

PAPER • OPEN ACCESS

Mechanistic and experimental models of cell migration reveal the importance of cell-to-cell pushing in cell invasion

To cite this article: Oleksii M Matsiaka *et al* 2019 *Biomed. Phys. Eng. Express* **5** 045009

View the [article online](#) for updates and enhancements.

Biomedical Physics & Engineering Express



PAPER

OPEN ACCESS

RECEIVED
28 February 2019

REVISED
17 April 2019

ACCEPTED FOR PUBLICATION
18 April 2019

PUBLISHED
27 May 2019

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Mechanistic and experimental models of cell migration reveal the importance of cell-to-cell pushing in cell invasion

Oleksii M Matsiaka¹, Ruth E Baker², Esha T Shah³ and Matthew J Simpson¹

¹ School of Mathematical Sciences, Queensland University of Technology (QUT), Brisbane, Queensland, Australia

² Mathematical Institute, University of Oxford, Oxford, United Kingdom

³ Ghrelin Research Group, Translational Research Institute, QUT, 37 Kent St, Woolloongabba, Queensland, Australia

E-mail: matthew.simpson@qut.edu.au

Keywords: cell migration, intercellular forces, cell spreading, cancer, wound healing, mathematical model

Supplementary material for this article is available [online](#)

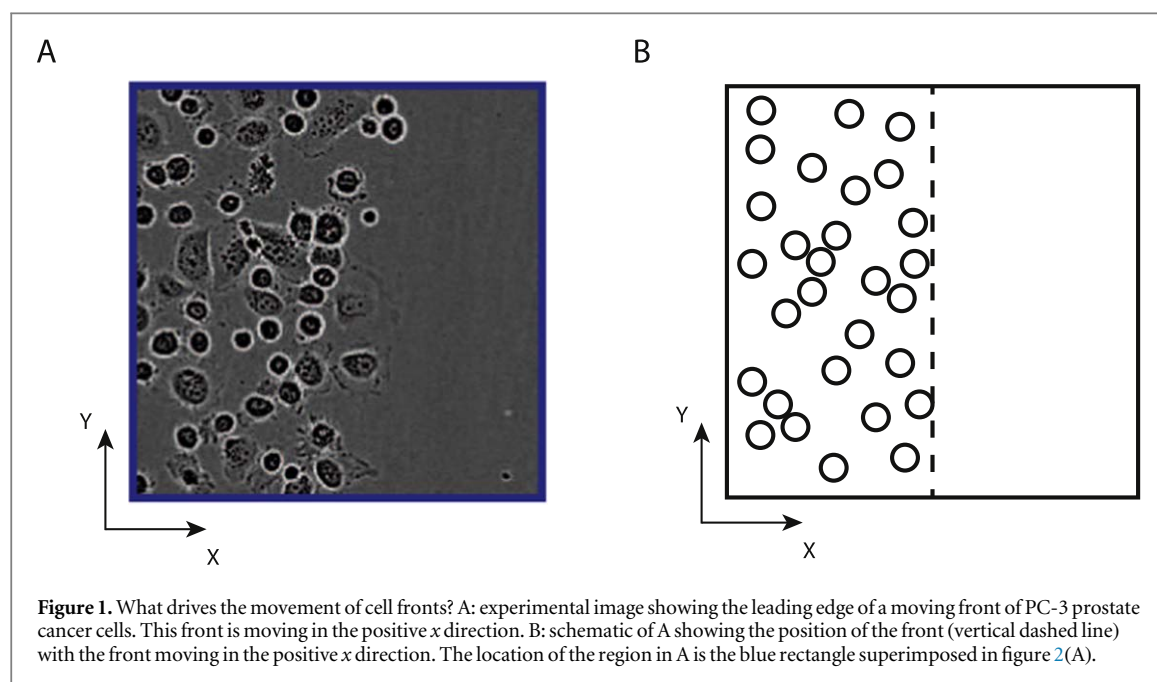
Abstract

Moving fronts of cells are essential for development, repair and disease progression. Therefore, understanding and quantifying the details of the mechanisms that drive the movement of cell fronts is of wide interest. Quantitatively identifying the role of intercellular interactions, and in particular the role of cell pushing, remains an open question. In this work, we report a combined experimental-modelling approach showing that intercellular interactions contribute significantly to the spatial spreading of a population of cells. We use a novel experimental data set with PC-3 prostate cancer cells that have been pretreated with Mitomycin-C to suppress proliferation. This allows us to experimentally separate the effects of cell migration from cell proliferation, thereby enabling us to focus on the migration process in detail as the population of cells recolonizes an initially-vacant region in a series of two-dimensional experiments. We quantitatively model the experiments using a stochastic modelling framework, based on Langevin dynamics, which explicitly incorporates random motility and various intercellular forces including: (i) long range attraction (adhesion); and (ii) finite size effects that drive short range repulsion (pushing). Quantitatively comparing the ability of this model to describe the experimentally observed population-level behaviour provides us with quantitative insight into the roles of random motility and intercellular interactions. To quantify the mechanisms at play, we calibrate the stochastic model to match experimental cell density profiles to obtain estimates of cell diffusivity, D , and the amplitude of intercellular forces, f_0 . Our analysis shows that taking a standard modelling approach which ignores intercellular forces provides a poor match to the experimental data whereas incorporating intercellular forces, including short-range pushing and longer range attraction, leads to a faithful representation of the experimental observations. These results demonstrate a significant role of cell pushing during cell front movement and invasion.

1. Introduction

Moving cell fronts occur during many physiological processes, such as wound healing, morphogenesis, and malignant invasion [1–4]. Typically, cell fronts are observed as advancing, sharp boundaries between densely occupied and vacant regions, or as a moving interface between two distinct populations of cells [5]. An example of the first scenario is wound healing, where populations of cells close and recolonize an initially vacant space [6, 7], as shown in figure 1. An advancing interface between two populations of cells is

often associated with malignant invasion into surrounding tissues [8, 9]. Improving our understanding of how cell populations spread can provide important, clinically-relevant information about the nature of moving cell fronts. Historically, moving cell fronts have been studied, both *in vitro* and *in vivo*, to provide both qualitative and quantitative information about the mechanisms that drive front movement. We note that quantifying the precise contributions of various cellular-level mechanisms that lead to population-level front behavior is a nontrivial task that requires the integration of many different types of experimental



data [10]. Often it is assumed that the movement of advancing cell fronts is driven by the combined effects of undirected cell migration and carrying capacity-limited cell proliferation [10]. At present, a fundamental question, which remains largely overlooked in the mathematical biology literature, is *what is the role of cell-to-cell pushing and how does it influence population-level front movement?*

Cell motility is a complicated process involving both the interplay and competition between various individual-level mechanisms [5, 10, 11]. One of the most well-studied individual-level cell motility mechanisms is lamellipodial cell migration where cells undergo undirected movement due to myosin-powered contractions of the actin network under the cytoplasmic membrane [12]. Since this process is observed in many cell types, it remains prevalent in many mathematical modelling frameworks. As such, the assumption that cells undergo Brownian motion is often invoked and cells are represented, either implicitly or explicitly, as non-interacting point particles that move according to a white Gaussian process [13]. While this approach is appealing due to its simplicity, it neglects the effects of intercellular interactions, such as adhesion and finite size (crowding) effects. The neglect of cell-to-cell adhesion can be problematic because it is known that mesenchymal cell types, such as keratinocytes, can be strongly affected by cell-to-cell adhesion during wound healing [14]. Furthermore, adherent cells can form clusters that exhibit qualitatively different behaviour from isolated cells [15–18]. For example, strongly adherent groups of cells can undergo movement mediated by adherens junctions and move as a whole in coordinated process known as *locomotion* [19, 20]. During this motion cells can change shape because of the imbalance between traction and friction forces induced by leader and follower cells [21]. A few

individual cells can heavily influence front expansion in a process known as *trailblazing* and form finger-like protrusions [22, 23]. Tightly packed interacting monolayers of cells can exhibit properties similar to the solid state matter with a transition to the fluid-like behaviour [24]

Arguably, some of the most striking examples of the front-like spreading of a cell population occur during embryonic development, such as neural crest cell invasion in the developing gut tissues, which is thought to arise as a consequence of combined undirected Brownian cell motility and carrying capacity limited cell proliferation [25–27]. However, previous investigations have made the point that short range cell pushing can also play a role in driving the movement of cell fronts in confined environments, such as living tissues [28]. Henceforth, we hypothesize that cells in a confined space may generate population pressure, driven by finite size effects and local repulsion, which can stimulate spatial expansion of the population.

Perhaps the most popular mathematical framework for modelling the movement of cell fronts involves using reaction-diffusion equations [29–31], including the Fisher-Kolmogorov equation, and generalisations thereof. Previously, the lack of individual-level experimental data meant that classical reaction-diffusion models were a useful way to conceptualize and simulate collective cell behaviour. However, with the increasing availability of individual-level information it is becoming important to develop mathematical models that provide both population-level information and individual-level information. There is a vast number of discrete models that focus on many individual-level features of cell behaviour [32–35]. The downside of many modelling frameworks presented in the literature is a large number of free parameters

which makes parameter estimation problems computationally intractable and simulations hard to interpret. For example, in [33] the authors develop a comprehensive mathematical model of stem cells dynamics in the intestinal crypt. This model requires the selection of 25 free parameters. Some of these parameters are not directly measurable, and other parameters in the model had not been estimated before this study. Consequently, quantifying the precise contribution of different factors that control cell colony behaviour and dynamics is very challenging. To deal with this question of over parameterization, we deal only with a fundamental model that is characterized by a minimal number of degrees of freedom that we can use experimental data to identify the parameters in the model and hence to quantitatively explore the roles of undirected migration and cell-to-cell pushing in our experiments.

Previously, cell pushing has been incorporated into lattice-based models of cell motility where agents move on a spatial domain that is represented as a regular lattice [36, 37]. However, these previous studies about the role of cell pushing are primarily theoretical studies that do not consider calibrating models to quantitatively match any experimental data. Our current work is the first attempt to incorporate short range cell pushing into a more realistic spatially continuous, off-lattice discrete model of cell migration. Importantly, we directly apply the model to quantitatively mimic a novel experimental data set. Figure 2 illustrates the IncuCyte ZOOMTM scratch assay experimental protocol that we use in this work, at $t = 0$ h and $t = 48$ h. In our experiments we isolate the role of cell migration from the effects of cell proliferation by working with a population of cells that has been pretreated with a chemotherapy drug, Mitomycin-C, to block DNA replication and, consequently, suppress proliferation [38]. This means that the number of cells present in the experimental field of view over the duration of the experiment remains approximately constant. However, a side effect of Mitomycin-C pretreatment is that individual cells increase in size during the experiment as the cells prepare to proceed through the cell cycle but are unable to divide [39]. This dynamic increase in cell size, which is typically neglected in previous modelling studies [39], can significantly influence intercellular interactions during the experiment and so we take great care to incorporate these effects into our mathematical model. Our approach for incorporating dynamic cell size effects in the model is justified by also working with a simpler model that neglects dynamic changes in cell size; we find that the simpler, standard model leads to a much poorer match with the experimental data.

