PNAS
www.pnas.org

Supplementary Information for

Cell division Rate Controls Cell Shape Remodeling in Epithelia

John Devany, Daniel M. Sussman, Takaki Yamamoto, M. Lisa Manning, Margaret L. Gardel*

*Correspondence should be addressed to: gardel@uchicago.edu

This PDF file includes:

Supplemental Information Text	p2-11
Figures S1 to S16	p12-27
Legends to movies S1 to S8	p28
SI References	p29-30

Other supplementary materials for this manuscript include the following:

Movies S1 to S8

Supplemental Information Text:

Methods:

Reagents: PND1184, Nocodazole, y27632, NSC23766, Mitomycin C, Human Transferrin, (3-Aminopropyl)trimethoxysilane were purchased from Sigma-Aldrich (Saint Louis, MO), Glutaraldehyde purchased from Electron Microscopy Sciences (Hatfield, PA), BD Collagen I, rat tail was purchased from BD Biosciences (San Jose, CA). 1X PBS, 1X DMEM, Fetal Bovine Serum, l-glutamine, Penicillin-Streptomycin, Trypsin EDTA were purchased from Corning Inc. (Tewksbury, MA), TBS, MnCl, NaOH were purchased from Fisher Scientific (Hampton, NH), Ro-3306 was purchased from Cayman Chemical (Ann Arbor,MI)

Cell culture: Madin-Darby Canine Kidney (MDCK) cells and Mouse Embryonic Fibroblasts (MEFs) were cultured in high-glucose DMEM supplemented with 10% FBS, 2mM L-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin at 37C and 5% CO2. Caco-2 cells were cultured in high-glucose DMEM supplemented with 10% FBS, 2mM L-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin and Human Transferin at 37C and 5% CO2. Cells were passaged using 0.25% trypsin EDTA every 2-3 days. Cells were checked for mycoplasma by Hoechst staining.

Stargazin-GFP MDCK cells we produced by transient transfection of WT MDCK cells with a PiggyBac-stargazin-GFP construct followed by selection by puromycin and subcloning. A clone with high expression of the marker and similar morphology to WT MDCK cells was selected for experiments. Stargazin-GFP was a gift from Michael Glotzer. Stargazin-halotag Caco-2 and MDCK cells were produced by lentiviral infection of WT CACO-2 and MDCK cells by a WPT-Stargazin-halotag construct packaged in 293T cells by a second generation lentiviral system with rev8.2 and VSVG. Viral supernatant was collected at 24, 48 and 72 hours then concentrated ~30x using Amicon Ultra-15 Centrifugal Filter Unit (100kDa). Cells were treated with virus and 8µg/ml polybrene overnight. Positive cells were isolated using a cell sorter. PIP-FUCCI MDCK cells were produced by lentiviral infection with virus packaged the same way. Cells were then selected using 800 µg/ml G418. pLenti-PGK-Neo-PIP-FUCCI was a gift from Jean Cook (Addgene plasmid # 118616 ; http://n2t.net/addgene:118616 ; RRID:Addgene_118616). Tet P27 1-176 cells were produced using the Lenti-X Tet-On 3G (Takara Bio). Human Snaptag-P27 1-176 was subcloned into the Tre3g vector. Cells were infected with lentivirus for both the Tet-on 3g and Tre3g snaptag-p27 1-176 plasmids then selected using 2 µg/ml puromycin and 800 µg/ml G418.

AminoSilane Glutaraldehyde modification of glass coverslips: Glass coverslips were modified as previously described to couple collagen gels to the surface of the glass (1). Coverslips were first cleaned by sonication in 70% and 100% ethanol solutions then dried with compressed air. We placed coverslips in a staining rack and submerged the rack in a solution of 2% (3-Aminopropyl)trimethoxysilane (APTMS) 93% propanol and 5% DI water for 10 minutes at room temperature while stirring. Staining racks were removed and washed in DI water 5 times then placed in a 37C incubator for 6 hours to allow the water to dry and aminosilane layer to cure. The staining racks were then submerged in 1% glutaraldehyde in DI water for 30 minutes while stirring. Then the samples were washed 3 times for 10 minutes in distilled water, air dried and stored at room temperature. Activated coverslips were used within 2 months of preparation.

Collagen gel preparation: 10x PBS, milli-q filtered water, a 5mg/ml collagen stock and 1N NaOH were mixed to generate a polymerization mix with 1xPBS and 2 or 4 mg/ml collagen at neutral pH. To visualize

collagen gel thickness we produced a fluorescently labeled collagen stock by mixing collagen with alexa647-NHS ester in 0.02M acetic acid overnight. We added fluorescently labeled collagen at a 20:1 ratio to unlabeled collagen in the polymerization mix. 70uL of the polymerization mix was added on to a 25mm round modified coverslip and quickly spread to coat the surface using a pipette tip. Samples were transferred to a humidified incubator at 37C to polymerize for 1 hour. After polymerization gels were washed 3 times in 1x PBS and it was verified that gels were still intact and adhered to the glass by a tissue culture microscope.

Glutaraldehyde crosslinked gels were prepared as above and crosslinked as previously described (2). Directly after polymerization and washing gels were incubated in 1xPBS containing 0.2% glutaraldehyde for 30 minutes. Gels were then washed quickly 3 times in 1xTBS, washed in 1x TBS at 1 hour intervals 5 times and left in 1x TBS overnight to quench excess crosslinking groups on the gel. The gels were then washed in 1x PBS three times. All gels were used within 2 days of polymerization.

Monolayer preparation: Monolayers were formed on collagen I gel to study monolayers under more physiologically relevant conditions than typical glass, plastic or hydrogel surfaces (3–7). Cells were seeded onto collagen gels at high density (~600,000 cells on 500mm^2 surface) such that cells coated ~70% of the gel surface at seeding. The sample was returned to the incubator overnight (8-12 hours) before the experiment. For inhibitor treated conditions the inhibitor was added 10-20 minutes after the cells were seeded on the gel to allow cells to attach before adding fluid volume. Before mounting samples, the monolayer was viewed under a tissue culture microscope and continuously covered a region 10-15 mm across. Samples were washed 3 times with 1x PBS and quickly mounted into a sealed round chamber with 1.5mL of culture media and with equal concentration of inhibitor to at seeding. Before taking each time lapse the collagen gel was verified by fluorescence microscopy to be homogeneous and 150-250 µm thick. The monolayer was confirmed to extend at least several hundred microns in each direction outside selected fields of view. By only observing cells more than a few hundred microns from a free edge we avoid the increased migration speed and correlation from cytoskeletal assemblies specific to wound healing (8, 9).

Fluorescence microscopy: Cells were imaged on an inverted epi-fluorescence microscope (Nikon TI-E, Nikon, Tokyo, Japan) with a 20x plan flour multi-immersion objective. Glycerol was used as an immersion medium to more closely match the index of refraction of the collagen gel. Images were acquired at 10 minute intervals in GFP, 642 and transmitted light channels using standard filter sets (Ex 490/30, Em 525/30, Ex 640/30, DAPI/FITC/TRITC/cy5 cube) (Chroma Technology, Bellows Falls, VT). Samples were mounted on the microscope in a humidified stage top incubator maintained at 37C 5% CO2. Images were acquired on either a Photometrics Coolsnap HQv2 CCD camera (Photometrics, Tucson, AZ) or Andor Zyla 4.2 CMOS camera (Andor Technology, Belfast, UK).

Image Segmentation: Images were segmented using custom MATLAB code. The main algorithm performs initial segmentation using the Phase Stretch Transform algorithm developed by the Asghari and Jalali (10). Phase stretch images were thresholded and skeletonized to obtain cell outlines. Broken edges in the skeleton were repaired using a modified implementation of edgelink developed by Peter Kovesi (11). After edge-linking the remaining unclosed portions of the path were removed. The interior of each cell is checked for high intensity features which typically indicate under-segmentation. Any region containing a high intensity region within the boundary are discarded. The above algorithm has 4 parameters describing the PST parameter set and the interior threshold. Parameters were first optimized by hand. We then randomly generated 1000 parameter sets around these values. We generated a rough ground truth segmentation by averaging over several parameter sets which we verified to segment the images well. The 1000 parameter sets were checked against this ground truth to choose a final parameter set. This parameter

set was used to segment all images analyzed in this paper. Examples of the segmented outlines overlaid on the fluorescence images are found in the supplement (Movie S8).

Cell tracking: Cell tracking was performed using established particle tracking methods (12). Cell centers were determined by taking the centroid of each region in the cell outlines generated as described above in Image Segmentation. The particle trajectories were compiled from these position measurements using SimpleTracker, a MATLAB function developed by Jean-Yves Tinevez (13). For each image in a given time series the average displacement was determined and was subtracted from the cell positions to account for stage drift between each frame. We determined a lower bound on the tracking error by tracking image series of fixed cells (Fig. S15)

For each movie the cell position, shape, magnitude of displacement at a 10-minute interval, and cell area were determined for each individual cell over all time points. The average magnitude of displacement was calculated for each field of view at each time point to obtain a cell speed. The inverse of the average area was calculated to determine cell density in each frame. We computed displacements squared at all subsequent images in the time series and averaged across the dataset to produce a mean squared displacement curve (supplemental figure 3d). Cell positions, velocities and shapes obtained from the tracking were used to compute all correlation functions described in the following sections.

Measurement of Cell Shape: We benchmarked a variety of algorithms for determining perimeter and area and found many give rise to large systematic errors. We have chosen to report the shape of a polygon reconstructed from identified cell vertex locations which we found to be the most accurate and robust metric (Fig S1, supplemental discussion). Shape parameter was computed for each cell based on the perimeter and area of a polygon constructed from cell vertex locations. The vertex locations parameterize the vertex model and therefore also allow for direct comparison with the model. Reconstructing a polygon from vertices also removes resolution dependent ambiguity in perimeter measurement (Coastline Paradox).

Vertex locations were found by locating branch points in the segmentation mask generated as described in Image Segmentation using the built in Matlab bwmorph function. Cells which do not have a complete set of segmented neighbors also do not have a complete set of vertices and therefore were discarded. We measured average cell shape for the full set of cells and cells remaining after discarding edge cells and found only a small difference in the average. We believe this difference comes from a small segmentation bias for cells whose outline is not correctly constrained by the neighbor. Therefore, the interior cell shapes are likely slightly more accurate. The shape parameter q = perimeter/sqrt(area) was measured by computing perimeter and area for the reconstructed polygon. The shape parameter is related to the another popular shape metric – circularity = 4*pi*area/perimeter^2 = 4*pi/q^2. Final cell shape was determined for each experiment by averaging the shape parameter for all fields of view measured to have a speed below 0.04μm/min. This final shape value is within a few percent of the value obtained by fitting the shape vs time curves to an exponential decay.

Simulation methods: We perform numerical simulations of monodisperse thermal Voronoi models, as described in (14). Briefly, we begin by writing down a dimensionless form for the standard vertex model energy, E= $\sum_{i=1}^{N}[k_A(a_i - a_0)^2 + (p_i - p_0)^2]$. This simple expression assigns an energy to the cells in a confluent monolayer in terms of their preferred geometry. The energy depends on the area a_i and perimeter p_i of each of the N cells (indexed by i), which are determined by a Voronoi tessellation of the cell positions. The unit of length in the simulations are defined such that the average cell area is unity, and we also set both the preferred area $a_0 = 1$ and the stiffness parameter $k_A = 1$. The preferred value for the cell perimeter, p_0, then constitutes the remaining control parameter which sets the target state of the monolayer.

