA). The centroids for BID-1870 and U0126 are slightly displaced in t_1 from DMSO by -0.30 and 0.14, respectively. The largest displacement of 0.85 in t_1 from DMSO again occurs for FMKMEA, corresponding to the strongest suppression of overall average velocity and distance invaded. The confidence ellipse for DMSO (red) has a \mathbf{P}_1 axis of 5.06, compared to 4.8 for BID1870, 4.98 for U0126 and 4.53 for FMKMEA. Overall, FKMEA results in the most compact statistical distribution, associated with decreased variation in migration behaviors.

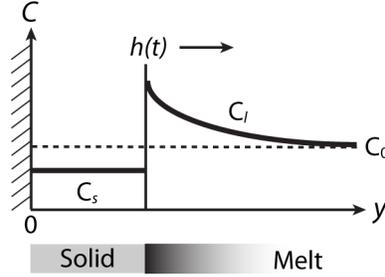
In summary, FMKMEA leads to the largest displacement in t_1 from DMSO for collective migration, individual migration and total migration, corresponding to strongest suppression of distance invaded and average velocity. FMKMEA also tends to reduce the \mathbf{P}_1 axis of the 95% confidence ellipse, indicating a more compact statistical distribution and decreased variability. These results based on principal component analysis are qualitatively consistent with the analyses based on solidification model as well as net migration flux. An interesting difference is that BID1870 exhibits a large displacement in t_3 relative to DMSO. This may occur due to an increase in path straightness, which would shift the centroid in the direction of $-\mathbf{P}_3$ and $-\mathbf{P}_1$.

Note SN4: Derivation of Binary Alloy Solidification Model

The heterogeneous migration dynamics associated with the epithelial-mesenchymal transition have analogies with the plane-front solidification of binary mixtures. Worster presented an analytical solution for this scenario (J. Fluid Mech 167, 481, 1986) as an extension of the Neumann solution of the classical Stefan problem for pure melts.

A semi-infinite region ($y > 0$) initially contains a binary melt with uniform composition of solute $C = C_0$ and a temperature T_∞ . At time $t = 0$, the temperature of the boundary is abruptly

cooled to (and maintained at) $T(y = 0) = T_B$. As a consequence, there is a moving solidification front $h(t)$ between the solid and melt phases.



Finally, the other boundary conditions at infinity are given by:

$$C(y \rightarrow \infty) = C_0, T(y \rightarrow \infty) = T_\infty \quad (1)$$

Since the transport of solute occurs only by molecular diffusion, the one-dimensional diffusion equation is given by:

$$\frac{\partial c_i}{\partial t} = D_i \frac{\partial^2 c_i}{\partial x^2} \quad (i = s, l) \quad (2)$$

Note that the $D_s \ll D_l$, so that diffusion through the melt is the dominant transport process compared to back diffusion into the solid. As a consequence, only diffusion in the melt phase will be considered ($i = l$). Similarly, since material is conserved across the interface,

$$(C - C_s)\dot{h} = -D_l \left. \frac{\partial C}{\partial x} \right|_{h^+} + D_s \left. \frac{\partial C}{\partial x} \right|_{h^-} \quad (3)$$

where the second term on the RHS of (3) is negligible.

Since there are no natural spatial or time scales in this problem, a similarity solution will be attempted. Let the new dependent variable be:

$$\eta = \frac{y}{\sqrt{4Dt}} \quad (4)$$

where the interface has position

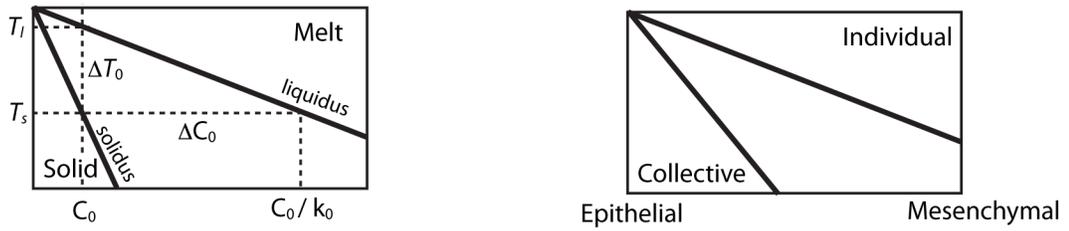
$$h(t) = 2\lambda\sqrt{Dt} \quad (5)$$

The governing equations reduce to ordinary differential equations in the similarity variable and can be integrated to give:

$$C = C_0 + (C_h - C_0) \frac{\text{erfc}(\eta)}{\text{erfc}(\lambda)}, \quad \lambda < \eta \quad (6)$$

where $C = C_s$, $\eta < \lambda$.

Since solute is usually more soluble in the melt phase than in an ordered solid phase, freezing leads to the rejection of some or all of the solute from the solid phase. The extent of this solubility is given by the distribution coefficient (at constant temperature and pressure) $k_0 = C_s / C_l$.



The jump in concentration at the solid/melt interface is thus given by:

$$\Delta C_0 = (C_l - C_s) = \frac{1-k_0}{k_0} C_0 \quad (7)$$

So the solution (6) can be given by:

$$C = C_0 \left(1 + \frac{1-k_0}{k_0} \left(\frac{\text{erfc}(\eta)}{\text{erfc}(\lambda)} \right) \right)$$

This solution provides sufficient scaling information for fitting the experimental data to a theoretical model. The constant λ must be solved numerically from the interfacial conditions for interfacial temperature, which is not relevant for the purposes of this paper.

The spatiotemporal positions of all cells in the epithelial and mesenchymal subpopulations were sorted into bins of 1 hr and 20 μm , respectively. The position of the interface was fit to the maximum of the mesenchymal distribution, yielding D_E . The mesenchymal distribution ahead of the interface (i.e. $y > \sqrt{4D_E t}$) was then fit to a complimentary error function $C_M \text{erfc} \left(\frac{y}{\sqrt{4D_M t}} \right)$. It should be noted that the interface between epithelial and mesenchymal cells is not completely

sharp, so the epithelial distribution was empirically fit to an expression of the form $C_E(1 - \exp[y - \sqrt{4D_E t}])$, which represents a translating “wave” front with some curvature. Note that these distributions implicitly assume a discontinuity, so that the epithelial distribution is 0 ahead of the interface and the mesenchymal distribution is 0 behind the interface.

Note SN5: Materials and Methods

Cell Culture:

MCF-10A: Cells were a gift from G. Smolen and J. Zhang. Cells were cultured in MCF-10A growth media as described (Debnath et al. Methods 30, 256, 2003): DMEM/F12 (Invitrogen), with 5% (v/v) horse serum (Invitrogen), 20 ng/mL epidermal growth factor (EGF, Peprotech), 0.5 mg/mL hydrocortisone (Sigma), 100 ng/mL cholera toxin (Sigma), 10 ug/mL insulin (Sigma) and 1% (v/v) penicillin-streptomycin. Cells were passaged at 1:20 split ratios every 3 days up to passage 20.

MDA-MB-231: Cells were a gift from F. Bersani. Cells were cultured in DMEM (Invitrogen) with 10% (v/v) FBS and 1% (v/v) penicillin-streptomycin. Cells were passaged at 1:10 split ratios every 4 days up to passage 20.

Nuclear staining: For imaging and visualization, cells were labeled by incubation with nuclear-staining 1:500 10mg/mL Hoechst 33258 (Invitrogen) in MCF-10A growth media for 1 hour at 37°C. Cells were then removed from the incubator and washed twice with 1X phosphate-buffered saline (PBS). Cells were detached from the surface using Accumax (Sigma) for 10 min at 37°C and quenched with MCF-10A resuspension media (DMEM/F12 (Invitrogen), with 20% (v/v) horse serum (Invitrogen), and 1% (v/v) penicillin-streptomycin). Cells were centrifuged at 1000 rpm for 5 minutes and then re-suspended in growth media to a working concentration of 3 million cells/mL.

Transfection:

To generate a potent reversible EMT-inducing stimulus in MCF-10A cells, a Snail-1 retroviral expression construct was used, with a fused estrogen receptor (ER) response element and a six amino acid substitution that confers constitutive activity through resistance to the inactivating

GSK3b phosphorylation (Hung et al. Nat Cell Biol 6, 931, 2004). Infection of non-transformed, immortalized mammary epithelial MCF10A cells with ER-Snail 6SA, followed by treatment with tamoxifen (4-OHT), triggered morphological and biomarker characteristics of EMT. For ease of visualization, GFP-overexpressing cell lines were used in some experiments that did not require immunostaining. These cells were slightly slower than the corresponding uncolored cells, but otherwise displayed qualitatively similar migration behaviors.

Device Fabrication:

Silicon Master Fabrication: All devices were fabricated using standard multilayer photolithography. Single-side polished <100> silicon wafers were cleaned in piranha solution (70% sulfuric acid, 30% hydrogen peroxide. *CAUTION: Piranha solution reacts violently with organic material and should be handled with extreme caution*). The first layer of SU-8 (SU-8 10, MicroChem) was spun-on, baked and exposed to define the channel features. The central cell reservoir was defined using a second layer of SU-8 (SU-8 50, MicroChem). Both layers were then developed using negative-resist developer for 2-5 minutes to generate the final master.

Soft Lithography: Elastomeric devices were fabricated using standard soft lithography, as described elsewhere (Irimia and Toner, Integr Biol 1, 506, 2009). Briefly, (poly)dimethylsiloxane (Sylgard 184, Dow Corning) prepolymer was mixed with its crosslinker at 10:1 (w/w) ratio. The mixture was poured onto the SU-8 master, degassed and allowed to cure in a conventional oven at 65 °C for 24 h. Devices were punched out and bonded to glass coverslip bottom 24-well plates (MatTek Corporation) after exposure to a 20-second, rf oxygen plasma (March Corporation).

Devices were immediately functionalized with extracellular matrix protein by pipetting 10 uL (100 ng/mL fibronectin, Sigma) into the device for 15 minutes. Each well was then filled with 1x PBS and stored at 4° C until use.

Prior to use, devices were prefilled with growth media for at least one hour. Approximately 30,000 cells were loaded into each device. Each well was then filled with growth media.

Additional solutions include 10 μ M BID-1870 (Stemgent), 10 μ M U0126 (EMD Biosciences) and 10 μ M FMK-MEA (a gift from J. Taunton, UCSF) and 5 μ g/mL CDH1.

Wound-Healing Assay:

The wells of a 24 well plate were each treated with 1x PBS, 10 μ g/mL fibronectin, or 10 μ g/mL collagen I, then incubated at 37°C for 1 hour. 10^5 cells were seeded in each well and left overnight in growth media. Once cells became confluent, they were Hoechst stained as described previously. Media was removed from the wells and a wound was made down the middle of each well through the confluent cell layer with a P200 tip. Each well was washed twice with 1x PBS. Growth media was added to each well. The plate was sealed with parafilm with slits to allow for gas exchange. The plate was placed in an incubating, temperature controlled microscope stage and imaged every 15 minutes for 24 hours, or immunostained at subsequent time points.

Low Density Cell Migration in 2D:

The wells of a 24 well plate were each treated with 1x PBS, 10 μ g/mL fibronectin, or 10 μ g/mL collagen I, then incubated at 37°C for 1 hour. 10^5 cells were seeded in each well and left overnight in growth media. Cells were then Hoechst stained as described previously. The plate was sealed with parafilm with slits to allow for gas exchange. The plate was placed in an incubating, temperature controlled microscope stage and imaged every 15 minutes for 24 hours.

Time-Lapse Microscopy:

Fluorescently-labeled cells were imaged using an inverted microscope (Nikon TiE) under environmentally controlled conditions (37° C and 5% CO₂, humidified). Each device in the multiwell plate was scanned every 15 minutes using a 10X (NA 0.45) long-working distance objective, while maintaining a consistent focus (Perfect Focus System, Nikon Elements). Fluorescence illumination was provided by a light-guide-coupled mercury lamp illumination source with built-in shutter (Nikon Intensilight). Emission spectra were captured using DAPI (Chroma 31000v2), FITC (Nikon 96320) and TRITC (Nikon 96305) filter sets. All images were acquired with 12-bit resolution using a cooled-CCD camera (Photometrics CoolSnap HQ2). Care was taken to ensure that all images were recorded with identical acquisition parameters (exposure time, camera gain/gamma control and microscope aperture settings).

Automated Cell Tracking.

Image Preprocessing: Cell detection was accomplished using the DAPI fluorescent signals of cellular nuclei. Time-lapse images in the DAPI channel were run through a pipeline developed in Cell Profiler version 10415 (Broad Institute). Images were converted from color to grayscale. Intensity variations were corrected using illumination correction; a smooth illumination function for each image was calculated by median filtering based on the intensity at each pixel of the image. The intensity of all pixels were adjusted by subtracting the illumination function and were subsequently rescaled.

Object Segmentation: Objects were identified from these adjusted images by using parameters of intensity, size, and shape. A background method was used to determine the foreground from the background of the image; the mode of the intensity histogram of each image was assumed to be the background of the image and a threshold value was set by manually setting a threshold correction factor to apply to the intensity mode. Only objects with intensities above the set threshold and sizes between 12 and 40 pixels (7.8 and 26 μm) were identified as nuclei. The shapes of the objects were used to distinguish clumped objects while the watershed method was used to segment the clumped objects into individual nuclei by determining the dimmer dividing lines between objects within the clumps.

Nuclei Tracking: The resultant intensities and spatial positions of the detected nuclei were loaded into u-track (Jaqaman et al. Nature Methods 2008), a MATLAB based multi-particle tracking software, to determine the trajectory of each object. This algorithm seeks to optimize the possible set of trajectories linking different points in time based on the minimization of a cost function. This program can account for cell division and occasional errors in segmentation using a user specified cost function.

Migration Analysis (see Note SN1):

Statistical Analysis: Distributions were checked for statistical significance using two-sample Kolmogorov-Smirnov tests as well as n-way ANOVA in MATLAB.

Plotting: Scatterplots were generated using the *scatplot* function (Alex Sanchez, MathWorks File Exchange)

Immunofluorescent Staining:

After time-lapse microscopy, immunofluorescent staining was performed to verify marker expression. Cells were washed in cold 1X PBS (with 20 mM MgCl₂ and 20 mM CaCl₂) for 2-3 minutes. Cells were then fixed for 1 hour at 4° C in a solution of 1X PBS, 4% formaldehyde, 20 mM MgCl₂ and 20 mM CaCl. The fixing solution was aspirated and wells were filled with 1X PBS and stored overnight at 4° C. Cells were washed with 1X PBS and blocked with 10% goat serum in PBS for 2 hours. Next, cells were washed with a solution of 1X PBS with 5% sodium acetate for 30 min. Each device was then filled with 20 uL of primary antibody solution: 1:500 250ug/mL CDH1 (BD Biosciences), 1:200 vimentin (Cell Signaling) and 1:600 phalloidin (Life Technologies) in 1X PBS with 1% milk and stored overnight at 4° C. Devices were washed three times with 1X PBS with 1% milk. 20 uL of solution of appropriately matched secondary antibodies: 1:500 2mg/mL Alexa Fluor 488 (Invitrogen), 1:500 2 mg/mL Alexa Fluor 555 (Invitrogen) and 1:500 2 mg/mL Alexa Fluor 647 (Invitrogen) in 1X PBS with 1% milk was added to each device for 1 hour at room temperature in the dark. Devices were then rinsed twice with 1X PBS for 10 minutes each. 1:500 10mg/mL Hoechst 33258 (Invitrogen) in PBS was added for half an hour at room temperature. Devices were again rinsed twice with 1X PBS for 10 minutes each.

Western Blot

Cells were harvested in 1x RIPA buffer containing 1X protease inhibitor cocktail (Complete™ EDTA-free, Roche). Cell lysates were cleared by centrifugation at 14,000 rpm for 10 min at 4°C. For immunoblotting analysis, lysates were loaded onto 4-15% SDS-PAGE gels (ReadyGel, Bio-Rad), and subsequently transferred onto Immobilon PVDF membrane (Millipore). Proteins were visualized with Western Lightning Plus chemiluminescence kit (Perkin Elmer). Antibodies used: CDH1 (610181, BD Biosciences), VIM (D21H3, Cell Signaling), FN1 (F3648, Sigma) and β-actin (ab6276, Abcam).

Western Blot on Nuclear/Cytoplasmic Markers:

Nuclear/cytoplasmic fractionation was performed using the NE-PER Nuclear and Cytoplasmic Extraction Reagents kit from Pierce Protein Biology Products (Catalog # 78835) according to manufacturer's protocol. Cells were lysed and subsequently analyzed with antibodies against Snail (R&D: AF3639), LDH (Santa Cruz: sc-133123) and HDAC1 (Santa Cruz: sc-7872).