This work is structured as follows. We begin by describing the IncuCyte ZOOMTM experimental protocol, the experimental data, and the procedure we use to process the experimental images. We then introduce the discrete mathematical model which accounts

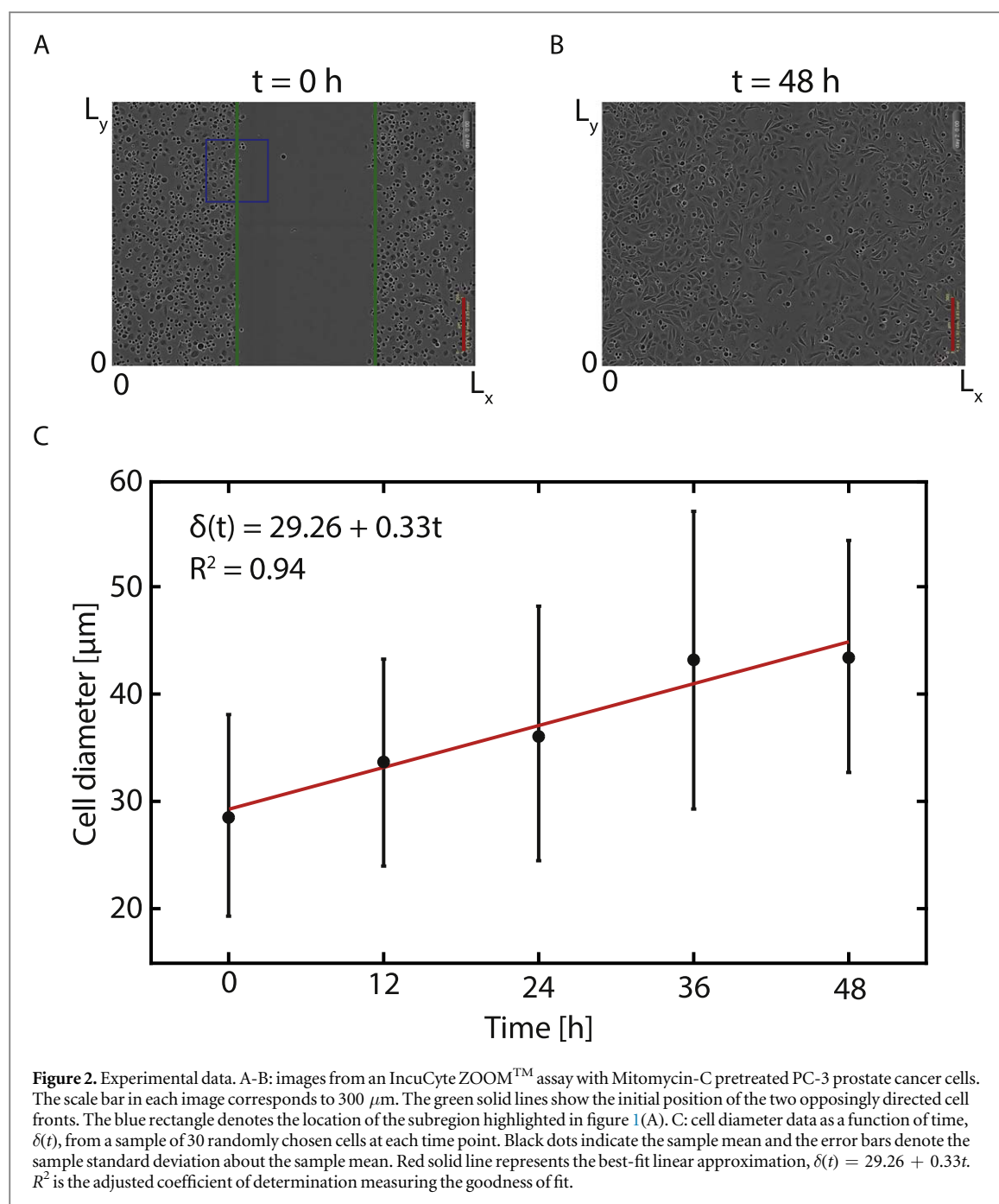
for random motility and intercellular interactions, including short range pushing and longer range attraction, as well as incorporating a mechanism for describing dynamic cell size changes. We refer to this model as *Model I* since it incorporates all four mechanisms that are thought to be relevant to the experimental system. To quantitatively explore the significance of these various cell-level mechanisms we systematically repeat the model calibration process for a range of simpler, more commonly used models. These simplified models account for: (i) random motility and intercellular forces (*Model II*); (ii) intercellular forces only (*Model III*); and (iii) random motility only (*Model IV*). We discuss the performance of each model when applied to the IncuCyte ZOOMTM data in the Results and Discussion. Finally, in the Conclusions we summarize our findings and discuss alternative applications and extensions of our modelling framework.

2. Materials and methods

2.1. IncuCyte ZOOMTM experimental data

Monolayer scratch assays are performed using the IncuCyte ZOOMTM system (Essen BioScience, MI USA [40]) as shown in figures 2(A)–(B). All experiments are performed using the PC-3 prostate cancer cell line [41] from the American Type Culture Collection (ATCC, Manassas, USA). The procedure of growing the cell culture in a flask is outlined in [42]. After growing, cells are removed from the flask using TrypLETM (Life Technologies) in phosphate buffered saline, resuspended in medium and seeded at a density of 20,000 cells per well in 96-well ImageLock plates (Essen BioScience) as shown in figure 3. The diameter of each individual well is 9000 μm .

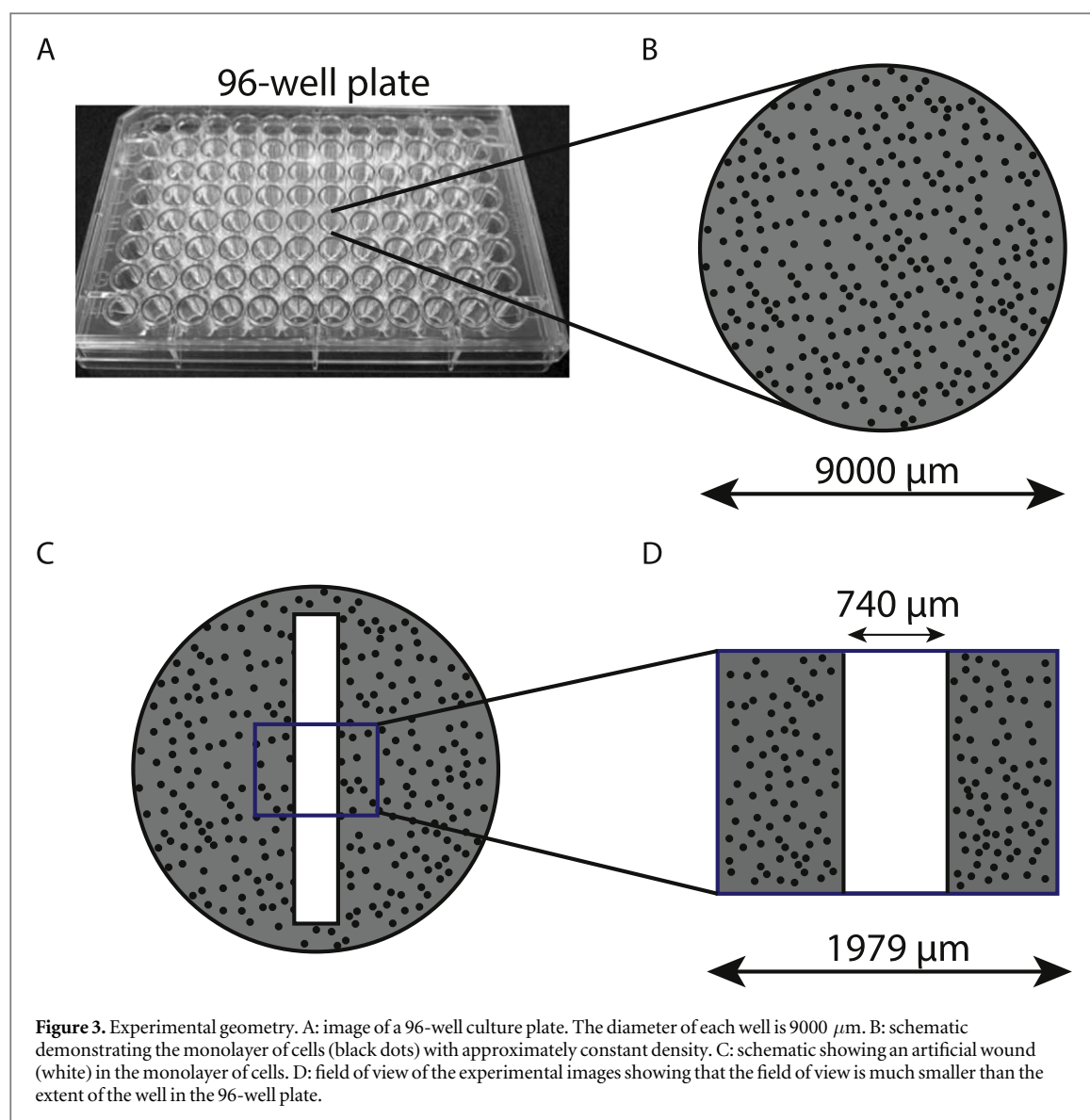
Mitomycin-C is added at a concentration of 10 g mL^{-1} for two hours before a scratch is made in the monolayer of cells [38, 43, 44]. A WoundMakerTM (Essen BioScience) is used to create identical scratches in the uniformly distributed populations. Medium is aspirated after scratching; each well is washed twice and refilled with fresh medium (100 μl). Plates are incubated in IncuCyte ZOOMTM and photographed every 2 hours for 48 hours. In total, these experiments are performed in eight of the 96 wells on the 96-well plate. After a preliminary visual inspection of the resulting eight experimental images, we selected four typical wells for analysis. Throughout this work we will refer to these four identically prepared experiments as Experiment A, B, C and D, respectively. All experiments are initiated with a monolayer of cells close to confluence. Even though the cell culture was treated with Mitomycin-C, which ensures the absence of proliferation and approximately constant cell count during an experiment, the monolayer of cells remains close to confluence throughout the duration of the experiment due to increase in the size of individual



cells (figure A.1, Supplementary materials available online at stacks.iop.org/BPEX/5/045009/mmedia).

By the end of the experiment, summarised in figures 2(A)–(B), we see that the Mitomycin-C pretreated cells have approximately doubled in size [39]. There is little information about the change in volume of cells since all data is in the form of 2D images. However, we note that, due to effects of Mitomycin-C pretreatment that allows cells to go through cell cycle to the point when they are ready to divide, the observed growth is rather volumetric and reflects actual change in cell size. To quantify this increase in cell size we randomly choose 30 cells from the experimental images at $t = 0, 12, 24, 36$ and 48 h and use these cells to estimate the average diameter as a function of time, $\delta(t)$.

To do this we estimate the area on the image occupied by each particular cell, and then convert this estimate of area into an equivalent diameter, $\delta = \sqrt{4A/\pi}$, where A is the area estimate. With 30 estimates of the diameter at $t = 0, 12, 24, 36$ and 48 h, we compute the sample mean and sample standard deviation at each time point and plot the data in figure 2(C). Visually we see that the average diameter appears to increase approximately linearly with time, and so we fit a linear model to the data. The cell diameter data and the linear model are shown in figure 2(C). Note that had our experiments been performed over a longer period of time it would be more appropriate to use a different model, such as the logistic growth function, to model the temporal cell size dynamics.



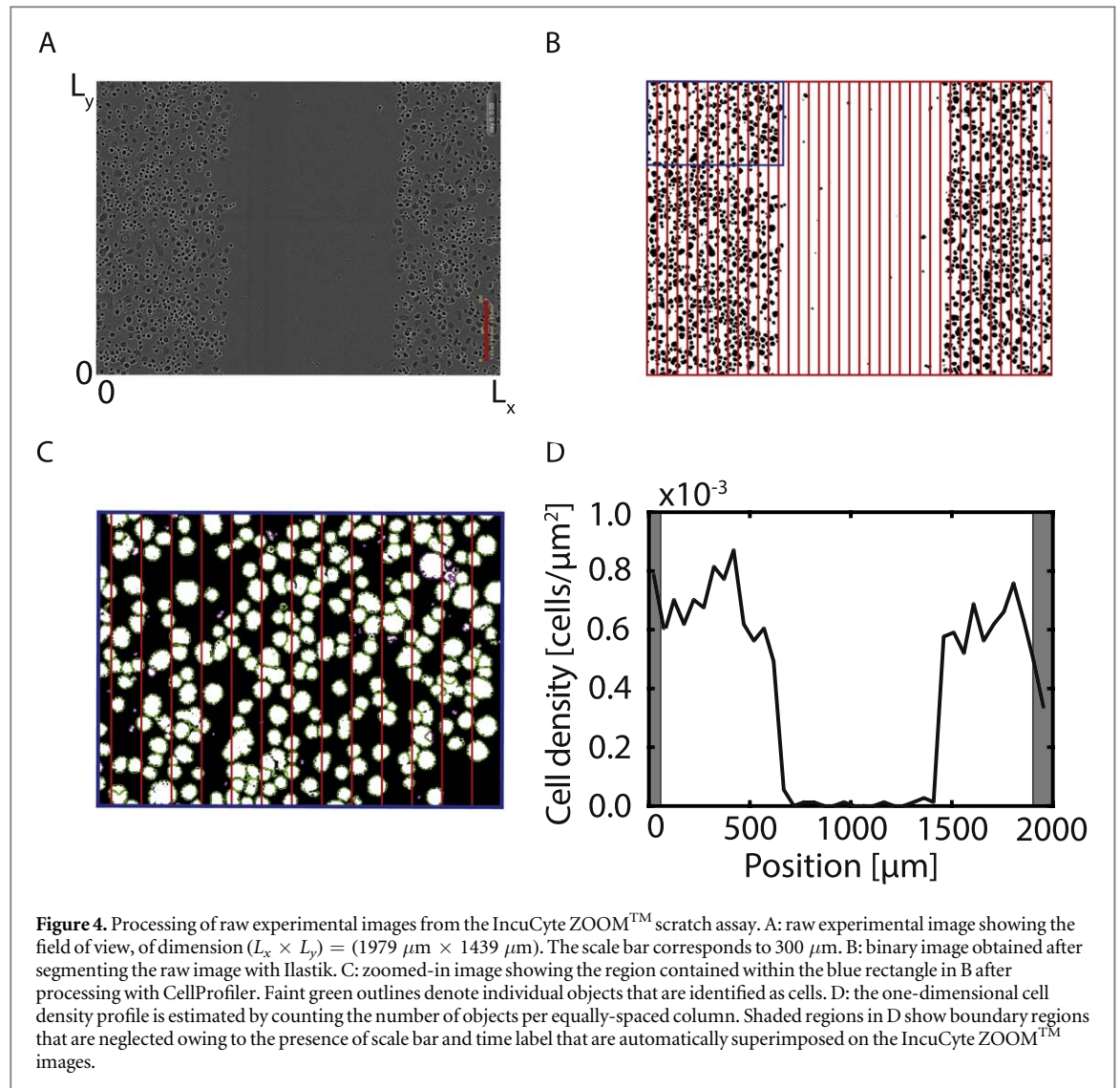
2.2. Image analysis

An example of a raw experimental image and a detailed description of the procedure we use to extract density information from that image are summarized in figure 4. Here, the size of the field of view is $(L_x \times L_y) = (1979\ \mu\text{m} \times 1439\ \mu\text{m})$, as shown in figure 4(A). Throughout this work we use data from four identically-prepared experimental replicates of the scratch assay. Experimental images at the beginning of the experiment for each of the four replicates are shown in figure 5. For completeness, the time series of experimental images for all four identically prepared experiments is given in Supplementary materials (figure A.1), giving a total of 20 experimental images.