We then use the cellGPU package to simulate overdamped Brownian dynamics of the model at different temperatures, T. The curves in Fig. 2c were created by performing an ensemble average over approximately 30 independent systems of $N = 1000$ cells at each (p_0, T) point in parameter space along the lines indicated in (Fig. 2b). Each system was initialized in a high-temperature configuration and then allowed to equilibrate at the target temperature for a large multiple of the system's characteristic relaxation time, as estimated from the data in Ref. (15); after this equilibration period the observed average shape parameter of the cells and mean-squared displacement at several typical time lags was evaluated.

Numerical simulations of the vertex model with fluctuating junctional tension are performed as previously described (16). The dimensionless energy of the vertex model is written as a function of the vertex coordinates $\{\vec{r_i}\}$; $\epsilon(\{\vec{r_i}\}) = (1/2)\sum_{\alpha=1}^{N}[k_\alpha(a_\alpha - a_0)^2 + (p_\alpha - p_0)^2] + \sum_{(i,j)} \Delta\lambda_{ij}(t) l_{ij}$. Here, α and N denote the label of each cell and the total number of cells, a_α and p_α are the area and perimeter of cell α. The preferered area and perimeter are denoted as a_0 and p_0, respectively. k_α is the stiffness parameter. In the non-dimensionalized form, we choose the length scale to satisfy the average cell area $\langle a_\alpha \rangle$ = 1. The edge length between the vertices i and j is denoted as l_{ij}, and the fluctuating line tension is introduced by $\Delta\lambda_{ij}(t)$. The dynamics of $\Delta\lambda_{ij}(t)$ is described by an Ornstein-Uhlenbeck process satisfying $\langle \Delta\lambda_{ij}(t) \rangle = 0$ and $\langle \Delta\lambda_{ij}(t_1)\Delta\lambda_{kl}(t_2) \rangle = \delta_{ik}\delta_{jl}\sigma^2\, e^{-|t_1 - t_2|/\tau}$. The vertex dynamics is described by the time-evolution equation $\eta(d\vec{r_i}/dt) = -\partial\epsilon(\{\vec{r_i}\})/\,\partial\vec{r_i}$. In this paper, we choose the $\sigma = 0.2, \tau = 1, k_\alpha = 1, a_0 = 1$, and $\eta = 1$, and we run simulations with 340 cells in a square simulation domain with periodic boundary conditions. As we for our Voronoi model simulations, we initialize the cellular configuration under a high-temperature condition ($\sigma = 0.35, \tau = 1$), and thermalize with the target parameters. After the thermalization, we begin setting the tension $\Delta\lambda_{ij}(t)$ of edges to zero permanently by randomly selecting a target edge every constant time interval τ^R.

Preparation of Fibroblast conditioned Medium: Fibroblast conditioned medium was prepared as described previously (17). We cultured a 10 cm dish of MEF cells to confluence. Cells were washed and 20 mL of fresh medium was added to the cells. 48 hours later the media was collected and centrifuged to remove any cells. Conditioned media was aliquoted and frozen at -20C then used within 1 month of preparation. The conditioned medium was mixed 50:50 with fresh culture medium and added to MDCK cells 15 minutes after plating. We tested that fibroblast conditioned medium caused cell scattering of small colonies on collagen coated glass substrates as previously described (18).

Cell division tracking: We found that our cell tracking consistently does not follow cell trajectories through a cell division because the distance threshold of trajectory linking is significantly larger than a cell radius. We exploit this to identify cell divisions and to measure shape change independent of cell division (Fig. 5d, Fig. S10). We identified cell divisions by identifying pairs of cells which appear in adjacent to each other in a frame after both cells were not present in the previous frame. We further filter out cells which are not of similar size to one another. We confirmed by inspection that this gives us a subset set of cells which have divided in the previous frame with few false positives. We then find the mother cell by looking several frames back for a cell near the centroid of the pair of daughter cells. We identify neighbors adjacent to both or one of the daughter cells and track each of these cells back in the trajectory to compare cell shapes before and after the division. We then average across all cell divisions in the dataset to produce the final values (figure 5e).

Calculation of mean squared displacements: Particle trajectories were compiled as described in the Particle Tracking section of the methods. We took each trajectory and decomposed it into non-overlapping sub trajectories starting from the initial time point with length ranging from tau = 10 minutes

to the full trajectory length. The average displacement squared for each sub trajectory was computed and averaged across the dataset (Fig. S5d). To compute the time averaged mean squared displacement averaging was done for all sub-trajectories of a single particle and each cell with a full trajectory is plotted (Fig S2a). To plot mean squared displacement against time we averaged for a given value of tau and t across the dataset (Fig. S2b).

Three Pixel vector method of measuring perimeter and area: The three pixel vector method was implemented as previously described (19). Briefly, the different ways of linking pixels are divided into 4 classes and 13 subclasses. Each subclass has a defined length giving a 1x13 vector. For each class there is a linking correction depending on the difference of the direction of the current and subsequent vector which is specified by a 4(classes)x16(directions) matrix. To determine the areas enclosed by each vector we sketched the set of 13 available classes of linking and computed the area of edge pixels and pixels outside the polygon for each case. We obtained the values [2,3,2,2,2,2,3/2,1,11/4,3,2,3/2,2]. We then inferred area corrections for each class from their respective perimeter corrections linking as follows : [0,0,0,0,1/2-2,1-1/2,2-2,0,2,2,2,2,1,1,1,1]; [0,0,1/2-2,1/2-2,2-5,0,0,2,2,2,1,1,1,1,0,0]; [0,-1,-1,2-2,0,0,0,0,2,1,1,1,1,0,0,0]; [0,0,0,1/2-2,2-2*1/2,2-2,0,2,2,2,2,1,1,1,1,0]. We compute the perimeter as described by Inoue et al. and using the same rules and values specified above plus the area of the interior pixels calculate the area of the polygon.

Field of View analysis: The shape vs speed for different size fields of view were computed by segmenting each field of view into a set of sub fields. To simplify this process first the field of view was truncated into a square to make sub division possible in such a way that the same set of cells are always measured. Then the field was subdivided into 1, 2x2, 3x3… 10x10 regions. We computed the average speed and shape in each sub-region, then binned the results of all regions according to average speed in the region. To characterize noise in the correlation we computed the average deviation of the derivative in a linear region of the curve for shape parameters ranging from 3.93 to 3.97.

Measurement of neighbor exchange rate

To measure a neighbor exchange rate, we detected 4 fold vertices and computed the ratio of 4 fold to 3 fold vertices per unit time. Such 4 fold vertices are not stable in the system and thus result in either a successful or attempted neighbor exchange. Our method does not detect the difference between successful neighbor exchanges and attempts which resolve in the original direction. A 4 cell vertex after formation was observed to resolve in either direction within at most 10 frames. Therefore, we iterated through each set of outlines and detected all 4 cell vertices. We discarded detections which happened within 10 pixels and 10 frames of a detected event to avoid double counting due to time delay between formation and resolution. This set of candidate events contained many false positives where the outline appeared to have a four-fold vertex but upon inspection of the raw data we observed a short 3 cell vertex. To remove these false positives, we manually sorted through the 4 cell vertex candidates and selected out real events.

Cell Doublet preparation and measurements

Cell doublets were produced by treating cells with trypsin until detachment from cell culture dishes. Around 15,000 cells were transferred to a PDMS well which was pretreated with 1% Pluronic f127 for 1 hour then washed 3 times with 1X PBS. Cells were incubated in the wells at 37C and 5% CO2 for 4 hours in the presence of inhibitors or doxycycline (100ng/ml) and halotag and snaptag ligands (1µM) then imaged at room temperature for less than 30 minutes. For the “Contact inhibited” condition cells were plated at nearly 100% confluence and grown for 3 days prior to detachment while other conditions were cultured under normal cell culture conditions. Cells were imaged by mounting PDMS wells on coverglass

and using imaging methods described above. The contact angle was measured between isolated pairs of cells manually using imageJ.

FUCCI Measurements

Cells were imaged in GFP and RFP channels similar to above methods. Images of FUCCI markers and cell boundaries were segmented using Phase Stretch Transform in Matlab. Each cell was identified using the cell boundaries and was determined to be GFP or RFP positive by measuring the intensity contained within the segmented images of each nuclear marker. The percent of cells in G1 was determined by taking the ratio of cells identified as only GFP positive to the cells identified as GFP positive, RFP positive and positive for both markers.

Additional Discussions

Shape Metric Benchmarking - Common Shape Metrics are resolution Dependent

In the initial analysis we noticed an unexpected density dependence of cell shape which had a different magnitude depending on which metric we used for measuring shape. We therefore decided to benchmark methods of measuring object shapes in digital images. Measuring the shape of polygonal objects projected onto a pixel grid is nontrivial because segmented edges are limited to single pixels while the actual object edge is a subpixel feature. Data is recoded on a 6.45µm/pixel camera at 20x magnification giving 0.3225 µm/pixel final resolution for a cell radius of ~15µm. Therefore, the edge of a cell contains few enough pixels that choice of shape metric is important. The simplest method to determine a length in an image made of pixels is to count the number of pixels along the perimeter. A more sophisticated method count pixels at 0, 90,180,270 degree angles (even) as a distance 1 and diagonal (odd) pixel connections as sqrt(2). This method is used to determine perimeter in the popular image processing software imageJ. This method overestimates the length where even and odd edges meet. Another method developed by Vossepoel uses a correction factor for pixels which change from odd to even with parameter values fit from simulated images of lines. This is used in the built in perimeter determination in regionprops in Matlab. This correction factor is not always accurate for small polygons. The corner correction can also be made by estimating that these connection have length sqrt(5) although this method does not correctly capture the length of some corner corrections. The Three Pixel Vector method developed by Inoue and Kimura explicitly implements all possible corner corrections (19). Object areas also can be measured in several ways. The most common method is to add up all the pixels in the polygon which the default method used in ImageJ and Matlab. This method overestimates the area of the polygon as it encloses area outside of the corrected perimeter contours described above. Instead the area can be measured by constructing the contours described above and finding the enclosed area.

To measure shape we implemented several of the methods described above. We also implemented an additional method which is specific to our dataset, where the objects can be defined by simple polygons constrained by the cell vertices. To measure this polygon shape we measure vertex locations estimated to the nearest pixel and reconstruct the polygon by connecting the vertices by straight lines. The vertex locations were determined by locating branch points in the cell outlines. One drawback of this method is that segmented cells which do not have a complete set of neighbors lack a complete set of vertices and must be discarded. For every cell with a complete set of neighbors, the vertices belonging to each cell were determined. The perimeter and area of a polygon defined by a set of vertex locations can be computed directly without lines being interpolated on to a grid of pixels. A representative image of the vertex locations on the membrane GFP image are displayed (Fig. S1b).

We compared the relationship between cell shape and density which reveals the presence of resolution dependent artifacts at increased density (Fig S1c). We observe a larger density dependence for more simple methods of perimeter and area determination which are known to produce large resolution dependent errors. The methods from three pixel vector closely match the vertex reconstruction – we infer that these are the two most accurate methods. We confirmed these findings by generating simulation images from the thermal voronoi model with known shape parameter and measuring the cell shapes (Fig S1d). This confirms that TPV and vertex reconstruction typically have less than 1% error. Finally, we

segmented images of fixed cells (Fig S9) and measured the cell shape over a trajectory. This shape should be constant because there is no motion in the images only small changes in signal to noise and small stage drifts. We compared the measured values across each trajectory to the mean and observe larger fluctuations in the TPV method (Fig S1e). This indicates that small differences in the segmentation boundaries can lead to relatively large errors in the final value of shape parameter. Therefore, the vertex reconstruction appears to be more robust to small changes in the signal. Although the vertex reconstruction requires discarding cells with an incomplete set of neighbors, for our 335x445 µm field of view and signal to noise the number of measurements (100-400 cells) is still large enough to provide a reliable average.

