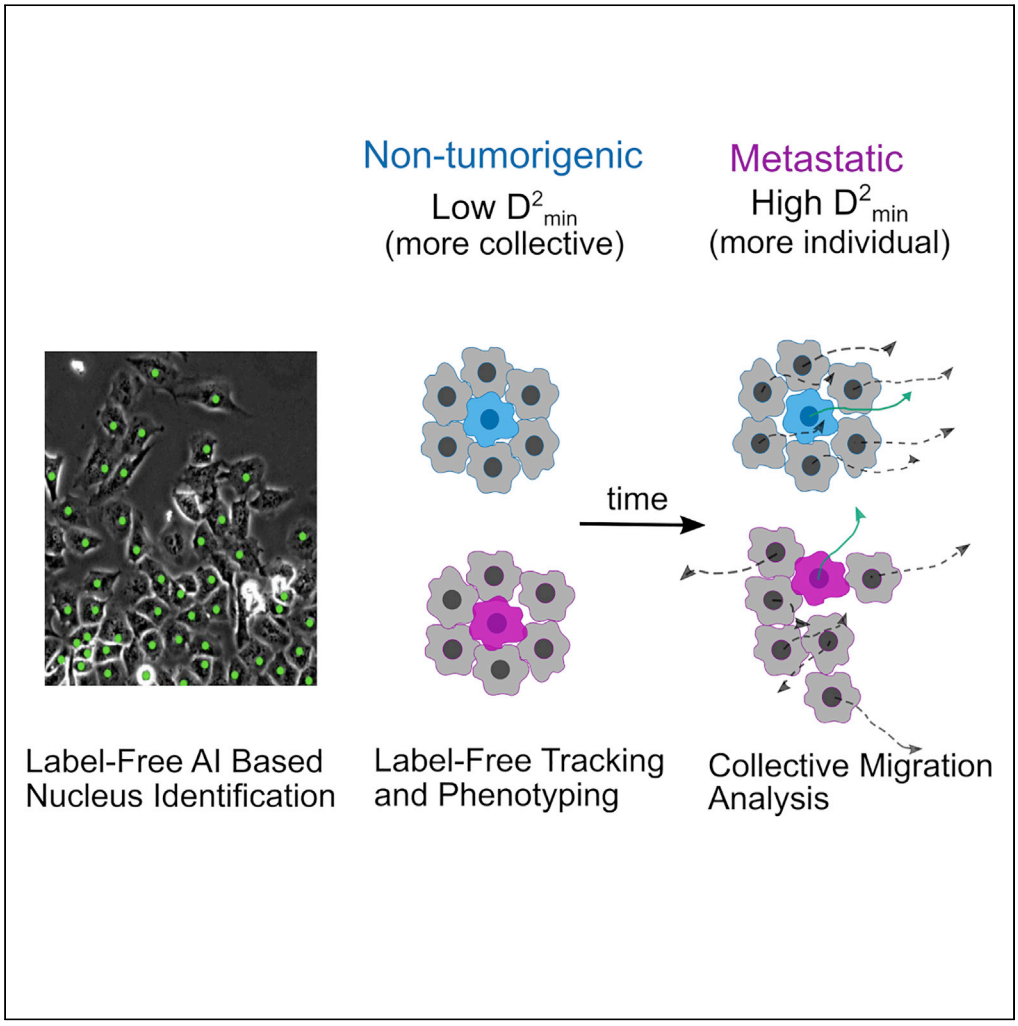
iScience

CellPress
OPEN ACCESS

Article

Label-free cell tracking enables collective motion phenotyping in epithelial monolayers



Shuyao Gu, Rachel M. Lee, Zackery Benson, ..., Stuart S. Martin, Joe Chalfoun, Wolfgang Losert

wlosert@umd.edu

Highlights

Versatile AI algorithm identifies individual cell tracks in phase contrast images

Motion of cells relative to nearby neighbors may indicate cancer progression

Gu et al., iScience *25*, 104678
July 15, 2022
https://doi.org/10.1016/j.isci.2022.104678

iScience

CellPress
OPEN ACCESS

Article

Label-free cell tracking enables collective motion phenotyping in epithelial monolayers

Shuyao Gu,[1] Rachel M. Lee,[2,3] Zackery Benson,[1] Chenyi Ling,[4] Michele I. Vitolo,[2,5] Stuart S. Martin,[2,5] Joe Chalfoun,[4] and Wolfgang Losert[1,3,6,*]

SUMMARY

Collective cell migration is an umbrella term for a rich variety of cell behaviors, whose distinct character is important for biological function, notably for cancer metastasis. One essential feature of collective behavior is the motion of cells relative to their immediate neighbors. We introduce an AI-based pipeline to segment and track cell nuclei from phase-contrast images. Nuclei segmentation is based on a U-Net convolutional neural network trained on images with nucleus staining. Tracking, based on the Crocker-Grier algorithm, quantifies nuclei movement and allows for robust downstream analysis of collective motion. Because the AI algorithm required no new training data, our approach promises to be applicable to and yield new insights for vast libraries of existing collective motion images. In a systematic analysis of a cell line panel with oncogenic mutations, we find that the collective rearrangement metric, $D^2_{min,}$ which reflects non-affine motion, shows promise as an indicator of metastatic potential.

INTRODUCTION

Breast cancer is one of the most common cancers in women. Many breast-cancer-related deaths result from metastasis, a complex, multistep process in which cancer cells migrate away from the primary tumor and eventually grow in other organs (Ferlay et al., 2015; Hagemeister et al., 1980; Klein, 2009). During the metastatic process, clusters of primary tumor cells can leave the original site, circulate in the bloodstream, and seed secondary tumor growth at distant sites. There is growing evidence that tumor cells that collectively migrate and enter the bloodstream as clusters (defined as any group larger than one cell) are 50–100 times more likely than individual cells to lead to metastasis (Aceto et al., 2014; Cheung et al., 2016; Wrenn et al., 2020). Characterization of collective motion is generally accomplished in one of two ways: PIV applied to phase contrast movies, which ignores individual cellular characteristics, or tracking applied to fluorescent labels, which tend to affect cell behavior. Here, we introduce a new artificial intelligence (AI)-based approach to track the motion of individual cells from phase contrast movies. This AI-enabled cell tracking allows us to characterize a shifting balance between individual motion and collective motility of mammary epithelial cells as they become more transformed, tumorigenic, and metastatic.

We use a commonly studied breast epithelial cell line (MCF10A), a highly metastatic cell line (MDA-MB-231), and a genetically defined model system for cancer and metastasis. The genetically defined model system is based on the MCF10A cell line and introduces phosphatase and tensin homolog deletion ($PTEN^{-/-}$) and/or overexpression of activated KRas (G12V), two common oncogenic mutations that have downstream impacts on the cytoskeleton. PTEN loss is found in 24% of breast cancers (López-Knowles et al., 2010) and has been associated with poorer prognoses in many solid tumors (Wang et al., 2018); KRas is a frequent driver of tumors across different types of human cancers (Herbst and Schlessinger, 2019; Sanchez-Vega et al., 2018). This model cell system relates our results to *in vivo* outcomes, since the cell lines have been characterized in terms of tumorigenicity (Thompson et al., 2015) and metastatic behavior (Yankaskas et al., 2019).