To process each experimental image we first separate the background of the image from the cells using Ilastik [45]. Ilastik is a machine learning tool that enables automatic object identification, and allows us to separate the cells in each image from the background. An example of a grey scale segmented image is

shown in figure 4(B). A visual comparison of the raw image in figure 4(A) with the segmented image in figure 4(B) confirms that the identification of cells from the background in the image is accurate. Since the density of cells in the original images is independent of the vertical coordinate, as shown in figure 5, we divide each image into 40 equally-spaced columns, $49.5\ \mu\text{m}$ wide each. We then use CellProfiler to automatically estimate the number of cells per column [46]. With this data we divide the number of cells per column by the area of each column to give an estimate of the cell density across the horizontal coordinate of the experimental images.

A typical column-averaged cell density profile, shown in figure 4(D), summarizes the spatial variations in density as a function of the horizontal coordinate. We employ this technique to extract density profiles at each time point, $t = 0, 12, 24, 36$, and $48\ \text{h}$, for each experimental replicate. The data presented in figure 4(D) shows the average cell density that we associate with the centre of each column. This



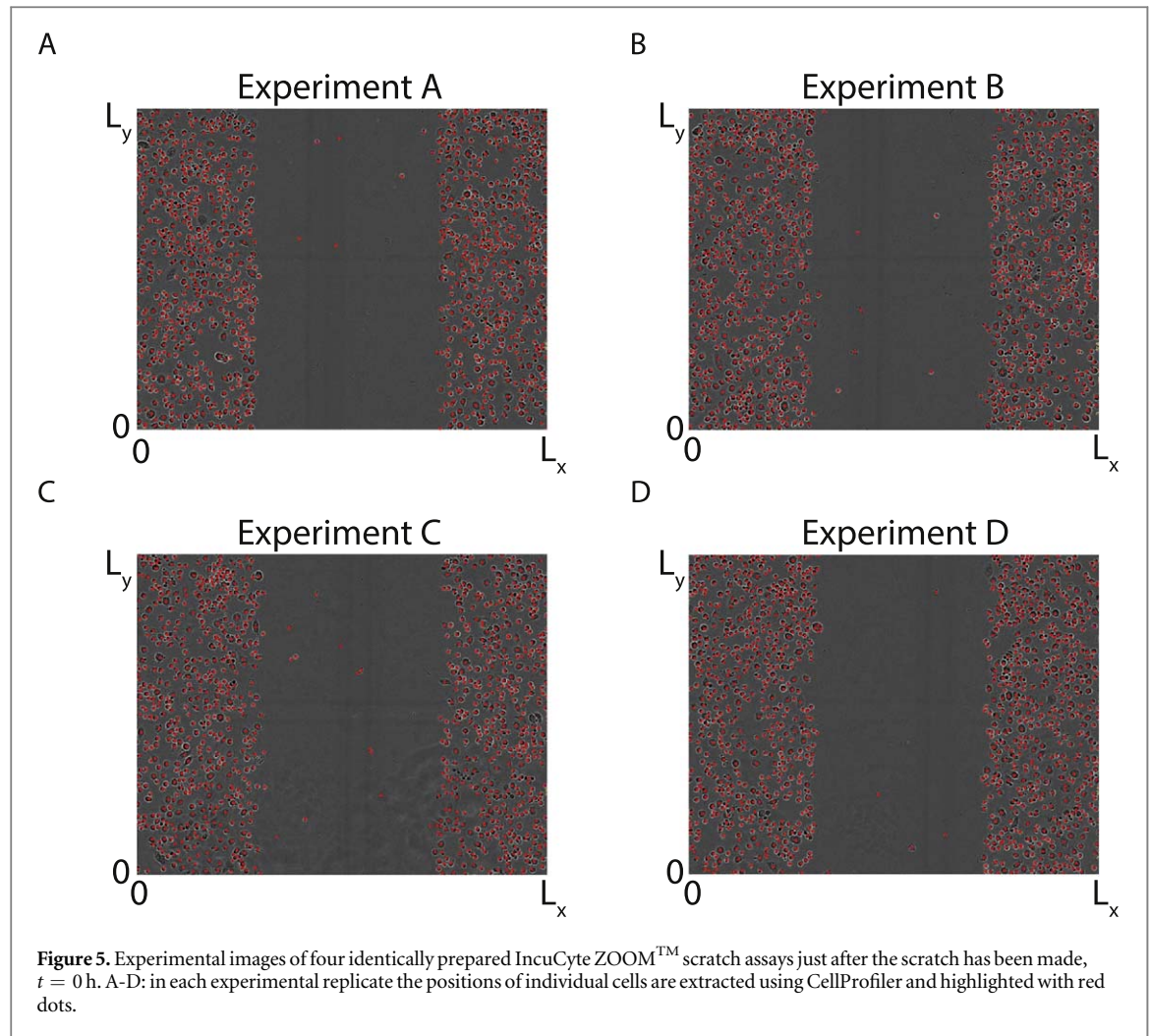
particular density profile is relatively noisy because it is associated with the single image in figure 4(A). Before we proceed, we discard density data from the two right-most columns of each image because the time label and scale bar are automatically superimposed on the IncuCyte ZOOM™ images and these objects partially obscure the numbers of cells in these subregions of each images. We also discard the left-most column from each image, which leaves us with a slightly smaller image that is discretized into 37 equally-spaced columns of width $49.5 \mu\text{m}$, and a reduced domain width of $1831 \mu\text{m}$. Next, to reduce the magnitude of the fluctuations in the cell density profile, we average the density profiles associated with each of the four experimental replicates to obtain a single averaged cell density profile as a function of time, as summarized in figure 6. For completeness, the relatively noisy column-averaged cell density profiles associated with each of the four individual experimental replicates are given in Supplementary materials (figure C.1).

Now that we have quantified our experimental observations in terms of the temporal variation in the column-averaged cell density profiles, further averaged

across four identically prepared experimental replicates, we will now attempt to use a suite of discrete mathematical models to mimic the experimental data set. To provide the most realistic discrete simulations, we always take care to initialize each discrete simulation using the exact same number and locations of cells that are present in the experimental images at $t = 0$ h (figure 5).

2.3. Discrete stochastic model

To simulate our experimental data set we use a two-dimensional discrete model of cell motility that incorporates random motility, intercellular interactions including both long range cell-to-cell adhesion and short range cell pushing, as well as capturing dynamic changes in cell size. We refer to this model as *Model I* since it describes the situation where all four processes are acting simultaneously. We choose to work with a discrete modelling framework because discrete individual-based approaches are more natural to compare with experimental images than continuum models [47, 48]. Such discrete individual-based models are used to study a range of cell biology phenomena, including malignant invasion [49], wound healing



[50], self-organization [35], angiogenesis [51] and embryonic development [25]. Since we work with an off-lattice discrete framework, each agent is allowed to move in any direction on a continuous domain. This off-lattice approach is more realistic than a simpler lattice-based model where the locations of agents are restricted to an artificial lattice structure [13, 35, 52, 53].

We begin by introducing Model I and then describe three simplifications of this model in which we systematically neglect certain features. To explore the relevance of the mechanisms inherent in these four models we carefully calibrate each model to match the density data summarized in figure 6 and quantitatively compare the results of the calibration procedure.

2.3.1. Model I: random motility, long range cell-to-cell adhesion, short range cell pushing and dynamic changes in cell size

A key novelty of our approach is that we simulate the dynamic change in agent diameter, $\delta(t)$. This approach is very different to standard approaches where agents in discrete models are thought of as having either no size or a constant size. Here, to match our experimental measurements presented in figure 2(C), we

assume that $\delta(t)$ increases linearly,

$$\delta(t) = 29.26 + 0.33t, \quad (1)$$

where t is time, measured in hours.

We employ a Langevin stochastic framework [54], where a population of N cells is modelled by a system of N first order stochastic differential equations

$$\frac{d\mathbf{u}^{(i)}}{dt} = \overbrace{\sum_{j \neq i} \mathbf{F}_{ij}(r, t)}^{\text{directed force}} + \overbrace{\xi_i}^{\text{random noise}}, \quad i = 1, \dots, N, \quad (2)$$

where $\mathbf{u}^{(i)}$ is the position vector of the i th agent on a two-dimensional domain, \mathbf{F}_{ij} is the deterministic interaction force between agents i and j that are separated by distance r [55], and ξ_i is a random stochastic force exerted on the i th agent. The stochastic force ξ_i is sampled from the Gaussian distribution [54] with zero mean and variance $2D/\Delta t$, where D is the diffusivity, and Δt is the time step used to numerically solve equation (2). Since the Langevin equation formalism does not include any inertial forces, this framework implicitly neglects agent acceleration. This assumption is reasonable at low Reynolds numbers, and is routinely invoked at cellular length scales [56]. The model given by equation (2) does not address possible directional bias near the edge of a wound resulting from the chemical signalling responses

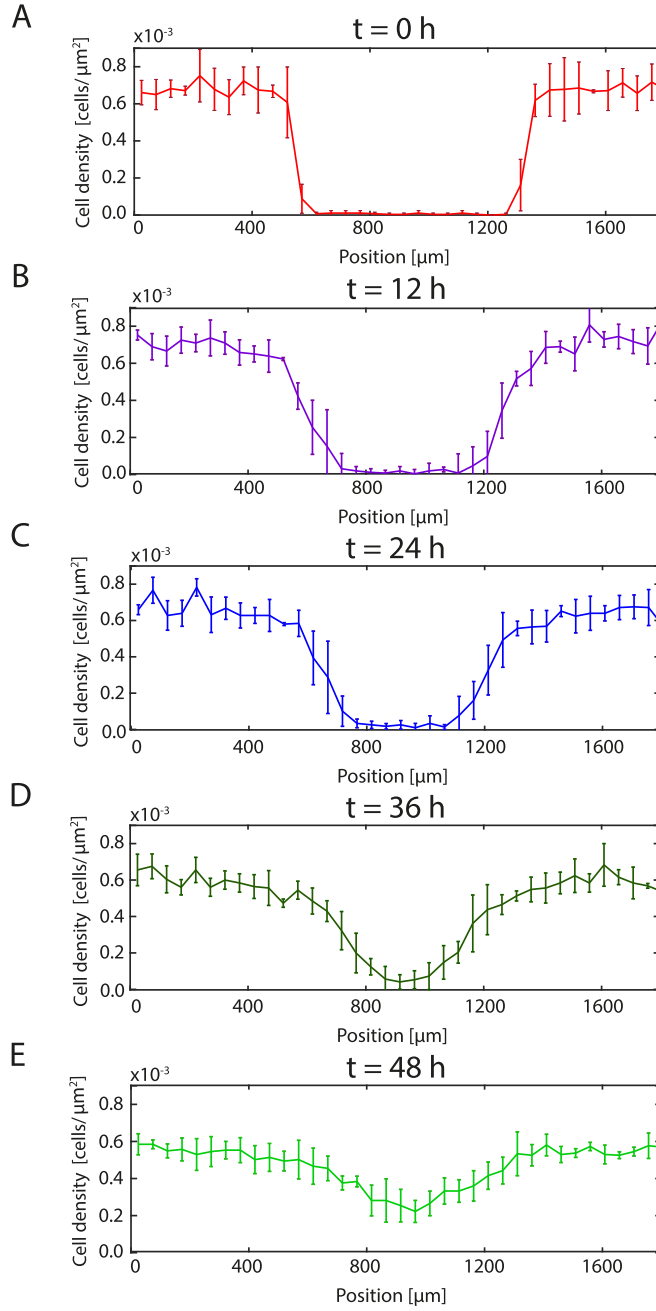


Figure 6. Averaged cell density profiles. A-E: averaged cell density profiles at $t = 0, 12, 24, 36$, and 48 h, as indicated. All profiles report the sample mean density computed using four identically prepared experimental replicates. The error bars denote the sample standard deviation.