Proliferation Assays:

MCF-10A cells were plated in 4 96-well tissue culture treated plates (Falcon) with 2.0×10^3 cells 100 μ L of growth media per well. Two cell conditions were plated (MCF-10A cells labeled by incubation with nuclear staining 1:500 10mg/mL Hoechst 33258 (Invitrogen) in MCF-10A growth media for 1 hour at 37°C, MCF-10A cells that had been incubated with tamoxifen (1:1000) in growth media for 72 hours and then labeled by incubation with nuclear staining 1:500 10mg/mL Hoechst 33258 (Invitrogen) in MCF-10A growth media for 1 hour at 37°C, All cells in one plate were fixed at each time point (0 hours, 24 hours, 48 hours, 72 hours) following the same fixing protocol used for immunofluorescent staining. After each plate was fixed it was stored at 4° C. After the 72 hour time point all cells were washed with 1X PBS then incubated in 1:5000 SYTO60 (Life Technologies) in 1X PBS at room temperature for 1 hour. Cells were washed two times with 1X PBS and then incubated in 1X PBS at 4° C. The intensity of each well was determined with a plate scanner exposing at 700 nm wavelength.

Proliferation was also inhibited using mitomycin C (Sigma; M4287). Cells were incubated in 25 μ g/mL mitomycin C in normal MCF10A growth media for 2 hours at 37 degrees C and 5% CO₂. Cells were then washed 2 times with 1X PBS. Migration assays were then performed as described.

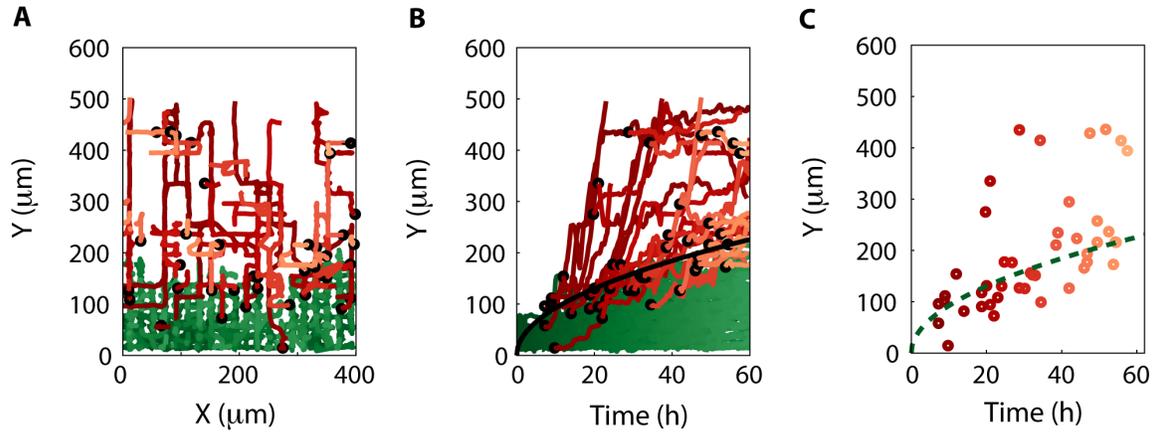


Figure S1: (A) X and Y coordinates of collective (green) and individual (red) migration trajectories. Black dots represent start of each trajectory. (B) Collective (green) and individual (red) migration trajectories in the Y direction as a function of time. Black dots represent the start of each trajectory. Black line corresponds to the collective front, which moves diffusively as $\sim\sqrt{t}$. (C) Reduced representation of B showing the position of the collective front (green dotted line) as well as the starting locations of all individual trajectories (red)

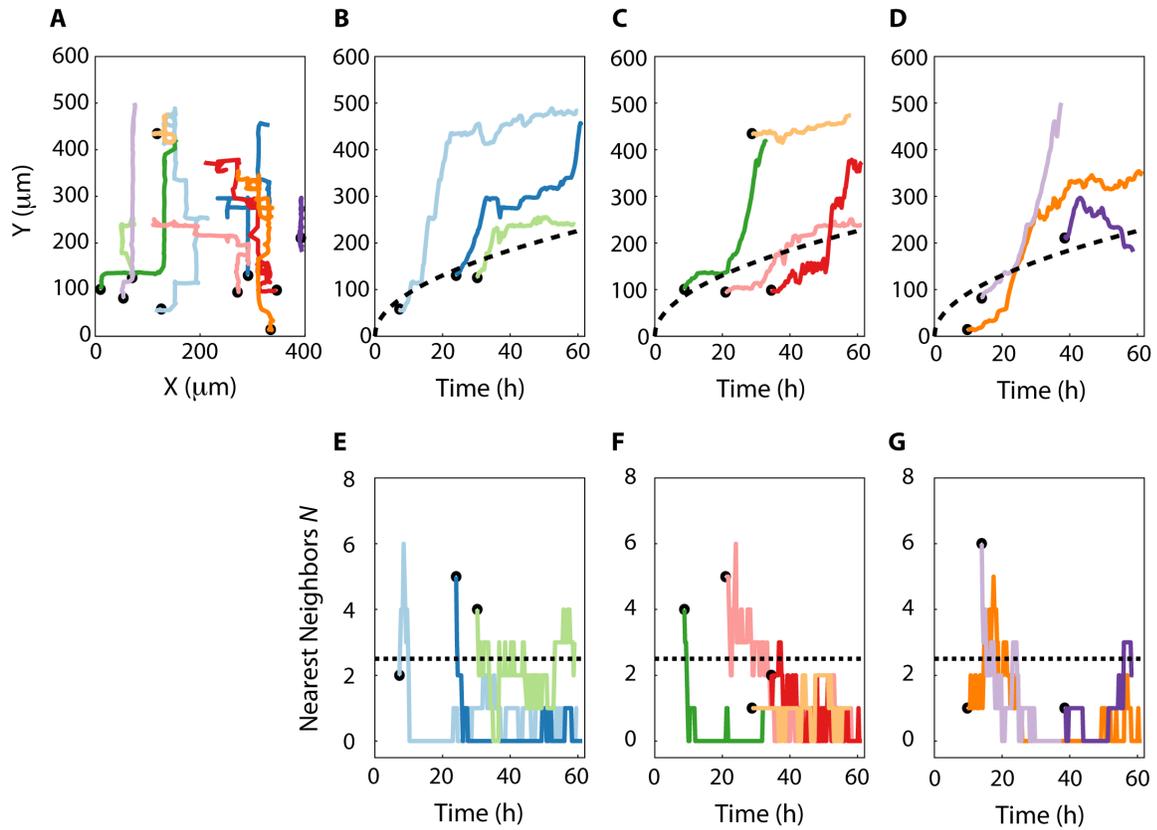


Figure S2: (A) X and Y coordinates of ten representative individual migration trajectories. Black dots represent start of each trajectory. (B-D) Various individual migration trajectories in the Y direction as a function of time. Black dots represent the start of each trajectory. Dashed black line corresponds to the collective front, which moves diffusively as $\sim\sqrt{t}$. (E-G) Corresponding numbers of nearest neighbors as a function of time. Black dots represent the start of each trajectory. Dotted black line corresponds to the cutoff of $N = 2.5$ neighbors used to classify collective or individual migration.

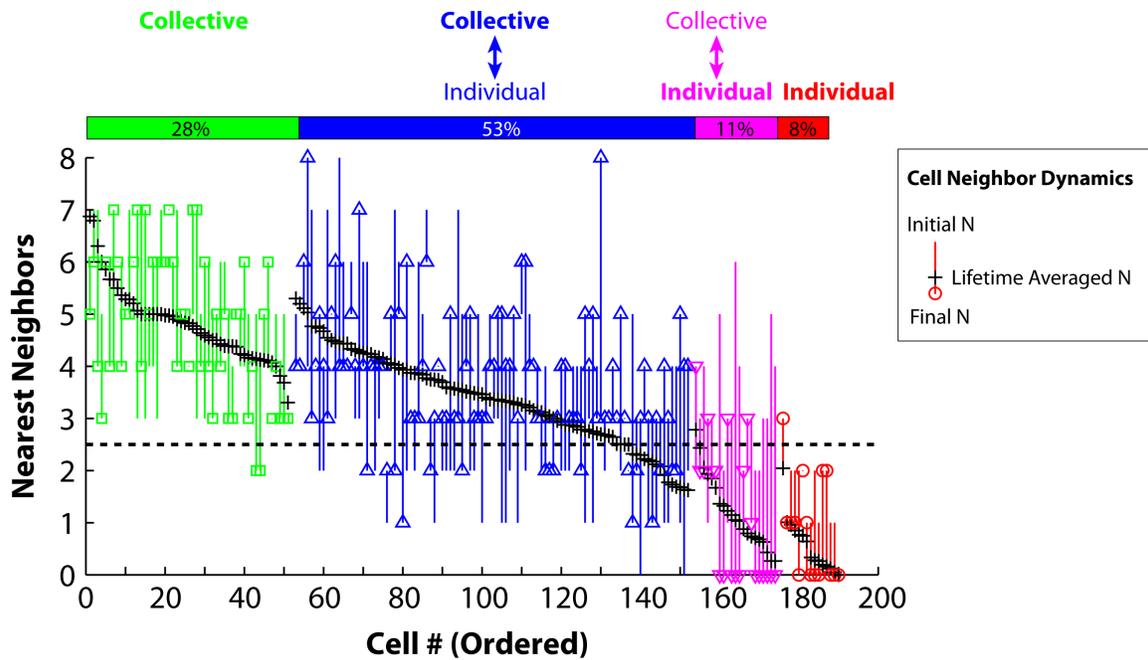


Figure S3: Cells could be further classified into always collective (green), mostly collective with some individual migration (blue), mostly individual with some collective migration (pink) and always individual (red). Each line maps out the initial and final number of neighbors, while the black cross represents the lifetime average.

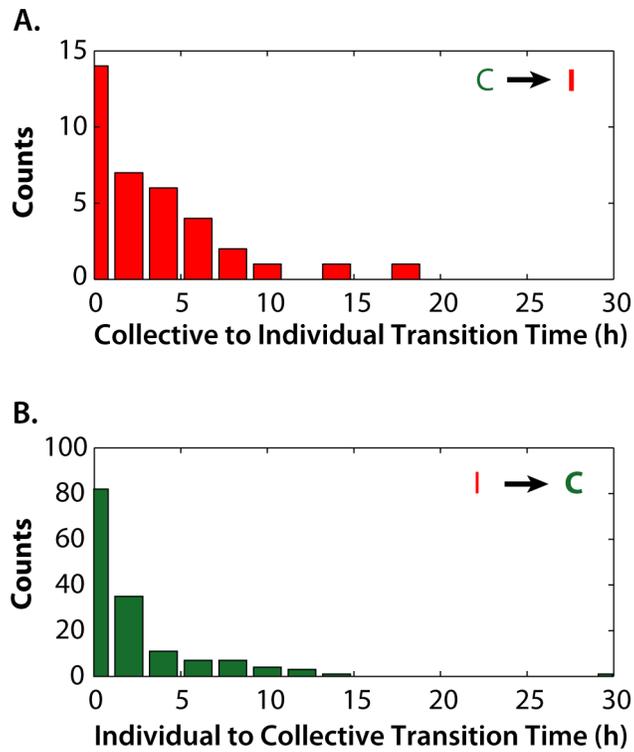


Figure S4 (A) Distribution of timescales for transition from collective migration ($N > 2.5$) to individual migration ($N < 1$). **(B)** Distribution of timescales for transition from individual migration ($N < 1$) to collective migration ($N > 2.5$).

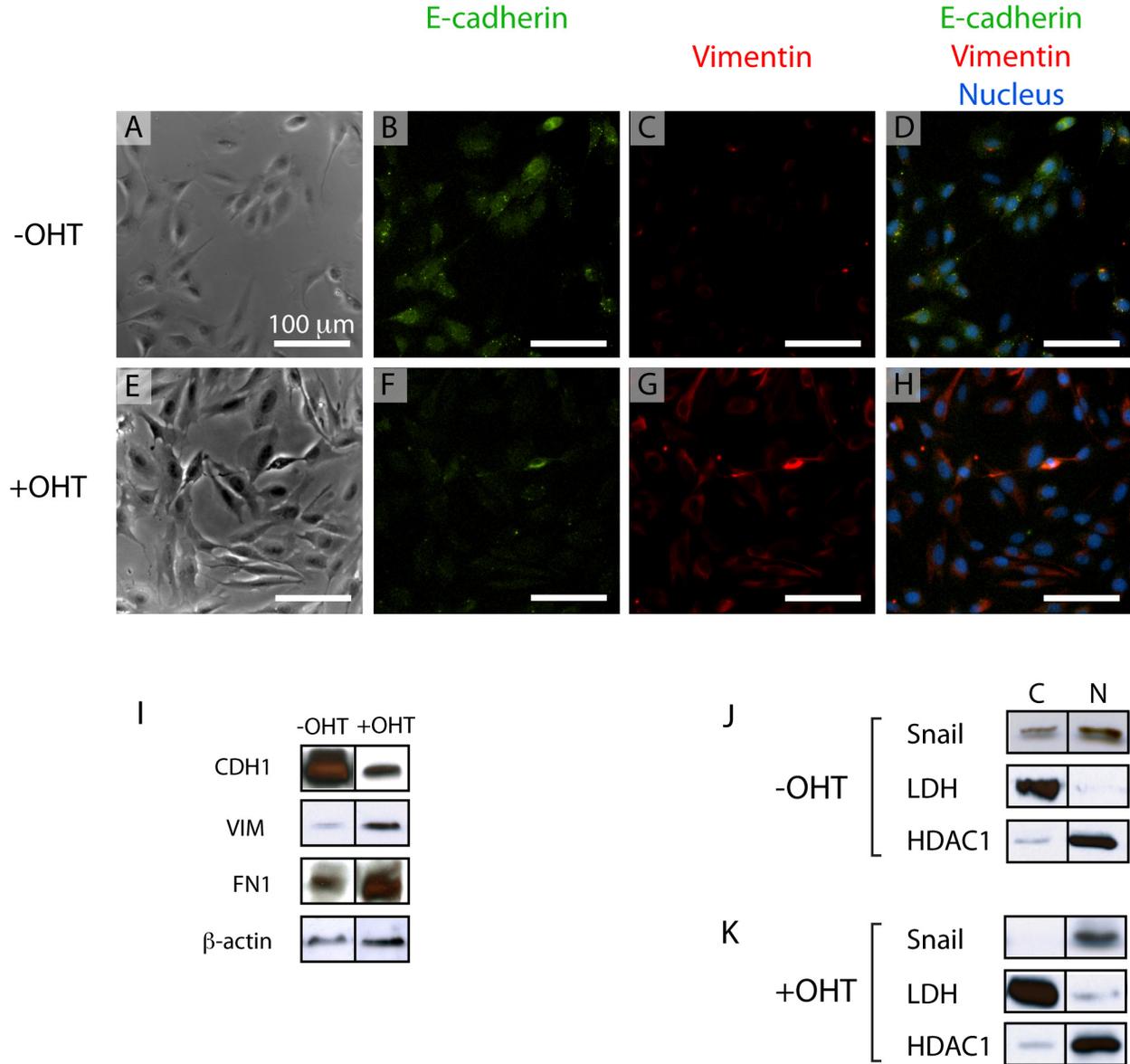


Figure S5: Induction of Snail in MCF-10A causes an epithelial-mesenchymal transition (EMT) with a concurrent loss of epithelial markers and gain of mesenchymal markers. **(A,E)** Phase contrast images of subconfluent MCF-10A after 72 h treatment with control **(A)** and OHT **(E)**. **(B-D, F-H)** Immunofluorescence images showing E-cadherin **(B,F)**, Vimentin **(C,G)**, Merge E-cadherin, vimentin and DAPI **(D,H)**. **(I)** Western blot of epithelial (CDH1) and mesenchymal (FN1 and SERPINE1) markers of un-induced (-OHT) and induced (+OHT) MCF-10A Snail cells. β -actin serves as loading control. MCF-10A were induced for 72 h. **(J, K)** Nuclear/cytoplasmic fractionation was performed on un-induced (-OHT) **(J)** and induced (+OHT) **(K)** MCF-10A cells and western blot analysis was performed. LDH used as cytoplasmic control and HDAC1 used as nuclear control. MCF-10A were induced for 72 h.