Cell Motility decreases with time

We observed that dynamics in the MDCK monolayers evolve with time. To confirm that individual cells show similar motile behavior we plot the time averaged mean squared displacement (TAMSD) (supplemental figure 2a). We observe qualitatively similar behavior for all cells in the monolayer. At short time scales the motion is nearly ballistic, however at later lag times the TAMSD plateaus for each cell. This indicates that at longer time scales most cells are confined by the neighboring cells and as a result do not end up traveling more than a few microns – only a fraction of a cell diameter. We also look at the time evolution of this motion by plotting the mean squared displacement as a function of time (supplemental figure 2b). We observe that for different lag times the mean squared displacement decreases. This shows that at later time points even for short time scales the diffusion is slower.

Correlation between shape and speed is not dependent on field of view size or lag time

From our vertex model, we expect that shape parameter is useful for describing dynamics in the system. In the model, passive and active forces are defined at the level of single cells and therefore we expect that behaviors are mainly dictated by a cell and its nearest neighbors. The relationship between shape and speed should not be dependent on how the system is measured, meaning that the size of the field of view or lag time is somewhat arbitrary. We first test the relationship between the correlation we observe and the field of view size. We subdivide the field of view into at 335µm square, 2x2 168µm squares, … 10x10 35µm squares (Fig. S3a). These squares contain the same data but if the properties of the system were not uniform, the smaller partitions may show different behaviors. If the forces are defined at the cellular scale, as we expect from the model, this partitioning should not change the relationship between shape and speed. We observe that across all partitions the same final relationship is recovered (Fig. S3b). However, we see that there is less noise in the correlation curve when partitioning into smaller regions. Because the correlation curve is not linear and the field of view is not uniform, as we average over a larger area the average shape does not capture local shape variation which results in regions much faster or slower than the average. We plot the fluctuations in the correlation curve with respect to the average to describe this effect (Fig S3c). As we reduce the size of the region we see the curve becomes more smooth until we reach a noise floor at ~100µm. We anticipate that if we had lower noise of cell displacements and a full segmentation these fluctuations would continue to decrease to the scale of a few interacting cells.

We also wanted to ensure that this correlation is not dependent on the lag time within a range of timescales. We plot the mean squared displacement as a function of the shape parameter at the beginning of each trajectory (Fig S3d). We see that for each lag time there is still a relationship between shape and speed. We plot the relationship between shape and speed for one field of view at different lag times and see similar behavior (Fig S3e). The relationship shifts down because the motion is diffusive so as lag time

is increased by a factor of two displacement increase by a factor of sqrt(2). We also see the same trend for the ensemble averaged correlations (Fig S3f).

Cytoskeletal rescue experiments do not restore relationship between shape and speed

Across inhibitors tested we examined the relationship of shape and speed. A subset of these conditions is plotted (Fig S8). In general, we see that the relative relationship between shape and speed for a given condition is similar however the absolute values are all shifted in the inhibitor cases. One potential concern in these experiments is that cytoskeletal polarization may be perturbed (20–24). Changes in cytoskeletal organization may affect how well the system is described by vertex models which assume the monolayer mechanics are dominated by cortical actin at the cell-cell junctions. We wanted to examine if these defects are related to changes in cell polarity observed for many of these inhibitors. We attempted rescue experiments based on published rescues of RAC and FAK knockdown experiments. It has been observed that RAC promotes apical basal polarity through its role in assembling laminin into the basement membrane (24). We seeded cells on a collagen gel with 1mg/ml matrigel to see if the inclusion of exogenous laminin in the matrix via the matrigel would rescue the relationship between shape and speed (Fig S9b). We observe the same relationship between shape and speed in the presence of matrigel. We also attempted inhibiting both FAK and ROCK at the same time. ROCK inhibition rescues wound healing in FAK knockdown cells and restores apical basal polarity (20, 25). We observe similar behavior in conditions with both inhibitors (Fig. S9a)

Neighbor exchanges rates are low

We measured the rate of neighbor exchanges for one wild type and one inhibitor dataset (Fig. S12a). We were not able to reliable measure neighbor exchange rates with automated analysis so we only made this measurement for these two datasets. These neighbor exchange rates range from 1-10 per hour per 1000 cells which seems low compared to observations from some phases of development where neighbor exchanges are important for tissue flow (26). We observe that neighbor exchange rates are shape dependent and lower in the inhibitor treated conditions at all shape parameters. This suggests that the rate of neighbor exchange also depends on the cell division fraction, or active stress in the system. We relate these neighbor exchange rates to observed velocity and observe a correlation (Fig S12b).

A metric q_{track} can be related to different modes of shape change

We wanted to confirm that the relationship between final shape and cell divisions was not simply the consequence of oriented cell division. We exploit the fact that q_{track}, the difference in cell shape between subsequent time steps ($q_{track}(t) = \langle q(t+1)-q(t)\rangle$), measures certain forms of cell shape change while ignoring others. For a given time step we directly observe, by tracking, a subset of the shape change which occurs. We denote the two types of cell shape change, division based (Q2) and deformation based (Q1 / q_{track}). The observed subset selectively ignores cell divisions, because division results in large enough displacements that the trajectory is broken in our tracking algorithm. q_{track} also ignore a subset of shape changes where the cell is not segmented in the initial or subsequent frame which are of either type (Fig S10a). Importantly, only type 1 shape changes are observed in q_{track}. Therefore, as long as there is not a difference between the behavior of cells which are not segmented we can obtain the average value of a Q1 shape change for a give frame. We can also determine the total shape change by measuring shape in each frame over all cells. We show that the ratio of these two, multiplied by a small correction factor (~1%), gives a relationship for the relative value of non-division based shape change (Fig S10b, S10c). If this value were on average 0 we could explain the correlation in figure S10d by oriented cell division. We observe values much larger than zero for this metric demonstrating there is an additional mechanism driven by cell divisions. Our calculation ignores the addition of a neighbor to cells adjacent to the

dividing cell. We confirm that the shape change by q_{track} is several times larger than the shape change which would result from the gain of an additional neighbor (Fig S10e). We observe that this metric is nearly 1 across all average shape parameters for three datasets with very different division rates (Fig. S10e). We show that for one condition q_{track} is similar to the total shape change except at early time points (Fig. S10f). We also observe that the direct shape change before and after cell division is shape dependent in this dataset consistent with the deviations in q_{track} we observe at high shape parameter (Fig. S10g).

q_{track} Derivation

The following derivation shows how the ratio of q_{track} to q_{total} can be interpreted in Fig S10. We see that the ratio is 1 if shape change through division is small or zero if shape change through division is equal to the total.

$\Delta Q_i = \Sigma \Delta q_i$

$\Delta Q = \Delta Q_1 + \Delta Q_2$

$n\langle\Delta q\rangle = n_1\langle\Delta q_1\rangle + n_2\langle\Delta q_2\rangle$

$\rho A\langle\Delta q\rangle = (\rho-\Delta\rho)A\langle\Delta q_1\rangle + 2\Delta\rho A\langle\Delta q_2\rangle$

$1 = (\rho-\Delta\rho)\langle\Delta q_1\rangle / \rho\langle\Delta q\rangle + 2\Delta\rho A\langle\Delta q_2\rangle / \rho A\langle\Delta q\rangle$

$1 = ((\rho-\Delta\rho)/\rho)(\Delta q_{track}/\Delta q_{total}) + \Delta Q_2/\Delta Q$

$((\rho-\Delta\rho)/\rho)(\Delta q_{track}/\Delta q_{total}) = 1 - \Delta Q_2/\Delta Q$

q_{track} values are consistent with junction length changes as the main source of shape change

We measured shape change along cellular trajectories $\Delta q_{track}(t) = \langle q(t+1)-q(t)\rangle$ which we show measures deformation based changes and ignores oriented division effects (Fig. S10b; S10c). We then compare this to the total shape change between time points $\Delta q_{total}(t) = q(t+1)-q(t)$ to measure the relative contribution of junction length changes and oriented division (Fig. S10e). We observe a slight discrepancy between these two metrics at early times in the experiment. At these time points cell aspect ratios may be large enough for oriented division to cause a net decrease in cell shape (Fig. S10d). However, we observe at later time points the q_{track} metric which ignores oriented division effects is sufficient to capture nearly all shape change in the monolayer. Across conditions with different division fraction junction length change consistently explains a majority of shape change in the monolayer (Fig. S10e). Therefore, the relationship in figure 5d implies that in monolayers with higher division fraction there is additional cell shape remodeling which occurs as a result of differences in cell mechanical properties or active stress caused by cell division.