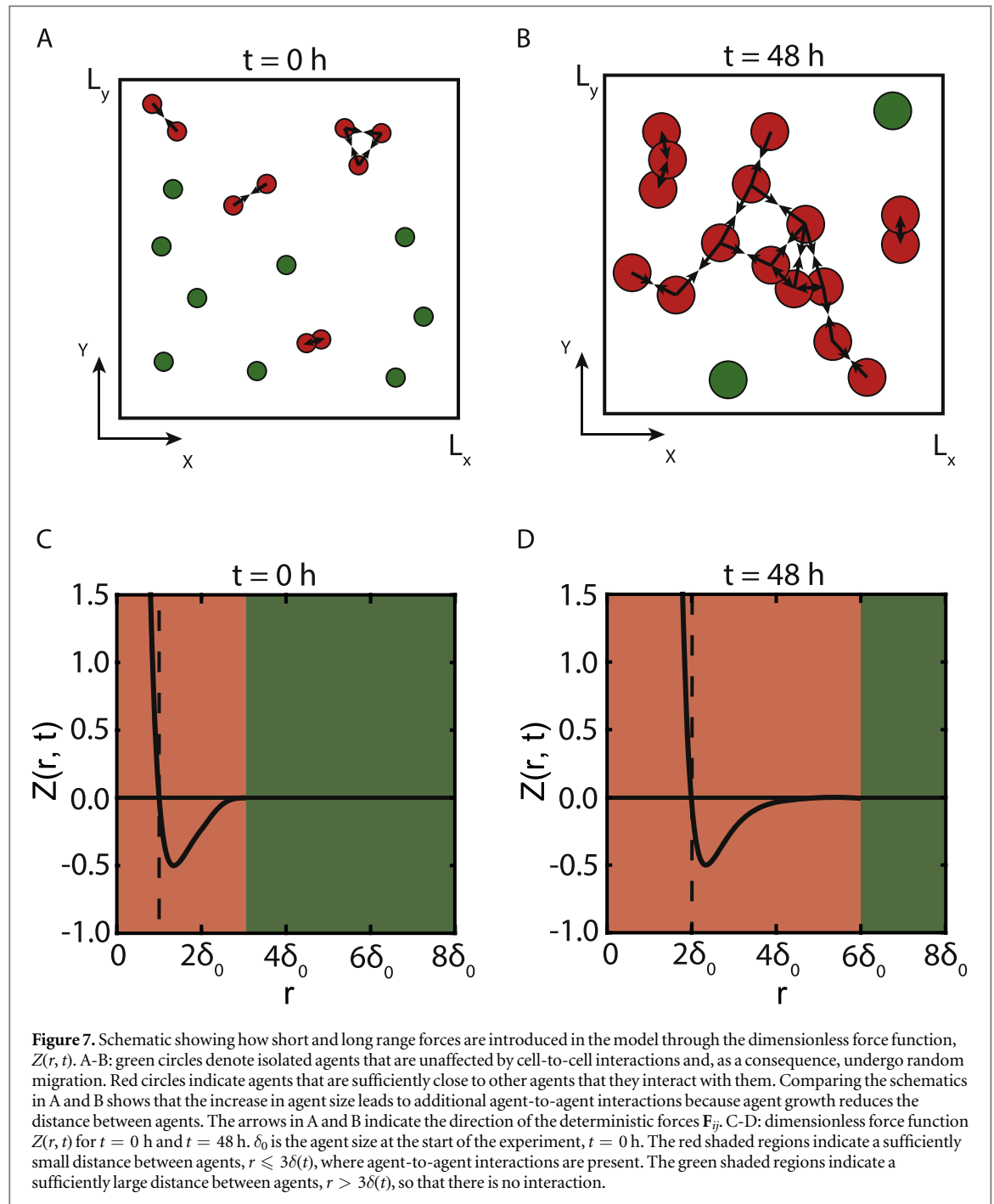
to wounding [57]. Including directional bias and chemotaxis will result in additional free parameters in the system [31], that we have little prior knowledge about, which, in turn, will make the parameter estimation problem significantly harder. Increasing number of parameters in the model requires increased complexity of the experimental data to avoid ambiguities in the parameter estimation [58]. As such, we focus on describing a cell population using the fundamental model where the cell migration is mediated by a directional movement due to cell-to-cell adhesive and repulsive forces, \mathbf{F}_{ij} , and the undirected stochastic force, $\boldsymbol{\xi}_i$.

The details of the deterministic interaction force, \mathbf{F}_{ij} , can be chosen to incorporate a range of relevant

phenomena such as long range attraction and short range repulsion [59], as illustrated in figure 7. In this work we specify that the interaction force, \mathbf{F}_{ij} , depends upon the distance between agents, r , and time, t , and is given by

$$\mathbf{F}_{ij}(r, t) = f_0 Z(r, t) \frac{\mathbf{u}^{(i)} - \mathbf{u}^{(j)}}{|\mathbf{u}^{(i)} - \mathbf{u}^{(j)}|}, \quad (3)$$

where f_0 is dimensional magnitude of the interaction force, $Z(r, t)$ is a dimensionless function describing how the interaction force depends on the separation between the agents, $r = |\mathbf{u}^{(i)} - \mathbf{u}^{(j)}|$. As with all models of the form of equation (2), the dimensional magnitude of the interaction force, f_0 , has dimension of velocity.



Previously, the interaction forces between cells in the epithelial monolayer have been inferred using a discrete framework in [48], where the authors explicitly demonstrate the significance of population pressure and mechanically driven population spreading [60]. The key novelty in the expression for the force function, equation (3), is the time dependent cell-to-cell interaction potential, $Z(r, t)$, figure 7. This approach allows us to explicitly parameterise the dynamic change in cell size and simulate mechanically driven cell pushing, which is in contrast to many models of cell motility and adhesion present in the literature where cells have constant size [61]. The cell-to-cell pushing is expressed by a combination of cell-to-cell repulsive forces incorporated in the force function, equation (3), dynamical change in cell

size, and cell random motility. Here, we address basic phenomenological principles of cell-to-cell adhesion without considering biochemical pathways that facilitate adhesion, such as E-cadherin protein junctions [62, 63].

In this work we use the dimensionless function $Z(r, t)$ to incorporate three main features of cell-to-cell interactions: (i) short range repulsion forces which mimic cell pushing; (ii) longer range attraction forces which mimic cell-to-cell adhesion; and (iii) dynamic changes in agent size. The short range repulsion forces can be interpreted as cell resistance to deformation, which leads to crowding and volume exclusion effects [64]. In contrast, the longer range attraction forces mimic intercellular attraction. These cell-to-cell attraction forces are thought to be a

predominant factor in the cell-to-cell adhesion [65]. To incorporate these effects we use a modified Morse potential [66],

$$Z(r, t) = \begin{cases} 2(\exp(-2a[r - \delta(t)]) - \exp(-a[r - \delta(t)])), & r < 2\delta(t), \\ 2(\exp(-2a[r - \delta(t)]) - \exp(-a[r - \delta(t)]))g(r, t), & 2\delta(t) \leq r \leq 3\delta(t), \\ 0, & r > 3\delta(t), \end{cases} \quad (4)$$

where $\delta(t)$ is the time-dependent agent diameter, $a > 0$ is a parameter that controls the shape of the force

function, and $g(r, t) =$

$(1 - \sin[(2\pi r - \pi\delta(t))/2\delta(t)])/2$ is the Tersoff cut-off function [67] which ensures a finite range of interactions. The cell-to-cell interaction range is finite, and set to three agent diameters [54]. As such, the interaction force is zero for separation distances of greater than three agent diameters. For all results presented here we set $a = 0.08$ which implies that the agents are relatively rigid and highly unlikely to overlap [66].

Schematics showing the key features of $Z(r, t)$ at $t = 0$ h and $t = 48$ h are shown in figures 7(C)–(D). Here we see that at short separation distances we have strong positive $Z(r, t)$, which captures short range repulsion and pushing, owing to finite size effects. Over longer separation distances we have smaller negative $Z(r, t)$ which models attraction, such as adhesion. Finally, over sufficiently large distances we have no interactions as $Z(r, t) = 0$. To capture the effects of the increase in cell size, all length scales in figures 7(C)–(D) are given in terms of the average cell diameter, $\delta(t)$, which can vary with time. In this work we use a linear function for $\delta(t)$ because this matches our experimental observations, however other functional forms for $\delta(t)$ are possible. In all simulations we apply the Langevin model on a domain of size $1831.5 \mu\text{m} \times 1439 \mu\text{m}$, which is the size of the experimental field of view after the boundary columns have been neglected.

2.4. Reduced models

- Model II: constant cell size ($\delta(t) = 29.26 \mu\text{m}$) with cell diffusion and intercellular forces
- Model III: variable cell size with intercellular forces but no cell diffusion, $D = 0$ and $f_0 > 0$
- Model IV: variable cell size with diffusion and without intercellular forces, $f_0 = 0$ and $D > 0$

To simulate Models I–IV we must apply appropriate boundary conditions to reflect the conditions relevant to the experiment. To determine these boundary conditions, we note that the experimental images, shown in figures 2(A)–(B), correspond to a

field of view that is much smaller than the spatial extent of the experiment. For example, the width of the field of view in figures 2(A)–(B) is $1979 \mu\text{m}$, which

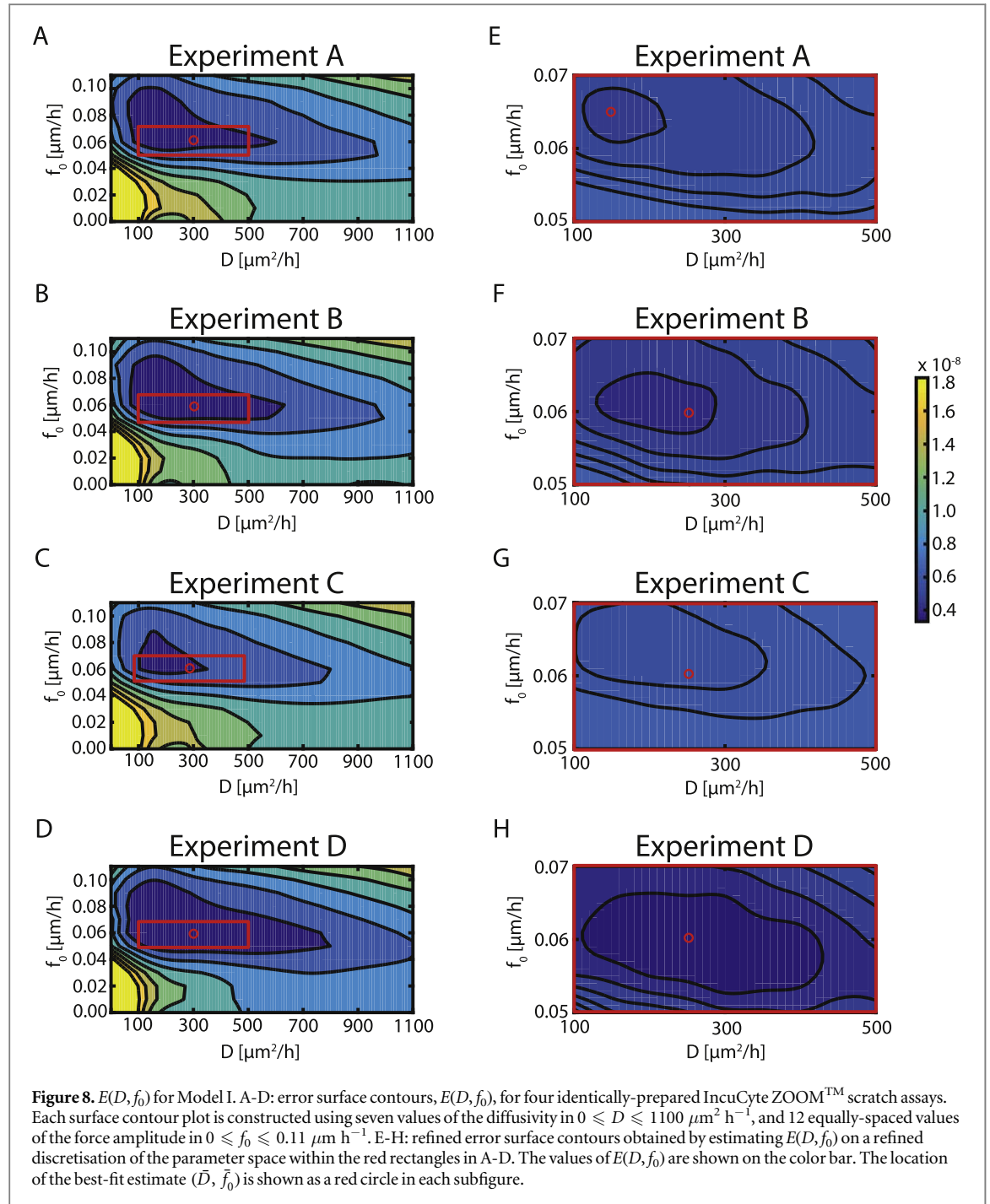
is much smaller than the diameter of the well in the 96-well plate ($9000 \mu\text{m}$), as shown in figure 3. The schematic in figure 3 is important because it emphasizes that the images from this kind of assay only show a small proportion of the population of cells present in the well. In particular, it is important to remember that the boundaries around the field of view are not physical boundaries since the spatially uniform population of cells extends far beyond the boundaries around the field of view. This means that whenever the density is below confluence, cells will migrate, in each direction, across the boundaries of the field of view. However, since the population of cells is placed uniformly into each well of the 96-well plate, the net flux of cells across the boundaries of the field of view will be approximately zero owing to symmetry. To justify zero net flux boundary conditions we also count a number of cells in each experimental replica at $t = 0, 12, 24, 36$, and 48 hours (figure B.1, Supplementary materials). These additional results demonstrate that cell counts remain approximately constant with average fluctuations of 8.1% of cell count at $t = 0$ h across four experiments. Similar experimental set up without Mitomycin-C pretreatment would have resulted in about 100% increase in cell count due to proliferation. Therefore, we impose zero net flux boundary conditions around all boundaries of the field of view [68]. We implement these boundary conditions by simply aborting any potential movement event that would take a particular agent across one of the boundaries.