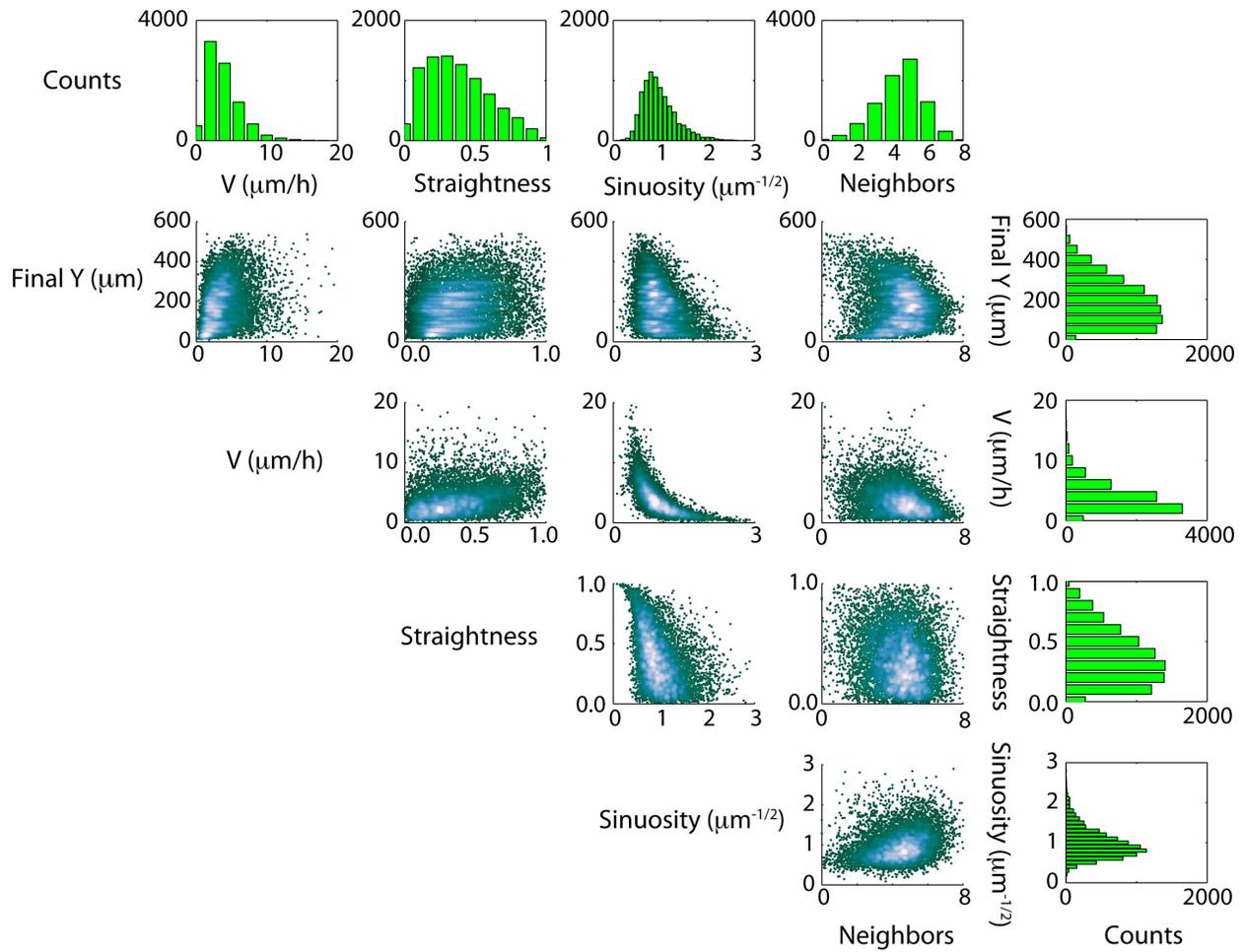


Figure S6. Scatterplot matrix of Final Y position, velocity, straightness, sinuosity and lifetime-averaged nearest neighbors for epithelial cells (MCF-10A).

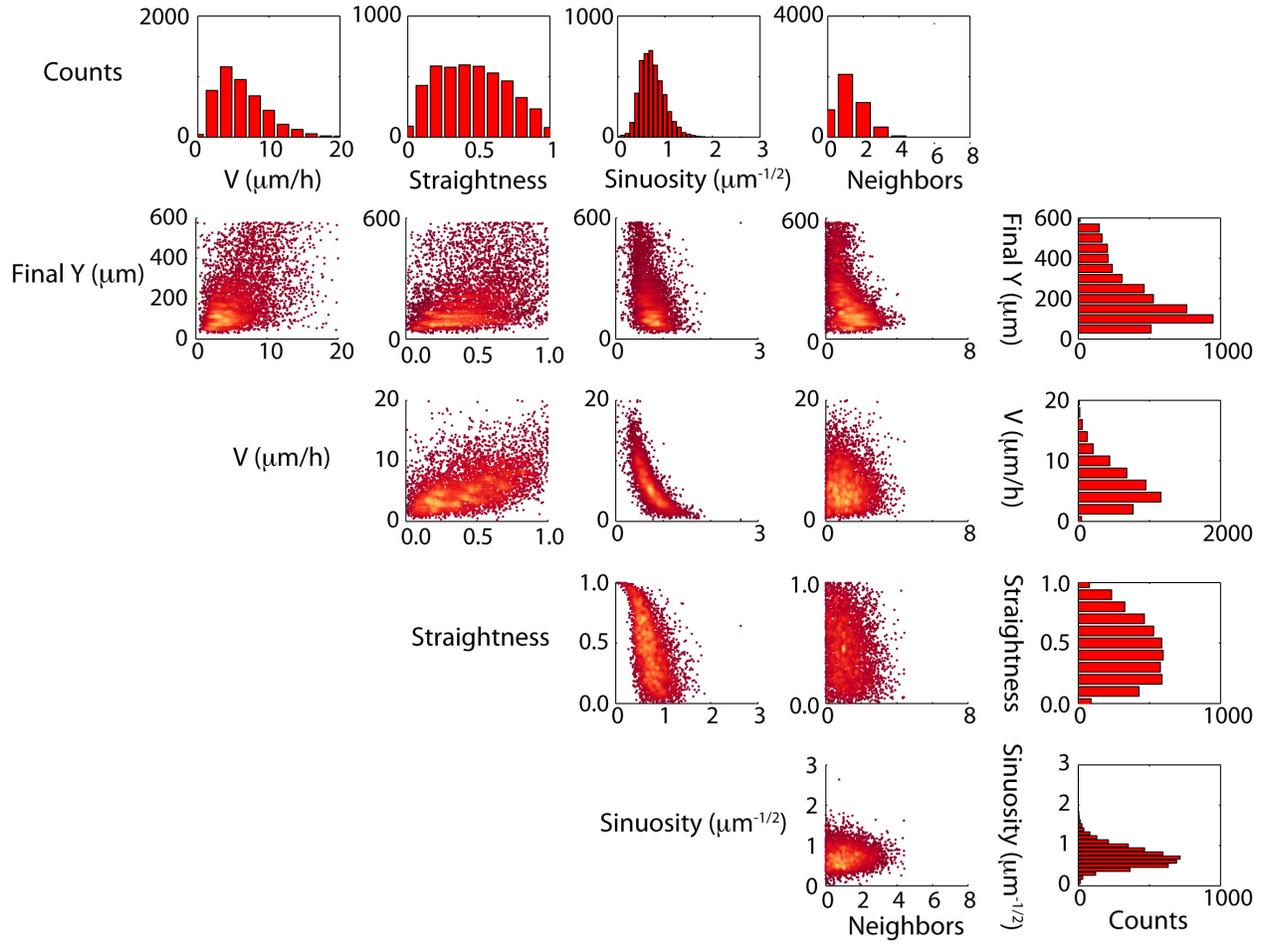


Figure S7. Scatterplot matrix of Final Y position, velocity, straightness, sinuosity and lifetime-averaged nearest neighbors for mesenchymal cells (MDA-MB-231).

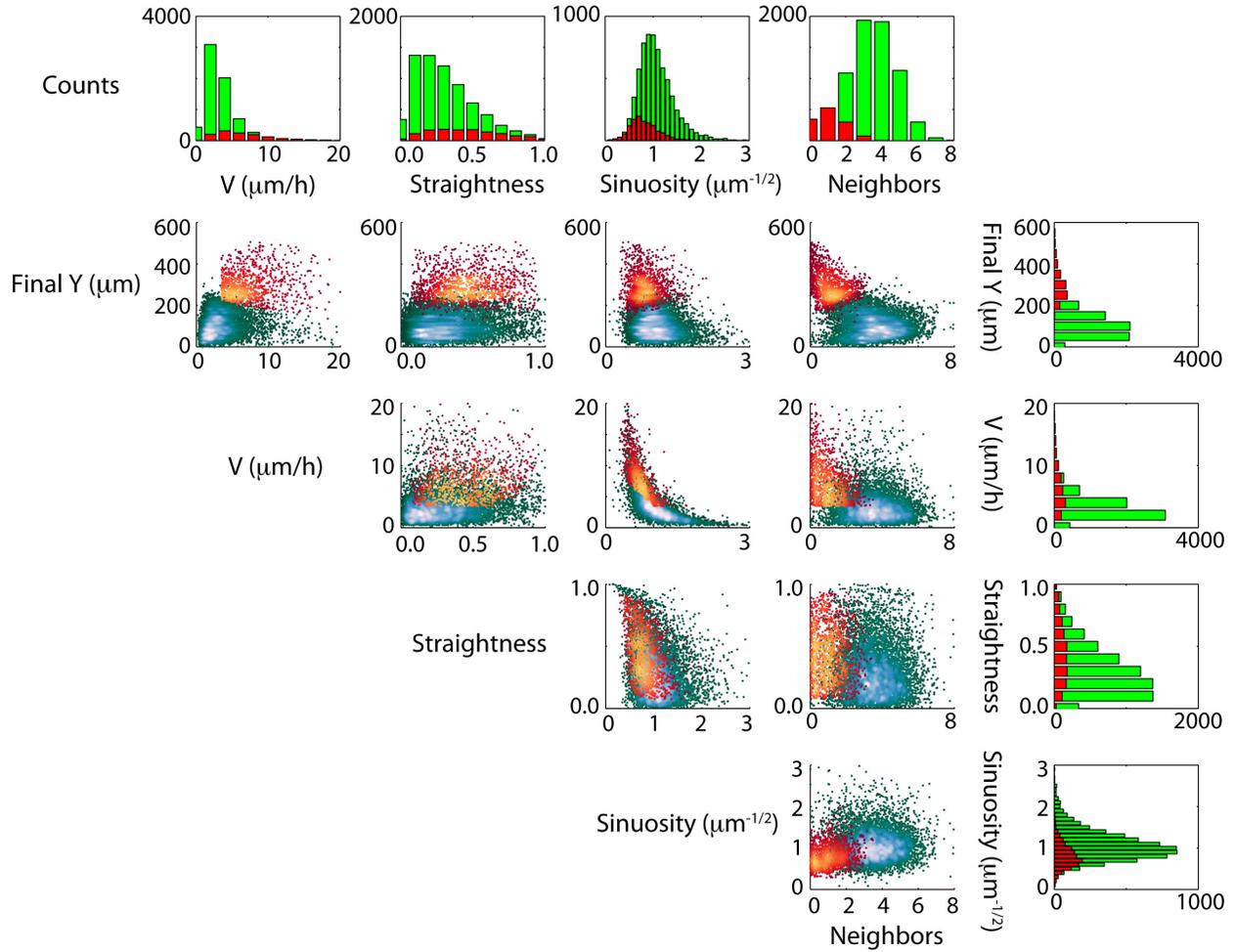


Figure S8. Scatterplot matrix of Final Y position, velocity, straightness, sinuosity and lifetime-averaged nearest neighbors for EMT-induced cells (MCF-10A Snail).

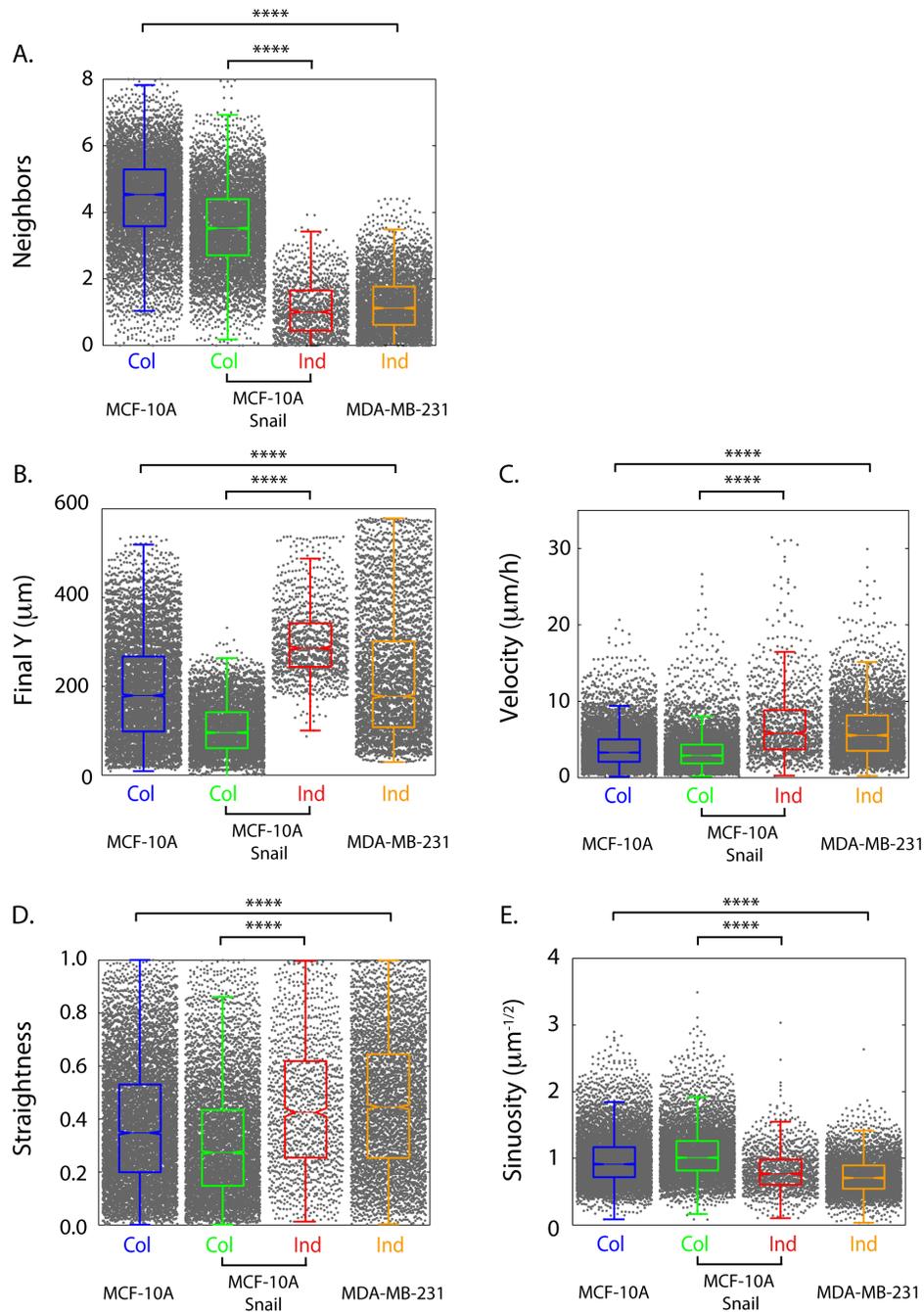


Figure S9. Comparison of quantitative metrics for collective (MCF-10A; MCF-10A Snail) and individual (MCF-10A Snail, MDA-MB-231) migration. **** denotes significance at $p < 10^{-6}$ by two-sample Kolmogorov-Smirnov test. **(A)** Cells that display collective migration encounter a lifetime average of ~ 4 nearest neighbors, while those that display individual migration encounter < 1 neighbor. **(B)** Collective migration is associated with decreased distance invaded into the device relative to individual migration. **(C)** Cells that display collective migration travel at half the average velocity of those that display individual migration. **(D, E)** Cells that display collective migration travel through less straight or more tortuous trajectories than those that display individual migration.