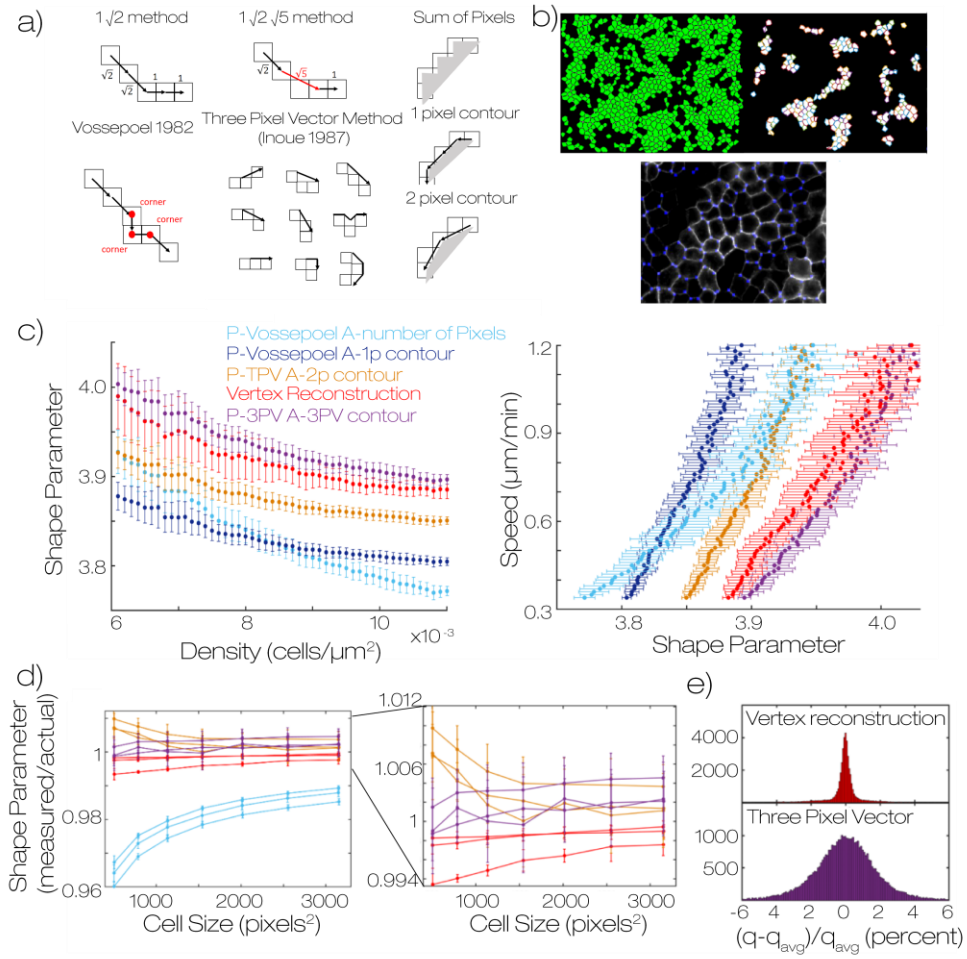

Fig. S1: *Shape metrics are resolution dependent*. (a) Schematics describing different popular methods for determination of perimeter and area of a polygon constructed from pixels. Methods depicted include the 1 root 2 method – default in imageJ, and voesspoel – default in Matlab regionprops. We also schematize several methods for determining polygon area. Sum of pixels is default in both imageJ and Matlab (b) An image of cell outlines (green) with marked vertex locations (pink). The cells which do not have a complete set of neighbors are discarded and remaining polygons are constructed from the vertices. We show an overlay of detected vertices on the raw data. (c) For one experimental dataset, WT on 2mg/ml collagen – we plot shape vs density for a variety of shape metrics and Speed vs shape for the same set of metrics. Errorbar represents the standard deviation of 60 fields of view. (d) We segmented simulation data and plot shape/actual shape vs area for this data for a variety of shape metrics for images of different resolution. We observe that some shape metrics show large errors and are highly resolution dependent. We show a zoom in of the top performing metrics. Error bar represents the standard deviation of three replicates (e) We measured shape of fixed cell data and used the two top performing metrics to measure cell shape. The fixed data is a time series of images where there is no cell motion but images show photobleaching and noise variation. We know the shapes are not changing with time, however the measured shape fluctuates due to changes in noise. We measure the distribution of measurements around the mean for the two metrics for each cell. Shapes reconstructed from vertices are much more robust to noise.

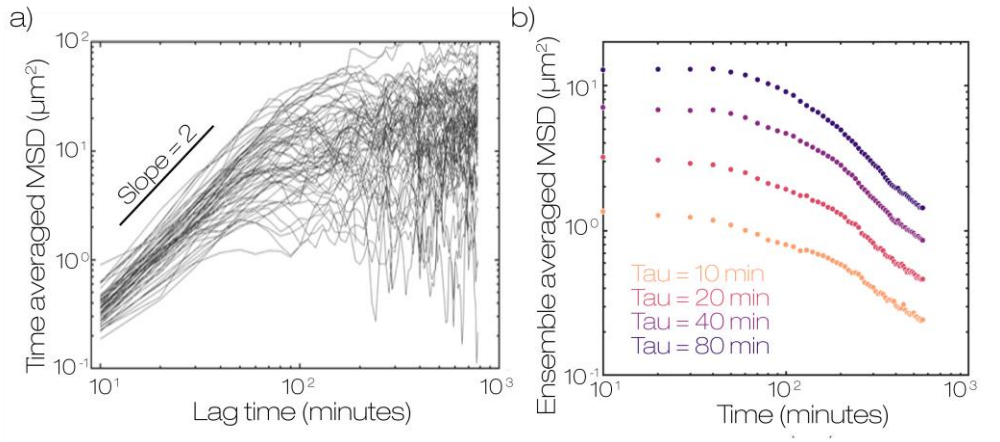

Fig. S2: *Mean square displacements are time dependent*. (a) Time averaged mean squared displacements for individual cells show a plateau at long time scales indicating cells are not freely moving at long times. A small representative subset of 100 cells from one field of view is displayed (b) mean squared displacement for different lag times decay with time indicating that cell motion is reduced at later times. Data is across all cells in 60 fields of view in a single single experiment on a 2mg/ml collagen gel under wild type conditions.

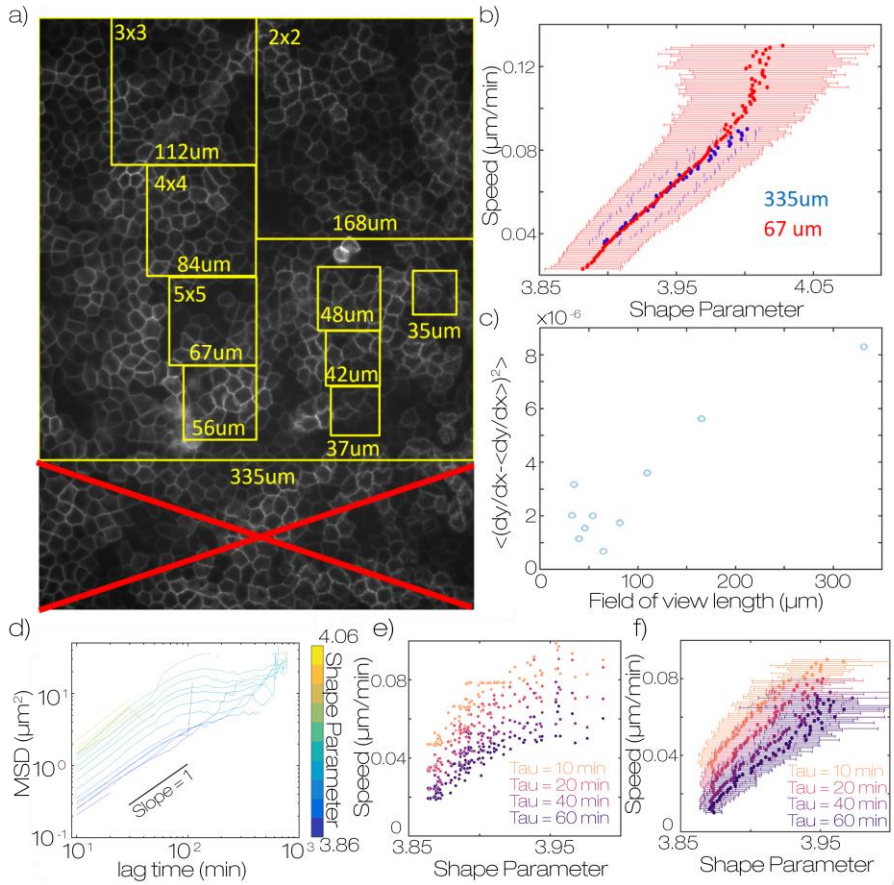

Fig. S3: *Correlation between shape and speed does not depend on the field of view size or time between images.* (a) representation of different field of view sizes. For this analysis the field of view was first truncated into a square window 335x335 µm. This field was subdivided into 2x2, 3x3… 10x10 sections and each section was averaged independently as if that was the full field of view. (b) speed vs shape parameter for the full field of view 335x335 µm and a 5x5 truncation 67x67µm. Both field of view sizes show the same final average. All other truncation sizes give the same average values. Errorbars represent the standard deviation of 60 and 1500 fields respectively (c) comparison of the derivative in the linear portion of the curve in b, from q=3.93 to q = 3.97. The value is larger for large fields of view because there can be greater heterogeneity within the same field of view. At lower values the linearity plateaus due to noise from averaging fewer cells. This shows that our data is consistent with energy which is defined at the single cell level. At ~100um or groups of a couple dozen cells we reach a noise floor. This analysis uses the TPV metric for cell shape to ensure there are enough cells within small fields of view to get a representative average. (d) Mean squared displacement for a dataset on a WT 2mg/ml collagen gel. Curves are binned according to the shape parameter at the initial time of each subtrajectory. Subtrajectories are not overlapping. Lag times are at 10 minute intervals. We observe that for any lag time there is lower displacement for lower shape parameter. (e) Shape parameter vs speed for a single field of view over 12 hours for different lag times. (f) Shape parameter vs speed averaged across all fields of view in the sample. We see that increasing the lag time shifts this curve downward consistent with the motion being diffusive. At longer times the distance traveled only increases by sqrt(t) as time increases by t leading to lower values of speed. The data appears to reach the noise floor at low shape parameter for larger tau.

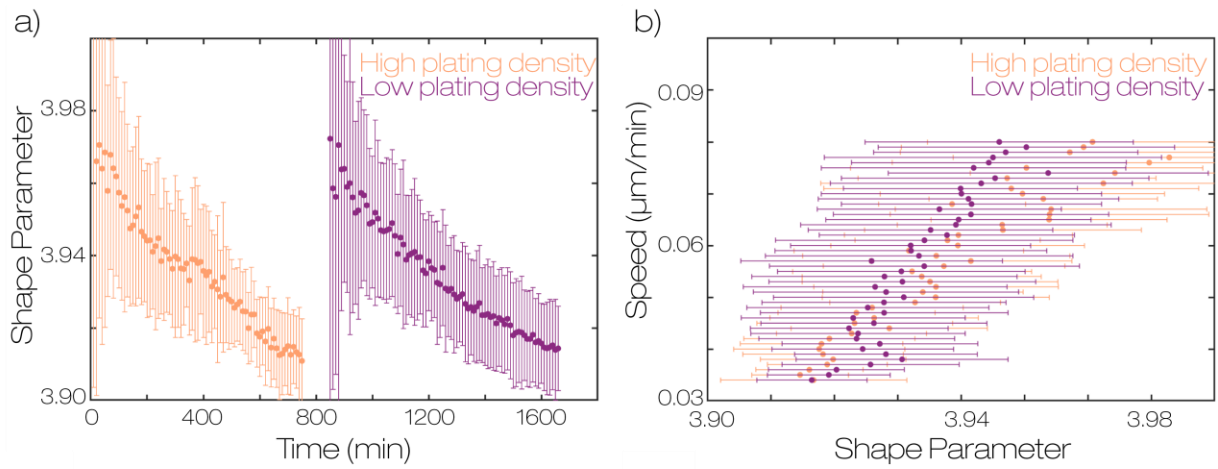

Fig. S4: *Monolayer remodeling does not depend on initial cell seeding density* (a) Cell Shape vs time for monolayers seeded at two different densities. Both samples were made at the same time with ~ 700K cells (high density) and ~350k cells (low density). The samples were imaged sequentially starting from the time point when the monolayer became confluent across most of the cover slip. (b) Relationship between cell shape and speed for the samples in a).

