Local cellular neighbourhood controls proliferation in cell competition

Anna Bove*[1,2], Daniel Gradeci*[1,3], Yasuyuki Fujita[4], Shiladitya Banerjee[#3,5], Guillaume Charras[#1,2,5], Alan R. Lowe[#1,6]

1. London Centre for Nanotechnology, University College London, 17-19 Gordon street, London WC1H 0AH, UK
2. Department of Cell and Developmental Biology, University College London, Gower street, London, WC1E 6BT, UK
3. Department of Physics and Astronomy, University College London, Gower street, London, WC1E 6BT, UK
4. Division of Molecular Oncology, Institute for Genetic Medicine, Hokkaido University, North 15 West 7, Kita-ku, Sapporo, Hokkaido, 060-0815, Japan
5. Institute for the Physics of Living Systems, University College London, Gower street, London, WC1E 6BT, UK
6. Institute for Structural and Molecular Biology, University College London, Gower street, London, WC1E 6BT, UK

*These authors contributed equally.

[#]correspondence to: SB: shiladitya.banerjee@ucl.ac.uk, GC: g.charras@ucl.ac.uk, AL: a.lowe@ucl.ac.uk

Running head: Single cell competition

Abstract

Cell competition is a quality control mechanism through which tissues eliminate unfit cells. Cell competition can result from short-range biochemical inductions or long-range mechanical cues. However, little is known about how cell-scale interactions give rise to population shifts in tissues, due to the lack of experimental and computational tools to efficiently characterise interactions at the single-cell level. Here, we address these challenges by combining long-term automated microscopy with deep learning image analysis to decipher how single-cell behaviour determines tissue make-up during competition. Using our high-throughput analysis pipeline, we show that competitive interactions between MDCK wild-type cells and cells depleted of the polarity protein scribble are governed by differential sensitivity to local density and the cell-type of each cell's neighbours. We find that local density has a dramatic effect on the rate of division and apoptosis under competitive conditions. Strikingly, our analysis reveals that proliferation of the winner cells is upregulated in neighbourhoods mostly populated by loser cells. These data suggest that tissue-scale population shifts are strongly affected by cellular-scale tissue organisation. We present a quantitative mathematical model that demonstrates the effect of neighbour cell-type dependence of apoptosis and division in determining the fitness of competing cell lines.

Introduction

Competition between cells is a phenomenon originally identified in development that results in the elimination of less fit cells (the loser cells) from a tissue (Levayer and Moreno, 2013; Vincent *et al.*, 2013; Merino *et al.*, 2016). The viability of loser cells depends strongly on context: when they are cultured alone, they thrive but, when they are cultured in a mixed population, they are eliminated by cells with greater fitness. Many of the mutations leading to competition give rise to a change in growth rate with the faster growing cells eventually eliminating the slower growing ones (Morata and Ripoll, 1975; Simpson and Morata, 1981; Oliver *et al.*, 2004). The relationship between cell competition and cancer is complex. In some cases, competition can confer protection against tumorigenesis by eliminating cells with oncogenic mutations (Hogan *et al.*, 2009; Tamori *et al.*, 2010; Norman *et al.*, 2012; Martins *et al.*, 2014; Merino *et al.*, 2015; Porazinski *et al.*, 2016) but, in other cases, cancer cells can turn competition to their advantage leading to field cancerization (Rhiner and Moreno, 2009; Fernandez *et al.*, 2016). In cancer, many different lineages with distinct mutations are present in a tumour (Sottoriva *et al.*, 2013; Navin *et al.*, 2011) and their interaction can be viewed as a competition. Recent computational work has shown that a small competitive advantage, linked for example to better survival in the presence of a drug, can lead to one lineage taking over to whole tumour (Waclaw *et al.*, 2015). Thus, a detailed understanding of cell competition, how it depends on local context and how competition is modulated by the environment, is of clear therapeutic interest.

Competition involves either the exchange of short-range biochemical information (biochemical competition) or longer-range mechanical cues (mechanical competition). In the former, competing cells must be in contact with one another and, as a result, cell death occurs only at the interface between cell types (Moreno *et al.*, 2002). In mechanical competition, cell death occurs because of differences in tolerance to cell density and the role of contact interaction is less clear. To date, most studies have quantified competition at the tissue-scale mainly concentrating on increases in apoptotic events and reporting these across tissues (Levayer *et al.*, 2015; Wagstaff *et al.*, 2016). Although these studies have yielded insight into the mechanisms of competition, many important aspects still remain unclear. Indeed, changes in population composition can arise not only from increases in apoptosis but also through changes in division rates. Furthermore, as competition is by nature context dependent, a tissue-scale description of outcome obscures key characteristics of how competition takes place at the single cell level. For example, recent work has revealed that, when loser cells have a limited fraction of their surface in contact with winner cells, they survive (Levayer *et al.*, 2015).

These examples highlight the need to quantitatively characterise the apoptosis and division rates of each cell type and investigate how local cell density and neighbourhood modulate these. Such characterisation presents several challenges. First, cell trajectories must be accurately tracked and the cell cycle states determined over long durations for several hundred cells. Second, new data analysis tools and metrics must be developed to characterise the spatiotemporal rules of competition. Here, we address these challenges by developing long-term automated microscopy together with deep learning image analysis to decipher single cell behaviours underlying population shifts during competition. As a test case for our analysis pipeline, we examine competition between MDCK cells depleted in the polarity protein scribble (scribble^{kd}) and wild-type cells (MDCK^{WT}).

Competition induced by loss of scribble is widely regarded as an example of mechanical competition (Norman *et al.*, 2012). Recent work has suggested that mechanical competition might stem from differences in the tolerances of each cell type to density. Indeed, when cultured in pure populations, the density in each cell population first increases before reaching a plateau (the homeostatic density). This homeostatic density is maintained over time because rate of cell death exactly

compensates rate of cell birth. Remarkably, scribble[kd] cells reached lower homeostatic densities than MDCK[WT] cells because of a 3-fold larger apoptosis rate (Wagstaff et al., 2016). In competitive conditions, MDCK[WT] cells compacted scribble[kd] cells causing an increase in their death rates. These results suggest changes in density may induce increases in apoptosis events that alone are sufficient to result in the elimination of the scribble[kd] cells. However, a detailed characterisation of the influence of density on apoptosis is lacking for a quantitative test of this hypothesis and the sensitivity of cell division to density has never been characterised.

Here, we take a first step in characterising apoptosis, cell division, and net growth as a function of local cell density. Next, we show that the local neighbourhood influences apoptosis and division, suggesting a role for inductive phenomena in mechanical competition and emphasising the importance of examining single-cell behaviours to characterise competition at the tissue-scale. Finally, we implement a simple numerical model of competition to investigate how the density dependence of cell division and apoptosis determine the time evolution of cell count and the overall population fitness. Our analytical tools can be generally applied to determine the dependence of cell competition on local density and local neighbourhood, as well as analyse cell interactions with relevance to cancer, developmental biology and stem cell biology.

Experimental and computational strategy

To characterise how single cell behaviours give rise to population shifts during cell competition, we developed a high-throughput imaging and analysis pipeline. We acquired time-lapse movies lasting several days imaging co-cultures of MDCK^{WT} cells and scribble^{kd} cells expressing H2b-GFP and H2b-RFP, respectively. We first confirmed that exposure of scribble^{kd} cells to tetracycline for 70h led to ~90% reduction in scribble expression, consistent with previous reports (Norman *et al.*, 2012) (**Fig S1E**). We assembled a low-cost cell imaging system for multi-position and multi-wavelength acquisition inside a standard CO_2 incubator. Using fast-switching LEDs for illumination and an automated stage, we acquire bright-field, GFP, and RFP images for up to 12 regions of (530 × 400) μm^2 with a 20x objective with a temporal resolution of 4 minutes for over 80 hours (**Fig 1A, B**). Regions can be sampled from multiple Petri dishes allowing competitions and controls to run side by side in identical conditions. In these conditions, the fate of loser cells (scribble^{kd}) surrounded by winner cells (MDCK^{WT}) can be easily traced as they undergo apoptosis prior to extrusion from the monolayer (**Fig 1B**).

Movies were then automatically analysed to track the position, state and lineage of the cells using deep learning based image classification in a first processing step. In this first step, cell nuclei are segmented **(Fig 1A-C)** and classified using a deep neural network to assign them one of five states: interphase, prometaphase, metaphase, anaphase or apoptosis (**Fig 1D-G**). The positions and states of nuclei are then linked into trajectories over time (**Fig 1H**) using a Bayesian tracking method followed by hypothesis-based optimisation to generate lineage trees (**Fig 2**). In a second processing step, we determined the rates of division, apoptosis, and net growth as a function of density and local neighbourhood. These data were then used in numerical models of cell competition. Further details of the segmentation and tracking algorithms along with comparison to existing approaches can be found in **Supplementary Information**.

Over the course of a single imaging experiment, we acquire 12 movies in parallel, following 800-1,000 cells per field of view for 800-1200 frames. The analysis of such an experiment results in the determination of the fate of 9,600-12,000 cells (800-1,000 cells x 12) and returns 640,000 to 1,200,000 (800-1,000 x 800-1,200) discrete cellular-scale observations (cells/time) per field of view. In the following, we define an "observation" as the detection of an object (a cell) at a given frame while an "event" refers to the detection of an apoptosis or a cell division.

Results

Cell count, cell cycle length, and apoptosis in pure and mixed populations

Following tracking and identification of cell-cycle state (**Supplementary movies**), we could generate lineage trees for each cell identifying its progeny as well as potential termination by apoptosis (**Fig 2A**). By measuring the temporal separation between the birth of a cell and that of its daughters on a lineage tree, we extract the duration of the cell-cycle at single-cell resolution (**Fig 2B**) and plot its distribution for the entire population (**Fig 2C**). For MDCK^{WT}, this yields a mean cell cycle time of 18 ± 3.2 h, consistent with other reports (Puliafito *et al.*, 2012, Gudipaty *et al.*, 2017) and validating our segmentation, classification and tracking steps. It is important to note that, despite accurate estimation of cell-cycle duration for the vast majority of the population, some errors subsist due to identity swaps at high cell densities or triple divisions that lead to under-estimation of the cell-cycle time (e.g. ~1.5% of cells have estimated cell-cycle times of 5-10 hours, **Fig 2C**).

As further validation of our algorithms, we analysed the change in cell numbers and the number of apoptoses in a competition assay in which MDCK^{WT} cells were mixed with scribble^{kd} cells. We mixed cells in a 90:10 MDCK^{WT}:scribble^{kd} ratio with an initial seeding density of 10^{-3} cells/µm^2. Over the course of 80 hours, scribble^{kd} and MDCK^{WT} proliferation differed markedly, with the scribble^{kd} count peaking after 40 hours before decreasing (**Fig 2D** inset) and MDCK^{WT} count showing a 7-fold increase (**Fig 2D**), consistent with previous work (Wagstaff *et al.*, 2016). Previous work has revealed increased cell death in the scribble^{kd} cells when in presence of MDCK^{WT} cells (Wagstaff *et al.*, 2016). To confirm this, we plotted the cumulative count of apoptosis events for each cell type (**Fig 2E**). This revealed that scribble^{kd} cells had significantly higher counts of apoptosis than MDCK^{WT} cells despite them being far scarcer, indicating a higher probability of apoptosis in scribble^{kd} cells.

In contrast, control experiments mixing H2b-GFP MDCK^{WT} with non-induced H2b-RFP scribble^{kd} in a 90:10 ratio showed a 2-fold increase in cell count for H2b-RFP scribble^{kd} and a 3-fold increase in MDCK^{WT} cells after 50h (**Fig S1A**). These increases were comparable to those recorded in pure populations of each cell type over a similar duration (**Fig S1D**). When we examined apoptosis in each cell population, we found that WT cells had significantly higher apoptosis counts, as would be expected from their 10-fold larger numbers (**Fig S1B**) and in contrast to what is observed upon depletion (**Fig 1G**). Together, these data confirmed that, when scribble depletion is not induced, scribble^{kd} and MDCK ^{WT} cells did not compete (Norman *et al.*, 2012).

Altogether, our large-scale quantification of growth rates and apoptosis in competitive and non-competitive conditions are consistent with previous findings (Wagstaff *et al.*, 2016), confirming that our analysis pipeline can provide reliable high-throughput automated quantification of single cell events during competition.

Probability of division and apoptosis depend on local cell density in competitive interactions

Previous experiments showed that scribble^{kd} cells are less tolerant to density than MDCK^{WT} cells resulting in a higher rate of apoptosis for a given density (Wagstaff *et al.*, 2016). During competition, MDCK^{WT} cells cause compaction of the scribble^{kd} cells inducing them to apoptose. However, the exact relationship between cell density and probability of apoptosis was not determined. Furthermore, although a decrease in probability of division with density would also contribute to competition, the influence of cell density on cell division has not been examined.

Using our analysis pipeline, we investigated the impact of cellular density on the probability of division and apoptosis. For this, we implemented a measure of single-cell density based on a Delaunay triangulation of the centres of mass of cell nuclei in each image frame (**Fig 3A**). In computing density, we verify that cells are in contact with one another using a distance threshold

(**Fig S2, S3**) and we exclude cells that are closer than 10 μm to the edge of the field of view because not all of their neighbours can be identified leading to an underestimation of density. We defined the local cellular density ρ as the sum of inverse areas of the triangles that share a common vertex with the cell of interest, as given by equation:

$$\rho = \sum_{i=1}^{n} \frac{1}{A(i)}$$

where *A(i)* is the area of the triangle *i,* possessing a vertex in the centre of the nucleus of the cell of interest (**Fig 3A**). Consistent with previous qualitative descriptions (Wagstaff *et al.*, 2016), the mean local density of non-induced scribble[kd] (tet-) and MDCK[WT] cells increases with time in pure populations, whereas the density of pure populations of induced scribble[kd] fluctuates around the initial density value (**Fig 3B**). After 80 h, the mean density of non-induced scribble[kd] and MDCK[WT] cells reaches a plateau (the homeostatic density) several fold higher (3.6 and 3-fold, respectively) than the density of induced scribble[kd] (**Fig 3B**). Thus, induced scribble[kd] cells possess a lower homeostatic density than MDCK[WT] or non-induced scribble[kd] cells. Strikingly, when induced scribble[kd] cells are placed in competition with MDCK[WT] cells, the temporal evolution of their density changes dramatically, augmenting 4-fold over 80 h and reaching a final density 1.4 fold higher than the average density of surrounding MDCK[WT] (**Fig 3C**). In contrast, when MDCK[WT] cells and non-induced scribble[kd] cells are placed in competitive conditions, they follow behaviours similar to those observed in pure populations (**Fig S1A**). Overall, our time-resolved mean local density measurements are consistent with the global density analysis previously performed by means of single–time-point quantification (Wagstaff *et al.*, 2016).

Having validated the efficacy of our approach to estimate density heterogeneity in a competing co-culture, we experimentally determined the dependency of proliferation and apoptosis probability on local cellular density. To do this, we discretized the local densities for each cell ID at each time frame into twenty bins; the middle bin corresponds to the mean local density of all cells and the first and last bins correspond to the minimum and maximum local density measured across the population. The probability of cells undergoing mitosis/apoptosis per cell per frame was then calculated for each bin as:

$$p^{div/apo}(\rho) = \frac{\sum_{\rho} f(div/apo)}{\sum_{\rho} f},$$

with $f(div/apo)$ the number of observed events (division or apoptosis) in each bin and f the total number of observations in that bin. The net growth per cell per frame $p^{net\ growth}$ is then defined as:

$$p^{net\ growth}(\rho) = p(\rho)^{div} - p(\rho)^{apo}$$

In pure populations, MDCK[WT] cells showed a high baseline probability of division that decreased when density reached $\rho \sim 10^{-2} \mu m^{-2}$ whereas this remained approximately constant at a 4-fold smaller level for scribble[kd] cells (**Fig 3D**). Probability of apoptosis for both cell types was similar, approximately 10-fold smaller than the probability of division, and increasing with density (**Fig 3E**). Overall, this resulted in net growths that were positive and relatively insensitive to density until $\rho \sim 10^{-2} \mu m^{-2}$ after which they decreased (**Fig 3F**). These data suggest that homeostatic density is set differently for each cell type: in MDCK[WT] cells, it is controlled by a density-dependent decrease in proliferation; while in scribble[kd] cells, it results from an increase in apoptosis with density.

As hypothesised previously (Levayer *et al.*, 2015; Wagstaff *et al.*, 2016), we found that the probability of apoptosis p^{apo} of scribble[kd] cells increased sharply with density for densities $\rho > 4 \times 10^{-3} \mu m^{-2}$ when placed in contact with MDCK[WT] cells (**Fig 3H, E**). For MDCK[WT] cells, p^{apo} was comparable to that observed in pure populations until $\rho \sim 10^{-2} \mu m^{-2}$, after which it increased sharply. For high densities ($\rho \geq 10^{-2} \mu m^{-2}$), the probability of apoptosis for scribble[kd] was ~2.5 fold higher than in MDCK[WT] cells. At the lowest densities $\rho < 3 \times 10^{-3} \mu m^{-2}$, the probability of apoptosis for scribble[kd] was similar in competitive and pure populations (Compare **Fig 3H** and **3E**). Interestingly, the local density observed in competitions covered a larger range than in pure populations (**Fig 3H, 3C**), perhaps because in pure populations scribble[kd] cells strive to preserve the low homeostatic density that they prefer. Together, these data directly show that apoptosis is upregulated with increasing density, as hypothesised previously (Wagstaff *et al.*, 2016).