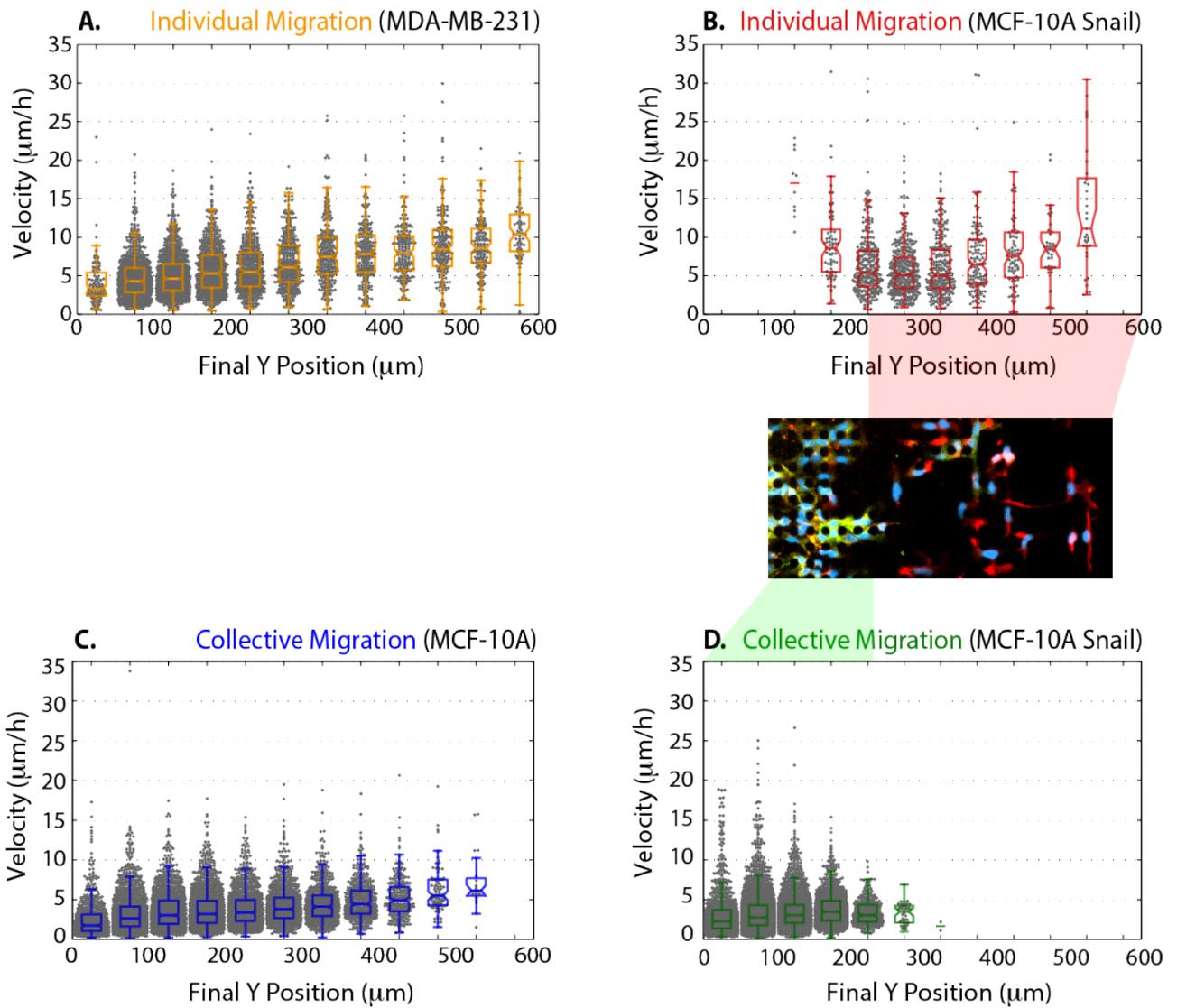


Figure S10. Comparison of migration velocity distributions with final Y position. For all cell types, the migration velocity decreases with distance from the front, likely due to increased crowding. **(A)** Individual migration (MDA-MB-231), **(B)** Individual migration (MCF-10A Snail). **(C)** Collective migration (MCF-10A), **(D)** Collective migration (MCF-10A Snail)

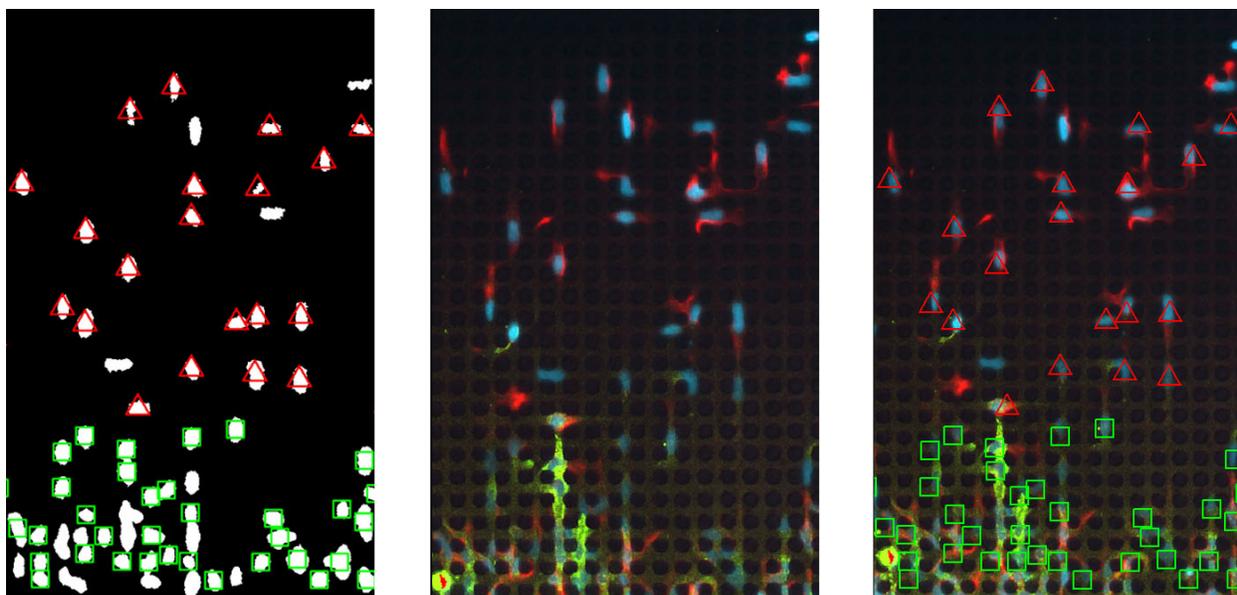


Figure S11. Representative data for induced EMT (MCF-10A Snail). Comparison of profiling based on (Left) collective and individual migration with (Middle) Immunostaining for epithelial (E-cadherin, green) or mesenchymal (vimentin, red) biomarkers). (Right) Merge migration behavior and immunostaining.

Profiled collective migration is 94% consistent with epithelial biomarker expression (E-cad), while profiled individual migration is 92% consistent with mesenchymal biomarker expression (vim). The disagreement is primarily due to cells at the migration front in the process of breaking away when the immunostaining occurred.

Table S1: Comparison of 848 total cells by migration behavior and biomarker expression

	Profiled Collective Migration	Profiled Individual Migration
Epithelial Biomarker (E-cad)	606	20
Mesenchymal Biomarker (Vim)	38	242

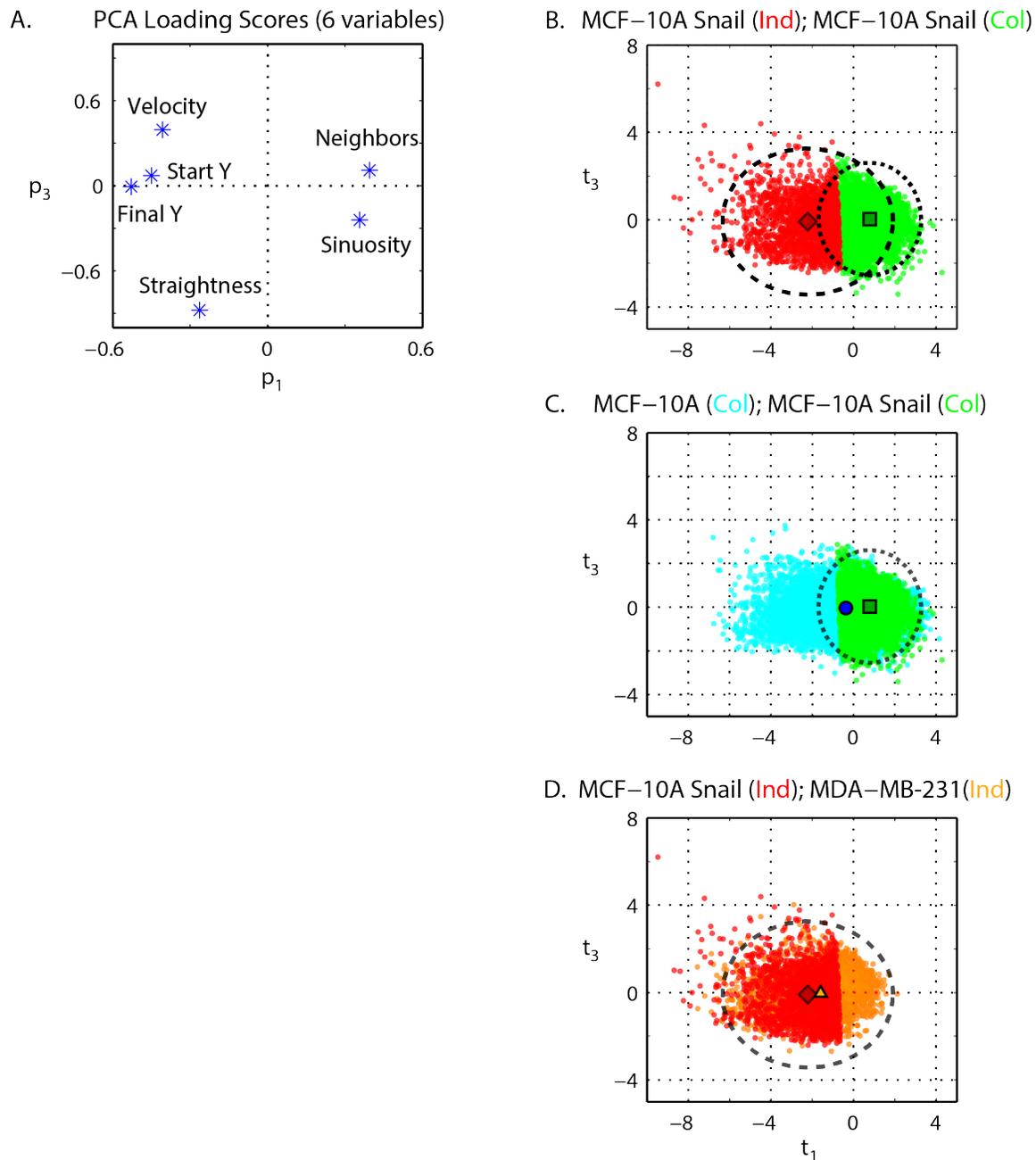


Figure S12. (A) Principle component analysis was used to generate a loading plot, indicating that individual migration is characterized by few nearest neighbors, low sinuosity, high average velocity, starting and final Y position, while collective migration is characterized by many nearest neighbors, high sinuosity, low average velocity, starting and final Y position (Note SN2). (B) A Gaussian mixture model was applied to cluster collectively or individually migrating subpopulations from the induced EMT population (MCF-10A Snail). (C) The projected migration behaviors of a purely epithelial population (MCF-10A) are comparable to the collectively invading subpopulation after induced EMT (MCF-10A Snail). (D) The projected migration behaviors of a purely mesenchymal population (MCF-10A) are comparable to the individually invading subpopulation after induced EMT (MCF-10A Snail).

Table S2A. Covariance(**X**) shows correlations between various migration metrics, but the overall distribution is not simply captured by any one metric or pair of metrics.

	Final Y	Velocity	Sinuosity	Straightness	Neighbors	Start Y
Final Y	1.00	0.35	-0.27	0.24	-0.50	0.88
Velocity		1.00	-0.65	0.23	-0.31	0.20
Sinuosity			1.00	-0.33	0.16	-0.11
Straightness				1.00	-0.19	0.08
Neighbors					1.00	-0.39
Start Y						1.00

Table S2B. 6x6 Loadings **V** corresponding to the principal components (eigenvectors) of **C**. The first 4 principal components were sufficient to capture the overall variance, so a 6x4 truncated loading **P** was used (black text), without including the last 2 principal components (gray text).

	V					
	P					
	<i>P₁</i>	<i>P₂</i>	<i>P₃</i>	<i>P₄</i>	<i>P₅</i>	<i>P₆</i>
Final Y	-0.53	0.34	-0.01	0.25	-0.02	-0.74
Velocity	-0.41	-0.44	0.39	-0.11	-0.69	0.07
Sinuosity	0.36	0.56	-0.24	-0.18	-0.68	-0.04
Straightness	-0.26	-0.34	-0.88	0.13	-0.15	0.09
Neighbors	0.40	-0.18	0.11	0.87	-0.21	-0.07
Start Y	-0.45	0.48	0.07	0.35	0.00	0.66

Table S2C. For each principal component (eigenvector) of **C**, the corresponding eigenvalues, the percentage of variance captured (eigenvalue normalized by the sum of all eigenvalues) and the cumulative percentage of variance captured. Note that the first 4 principal components capture ~93% of the variance (black text), so the remaining 2 are not needed (gray text).

Principal Component	<i>Eigenvalues of Cov(X)</i>	<i>% Variance Captured</i>	<i>Cumulative % Variance</i>
<i>P₁</i>	2.69	44.89	44.89
<i>P₂</i>	1.39	23.20	68.09
<i>P₃</i>	0.83	13.76	81.85
<i>P₄</i>	0.68	11.30	93.15
<i>P₅</i>	0.32	5.33	98.48
<i>P₆</i>	0.09	1.52	100.00

Table S3D. Centroids and axes of 95% Confidence Ellipses. Note that since the maximum variance occurs along *P₁*, the centroids of each clustered subpopulation is displaced primarily along *P₁* and not much along *P₃*

		Centroid <i>P₁</i>	Centroid <i>P₃</i>	Axis <i>P₁</i>	Axis <i>P₃</i>
Individual	MCF-10A Snail	-2.22	-0.09	4.12	3.34
Individual	MDA-MB-231	-1.59	0.04	-	-
Collective	MCF-10A Snail	0.79	0.03	2.48	3.08
Collective	MCF-10A	-0.37	-0.03	-	-

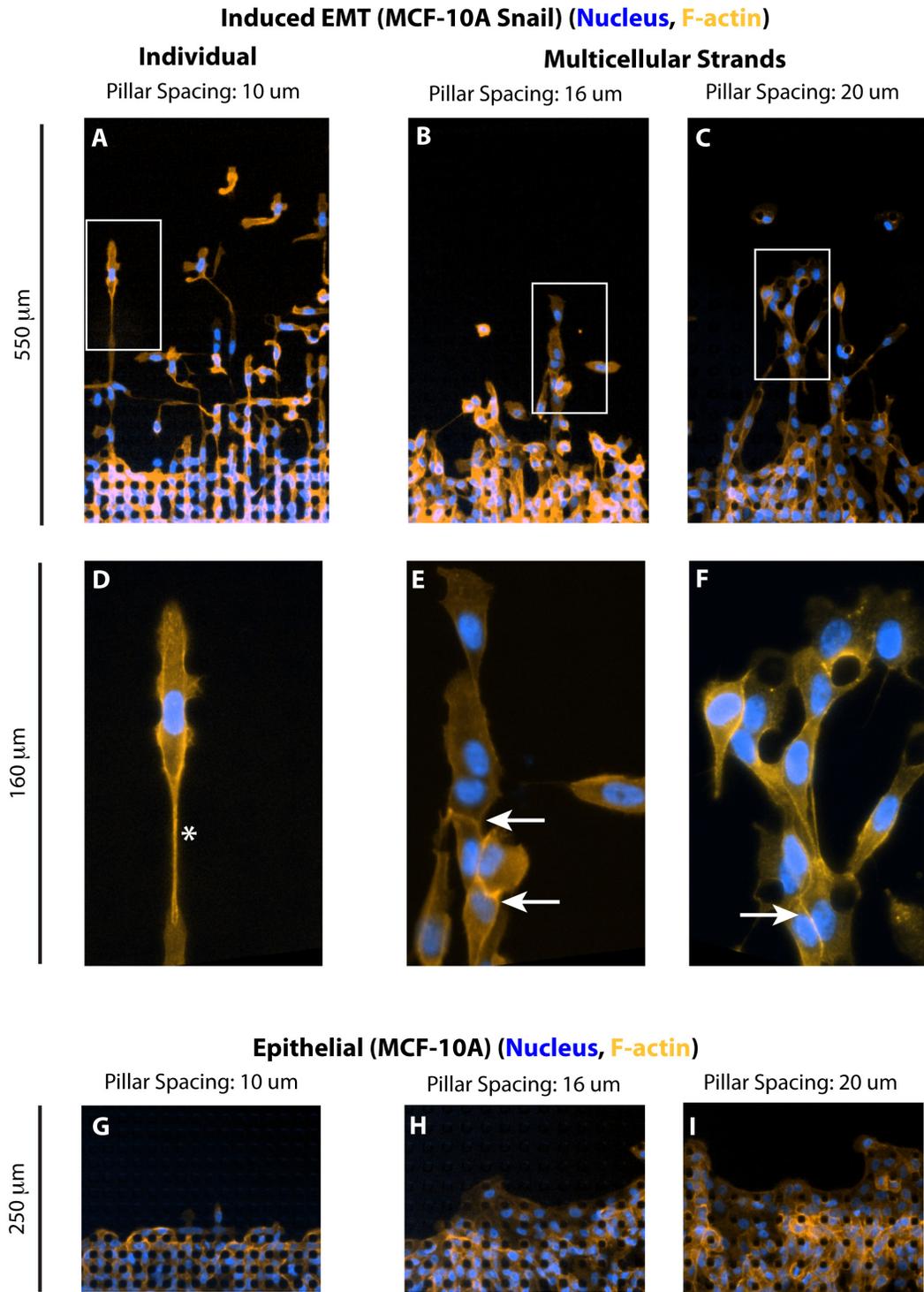
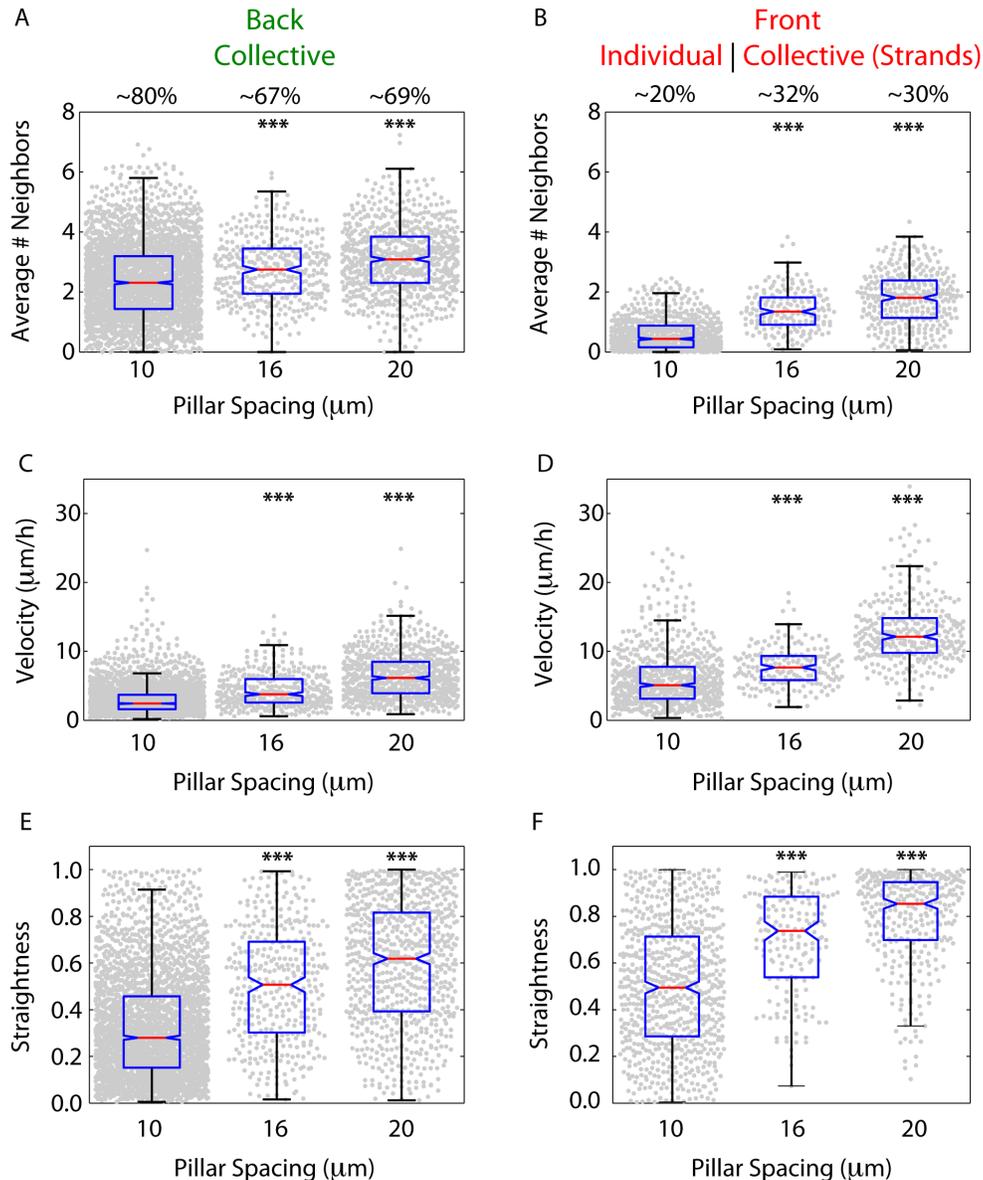