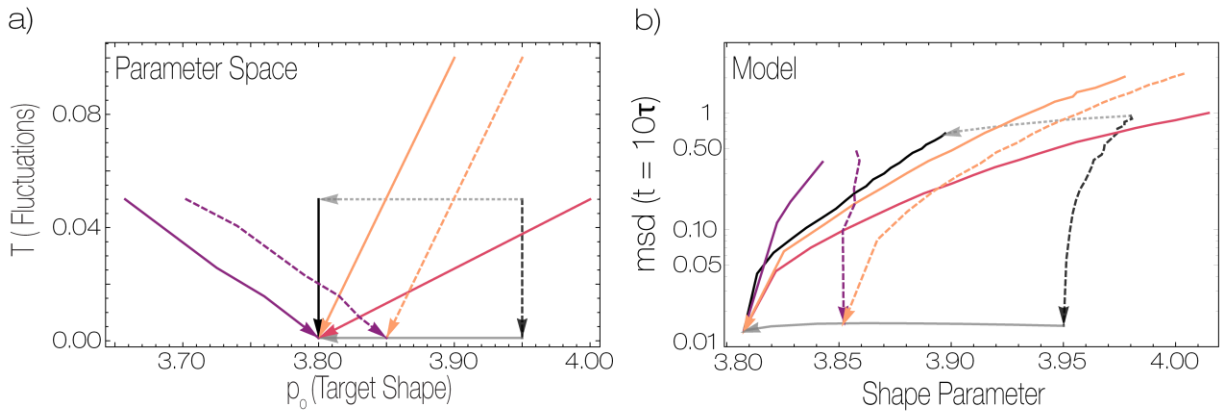

Fig. S5: *Additional modeling of relationship between cell shape and speed in Active Vertex models.* (a) Parameters of the thermal Voroni model were varied along several representative curves. Along these curves a simulated monolayer was equilibrated at each point then measurements were made on the monolayer. Solid curves approach zero temperature at a value of p_0 where the tissue is rigid, while dashed curves approach T=0 at a value of p_0 where the tissue is floppy. Colors represent trajectories with different slopes. Dashed lines indicated different values of p_0 approaching T=0. Arrows indicate the order of simulations along a trajectory (c) Observed values of speed (quantified by MSD in a given time window) and shape corresponding to the parameter space trajectories shown in panel b). MSD is given in units of $\sqrt{\text{Cell Area}}$ over a time window of 10 natural time units

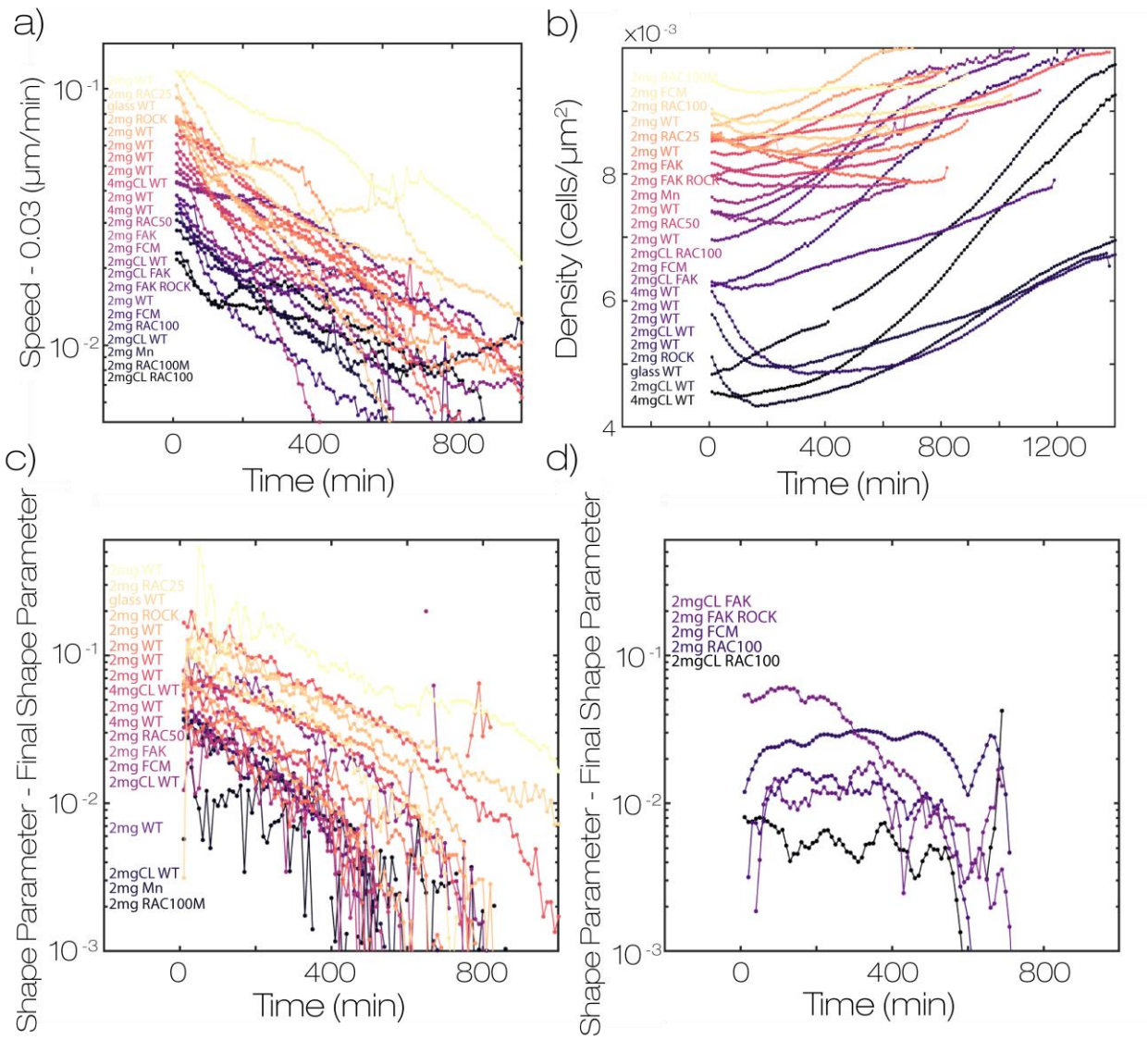

Fig. S6: *Shape, Speed and cell area decrease with time across all conditions* (a) plot of speed vs time for all conditions in figure 2b. The final speed for all the time series plateau at 0.03um/min so this value is subtracted from each time series. Curves are colored in order of speed at time = 0 min. The same color map is used in b ,c (b) density vs time for all datasets. The data is colored by initial density and differs in color map from a-c. (c) plot of shape-final shape for all data. Each time series decays to the final shape with similar kinetics. (d) several datasets were too noisy and were moved to a different plot for clarity. These datasets were smoothed temporally with a 10-point window to reduce noise. These data slightly reduce in shape with time however the difference in initial and final shape are comparable to the noise in our measurement. All data points represent the time average of at least 30 fields of view in one sample.

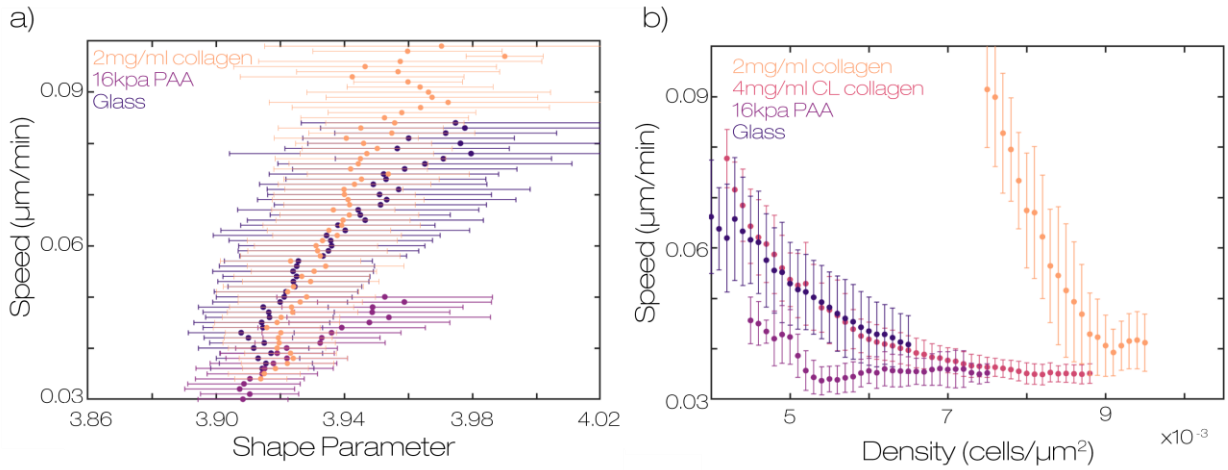

Fig. S7: *Relationship between cell shape and speed is independent of substrate stiffness* (a) relationship between shape and speed on division rate matched conditions on collagen, 16KPa polyacrylamide, and Glass (b) relationship between speed and cell density on substrates with different stiffness. Error bars represent binned averages at each speed for at least 30 fields of view over at least 60 time points.

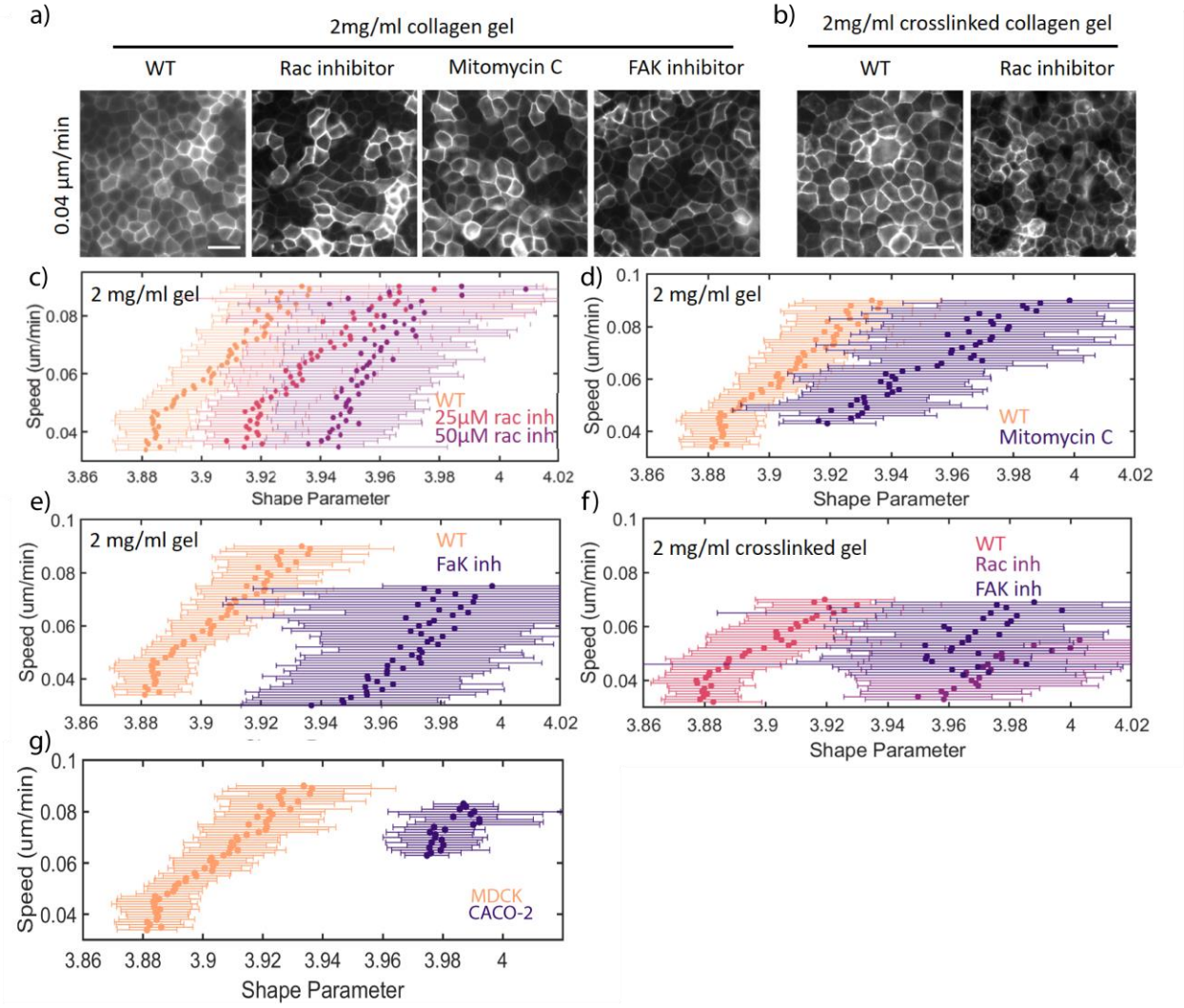

Fig. S8: *Shift in relationship between shape parameter and speed is qualitatively similar across conditions.* (a) Representative images for several inhibitor conditions at low average speed. (b) Images on stiff substrates (c-e) Speed vs Shape parameter curves for different inhibitors. Similar shifting behavior is observed across these conditions. (f-g) Similar behavior is also observed on stiff substrates and for CACO-2 epithelial cells. Error bars represent binned averages at each speed for at least 30 fields of view over at least 60 time points.

