Continuum descriptions of spatial spreading for heterogeneous cell populations: theory and experiment

Oleksii M Matsiaka¹, Ruth E Baker², *Matthew J Simpson¹

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

² Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, United Kingdom.

Abstract

Variability in cell populations is frequently observed in both *in vitro* and *in vivo* settings. Intrinsic differences within populations of cells, such as differences in cell sizes or differences in rates of cell motility, can be present even within a population of cells from the same cell line. We refer to this variability as cell *heterogeneity*. Mathematical models of cell migration, for example, in the context of tumour growth and metastatic invasion, often account for both undirected (random) migration and directed migration that is mediated by cell-to-cell contacts and cell-to-cell adhesion. A key feature of standard models is that they often assume that the population is composed of identical cells with constant properties. This leads to relatively simple single-species *homogeneous* models that neglect the role of heterogeneity. In this work, we use a continuum modelling approach to explore the role of heterogeneity in spatial spreading of cell populations. We employ a three-species heterogeneous model of cell motility that explicitly incorporates different types of experimentally-

Preprint submitted to Journal of Theoretical Biology

16 August 2019

motivated heterogeneity in cell sizes: (i) monotonically decreasing; (ii) uniform; (iii) non-monotonic; and (iv) monotonically increasing distributions of cell size. Comparing the density profiles generated by the three-species heterogeneous model with density profiles predicted by a more standard single-species homogeneous model reveals that when we are dealing with monotonically decreasing and uniform distributions a simple and computationally efficient single-species homogeneous model can be remarkably accurate in describing the evolution of a heterogeneous cell population. In contrast, we find that the simpler single-species homogeneous model performs relatively poorly when applied to non-monotonic and monotonically increasing distributions of cell sizes. Additional results for heterogeneity in parameters describing both undirected and directed cell migration are also considered, and we find that similar results apply.

Key words: Cell migration; heterogeneity; continuum description; scratch assay

^{*} Corresponding author

Email address: matthew.simpson0qut.edu.au, *Telephone* + 617 31385241, Fax + 617 3138 2310 (*Matthew J Simpson¹).

1 **1** Introduction

In vitro cell migration experiments play an important role in the discovery and testing of putative drug treatments, the study of malignant tumour growth and metastasis, as well as tissue regeneration and repair (Savla et al., 2004; Sengers et al., 2007; Tremel et al., 2009; Sarapata and de Pillis, 2010; Gerlee, 2013; Edmondson et al., 2014; Shah et al., 2016). Mathematical models 6 of many biological processes involved in these experiments normally require 7 certain assumptions to make the problem mathematically and computationally tractable. When modelling large populations of cells, one of the most 9 intuitive approaches is to assume that all cells have fixed properties, such 10 as assuming all cells have constant size and constant diffusivity (Sherratt and 11 Murray, 1990; Galle et al., 2005; Simpson et al., 2013). In this framework a cell 12 population is considered to be a *homogeneous* population, and single-species 13 homogeneous models are routinely invoked (Maini et al., 2004a; Maini et al., 14 2004b; Sepulveda et al., 2013; Simpson et al., 2013; George et al., 2017; Vo et 15 al., 2015). Single-species homogeneous models are much less computationally 16 expensive than more elaborate multi-species heterogeneous models and, as a 17 result, are frequently used relative to multi-species counterparts. In addition, 18 multi-species frameworks usually involve a significantly larger number of free 19 model parameters that we may have little prior knowledge about and so the 20 process of calibrating multi-species heterogeneous models to match experimen-21 tal observations is significantly more challenging than calibrating single-species 22 homogeneous models. This is an important consideration because it is well-23 known that parameterising mathematical models of biological processes can 24 be challenging, often requiring computationally-intensive methods (Pozzobon 25



²⁶ and Perré, 2018; Warne et al. 2019).

Fig. 1. Heterogeneity in a population of PC-3 prostate cancer cells (Kaighn et al., 1979). (a) Experimental image of an advancing cell population and corresponding cell size distribution. The red solid line denotes position of the leading edge. (b) Detailed image of the subregion denoted in the blue rectangle in Figure 1(a). (c) Cell size distribution with a bin size of $15 \,\mu$ m. The cell size distribution is obtained from the sample of 184 cells randomly selected from the population. (d) Cell size distribution with a bin size of $2.3 \,\mu$ m. The histogram in Figure 1(d) is constructed using the same sample of 184 cells.

Although *heterogeneity* in cell populations is frequently observed in experiments, there is relatively little guidance or consensus in the literature about how to incorporate such heterogeneity into the mathematical models used to replicate and predict such experiments (An et al., 2001; Altschuler et al., 2010; Menon et al., 2018). Figure 1(a)-(b) shows a typical experiment where we can clearly visually observe cells of different sizes. The measured cell size distribution in Figure 1(c) quantifies this heterogeneity in cell sizes and raises the