Figure S13: Increasing pillar spacing decreases individual scattering and increases multicellular strand formation. Immunostaining for F-actin and Nucleus. Induced EMT (MCF-10A Snail) at pillar spacing of 10 μm (A, D), 16 μm (B, E) and 20 μm (C, F). * indicates actin stress fibers, arrow indicates cell-cell junctions. Epithelial (MCF-10A) at pillar spacing of 10 μm (G), 16 μm (H) and 20 μm (I).

Induced EMT (MCF-10A Snail)



* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Figure S14. Comparison of quantitative metrics for Induced EMT (MCF-10A Snail) in different pillar spacings (10 μm , 16 μm , 20 μm). Overall, as pillar spacing is increased, there is still collective sheet migration at the rear (A) but a transition from individual scattering to multicellular strands (B), with increasing numbers of nearest neighbors. The distribution of cells at the front also shifts from ~20% for 10 μm spacing to ~30% for 16 μm and 20 μm spacing. The collective sheet migration still displays reduced velocity and straightness (C, E) relative to the individual or multicellular strands (D, F). A, B. Comparison of lifetime averaged nearest neighbors for front and back cells. C, D. Comparison of average velocities for front and back cells. E, F. Comparison of straightness for front and back cells.

Epithelial (MCF-10A)

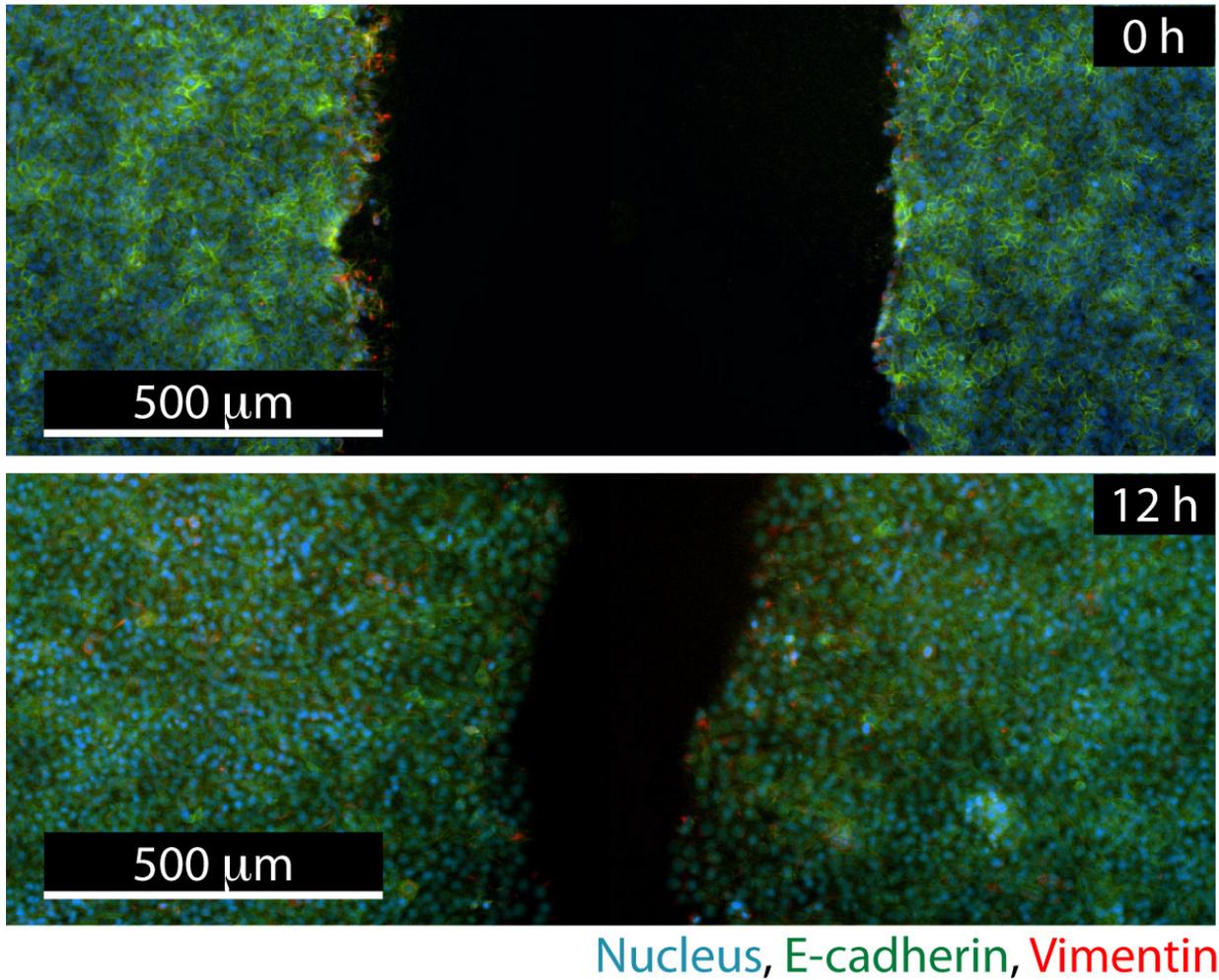
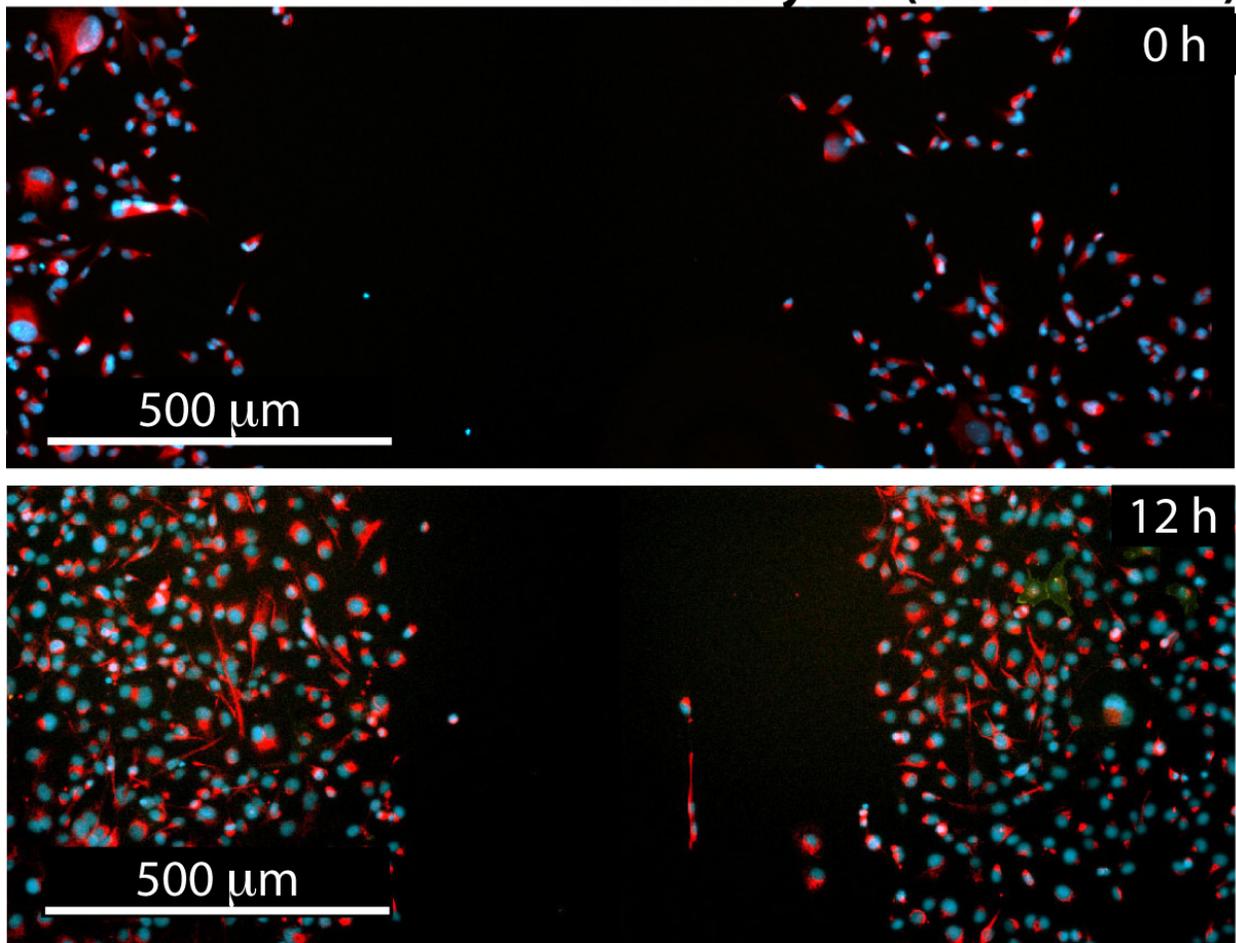


Figure S15. Immunostaining of classical wound healing assays reveals a bias towards sheet-like collective migration with limited scattering. Epithelial populations (MCF-10A) express E-cadherin (green).

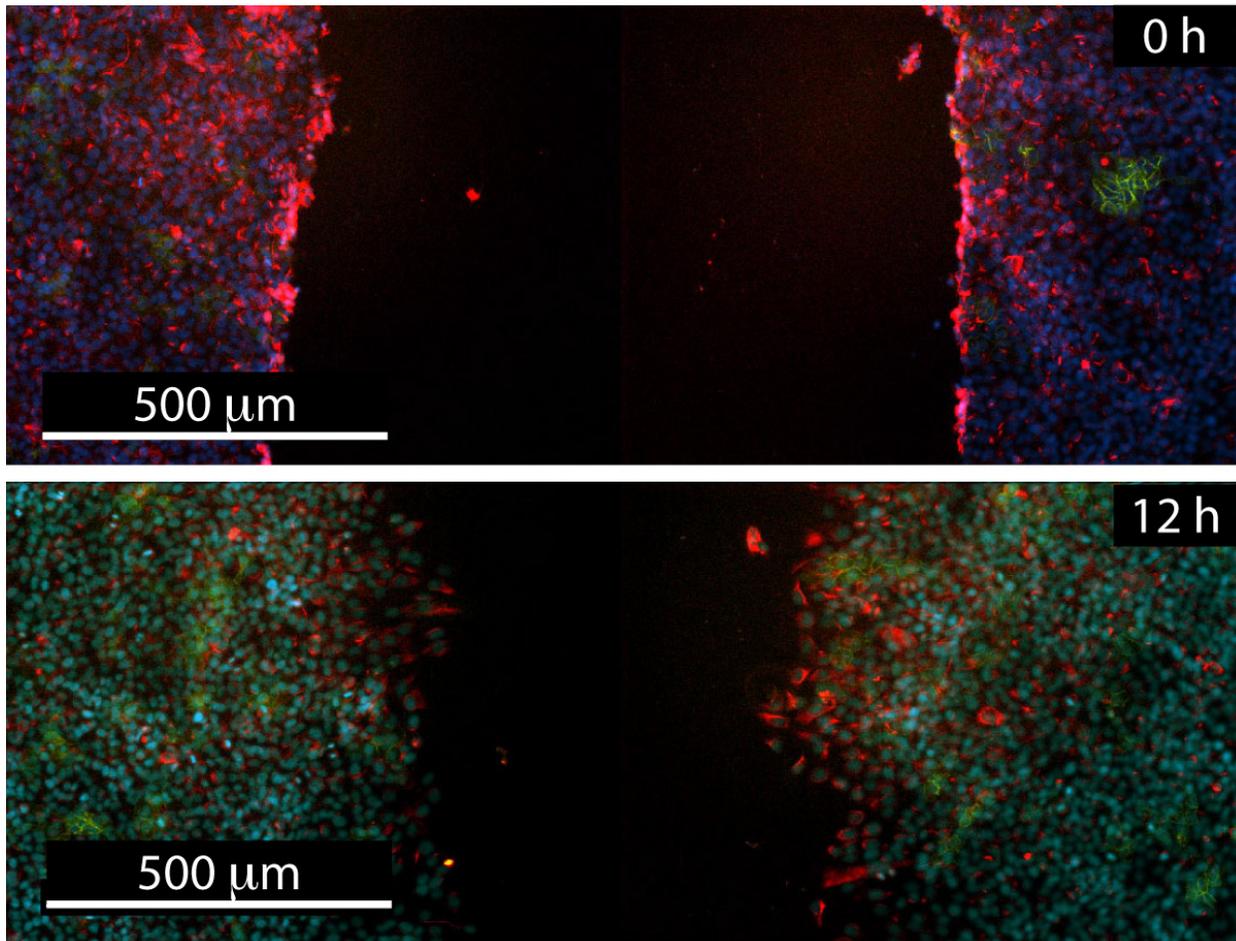
Mesenchymal (MDA-MB-231)



Nucleus, E-cadherin, Vimentin

Figure S15 (cont). Immunostaining of classical wound healing assays reveals a bias towards sheet-like collective migration with limited scattering. Mesenchymal populations (MDA-MB-231) express vimentin (red).

Induced EMT (MCF-10A Snail)



Nucleus, E-cadherin, Vimentin

Figure S15 (cont). Immunostaining of classical wound healing assays reveals a bias towards sheet-like collective migration with limited scattering. After EMT (MCF-10A Snail), cells initially express vimentin (red) but begin to express E-cadherin (green) after 12 h.