In a previous study, we imaged the motion of groups of cells from these cell lines in a well-defined *in vitro* assay (Lee et al., 2013) and characterized the motion using particle image velocimetry (PIV). We found that deletion of PTEN caused cells to move more collectively, while KRas activation had the opposing effect (Lee et al., 2021a). However, when both mutations were present, the phenotype induced by KRas dominated and the KRas and KRas/$PTEN^{-/-}$ cell lines displayed similar dynamics (Lee et al., 2021a; Yankaskas et al., 2019), even though they have significantly different *in vivo* outcomes. One might think that

[1]Department of Physics, University of Maryland, College Park, MD 20742, USA

[2]Marlene and Stewart Greenebaum NCI Comprehensive Cancer Center, University of Maryland School of Medicine, Baltimore, MD 21201, USA

[3]Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA

[4]Software and Systems Division, Information Technology Lab, NIST, Gaithersburg, MD 20899, USA

[5]Departments of Pharmacology and Physiology, University of Maryland School of Medicine, Baltimore, MD 21201, USA

[6]Lead contact

*Correspondence: wlosert@umd.edu

https://doi.org/10.1016/j.isci.2022.104678

malignancy would lead to higher cell motility; however, these model cell lines and others with characterized *in vivo* behavior do not show a clear correlation between cell speed and malignancy (Lee et al., 2013; Weiger et al., 2013). One limitation of PIV is that it characterizes motion based on the movement of all features in an image. Thus, PIV captures overall movement visible in an image, including the motion of the many organelles visible in a typical phase contrast image, which may obscure the full motion of each cell's nucleus. However, the cell nucleus serves as the main reference point of a cell, steering cell motion (Renkawitz et al., 2019) and defining orientation when compared to the microtubule-organizing center (Desai et al., 2009).

In this study, we introduce an AI-enabled (U-Net) cell nucleus segmentation tool for epithelial cells (Ling et al., 2020) that allows us to track cell nuclei and study their dynamics. The AI model was trained on stem cell images with Lamin B1 stain for automated nucleus detection (Ling et al., 2020) and did not require further training to perform nucleus segmentation on the studied epithelial cell sheets. Additionally, our AI-enabled single-cell tracking allows us to characterize motion of cells relative to their neighbors. In particular, we borrowed a metric from soft matter physics, D^2_{min} (Crocker and Grier, 1996; Falk and Langer, 1998; Lee et al., 2021a), to describe the extent of similarity of motion between the cell of interest and its closest neighbors. This ability to assess how closely a cell follows its neighbors yields a phenotype of cell rearrangements during collective migration capable of differentiating the metastatic KRas/PTEN$^{-/-}$ from the non-metastatic cell lines. AI-enabled tracking and quantification of collective rearrangements thus provide new tools for quantifying metastatic potential and could be used to improve patient diagnosis or to develop novel therapeutic targets.

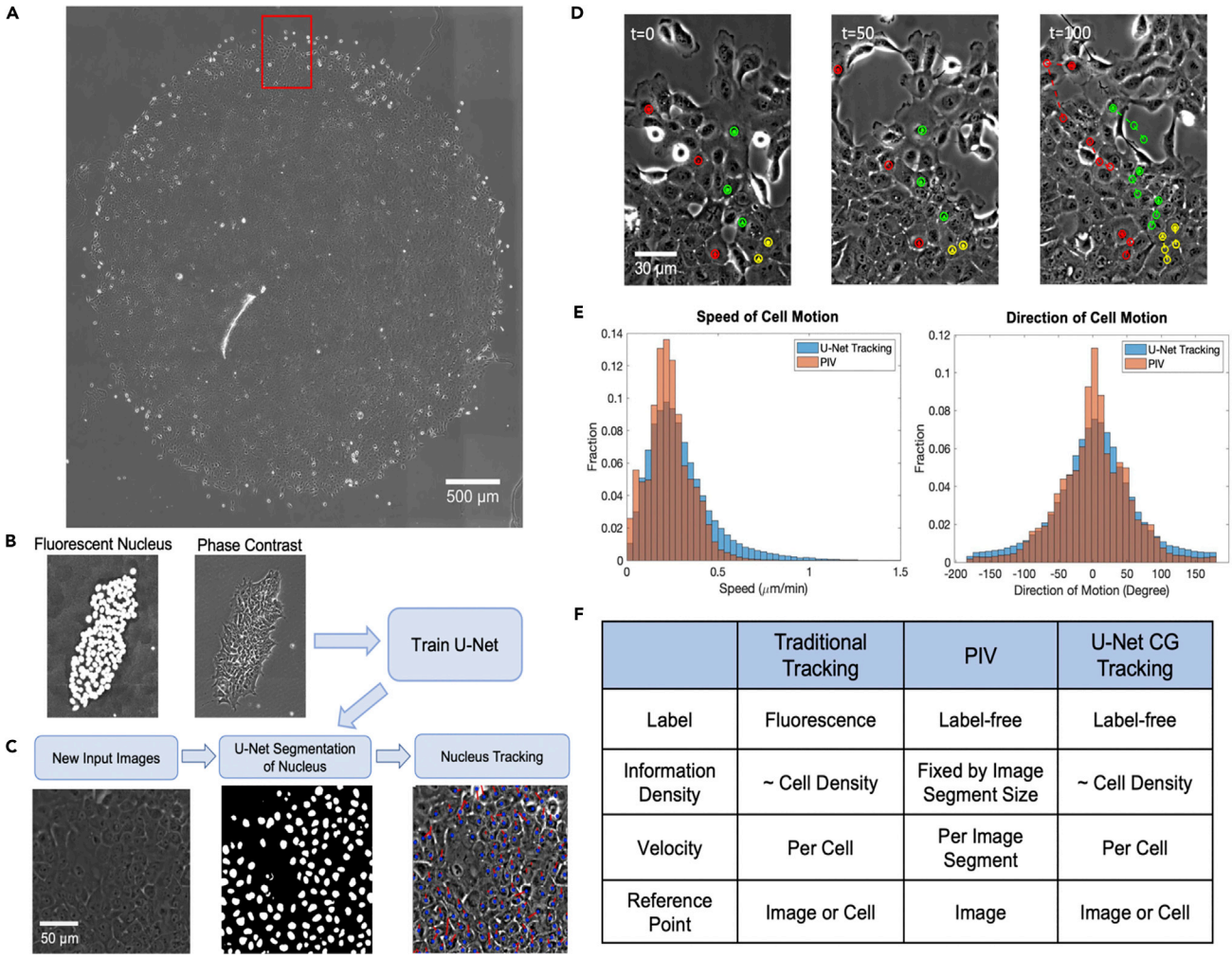
RESULTS

AI-based analysis workflow enables label-free single-cell tracking

Our analysis focuses on the collective dynamics of a migrating monolayer of epithelial cells (Figure 1A) imaged with phase-contrast microscopy. Nuclear segmentation and tracking are a powerful approach to studying cell migration dynamics. Here, we show that deep learning (U-Net)-based inferences of nuclear positions can provide a substitute for fluorescently labeled nuclei. We use a successful image analysis deep learning model, U-Net, that was originally trained on phase contrast images of migrating sheets of stem cells (a human iPSC clonal line) and ground truth nucleus data (Figure 1B) (Ling et al., 2020). We find that the model can identify the locations of cell nuclei from only phase contrast images of other mammary epithelial cell lines, such as the MCF10A epithelial cell line (Figure 1C), without requiring any training on these epithelial cell lines. This U-Net model was made robust to noise blurriness and intensity changes that one might encounter in live image acquisition by adding Gaussian noise, contrast and brightness change, blur, and scaling during the augmentation step in the training process after segmentation. Once segmented, the images are fed into a tracking algorithm (Crocker and Grier, 1996; Blair and DufresneCrocker-Grier), which identifies a center for each segmented nucleus and maps these positions from frame to frame into a cellular track, allowing us to follow individual cell nuclei over hundreds of frames (Figure 1D).

The trajectories of cell nuclei using this approach are validated by a comparison to the dynamics obtained in prior studies on the same data using particle image velocimetry (PIV) analysis (Lee et al., 2021a) as shown in Figure 1E. The two methods produce similar speed and direction of motion distributions, with our U-Net-based tracking exhibiting longer tails at higher velocities and larger angles. Bhattacharyya coefficients quantify the overlap between distributions and range between 0 (no overlap) and 1 (complete overlap). The distributions shown in Figure 1E have Bhattacharyya coefficients of 0.971 (speed) and 0.991 (direction), which indicates agreement between the PIV and U-Net results (Kailath, 1967). We note that the angles are computed relative to the direction of monolayer expansion. The small differences observed in the tails of the distributions agree with a previous comparison of PIV and tracking (Lee et al., 2013), which found that the motion captured by PIV is smoother than the motion of nuclei.