To date, analysis of competition has mostly focused on increases in the probability of apoptosis. Interestingly, the analysis of division probabilities in competitive conditions revealed clear differences in behaviour between scribble[kd] and MDCK[WT]. MDCK[WT] cells behaved as in pure populations, with p^{div} remaining largely insensitive to density before decreasing for $\rho \geq 10^{-2} \mu m^{-2}$ (**Fig 3D,G**). In contrast, scribble[kd] cells showed clear sensitivity to local density with similar p^{div} than in pure populations at low density but a larger p^{div} for densities $\rho \geq 10^{-2} \mu m^{-2}$ (**Fig 3D,G**). In both pure populations and competitions, p^{div}_{WT} was larger than $p^{div}_{scribble}$ until $\rho \sim 10^{-2} \mu m^{-2}$. Together, the increase in p^{apo} with density and the decrease in p^{div} with density resulted in net growth that decreased sharply for densities $\rho \geq 10^{-2} \mu m^{-2}$ (**Fig 3I**). Overall, the MDCK[WT] cells dominate the competition at all densities, with a net growth of 0.2% that drops to 0.025% for the largest densities present in our experiments due to the combination of a decrease in p^{div} and an increase in p^{apo}. Thus, our single-cell analysis emphasises the importance of considering division as well as apoptosis when examining cell competition.

Single cell analysis reveals the presence of local neighbourhood effect in net-growth

The induction of behavior in one cell lineage by contact with another is a central concept in cell competition (Vincent *et al.*, 2013). To detect inductive behaviours during cell competition, we categorise each division and apoptosis as a function of the number of neighbours of each type. To do this, we implemented a neighbourhood-based distance algorithm to retrieve the interaction network for each cell at each time point. First, we used the localization of centroids to infer the Voronoi diagram (Barber *et al.*, 1996) in each frame (**Fig 4A, Fig S2-S3**) and verified its accuracy in determining the number of neighbours of each cell (**Fig S8**). For each cell ID, we compute the total number of neighbours and the number belonging to each cell lineage. Such calculations exclude cells closer than 10 µm to the edge of the field of view because their entire neighbourhood cannot be identified. This information enables the generation of "neighbourhood plots", where the value of a parameter of interest is colour-coded and placed in a grid as a function of the number of scribble[kd] and MDCK[WT] neighbours, respectively on the x-axis and y-axis. The measurement at each grid position is typically computed from >500 cells and often 10^4-10^5 cells (**Fig S4**). In our diagrams, we annotated with an asterisk grid positions populated by more than 500 observations, but for which no event (e.g. a division or an apoptosis) of interest was detected. Thus, in these positions, we provide an upper bound ($1/N_{observations}$) for the probability of the event.

Next, we employed neighbourhood plots to investigate how proliferation, apoptosis and the resulting net-growth depend on the local neighbourhood (**Fig 4B-D**). A diagram with uniform colour indicates a behavior independent on the composition of the cell neighborhood, whereas a diagram

showing asymmetry about the diagonal identifies a behavior dependent on neighborhood. To provide a metric for asymmetry, we computed:

$$s = \left| \sum U - \sum L \right|$$

where *U* and *L* are the upper and lower triangular matrices of the neighbourhood plot (See Methods, **Fig 4E**). With this metric, the higher the *s* value the higher the inductive effect (**Fig 4F**). The probability of division (p^{div}) of MDCK^WT is strongly influenced by neighborhood, being higher in scribble^kd dominated neighborhoods (top panel, lower quadrant, **Fig 4B**). In contrast, p^{div} for scribble^kd is insensitive to neighbourhood (bottom panel, **Fig 4B**). Apoptosis diagrams displayed higher symmetry (**Fig 4F**), with p^{apo} higher in scribble^kd cells than MDCK^WT cells for most grid positions (**Fig 4C**). For both cell types, p^{apo} are lower than p^{div} by approximately an order of magnitude (**Fig 4B,C**), consistent with Fig 3. As a result, the net growth neighbourhood plots reflect the prevalent contribution of division (**Fig 4D**). Net growth is positive and highest in the bottom right corner of the neighbourhood plot for MDCK^WT cells (**Fig 4D**, top panel), while scribble^kd have either zero or negative net-growth in the upper left corner (**Fig 4D**, bottom panel), the region dominated by MDCK^WT. Similar plots for mixed populations of non-induced scribble^kd (tet-) and MDCK^WT show less sensitivity to neighbourhood, with much lower degree of asymmetry in apoptosis, division and net growth (**Fig S5**). In particular, the non-induced scribble^kd (tet-) have a positive and quite uniform net-growth across the entire grid, signifying that their behaviour is independent on the composition of the cell neighbourhood.

Rate equation model of density-dependent growth and death quantitatively reproduces competition dynamics

Previous work has suggested that mechanical competition may be the result of cell autonomous increases in apoptosis with density (Wagstaff *et al.*, 2016). Here, we test this hypothesis by developing a quantitative model based on our experimental findings. To this end, we implemented a coupled rate-equation model to investigate how the density dependence of cell division and apoptosis determine the time evolution of cell count and the overall population fitness. In this model, the density of the MDCK^WT (WT) and scribble^kd cells (KD) increases at a rate proportional to the density-dependent division rate (f_{div}^{wt}, f_{div}^{kd}), and decreases proportionally with the density-dependent death rate (f_{apo}^{wt}, f_{apo}^{kd}), as given by Eqs (1) and (2). The rate equations for cell counts N_{wt} and N_{kd} are dependent on the cell density, as given by Eqs (3) and (4).

$$\frac{\partial \rho_{wt}}{\partial t} = f_{div}^{wt}(\rho_{wt} + (1-a)\rho_{kd}) \cdot (\rho_{wt} + \rho_{kd}) - f_{apo}^{wt}(\rho_{wt} + \rho_{kd}) \cdot \rho_{wt} \qquad (1)$$

$$\frac{\partial \rho_{kd}}{\partial t} = f_{div}^{kd}(\rho_{wt} + \rho_{kd}) \cdot (\rho_{wt} + \rho_{kd}) - f_{apo}^{kd}(\rho_{wt} + \rho_{kd}) \cdot \rho_{kd} \qquad (2)$$

$$\frac{\partial N_{wt}}{\partial t} = [f_{div}^{wt}(\rho_{wt} + (1-a)\rho_{kd}) - f_{apo}^{wt}(\rho_{wt} + \rho_{kd})] \cdot N_{wt} \qquad (3)$$

$$\frac{\partial N_{kd}}{\partial t} = [f_{div}^{kd}(\rho_{wt} + \rho_{kd}) - f_{apo}^{kd}(\rho_{wt} + \rho_{kd})] \cdot N_{kd} \qquad (4)$$

Equations (1) and (2) describe the temporal evolution of local density of MDCK^WT and scribble^kd, respectively. Equations (3) and (4) describe the temporal evolution of cell count of MDCK^WT and scribble^kd, respectively. We solve Equations (1)-(4) simultaneously for the four unknowns (ρ_{wt}, ρ_{kd}, N_{wt}, N_{kd}), as since the local density ρ and cell count N are not trivially related. To describe the density dependence of birth and death rates of MDCK^WT and scribble^kd cells, we fit logistic curves to the experimental data in **Fig 3G-H** (**Fig S6C-D, Fig S6F**). The analytical form for the division rates of

scribble^{kd} (f_{div}^{kd}) and MDCK^{WT} (f_{div}^{wt}) are determined by fitting a Gaussian and a logistic function, respectively, to the experimental data in **Fig 3G** (**Fig S6A-B, Fig S6F**). The resultant input functions to the model are shown in **Fig 5A**. Based on our experimental finding (**Fig 4B-C**) that the MDCK^{WT} cells exhibit an asymmetric neighborhood dependence of division rate, we introduce a parameter, *a*, describing the degree of asymmetric dependence of MDCK^{WT} division rate on the densities of MDCK^{WT} and scribble^{kd} cells. The coupled equations (1)-(4) are solved numerically to reproduce the temporal evolution of cell counts of MDCK^{WT} and scribble^{kd} cells (**Fig 5B-C, Fig S6E**), subject to the experimental initial conditions for the cell count and density of the two cell types.

We hypothesize three different models for competitive interaction between the MDCK^{WT} and the scribble^{kd} cells, modulated by the asymmetry parameter, *a*. First, we consider an uncoupled model (*a*=1), where the growth rates of MDCK^{WT} and scribble^{kd} cells are independent of each other. Second, we examine a symmetric interaction model (*a*=0), where the evolution of cell density and count are equally affected by the densities of the MDCK^{WT} and scribble^{kd} cells. Finally, we investigate an asymmetric interaction model, with *a*>0 treated as a fitting parameter, such that the division rate of MDCK^{WT} are enhanced in the presence of scribble^{kd} neighbours (*a*>1).

Both the symmetric model (**Fig 5B**) and the uncoupled model (**Fig S6E**) fail to quantitatively reproduce the experimental cell count, predicting a lower count of MDCK^{WT} cells. By contrast, the asymmetric model (*a*=2.6), quantitatively reproduces the experimentally observed cell counts (**Fig 5C**), by accounting for the enhancement of division rate of MDCK^{WT} when in neighbourhoods with high proportions of scribble^{kd} cells (**Fig 4B, 4F**). This result reinforces our hypothesis that the MDCK^{WT} cells exhibit division induction dependent on local neighborhood. The model further predicts how the net growth of the competing cell lines varies with the densities of MDCK^{WT} and scribble^{kd} cells (**Fig 5D-E**). First, the net growth of the MDCK^{WT} is larger than that of the scribble^{kd} cells, as seen experimentally (**Fig 4D**). Second, in agreement with the experimental data (**Fig 4D**-bottom) the net growth of scribble^{kd} cells exhibit a symmetric dependence on ρ_{WT} and ρ_{kd} (**Fig 5E**). Finally, the relative fitness (net growth of the MDCK^{WT} – net growth of the scribble^{kd}) heat map (**Fig 5F)** predicts that the net growth of the MDCK^{WT} cells is always larger than scribble^{kd} cells except in a small region with high density of scribble^{kd} cells and low density of MDCK^{WT} (dashed line, **Fig 5F**). This prediction quantitatively delineates the influence of asymmetric density dependence of MDCK^{WT} division rate on the overall fitness landscape.

Discussion

Competition is a process during which two (or more) cell types interact and whose outcome is the elimination of the less fit cells. Previous work has shown that competition can arise through biochemical induction via intercellular contact or through different tolerances to cell density (Levayer and Moreno, 2013; Vincent *et al.*, 2013). Furthermore, as competition typically takes place over several days, population shifts may result from changes in division rate as well as apoptosis rate. Thus, quantitatively characterising cell competition necessitates high-throughput automated analysis strategies to mine long duration time-lapses of cell interactions.

Here we describe a new high-throughput analysis pipeline for characterising the single cell behaviours giving rise to population shifts during cell competition. We developed a low-cost time-lapse acquisition system for imaging cells over long durations (up to 80 h) that we coupled with an image analysis pipeline that tracks cells, automatically annotates cell-cycle state, and generates lineage trees for each cell at each time point. Next, we designed tools to investigate how key parameters in competition, such as the probabilities of apoptosis and division, are affected by the local cellular density as well as the composition of the local cellular neighbourhood. To benchmark our analysis, we examined the interaction between MDCK^{WT} cells and scribble^{kd} cells, as previous

work has highlighted it as an example of mechanical competition resulting from differential sensitivity of the cell lines to cell density (Wagstaff *et al.*, 2016).

Previous work showed that scribble[kd] cells are more sensitive to cell density leading to apoptosis and suggested that compaction caused by the MDCK[WT] cells leads to their eventual elimination (Wagstaff *et al.*, 2016). We found that, when scribble[kd] cells are placed in the presence of MDCK[WT] cells, their local density increases three-fold compared to when they are in pure populations. Using our automated analysis pipeline, we quantitatively characterised the dependency of the probability of apoptosis and division as a function of density. Strikingly, interaction with MDCK[WT] cells causes the probability of apoptosis of scribble[kd] cells to increase sharply at higher densities. As the probability of division is approximately one order of magnitude larger than the probability of apoptosis, any effect on division will be the dominant effect on competition. For instance, at high density ($\sim 10^{-2}$ μm^{-2}), the division rate for scribble[kd] cells is higher than the division rate of MDCK[WT] cells and higher than in pure populations of scribble[kd] cells. Despite this, the net growth of scribble[kd] cells at high density remains lower than MDCK[WT] because of a concomitant increase in the probability of apoptosis of scribble[kd] cells. Therefore, in the range of density explored in our experiments, there is no regime where scribble[kd] cells have higher net growth than MDCK [WT] cells. These competition-specific changes were intriguing because they did not fit in a framework where cell density is the main predictor of competition outcome and suggested that other factors may participate.

We addressed this question by using our single-cell analysis approach to investigate the impact of the local neighbourhood make-up on population dynamics. For this, we generated neighbourhood plots, which display how the probability of apoptosis or division depends on the number and cell-type of neighbours. Our neighbourhood plots suggest that apoptosis is increased in scribble[kd] cells possessing many neighbours, consistent with the notion that it increases at high density. Our metric for asymmetry revealed that apoptosis in scribble[kd] cells was more sensitive to neighbourhood than apoptosis in MDCK[WT] cells. However, the most striking neighbourhood dependence was revealed in neighbourhood plots of division in MDCK[WT] cells. Interestingly, we found that the probability of division is significantly higher for MDCK[WT] cells in a neighbourhood mostly populated by scribble[kd] cells. Thus, proliferation also seems to be strongly affected by local cellular neighbourhood and surprisingly, this is the case in the winner cell type. These data suggest that some inductive cell behaviour may be at play in this competition, something that is a well-known marker of cell competition in its traditional definition (Vincent *et al.*, 2013). In addition, our neighbourhood analysis can be extended to include a temporal aspect, so that changes in competition at high or low cell density can be assessed, for example (**Fig S7**). Time-resolved neighbourhood plots may also enable comparison of competition before and after drug treatment.

To explore the dependence of apoptosis and division on neighbour cell-type evident from our experimental data, we introduced a coupled rate equation model for the evolution of cell count, where rates of division and apoptosis depended on cell density. We found that the scenario that best simulated the experimental cell counts assumed an asymmetric dependency of the division rate of MDCK[WT] on the density of scribble[kd] cells. There is a clear difference when we compare this scenario to the symmetric interaction model, which underestimates the temporal evolution of MDCK[WT] cell count.

Interestingly, our numerical simulation shows that scribble[kd] cells may outcompete MDCK[WT] cells in a region of high scribble[kd] density and low MDCK[WT] density (bottom right hand corner of figure **5F**). Such regime is never observed in our experimental conditions and would require external manipulation to be applied.

Initial seeding density is a key parameter in the competition phenomenon described here. All of our experiments were performed with an initial density of 10^{-3} cells/μm^2 and a 90:10 ratio of MDCK^{WT} and scribble^{kd} cells. Future experiments will be necessary to assess how the competition outcome depends on initial seeding density and seeding ratios of the competing lines.

One intriguing question arising from our analysis is to understand if increased division of MDCK^{WT} cells in majority scribble^{kd} neighbourhoods is cause or consequence of increased apoptosis in scribble^{kd} cells. Further experimental work will be needed to understand the molecular mechanisms underlying the sensitivity of mitotic behaviour to density and neighbourhood and provide a dynamic characterization of the molecular changes occurring at the interface between cell types.

Our quantitative analysis of competition has suggested original hypotheses underlying the eventual elimination of loser cells from the tissue and emphasises the need to examine how the probability of division changes with density and neighbourhood. In addition to competition, our pipeline and characterisation tools are broadly applicable to any interaction between cell types leading to outcomes such as division, death, or differentiation in processes such as cancer, stem cell biology, and development.

Materials and Methods

Cell culture. The MDCK cell lines used for this study (MDCK[WT] and pTR scribble shRNA, scribble[kd]) were generated as described in (Norman et al., 2012). All cell lines used in this publication have been tested for mycoplasma infection and were found to be negative (MycoAlert Plus Detection Kit, Lonza, LT07-710).

MDCK cells were grown in DMEM (Thermo-Fisher) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich), Hepes buffer (Sigma-Aldrich), and 1% Penicillin/Streptomycin in a humidified incubator at 37C with 5% CO_2. The scribble[kd] cells were cultured as wild-type cells, except that we used tetracycline-free bovine serum (Clontech, 631106) to supplement the culture medium. For inducing expression of scribble shRNA, doxycycline (Sigma-Aldrich, D9891-1G) was added to the medium at a final concentration of 1 μg/ml.

To enable visualisation of nucleic acid organisation during the cell cycle, we established cell lines stably expressing fluorescently tagged histone markers. Use of different fluorescent proteins enabled us to distinguish the two competing cell types and allowed for accurate segmentation. To do this, we transduced MDCK[WT] cells with lentiviruses encoding H2B-GFP (Addgene, plasmid #25999) and the scribble[kd] cells with lentiviruses encoding H2B-RFP (Addgene, plasmid #26001). After transduction, cells were sorted using fluorescence activated flow cytometry based on GFP or RFP fluorescence to yield populations with homogenous levels of fluorescence.