question if the most straightforward approach of applying a single-species ho-34 mogeneous model can be reasonably used to predict the spatial spreading of 35 this clearly heterogeneous population. In addition to the clear visual hetero-36 geneity in cell sizes, it could be relevant to consider that cells of different sizes 37 can exhibit different behaviour such as different rates of motility, or differ-38 ent mechanical properties including resistance to deformation and adhesion. 39 Therefore, it could be possible that there are multiple types of heterogeneity 40 acting in even this very simple experiment. Previously, heterogeneity in cell 41 populations has been introduced in both discrete and continuum models of 42 cell motility (Simpson et al., 2014; Jin et al., 2016b; Sundstrom et al., 2016; 43 Matsiaka et al., 2017). Previous work has also attempted to estimate parame-44 ters in heterogeneous models that describe glioblastoma progression (Rutter et 45 al., 2018). However, these previous modelling studies do not address the basic 46 question of identifying whether it is absolutely necessary to apply a multi-47 species heterogeneous models to obtain a faithful description of the behaviour 48 of the heterogeneous population and whether different forms of heterogeneity 49 affect the answer to this fundamental question. 50

In our work we use an experimentally-motivated approach to investigate the 51 role of heterogeneity in two-dimensional scratch assays, and we compare the 52 performance of a single-species homogeneous model relative to a heteroge-53 neous multi-species model. We use numerical solutions of the multi-species 54 heterogeneous model to produce synthetic test data that we use to investigate 55 the performance of a simpler single-species homogeneous model. To mimic 56 experimental data, such as depicted in Figure 1, we use the multi-species 57 continuum approach introduced by Matsiaka et al. (2017). To keep our work 58 tractable, we describe the heterogeneity by dividing the total population into 59

three subpopulations with varying properties. The choice of working with three 60 subpopulations allows us to keep the model computationally tractable while 61 capturing important differences in the population properties, as illustrated 62 in Figure 1(d). Throughout this work we consider four distinct distributions 63 of cell sizes: (i) monotonically decreasing (Set Ia); (ii) uniform (Set Ib); (iii) 64 non-monotonic (Set Ic); and (iv) monotonically increasing (Set Id). The mono-65 tonically decreasing distribution, as shown in Figure 3(a), is a fairly accurate 66 approximation of the experimentally observed cell size distribution in Figure 67 1(d). The other three kinds of distributions are included in our work for com-68 pleteness. Our findings suggest that, for certain cell size distributions, namely 69 monotonically decreasing and uniform distributions, the single-species homo-70 geneous model performs remarkably well with an excellent match between the 71 density profiles generated by the three-species heterogeneous model and den-72 sity profiles predicted by its single-species homogeneous analogue. Therefore, 73 our results imply that applying a single-species homogeneous model to describe 74 experiments with monotonically decreasing or uniform cell size distributions 75 might be sufficient for accurately predicting population-level behaviour. In 76 contrast, the data with non-monotonic and monotonically increasing cell size 77 distributions might require the application of multi-species models to account 78 for differences in population. 79

This manuscript is organised in the following way. In Section 2 we describe experimental data for a series of two-dimensional scratch assays that clearly involve a significant level of heterogeneity among the population. In Section 3 we introduce a mathematical model of the cell motility and adhesion. In particular, we focus on two analogues of the mathematical model: (i) a three-species heterogeneous model of cell motility where parameters including cell size, cell

diffusivity and cell adhesion strength can vary between the subpopulations; 86 and (ii) a more traditional single-species homogeneous model of cell motility 87 where all cells in the population are treated as having the same cell size, cell 88 diffusivity and cell adhesion strength. Results in Section 4 compare perfor-89 mance of the single-species homogeneous model as applied to data generated 90 using the three-species heterogeneous model for different cell size distributions. 91 Additional results presented in the Supplementary Material explore the role 92 of: (i) heterogeneity in undirected (diffusive) migration, Set II; and (ii) hetero-93 geneity in directed (adhesion/cell-to-cell contacts) migration, Set III. Finally, 94 in Section 5 we summarise our result and propose potential extensions. 95

96 2 Experimental data

Monolayer scratch assays are performed using the IncuCvte ZOOMTM svs-97 tem (Essen BioScience). In all experiments we use the PC-3 prostate cancer 98 cell line (Kaighn et al., 1979) from the American Type Culture Collection 90 (ATCCTM, Manassas, USA). After growing, cells are removed from the flask 100 using TrypLETM (ThermoFisher Scientific) in phosphate buffered saline, re-101 suspended in growth medium and seeded at a density of 20,000 cells per well in 102 96-well ImageLock plates (Essen BioScience). The diameter of each individual 103 well is 9000 μ m. 104

Mitomycin-C is added at a concentration of 10 g/mL for two hours before a scratch is made in the monolayer of cells (Sadeghi et al., 1998). Mitomycin-C is a chemotherapy drug that blocks DNA replication and, consequently, stops proliferation. As a result of treatment the number of cells in the assay remains approximately constant since cells neither proliferate or die on the timescale of

the experiment. Often scratch assays are performed using mitomycin-C treated 110 cells so that the experiment focuses only upon the role of cell migration as 111 opposed to the combined effects of cell migration and cell proliferation. A 112 WoundMakerTM (Essen BioScience) is used to create identical scratches in 113 the uniformly distributed populations. Medium is aspirated after scratching; 114 each well is washed twice and refilled with fresh medium (100 μ L). Plates are 115 incubated in the IncuCyte ZOOMTM and photographed every two hours for 116 48 hours. In total, these experiments are performed in eight of the 96 wells 117 on the 96-well plate. In our work we use one of the experimental replicates at 118 t = 0 h, shown in Figure 1, to quantify the heterogeneity in a cell population. 119