Our U-Net-based cell tracking approach from phase contrast images complements both PIV and traditional fluorescent nucleus tracking as illustrated in a comparison table (Figure 1F). PIV is widely used because it can be applied to label-free images, such as the phase contrast images studied in this current work. Fluorescent labeling of cell nuclei can provide additional information about individual cell behavior but often suffers from issues of phototoxicity (Phototoxicity revisited, 2018). Our label-free U-Net cell tracking approach yields the best of both worlds because it allows for nuclear tracking without interfering with long-term behavior or decreasing the survival rate of cells.



F

	Traditional Tracking	PIV	U-Net CG Tracking
Label	Fluorescence	Label-free	Label-free
Information Density	~ Cell Density	Fixed by Image Segment Size	~ Cell Density
Velocity	Per Cell	Per Image Segment	Per Cell
Reference Point	Image or Cell	Image	Image or Cell

Figure 1. AI-based segmentation and tracking workflow captures collective migration behavior

(A) A tiled image of a migrating cell sheet. The red box indicates the size of the field of view captured over time.

(B) An example of human iPSC clonal line images used as the training set for U-Net.

(C) Computational workflow: an input image is segmented using U-Net and then the centroids of the segmented cell nuclei are tracked using the Crocker-Grier algorithm.

(D) Tracks of 8 cells from t = 0 to frame 100 (3 min per frame, frames t = 0, 50, and 100 shown).

(E) Comparisons of tracking results with PIV results.

(F) Table comparing benefits of traditional tracking, PIV, and U-Net Crocker-Grier tracking.

A quantitative metric for cell rearrangements distinguishes tumorigenic and metastatic cell lines

With the ability to follow individual cells in a sheet, we can compute how cells move relative to their neighbors. One measure of collective rearrangements, D^2_{min} (Equation 1), has been used in models of elasto-plastic deformation of amorphous solids or granular materials (Utter and Behringer, 2008) to describe the motion of particles relative to their neighbors. Analyzing a cluster of N particles, D^2_{min} characterizes departures from the local affine deformation by fitting overall motion of the cluster over a time Δt to an affine deformation and reporting the residue.

$$D^2_{i,\,\min} = \frac{1}{N}\min\left\{\sum_{j=1}^{N}\left[\vec{d}_{ij}(t_1) - E_i\,\vec{d}_{ij}(t_0)\right]^2\right\} \qquad \text{(Equation 1)}$$

In Equation (1), $D^2_{i,\,min}$ is the D^2_{min} value for cell i at time $t_1 = t_0 + \Delta t$; N is the number of neighboring cells for cell i; $\vec{d}_{ij}(t_1)$ and $\vec{d}_{ij}(t_0)$ are the relative spacing between cell i and neighboring cell j at times

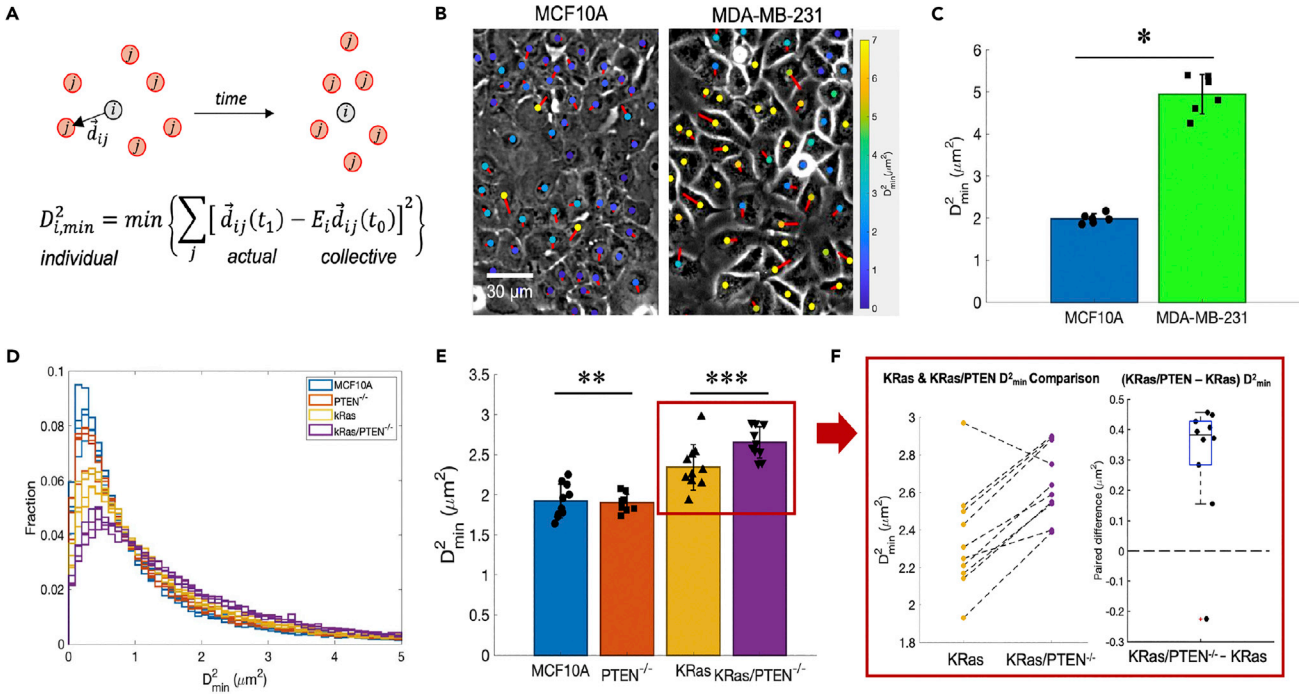


Figure 2. Cell rearrangements distinguish tumorigenic and metastatic cell lines

(A) A schematic of how D^2_{min} is calculated.

(B) Sheets of migrating cells labeled by their motion and rearrangements. Colorful dots represent D^2_{min} values for individual cells, with their values indicated by the color bar. Red arrows show the direction of cell motion.

(C) Comparison of average D^2_{min} values for MCF10A and MDA-MB-231 cells over 6 independent experiments.

(D) Comparison of D^2_{min} distribution of MCF10A mutants in one set of experiments (4 technical replicates of each mutant).

(E) Average D^2_{min} for each mutant over 10 independent experiments.

(F) Pairwise comparison of KRas and KRas/PTEN$^{-/-}$ D^2_{min}. (left) Dashed lines connect experiments performed on the same 12 well plate. (right) The central line indicates the median value of the paired difference, while the top and bottom edges of the box indicate the 75th and 25th percentiles.

Error bars in c, e represent the standard deviation across independent replicates. Data are represented as mean ± SD "*" = $p < 0.001$, "**" = $p = 0.78$, "***" = $p = 0.01$. Two-sample t-test.

t_1 and t_0, respectively; and E_i is the strain tensor for the neighbor cell group around cell i. We chose N = 6 under the assumption that epithelial cells have on average six neighbors in a confluent sheet. D^2_{min} (Chen et al., 2010; Falk and Langer, 1998) (Equation 1) is a least-squared calculation in which the final relative spacing between cells, $d_{ij}(t_1)$, is predicted via a linear transformation, E_i, of the initial relative spacing $d_{ij}(t_0)$, (Figure 2A). The resulting difference is called non-affine (non-linear) motion, whereas the $E_i d_{ij}(t_0)$ is the affine motion. E_i is fit to minimize the difference between the affine motion and the actual motion. The affine component, which reflects collective motion, is essentially how well the displacements of the neighboring cells can predict the motion of the central cell. A lower D^2_{min} value indicates the cell motion is more affine, meaning the cell more closely follows its neighbors' motion and is moving more collectively. A higher value of D^2_{min} thus implies a more individualistic or non-affine motion.

Using our proposed U-Net-based segmentation and tracking workflow, we calculate D^2_{min} for both MCF10A (non-tumorigenic) and MDA-MB-231 (metastatic) breast cells. This reveals strong differences in their rearrangement behavior. Figure 2B shows single frames from a time-lapse series in which cell nuclei are marked using colorful dots according to their respective D^2_{min} value. The non-tumorigenic MCF10A cells behave more collectively (lower D^2_{min} values) compared to the metastatic MDA-MB-231 cells (Figure 2C).