Western-Blotting. We performed Western Blotting on MDCK H2b-GFP cells, non-induced MDCK scribble[kd] H2b-RFP (tet-) and induced MDCK scribble[kd] H2b-RFP cells. Induction of scribble shRNA was carried out as described in (Norman et al., 2012). Briefly, cells were induced with 1 μg/ml doxycycline for 70 hours before lysis. For preparation of protein extracts, cells were placed on ice and washed with cold PBS. After removal of PBS, the cells were lysed using RIPA lysis buffer (Santa Cruz Biotechnology) to which protease and phosphatase inhibitors were added at appropriate concentrations. The lysates were clarified by centrifugation at 8,000g for 4 min at 4C, diluted 1:1 with ×2 Laemmli buffer (Sigma-Aldrich), denatured for 5 min at 95C and loaded onto NuPage 4–20% gradient gels (Bio-Rad). For immuno-blotting, we used Goat anti-Scribble primary antibody (1:500, Santa Cruz sc-11048) and Mouse anti-GAPDH (1:1000, Novus Biologicals NB300-221) as loading control. For secondary antibodies we used HRP coupled anti-Mouse (GE Healthcare, NXA931) and HRP coupled anti-Goat (Abcam, ab 97110). All HRP coupled secondary antibodies were used at 1:10000 dilution. Protein bands were visualised using ECL Detection kit (GE Healthcare).

Wide-field microscopy. A custom-built automated epifluorescence microscope was built inside a standard CO_2 incubator (Thermo Scientific Heraeus BL20) that maintained the temperature at 37°C and in a 5% CO_2 atmosphere. The microscope comprised a high performance motorized stage (Prior Proscan III, H117E2IX), with a motorised focus controller (Prior FB201 and PS3H122R) and a 9.1MP CCD camera (Point Grey GS3-U3-91S6M). Brightfield illumination was provided using a green LED (Thorlabs M520L3, 520nm). Fluorescence illumination in two channels, GFP and mCherry/RFP, was via a blue (Thorlabs M470L3, 470nm) or yellow (Thorlabs M565L3, 565nm) LED respectively. These were combined using a dichroic beamsplitter (Semrock), and focussed onto the back focal plane of a 20x air objective (Olympus 20x, 0.4NA) in an epifluorescence configuration. The camera and the LEDs were synchronised using TTL pulses from an external D/A converter (Data Translation DT9834). A custom built humidified chamber maintained the humidity within the sample chamber and was fitted with a thermocouple and humidity sensor to continuously monitor the environment. The microscope setup was controlled via custom written software in Python and C++.

Long-term live imaging and competition assay. Cell competition assays were carried out in 35 mm glass bottom Petri-dishes (WillCo). At the start of each experiment, cells were seeded at an initial density of 1×10^{-3} cells/μm^2. MDCK^{WT} cells expressing H2b-GFP were mixed with scribble^{kd} H2b-RFP cells at a ratio of 90:10. In some experiments, the expression of scribble shRNA had been induced in scribble^{kd} cells by exposure to doxycycline for 70 hours before seeding. In other experiments, the cells were maintained in tetracycline free medium to prevent scribble shRNA induction.

Imaging was started 2–3 h after seeding. Imaging medium used during the assay was phenol red free DMEM (Thermo Fisher Scientific, 31053) supplemented with tetracycline-free bovine serum, Hepes, antibiotics and, for experiments involving induction, doxycycline at the dose indicated above. Multi-location imaging was performed inside the incubator-scope for duration of 50-80 h. Bright field, GFP and RFP fluorescence images were acquired with a frequency of 1 frame every 4 minutes for each position.

Image processing and cell tracking. After having acquired time-lapse movies of cells using the incubator microscope, we segment the images into foreground (cells) and background. Several pre-processing steps were performed to restore the image. Flat-field illumination correction and CCD "hot pixels" were removed.

Following image restoration, segmentation of the fluorescence images was performed using a Gaussian Mixture Model (GMM). Briefly, the combined intensity histogram of three images taken from the beginning, middle and end of the movie were fitted to a GMM using the Expectation Maximisation (EM) algorithm to learn the appropriate parameters (Xu and Jordan, 1996). The intensity distribution was described as a weighted sum of n normal distributions:

$$P(x|\theta) = \sum_{k=1}^{n} \lambda_k \, \mathcal{N}_k(\mu_k, \sigma_k^2)$$

Where θ represents the learned parameters for the n models: λ_k is the normalised weight, μ_k the mean intensity and σ_k^2 the variance for each normal distribution in the mixture model. We typically used n=3, and separate parameters were learnt for the GFP and RFP fluorescence movies. In general, when ordered by increasing μ_k, the three normal distributions reflect the intensity distributions of background, interphase and mitotic/apoptotic cells. The output of the segmentation method is a binary classification of the image into background and cells. Dense regions of cells were separated using either a marker controlled watershed transform, a custom written object splitting algorithm based on calculating regions of concavity in convex objects (Wienert *et al.*, 2012) or a hybrid of both methods.

Next we use an additional merging step to recombine fragments arising from over-segmentation of nuclei with a weak-fluorescence signal. The algorithm attempts to find the best possible hypothesis for merging the objects, based on separation distance and image features. This works in several phases. First, a Delaunay graph is calculated to make putative clusters of fragments. Second, hypotheses for combinations of fragments constituting a single object are constructed. Each hypothesis has an equal prior probability. Third, successive Bayesian updates are performed using the separation distance and image feature information. Finally, the algorithm selects the merging hypotheses with the highest posterior probabilities. This algorithm has the advantage of not merging apoptotic fragments with non-apoptotic nuclei.

Once the objects are segmented and split/merged, each nuclear marker in the original image is classified according to the position in the cell cycle. We use the following classes for simplicity:

Interphase, Pro(meta)phase, Metaphase, Anaphase/Telophase and Apoptotic. We trained a deep convolutional neural network (CNN) to perform object classification.

Our CNN architecture was broadly based on the LeNet-5 architecture (Lecun *et al.*, 1998), and consists of several layers of 3x3 convolution, rectified linear units (ReLU) (Nair and Hinton, 2010) and 2x2 max-pooling units (Scherer *et al.*, 2010), which decrease spatial dimensionality and increase the number of filters. These layers are followed by several fully connected layers, which reduce the output to a one-dimensional tensor representing the five mutually exclusive cell-cycle classes. A final Softmax layer (normalized exponential function) returns the output probabilities for each class.

In order to train the deep CNN, we generated three datasets: (i) training (ii) test (iii) validation. Each set has an identical number of training examples that are shuffled, class balanced, and augmented (rotations, noise, translations) to yield a large number of training examples. CNNs were implemented in *Caffe* (Jia et al., 2014) or *TensorFlow* (Abadi et al., 2016). Training was performed using a momentum optimizer with an exponentially decaying learning rate until convergence. We measured the accuracy of the CNN classification by calculating a confusion matrix that compares a ground truth based on human operator classification and the CNN prediction using the validation data set. Following the training steps, the CNN achieved an overall accuracy of >99% (**Fig 1F**).

We also trained a non-linear Support Vector Machine (SVM) with a Radial Basis Function using image features such as fluorescence intensity, intensity gradient, HoG features (Dalal and Triggs, 2005), orientation, eccentricity and texture. Although the SVM performed well at cell-cycle state classification, it did not match the performance of the CNN, particularly with apoptosis detection, with a maximum accuracy of ~80% (Not shown). We utilised the CNN for all further data analysis.

Next, classified and segmented objects are assembled into tracks. The tracking algorithm assembles reliable sections of track that do not contain cell division events (tracklets). Each new tracklet initiates a probabilistic model in the form of a Kalman filter (Kalman, 1960), and utilises this to predict future states (and error in states) of each of the objects in the field of view. We assign new observations to the growing tracklets (linking) is performed by evaluating the posterior probability of each potential linkage from a Bayesian belief matrix for all possible linkages (Narayana and Haverkamp, 2007). The best linkages are those with the highest posterior probability. Despite the high instantaneous accuracy of the CNN classification, occasional errors occur. We correct errors using a temporal model of the cell cycle implemented as a Hidden Markov Model (HMM) comprising interphase, the three states of mitosis, and a dead-end state of apoptosis. Any tracklets containing a metaphase to anaphase transition are split into separate tracks so that they can be labeled as division events in later steps of the algorithm.

The tracklets are then assembled into lineage trees by using multiple hypothesis testing and integer programming (Al-Kofahi et al., 2006; Bise et al., 2011) to identify a globally optimal solution. We build upon this previous work to incorporate hypotheses specific to apoptosis/extrusion and use additional geometric features and CNN classifications in the hypothesis generation. The following hypotheses are generated: (i) true positive track (ii) false positive track (iii) initializing at the beginning of the movie or near the edge of the FOV, (iv) termination at the end of the movie or near the edge of the FOV (v) a merge between two tracklets (vi) a division event or (vii) an apoptotic event. The likelihood of each hypothesis is calculated for some or all of the tracklets based on heuristics. The global solution identifies a sequence of high likelihood hypotheses that accounts for all observations. Having identified the global solution, the fates of each cell are updated, tracks are merged and lineage trees are generated using a Breadth First Search (BFS) to traverse the trees.

At the end of the image processing and tracking steps, the xyt-position, cell-cycle state, and lineage of each cell in the field of view has been determined.

All code was implemented in Python and C/C++ using CVXOPT, GLPK, Numpy, Scipy, TensorFlow and Caffe libraries. All image processing was performed on a Dell Precision workstation running Ubuntu 16.04LTS with 32Gb RAM and a NVIDIA GTX1080 GPU. Computational time is of the order of minutes to hours depending on the complexity of the data.

Post processing analysis. The output of the tracking software is a table containing a time-resolved list of unique cell IDs; for each cell ID, the software saves the centroid coordinates, the assigned cell cycle state and lineage information (mother ID).

To investigate the role of local neighbourhood in cell competition, we implemented a neighbourhood-based distance algorithm to retrieve the cellular interaction network (**Fig 3A**). A custom written *MATLAB* (MathWorks) script was created to calculate the Voronoi diagram (Barber et al., 1996) using the known localization of cell centroids in each frame. The distance between Voronoi cells was computed (**Fig S2**) and compared to a threshold value (D_{thresh}). We calculated the mean value of inter-nuclear separation over time, determining D_{thresh} to be 30 μm for MDCK ^{WT} and 60 μm for induced MDCK *scribble*^{kd}. Among the neighbouring Voronoi cells centred at a distance below D_{thresh}, we defined true neighbours as those cells sharing one common vertex with the target cell. This definition is important to detect and remove neighbours situated too far from the cell of interest to truly interact with it. Thus, for each cell ID, we could compute the number of neighbours and the fraction of neighbours belonging to each cell lineage. This allowed investigation of the dependency of proliferation and apoptosis on local neighbourhood (**Fig 3B-C**).

To investigate how apoptosis and division depended on local cell density, we implemented a custom-written *MATLAB* script to computing a local cellular density measurement based on the Delaunay triangulation of nuclei centres of mass in each image frame (**Fig 3A**). We defined the local cellular density as the sum of inverse areas of the triangles that share a common vertex with the cell of interest, as given by equation:

$$\rho = \sum_{i=1}^{n} \frac{1}{A(i)}$$

where *A(i)* is the area of the triangle *i* sharing a vertex with the target cell. Local cell density (ρ) was computed for each cell ID and averaged among cells of the same lineage at each time point. Such average was then plotted as a function of time for each cell type separately in mixed populations. We discretized the local densities for each cell ID at each time frame into twenty bins; the middle bin corresponds to the mean local density of all cells and the first and last bin correspond to the minimum and maximum local density measured across the population. The probability of cells undergoing mitosis/apoptosis per cell per frame was then calculated for each bin as:

$$p^{div/apo}(\rho) = \frac{\sum_{\rho} f(div/apo)}{\sum_{\rho} f},$$

with $f(div/apo)$ the number of observed events (division or apoptosis) in each bin and f the total number of events. The net growth per cell per frame is then defined as:

$$p^{net\ growth}(\rho) = p(\rho)^{div} - p(\rho)^{apo}$$

Error bars correspond to the inverse of the number of observations for each data point. For generating the neighbourhood diagrams shown in **Fig 4**, we implemented a custom-written MATLAB script to compute the interaction network for each cell at each time point, based on a Voronoi tessellation. For each cell ID, we compute the total number of neighbours and the number belonging to each cell lineage, MDCK[WT] or scribble[KD]. We categorise each division and apoptosis event as a function of the number of neighbors of each type. We color-code the probability of division, apoptosis and net-growth and display these parameters as function of neighborhood by placing them in a grid, where the x-axis and y-axis respectively represents the number of scribble[kd] and MDCK[WT] neighbours. The measurement within each grid position is computed from $N_{observations}$ >500 cells. We annotated with an asterisk grid positions populated by more than 500 observations, but for which no event (e.g. a division or an apoptosis) of interest was detected. In these positions, we provide an upper bound ($1/N_{observations}$) for the probability of the event. We analyzed the symmetry of the neighborhood plots by calculating an asymmetry parameter *s*:

$$s = \left| \sum U - \sum L \right|$$

where *U* and *L* are the upper and lower triangular matrices of the neighbourhood plot.

Numerical simulations. The experimental data for the density dependence of the probability of division and apoptosis were fitted with analytic functions. These functions correspond to the density dependent division and apoptotic rate for both respective cell types used in the model (f_{div}^{wt}, f_{apo}^{wt}, f_{div}^{kd}, f_{apo}^{kd}). The mathematical form of the fitting functions were chosen to minimise the number of parameters (**Fig 5A, S5A-D**), while providing an accurate fit that converge at extremities. For density dependent apoptotic rate of MDCK[WT] and density dependent apoptotic rate of scribble[KD], logistic functions satisfied these criteria. For density dependent division rate of wild-type and scribble[KD], Gaussian functions best described the experimental trends (**S5 A-B**). We numerically solved the four coupled rate equations (Eq. 1-4) for the density of MDCK[WT], density of scribble[KD], cell count of MDCK[WT] and the cell count of scribble[KD] using *Mathematica* (Wolfram Research Inc.), and plotted the normalised cell counts in Fig 5c. We studied two interacting limits of the rate equation models, one describing a symmetric interaction of local densities of MDCK[WT] and scribble[KD] (a=0) and the other describing asymmetric interaction (a>0 in equations 1-4). We found that the asymmetric model best replicated the experimental findings. For figure 5d-e, the heat maps of net growth against density of MDCK[WT] and density of scribble[KD] were then plotted by using the built in Density Plot function in *Mathematica*, where the net growth of a cell type is calculated as the density dependent division rate minus the density dependent apoptotic rate. The relative fitness is calculated as the net growth of MDCK[WT] minus the net growth of scribble[KD].

Software availability. *MATLAB* scripts for analysis of cell trajectories are available at https://github.com/quantumjot/CellTracking. The Bayesian tracking library is available at https://github.com/quantumjot/BayesianTracker.

Acknowledgements

References

Abadi, M., Agarwal, A., Barham, P., and Brevdo…, E. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. In arXiv.

Al-Kofahi, O., Radke, R.J., Goderie, S.K., Shen, Q., Temple, S., and Roysam, B. (2006). Automated cell lineage construction - A rapid method to analyze clonal development established with murine neural progenitor cells. Cell Cycle *5*, 327-335.

Barber, C.B., Dobkin, D.P., and Huhdanpaa, H. (1996). The Quickhull algorithm for convex hulls. Acm T Math Software *22*, 469-483.

Bise, R., Yin, Z.Z., and Kanade, T. (2011). Reliable Cell Tracking by Global Data Association. I S Biomed Imaging, 1004-1010.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P., and Sabatini, D.M. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol *7*.

Dalal, N., and Triggs, B. (2005). Histograms of oriented gradients for human detection. In … and Pattern Recognition.

Fernandez, L.C., Torres, M., and Real, F.X. (2016). Somatic mosaicism: on the road to cancer. Nat Rev Cancer *16*, 43-55.

Held, M., Schmitz, M.H., Fischer, B., Walter, T., Neumann, B., Olma, M.H., Peter, M., Ellenberg, J., and Gerlich, D.W. (2010). CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. Nat Methods *7*, 747-754.

Hilsenbeck, O., Schwarzfischer, M., Skylaki, S., Schauberger, B., Hoppe, P.S., Loeffler, D., Kokkaliaris, K.D., Hastreiter, S., Skylaki, E., Filipczyk, A., Strasser, M., Buggenthin, F., Feigelman, J.S., Krumsiek, J., van den Berg, A.J.J., Endele, M., Etzrodt, M., Marr, C., Theis, F.J., and Schroeder, T. (2016). Software tools for single-cell tracking and quantification of cellular and molecular properties. Nat Biotechnol *34*, 703-+.

Hogan, C., Dupre-Crochet, S., Norman, M., Kajita, M., Zimmermann, C., Pelling, A.E., Piddini, E., Baena-Lopez, L.A., Vincent, J.P., Itoh, Y., Hosoya, H., Pichaud, F., and Fujita, Y. (2009). Characterization of the interface between normal and transformed epithelial cells. Nat Cell Biol *11*, 460-467.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014). Caffe: Convolutional Architecture for Fast Feature Embedding. In arXiv.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. In Journal of basic Engineering.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. P Ieee *86*, 2278-2324.

Levayer, R., Hauert, B., and Moreno, E. (2015). Cell mixing induced by myc is required for competitive tissue invasion and destruction. Nature *524*, 476-480.

Levayer, R., and Moreno, E. (2013). Mechanisms of cell competition: themes and variations. J Cell Biol *200*, 689-698.