To quantify the heterogeneity in cell size we randomly select 184 cells from 120 the experimental image in Figure 1(a) at t = 0 h. Assuming each cell can be 121 treated as a disc, we estimate the equivalent diameter of each individual cell 122 using the following approach. First, we use the histogram tool in Photoshop 123 CS5 to count a number of pixels in the area occupied by each individual cell. 124 The pixel count is converted to an area, A. Second, we estimate the equivalent 125 diameter, $\delta = \sqrt{4A/\pi}$ and use this data to produce histograms to illustrate 126 and visualise the variability in cell size within the experiment. The result-127 ing cell size distribution, presented as a histogram constructed with bin width 128 $2.3 \,\mu\text{m}$, is shown in Figure 1(c). The bin width $2.3 \,\mu\text{m}$ is chosen to demonstrate 129 the fine structure within the cell population that is not normally incorporated 130 in mathematical models of cell migration. However, the computational sim-131 ulation of a population with the cell size distribution shown in Figure 1(c) 132 is impractical since it would require significant computational resources to 133 simulate the dynamics of 17 distinct subpopulations. As a compromise, we 134 increase the bin width to reduce the number of distinct subpopulations while 135

still retaining a sufficient number of bins to allow us to broadly characterise the heterogeneity in the population. Figure 1(d) demonstrates the histogram of cell sizes constructed using the same sample of cells with a larger bin size width of 15 μ m. Here, we have three subpopulations that capture the key trends in the heterogeneity in Figure 1(c) without needing to deal with 17 distinct subpopulations.

In this work we use experimental data to extract the cell size distribution 142 at t = 0 h and use this data to generate the initial conditions in the three-143 species heterogeneous model (Set Ia, Figure 3). An interesting side effect of 144 Mitomycin-C pretreatment is that cells increase in size abnormally fast com-145 pared to similar experiments without pretreatment (Matsiaka et al., 2018). As 146 a result of pretreatment, the cell size distribution changes significantly with 147 time, which, in turn, represents an additional degree of freedom in the prob-148 lem. To keep our work tractable, we consider the most fundamental problem 149 where we treat the cell size distribution as being constant through time, and 150 we leave an extension to the case where the cell size distribution varies with 151 time for future analysis. 152

153 **3** Mathematical model

¹⁵⁴ Discrete, stochastic models are often used to describe the spatial spreading of
^a population of cells, especially when the population of cells is not too large.
^{Here,} cells move and interact with each other via predefined force function, as
^{illustrated} schematically in Figure 2 (Newman and Grima, 2004; Callaghan et
^a al., 2006; Hasenauer et al., 2011; Frascoli et al. 2013; Osborne et al., 2017).
^{This} approach is *individual-based* in the sense that knowledge about the move-



Fig. 2. (a) An idealisation of the front-like distribution of cells in the experimental design shown in Figure 1(a). Here all cells are of constant size. F_{ij} is the interaction force between cell *i* and cell *j*. The vertical dashed line represents the approximate leading edge of the population. (b) A typical cell-to-cell interaction force function in the form of the modified Morse potential, Z(r), (Equation (3.7)) used to mimic adhesion and repulsion between individual cells. The vertical dashed line represents the diameter of individual agents, δ . The horizontal line at Z(r) = 0 shows the change from long-range attraction (Z(r) < 0 for $r > \delta$) to short-range repulsion (Z(r) > 0 for $r < \delta$).

¹⁶⁰ ment of each individual is essential to infer the evolution of a density on the ¹⁶¹ population-level scale. One of the most popular individual-based modelling ¹⁶² approaches makes the assumption that the motion of each cell can be de-¹⁶³ scribed by a Langevin stochastic differential equation (Newman and Grima, ¹⁶⁴ 2004; Middleton et al., 2014). As such, the system of N cells is described by ¹⁶⁵ a system of N stochastic differential equations of the form

$$\frac{\mathrm{d}\vec{x}_i}{\mathrm{d}t} = \sum_{i \neq j} \vec{F}_{ij} + \vec{\xi}_i, \qquad (3.1)$$

where \vec{x}_i is the position vector of the *i*th cell, \vec{F}_{ij} is the interaction force between cells *i* and *j*, and $\vec{\xi}_i$ is the random stochastic force acting upon cell *i* (Middleton et al., 2014; George et al., 2017; Osborne et al., 2017). The interac-

tion force, \vec{F}_{ij} , can be used to parametrise various features of cell populations, 169 including heterogeneity. In fact, it is relatively straightforward to model het-170 erogeneity in cell sizes in a discrete framework since the interaction force, \vec{F}_{ij} , 171 can be chosen to explicitly include the cell size as a parameter (Matsiaka et al., 172 2018). Here we can easily differentiate the population into an arbitrary num-173 ber of subpopulations by assigning the value of the cell size to each member of 174 the population. Despite the many advantages of this kind of individual-based 175 modelling approach, such individual-based models are computationally inef-176 ficient as the number of cells, N, increases. This is because the computation 177 time required to simulate such models increases with N. 178