We also applied our U-Net cell tracking workflow, without additional training of the AI algorithm, on a genetically defined cancer progression model system based on the MCF10A cells (Thompson et al., 2015) (Video S1). In this model system, cells with PTEN deleted (PTEN$^{-/-}$) have previously been found to

remain dormant *in vivo*. KRas-activated cells also usually stay dormant *in vivo*; however, sometimes the KRas cells can form primary tumors (poorly tumorigenic) (Lee et al., 2021a; Thompson et al., 2015). The double mutant (KRas/PTEN$^{-/-}$) is aggressively tumorigenic (Thompson et al., 2015) and metastatic (Yankaskas et al., 2019). Figure 2D showcases the D^2_{min} distribution from four technical replicates for each cell line from one experiment. The D^2_{min} distribution for each cell line is consistent across technical replicates and the combined experiment shows the overall trend that the original MCF10A and PTEN$^{-/-}$ mutants have smaller D^2_{min} values, which means that their motion is more collective, and cells tend to follow their neighbors' motion. The overall results from 10 independent experiments (Figure 2E) confirm the poorly tumorigenic KRas and metastatic KRas/PTEN$^{-/-}$ mutants have higher D^2_{min} than the non-tumorigenic MCF10A and dormant PTEN$^{-/-}$ mutants. To further differentiate the KRas and KRas/PTEN$^{-/-}$ mutants, a pairwise comparison was performed (Figure 2F). A blocked experimental design allows us to pairwise compare the D^2_{min} value within each experiment, depicted by dashed line connecting the raw values (Figure 2F, left), and by directly subtracting the KRas D^2_{min} value from the KRas/PTEN$^{-/-}$ D^2_{min} value (Figure 2F, right). The results show that most experiments yield the metastatic KRas/PTEN$^{-/-}$ having higher D^2_{min} than the poorly tumorigenic KRas mutant. Therefore, the KRas/PTEN$^{-/-}$ mutant motion is more individual (non-affine) than the KRas mutant, and D^2_{min} can distinguish the poorly tumorigenic KRas and metastatic KRas/PTEN$^{-/-}$ cell lines.

Affine motion quantifies changes in collective behavior

To further study the collective component of cell motion, we extract affine dilation and deformation components from the minimized linear transformation matrix E_i (Equation 1) (Falk and Langer, 1998). We decompose the 2D matrix E_i into a product of two matrices: a pure rotation matrix $R(\theta)$ and a symmetric strain tensor F. The average of the eigenvalues of the strain tensor F corresponds to affine dilation, while the difference is the affine deformation. Physically, the two eigenvalues describe the contraction or expansion in the direction of their respective eigenvectors. Affine dilation represents the collective contraction or expansion of the group of 6 neighbors. Affine deformation describes whether the cells' neighbors are compressed or stretched. Schematics shown in Figure 3A illustrate examples of affine dilation and deformation in a 6-neighbor system.

We know from analysis of the nuclear tracks that the metastatic MDA-MB-231 cells have a larger average motion than the non-tumorigenic MCF10A cells (Figure S1). Affine calculations reveal that the MDA-MB-231 cells also display larger deformations and dilations (Figure 3B). Thus, the larger average total motion observed in the MDA-MB-231 cells has a collective component. The MCF10A mutants have similar affine dilation and deformation (Figure 3C). Combining data from all 10 experiments (Figure 3D), the magnitude of affine/collective motion among MCF10A mutants is similar.

Next, we compare the contribution of the two components of motion, affine (collective) and non-affine (individual), to the cell sheet's direction of motion. Angular distributions of both components (Figure 3E) show that the non-affine motion is roughly uniformly distributed in all directions, while the directions of affine motion favor the directional cue provided by empty space in the experimental system (roughly 90 degrees in the figures). This shows that the collective migration of cell sheets is mostly led by affine motion, where individual cells more closely follow their neighbors' motion. The non-tumorigenic MCF10A and dormant PTEN$^{-/-}$ cell lines show much more directional affine motion than the poorly tumorigenic and metastatic cell lines.

We calculate angular deviation for both affine and non-affine motion among all cell lines to quantify the spread of these distributions. Angular deviation values range from 0 to $\sqrt{2}$, with 0 meaning all cells are moving in one direction and higher values indicating a more dispersed set of migration directions. The non-affine motion of all cell lines has an angular deviation near $\sqrt{2}$, which confirms the observation that non-affine motion has no directional preference (Figure 3E). Meanwhile, the angular deviation differs for the affine (collective) motion across the different cell lines. The affine motion of the MCF10A cells is much more directional than the metastatic MDA-MB-231 cells (Figure 3F). Among the MCF10A model system, KRas and KRas/PTEN$^{-/-}$ have significantly higher affine motion angular deviations compared to MCF10A and PTEN$^{-/-}$, indicating the collective motions of the poorly tumorigenic (KRas) and metastatic (KRas/PTEN$^{-/-}$) cell lines are less directional (Figure 3G). A pairwise comparison to MCF10A shows that PTEN$^{-/-}$ cells have the lowest angular deviation (Figure 3H), indicating that PTEN deletion leads to more directional movement consistent with prior PIV analysis (Lee et al., 2013). We note that pairwise