Martins, V.C., Busch, K., Juraeva, D., Blum, C., Ludwig, C., Rasche, V., Lasitschka, F., Mastitsky, S.E., Brors, B., Hielscher, T., Fehling, H.J., and Rodewald, H.R. (2014). Cell competition is a tumour suppressor mechanism in the thymus. Nature *509*, 465-470.

Merino, M.M., Levayer, R., and Moreno, E. (2016). Survival of the Fittest: Essential Roles of Cell Competition in Development, Aging, and Cancer. Trends Cell Biol *26*, 776-788.

Merino, M.M., Rhiner, C., Lopez-Gay, J.M., Buechel, D., Hauert, B., and Moreno, E. (2015). Elimination of unfit cells maintains tissue health and prolongs lifespan. Cell *160*, 461-476.

Morata, G., and Ripoll, P. (1975). Minutes: mutants of drosophila autonomously affecting cell division rate. Dev Biol *42*, 211-221.

Moreno, E., Basler, K., and Morata, G. (2002). Cells compete for decapentaplegic survival factor to prevent apoptosis in Drosophila wing development. Nature *416*, 755-759.

Nair, V., and Hinton, G.E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. Proceedings of the 27th International Conference on Machine Learning.

Narayana, M., and Haverkamp, D. (2007). A Bayesian algorithm for tracking multiple moving objects in outdoor surveillance video. Proc Cvpr Ieee, 3396-+.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndee, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., Muthuswamy, L., Krasnitz, A., McCombie, W., Hicks, J., and Wigler, M. (2011). Tumour evolution inferred by single-cell sequencing. Nature *472*, 90-4.

Norman, M., Wisniewska, K.A., Lawrenson, K., Garcia-Miranda, P., Tada, M., Kajita, M., Mano, H., Ishikawa, S., Ikegawa, M., Shimada, T., and Fujita, Y. (2012). Loss of Scribble causes cell competition in mammalian cells. J Cell Sci *125*, 59-66.

Oliver, E.R., Saunders, T.L., Tarle, S.A., and Glaser, T. (2004). Ribosomal protein L24 defect in belly spot and tail (Bst), a mouse Minute. Development *131*, 3907-3920.

Porazinski, S., de Navascues, J., Yako, Y., Hill, W., Jones, M.R., Maddison, R., Fujita, Y., and Hogan, C. (2016). EphA2 Drives the Segregation of Ras-Transformed Epithelial Cells from Normal Neighbors. Curr Biol *26*, 3220-3229.

Prince, S.J.D. (2012). Computer vision : models, learning, and inference. Cambridge University Press: New York.

Rhiner, C., and Moreno, E. (2009). Super competition as a possible mechanism to pioneer precancerous fields. Carcinogenesis *30*, 723-728.

Scherer, D., Muller, A., and Behnke, S. (2010). Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. Lect Notes Comput Sc *6354*, 92-101.

Simpson, P., and Morata, G. (1981). Differential mitotic rates and patterns of growth in compartments in the Drosophila wing. Dev Biol *85*, 299-308.

Sottoriva, A., Spiteri, I., Piccirillo, S., Touloumis, A., Collins, V., Marioni, J., Curtis, C., Watts, C., and Tavaré, S. (2013). Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. PNAS *110*, 4009-14.

Tamori, Y., Bialucha, C.U., Tian, A.G., Kajita, M., Huang, Y.C., Norman, M., Harrison, N., Poulton, J., Ivanovitch, K., Disch, L., Liu, T., Deng, W.M., and Fujita, Y. (2010). Involvement of Lgl and Mahjong/VprBP in Cell Competition. Plos Biol *8*.

Vincent, J.P., Fletcher, A.G., and Baena-Lopez, L.A. (2013). Mechanisms and mechanics of cell competition in epithelia. Nat Rev Mol Cell Biol *14*, 581-591.

Waclaw, B., Bozic, I., Pittman, M., Hruban, R., Vogelstein, B., and Nowak, M. (2015). A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. Nature *525*, 261-4.

Wagstaff, L., Goschorska, M., Kozyrska, K., Duclos, G., Kucinski, I., Chessel, A., Hampton-O'Neil, L., Bradshaw, C.R., Allen, G.E., Rawlins, E.L., Silberzan, P., Carazo Salas, R.E., and Piddini, E. (2016). Mechanical cell competition kills cells via induction of lethal p53 levels. Nat Commun *7*, 11373.

Wienert, S., Heim, D., Saeger, K., Stenzinger, A., Beil, M., Hufnagl, P., Dietel, M., Denkert, C., and Klauschen, F. (2012). Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach. Sci Rep *2*, 503.

Winter, M., Wait, E., Roysam, B., Goderie, S.K., Ali, R.A.N., Kokovay, E., Temple, S., and Cohen, A.R. (2011). Vertebrate neural stem cell segmentation, tracking and lineaging with validation and editing. Nat Protoc *6*, 1942-1952.

Xu, L., and Jordan, M.I. (1996). On convergence properties of the EM algorithm for gaussian mixtures. Neural Comput *8*, 129-151.
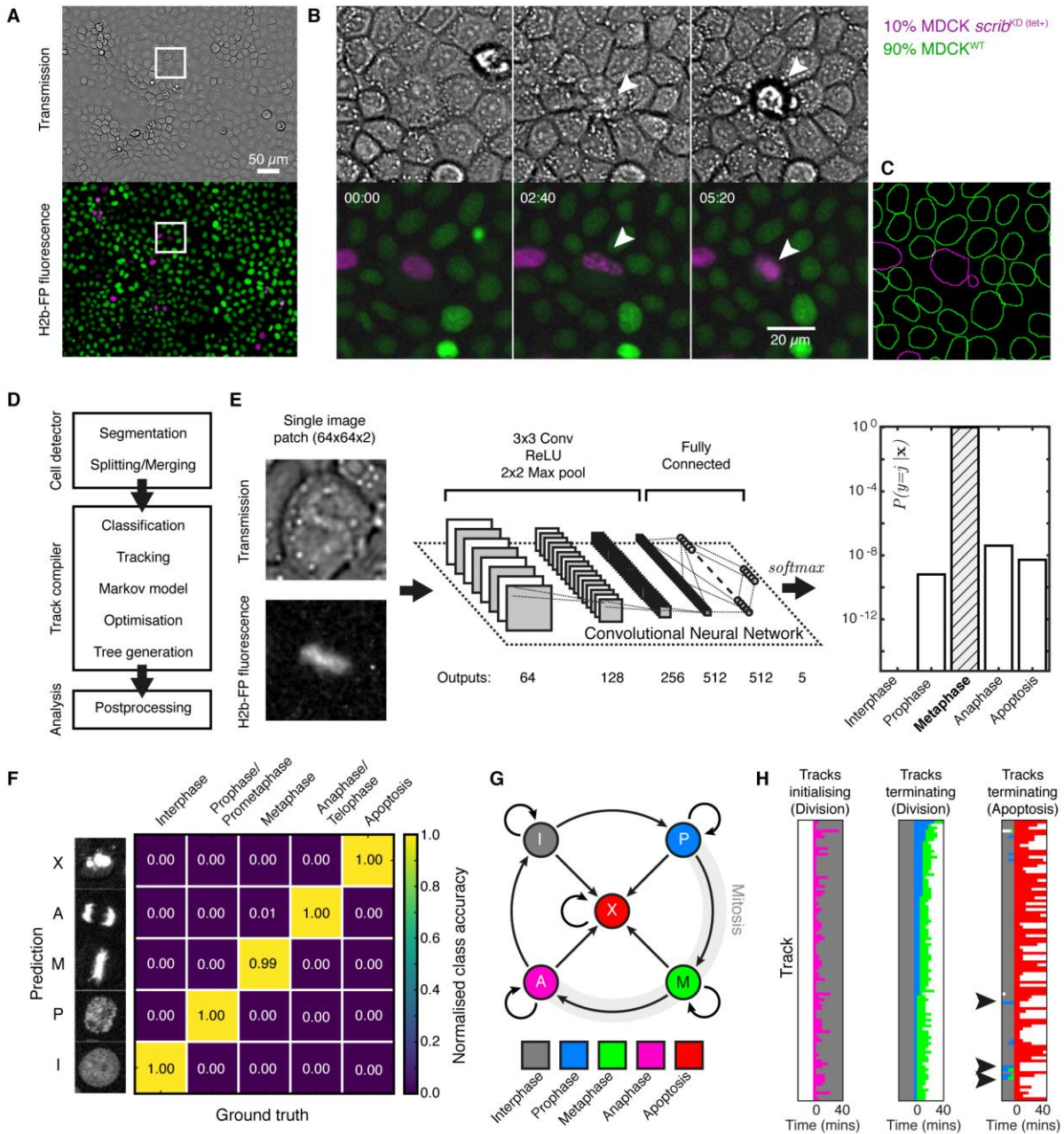
Figure 1: High-throughput image analysis pipeline for long-term analysis of single cell dynamics. A Representative image containing H2b-GFP MDCK[WT] (green) and H2b-RFP scribble[kd] (magenta) cells mixed in an initial 90:10 ratio. The image corresponds to one field of view (1600 x 1200 pixels, 530μm x 400μm) acquired by wide-field epifluorescence using 20x magnification. The region of interest in the white rectangle is shown in B. scale bar = 50μm. **B** Time series from the competition assay in the region boxed in A. The white arrow indicates a scribble[kd] cell surrounded by MDCK[WT] neighbours that undergoes apoptosis. The acquisition of both transmission and fluorescence images enables detection of the apoptotic fragmentation of the loser's nucleus happening prior to extrusion of the cell body from the monolayer. Timings are indicated in the top left corner in hours and minutes. **C** Segmentation of the final image in B. MDCK[WT] cells are outlined in green and scribble[kd] cells in magenta. **D** Flow chart of the computational pipeline implemented for the study of competition dynamics. The strategy is based on segmentation of individual cells (cell detector), automatic annotation of morphological classes related to cell cycle state and apoptosis (track compiler), and post-processing analysis of single-cell tracking data. **E** Convolutional Neural Network architecture (CNN) for object classification. The CNN inputs are single-object patches, both in the transmission (BF) and fluorescence channels (Left). The CNN stacks together four types of layers: convolutional/ReLU/max-pooling and fully connected layers (middle). The CNN transforms the input image layer by layer from the original pixel values to the final class scores with the highest score reflecting the most probable classification of the image data (right). **F** Confusion matrix showing the matching of human annotations versus the annotation of the CNN system. **G** Hidden Markov model (HMM) used for modelling progression through the cell cycle. The figure depicts the permitted directional transitions between five classes (interphase, prometaphase, metaphase, anaphase and apoptosis). **H**

22

Automated annotation of cell trajectories over time. A random selection of 100 trajectories (rows) is aligned and shown over a 40 minute period. Colours refer to state labels as defined in F. Left: Tracks following division start with anaphase before proceeding to interphase. Middle: Tracks terminating in a division are preceded by interphase before going through prometaphase and metaphase. Right: Tracks terminating with apoptosis are often preceded by interphase, but can arise through failed division events (highlighted with arrows).
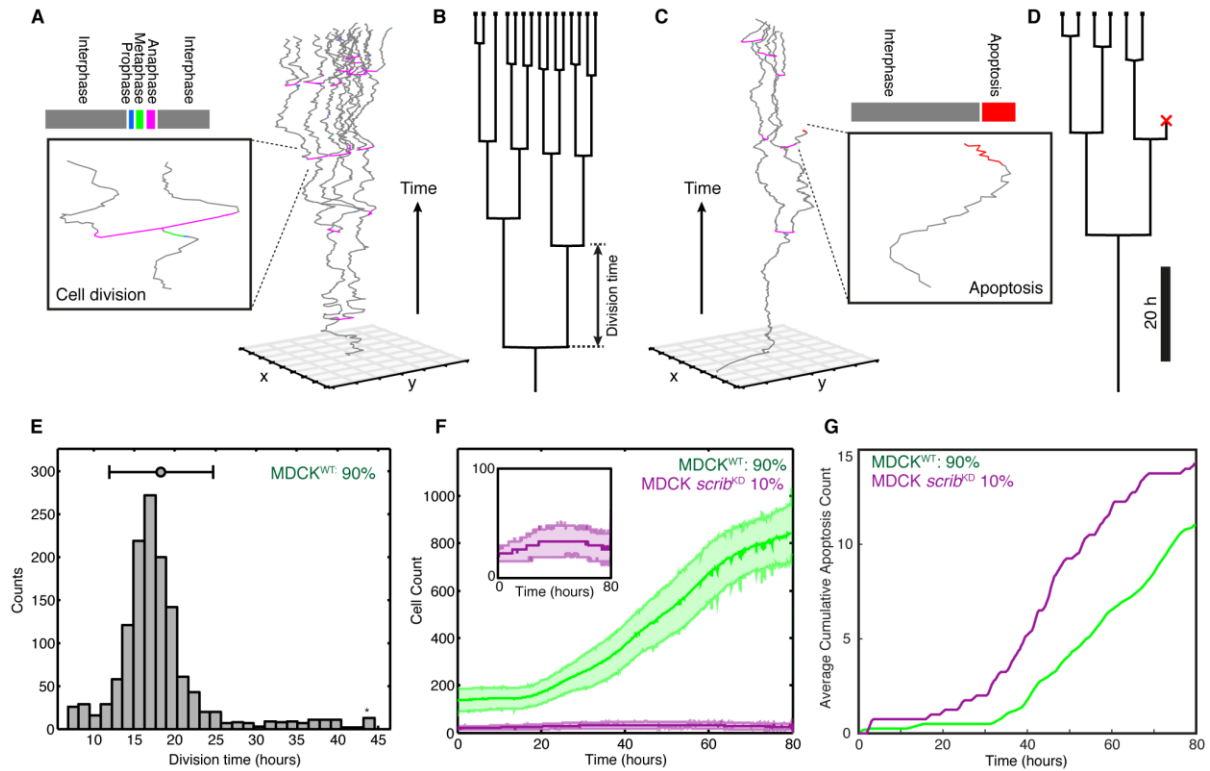
Figure 2: Cell lineage, cell cycle duration and apoptosis detection during cell competition. A Representative cell tracks assembled into lineages trees as result of the global optimization. The original progenitor cell is at the bottom of the diagram, time is on the z axis and the position of all progeny is plotted in the x-y plane. Tracks in which the metaphase to anaphase transition is present are split, and labelled as division events. In our classification, the metaphase state corresponds to a branching point, and new tracks start with the anaphase state. Each cell is assigned a unique cell ID (not shown in the figure). B Representative lineage tree for the cells shown in A. Cell cycle time can be measured for each individual cell as the time between consecutive division events on the lineage tree. C Representative cell tracks assembled in to a lineage tree, showing a single apoptotic termination event, following multiple observations of the apoptotic state. The apoptotic states are labelled in red. D The lineage tree of the cells shown in C. E Histogram of cell-cycle time measured for a population of MDCK^{WT} showing a mean value of 18± 3.2 h. The distribution represents a population of at least 250 cells in the same field of view followed for 80h. A total of 1326 division events were observed. F Proliferation profiles of MDCK^{WT} (green) and scribble^{kd} (magenta) during a competition. Data are pooled from three biological replicates imaging 4 fields of view for each replicate. The solid line indicates the mean of the experiments and the shaded area indicates the standard deviation. The inset shows the evolution of cell count in the scribble^{kd} cells on a smaller scale. G Quantification of apoptotic events for MDCK^{WT} (green) and scribble^{kd} (magenta) during competition. The number of apoptoses is detected and averaged across the 4 areas imaged during one competition assay.
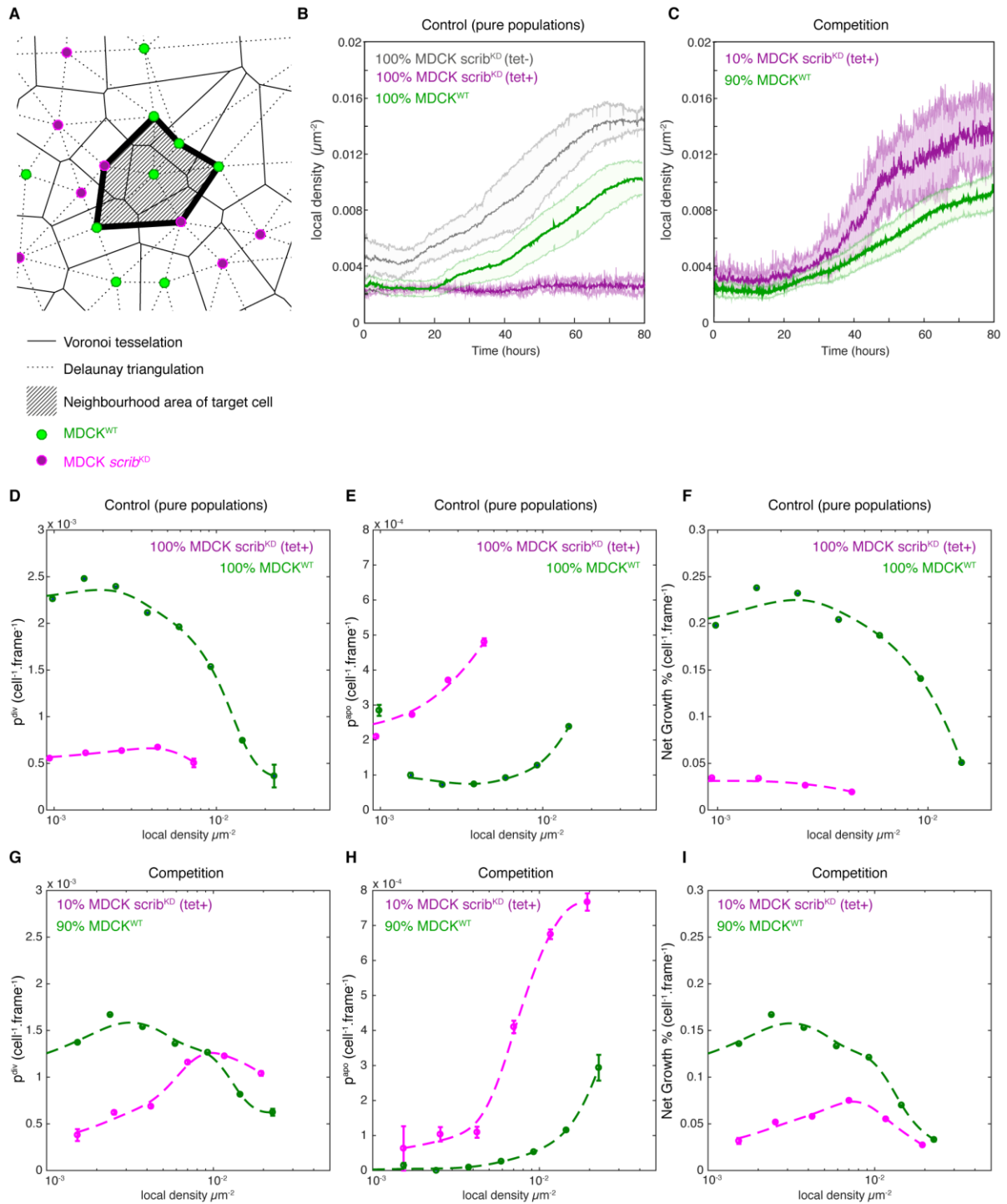
Figure 3: Probability of apoptosis and division are sensitive to local density in competitive conditions. A Definition of local density. The centre of the nucleus of each cell is represented by a filled circle (green: MDCK[WT] and magenta: scribble[kd]). Nuclei centres are used to construct a Delaunay triangulation (dashed lines) and a Voronoi tesselation (solid lines). Local density was defined as the sum of inverse areas of the triangles sharing a common vertex at the nucleus of the cell of interest. The area taken into account in this example is shown in grey. **B-C** The local cellular density was computed for each cell ID and averaged among cells of the same lineage at each time point. The solid line indicates the mean of the experiments and the shaded area indicates the standard deviation. Data are pooled from three biological replicates. **B** Temporal evolution of local density for pure populations of MDCK[WT] (green), non-induced scribble[kd] (tet-, black), and induced scribble[kd] (tet+, magenta). **C** Temporal evolution of local density for MDCK[WT] (green) and scribble[kd] (magenta) seeded at a 90:10 ratio during a competition assay. **D-E-F** Probability of division (p^{div}), apoptosis (p^{apo}), and net growth ($p^{div} - p^{apo}$) per cell per frame as function of local density computed for induced scribble[kd] (tet+, magenta) and MDCK[WT] pure populations. **G-H-I** Probability of division, apoptosis, and net growth per cell per frame for MDCK[WT] and induced