3. Results and discussion

To quantitatively compare the suitability of Models I–IV to describe our experimental data set, we calibrate each model to provide the best match to the experimental density data. For Model I our aim is to estimate the two model parameters, (D, f_0) , that lead to the best match with the experimental measurements. To facilitate this we introduce a measure of the discrepancy between the experimental data and the model solution,

$$E(D, f_0) = \frac{1}{148} \sum_{j=1}^4 \sum_{i=1}^{37} [p_{\text{data}}(x_i, t_j) - p_{\text{model}}(x_i, t_j)]^2, \quad (5)$$

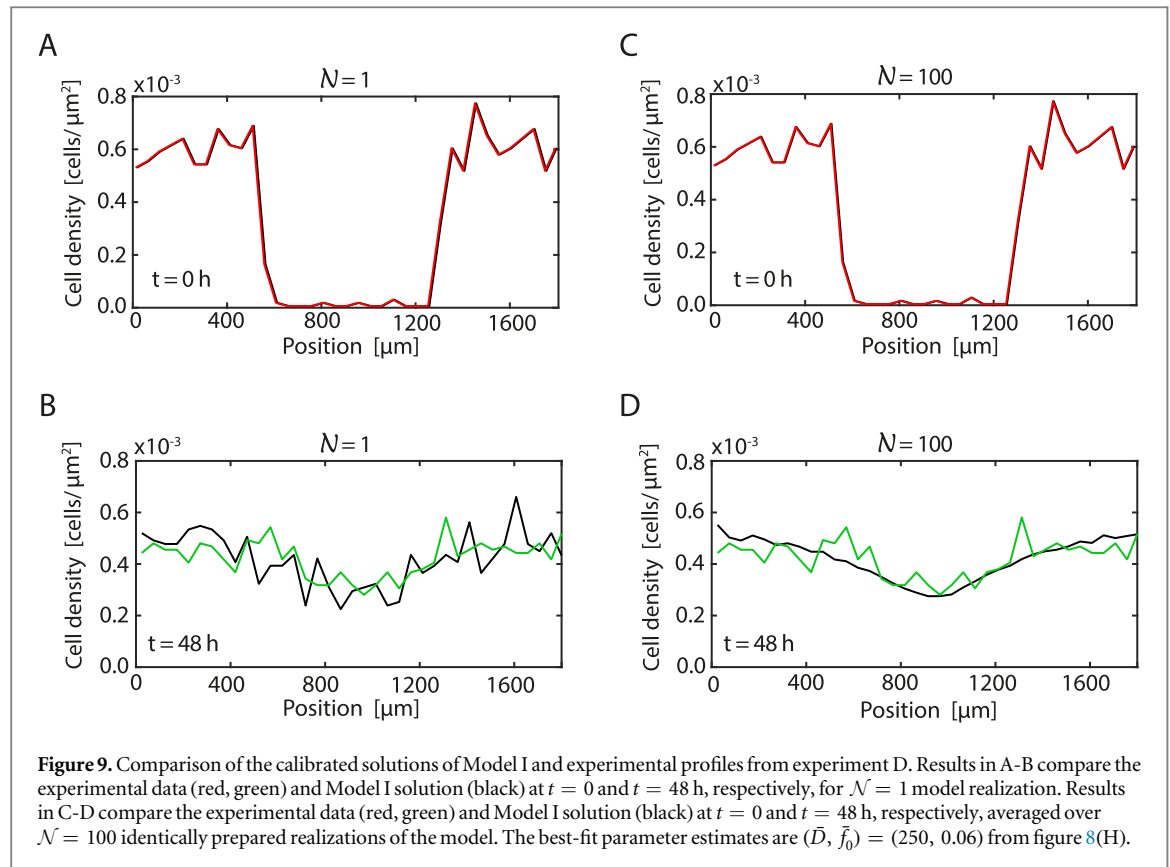
where $E(D, f_0)$ measures the discrepancy between the experimental cell density, $p_{\text{data}}(x_i, t_j)$, and the density



predicted by the model, $p_{\text{model}}(x_i, t_j)$, for given values of D and f_0 . The index i denotes column number along the x coordinate, and index j denotes time so that $j = 1, 2, 3$ and 4 correspond to the experimental time points $t = 12, 24, 36$, and 48 h, respectively. The experimental density estimates, $p_{\text{data}}(x_i, t)$, correspond to the averaged experimental density, where the average is taken across all four identically-prepared experimental replicates. We find that it is necessary to estimate $E(D, f_0)$ using averaged experimental data, rather than working with the four experimental data sets separately since the fluctuations in the data from the individual replicates lead to sufficiently large fluctuations in our estimates of $E(D, f_0)$. We will refer to the function $E(D, f_0)$ as an *error surface*. We will

visualize the surface and seek to find values of D and f_0 that minimize $E(D, f_0)$, and we denote these estimates as \bar{D} and \bar{f}_0 , respectively. All results, for Models I–IV, are obtained by numerically integrating the governing equations in two-dimensional space using a forward Euler method with a constant time step of duration $\delta t = 0.02$ h. This choice of time step is sufficiently small that our results are grid-independent. We then construct one-dimensional density distributions from the model output, $p_{\text{model}}(x_i, t_j)$, using the same procedure that is used to convert the distribution of cells in the two-dimensional experimental images into one-dimensional density profiles.

Since we have access to four identically prepared initial conditions for our experimental data set, each



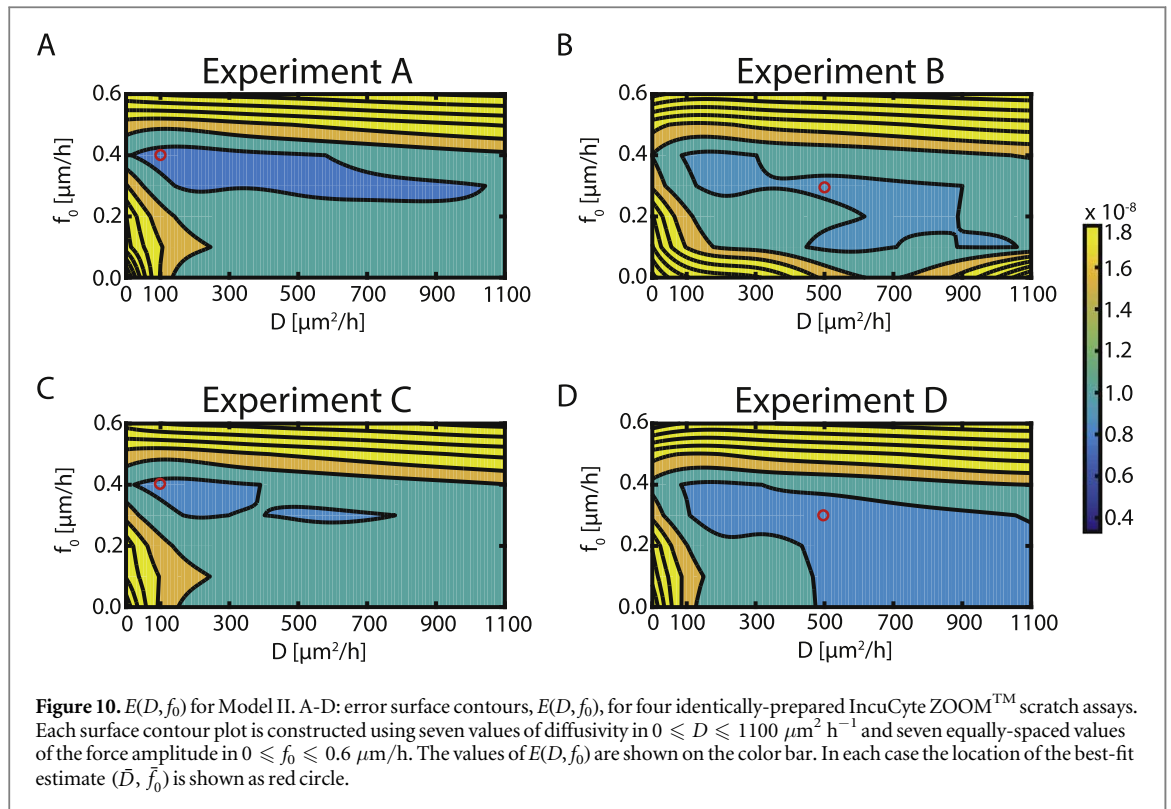
time we attempt to match the models with the averaged experimental data we repeat the process four times using the four different choices of initial conditions associated with experimental data sets A, B, C and D. This approach means that we can estimate and visualize the error surface four times for each particular model. A mathematical model that is compatible with the data ought to lead to estimates of D and f_0 that are consistent across the four initial conditions, and so we will examine Models I–IV to see whether they are capable of providing parameter estimates that are consistent across the four different initial conditions.

Previous modelling studies based on using reaction-diffusion equations indicate that estimates of diffusivity for PC-3 prostate cancer cells varies significantly, from about $300 \mu\text{m}^2 \text{h}^{-1}$ for low density conditions to approximately $1000 \mu\text{m}^2 \text{h}^{-1}$ for high density conditions [42]. Consequently, we focus our search for parameter estimates in the interval $0 \leq D \leq 1100 \mu\text{m}^2 \text{h}^{-1}$. In contrast, there are no previous estimates of f_0 for the PC-3 cell line. Since we have little initial guidance about an appropriate choice of f_0 , we first conducted a series of preliminary simulations (not shown) to determine an acceptable range of f_0 for each model. This exercise suggests that acceptable ranges are approximately $0 \leq f_0 \leq 0.11 \mu\text{m} \text{h}^{-1}$ for Model I; $0 \leq f_0 \leq 0.6 \mu\text{m} \text{h}^{-1}$ for Model II; and $0 \leq f_0 \leq 0.11 \mu\text{m} \text{h}^{-1}$ for Model III.

To estimate \bar{D} and \bar{f}_0 for Model I we estimate discrete values of $E(D, f_0)$ by using a series of numerical

solutions of equation (2) over many (D, f_0) pairs. For each parameter pair we generate an ensemble of 100 identically prepared realizations of the stochastic model and then estimate $E(D, f_0)$ by averaging the density data from the 100 identically prepared realizations. Results in figures 8(A)–(D) show the error surfaces constructed using seven values of diffusivity, $D = 0, 100, 300, 500, 700, 900, 1100 \mu\text{m}^2 \text{h}^{-1}$ and 12 equally-spaced values of the force amplitude in the interval $0 \leq f_0 \leq 0.11 \mu\text{m} \text{h}^{-1}$. All contour lines in figure 8 are obtained using the Matlab spline interpolation function `griddedInterpolant`. Since the initial positions of agents in the discrete simulations exactly correspond to the positions of cells in experimental images, given in figure 5, the magnitude of the fluctuations in the experimental data are consistent with the magnitude of the fluctuations from one realization of the stochastic model. In general, the magnitude of the fluctuations in the stochastic model decreases as the number of realisations increases, $\sim 1/\sqrt{\mathcal{N}}$. Therefore, our choice of using $\mathcal{N} = 100$ realizations leads to averaged discrete density profiles with fluctuations that are approximately one order of magnitude smaller than fluctuations in the experimental data.