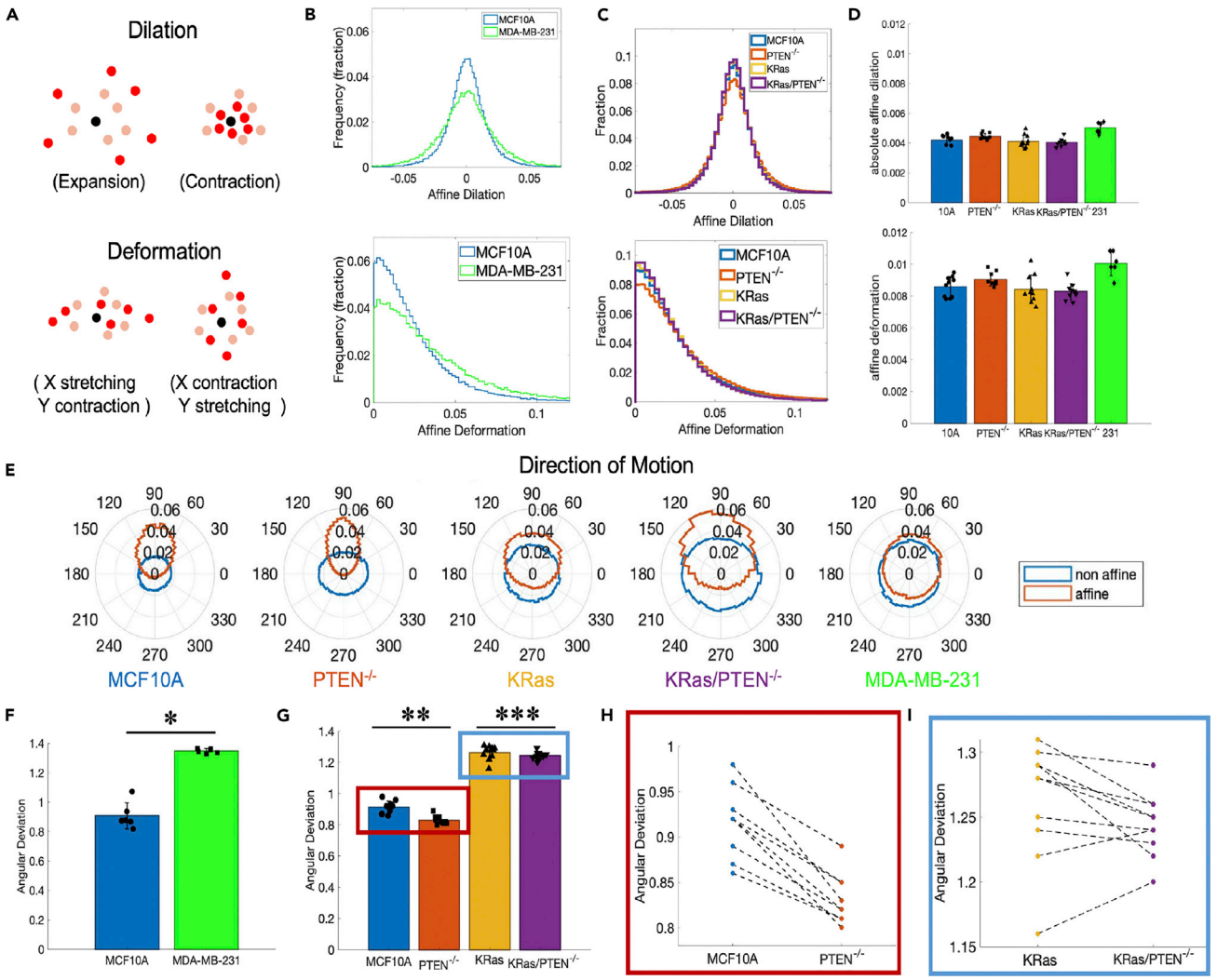


Figure 3. The affine component of motion describes collective behavior

(A) Schematics of affine dilation (top) and affine deformation (bottom), with pink dots indicating the original positions of nuclei and red dots representing their positions at a later time.

(B) Representative distributions of affine deformation and dilation comparing the MCF10A and MDA-MB-231 cells.

(C) Representative distributions of affine deformation and dilation comparison for the MCF10A mutants.

(D) The absolute value of affine dilation and average affine deformation for MCF10A mutants (10 independent experiments) and 231 cells (6 independent experiments).

(E) The direction of affine and non-affine motion for representative experiments for the MCF10A mutants and MDA-MB-231.

(F) MCF10A and MDA-MB-231 affine angular deviation.

(G) MCF10A mutants' affine angular deviation.

(H) Pairwise comparison of MCF10A and PTEN$^{-/-}$.

(I) Pairwise comparison of KRas and KRas/PTEN$^{-/-}$.

Error bars represent the standard deviation across independent replicates. Data are represented as mean ± SD. "*" = $p < 0.001$, "**" = $p < 0.001$, "***" = $p = 0.32$. Two-sample t test.

comparison of KRas and KRas/PTEN$^{-/-}$ mutants shows that PTEN deletion does not have a similar effect of increasing directionality of collective motion in KRas mutants (Figure 3I).

Quantification of cell rearrangements by D^2_{min} is robust to tracking errors

To test the robustness of our findings, 20% of the cells were randomly removed in each frame to simulate a cell segmentation error situation, which might occur with non-ideal imaging or segmentation

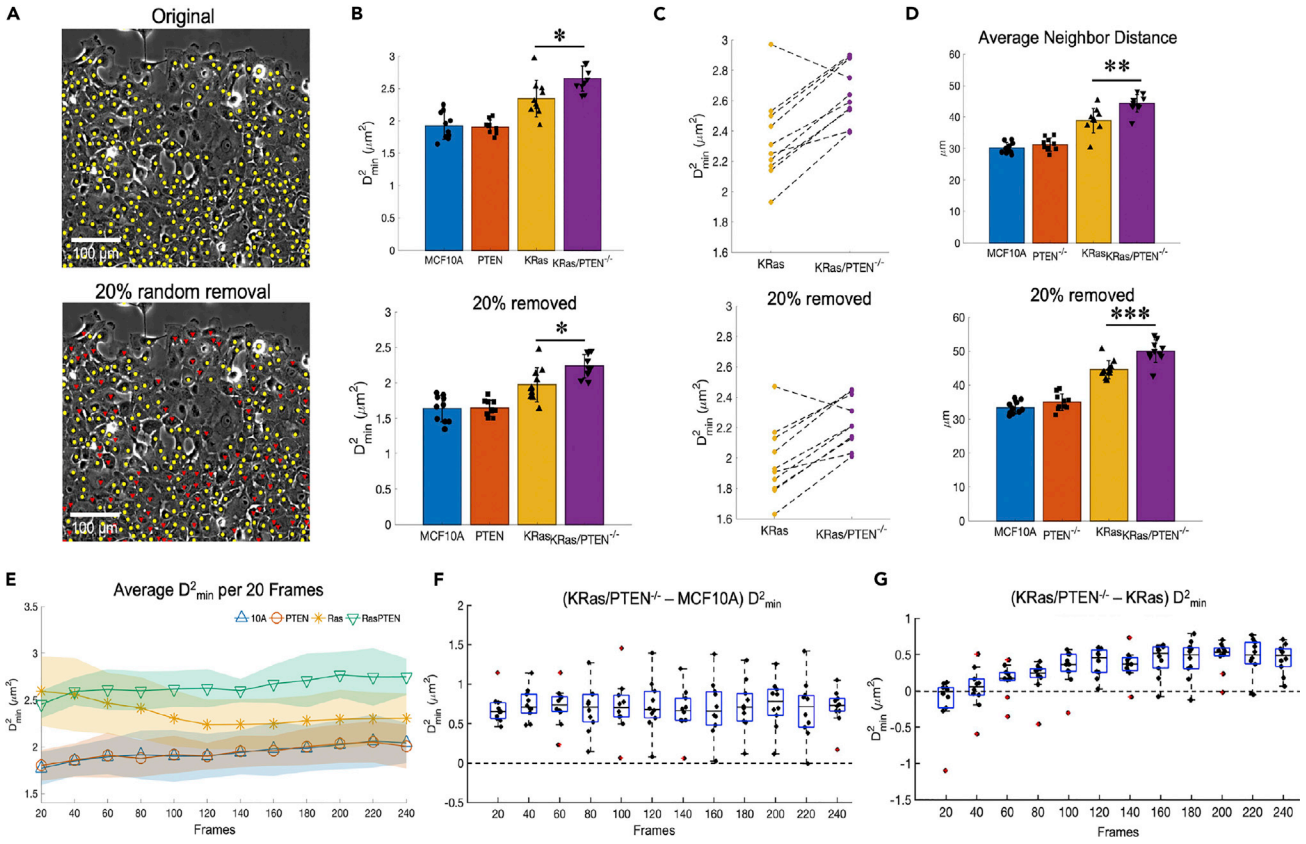


Figure 4. D^2_{min} is robust to potential tracking errors

(A) Top: a sheet of MCF10A cells, with all U-Net-segmented nuclei labeled in yellow; bottom: the same sheet of MCF10A cells, with 20% of nuclei randomly removed from tracking (indicated in red).

(B) Top: average D^2_{min} for 10 independent experiments from U-Net-segmented nuclei; bottom: average D^2_{min} with 20% of nuclei randomly removed in each frame.

(C) Pairwise D^2_{min} comparison of KRas and KRas/PTEN$^{-/-}$.

(D) Average neighbor distance.

(E) Average D^2_{min} per 20 frames (60 min) for the MCF10A mutants over time (shaded regions represent the standard deviation across independent replicates).

(F) Average D^2_{min} difference between KRas/PTEN$^{-/-}$ and original MCF10A per 20 frames (60 min) over time.

(G) Average D^2_{min} difference between KRas/PTEN$^{-/-}$ and KRas per 20 frames (60 min) over time. For boxplots, the central line indicates the median value, while the top and bottom edges of the box indicate the 75th and 25th percentiles.