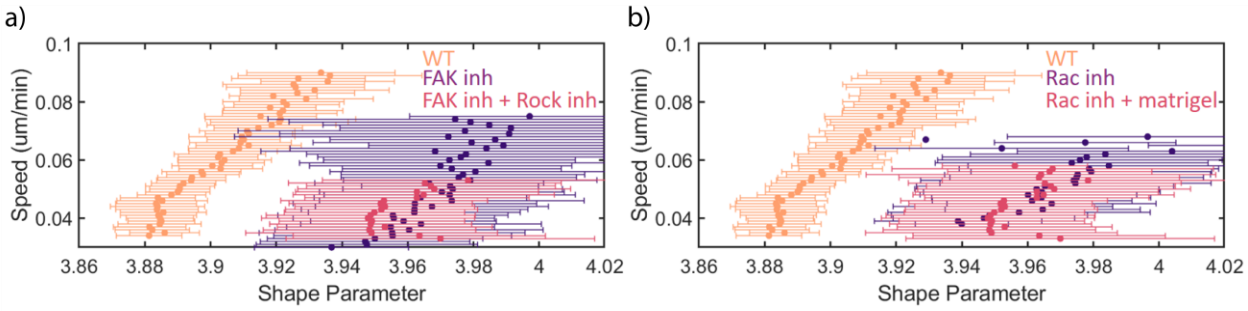

Fig. S9: *Rescue experiments do not restore shape speed correlation.* (a) Comparison of shape speed correlation for cells treated with a FAK inhibitor and with the addition of a Rock inhibitor. Rock inhibition has been observed to rescue the polarity of 3D MDCK cultures with FAK knockdown. Rock inhibition has also been observed to restore collective motility in FAK knockdown cells (see discussion). (b) Comparison of shape speed correlation for cells treated with a RAC inhibitor on collagen gels and collagen gels with 1mg/ml matrigel. RAC is required for the assembly of laminin at the basal surface of 3D cultures of MDCK cells. The polarity defect of RAC knockdown cells can be rescued by providing laminin, a component of matrigel, in the substrate (see discussion). Neither experiment significantly restored the correlation between shape and speed. Error bars represent binned averages at each speed for at least 30 fields of view over at least 60 time points.

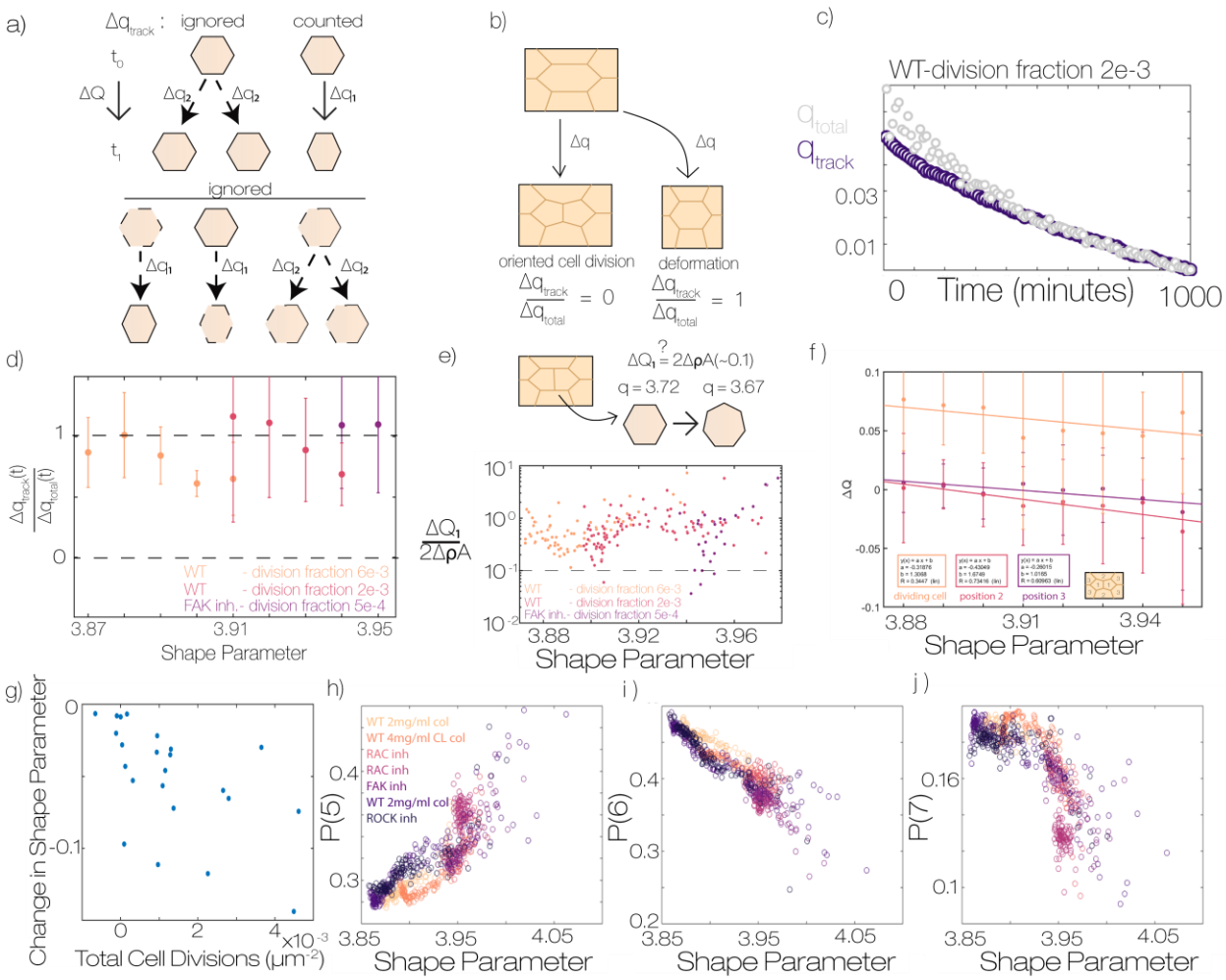

Fig. S10: *Oriented division is insufficient to explain differences in monolayer remodeling* (a) Schematic of different modes of shape change – either shape changes during a cell division Q2 or within a trajectory Q1. There are additional q1 type and q2 type shape changes which are ignored by cqq. (b) schematic of different sources of cell shape change and their relative values of q_{track} (c) for a single condition we plot q_{track} and q_{total} over time. We observe that for most of the experiment these metrics are the same. (d) ratio of q_{track} to q_{total} for three different conditions with variable division fraction noted in the legend. We observe in all cases that the ratio is close to 1 independent of cell shape. Error bars represent standard error of each bin. (e) The q_{track} metric has ignored the gain of neighbors for cells adjacent to division. We show that the value of q_{track} is several fold larger than this effect. Points represent time averaged values for at least 50 fields of view (f) average shape change per cell division from division tracking data as a function of the average cell shape. We observe larger decreases in cell shape when for the mother cell and neighbors when average cell shape parameter is large. Error bars represent standard deviation of divisions in each bin (h) plot of total cell division vs total shape change. If cell divisions cause all shape change, we would expect a strong correlation between these variables. Each point represents a different condition from Fig 5a. (g-h) relationship between shape parameter and the fraction of cells with either 5,6, or 7 neighbors for different conditions. Across these conditions we observe a common relationship between the topology of the monolayer and the shape parameter.

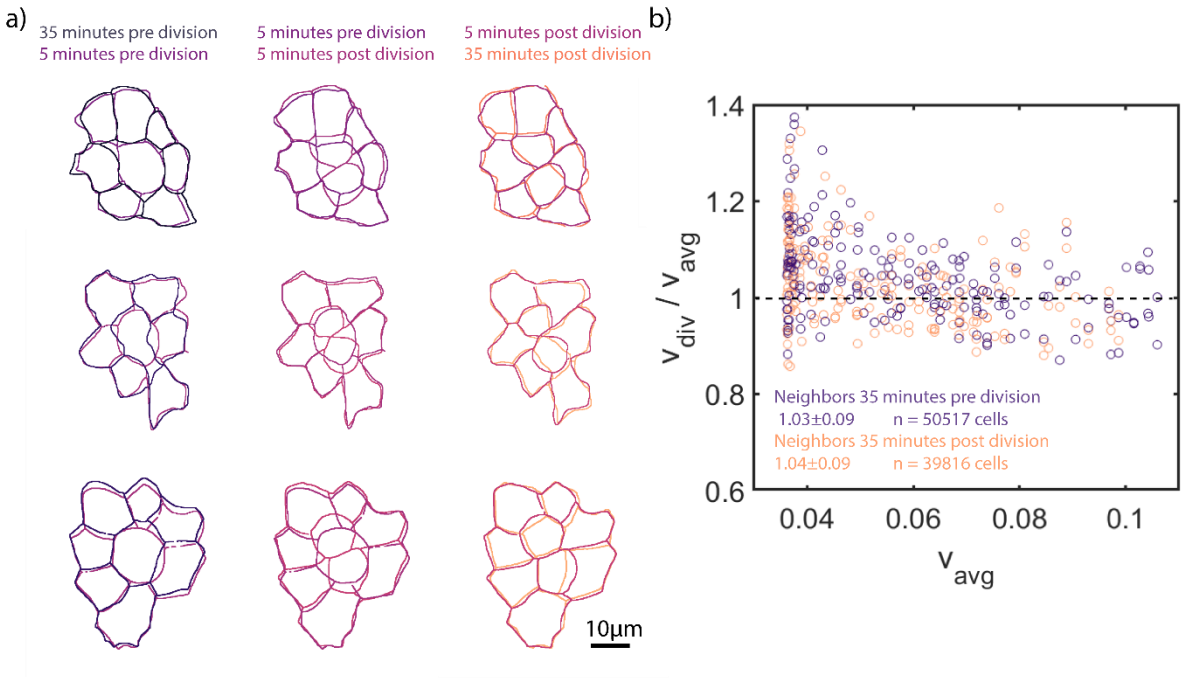

Fig. S11: *Cell divisions do not produce large deformations of the monolayer* (a) representative outlines of dividing cells in the monolayer during mitotic rounding, cytokinesis and reintegration of daughter cells into the monolayer. (b) Measurement of neighbor cell speed before and after cell divisions. All cells adjacent to a detected cell division are averaged for a given time point and compared to the average cell speed at that time point. Cell speed is the displacement of cells over a 10 minute lag time.