scribble[kd] during competition. Data are pooled from 8 fields of view from two biological replicates. **D-I** Data points are indicated by solid circles. Each data point is computed from N>500 observations. Trend lines computed using smoothing splines are plotted as dashed lines. In panels D-F, the mean local density used to define the bins is 2.6×10^{-3} μm^{-2} for scribble[kd] cells and 6×10^{-3} μm^{-2} for MDCK[WT]. In panels G-I, the mean local density value is 6×10^{-3} μm^{-2} for both cell types. For each data point, whiskers indicate the inverse of the number of observations as an estimate of accuracy in determining the probability.
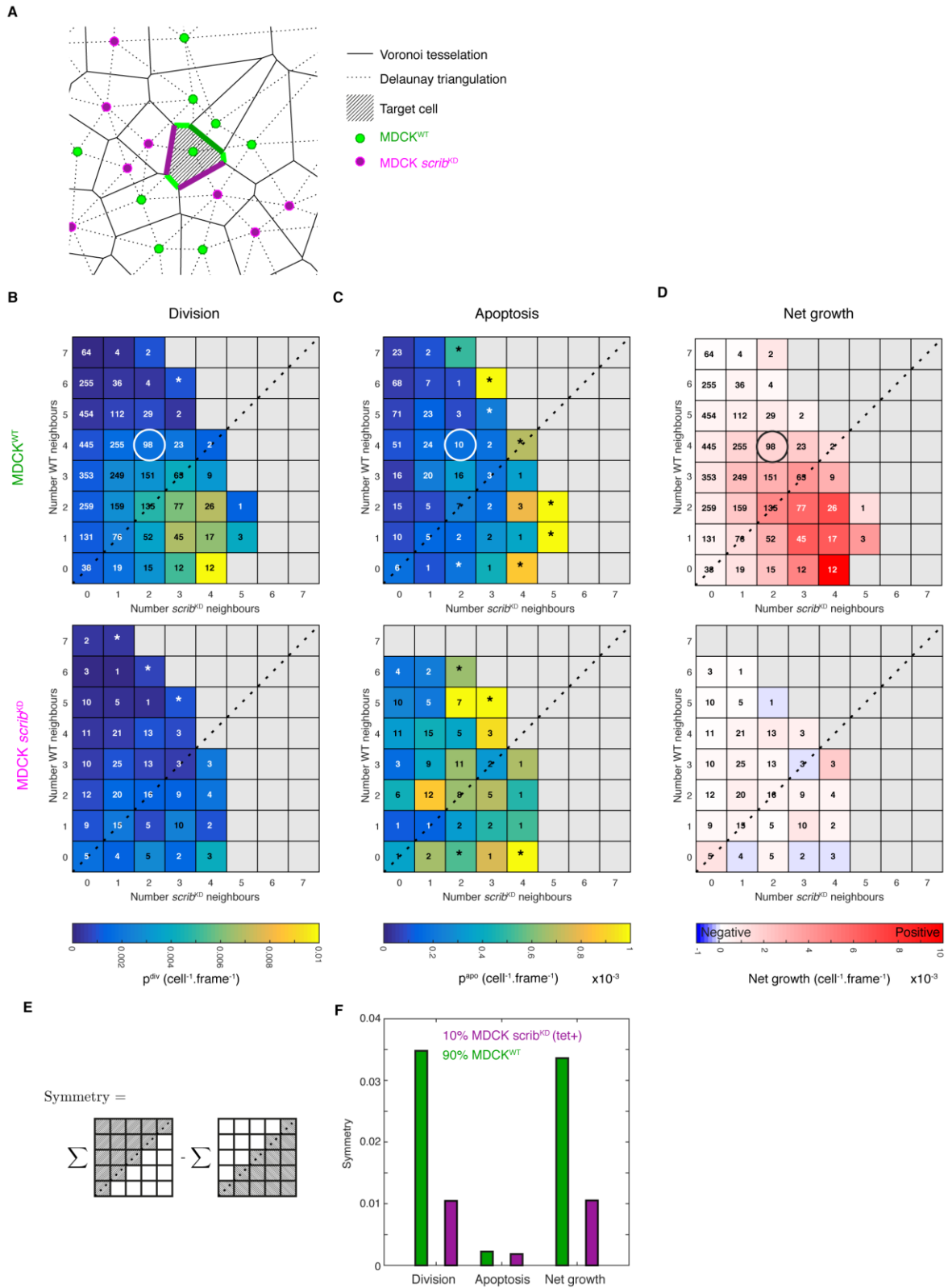
Figure 4: Probability of apoptosis, division, and net growth are sensitive to the composition of local neighbourhood. A Definition of local neighbourhood. The centre of the nucleus of each cell is represented by a filled circle (green: MDCK[WT] and magenta: scribble[kd]). Nuclei centres are used to construct a Voronoi tesselation (solid lines). To determine the composition of local neighbourhood, we determine how many intercellular junctions the cell of interest (grey) has in common with the MDCK[WT] cells (junctions shown in green) or scribble[kd] cells (junction shown in magenta). The position of the cell of interest depicted here is shown by the grid position circled with the white frame in **B** in the neighbourhood plot for MDCK[WT] cells. **B-C-D** Neighbourhood plots showing the colour coded probability per cell per frame of Division,

Apoptosis and Net-growth for MDCK^{WT} (top) and scribble^{kd} (bottom) during competition. The number of scribble^{kd} neighbours is shown on the x-axis and the number of MDCK^{WT} cells is shown on the y-axis. The diagonal (black dashed line) indicates grid positions with equal numbers of MDCK^{WT} and scribble^{kd} neighbours. Numbers in each grid position indicate the number of detected events (division/apoptosis). The total number of observations for each grid position is shown in **Fig S4**. Measurements for each grid position are typically computed from observations of >500 cells. The probability for each grid position is defined as p=N_{events}/$N_{observations}$. Grid positions for which many observations were made but no event detected are marked by an asterisk and coloured as 1/$N_{observations}$ to provide an upper-bound for the probability in that position. Data are pooled from 12 time-lapse movies from three biological replicates. **E** Definition of *s* as parameter used for calculating the symmetry of neighbourhood plots around the diagonal. For each neighborhood plot, *s* is computed as difference between the sum of the lower quadrant and the sum of the upper quadrant for grid positions ranging from [0,0] to [4,4] inclusive. **F** Symmetry calculation performed on checkerboard plot shown in **B-D**. The symmetry defines whether the behaviour is cell-autonomous (low *s* values) or dependent on the cell-type of neighbours (high *s* values).
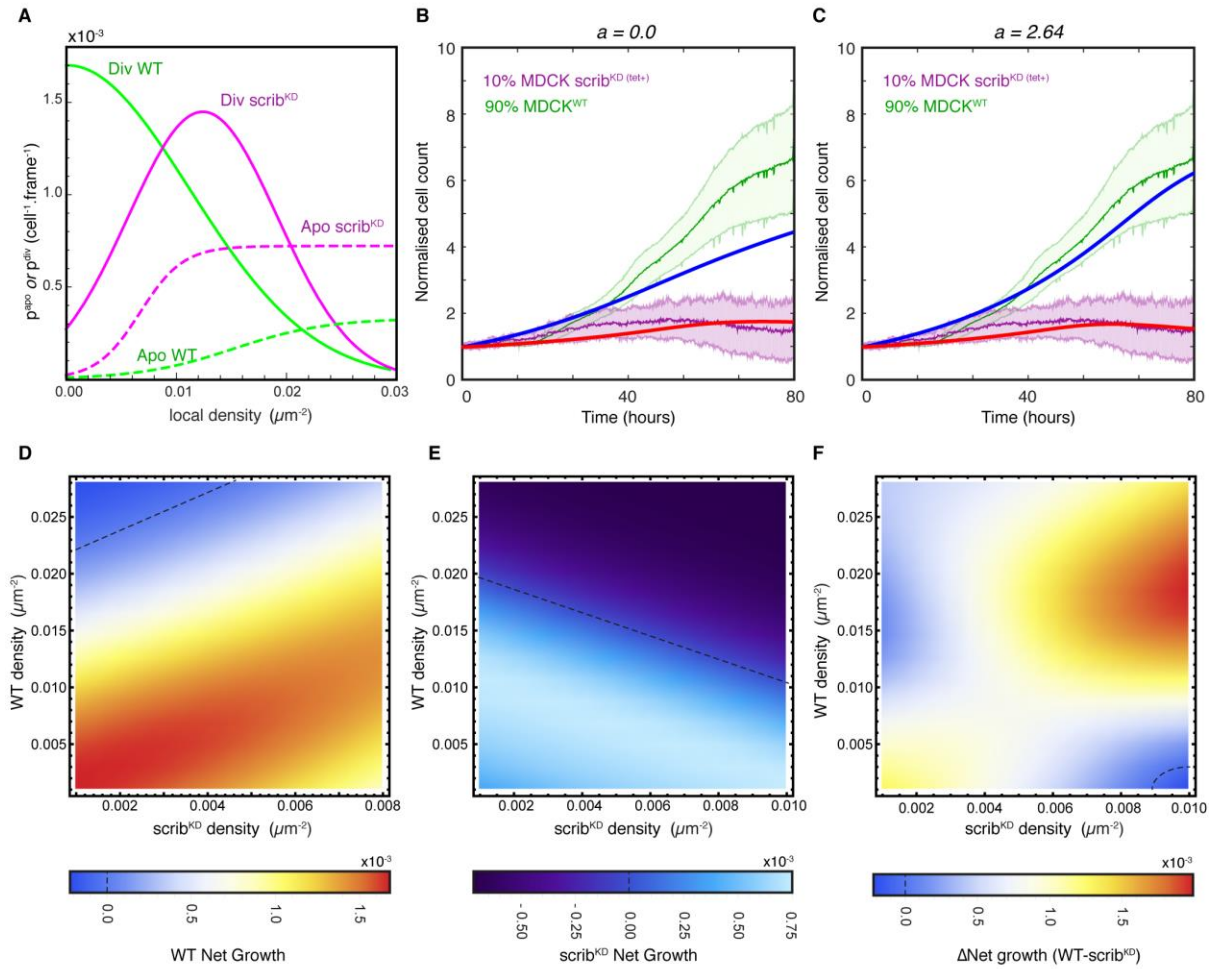
Figure 5: A density dependent rate equation model for cell competition. A Probability (per unit time per cell) of apoptosis (dashed lines) and probability of division (solid lines) for MDCK[WT] (green lines) and scribble[kd] (magenta lines) cells determine the rates f_{apo} and f_{div} for our model. These functions were fitted to the experimental data in Fig **3G-I, S5A-D**. **B** Temporal evolution of cell count predicted by the symmetric interaction model (a=0.0) initialised with the mean experimental cell count at t=0 for MDCK[WT] (solid blue line) and scribble[kd] (solid red line). The model curves are overlaid with the experimental cell count from Fig **2D** for MDCK[WT] (green) and scribble[kd] cells (magenta). **C** Temporal evolution of cell count predicted by the asymmetric interaction model (a=2.64) initialised with the mean experimental cell count at t=0 for MDCK[WT] (solid blue line) and scribble[kd] (solid red line). **D-E** Heat maps of net growth as a function of local density of MDCK[WT] cells on the y-axis and local density of scribble[kd] cells on the x-axis for the asymmetric interaction model for MDCK[WT] cells (a=2.64, D) and scribble[kd] cells (a=2.64, E). Warm colours indicate high net growths while cold colours indicate low net growths. Dashed line indicates the contour of zero net growth, or density homeostasis. **F** Relative fitness landscape for the asymmetric density dependent model (a=2.64), defined as the net growth of the MDCK[WT] cells minus the net growth of the scribble[kd] cells. MDCK[WT] cells have a higher fitness than scribble[kd] cells, except within the region delineated by the dashed line.

Supplemental Materials
Molecular Biology of the Cell

Bove et al.