Error bars in B, D represent the standard deviation across independent replicates. Data are represented as mean ± SD. Data are represented as mean ± SD. "*" = p = 0.01, "**" = p = 0.02, "***" = $p < 0.001$. Two-sample t test.

parameters. The U-Net applied here was trained on a different cell line, without any retraining. We manually segmented a set of images and found a segmentation accuracy of at least 80% with respect to detecting all cell nuclei in the sample (Figure S2). As such, the removal of 20% of the tracked cells is a good simulation to test the measurement robustness of our pipeline to expected segmentation errors. Figure 4A shows the original data, with each cell nucleus (segmented by U-Net) labeled in yellow and a frame where 20% of the identified nuclei are labeled for removal (red). Our D^2_{min} analysis is still able to distinguish between the KRas and KRas/PTEN$^{-/-}$ cell lines, as shown in Figures 4B and 4C. With the removal of 20% of cells, the D^2_{min} of these four cell lines maintain the same overall trends, including KRas/PTEN$^{-/-}$ consistently exhibiting higher D^2_{min} than the cell line mutated in KRas alone. The average neighbor distance (Figure 4D) slightly increases and is accompanied by a slightly smaller D^2_{min} for all cell lines with 20% cell removal. In other words, the decrease of cell density from artificially removing cells leads to lower D^2_{min}. So, the larger neighbor distance seen in KRas/PTEN$^{-/-}$ cells cannot explain the higher D^2_{min} found in these cells. This means that high D^2_{min} for KRas/PTEN$^{-/-}$ reflects

individualistic motion. Neighbor distances, or cell density might be impacted by tracking accuracy; however, Figures 4A and 4B show that the D^2_{min} metric can distinguish poorly tumorigenic from metastatic cell lines even when tracking accuracy is artificially lowered.

Next, we investigate the change in motion with time over the 12 h timescale of our experiments. We segmented the data into 20-frame pieces and found the average D^2_{min} over each time segment, as shown in Figure 4E. The shaded regions in the plot are the standard deviations for D^2_{min} values over 10 experiments. The poorly tumorigenic KRas and metastatic KRas/PTEN$^{-/-}$ have consistently higher D^2_{min} compared to the non-tumorigenic MCF10A and dormant PTEN$^{-/-}$ cells at all timepoints. The KRas cells' D^2_{min} value decreases over the first 120 frames, and then stabilizes; the KRas/PTEN$^{-/-}$ mutant D^2_{min} values slightly increase over time. These two time-related behaviors lead to the KRas/PTEN$^{-/-}$ cells having a higher time-averaged D^2_{min} value, as shown in Figures 2E and 2F. Over the 12 h of imaging (240 frames), the metastatic KRas/PTEN$^{-/-}$ consistently has higher D^2_{min} than the non-tumorigenic MCF10A (Figure 4F), indicating that D^2_{min} is a robust comparison between non-tumorigenic and metastatic cells. Comparing the KRas/PTEN$^{-/-}$ and KRas cells (Figure 4G), we find that their motion is initially similar, but differences emerge 120 frames (6 h) after the start of imaging, or 30 h after cell plating. At later times, 20 frames (60 min) of imaging are sufficient for D^2_{min} to distinguish the KRas/PTEN$^{-/-}$ and KRas cell lines.

DISCUSSION

We present an AI-enabled label-free nuclear tracking method capable of identifying the motion of cell nuclei and use this information to quantify the collective behavior of panels of cell lines. We study the MCF10A epithelial cell line and its variants with cancer-related mutations, specifically contrasting KRas activation alone and in conjunction with PTEN deletion. These two oncogenic mutations yield different outcomes *in vivo* but their associated motion characteristics have been indistinguishable when analyzed with prior methods (Lee et al., 2021a; Yankaskas et al., 2019).

Our nuclear segmentation is done by an AI model previously trained on a human iPSC clonal line (Ling et al., 2020). The parameters of the tracking algorithm maintain a temporary memory of cells that disappear in a frame or two due to imaging or segmentation issues and remove tracks that are too short, improving successful cell tracking and reducing noise. With these quality-control features, we can segment and track on average 80% of the nuclei in an image for MCF10A and MDA-MB-231 cell lines without further training on U-Net (Figure S2).

The successful application of the AI model on these two cell lines suggests that our AI model and tracking algorithm are versatile and can be applied to other epithelial cell lines without an extra training set. Moreover, random removal of 20% of the nuclei yielded consistent D^2_{min} results (Figure 4), suggesting that we are tracking a large enough fraction of cells for robust calculations not prone to change with tracking errors. Being versatile and robust to errors, we can use our workflow to track and analyze the collective motion for a variety of different types of cells.

Based on measuring the individual motion of cells, D^2_{min} is capable of clearly distinguishing non-tumorigenic and tumorigenic cells and is also able to distinguish the poorly tumorigenic KRas and the metastatic KRas/PTEN$^{-/-}$ cells (Figure 2F). The D^2_{min} calculation (Equation 1), provides multiple metrics for understanding collective migration behavior. In addition to D^2_{min}, which measures the balance of collective vs individual behavior, the calculation quantifies "affine" motion characteristics, namely, the magnitudes and preferred directions of dilations and deformations. Both our previous PIV and current U-Net tracking results show that PTEN deletion makes cells more collective and more directional (Figure 3), while KRas activation makes cells less collective (Figures 2 and 4). However, D^2_{min} highlights the differences between the poorly tumorigenic KRas and metastatic KRas/PTEN$^{-/-}$ cell lines (Figure 2), despite similarities in their collective behavior (Figure 3). The metastatic cell line (KRas/PTEN$^{-/-}$) has the most individual motion, which leads to more local rearrangements in the sheet, impacting the overall collective migration of the cell sheet. This individual motion may be the result of a loss in cell adhesion in tumor progression (Cavallaro and Christofori, 2004), favoring tumor invasion and metastasis. Our analysis of collective and individual motions has the potential to be applied to distinguish non-tumorigenic, tumorigenic, and metastatic cells for easier and faster diagnosis.

In summary, we introduce an AI-enabled label-free nucleus tracking workflow for phase-contrast imaging that can follow the motion of cells without the need for new AI training data. The comprehensive assessment of cell motion is critical in enabling us to characterize collective motion using D^2_{min}, which provides a local metric of collective motion that shows differences between metastatic and non-metastatic cell lines. Our results indicate that AI-enabled collective motion measurements may provide a new perspective on the identification and diagnosis of metastatic risk.

NIST disclaimer

Commercial products are identified in this document in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose.

Limitations of the study

Comparing the AI-based nucleus finding algorithm and tracking algorithm together to a traditional method is unfortunately not possible in our system as fluorescent tracking can affect cell behavior, in particular the long-term behavior of cells after repeated light exposure.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell migration dataset
- METHOD DETAILS
  - Image and migration analysis
- D^2_{MIN} AND AFFINE ANALYSIS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104678.

ACKNOWLEDGMENTS

This work was supported by NIH grant R01-CA154624, AFOSR grant FA9550-16-1-0052, American Cancer Society Research Scholar Grant RSG-18-028-01-CSM.

AUTHOR CONTRIBUTIONS

Conceptualization, S.G., R.M.L., and W.L.; Methodology, S.G., R.M.L., and W.L.; Software, S.G., R.M.L., C.L., and J.C.; Formal Analysis, S.G. and R.M.L.; Resources, R.M.L., Z.B., and J.C.; Writing - Original Draft, S.G. and R.M.L.; Writing – Reviewing & Editing, S.G., R.M.L., Z.B., M.I.V., S.S.M., J.C., and W.L.; Supervision, W.L.; Funding Acquisition, W.L.

DECLARATION OF INTERESTS

The $PTEN^{-/-}$ cells are licensed by Horizon Discovery Ltd. (Cambridge, UK). Dr. Vitolo receives compensation from the sale of these cells.

Received: January 1, 2022
Revised: March 28, 2022
Accepted: June 23, 2022
Published: July 15, 2022

REFERENCES

Aceto, N., Bardia, A., Miyamoto, D., Donaldson, M., Wittner, B., Spencer, J., Joel Yu, M., Pely, A., Engstrom, A., Zhu, H., et al. (2014). Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell *158*, 1110–1122. https://doi.org/10.1016/j.cell.2014.07.013.

Blair, D.L., and Dufresne, E.R.. Crocker-grier particle tracking algorithm. http://site.physics.georgetown.edu/matlab/.

Cavallaro, U., and Christofori, G. (2004). Multitasking in tumor progression: signaling functions of cell adhesion molecules. Ann. N. Y. Acad. Sci. *1014*, 58–66. https://doi.org/10.1196/annals.1294.006.

Chen, D., Semwogerere, D., Sato, J., Breedveld, V., and Weeks, E.R. (2010). Microscopic structural relaxation in a sheared supercooled colloidal liquid. Phys. Rev. *81*, 011403. https://doi.org/10.1103/physreve.81.011403.