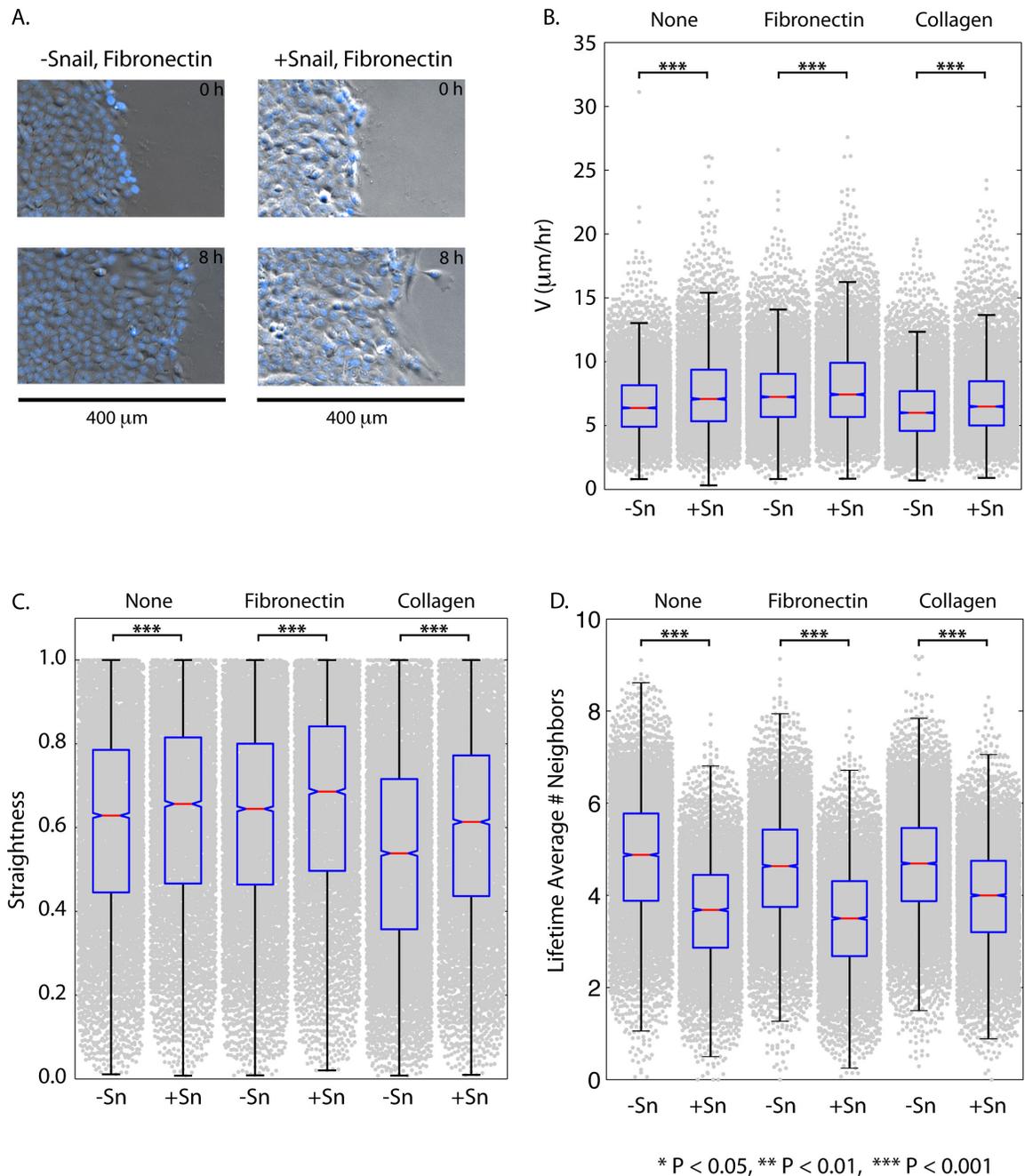


Figure S16. Comparison of quantitative metrics for 2D wound healing assays (MCF-10A; MCF-10A Snail) on different ECM proteins. Overall, induced EMT (MCF-10A Snail) cells exhibit increased velocities, straightness and decreased number of nearest neighbors relative to epithelial (MCF-10A cells). A. Representative time-lapse images. B. Comparison of average velocities. C. Comparison of straightness. D. Lifetime averaged number of nearest neighbors.

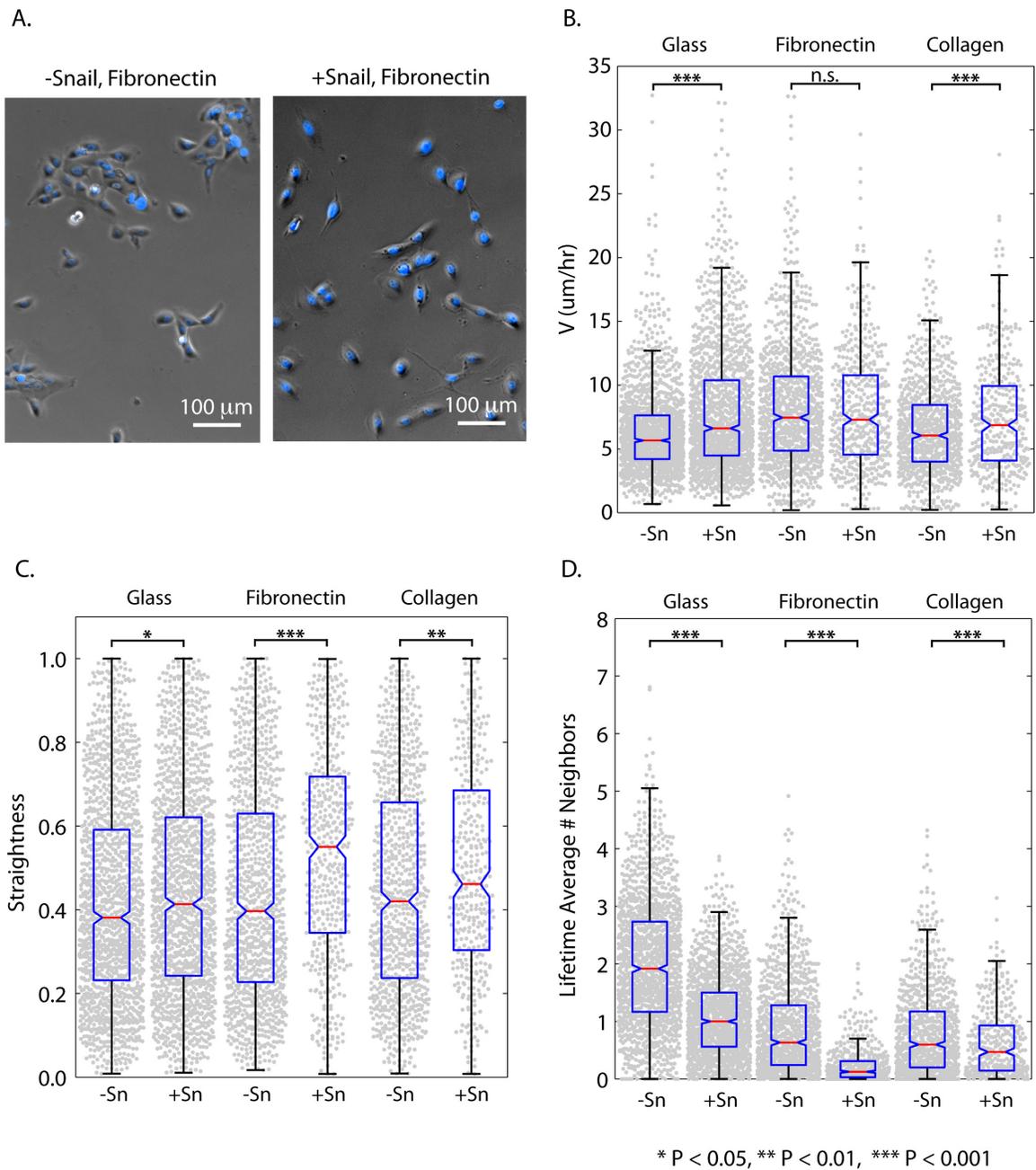
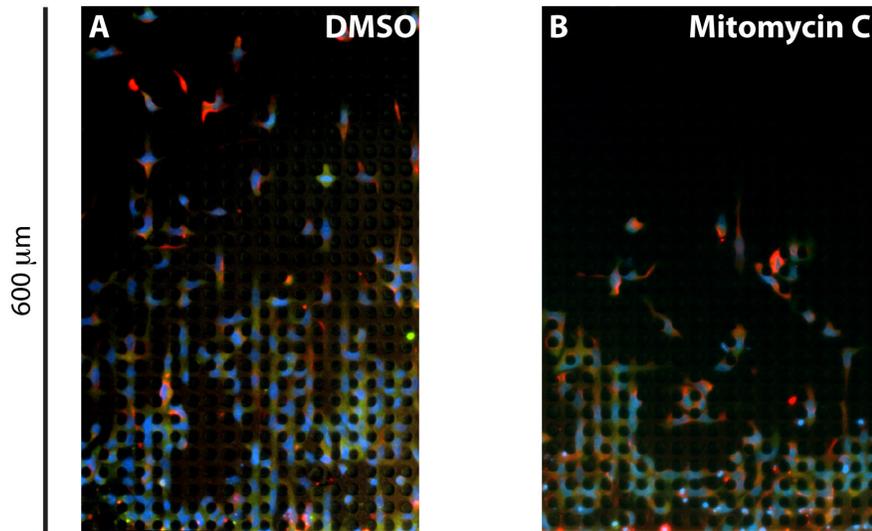


Figure S17. Comparison of quantitative metrics for isolated cell migration in 2D (MCF-10A; MCF-10A Snail) on different ECM proteins. Overall, induced EMT (MCF-10A) Snail cells exhibit increased velocities, straightness and decreased number of nearest neighbors relative to epithelial (MCF-10A) cells. However, there is no statistically significant difference in velocity on fibronectin, perhaps due to strong 2D adhesions. A. Representative images of cell morphology. B. Comparison of average velocities. C. Comparison of straightness. D. Lifetime averaged number of nearest neighbors.

Induced EMT (MCF-10A Snail) (Nucleus, E-cadherin, Vimentin)



Epithelial (MCF-10A) (Nucleus, E-cadherin, Vimentin)

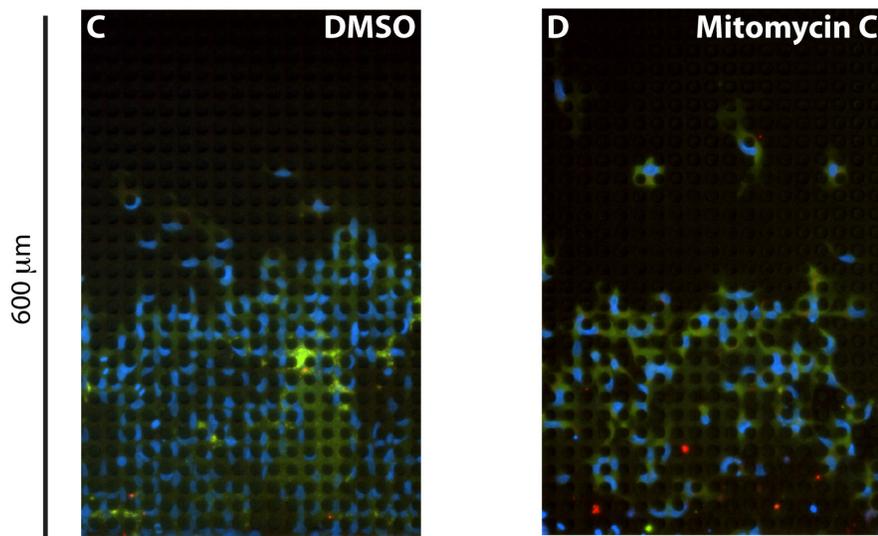


Figure S18. Individual scattering and collective migration is still observed after inhibition of proliferation with the DNA-alkylating agent mitomycin C (MMC) (25 $\mu\text{g}/\text{mL}$). Nevertheless, there is some reduction in overall distance migrated and cell density. **(A)** Induced EMT (MCF-10A Snail) control condition, **(B)** Induced EMT (MCF-10A Snail) with mitomycin C treatment, **(C)** Epithelial (MCF-10A) control condition, **(D)** Epithelial (MCF-10A) with mitomycin C treatment.

Induced EMT (MCF-10A Snail)

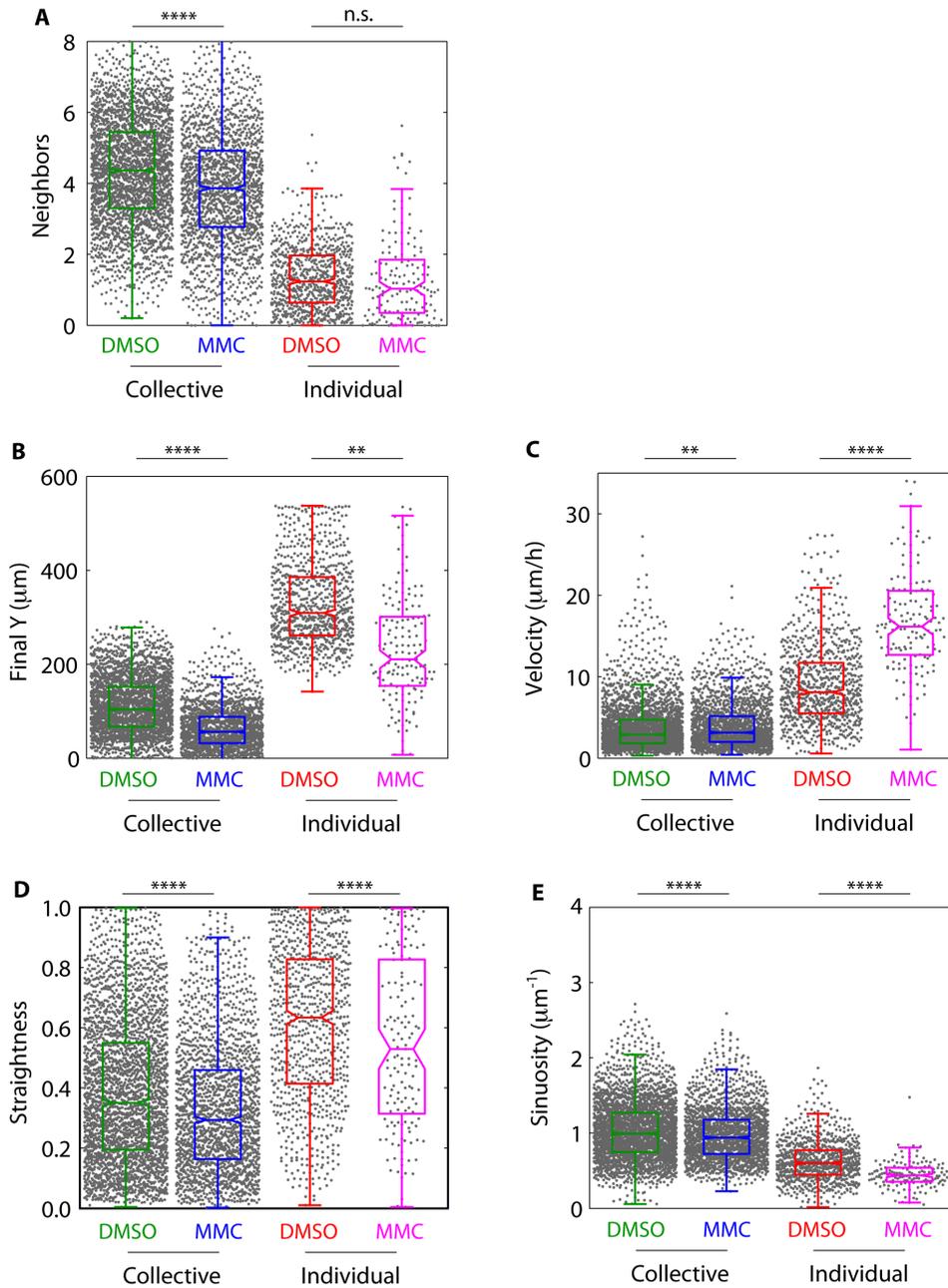


Figure S19. Comparison of metrics for collective and individual migration with induced EMT (MCF-10A Snail) cells after inhibition of proliferation with mitomycin C (MMC) (25 $\mu\text{g}/\text{mL}$). After MMC treatment, individual cells are still scattered, albeit with lower cell densities and reduced distance traveled. **(A)** Lifetime-averaged nearest neighbors: collectively migrating cells are less dense after MMC treatment, whereas individually migrating cells migrate with a comparable number of neighbors. **(B)** Overall distance traveled into the device decreases after MMC treatment. **(C)** Velocity of collective migration remains similar after MMC treatment, although velocity of individual migration is increased. **(D, E)** Straightness decreases after MMC treatment. ** $p < 0.01$, **** $p < 10^{-6}$ by two-sample Kolmogorov-Smirnov test.