Local cellular neighbourhood controls proliferation in cell competition

Supplementary Information

Anna Bove*[1,2], Daniel Gradeci*[1,3], Yasuyuki Fujita[4], Shiladitya Banerjee[#3,5], Guillaume Charras[#1,2,5], Alan Lowe[#1,6]
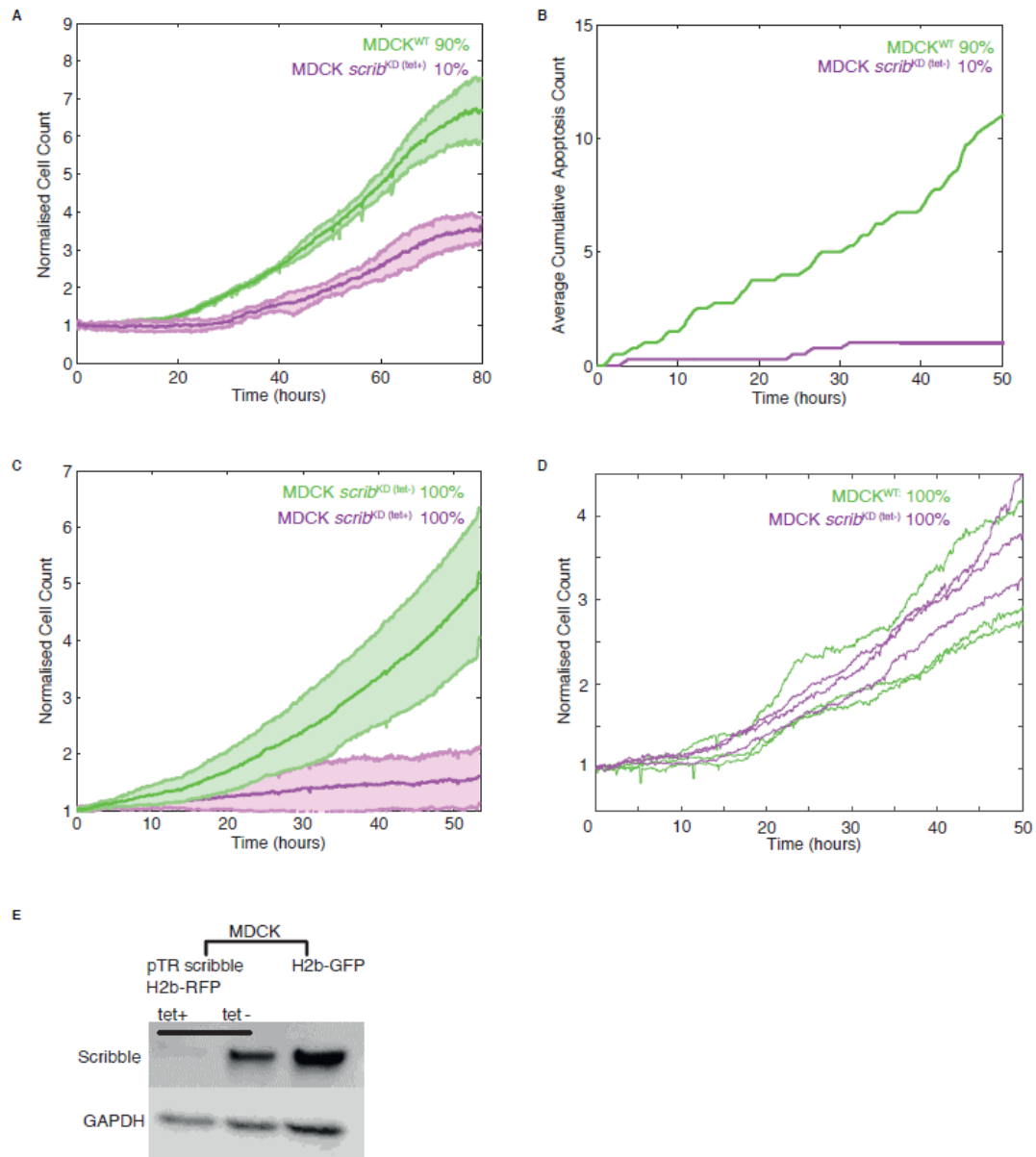
Figure S1 Proliferation and apoptosis quantification in non-competitive conditions. A Proliferation profiles of MDCK [WT] (green) and non-induced scribble[kd] (magenta) mixed in 90:10 relative ratios. Data are normalized to the initial cell count and pooled from two biological replicates imaging 4 fields of view for each replicate. The solid line indicates the mean of the experiments and the shaded area indicates the standard deviation. **B** Quantification of apoptotic events for MDCK [WT] (green) and non-induced scribble[kd] (magenta) mixed in 90:10 relative ratios. The number of apoptoses is detected and averaged across the four areas imaged during one control experiment. The number of apoptoses for MDCK [WT] and non-induced scribble[kd] cells evolved in proportion to the initial seeding ratio. **C** Quantification of growth of pure non-induced scribble[kd] (tet-, green) and induced scribble[kd] (tet+, magenta) populations. Plots are normalised to the initial cell count and are pooled from three biological replicates, imaging three fields of view each. The solid line indicates the mean of the experiments and the shaded area indicates the standard deviation. **D** Proliferation profiles of non-induced scribble[kd] (tet-, magenta) and MDCK [WT] (green) cells in pure populations. The three lines show the cell count in each field of view imaged during one control experiment. **E** Immunoblot showing scribble and GAPDH expression for induced scribble[kd] cells (tet+), non-induced scribble[kd] cells (tet-), and MDCK [WT] cells expressing H2b-GFP. Induction of scribble shRNA expression results in a significant decrease in scribble expression levels.
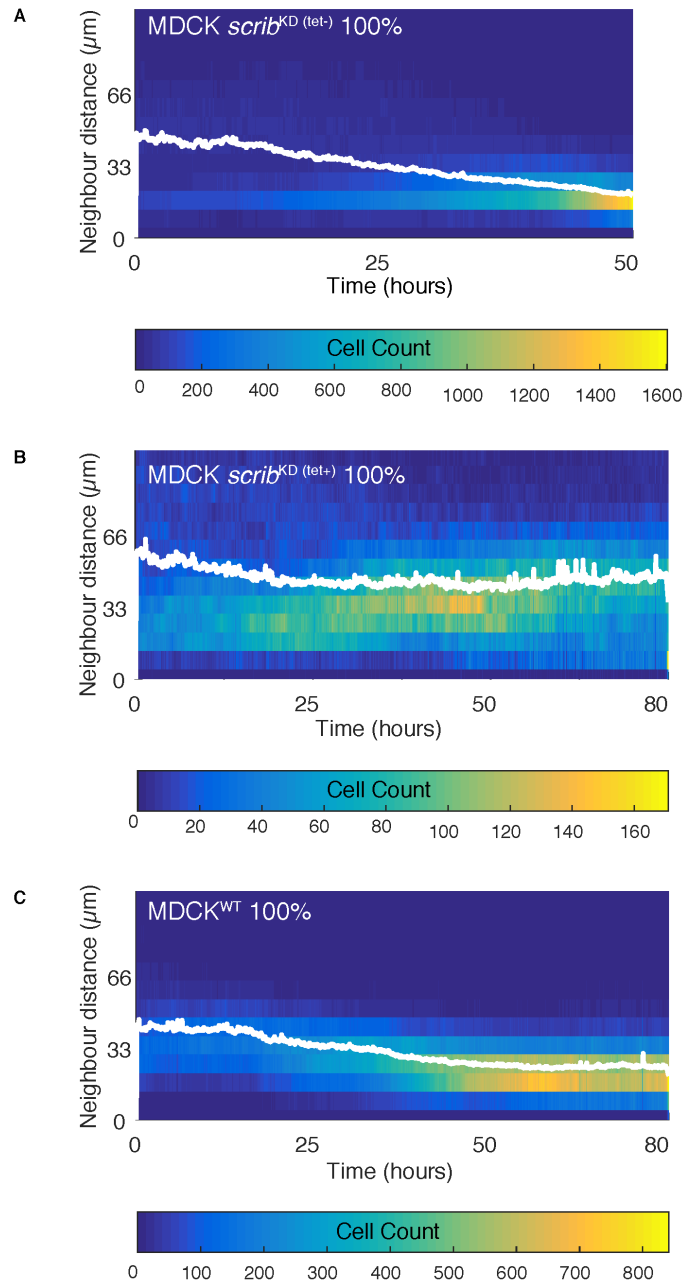
Figure S2: Plot of inter-nuclear separation over time. Two-dimensional histograms of the time evolution of edge lengths in the Delaunay graph constructed from cell centroids. The mean value of edges lengths over time is displayed as solid white line. The time evolution is used to calculate the threshold value (D_{thresh}) for interacting cells used in neighbourhood estimation. We determined D_{thresh} to be **A** ~30 μm for non-induced MDCK scribble[kd] (tet-), **B** ~60 μm for MDCK scribble[kd] (tet+) in which knock-down of scribble had been induced, and **C** ~30 μm for MDCK[WT]. These threshold values correspond to the respective average cell diameter in pre-confluency conditions, which describes the maximum distance between two cells in contact. Any distance larger than this would imply the presence of free space between cells, signifying they cannot be classified as neighbours. The plots in Figure S2 show that the average inter-nuclear distance is larger the threshold values at low densities due to the presence of free-space between some of the cells in the triangulation. This distance shortens over time below the threshold, as cell numbers increases and the cells reach confluency.
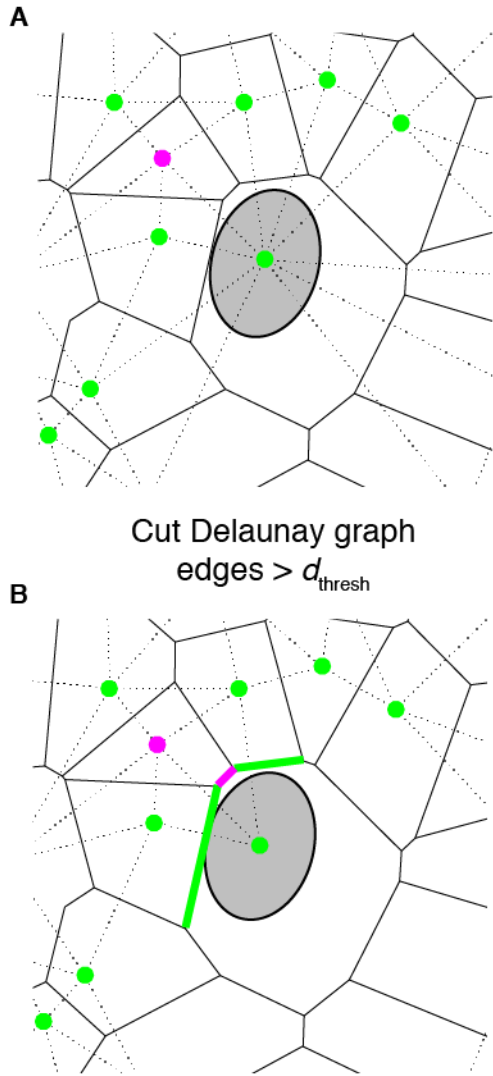
A