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., and Ewald, A.J. (2016). Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. Proc. Natl. Acad. Sci. USA *113*, E854–E863. https://doi.org/10.1073/pnas.1508541113.

Crocker, J.C., and Grier, D.G. (1996). Methods of digital Video microscopy for colloidal studies. J. Colloid Interface Sci. *179*, 298–310. https://doi.org/10.1006/jcis.1996.0217.

Desai, R.A., Gao, L., Raghavan, S., Liu, W.F., and Chen, C.S. (2009). Cell polarity triggered by cell-cell adhesion via E-cadherin. J. Cell Sci. *122*, 905–911. https://doi.org/10.1242/jcs.028183.

Falk, M.L., and Langer, J.S. (1998). Dynamics of viscoplastic deformation in amorphous solids. Phys. Rev. *57*, 7192–7205. https://doi.org/10.1103/physreve.57.7192.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer *136*, E359–E386. https://doi.org/10.1002/ijc.29210.

Hagemeister, F.B., Buzdar, A.U., Luna, M.A., and Blumenschein, G.R. (1980). Causes of death in breast cancer a clinicopathologic study. Cancer *46*, 162–167. https://doi.org/10.1002/1097-0142(19800701)46:1<162::aid-cncr2820460127>3.0.co;2-b.

Herbst, R.S., and Schlessinger, J. (2019). Small molecule combats cancer-causing KRAS protein at last. Nature *575*, 294–295. https://doi.org/10.1038/d41586-019-03242-8.

Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. Commun. *15*, 52–60. https://doi.org/10.1109/tcom.1967.1089532.

Klein, C.A. (2009). Parallel progression of primary tumours and metastases. Nat. Rev. Cancer *9*, 302–312. https://doi.org/10.1038/nrc2627.

Lee, R.M., Kelley, D.H., Nordstrom, K.N., Ouellette, N.T., and Losert, W. (2013). Quantifying stretching and rearrangement in epithelial sheet migration. New J. Phys. *15*, 025036. https://doi.org/10.1088/1367-2630/15/2/025036.

Lee, R.M., Vitolo, M.I., Losert, W., and Martin, S.S. (2021a). Distinct roles of tumor associated mutations in collective cell migration. Sci. Rep. *11*, 10291. https://doi.org/10.1038/s41598-021-89130-6.

Lee, R.M., Vitolo, M.I., Losert, W., and Martin, S.S. (2021b). idr0111-lee-cellmigration. https://doi.org/10.17867/10000163.

Ling, C., Halter, M., Plant, A., Majurski, M., Stinson, J., and Chalfoun, J. (2020). Analyzing U-net robustness for single cell nucleus segmentation from phase contrast images (IEEE), pp. 4157–4163.

López-Knowles, E., O'Toole, S.A., Mcneil, C.M., Millar, E.K.A., Qiu, M.R., Crea, P., Daly, R.J., Musgrove, E.A., and Sutherland, R.L. (2010). PI3K pathway activation in breast cancer is associated with the basal-like phenotype and cancer-specific mortality. Int. J. Cancer *126*, 1121–1131. https://doi.org/10.1002/ijc.24831.

Phototoxicity revisited (2018). Nat. Methods *15*, 751. https://doi.org/10.1038/s41592-018-0170-4.

Renkawitz, J., Kopf, A., Stopp, J., de Vries, I., Driscoll, M.K., Merrin, J., Hauschild, R., Welf, E.S., Danuser, G., Fiolka, R., and Sixt, M. (2019). Nuclear positioning facilitates amoeboid migration along the path of least resistance. Nature *568*, 546–550. https://doi.org/10.1038/s41586-019-1087-5.

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafinia, S., et al. (2018). Oncogenic signaling pathways in the cancer genome atlas. Cell *173*, 321–337.e10. https://doi.org/10.1016/j.cell.2018.03.035.

Thompson, K.N., Whipple, R.A., Yoon, J.R., Lipsky, M., Charpentier, M.S., Boggs, A.E., Chakrabarti, K.R., Bhandary, L., Hessler, L.K., Martin, S.S., and Vitolo, M.I. (2015). The combinatorial activation of the PI3K and Ras/MAPK pathways is sufficient for aggressive tumor formation, while individual pathway activation supports cell persistence. Oncotarget *6*, 35231–35246. https://doi.org/10.18632/oncotarget.6159.

Utter, B., and Behringer, R.P. (2008). Experimental measures of affine and nonaffine deformation in granular shear. Phys. Rev. Lett. *100*, 208302. https://doi.org/10.1103/physrevlett.100.208302.

Wang, X., Cao, X., Sun, R., Tang, C., Tzankov, A., Zhang, J., Manyam, G.C., Xiao, M., Miao, Y., Jabbar, K., et al. (2018). Clinical significance of PTEN deletion, mutation, and loss of PTEN expression in de novo diffuse large B-cell lymphoma. Neoplasia *20*, 574–593. https://doi.org/10.1016/j.neo.2018.03.002.

Weiger, M.C., Vedham, V., Stuelten, C.H., Shou, K., Herrera, M., Sato, M., Losert, W., and Parent, C.A. (2013). Real-time motion analysis reveals cell directionality as an indicator of breast cancer progression. PLoS One *8*, e58859. https://doi.org/10.1371/journal.pone.0058859.

Williams, E., Moore, J., Li, S.W., Rustici, G., Tarkowska, A., Chessel, A., Leo, S., Antal, B., Ferguson, R.K., Sarkans, U., et al. (2017). Image Data Resource: a bioimage data integration and publication platform. Nat. Methods *14*, 775–781. https://doi.org/10.1038/nmeth.4326.

Wrenn, E.D., Yamamoto, A., Moore, B.M., Huang, Y., Mcbirney, M., Thomas, A.J., Greenwood, E., Rabena, Y.F., Rahbar, H., Partridge, S.C., and Cheung, K.J. (2020). Regulation of collective metastasis by nanolumenal signaling. Cell *183*, 395–410.e19. https://doi.org/10.1016/j.cell.2020.08.045.

Yankaskas, C.L., Thompson, K.N., Paul, C.D., Vitolo, M.I., Mistriotis, P., Mahendra, A., Bajpai, V.K., Shea, D.J., Manto, K.M., Chai, A.C., et al. (2019). A microfluidic assay for the quantification of the metastatic propensity of breast cancer specimens. Nat. Biomed. Eng. *3*, 452–465. https://doi.org/10.1038/s41551-019-0400-9.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Cell migration imaging dataset	Lee et al., 2021b	Imaging Data Resource: https://doi.org/10.17867/10000163a
Software and algorithms		
Crocker-Grier Tracking Algorithm	Daniel Blair and Eric Dufresne	http://site.physics.georgetown.edu/matlab/
ImageJ	Schneider et al. (2012)	https://imagej.nih.gov/ij/
MATLAB	The MathWorks, Inc	https://www.mathworks.com/?s_tid=gn_logo
Cell Counter plugin for ImageJ	Kurt De Vos	https://imagej.nih.gov/ij/plugins/cell-counter.html
U-Net model	Ling et al. (2020)	N/A
Trained U-Net and D^2_{min} analysis code	This paper	https://github.com/losertlab/UNet-D2min

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Wolfgang Losert (wlosert@umd.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- MCF10A and MDA-MB-231 data underlying the figures are available in Data S1. The images analyzed in this study are available in the Image Data Resource (Williams et al., 2017). DOIs are listed in the key resources table.
- The code used for analysis in this study has been deposited at GitHub and will be made publicly available upon acceptance for publication. DOIs for the analysis code are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell migration dataset

No cell culture or imaging experiments were conducted for this study. The images of cell migration used in this work are a freely available dataset in the Image Data Resource (Lee et al., 2021b). This previously published dataset (Lee et al., 2021a) was collected on the MCF10A human breast epithelial cell line, MCF10A PTEN$^{-/-}$, MCF10A KRas(G12V), MCF10A KRas(G12V)/PTEN$^{-/-}$ (Thompson et al., 2015), and the MDA-MB-231 breast cancer cell line. MCF10A cells were maintained in DMEM/F12 medium supplemented with 5% horse serum, 10 μg/mL insulin, 0.5 μg/mL hydrocortisone, and 100 ng/mL cholera toxin. For comparisons between the MCF10A, MCF10A PTEN$^{-/-}$, MCF10A KRas(G12V), and MCF10A KRas(G12V)/PTEN$^{-/-}$ cells, the media was additionally supplemented with 0.5 μg/mL puromycin to select for LifeAct-GFP. MDA-MB-231 cells were maintained in DMEM supplemented with 10% FBS.