In contrast, continuum models based on partial differential equations (PDEs) 179 are much more convenient to model large cell populations because the time 180 taken to solve continuum PDE models is independent of the size of the pop-181 ulation (Sherratt and Murray, 1990; Sheardown and Cheng, 1995; Cai et al., 182 2007; Wise et al., 2008). Often, PDE models are derived using continuum-183 limit approximations of underlying discrete models and, as such, are able to 184 retain certain features of a discrete model (Middleton et al., 2014; O'Dea and 185 King, 2012). In this work we focus on a continuum model that is derived 186 by taking the limit of a three-species heterogeneous individual-based model 187 (Matsiaka et al., 2017). This approach allows us to conceptually incorporate 188 key features of the heterogeneous cell populations into a discrete modelling 189 framework, and then using a computationally efficient approach to solve the 190 resulting continuum-limit PDE description of the underlying heterogeneous 191 model. 192

¹⁹³ We note that, due to the geometry of experiments presented in Figure 1, we

are interested in the net movement of cells in only one direction, in this case 194 the horizontal direction (Jin et al., 2016a). This is due to the fact that the 195 net flux of cells in the vertical direction is, on average, zero because of the 196 symmetry in the initial conditions of a scratch assay. Consequently, we focus 197 on a one-dimensional continuum model and consider the evolution of the total 198 cell population in the horizontal direction only. The use of a one-dimensional 199 framework to describe two-dimensional scratch assays has been previously 200 demonstrated to be a convenient approach to reduce the computational com-201 plexity while still describing the key features of the experiment (Matsiaka et 202 al., 2018). 203

Here we employ a mean field model describing the spatial spreading of a population of cells composed of three distinct subpopulations. In one-dimension,
the model can be written as

$$\frac{\partial p^{(1)}(x,t)}{\partial t} = D_1 \Delta p^{(1)}(x,t) + \nabla (p^{(1)}(x,t) V^{(1,1)}(x,t)) \qquad (3.2)
- \sum_{i=1}^3 n_i \nabla (p^{(1)}(x,t) V^{(1,i)}(x,t)),
\frac{\partial p^{(2)}(x,t)}{\partial t} = D_2 \Delta p^{(2)}(x,t) + \nabla (p^{(2)}(x,t) V^{(2,2)}(x,t)) \qquad (3.3)
- \sum_{i=1}^3 n_i \nabla (p^{(2)}(x,t) V^{(2,i)}(x,t)),
\frac{\partial p^{(3)}(x,t)}{\partial t} = D_3 \Delta p^{(3)}(x,t) + \nabla (p^{(3)}(x,t) V^{(3,3)}(x,t))
- \sum_{i=1}^3 n_i \nabla (p^{(3)}(x,t) V^{(3,i)}(x,t)),
V^{(l,i)}(x,t) = \int_{\Omega} F^{(l,i)}(x-y) p^{(i)}(y,t) \, \mathrm{d}y, \qquad (3.5)$$

where $p^{(1)}(x,t)$, $p^{(2)}(x,t)$, and $p^{(3)}(x,t)$ are the cell densities associated with each subpopulation and depend on position x and time t. In this heterogeneous model, D_1 , D_2 , and D_3 are diffusivities of subpopulations 1, 2, and 3, n_1 , n_2 , and n_3 are the numbers of cells in each subpopulation, and $V^{(l,i)}(x,t)$ is the velocity field of subpopulation l induced by subpopulation i (Matsiaka et al., 2017). The diffusivity constants parameterise the undirected migration of each subpopulation and the velocity fields describe the directed migration of each subpopulation that is driven by a combination of cell-to-cell adhesion and crowding effects.

The interaction force between subpopulations l and i that describes directed migration is given by

$$F^{(l,i)}(x-y) = f_0^{(i)} \mathcal{Z}_i(r) \operatorname{sgn}(x-y),$$
(3.6)

where $f_0^{(i)}$ is the dimensional amplitude of the interaction force acting on 218 subpopulation $i, \mathcal{Z}_i(r)$ is a dimensionless function that parametrises different 219 features of the cell-to-cell interactions, and sqn is the signum function. We 220 choose to include long-range attraction that models cell-to-cell adhesion, and 221 a short-range repulsion that reflects volume exclusion effects (Frascoli et al., 222 2013; Painter et al., 2010). A number of different phenomenological laws, $\mathcal{Z}_i(r)$, 223 are used to model repulsive and adhesive intercellular forces (Murray et al., 224 2009; Jeon et al., 2010; Middleton et al., 2014). In our work we adopt modified 225 Morse potential in the form 226

$$\mathcal{Z}_{i}(r) = \begin{cases} 2\Big(\exp\left[-2a\left(r-\delta_{i}\right)\right] - \exp\left[-a\left(r-\delta_{i}\right)\right]\Big), & r < 2\delta_{i}, \\ 2\Big(\exp\left[-2a\left(r-\delta_{i}\right)\right] - \exp\left[-a\left(r-\delta_{i}\right)\right]\Big)g_{i}(r), & 2\delta_{i} \le r \le 3\delta_{i}, \\ 0, & r > 3\delta_{i}, \end{cases}$$