Cut Delaunay graph
edges > d_{thresh}

B

Figure S3: Example of neighbourhood estimation before cells have reached confluence. **A** True neighbours are those that share a common edge with the Voronoi polygon of the target cell (centre, grey ellipse MDCK^{WT}), and whose separation distance from the target cell is below the experimentally determined D_{thresh} value (**Fig S2**). In the diagram, the dotted lines link the centroid of neighbours. **B** In this case the target cell has three neighbours, two MDCK^{WT} (green) and one scribble^{KD} (magenta). In this diagram, neighbours with separation distances above D_{thresh} have been removed.
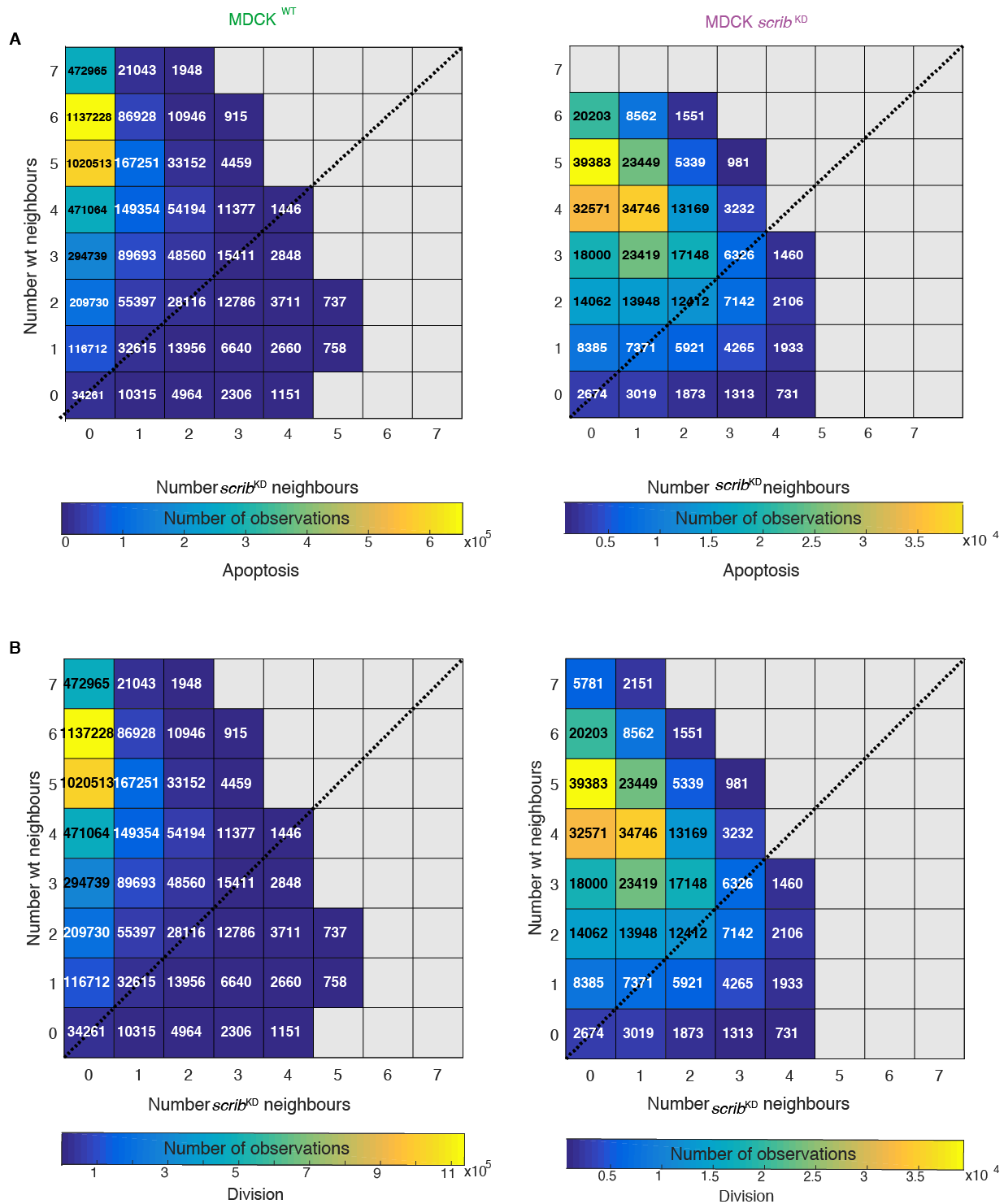
Figure S4: Number of observations in each position of neighbourhood plots. (A,B) Neighbourhood plots showing the colour coded number of observations (cell per frame) for MDCK[WT] (left panels) and scribble[kd] (right panels) during competition. The number of scribble[kd] neighbours is shown on the x-axis and the number of MDCK[WT] neighbours is shown on the y-axis. The diagonal (black dashed line) indicates grid positions with equal numbers of MDCK[WT] and scribble[kd] neighbours. Numbers in each grid position indicate the number of observations for MDCK[WT] and MDCK scribble[kd] cells with that particular neighbourhood. Data are pooled from 12 time-lapse movies from three biological replicates.
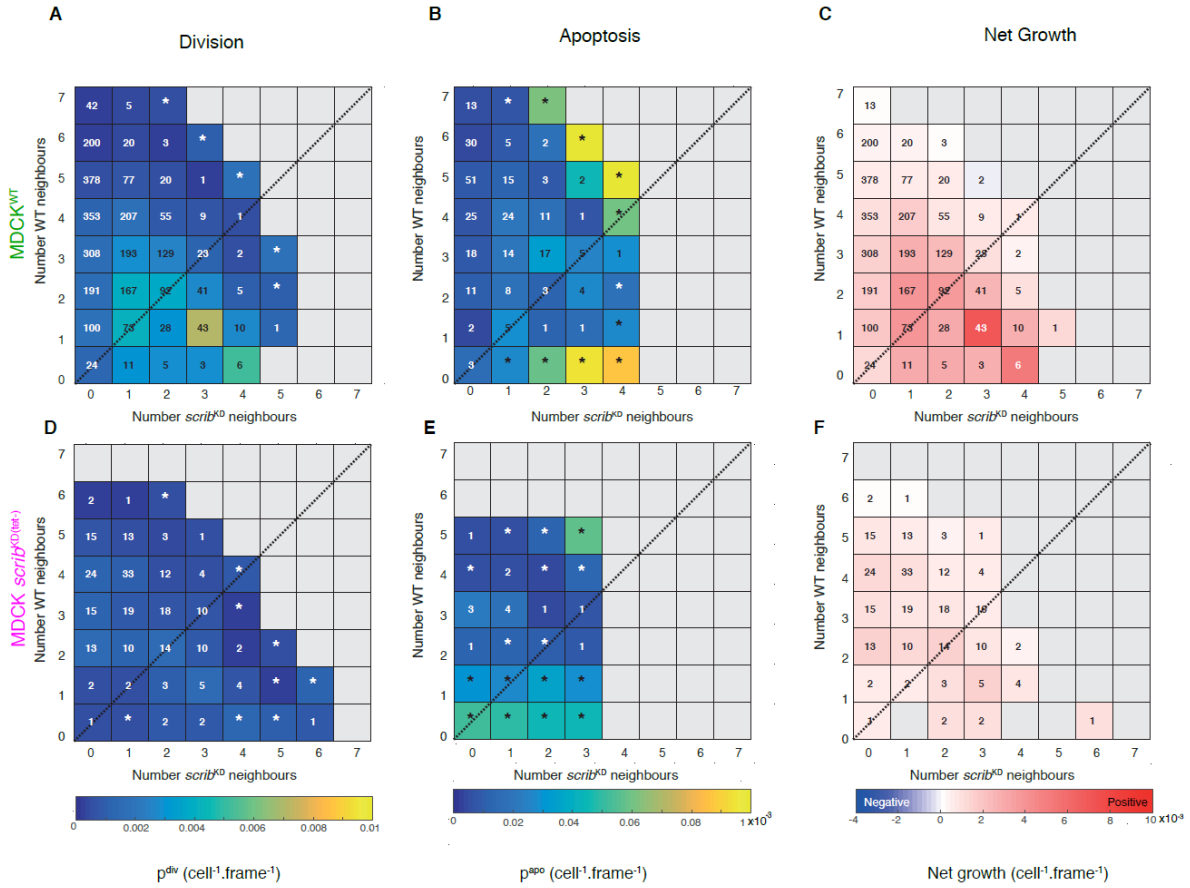
Figure S5: Neighbourhood plots for apoptosis, division, and net growth probabilities in non-competitive conditions. Neighbourhood plots showing the colour coded probability per cell per frame of division, apoptosis and net-growth for MDCK[WT] (**A,B,C**) and non-induced scribble[kd] (**D,E,F**) mixed in 90:10 relative ratios. Data are pooled from 7 time-lapse movies from two biological replicates. Colour scales are the same as used in Fig 4.
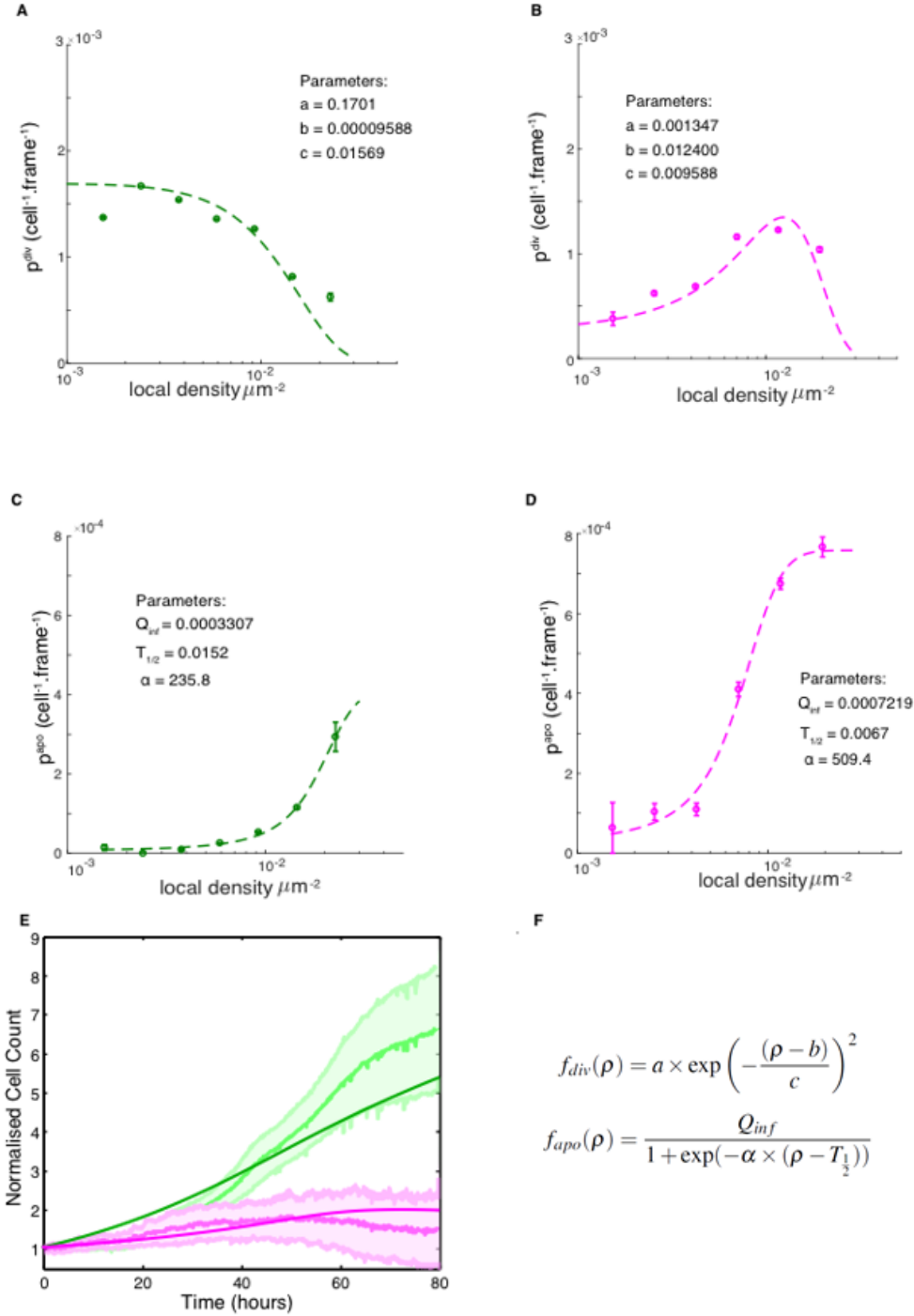
Figure S6: Analytical functions for fitting the density dependence of apoptosis and cell division. (**A,B**) Probability of division for MDCK[WT] and scribble[kd] as a function of local density fitted with Gaussian functions (dashed line) used in the rate equation model with parameters relating to equation (1) in **F**. (**C,D**) Probability of apoptosis for MDCK[WT] and scribble[kd] as a function of local density fitted with Logistic functions (dashed line) used in model with parameters relating to equation (2) in **F**. **E** Temporal evolution of cell count predicted by the uncoupled model (*a*=1) initialised with the mean experimental cell count at t=0 for MDCK[WT] (solid green line) and scribble[kd] (solid magenta line), overlaid with experimental cell count from for MDCK[WT] (green) and scribble[kd] cells (magenta). **F** Equations used for fitting the experimental data relating probability of apoptosis and division to local density. Equation 1 is used to fit graphs A, and B. Equation 2 is used to fit graphs C and D. On (**A-D**), the values of the parameters leading to the best fit are indicated on the graph.
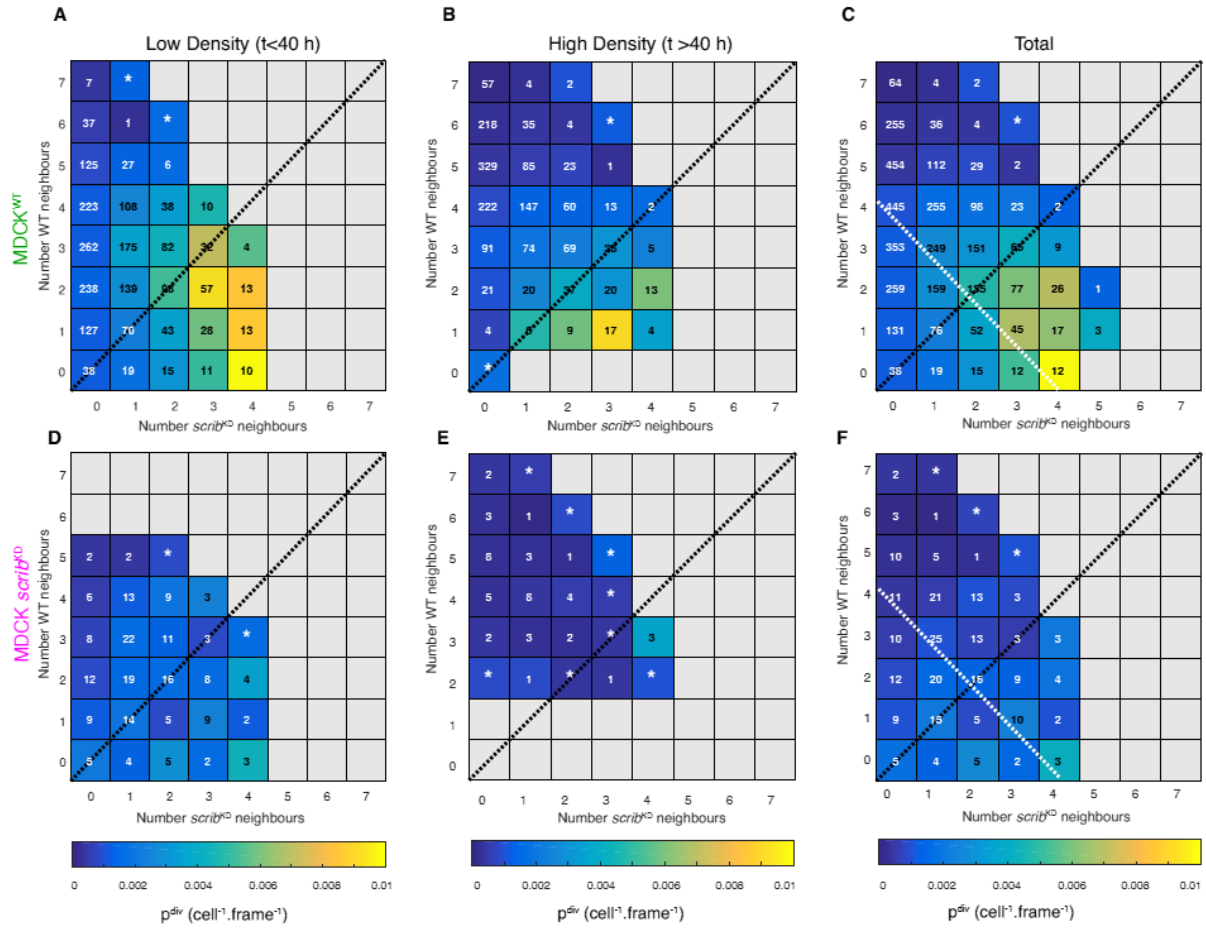
Figure S7: Time-resolved neighbourhood analysis Neighbourhood plots showing probability of division for MDCK[WT] (**A, B**) and Scribble[kd] (**D,E**). Data were split between low (t<40 h, **A,D**) and high (t>40 h, **B,E**) density; the resulting plots are displayed alongside with neighbourhood plots taking into account the whole duration of the experiment as shown in the original version of the manuscript (**C,F**). (**A-F**) The diagonal (black dashed line) indicates grid positions with equal numbers of MDCK[WT] and scribble[kd] neighbours. (**C,F**) The white dashed line separates grid positions representative of the low density (lower quadrant) from those representative of high density (upper quadrant).
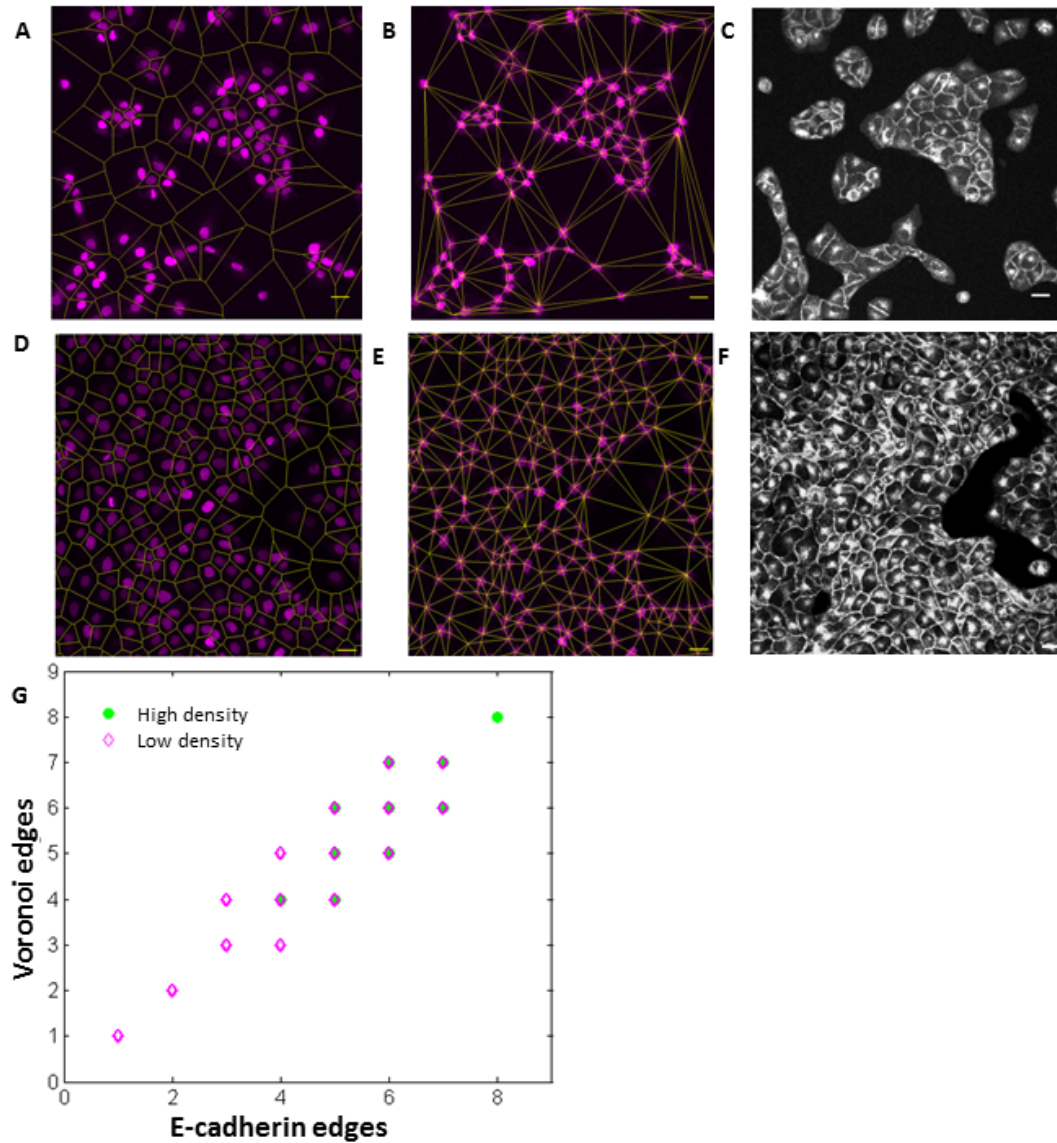
Figure S8: Validation of neighbourhood analysis using the Voronoi method. We assessed the accuracy of our neighbourhood and density analysis by comparing the number of neighbours determined from the Voronoi diagram (number of Voronoi edges) to the number of neighbours determined manually. In order to do so, we fixed and imaged MDCK cells stably expressing the H2B nuclear marker together with an E-cadherin marker, which localizes at adherent junctions. We considered both high and low density. High density (**D,E**) and low density (**A,B**) images of MDCK stably expressing H2b-RFP (magenta, **A-B, D-E**) and E-cadherin GFP (grey, **C,F**). Scale bar = 25µm. We applied the Voronoi tessellation and Delaunay triangulation algorithms to the nuclei images (**A-B, D-E**). The Voronoi diagram and the Delaunay triangulation are highlighted in yellow. Finally, we manually scored the number of E-cadherin delimited edges for each cell as a ground truth. Next, we determined the correlation between the edges detected via the Voronoi method and manually, at high and low densities. We plotted the number of neighbours determined using the Voronoi method on the y-axis as a function of the number of manually determined number of neighbours on the x-axis. We calculated the Pearson coefficient (r) and obtained r=0.832 for high density and r=0.951 for low density, indicating that the Voronoi method provides a reliable estimate of the number of neighbours.

Supplementary Information: Local cellular neighbourhood controls proliferation in cell competition

Anna Bove, Daniel Gradeci, Yasuyuki Fujita, Shiladitya Banerjee, Guillaume Charras, Alan R. Lowe

Supplementary Note 1

Several software packages have been developed for the purposes of single-cell tracking. These include tTy [1], CellProfiler [2], CellCognition [3] and LEVER [4] amongst others (reviewed in [1]). In general, they provide manual or semi-automated image segmentation and tracking over several division cycles, and have been, in most cases, tailored to the study of proliferation and differentiation processes.

Here we have additional experimental considerations in studying cell competition, in that we must accurately identify not only division events but also apoptotic events. Not only does this require robust cell tracking but also accurate identification of cell-cycle state. In our experience, previous software packages tend to over-estimate or incorrectly assign cell division events to account for apoptosis events, or prove to be too laborious for annotating the vast amount of data generated during imaging. Therefore, we developed our own software that takes advantage of recent advances in machine learning (in particular deep learning) and optimized our tracking algorithms to account for cell death as well as proliferation events. In this way we have created a computational strategy for characterising apoptosis, cell division, and net growth as a function of local cell neighbourhood.

To segment, track and classify cell-cycle state and apoptosis automatically, we implemented a novel multi-layered computational pipeline that first detects individual cells and then classifies their cell-cycle stage based on the combination of cell and histone morphology (**Fig 1C-D**).

Image segmentation. In the Cell Detector step (**Fig 1D**), fluorescence images are segmented into foreground (cells) and background. After having acquired time-lapse movies of cells using the incubator microscope, we restore the images by

performing flat-field illumination correction and removing CCD hot pixels. Following image restoration, segmentation of the fluorescence images was performed using a Gaussian Mixture Model (GMM). Briefly, the combined intensity histogram of three images taken from the beginning, middle and end of the movie were fitted to a GMM using the Expectation Maximisation (EM) algorithm to learn the appropriate parameters [5,6]. The intensity distribution was described as a weighted sum of n normal distributions:

$$P(\mathbf{x} \mid \theta) = \sum_{k=1}^{n} \lambda_k N_k(\mu_k, \sigma_k^2) \tag{1}$$

where θ represents the learned parameters for the n models: λ_k is the normalised weight, μ_k the mean intensity and σ_k^2 the variance for each normal distribution in the mixture model. We typically used n=3, and separate parameters were learnt for the GFP and RFP fluorescence movies. In general, when ordered by increasing μ_k, the three normal distributions reflect the intensity distributions of background, interphase and mitotic/apoptotic cells. The output of the segmentation method is a binary classification of the image into background and cells. Dense regions of cells were separated using either a marker controlled watershed transform, a custom written object splitting algorithm based on calculating regions of concavity in convex objects [7] or a hybrid of both methods.

Next we use an additional merging step to recombine fragments arising from over-segmentation of nuclei with a weak-fluorescence signal. We developed an algorithm that attempts to find the best possible hypothesis for merging the objects, based on separation distance and image features. This works in several phases. First, a Delaunay graph is calculated to make putative clusters of fragments. Second, hypotheses for combinations of fragments constituting a single object are constructed. Each hypothesis has an equal prior probability. Third, successive Bayesian updates are performed using the separation distance and image feature information. Finally, the algorithm selects the merging hypotheses with the highest posterior probabilities. This algorithm has the advantage of not merging apoptotic fragments with non-apoptotic nuclei.