For this published dataset, cells were plated in 12 well glass-bottom plates using a complete block design, which allows for pairwise comparisons between cell sheets imaged on the same plate. Each well was coated with 3.25 μg/cm^2 collagen IV by placing the plate on ice and leaving 500 mL of collagen IV solution (3.25 μg/cm^2 collagen IV in 0.05 M HCl) on each well for 1 h before rinsing twice with water. A 5 μL drop of 1.5 × 10^6 cells/mL was plated in the center of each well, allowed to adhere for approximately 45 min at 37°C, and then rinsed to remove non-adherent cells. This created a circular cell sheet, which was then allowed to adhere overnight before imaging, followed by a medium change 1 h before the start of imaging. An initial snapshot

of the entire cell sheet was recorded by tiling multiple positions before focusing on ROIs for time lapse imaging. Cells were then imaged every 3 min for 12 h using a 10× phase-contrast objective (Nikon CFI Plan Fluor, NA 0.3) on a PerkinElmer Ultra-View VoX spinning-disk confocal with a Tokai Hit incubator chamber (37°C and 5% CO2) and a motorized stage for acquiring multiple positions over time. Images were acquired using a Hamamatsu C10600-10B ORCA-R2 camera (12 bit images 1024 pixels × 1344 pixels, 0.582 μm/pixel) using PerkinElmer's Volocity software (version 6.4.0). The initial analysis of these images is described in (Lee et al., 2021a) and the image files are available in the Image Data Resource (Williams et al., 2017) under DOI [https://doi.org/10.17867/10000163a] (Lee et al., 2021b).

METHOD DETAILS

Image and migration analysis

Cell nuclear segmentation from phase contrast images was performed by a pre-trained U-Net model (Ling et al., 2020), which generated labeled binary images of inferred connected objects representing nuclei. We used a custom MATLAB code to filter out connected objects that are significantly smaller than the average physiological nucleus size; this step filters out imaging and segmentation noise. The centroid of each nucleus was found using the regionprops function in MATLAB. The Crocker-Grier algorithm (Crocker and Grier, 1996), as implemented by Blair & Dufresne for MATLAB (Blair and Dufresne), performs nuclear tracking on the centroids and outputs the tracks of each nucleus for each frame. A custom MATLAB neighbor-finding code (Lee et al., 2013) was used to find the closest 6 neighbors for each cell nucleus in each frame. This output was then used to calculate D^2_{min} and the affine components of motion as described in the section below.

Angular deviation, s, was calculated as $s = \sqrt{2(1-R)}$, where each tracking vector was transformed into a unit vector in the two-dimensional plane, $r_i = \begin{pmatrix} \cos\theta_i \\ sin\theta_i \end{pmatrix}$, and R is the length of the vector average of the direction vectors: $R = \|\frac{1}{N}\sum_i r_i\|$.

D^2_{MIN} AND AFFINE ANALYSIS

$$D^2_{i,\,min} = \frac{1}{6}\min\left\{\sum_{j=1}^{6}\left[\vec{d}_{ij}(t_1) - E_i\,\vec{d}_{ij}(t_0)\right]^2\right\} \qquad \text{(Equation 1)}$$

D^2_{min} is calculated using Equation (1) by minimizing the difference between the final relative position of a cell and its neighbors, $\vec{d}_{ij}(t_1)$, and the affine component, which is a linear transformation of the initial relative position, $E_i\,\vec{d}_{ij}(t_0)$.

The 2 × 2 linear transformation matrix E that is fit in the D^2_{min} equation was decomposed to further describe the affine components of motion. We rewrite the matrix as a product of a pure rotation $R(\theta)$ and a symmetric deformation matrix F: $E = R(\theta)F$ which we can expand as shown in Equation (2). The angle of rotation, θ, can then be found by substituting the elements of E into Equation (3).

$$E = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}\begin{bmatrix} f_1 & 0 \\ 0 & f_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta)f_1 & \sin(\theta)f_2 \\ -\sin(\theta)f_1 & \cos(\theta)f_2 \end{bmatrix} = \begin{bmatrix} e_1 & e_2 \\ e_3 & e_4 \end{bmatrix} \qquad \text{(Equation 2)}$$

$$\theta = tan^{-1}\left(\frac{e_2 - e_3}{e_1 + e_4}\right) = tan^{-1}\left(\frac{\sin(\theta)f_2 - (-\sin(\theta)f_1)}{\cos(\theta)f_1 + \cos(\theta)f_2}\right) = tan^{-1}\left(\frac{\sin(\theta)(f_1+f_2)}{\cos(\theta)(f_1+f_2)}\right) = tan^{-1}\left(\frac{\sin(\theta)}{\cos(\theta)}\right) \qquad \text{(Equation 3)}$$

The deformation matrix F can then be found using $F = R^{-1}(\theta)E$. We further define a strain tensor $\varepsilon = F - I$. The subtraction of the identity matrix from F ensures that its eigenvalues would be 0 if no deformation, positive if stretching, and negative if compressing. The affine deformation is calculated as the difference between the two eigenvalues of matrix F, and affine dilation is the average of the eigenvalues. Examples of θ, $R(\theta)$, F, and ε are given in Figure S3.

For instances of skipped frames, we keep a memory of disappearing cells in the tracking algorithm, however, metrics like velocity and D^2_{min} are calculated for frames of the track where the previous and later frames are not missing data.

The D^2_{min} calculation is done in the reference frame of an individual cell. Therefore, to extract the affine and non-affine motion components of the lab frame motion, we performed a change of reference frame. For example, we calculated the D^2_{min} value for a cell with 6 neighbors from a time t_1 to t_2. Then the expected difference in motion between cell i and neighbors j at t_2 is $E_i\ \vec{d}_{ij}(t_1)$, where $\vec{d}_{ij}(t_1)$ is the difference motion at time t_1: $\vec{d}_{ij}(t_1) = r_i(t_1) - r_j(t_1)$. Thus, the expected (affine) neighbor position in the lab frame at time t_2 is given by Equation 4.

$$r_{j,affine}(t_2) = E_i\ \vec{d}_{ij}(t_1) + \vec{d}_{ij}(t_1) + r_i(t_2) = E_i\ \vec{d}_{ij}(t_1) + r_j(t_1) + (r_i(t_2) - r_i(t_1)) \quad \text{(Equation 4)}$$

Equation (4) is the neighbor location in the lab frame plus the cell displacement over time plus the expected motion difference between cell and neighbors. We then can find the difference between the expected and actual positions for neighbors in the lab frame. Then, taking the average of over 6 neighbors, we get the non-affine motion of the cell compared to its neighbors. To get the affine component of the motion, we simply subtract the calculated non-affine motion from the actual (total) motion.

QUANTIFICATION AND STATISTICAL ANALYSIS

For each metric, means shown in the respective figures were calculated from N = 10 independent experiments for the MCF10A genetically defined model system and N = 6 for comparisons to the MDA-MB-231 cells. Error bars represent the standard deviation across independent replicates unless otherwise noted. In boxplot figures, the central line indicates the median value, while the top and bottom edges of the box indicate the 75^{th} and 25^{th} percentiles. Outliers on the boxplots (red plus symbols) were defined as values greater than 1.5 times the interquartile range (IQR indicated by the boxplot whiskers). Statistical signification test presented as asterisks are calculated using two-sample t test. All the calculations are done using MATLAB. Underlying data for quantitative figures is presented in tabular format in Data S1.