Preliminary estimates of \bar{D} and \bar{f}_0 are obtained by evaluating the error surface across a relatively coarse discretisation of the parameter space. These estimates are shown in figures 8(A)–(D) as red circles. We then refine our estimates of \bar{D} and \bar{f}_0 by considering a



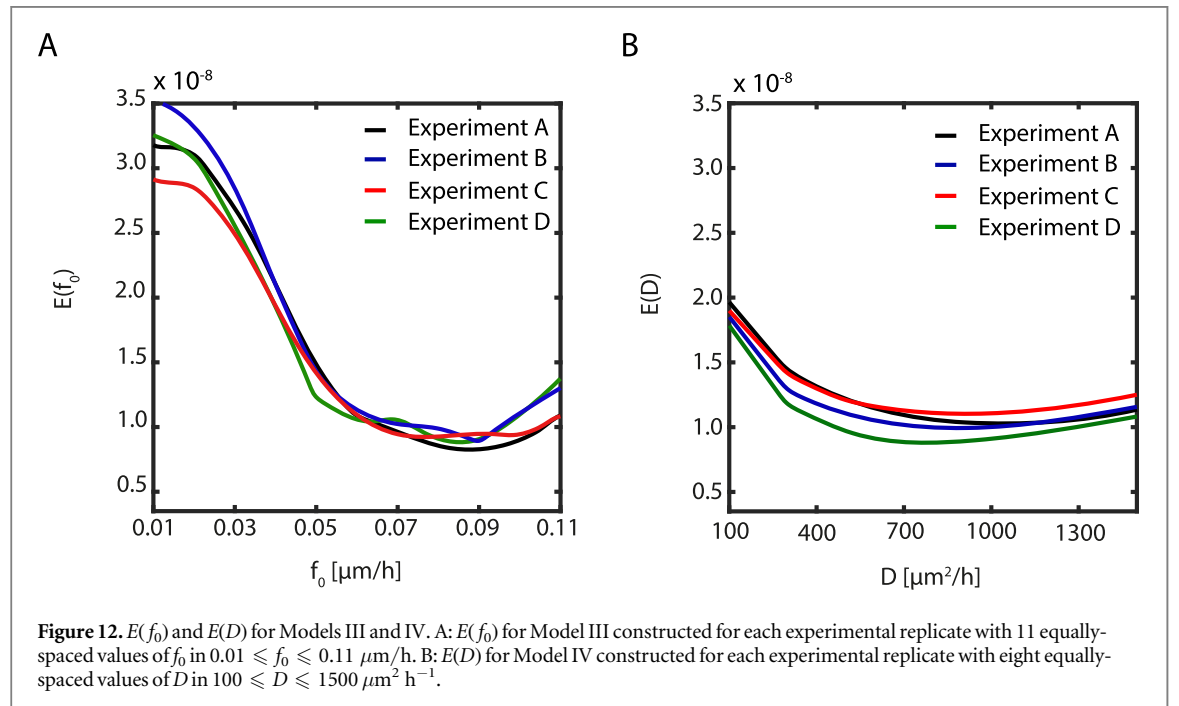
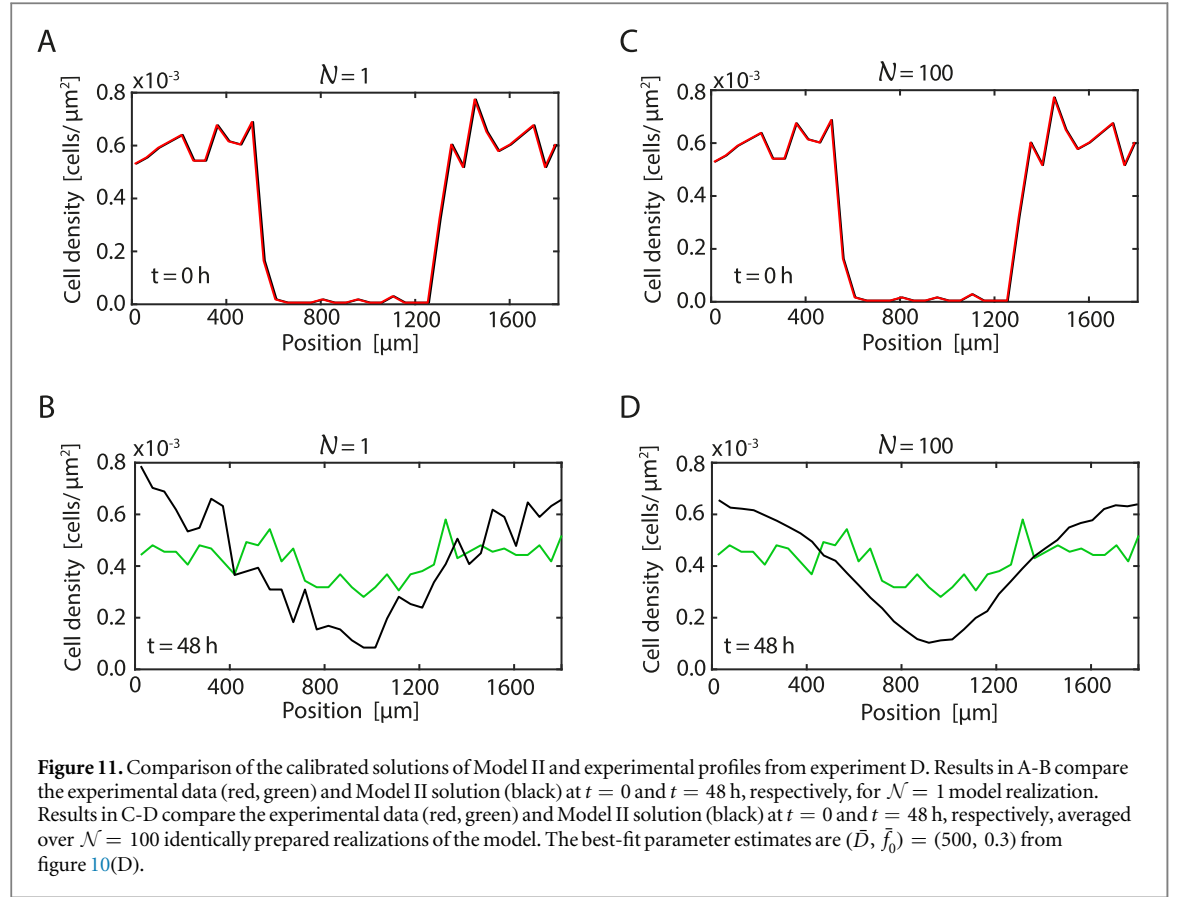
refined discretisation of a subregion surrounding each red circle in figures 8(A)–(D). These subregions are shown as red rectangles in figures 8(A)–(D). Refined plots of the error surface in figures 8(E)–(H) are obtained by calculating $E(D, f_0)$ across five equally-spaced values of D in $100 \leq D \leq 500 \mu\text{m}^2 \text{h}^{-1}$ and five equally-spaced values of f_0 in $0.05 \leq f_0 \leq 0.07 \mu\text{m/h}$. Each individual plot of $E(D, f_0)$ in figures 8(E)–(H) shows that we have a well-defined minimum from which we can estimate \bar{D} and \bar{f}_0 . Comparing data in figures 8(E)–(H) shows that our estimates of \bar{D} and \bar{f}_0 are consistent between the four experimental replicates. In fact, three of the four refined plots give remarkably consistent estimates of $(\bar{D}, \bar{f}_0) = (250, 0.06)$. Only one of the experimental replicates, shown in figure 8(E), gives slightly different parameter estimates, $(\bar{D}, \bar{f}_0) = (150, 0.065)$.

To visualize the quality of match between the experimental data and discrete profiles predicted by the stochastic model we use the best-fit parameter estimates $(\bar{D}, \bar{f}_0) = (250, 0.06)$ for experiment D. First, we solve Model I with $(\bar{D}, \bar{f}_0) = (250, 0.06)$ and calculate the density profiles from the simulations as before. Second, we superimpose density profiles from the discrete simulations with the experimental density distributions, as shown in figure 9. The choice of initial conditions in the stochastic model guarantees that we have an exact match between the experimental density profile and the simulation density profiles at $t = 0$ h. However, since we are dealing with stochastic experimental data and a stochastic mathematical model we do not expect there to be an exact match at later times. Results in figures 9(A)–(B) show that we have a

reasonable match between the calibrated simulation results and the experimental density data for a single realization of the stochastic model. Similarly, results in figures 9(C)–(D) show that we also have a good match between the simulation results and the experimental density data over 100 identically prepared realizations of the stochastic model where we see that the magnitude of the fluctuations in the averaged stochastic data are reduced.

We now turn our attention to calibrating Model II to match the experimental data. To achieve this we repeat the exact same calibration process except that we implement Model II with constant agent size. Comparing results in figures 9(A)–(D) and 10(A)–(D) shows that the best fit parameters in Model II lead to a larger value of $E(D, f_0)$. Furthermore, comparing estimates of \bar{D} and \bar{f}_0 for Model II between the four identically prepared experimental data sets shows that we have a much higher degree of variability between our parameter estimates for Model II than we did for Model I. Overall, these results suggest that Model I is more consistent with our experimental data than Model II, and so we do not proceed with any further refinement of our parameter estimates for Model II. This result shows that the traditional approach of neglecting the dynamical changes in cell size has clear impact on the ability of the model to describe the behaviour of the entire cell population.