$$(3.7)$$

where a is the parameter that controls the shape of the force function, δ_i is the 227 cell size in the subpopulation i, i = 1, 2, 3, and r = |x - y|. We fix the value of 228 the shape parameter at $a = 0.08 \,\mu \text{m}^{-1}$ (Matsiaka et al., 2017). The function 229 $g_i(r) = \left(1 - \sin\left[(2\pi r - \pi\delta_i)/2\delta_i\right]\right)/2$ is the Tersoff cut-off function introduced 230 to impose a finite range of intercellular interactions (Tersoff, 1988). A sketch 231 of the potential function given by Equation (3.7) for different values of the 232 parameter a is shown in Figure 2(b) confirming that this potential function 233 describes short range repulsion, longer range attraction and no interactions at 234 over much longer distances. In summary, the key parameters in the hetero-235 geneous three-species model are: (i) the cell sizes, δ_1 , δ_2 and δ_3 ; (ii) the cell 236 diffusivities, D_1 , D_2 and D_3 ; and (iii) the amplitudes of interaction forces, $f_0^{(1)}$, 237 $f_0^{(2)}$ and $f_0^{(3)}$. In this work we will systematically explore how heterogeneity 238 in each of these three key parameters influences whether we need to consider 239 a complex heterogeneous multi-species model or whether we can describe the 240 spatial spreading of a cell population using relatively simple homogeneous, 241 single-species models. Since our experimental data in Figure 1 allows us to 242 explicitly characterise the heterogeneity in cell size, all results in the main 243 document focus on cell size. Additional results in the Supplementary Mate-244 rial focus on heterogeneity in diffusivity and amplitude of interaction forces 245 to provide additional insight into the role of heterogeneity in these kinds of 246 experiments. 247

²⁴⁸ We define the total density of the heterogeneous population as

$$\mathcal{P}(x,t) = \sum_{i=1}^{3} \left[p^{(i)}(x,t) \right], \tag{3.8}$$

where $p^{(i)}(x,t)$ is the cell density of subpopulation i = 1, 2, 3 predicted by

Equations (3.2)-(3.4), and $\mathcal{P}(x,t)$ is the total cell density. It is important to interpret the solutions of Equations (3.2)-(3.4) in terms of total cell density since standard experimental protocols do not normally facilitate the measurement of spatial and temporal distributions of various subpopulations (Cai et al., 2007; Treloar et al., 2014).

We can reduce the three-species heterogeneous system of equations, Equations (3.2)-(3.4), to obtain a single-species homogeneous model in the form,

$$\frac{\partial P(x,t)}{\partial t} = \bar{D}\Delta P(x,t) - (N-1)\nabla \Big(P(x,t)V(x,t)\Big),\tag{3.9}$$

where P(x,t) is the cell density of the total population, $N = \sum_{i=1}^{3} n_i$ is the 257 total number of cells in the population. Here we assume that the cell size, 258 diffusivity and strength of the interaction force for each population is constant, 259 giving $\delta_i = \bar{\delta}$, $D_i = \bar{D}$, and $f_i^{(i)} = \bar{f}_0$ for i = 1, 2, 3. The key differences 260 between the homogeneous single-species model, Equation (3.9), and the three-261 species heterogeneous model, Equations (3.2)-(3.4) are: (i) the three-species 262 heterogeneous model incorporates three advection-diffusion equations while 263 the single-species homogeneous model is given by a single advection-diffusion 264 equation; (ii) the three-species heterogeneous model contains up to nine free 265 parameters as opposed to three parameters in the single-species homogeneous 266 model. 267

The initial conditions in all simulations are chosen to mimic a cell front, such as that shown in our experimental data set, Figure 1(a). As such, we adopt ²⁷⁰ an initial cell distribution in the form of the one-dimensional step function,

$$P(x,0) = \mathcal{P}(x,0) = \begin{cases} 23.9 \times 10^{-3} \text{ cells}/\mu\text{m}, & 0 \,\mu\text{m} < x < 1000 \,\mu\text{m}, \\ 0 & \text{cells}/\mu\text{m}, & 1000 \,\mu\text{m} < x < 2000 \,\mu\text{m}, \end{cases}$$
(3.10)

on $0 < x < 2000 \,\mu\text{m}$, which is consistent with a length-scale of a typical in 271 vitro experiment (Jin et al., 2016a). The initial cell distribution in the hetero-272 geneous model is given by the sum of initial densities of three subpopulations, 273 $\mathcal{P}(x,0) = \sum_{i} p^{(i)}(x,0)$, where the density of each subpopulation, $p^{(i)}(x,0)$, 274 varies between each cell size distribution and can be inferred from the his-275 tograms in Figure 3(a). The value of the initial density of the total population 276 is chosen to represent fairly confluent population of cells. For example, the 277 simulation of the three-species population with the monotonically decreasing 278 cell size distribution, Set Ia, is initiated with the confluence level of approx-279 imately 65% of maximum packing density, which is fairly typical for scratch 280 assay experiments (Jin et al., 2016; Matsiaka et al., 2017). We note that the 281 boundary of the experimental image in Figure 1(a) is not a physical boundary 282 and cells can freely move across this boundary because the image captures 283 only a small fraction of a much larger experimental domain (Simpson et al., 284 2018). During the experiment, cells freely migrate, in each direction, across 285 the boundary. However, since the density of cells away from the scratch is 286 spatially uniform, the net flux of cells across the boundary of the image is 287 zero. To capture this situation we impose zero net flux boundary conditions 288 at $x = 0 \,\mu m$ and $x = 2000 \,\mu m$. 289