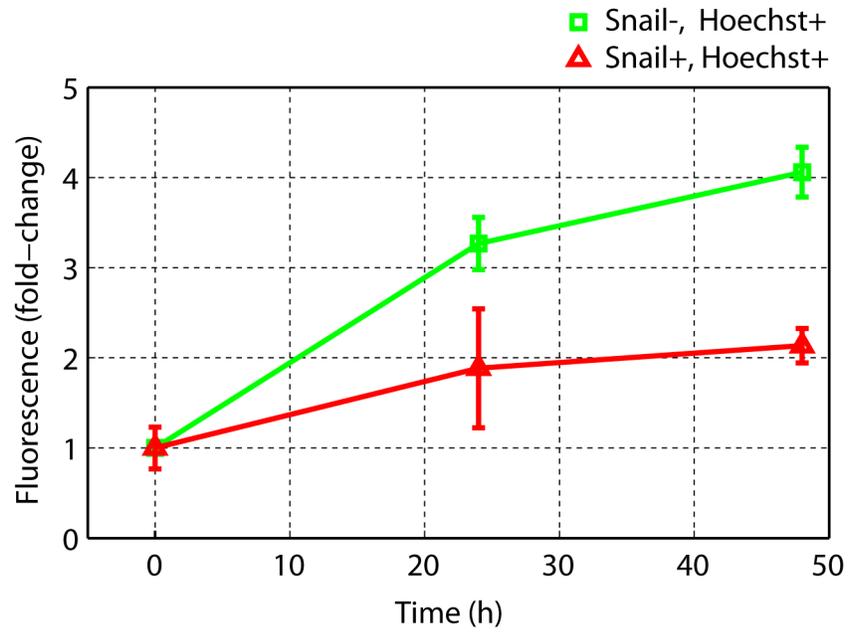


Figure S20. Proliferation assay for Snail- and Snail+ cells with Hoechst staining. Snail- cells divide approximately every 24 h, while Snail+ cells divide approximately every 48 h.

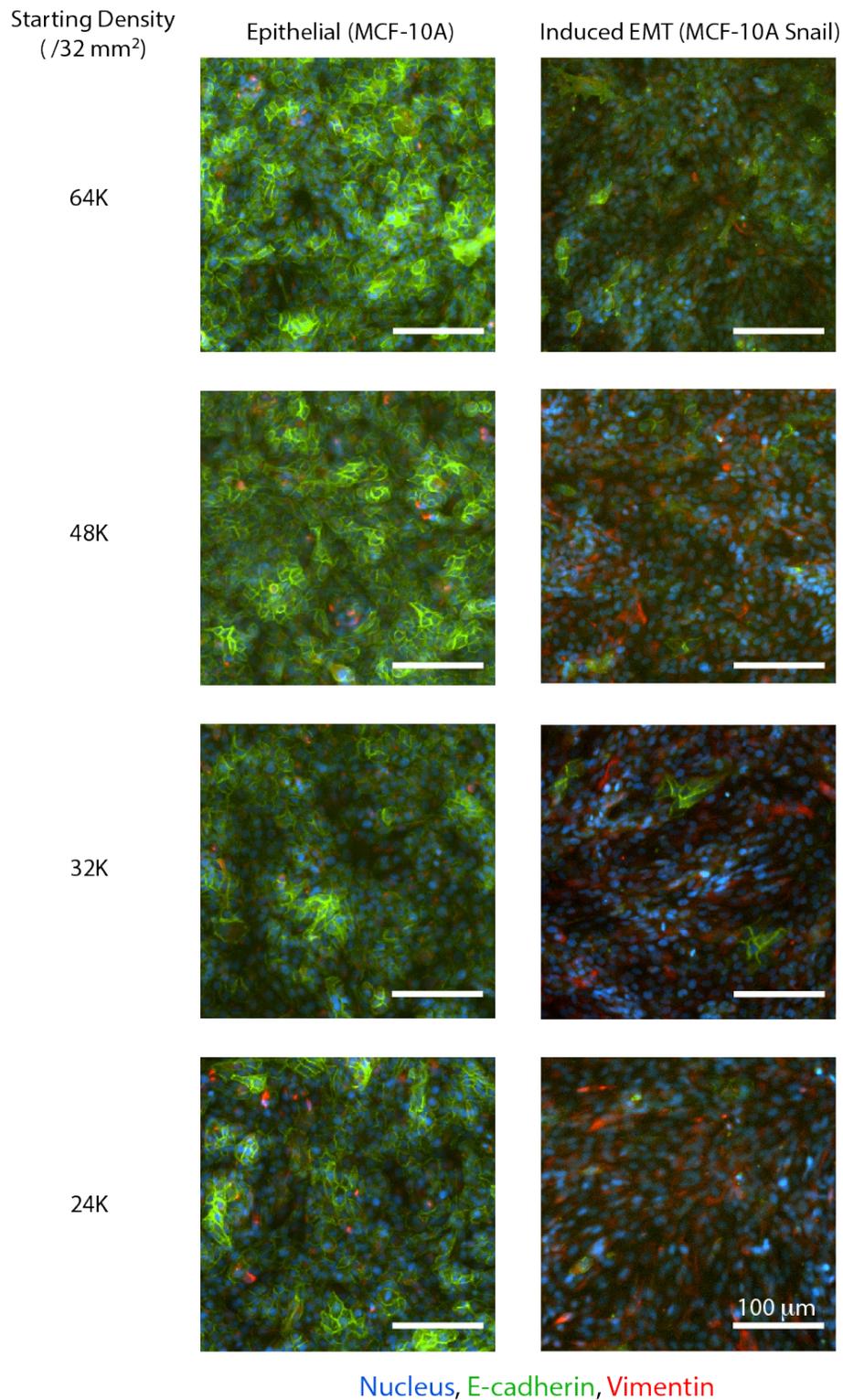


Figure S21: Epithelial (E-cadherin) and mesenchymal (vimentin) biomarker expression after 72 h at varying initial plating densities. MCF-10A Snail show increasing E-cadherin expression and decreasing vimentin expression with increasing cell density.

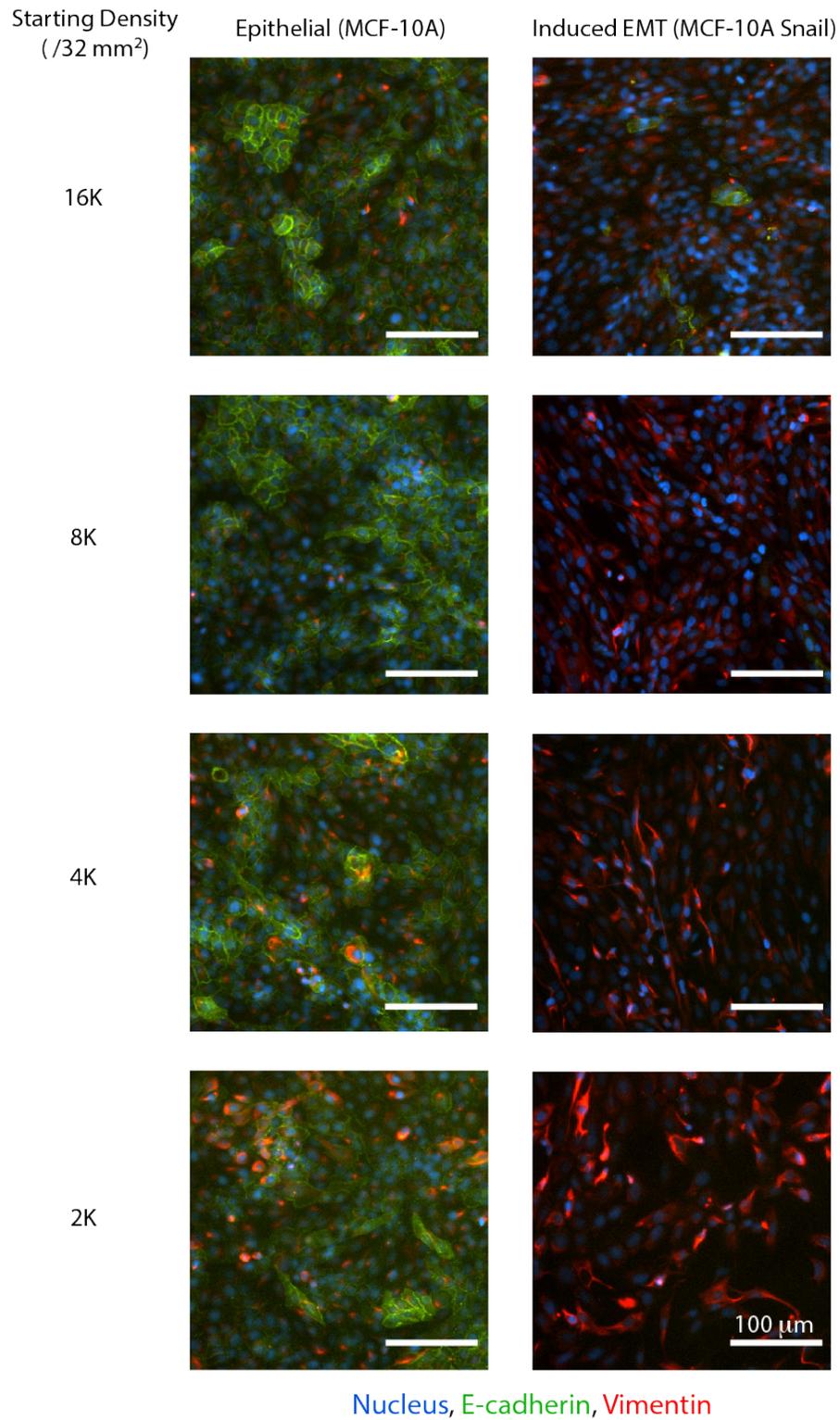


Figure S21: (cont): Epithelial (E-cadherin) and mesenchymal (vimentin) biomarker expression after 72 h at varying initial plating densities. MCF-10A Snail show increasing E-cadherin expression and decreasing vimentin expression with increasing initial cell density.

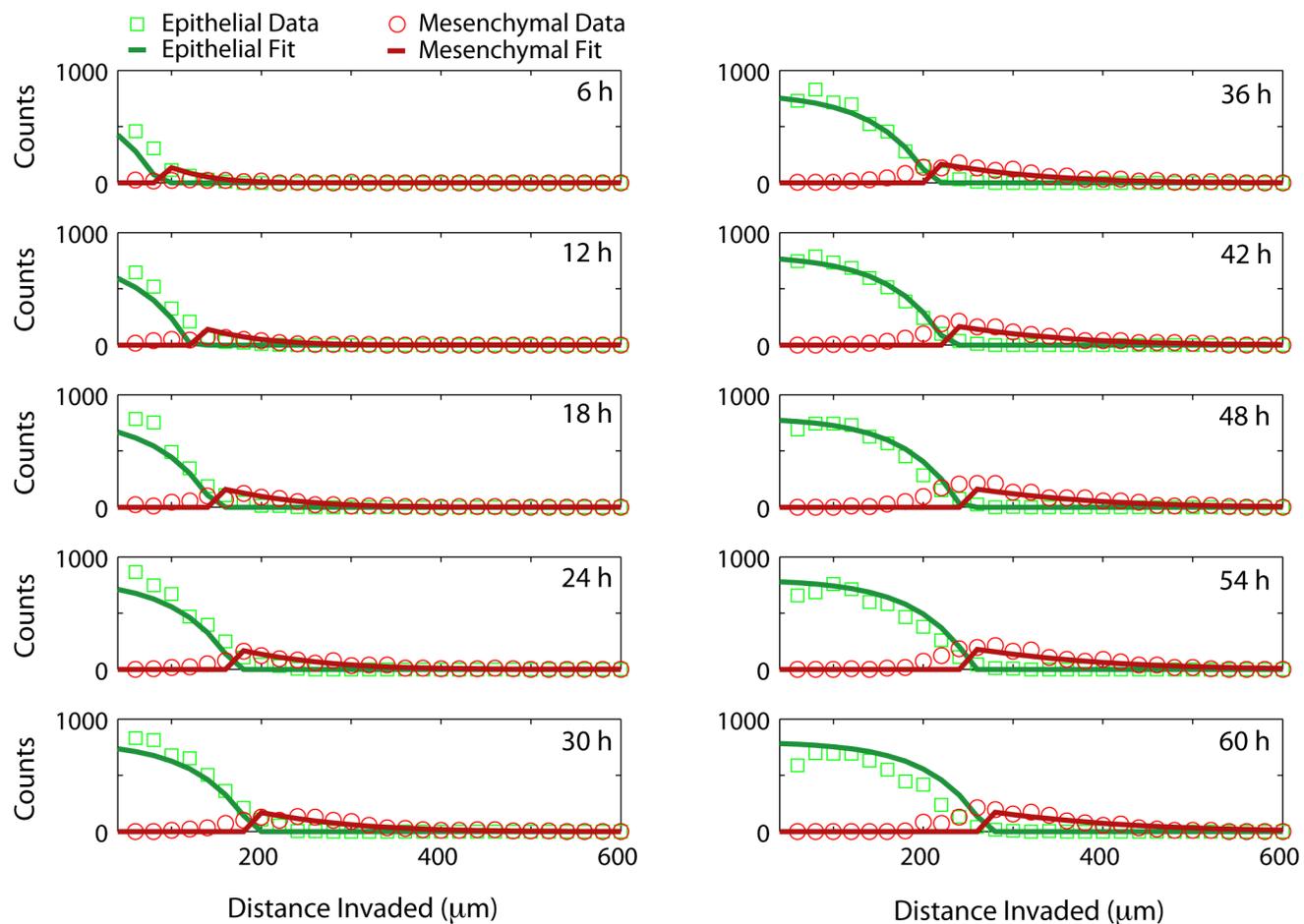


Figure S22. Best fits of solidification model show good agreement with experimental data for the control condition (DMSO) with the induced EMT (MCF-10A Snail) cells. $D_E = 306 \mu\text{m}^2/\text{h}$, $D_M = 545 \mu\text{m}^2/\text{h}$, $C_0 = 622$.

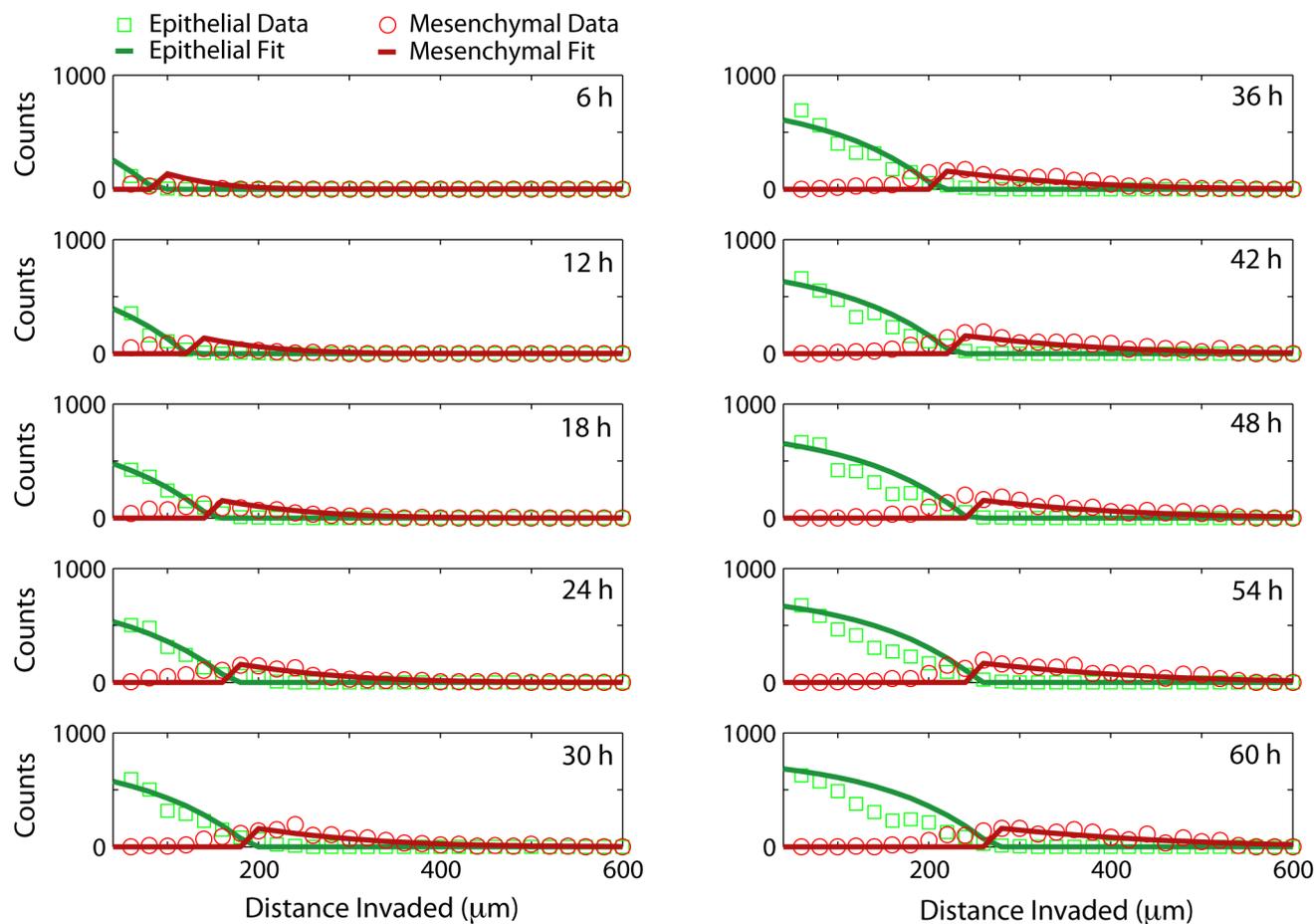


Figure S23. Best fits of solidification model show good agreement with experimental data for BID-1870 with the induced EMT (MCF-10A Snail) cells. $D_E = 196 \mu\text{m}^2/\text{h}$, $D_M = 676 \mu\text{m}^2/\text{h}$, $C_0 = 475$.