Object classification using deep neural networks. In the Track compiler step, each nucleus is assigned a type (scribble^{kd} or MDCK^{WT}) according to its fluorescence emission and is classified into one of five different classes reflecting the cell cycle or apoptotic-state. The five states are: interphase, pro(meta)phase, metaphase, anaphase/telophase or apoptosis. Automatically assigning class labels to images (image classification) is a task well suited to machine learning approaches. In recent years, a deep learning approach known as convolutional neural networks (CNN) have demonstrated great utility with significantly improved ac-

2

curacy over conventional approaches.

We implemented a deep CNN to automatically perform the cell-cycle state assignment of each detected cell (**Fig 1E**). A fully trained CNN takes a multi-dimensional image as input (such as the transmission and fluorescence images in our case, **Fig 1E**, left) and transforms it through a sequence of operations (**Fig 1E**, middle) that yields an output label. Our CNN architecture (**Fig 1E**) was broadly based on the LeNet-5 architecture [8], and consists of several layers of 3x3 convolution, rectified linear units (ReLU) [9] and 2x2 max-pooling units [10], which decrease spatial dimensionality and increase the number of filters. In this architecture, early layers of the network are sensitive to small image features (a small receptive field), and later layers to gross morphology in the image (larger receptive field). The combined effect is that the network utilises information from multiple spatial scales. Next, several fully connected layers reduce the output of the internal layers into a mapping to the output label - in our case, the five cell-cycle classes. The final output represents the probability of the image belonging to each of the five classes that we define (**Fig 1E**, right). The classification label is simply the label with the highest probability.

Training the neural network. The deep CNN needs to be trained to perform this task. We generated a ground truth dataset for training, by hand annotating thousands of regions of interest (ROI), using both transmission (BF) and fluorescence channels (**Fig 1E**, left). Each ROI, centered on the nucleus of the cell, was assigned one of five cell-cycle states. From this training set, we generated three subsets for: (i) training, (ii) testing, and (iii) validation. Each set had an identical number of training examples, with their order shuffled and the class numbers balanced such that all classes had an identical number of occurrences. Care was taken to annotate images that represented the diversity of the dataset and were temporally non-sequential to prevent over-training. To increase the number of training examples, we augmented the examples by introducing random transformations, rotations and noise, yielding 15,535 training examples per class. We used dropout (50% while training) to prevent over-fitting. CNNs were implemented in *Caffe* [11] or *TensorFlow* [12]. Training was performed using a momentum optimizer with an exponentially decaying learning rate until convergence.

Having trained the deep CNN, we determined the performance by calculating the normalized accuracy for each class, presented as a confusion matrix. Here we confront the prediction of the CNN to ground truth using the validation dataset (**Fig 1F**). Correct predictions are located on the diagonal of the matrix and, following completion of training, the CNN possessed an accuracy > **99**% for all states based on instantaneous image features. Such performance is remarkable particularly for the detection of the apoptotic state, which possesses a far broader

distribution of image features than other states.

We also trained a classifier based on a non-linear Support Vector Machine (SVM) [13] with a Radial Basis Function (RBF). We used image features such as fluorescence intensity, intensity gradient, Histogram of Oriented Gradients (HoG) features [14], orientation, eccentricity and texture. Although the SVM performed well at cell-cycle state classification, it did not match the performance of the CNN, particularly with apoptosis detection, with a maximum accuracy of $\sim 80\%$ (Not shown). We utilised the CNN for all further data analysis.

Cell tracking. Next, classified and segmented objects are assembled into tracks. The tracking algorithm assembles reliable sections of track that do not contain cell division events (tracklets). Each new tracklet initiates a probabilistic model in the form of a Kalman filter [15], and utilises this to predict future states (and error in states) of each of the objects in the field of view. We assign new observations to the growing tracklets (linking) is performed by evaluating the posterior probability of each potential linkage from a Bayesian belief matrix for all possible linkages [16]. The best linkages are those with the highest posterior probability. Despite the high instantaneous accuracy of the CNN classification, occasional errors occur. We correct errors using a temporal model of the cell cycle implemented as a Hidden Markov Model (HMM) [3] comprising interphase, the three states of mitosis, and a dead-end state of apoptosis (**Fig 1G**). Any tracklets containing a metaphase to anaphase transition are split into separate tracks so that they can be labeled as division events in later steps of the algorithm.

The tracklets are then assembled into lineage trees by using multiple hypothesis testing and integer programming [17,18] to identify a globally optimal solution. We build upon this previous work to incorporate hypotheses specific to apoptosis/extrusion and use additional geometric features and CNN classifications in the hypothesis generation. The following hypotheses are generated: (i) true positive track (ii) false positive track (iii) initializing at the beginning of the movie or near the edge of the FOV, (iv) termination at the end of the movie or near the edge of the FOV (v) a merge between two tracklets (vi) a division event or (vii) an apoptotic event. The likelihood of each hypothesis is calculated for some or all of the tracklets based on heuristics. The global solution identifies a sequence of high likelihood hypotheses that accounts for all observations. Having identified the global solution, the fates of each cell are updated, tracks are merged and lineage trees are generated using a Breadth First Search (BFS) to traverse the trees.

As displayed from the alignment of 100 randomly selected annotated trajectories followed over 40 minutes, the software is correctly assigning the anaphase state to new tracks initialized after a division event (**Fig 1H**, left). Likewise, trajectories terminating in a division event contain the correct class sequence motif

4

of Interphase-Prometaphase-Metaphase (**Fig 1H**, middle). Trajectories terminating in apoptosis often follow a prolonged interphase sequence, although in some circumstances appear to arise following a failed division event (**Fig 1H**, right).

The final output of the pipeline is a fully annotated time-lapse sequence yielding the centroid, cell type, state and lineage/fate of every cell. Although segmentation of the cell boundaries would provide useful additional information, we found that, in general, the result was error prone and not robust, due to the well-known difficulty in unambiguously determining cell boundaries in the absence of additional markers. In order to assess the overall quality of the tracking output we calculated the Multiple Object Tracking Accuracy (MOTA) [20], which yields a mean accuracy of $\mathbf{93.3 \pm 1.8}\%$ over the five datasets tested (Supplementary Note 2). Combined with the deep CNNs classification accuracy, these metrics suggest that we are able to accurately detect, classify and track thousands of cells over 80 hours of time-lapse imaging.

Supplementary Note 2

Quantitative assessment of cell tracking precision and accuracy. The accuracy of our image analysis pipeline can be thought of as being determined by several different sources of error:

1. Accuracy and precision of cell detection

2. Accuracy of cell state estimation

3. Accuracy of cell tracking

In this supplementary note we will quantify the accuracy and precision of detection and tracking (Items 1 and 3). We took five different short time-lapse sequences (consisting of 20-30 frames and $\geq$50 cells), and split these between three users to manually annotate the images, marking the cell centroids. We then used our pipeline to automatically track the cells. We define four different types of observations in the tracking output (**Fig S9**):

- **True positive** (TP, hit): This represents a true detection where the same object is found in the output and the manual annotation.

- **False positive** (FP, false alarm): These are object detections in the output that do not appear in the manual annotation.

- **False negative** (FN, miss): This represents a missed detection, not found in the output but present in the manual annotation.

- **Identity swap** (IS, mismatch): This is a tracking error where the unique identity of two tracks are swapped (perhaps by two objects crossing or by close proximity) or an incorrect linkage event, for example during cell division.

Next, we determined correspondences between the tracking output and the manual annotations. Briefly, a cost matrix representing the Euclidean distance between all pairs of detections and manual annotations is calculated. Matches are determined using the Jonker and Volgenant solver [19]. The solver finds the lowest cost solution, assigning a match (correspondence) between pairs of tracked objects and manually annotated objects that minimises the distance between them. Correspondences falling within a threshold distance limit are considered to be true matches. Although the global optimisation step of the tracker removes false positive tracks, we have included them in the analysis to give an accurate estimation of the errors in the segmentation steps. Finally, the tracks were inspected for identity
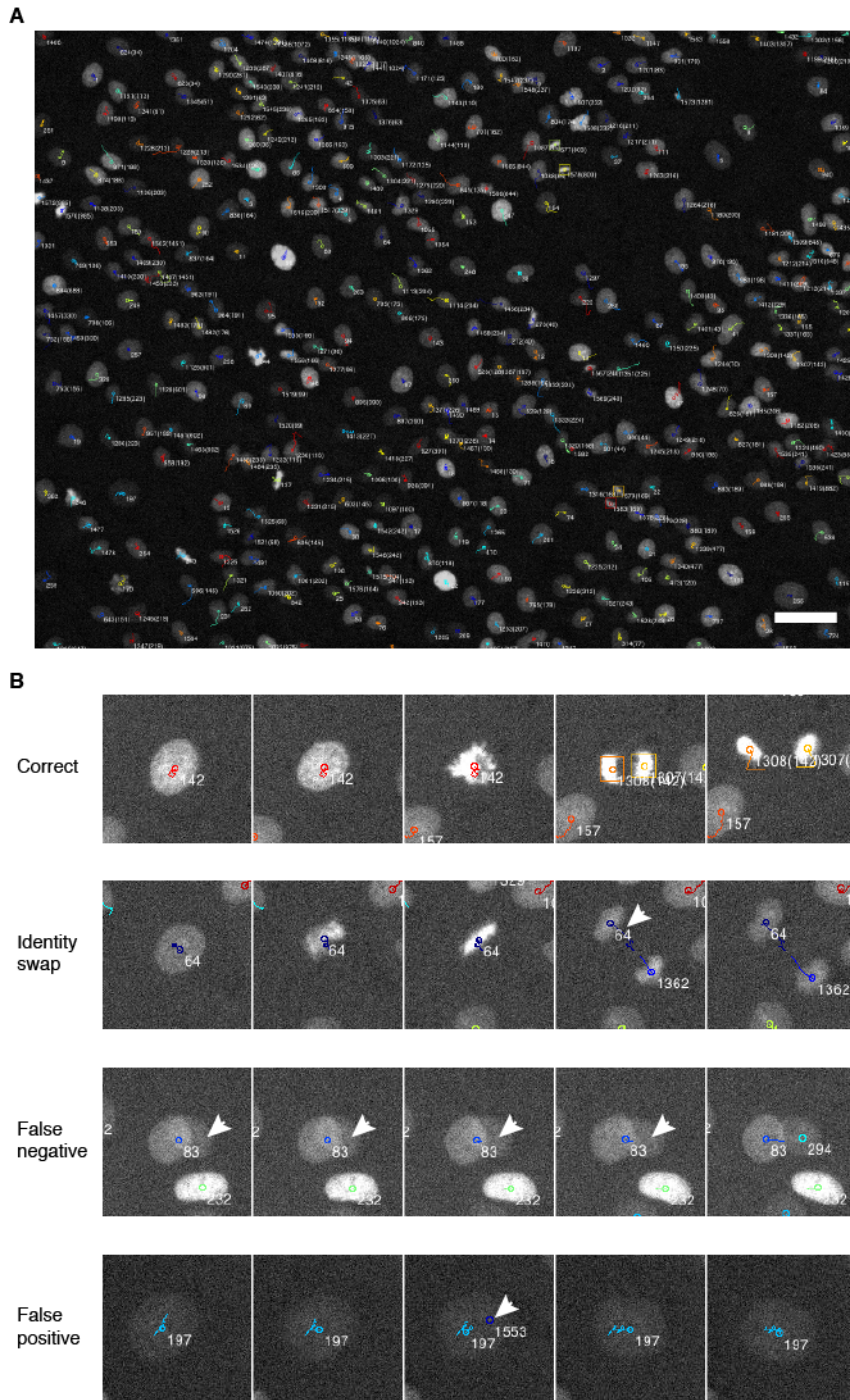
Figure S9: **Examples of tracking errors. A** Representative image of one field of view (1600 × 1200 pixels, 530μm × 400μm) showing tracking of hundreds of cells. **B** Examples of different types of tracking errors. White arrows show where incorrect detections or tracking errors occur. Numbers represent unique cell identifiers.

swap errors, which were counted manually.

We implemented the widely used Multiple Object Tracking (MOT) metrics to provide a quantitative assessment of the tracking performance [20]. The two MOT metrics used are:

Multiple Object Tracking Precision (MOTP):

$$MOTP = \frac{\sum_{i,t} d_t^i}{\sum_t c_t} \tag{2}$$

Where d_t^i is the Euclidean distance between manual annotation and tracker output i at time t and c_t is the number of matches at time t. This metric represents how precisely tracking algorithm can determine the position of an object. It is the ratio of the total error in position to the number of true positive correspondences between the tracker and manual annotation. In reality, the tracker is accurately measuring the centroid of the object, whereas the user is estimating it, so this value represents the user error and variability in determining the centroids in this case.

Multiple Object Tracking Accuracy (MOTA):

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IS_t)}{\sum_t g_t} \tag{3}$$

Where FN_t, FP_t and IS_t are the number of false negative, false positive and identity swaps observed at time t, and g_t is the total number of true observations at time t. This metric represents the errors associated with detecting objects and accurately keeping track of them, independent of the ability to precisely localise them. Overall we achieve a tracking accuracy (MOTA) of **93.3±1.8**% as shown in Table 1.

Image seq.	User	$N_{true}^{\ddagger}$	False positive	False negative	True positive	Identity swap	$MOTP$ (pixels)	$MOTA$
1	A	1099	35	39	1060	1	2.69	93.1%
2	B	795	27	17	778	0	3.33	94.4%
3	B	856	16	1	855	0	3.87	98.0%
4	B	931	33	19	912	0	3.88	94.3%
5	C	732	59	46	686	0	4.51	85.6%
Total	-	4413	170	122	4291	1	**3.61**	**93.3%**

Table 1: **Cell tracking quality assessment** Quality metrics calculated from five sequences by three separate users. $^{\ddagger}N_{true}$ is the total number of true observations in the sequence or $\sum_t g_t$.

Supplementary Note 3

Source code availability. MATLAB scripts for analysis of cell trajectories are available at `https://github.com/quantumjot/CellTracking`. The Bayesian tracking library is available at `https://github.com/quantumjot/BayesianTracker`.

References

[1] O. Hilsenbeck, M. Schwarzfischer, S. Skylaki, B. Schauberger, P. S. Hoppe, D. Loeffler, K. D. Kokkaliaris, S. Hastreiter, E. Skylaki, A. Filipczyk, M. Strasser, F. Buggenthin, J. S. Feigelman, J. Krumsiek, A. J. J. van den Berg, M. Endele, M. Etzrodt, C. Marr, F. J. Theis, and T. Schroeder, "Software tools for single-cell tracking and quantification of cellular and molecular properties," *Nature Biotechnology*, vol. 34, pp. 703–6, Jul 2016.

[2] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland, and D. M. Sabatini, "Cellprofiler: image analysis software for identifying and quantifying cell phenotypes," *Genome Biol*, vol. 7, p. R100, Dec 2006.

[3] M. Held, M. H. A. Schmitz, B. Fischer, T. Walter, B. Neumann, M. H. Olma, M. Peter, J. Ellenberg, and D. W. Gerlich, "Cellcognition: time-resolved phenotype annotation in high-throughput live cell imaging," *Nature methods*, vol. 7, pp. 747–754, Sep 2010.

[4] M. Winter, E. Wait, B. Roysam, S. K. Goderie, R. A. N. Ali, E. Kokovay, S. Temple, and A. R. Cohen, "Vertebrate neural stem cell segmentation, track-

ing and lineaging with validation and editing," *Nat Protoc*, vol. 6, pp. 1942–52, Nov 2011.

[5] L. Xu and M. Jordan, "On convergence properties of the em algorithm for gaussian mixtures," *Neural computation*, Dec 1996.

[6] S. Prince, "Computer vision: models, learning, and inference," *Cambridge University Press*, Dec 2012.

[7] S. Wienert, D. Heim, K. Saeger, A. Stenzinger, M. Beil, P. Hufnagl, M. Dietel, C. Denkert, and F. Klauschen, "Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach," *Scientific Reports*, vol. 2, p. 503, Dec 2012.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, Dec 1998.

[9] V. Nair and G. Hinton, "Rectified linear units improve restricted boltzmann machines," *Proceedings of the 27th international Conference on Machine Learning*, Dec 2010.

[10] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," *Artificial Neural Networks–ICANN 2010*, Dec 2010.

[11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv*, vol. cs.CV, Jun 2014.

[12] M. Abadi, A. Agarwal, P. Barham, and E. Brevdo. . . , "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv: . . .*, Dec 2016.

[13] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, Jan 1995.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec 2005.

[15] R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, Dec 1960.

[16] M. Narayana and D. Haverkamp, "A bayesian algorithm for tracking multiple moving objects in outdoor surveillance video," *Computer Vision and Pattern Recognition*, Dec 2007.

[17] O. Al-Kofahi, R. Radke, S. Goderie, Q. Shen, S. Temple, and B. Roysam, "Report automated cell lineage construction," *Cell Cycle*, vol. 5, no. 3, pp. 327–335, 2006.

[18] R. Bise, Z. Yin, and T. Kanade, "Reliable cell tracking by global data association," pp. 1004–1010, 2011.

[19] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The clear mot metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, Jan 2008.

[20] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, 1987.