All continuum results for single-species homogeneous and three-species heterogeneous models, given by Equation (3.9) and Equations (3.2)-(3.4), respectively, are solved numerically using the method of lines with $\Delta x = 4 \,\mu \text{m}$ and $\Delta t = 0.005 \,\text{h}$ on $0 < x < 2000 \,\mu \text{m}$ (Matsiaka et al., 2017). We find that this choice of spatial and temporal discretisations are sufficiently fine to produce grid independent results. The detailed discretisation scheme used in this work is presented in the Supplementary Material.

²⁹⁷ 4 Results and Discussion

To investigate the ability of a single-species homogeneous model to capture 298 the behaviour of the three-species heterogeneous analogue, we consider a series 299 of case studies. In these case studies we vary only one parameter at a time to 300 simplify our analysis and to focus on the impact of each individual parameter. 301 Another approach would be to use the mathematical models to explore hetero-302 geneity multiple parameter at the same time. However, in this first instance, 303 we prefer to take a more fundamental approach and examine the role of het-304 erogeneity in each parameter separately. In the first set of experiments, Set I, 305 we vary the cell size, $\bar{\delta}$, while keeping \bar{D} and \bar{f}_0 fixed at $\bar{D} = 250 \,\mu \text{m}^2/\text{h}$ and 306 $\bar{f}_0 = 1.0 \,\mu\text{m/h}$. The values of D_i and $f_0^{(i)}$ in the heterogeneous three-species 307 model are fixed at $D_i = 250 \,\mu\text{m}^2/\text{h}$ and $f_0^{(i)} = 1.0 \,\mu\text{m/h}$ for i = 1, 2, 3. These 308 values of diffusivity and amplitude of cell-to-cell interaction forces are based 309 on detailed experimental measurements reported previously (Matsiaka et al., 310 2019). 311

There are number of ways to quantify performance of the single-species homogeneous model in our framework. The position of the leading edge of the spreading population is routinely used by experimentalists to provide quantitative insights into the rate of spatial spreading of a cell population (Treloar and Simpson, 2013; Johnston et al., 2014; Kollimada et al., 2016; Nardini et al., 2016; Bobadilla et al., 2019). Therefore, we quantify the discrepancy between the solution of the heterogeneous three-species model and the homogeneous single-species model using an error measure, $E(\bar{\delta})$, associated with the position of the leading edge,

$$E(\bar{\delta}) = \frac{1}{\alpha} \sum_{j} \left[\mathcal{S}(t_j) - S(t_j) \right]^2, \tag{4.1}$$

where $\mathcal{S}(t_j)$ is the position of the leading edge according to the three-species 321 heterogeneous model at time t_j , $S(t_j)$ is the position of the leading edge pre-322 dicted by the single-species homogeneous model, and $\alpha = 49$ is the number of 323 discrete time points we use to compute $E(\bar{\delta})$. In both scenarios the position of 324 the leading edge is computed as the coordinate on the one-dimensional domain 325 where the density is 1% of the initial density (Treloar and Simpson, 2013). An 326 alternative approach is to use an error measure based on the discrepancy be-327 tween cell density profiles. At first, this approach of using the entire cell density 328 profile might be thought to be preferable to working with leading edge data 329 since density profiles incorporate much more detailed spatial information than 330 just using the position of the leading edge. However, extracting the density 331 data from experiments is much more tedious because it often involves manual 332 cell counting in regions where cell densities are high and this is both difficult 333 to reproduce and very time consuming (Treloar et al. 2014). Therefore, to keep 334 our work as practical as possible, here we present only results with an error 335 measure solely based on the leading edge data. Additional result that measure 336 the discrepancy between the models using the entire density information are 337 presented in the Supplementary Material (Figure A.1 and Figure A.2), and 338 we find that this more complicated approach gives very similar results to the 339

leading edge data. Therefore, in this work, we focus on the using leasing edgedata.

The experimental distribution of cell sizes in Figure 1(d) provides insights into 342 potential choices of the cell size distribution in Equations (3.2)-(3.4). Here we 343 define three subpopulations based on the equivalent cell size: small ($\delta_1 = 18$ 344 μ m), medium ($\delta_2 = 34 \ \mu$ m), and large cells ($\delta_3 = 50 \ \mu$ m). For simplicity, we 345 set the fractions of small and medium cells to be equal and refer to this distri-346 bution as a monotonically decreasing distribution of cell sizes (Set Ia, Figure 347 3). After considering the experimentally-motivated monotonically decreasing 348 distribution, we then systematically explore: (i) uniform (Set Ib, Figure 4), 349 (ii) non-monotonic (Set Ic, Figure 5), and (iii) monotonically increasing dis-350 tributions (Set Id, Figure 6). 351