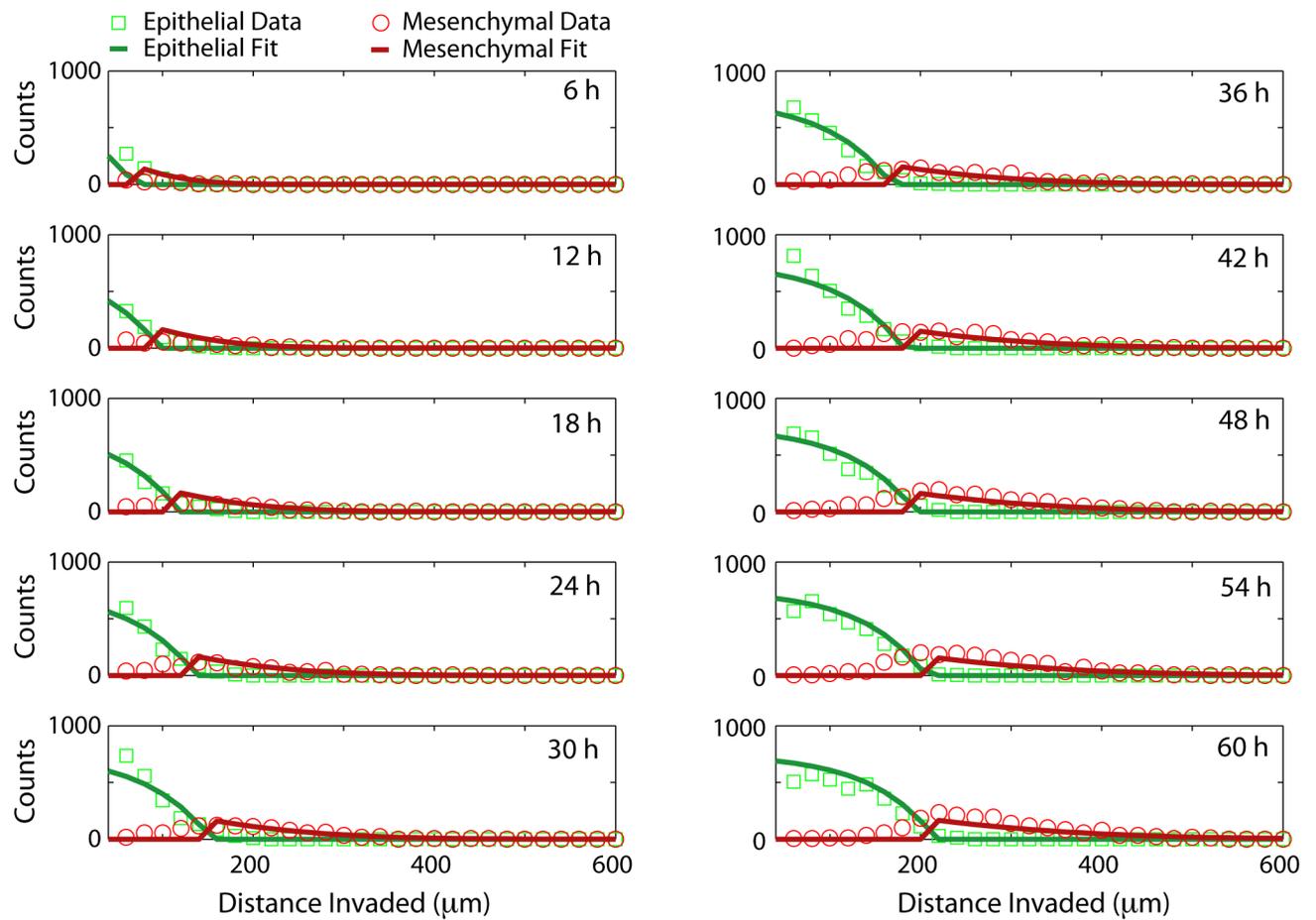


Figure S24. Best fits of solidification model show good agreement with experimental data for U0126 with the induced EMT (MCF-10A Snail) cells. $D_E = 210 \mu\text{m}^2/\text{h}$, $D_M = 488 \mu\text{m}^2/\text{h}$, $C_0 = 466$.

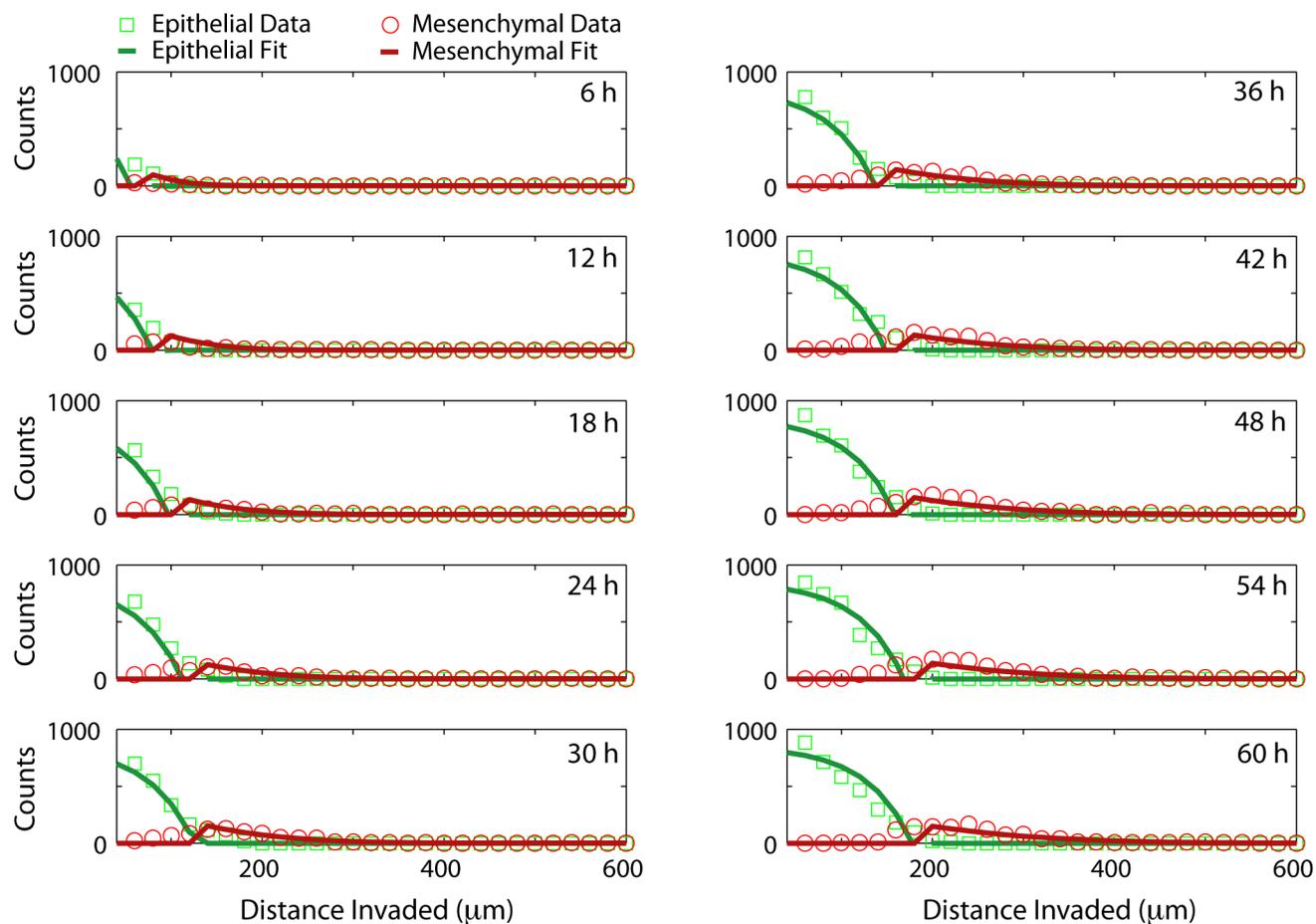


Figure S25. Best fits of solidification model show good agreement with experimental data for FMKMEA with the induced EMT (MCF-10A Snail) cells. $D_E = 156 \mu\text{m}^2/\text{h}$, $D_M = 328 \mu\text{m}^2/\text{h}$, $C_0 = 474$.

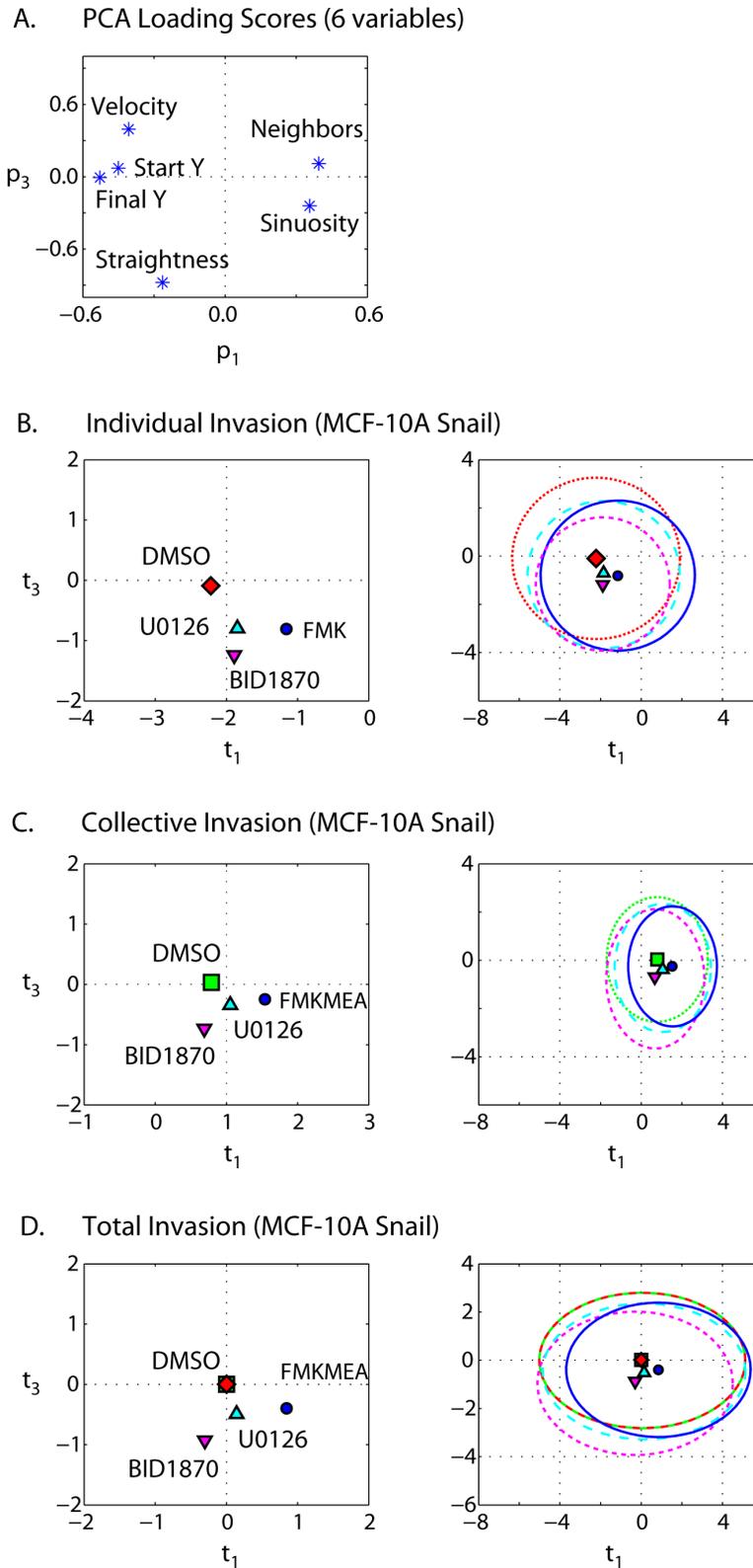


Figure S26. (A) Principle component analysis was used to generate a loading plot, indicating that individual migration is characterized by few nearest neighbors, small sinuosity, and high

average velocity and high start and final Y position, while collective migration is characterized by many nearest neighbors, high sinuosity, low average velocity and low start and final Y position (Note SN3). Centroid and 95% confidence ellipses for migration of induced EMT (MCF-10A Snail) when treated with DMSO, BID1870, U0126 and FMKMEA. **(B)** Individual migration **(C)** Collective migration **(D)** Overall collective and individual migration

Table S3A. 4x4 Covariance of truncated loadings T_L for BID-1870, MCF-10A Snail

	P_1	P_2	P_3	P_4
P_1	2.39	0.19	0.28	-0.23
P_2	0.19	1.70	0.37	-0.02
P_3	0.28	0.37	0.91	-0.08
P_4	-0.23	-0.02	-0.08	0.63

Table S3B. For each principal component (eigenvector) of C for BID-1870, MCF-10A Snail, the corresponding eigenvalues, the percentage of variance captured (eigenvalue normalized by the sum of all eigenvalues) and the cumulative percentage of variance captured. Note that the first 4 principal components capture ~93% of the variance (black text), so the remaining 2 are not needed (gray text).

Principal Component	Eigenvalues of Cov(X)	% Variance Captured	Cumulative % Variance
P_1	2.39	39.33	39.33
P_2	1.70	27.93	67.26
P_3	0.91	15.01	82.27
P_4	0.63	10.31	92.58
P_5	0.31	5.12	97.70
P_6	0.14	2.30	100.00

Table S3C. 4x4 Covariance of truncated loadings T_L for U0126, MCF-10A Snail

	P_1	P_2	P_3	P_4
P_1	2.68	0.24	0.28	0.02
P_2	0.24	1.31	0.19	-0.09
P_3	0.28	0.19	0.86	-0.06
P_4	0.02	-0.09	-0.06	0.63

Table S3D. For each principal component (eigenvector) of C for U0126, MCF-10A Snail, the corresponding eigenvalues, the percentage of variance captured (eigenvalue normalized by the sum of all eigenvalues) and the cumulative percentage of variance captured. Note that the first 4 principal components capture ~93% of the variance (black text), so the remaining 2 are not needed (gray text).

Principal Component	Eigenvalues of Cov(X)	% Variance Captured	Cumulative % Variance
P_1	2.68	45.77	45.77
P_2	1.31	22.26	68.02
P_3	0.86	14.59	82.61
P_4	0.63	10.82	93.43
P_5	0.27	4.68	98.11
P_6	0.11	1.89	100.00

Table S3E. 4x4 Covariance of truncated loadings T_L for FMKMEA, MCF-10A Snail

	P_1	P_2	P_3	P_4
P_1	2.16	0.34	0.21	0.26
P_2	0.34	1.27	0.10	-0.10
P_3	0.21	0.10	0.82	0.03
P_4	0.26	-0.10	0.03	0.72

Table S3F. For each principal component (eigenvector) of C for FMKMEA, MCF-10A Snail, the corresponding eigenvalues, the percentage of variance captured (eigenvalue normalized by the sum of all eigenvalues) and the cumulative percentage of variance captured. Note that the first 4 principal components capture ~92% of the variance (black text), so the remaining 2 are not needed (gray text).

Principal Component	Eigenvalues of Cov(X)	% Variance Captured	Cumulative % Variance
P_1	2.16	40.00	40.00
P_2	1.27	23.54	63.54
P_3	0.82	15.08	78.62
P_4	0.72	13.27	91.89
P_5	0.36	6.69	98.58
P_6	0.08	1.42	100.00

Table S3G. Centroids and axes of 95% Confidence Ellipses for DMSO and Rsk inhibitors with MCF-10A Snail

		Centroid P_1	Centroid P_3	Axis P_1	Axis P_3
Individual	DMSO	-2.22	-0.09	4.12	3.34
Individual	BID 1870	-1.87	-1.25	3.29	2.76
Individual	U0126	-1.85	-0.77	3.74	3.05
Individual	FMKMEA	-1.15	-0.81	3.79	3.11
Collective	DMSO	0.79	0.03	2.48	3.08
Collective	BID 1870	0.69	-0.76	2.41	3.31
Collective	U0126	1.06	-0.32	2.36	2.65
Collective	FMKMEA	1.54	-0.25	2.18	2.82
All	DMSO	0.00	0.00	5.06	2.80
All	BID 1870	-0.30	-0.95	4.8	2.97
All	U0126	0.14	-0.46	4.98	2.81
All	FMKMEA	0.84	-0.39	4.53	2.78