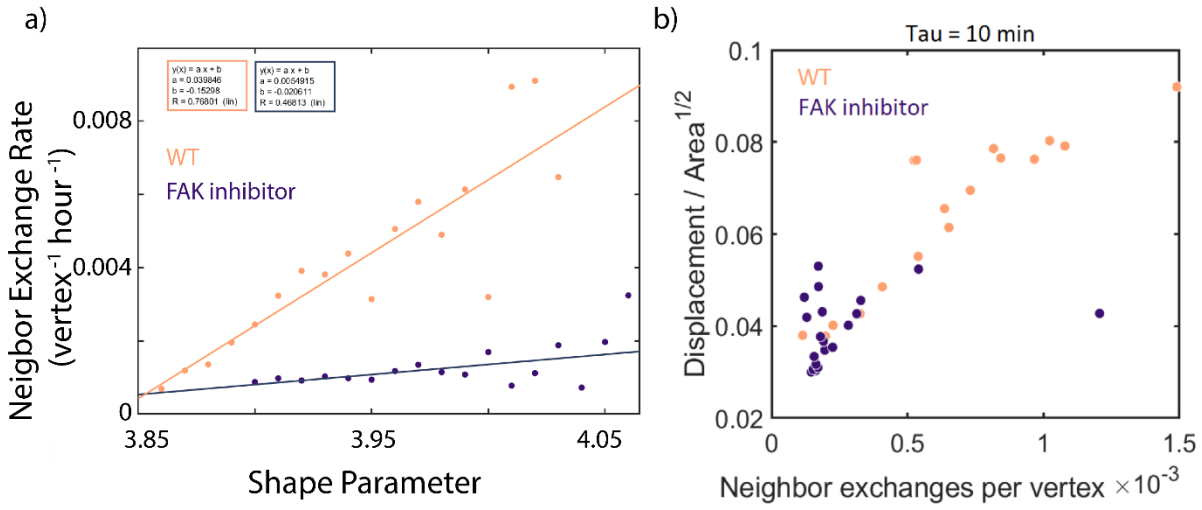

Fig. S12*: Neighbor exchange rates depend on shape and cell division rates* (a) Neighbor exchange rate vs shape for WT and FAK inhibited cells on 2mg/ml collagen gels. Neighbor exchanges were counted by identifying 4 fold vertices in the data. A 4-fold vertex represents an unstable configuration in the system and therefore will either resolve in the opposite direction it was formed (successful exchange) or back in the same direction (failed neighbor exchange). Our method does not differentiate the two types of neighbor exchange. Each point is the ratio of four fold to three fold vertices detected at the given average shape parameter across at least 50 fields of view and 80 time points in each condition. (b) relationship between neighbor exchange rates and measured cell speed for data points in a)

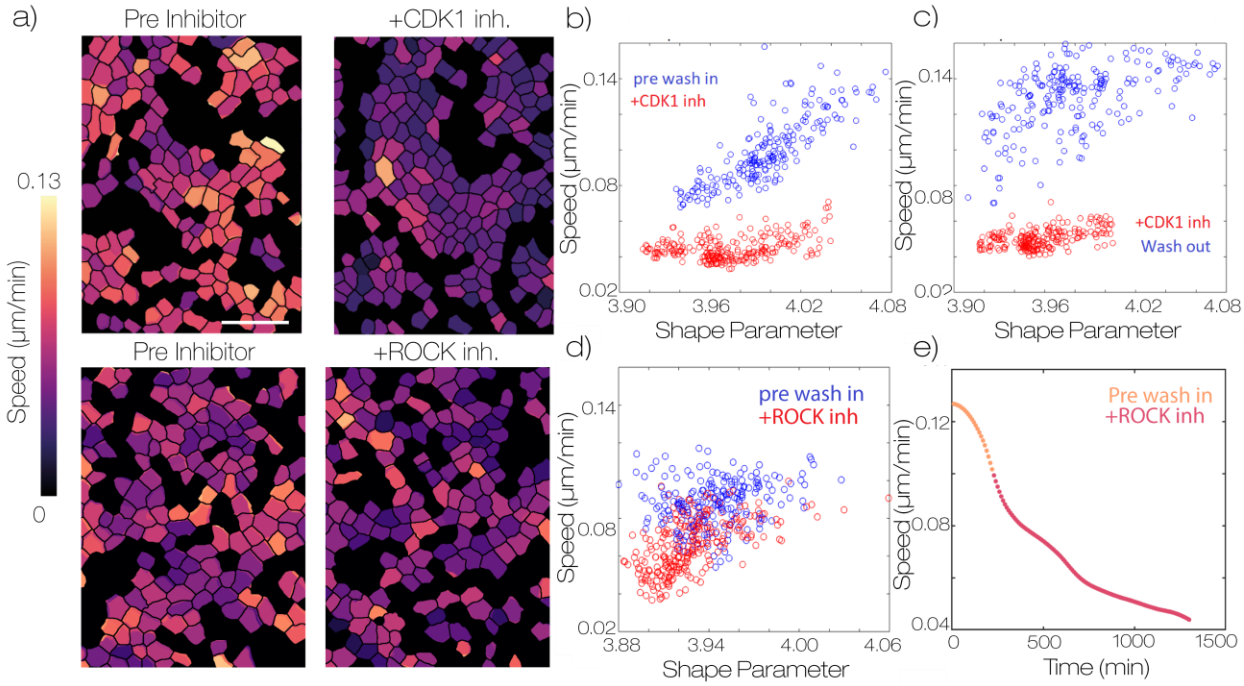

Fig. S13: *Inhibition of ROCK does not lead to large reduction in cell motility* (a) representative maps of cell speed before and after adding inhibitors of CDK1 (5µM RO-3306) and ROCK (20 µM y27632). Scale bar is 50 microns (b) measurements of cell shape vs speed for each field of view 100 minutes before and after adding the CDK1 inhibitor (c) measurements of cell shape vs speed for each field of view 100 minutes before and after washing out the CDK1 inhibitor (d) measurements of cell shape vs speed for each field of view 100 minutes before and after adding the ROCK inhibitor (e) plot of average cell speed vs time for samples treated with ROCK inhibitor.

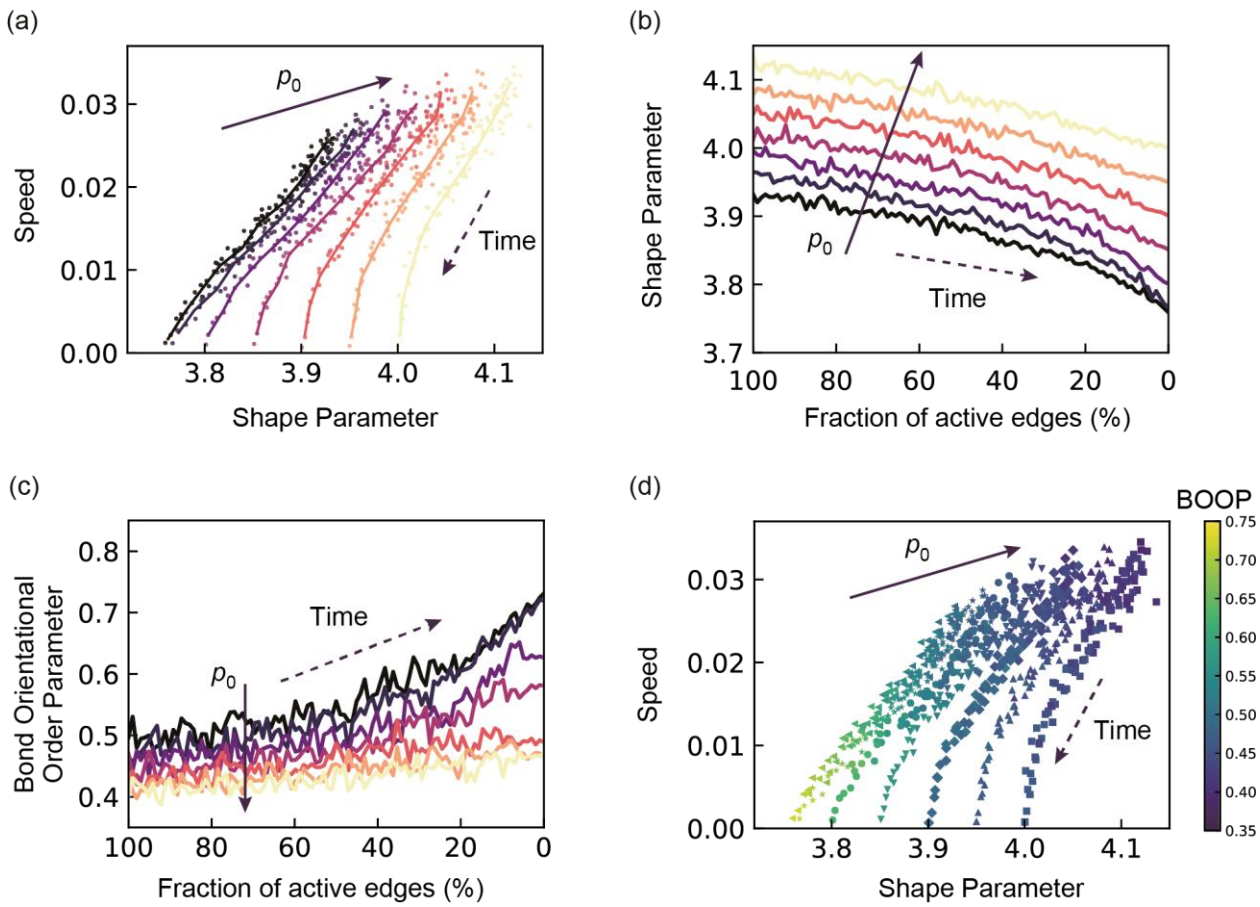

Fig. S14: *Time-evolution of the average cellular speed, the average shape parameter and the bond orientational order parameter (BOOP) for different shape indices, p_0, with $\tau^R=1$*

(a) The average cellular speed is plotted against the average shape parameter. The curves are obtained by smoothing the data points. (b) The average shape parameter is plotted against the fraction of active edges. (c) The BOOP is plotted against the fraction of active edges. (d) The average cellular speed is plotted against the average shape parameter with the color mapping the value of the BOOP. The color is mapped as shown in the color bar. In (a-c), p_0 = [3.70,3.75,3.80,3.85,3.90,3.95,4.0] from dark color to light color. In (d), the data for each p_0 is respectively represented by left-pointing triangle (3.70), star (3,75), filled circle (3.80), down-pointing triangle (3.85), diamond (3.90), up-pointing triangle (3.95) and square (4.00) symbols. In (a-d), the direction of time is indicated by dashed arrows.