20

Fig. 3. Set Ia. Heterogeneity in cell sizes: monotonically decreasing distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (3.2)-(3.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.472$; (iii) $n_3/N = 0.056$. (b) Leading edge as predicted by the three-species heterogeneous model, S(t) (solid red), and the best-fit approximation given by the single-species homogeneous model, S(t) (blue dashed). (c) Error measure, $E(\bar{\delta})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 28 \,\mu$ m. (d)-(e) Cell density profiles predicted by the three-species heterogeneous model calibrated with the best-fit value of δ , P(x,t) (solid red), superimposed with density profiles given by the single-species homogeneous model calibrated with the best-fit value of δ , P(x,t) (solid blue). The continuum results for both models are presented at t = 0, 12, 24, 36, and 48 h. Black arrows denote the direction of increasing time. Results in (e) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (d).

Figure 3(b) compares the leading edge prediction, $\mathcal{S}(t)$, given by the three-352 species heterogeneous model with the associated best-fit match, S(t), pre-353 dicted by the single-species homogeneous model. Our systematic computation 354 of the error measure, $E(\bar{\delta})$, demonstrates a clear minimum which ensures the 355 unique choice of a best-fit cell size, $\overline{\delta}$. Results in Figure 3(d) superimposes 356 the solution of the three-species heterogeneous model with the solution of the 357 single-species homogeneous model parameterised with the best fit cell size. 358 Comparing the time evolution of the spreading density profiles in Figure 3(d)359 (with additional details at the leading edge shown in the magnified region in 360 Figure 3(e)) we see that the appropriately parameterised single-species homo-361 geneous model captures the temporal evolution of the spreading profile given 362 by the heterogeneous model remarkably accurately. In particular, the den-363 sity profiles predicted by the single-species homogeneous model match both 364 the position and shape of the density profiles generated by the three-species 365 heterogeneous model. These results imply that in this case it would be rea-366 sonable to use a much simpler single-species homogeneous model to describe 367 and predict this spatial spreading. 368

Visual inspection of the results in Figures 3 - 6 suggests that we can always 369 find a unique, well-defined value of the cell size in the single-species homo-370 geneous model to provide an accurate prediction of the temporal evolution 371 of the position of a leading edge of the spreading heterogeneous cell popu-372 lations regardless of the underlying cell size distribution in the three-species 373 heterogeneous model (Figures 3(b)-6(b)). In contrast, the quality of match 374 between the shape of the density profiles for the three-species heterogeneous 375 model and the single-species homogeneous model varies significantly between 376 different cell size distributions. For example, the experimentally motivated dis-377

tribution in Figure 3(a) (Set Ia) leads to a remarkably good match between the 378 three-species heterogeneous model and the single-species homogeneous model. 379 Similarly, the uniform distribution shown in Figure 4(a) (Set Ib) also leads to 380 a reasonably good quality of match between two different models. In contrast, 381 the density profiles associated with the non-monotonic cell size distribution 382 (Figure 5, Set Ic) and monotonically increasing cell size distribution (Figure 383 6, Set Id) show a relatively poor match. In these cases, it would seem prudent 384 not to use a simpler single-species homogeneous model to simulate and predict 385 these experiments. 386

The values of the cell size, $\overline{\delta}$, that produce best match between the single-387 species homogeneous and three-species heterogeneous models vary significantly 388 between different cell size distributions. For example, the best-fit value of the 389 cell size for the uniform distribution (Figure 4, Set Ib), $\bar{\delta} = 36 \,\mu\text{m}$, is quite 390 close to the weighted average value of $34 \,\mu m$ for the distribution in Figure 391 4(a). This indicates that the choice of a simple weighted average of the cell 392 sizes might be a reasonable way to to parameterise the single-species homo-393 geneous model if the experimentally observed distribution is close to uniform. 394 We observe similar agreement for best-fit values of the cell size in the case of 395 monotonically decreasing (Set Ia) and monotonically increasing (Set Id) cell 396 size distributions, shown in Figure 3 and Figure 6, respectively. In contrast, 397 the best-fit value of the cell size for the non-monotonic distribution (Set Ic), 398 $\bar{\delta} = 40 \,\mu\text{m}$, differs significantly from the weighted average of $34 \,\mu\text{m}$. Therefore, 399 these results suggest that great care ought to be exercised when taking a dis-400 tribution of parameter values and attempting to select the most appropriate 401 single representative value of that parameter. 402



Fig. 4. Set Ib. Heterogeneity in cell sizes: uniform distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (3.2)-(3.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.33(3)$; (ii) $n_2/N = 0.33(3)$; (iii) $n_3/N = 0.33(3)$. (b) Leading edge as predicted by the three-species heterogeneous model, S(t) (solid red), and the best-fit approximation given by the single-species homogeneous model, S(t) (blue dashed). (c) Error measure, $E(\bar{\delta})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 36 \,\mu$ m. (d)-(e) Cell density profiles predicted by the three-species heterogeneous model calibrated with the best-fit value of $\bar{\delta}$, P(x, t) (solid red), superimposed with density profiles given by the single-species homogeneous model calibrated with the best-fit value of $\bar{\delta}$, P(x, t) (solid blue). The continuum results for both models are presented at t = 0, 12, 24, 36, and 48 h. Black arrows denote the direction of increasing time. Results in (e) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (d).