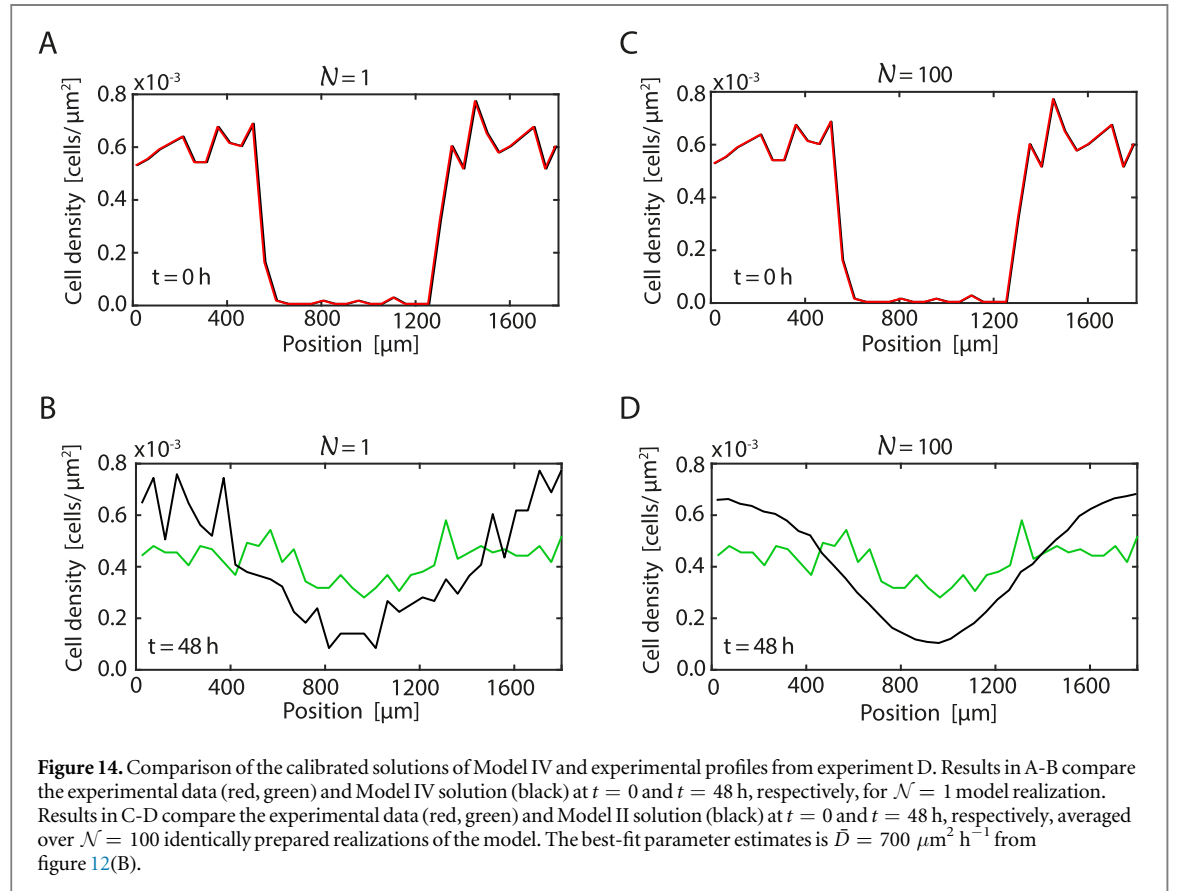
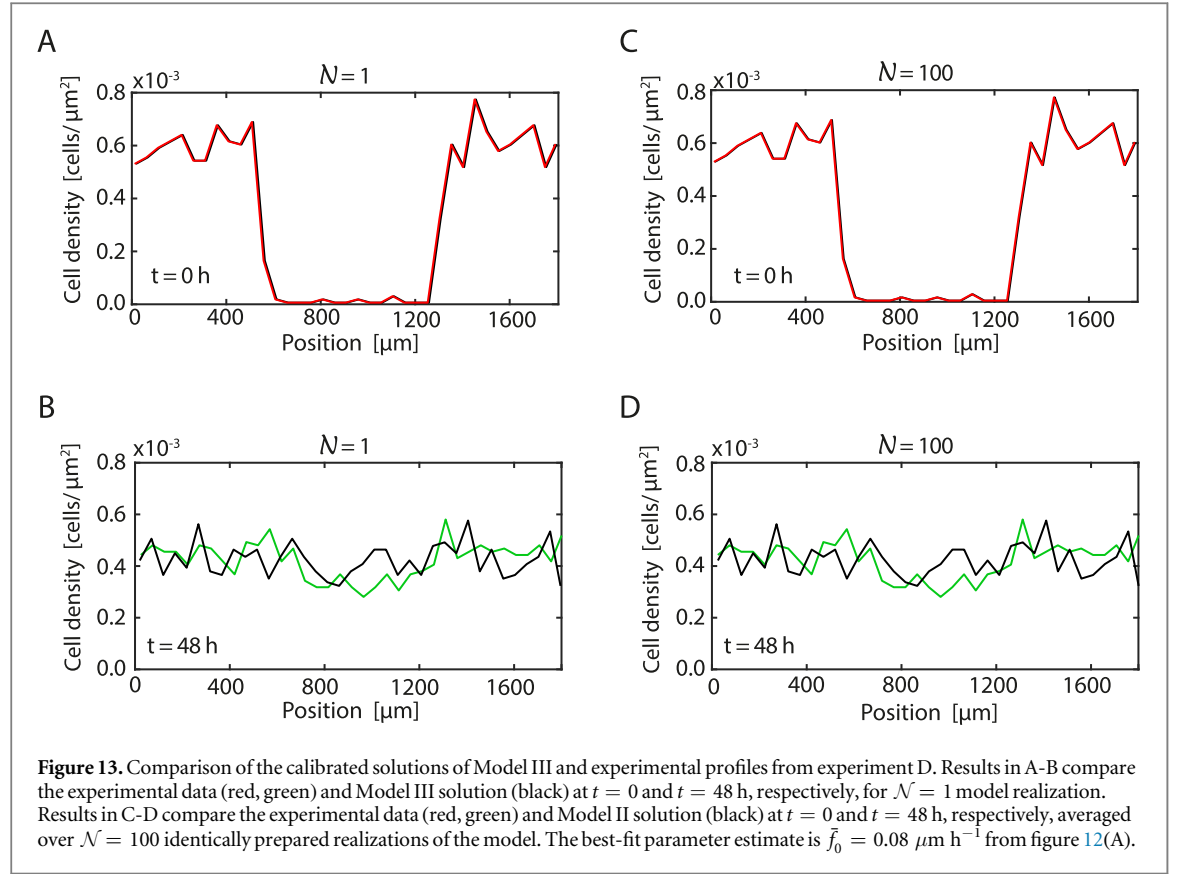
To visualize the quality of match between the experimental data and best-fit density profiles predicted by Model II we use the best-fit parameter estimates $(\bar{D}, \bar{f}_0) = (500, 0.3)$ for experimental replicate D. Again, we solve Model II with these parameter



estimates and then estimate the density profiles from those simulations. Results in figure 11 show the density profiles from the discrete simulations superimposed on the corresponding experimental density distributions. Visually, we see that the quality of match in figures 11(B) and (D) is notably poorer than the quality of match in figures 9(B) and (D). This visual

difference is consistent with the quantitative differences in $E(D, f_0)$ in figures 8 and 10.

Finally, we calibrate Models III and IV to match the experimental data. Note that Model III neglects the role of random motility so our parameter estimation involves just one parameter, f_0 . Similarly, Model IV neglects the role of intercellular forces so our



parameter estimation involves just one parameter, D . Following a now familiar procedure, we compute measures of discrepancy, $E(f_0)$ and $E(D)$, for Models

III and IV, respectively. Results in figure 12(A) show a well-defined minimum for each of the four experimental replicates for Model III, however the best-fit

estimates are in the range $0.07 \leq f_0 \leq 0.1 \mu\text{m}/\text{h}$, which is approximately double the estimate we identified previously for Model I. In contrast, results in figure 12(B) show that we have relatively poorly-defined minimum for all four experimental replicates for Model IV. In this case the best-fit estimates are in the range $700 \leq D \leq 1000 \mu\text{m}^2/\text{h}$ which is approximately four times greater than the estimates we obtained for Model I.

Results in figure 13 compare the discrete density profiles obtained using Model III parameterized with the best-fit estimate $f_0 = 0.09 \mu\text{m h}^{-1}$ and the experimental density profiles from experiment D. We note that Model III is deterministic and produces the same density distribution regardless of the number of model realizations. The quality of match between calibrated Model III and the experimental data for experimental replicate D is reasonable, however the value of $E(\bar{f}_0)$ is greater than the value of $E(\bar{D}, \bar{f}_0)$ for Model I, thereby indicating that Model I produces an improved match to the experimental data. Results in figure 14 compares discrete density profiles obtained using Model IV parameterized with the best-fit estimate $D = 700 \mu\text{m}^2 \text{h}^{-1}$ where we see no improvement in the quality of match between the calibrated mathematical model and the experimental data relative to Model I.

All results presented in this section of the main document focus on comparing the quality of match between Models I–IV and the experimental data using experimental replicate D. Similar comparisons between the best-fit solution of Models I–IV and experimental data from experimental replicates A, B and C are given in Supplementary materials (figures D.1–D.12). These additional comparisons are consistent with the comparisons made here in the main document.

4. Conclusions

In this work we use a combined experimental-mathematical modelling approach to quantitatively explore the contribution of cell-to-cell pushing and random motility in driving the movement of cell fronts. We perform a series of IncuCyte ZOOMTM scratch assay experiments in which cells are pretreated with the chemotherapy drug, Mitomycin-C. This approach is useful because Mitomycin-C suppresses proliferation, thereby allowing us to focus on the role of cell migration in the experiments.

We quantitatively assess the role of cell-to-cell interactions, including short range pushing and longer range adhesion, by calibrating an off-lattice discrete stochastic model to match our experimental data set. The mathematical model that we use accounts for random cell motility, long range cell-to-cell attraction (adhesion), short range cell-to-cell repulsion (pushing) and dynamic cell size changes. We refer to this model as *Model I*. To explore the significance of these

various cell-level mechanisms we systematically repeat the model calibration process for a range of simpler, more commonly used models. These simplified models account for: (i) random cell motility, long range cell-to-cell attraction (adhesion) and short range cell-to-cell repulsion (pushing) without any dynamical changes in cell size (*Model II*); (ii) long range cell-to-cell attraction (adhesion) and short range cell-to-cell repulsion (pushing) (*Model III*); and (iii) random cell motility only (*Model IV*).

The novelty in our work resides in the use of the time-dependent cell-to-cell interaction function, equation (3), that allows us to model the dynamical increase in the cell size during the experiments. In contrast, in most of the existing literature it is commonly assumed that cell size remains constant over the time scale of the experiments. We focus on examining the role of the cell-to-cell pushing as a result of combined effects of cell-to-cell repulsion, motility, and dynamic increases in cell size. The role of cell-to-cell pushing is not normally discussed when individual-based models are applied to the experimental data. Here, we use experimental data with cells that grow in size as a result of Mitomycin-C treatment to suppress proliferation. As such, the effects of pushing are much more pronounced in our experiments, as opposed to other experiments where the dynamic change in cell size is not as important.

The phenomenological principles of cell-to-cell interactions modelled in our work are applicable to a range of cell types. However, different cell types can have different balances of mechanisms. For example, epithelial skin cells are normally tightly packed in thin sheets with strong cell-to-cell bonds, while mesenchymal stem cells are highly motile. Consequently, applying our model to a specific cell type would require specific adjustments. Also, we do not consider biochemical pathways of cell-to-cell adhesion and motility, such as E-Cadherin junctions, trailblazing and leader-follower interactions. The leader-follower interactions have a potential to severely influence estimates of the leading edge position, which is normally used to evaluate the rate of spatial spreading of the cell colony. Incorporating these effects is theoretically possible, however, they are not always pronounced in the experimental results and are case specific.

The crucial advantage of our framework is that even in the most advanced *Model I* cell motility is parametrised by only two free parameters. This is important because we are able to explore parameter space reasonably efficiently. In contrast, many discrete models presented in the literature are not as much plausible for calibrating experimental data while addressing same aspects of individual-level behaviour. We note that there is still a certain degree of freedom associated with the choice of cell-to-cell interaction function. However, our choice is fairly typical and is suitable for our purposes.

Comparing the calibration of these four models to our experimental data provides insight into which model provides the most faithful representation of the experimental observation. Comparing estimates of $E(\bar{D}, \bar{f}_0)$ between the four models shows that Model I provides the best match to the experimental data. This result suggests that properly accounting for random motility, intercellular forces, including short range pushing and longer range attraction, as well as dynamic changes in cell size, are important for this fairly typical experimental protocol. In contrast, calibrating Models II–IV to the data always provides estimates of model parameters that give the best match to the experimental data, but this does not mean that these simpler frameworks are the best model of the underlying biological processes. This is an important result because often in the mathematical biology literature a single type of model will be used to mimic an experiment, without investigation of the more important question of whether that model provides a reasonable description of the underlying biological processes. Here, by systematically comparing the performance of four different, but related, mathematical models to our novel experimental data set, we provide insight into the underlying biological mechanisms in a way that is not possible when working with a single mathematical model in isolation. This approach of calibrating multiple competing mathematical models to match a single data set is a useful way to provide insight into the underlying biological processes, as well as providing insight into the important question of model selection [42, 69].

There are many ways in which our study could be extended to provide additional insight. A key simplifying assumption that we make in our modelling of the experiments is that we give all agents in the simulations the same size at the beginning of the simulation, $\delta(0) = \delta_0$. While our estimates of this initial size are based on experimental estimates given in figure 2(C), our approach is to represent the distribution of observed cell sizes using a sample mean. A close examination of the experimental images in figure 5 shows that there is considerable variability in the distribution of cell sizes at the beginning of the experiment. This variability is captured in the error bars in figure 2(C) but neglected in our analysis. Therefore, a reasonable extension of the current analysis would be to incorporate this initial variability into the stochastic models with a view to understanding how this initial variability influences the population-level motion of the cell fronts. In our work we do not analyse cell-to-cell correlations or spatial structure [70]. Previously, velocity-velocity correlations have been used to study cell swarming and mechanical waves during tissue expansion and similar analysis could be applied to the experimental data used in our work [44, 71, 72]. We leave these extensions for future consideration.

Acknowledgments

This work is supported by the Australian Research Council (DP170100474). REB is a Royal Society Wolfson Research Merit Award holder, would like to thank the Leverhulme Trust for a Research Fellowship and also acknowledges the BBSRC for funding via grant no. BB/R000816/1. We appreciate helpful comments provided by three anonymous referees.