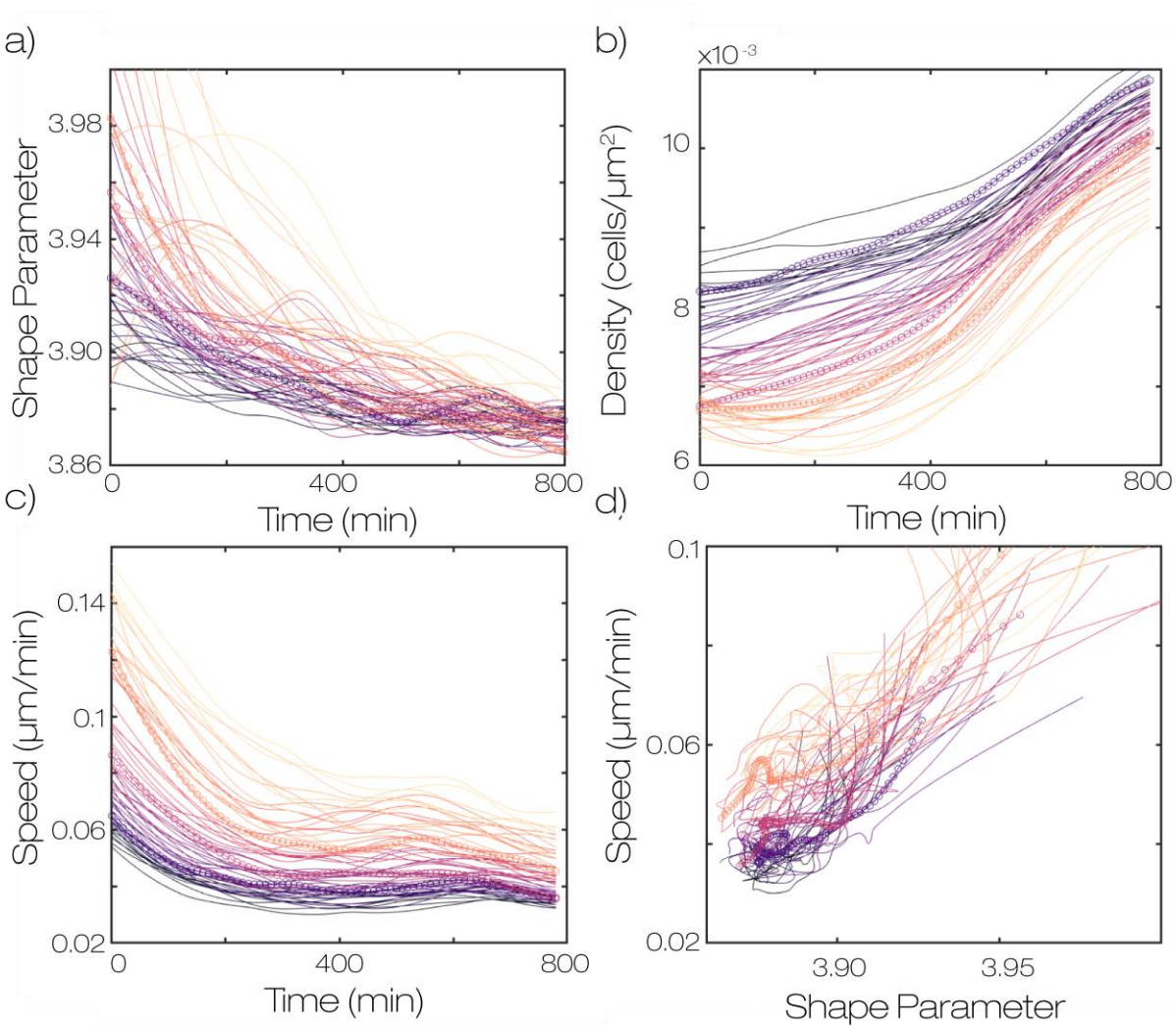

Fig. S15: *Different fields of view across the sample are qualitatively similar to the mean.* (a) values of shape parameter for each field of view over time (b) density against time for the same set of fields of view. (c) speed vs time for the same set of fields of view (d) shape vs speed for the same set of fields of view. Time points are taken every 10 minutes. Data is colored according to the initial speed in the field of view. Colors are consistent across all panels. We see that the datasets which have larger initial speed also have larger initial shape and lower initial density.

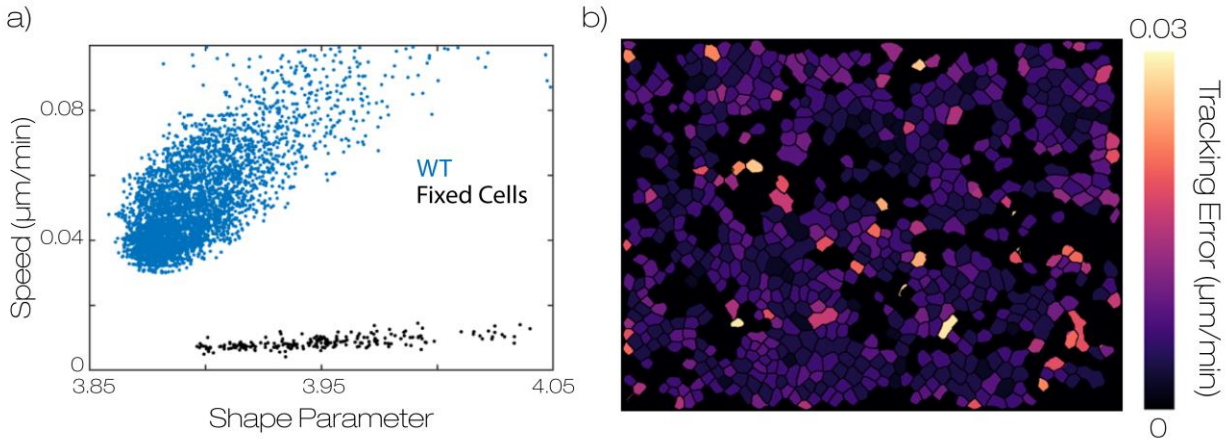

Fig. S16: *Characterization of lower bound for noise floor* (a) relationship between shape and speed in WT data on a 2mg/ml collagen gel and under the same conditions but fixed in 4% PFA for 10 minutes before imaging. We observe an estimated noise floor with minor shape dependence and average value well below the speeds measured in the experiment. The noise floor may be larger due to larger fluctuations in protein levels and cell heights in live samples. Points represent the average cell shape and speed at a single field of view and time point (b) heat map of perceived displacements in fixed data. Measurement error is fairly homogeneous and very few single cells tracking errors comparable to experimental displacements.

Legends for Movies S1 to S8

Movie S1: Representative time series from a WT monolayer on a 2mg/ml collagen gel. Tracked displacements and cell shapes are plotted as colormaps on the outline generated from segmentation. Time step is 10 minutes.

Movie S2: Representative time series from a WT monolayer on a 2mg/ml collagen gel (left) and on a 4mg/ml crosslinked gel (right). Time step is 10 minutes.

Movie S3: Representative time series from a monolayer treated with 100 µM RAC inhibitor NSC23766 on a 2mg/ml collagen gel. Tracked displacements and cell shapes are plotted as colormaps on the outline generated from segmentation. Time step is 10 minutes.

Movie S4: Representative time series from a WT condition (left) and with 50:50 Fibroblast conditioned medium to normal culture medium (right). Time step is 10 minutes.

Movie S5: Representative time series showing wash in and wash out of the CDK1 inhibitor RO-3306. The inhibitor was added at 5µM at 5:00 and washed out of the imaging chamber at 10:00. Colormap of cell speed is plotted on the segmented cell outlines. Time step is 10 minutes.

Movie S6: Representative time series showing behavior after addition of microtubule destabilizing compound Nocodazole. Nocodazole was added at 300nM at 4:00. Time step is 10 minutes.

Movie S7: Representative time series showing behavior after addition of DNA crosslinker Mitomycin C. Mitomycin C was added at 20µg/ml at 7:30. Time step is 10 minutes.

Movie S8: Representative movie showing the segmentation of a time series. Segmented cells are represented in pink and overlaid on the GFP signal. Time steps is 10 minutes. Movie duration is 800 minutes.

SI References

1. Zhu M, Lerum MZ, Chen W (2012) How to prepare reproducible, homogeneous, and hydrolytically stable aminosilane-derived layers on silica. *Langmuir* 28(1):416–423.

2. Lang NR, et al. (2015) Biphasic response of cell invasion to matrix stiffness in three-dimensional biopolymer networks. *Acta Biomater* 13:61–67.

3. Yu HH, Dohn MR, Markham NO, Coffey RJ, Reynolds AB (2016) p120-catenin controls contractility along the vertical axis of epithelial lateral membranes. *J Cell Sci* 129(1):80–94.

4. Simons K, Fuller SD (1985) Cell surface polarity in epithelia. *Annu Rev Cell Biol* 1:243–288.

5. Bissell MJ (1981) The differentiated state of normal and malignant cells or how to define a "normal" cell in culture. *Int Rev Cytol* 70:27–100.

6. Walpita D, Hay E (2002) Studying actin-dependent processes in tissue culture. *Nat Rev Mol Cell Biol* 3(2):137–141.

7. Elsdale T, Bard J (1972) Collagen substrata for studies on cell behavior. *J Cell Biol* 54(3):626–637.

8. Ng MR, Besser A, Danuser G, Brugge JS (2012) Substrate stiffness regulates cadherin-dependent collective migration through myosin-II contractility. *J Cell Biol* 199(3):545–563.

9. Trepat X, et al. (2009) Physical forces during collective cell migration. *Nat Phys* 5:426–430.

10. Asghari MH, Jalali B (2015) Edge detection in digital images using dispersive phase stretch transform. *Int J Biomed Imaging* 2015:687819.

11. Peter's Functions for Computer Vision Available at: https://www.peterkovesi.com/matlabfns/ [Accessed January 27, 2019].

12. Crocker JC, Grier DG (1996) Methods of Digital Video Microscopy for Colloidal Studies. *J Colloid Interface Sci* 179(1):298–310.

13. Simple Tracker - File Exchange - MATLAB Central Available at: https://www.mathworks.com/matlabcentral/fileexchange/34040-simple-tracker [Accessed October 2, 2017].

14. Sussman DM (2017) cellGPU: Massively parallel simulations of dynamic vertex models. *Comput Phys Commun* 219:400–406.

15. Sussman DM, Paoluzzi M, Cristina Marchetti M, Lisa Manning M (2018) Anomalous glassy dynamics in simple models of dense biological tissue. *EPL* 121(3):36001.

16. Yamamoto T, Sussman DM, Shibata T, Manning ML (2020) Non-monotonic fluidization generated by fluctuating edge tensions in confluent tissues. *arXiv*.

17. Montesano R, Schaller G, Orci L (1991) Induction of epithelial tubular morphogenesis in vitro by fibroblast-derived soluble factors. *Cell* 66(4):697–711.

18. Stoker M, Gherardi E, Perryman M, Gray J (1987) Scatter factor is a fibroblast-derived modulator of epithelial cell mobility. *Nature* 327(6119):239–242.

19. Inoue K, Kimura K (1987) A method for calculating the perimeter of objects for automatic recognition of circular defects. *NDT International* 20(4):225–230.

20. Bryant DM, et al. (2014) A molecular switch for the orientation of epithelial cell polarization. *Dev Cell* 31(2):171–187.

21. Yu W, et al. (2003) Hepatocyte growth factor switches orientation of polarity and mode of movement during morphogenesis of multicellular epithelial structures. *Mol Biol Cell* 14(2):748–763.

22. Akhtar N, Hotchin NA (2001) RAC1 regulates adherens junctions through endocytosis of E-cadherin. *Mol Biol Cell* 12(4):847–862.

23. Playford MP, Vadali K, Cai X, Burridge K, Schaller MD (2008) Focal adhesion kinase regulates cell-cell contact formation in epithelial cells via modulation of Rho. *Exp Cell Res* 314(17):3187–3197.

24. O'Brien LE, et al. (2001) Rac1 orientates epithelial apical polarity through effects on basolateral laminin assembly. *Nat Cell Biol* 3(9):831–838.

25. Chen B-H, Tzen JTC, Bresnick AR, Chen H-C (2002) Roles of Rho-associated kinase and myosin light chain kinase in morphological and migratory defects of focal adhesion kinase-null cells. *J Biol Chem* 277(37):33857–33863.

26. Blankenship JT, Backovic ST, Sanny JSP, Weitz O, Zallen JA (2006) Multicellular rosette formation links planar cell polarity to tissue morphogenesis. *Dev Cell* 11(4):459–470.