ORCID iDs

Matthew J Simpson  <https://orcid.org/0000-0001-6254-313X>

References

- [1] Tao H, Berno A J, Cox D R and Frazer K A 2007 *In vitro* human keratinocyte migration rates are associated with SNPs in the KRT1 Interval *PLoS ONE* **2** e697
- [2] Friedl P, Hegerfeldt Y and Tusch M 2004 Collective cell migration in morphogenesis and cancer *Int. J. Dev. Biol.* **48** 441–9
- [3] Ramis-Conde I, Chaplain M A J and Anderson A R A 2008 Mathematical modelling of cancer cell invasion of tissue *Math. Comput. Model.* **47** 533–45
- [4] Kabla A 2012 Collective cell migration: leadership, invasion and segregation *J. Royal Soc. Interface* **9** 3268–78
- [5] Hakim V and Silberzan P 2017 Collective cell migration: a physics perspective *Rep. Prog. Phys.* **80** 076601
- [6] Maini P K, McElwain D L S and Leavesley D 2004 Travelling waves in a wound healing assay *Appl. Math. Lett.* **17** 575–80
- [7] Jin W, Penington C J, McCue S W and Simpson M J 2016 Stochastic simulation tools and continuum models for describing two-dimensional collective cell spreading with universal growth functions *Phys. Biol.* **13** 056003
- [8] Eves P, Katerinaki E, Simpson C, Layton C, Dawson R, Evans G and Mac Neil S 2003 Melanoma invasion in reconstructed human skin is influenced by skin cells—investigation of the role of proteolytic enzymes *Clin. Exp. Metastasis* **20** 685–700
- [9] Lowengrub J S, Frieboes H B, Jin F, Chuang Y-L, Li X, Macklin P, Wise S M and Cristini V 2010 Nonlinear modelling of cancer: bridging the gap between cells and tumours *Nonlinearity* **23** R1–9
- [10] Treloar K K, Simpson M J, Haridas P, Manton K J, Leavesley D I, McElwain D L S and Baker R E 2013 Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies *BMC Syst. Biol.* **7** 137
- [11] Treloar K K, Simpson M J, McElwain D L S and Baker R E 2014 Are *in vitro* estimates of cell diffusivity and cell proliferation rate sensitive to assay geometry? *J. Theor. Biol.* **356** 71–84
- [12] Nickaen M, Novak I L, Pulford S, Rumack A, Brandon J, Slepchenko B M and Mogilner A 2017 A free-boundary model of a motile cell explains turning behavior *PLoS Comput. Biol.* **13** e1005862
- [13] Codling E A, Plank M J and Benhamou S 2008 Random walk models in biology *J. R. Soc. Interface* **5** 813–34
- [14] Nardini J T, Chapnick D A, Liu X and Bortz D M 2016 Modeling keratinocyte wound healing dynamics: cell-cell adhesion promotes sustained collective migration *J. Theor. Biol.* **400** 103–17
- [15] Steinberg M S 1996 Adhesion in development: an historical overview *Dev. Biol.* **180** 377–88
- [16] Nadell C D, Foster K R and Xavier J B 2010 Emergence of spatial structure in cell groups and the evolution of cooperation *PLoS Comput. Biol.* **6** e1000716

- [17] Painter K J, Armstrong N J and Sherratt J A 2010 The impact of adhesion on cellular invasion processes in cancer and development *J. Theor. Biol.* **7** 9720
- [18] Carmona-Fontaine C et al 2011 Complement fragment C3a controls mutual cell attraction during collective cell migration *Dev. Cell.* **21** 1026–37
- [19] George M, Bullo F and Campas O 2017 Connecting individual to collective cell migration *Sci. Rep.* **264** 1057–67
- [20] Etienne-Manneville S 2014 Neighborly relations during collective migration *Curr. Opin. Cell Biol.* **30** 51–9
- [21] Treppe X, Wasserman M R, Angelini T E, Millet E, Weitz D A, Butler J P and Fredberg J J 2009 Physical forces during collective cell migration *Nat. Phys.* **5** 426–30
- [22] McLennan R et al 2015 Neural crest migration is driven by a few trailblazer cells with a unique molecular signature narrowly confined to the invasive front *Development* **142** 1–12
- [23] Köpf M H and Pismen L M 2013 A continuum model of epithelial spreading *Soft. Matter.* **9** 3727–34
- [24] Park J-A, Atia L, Mitchel J A, Fredberg J J and Butler J P 2016 Collective migration and cell jamming in asthma, cancer and development *J. Cell Sci.* **129** 3375–83
- [25] Simpson M J, Zhang D C, Mariani M, Landman K A and Newgreen D F 2007 Cell proliferation drives neural crest cell invasion of the intestine *Dev. Biol.* **302** 553–68
- [26] Weijer C J 2009 Collective cell migration in development *J. Cell Sci.* **122** 3215–23
- [27] Simpson M J, Landman K A and Newgreen D F 2006 Chemotactic and diffusive migration on a nonuniformly growing domain: numerical algorithm development and applications *J. Comput. Appl. Math.* **192** 282–300
- [28] Hawkins R J, Piel M, Faure-Andre G, Lennon-Dumenil A M, Joanny J F, Prost J and Voituriez R 2009 Pushing off the walls: a mechanism of cell motility in confinement *Phys. Rev. Lett.* **102** 058103
- [29] Fisher R A 1937 The wave of advance of advantageous genes *Ann. Eug.* **7** 355–69
- [30] Sherratt J A and Murray J D 1990 Models of epidermal wound healing *Proc. Biol. Sci.* **241** 29–36
- [31] Johnston S T, Shah E T, Chopin L K, McElwain D L S and Simpson M J 2015 Estimating cell diffusivity and cell proliferation rate by interpreting IncuCyte ZOOM™ assay data using the Fisher-Kolmogorov model *BMC Syst. Biol.* **9** 38
- [32] Simpson M J, Landman K A, Hughes B D and Newgreen D F 2006 Looking inside an invasion wave of cells using continuum models: proliferation is the key *J. Theor. Biol.* **243** 343–60
- [33] Buske P, Galle J, Barker N, Aust G, Clevers H and Loeffler M 2011 A comprehensive model of spatio-temporal stem cell and tissue organisation in the intestinal crypts *PLoS Comput. Biol.* **7** e1001045
- [34] Gardiner B S, Wong K K L, Joldes G R, Rich A J, Tan C W, Burgess A W and Smith D W 2015 Discrete element framework for modelling extracellular matrix, deformable cells and subcellular components *PLoS Comput. Biol.* **11** e1004544
- [35] Osborne J M, Fletcher A G, Pitt-Francis J M, Maini P K and Gavaghan D J 2017 Comparing individual-based approaches to modelling the self-organization of multicellular tissues *PLoS Comput. Biol.* **13** e1005387
- [36] Nan P, Walsh D M, Landman K A and Hughes B D 2018 Distinguishing cell shoving mechanisms *PLoS One* **13** e0193975
- [37] Yates C, Parker A and Baker R E 2015 Incorporating pushing in exclusion-process models of cell migration *Phys. Rev. E* **91** 052711
- [38] Sadeghi H M, Seitz B, Hayashi S, LaBree L and McDonnell P J 1998 *In vitro* effects of Mitomycin-C on human keratinocytes *J. Refract. Surg.* **14** 534–40
- [39] Simpson M J, Treloar K K, Binder B J, Haridas P, Manton K J, Leavesley D I, McElwain D L S and Baker R E 2013 Quantifying the roles of cell motility and cell proliferation in a circular barrier assay *J. Royal Soc. Interface* **10** 20130007
- [40] Essen BioScience: incuCyte ZOOM live cell imaging. Available from: <http://www.essenbioscience.com/essen-products/incucyte/> (Accessed: April 2019).
- [41] Kaighn M E, Narayan K S, Ohnuki Y, Lechner J F and Jones L W 1979 Establishment and characterization of a human prostatic carcinoma cell line (PC-3) *Invest. Urol.* **17** 16–23
- [42] Jin W, Shah E T, Penington C J, McCue S W, Chopin L K and Simpson M J 2016 Reproducibility of scratch assays is affected by the initial degree of confluence: experiments, modelling and model selection *J. Theor. Biol.* **390** 136–45
- [43] Kumari R, Sharma A, Ajay A K and Bhat M K 2009 Mitomycin C induces bystander killing in homogeneous and heterogeneous hepatoma cellular models *Mol. Cancer* **8** 87
- [44] Thili S, Gauquelin E, Li B, Cardoso O, Ladoux B, Delanoë-Ayari H and Graner F 2018 Collective cell migration without proliferation: density determines cell velocity and wave velocity *R. Soc. Open Sci.* **5** 172421
- [45] Sommer C, Strähle C, Köthe U and Hamprecht F A 2011 *Ilastik: Interactive learning and segmentation toolkit In: Biomedical imaging: From nano to macro Proceeding of Eighth IEEE International Symposium on Biomedical Imaging*
- [46] Carpenter A E et al 2016 CellProfiler: image analysis software for identifying and quantifying cell phenotypes *Genome Biol.* **7** R100
- [47] Sepulveda N, Petitjean L, Cochet O, Grasland-Mongrain E, Silberzan P and Hakim V 2016 Collective cell motion in an epithelial sheet can be quantitatively described by a stochastic interacting particle model *PLoS Comput. Biol.* **9** e1002944
- [48] Chiou K K, Hufnagel L and Shraiman B I 2012 Mechanical stress inference for two dimensional cell arrays *PLoS Comput. Biol.* **8** e1002512
- [49] Zhu J, Liang L, Jiao Y, Liu L and (US-China Physical Sciences-Oncology Alliance) 2015 Enhanced invasion of metastatic cancer cells via extracellular matrix interface *PLoS One* **10** e0118058
- [50] De Jesus A M, Aghvami M and Sander E A 2016 A combined *in vitro* imaging and multi-scale modeling system for studying the role of cell matrix interactions in cutaneous wound healing *PLoS One* **11** e0148254
- [51] Peirce S M 2008 Computational and mathematical modeling of angiogenesis *Microcirculation* **15** 739–51
- [52] Ermentrout G B and Edelstein-Keshet L 1993 Cellular automata approaches to biological modeling *J. Theor. Biol.* **160** 97–133
- [53] Wynn M L, Rupp P, Trainor P A, Schnell S and Kulesa P M 2013 Follow-the-leader cell migration requires biased cell-cell contact and local microenvironmental signals *Phys. Biol.* **10** 035003
- [54] Middleton A M, Fleck C and Grima R 2014 A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion *J. Theor. Biol.* **359** 220–32
- [55] Mousavi S J, Doblare M and Doweidar M H 2014 Computational modelling of multi-cell migration in a multi-signalling substrate *Phys. Biol.* **11** 026002
- [56] Berg H C 1993 *Random Walks in Biology* (Princeton: Princeton University Press)
- [57] Nikolic D L, Boettiger A N, Bar-Sagi D, Carbeck J D and Shvartsman S Y 2006 Role of boundary conditions in an experimental model of epithelial wound healing *Am. J. Physiol. Cell Physiol.* **291** C68–75
- [58] Warne D J, Baker R E and Simpson M J 2019 Using experimental data and information criteria to guide model selection for reaction-diffusion problems in mathematical biology *Bull. Math. Biol.* (<https://doi.org/10.1007/s11538-019-00589-x>)
- [59] Matsiaka O M, Penington C J, Baker R E and Simpson M J 2018 Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics *Bull. Math. Biol.* **80** 738–57
- [60] Ranft J, Basan M, Elgeti J, Joanny J-F, Prost J and Jülicher F 2010 Fluidization of tissues by cell division and apoptosis *PNAS* **107** 20863–8
- [61] Huang S, Brangwynne C P, Parker K K and Ingber D E 2005 Symmetry-breaking in mammalian cell cohort migration during tissue pattern formation: role of random-walk persistence *Cell Motil. Cytoskeleton* **61** 201–13

- [62] Cai D, Chen S C, Prasad M, He L, Wang X, Choesmel-Cadamuro V, Sawyer J K, Danuser G and Montell D J 2014 Mechanical feedback through E-cadherin promotes direction sensing during collective cell migration *Cell* **157** 1146–59
- [63] Peglion F, Lense F and Etienne-Manneville S 2014 Adherens junction treadmilling during collective migration *Nat. Cell Biol.* **16** 639–51
- [64] Bruna M and chman S J 2012 Excluded-volume effects in the diffusion of hard spheres *Phys. Rev. E* **85** 011103
- [65] De Palo G, Yi D and Endres R G 2017 A critical-like collective state leads to longrange cell communication in Dictyostelium discoideum aggregation *PLoS Biol.* **15** e1002602
- [66] Matsiaka O M, Penington C J, Baker R E and Simpson M J 2017 Continuum approximations for lattice-free multi-species models of collective cell migration *J. Theor. Biol.* **422** 1–11
- [67] Tersoff J 1988 New empirical approach for the structure and energy of covalent systems *Phys. Rev. B* **37** 6991–7000
- [68] Simpson M J, Jin W, Vittadello S T, Tambyah T A, Ryan J M, Gunasingh G, Haass N K and McCue S W 2018 Stochastic models of cell invasion with fluorescent cell cycle indicators *Physica A* **510** 375–86
- [69] Ciupe M S, Bivort B L, Bortz D M and Nelson P W 2006 Estimating kinetic parameters from HIV primary infection data through the eyes of three different mathematical models *Math. Biosci.* **200** 1–17
- [70] Binder B J and Simpson M J 2013 Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions *Phys. Rev E* **88** 022705
- [71] Serra-Picamal X *et al* 2012 Mechanical waves during tissue expansion *Nat. Phys.* **8** 628
- [72] Zaritsky A *et al* 2014 Propagating waves of directionality and coordination orchestrate collective cell migration *PLoS Comput. Biol.* **10** e1003747