Fig. 5. Set Ic. Heterogeneity in cell sizes: non-monotonic distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (3.2)-(3.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.472$; (ii) $n_2/N = 0.056$; (iii) $n_3/N = 0.472$. (b) Leading edge as predicted by the three-species heterogeneous model, S(t) (solid red), and the best-fit approximation given by the single-species homogeneous model, S(t) (blue dashed). (c) Error measure, $E(\bar{\delta})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 40 \,\mu$ m. (d)-(e) Cell density profiles predicted by the three-species heterogeneous model calibrated with the best-fit value of $\bar{\delta}$, P(x,t) (solid red), superimposed with density profiles given by the single-species homogeneous model calibrated with the best-fit value of $\bar{\delta}$, P(x,t) (solid blue). The continuum results for both models are presented at t = 0, 12, 24, 36, and 48 h. Black arrows denote the direction of increasing time. Results in (e) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (d).



Fig. 6. Set Id. Heterogeneity in cell sizes: monotonically increasing distribution. (a) Cell size distribution adopted in the three-species heterogeneous model, Equations (3.2)-(3.4). Here the proportions of cells of different sizes are set to: (i) $n_1/N = 0.056$; (ii) $n_2/N = 0.472$; (iii) $n_3/N = 0.472$. (b) Leading edge as predicted by the three-species heterogeneous model, S(t) (solid red), and the best-fit approximation given by the single-species homogeneous model, S(t) (blue dashed). (c) Error measure, $E(\bar{\delta})$, between the position of the leading edge given by the three-species heterogeneous model and the position predicted by the single-species homogeneous model as a function of cell size, $\bar{\delta}$. The black arrow denotes the best-fit value of cell size, $\bar{\delta} = 41 \,\mu$ m. (d)-(e) Cell density profiles predicted by the three-species heterogeneous model calibrated with the best-fit value of $\bar{\delta}$, P(x, t) (solid red), superimposed with density profiles given by the single-species homogeneous model calibrated with the best-fit value of $\bar{\delta}$, P(x, t) (solid blue). The continuum results for both models are presented at t = 0, 12, 24, 36, and 48 h. Black arrows denote the direction of increasing time. Results in (e) show a close-up comparison right near the leading edge, denoted by the gray shaded region in (d).

In addition to the results in Figures 3 - 6 exploring the role of heterogeneity 403 in cell size, we present an additional suite of results where we systematically 404 explore the role of heterogeneity in diffusivity (Set II) and amplitude of interac-405 tion forces (Set III) while keeping the cell size constant in all subpopulations. 406 These additional results are presented in the Supplementary Material doc-407 ument. Both Set II and Set III data sets demonstrate exceptional quality of 408 match between the three-species heterogeneous simulation data and its best-fit 409 single-species homogeneous equivalent. Again, these additional results provide 410 guidance about when it is reasonable to approximate a more complicated het-411 erogeneous mathematical model with a simpler single-species homogeneous 412 model. 413

414 5 Conclusions

In this work, we explore the role of heterogeneity in the context of study-415 ing how an initially confined population of cells can spread into surround-416 ing initially unoccupied regions, as in the case of a scratch assay. We use 417 a three-species heterogeneous model of cell motility, account for undirected 418 cell motility, short range repulsion (crowding) and longer range adhesion, to 419 capture experimentally observed heterogeneity in cell sizes from a new exper-420 imental data set from a two-dimensional scratch assay as shown in Figure 1. 421 Our continuum models account for the undirected random motility, cell-to-cell 422 adhesion, and cell crowding. The single-species homogeneous model is applied 423 to each set of three-species heterogeneous simulation data in an attempt to 424 match cell density profiles. 425

⁴²⁶ To analyse the performance of the single-species homogeneous model to cap-

ture data from our three-species heterogeneous model we consider four dif-427 ferent cell size distributions: (i) monotonically decreasing distribution, (ii) 428 uniform distribution, (iii) non-monotonic distribution, and (iv) monotonically 429 increasing distribution. Overall, for a set of experimentally-motivated parame-430 ter combinations, we find that the standard single-species homogeneous model 431 is able to accurately predict the position of the leading edge for all case stud-432 ies presented. However, the quality of the match between the shape of the 433 density profiles varies significantly depending on the details of the form of 434 the heterogeneity present. For example, the monotonically decreasing distri-435 bution (Set Ia) demonstrates remarkable goodness of fit between the two sets 436 of density profiles, as shown in Figure 3(d). This result is important because 437 the monotonically decreasing cell size distribution is chosen to mimic the dis-438 tribution of the cell sizes observed in our new experimental data set, shown 439 in Figure 1. Similarly, the homogeneous distribution, Figure 4, shows that 440 single-species homogeneous model is able to accurately replicate the three-441 species heterogeneous model results. This is an expected result because in this 442 special case the cells of each subpopulation are the same size. In contrast, 443 the single-species homogeneous model does not perform so well when applied 444 to both non-monotonic and monotonically increasing distributions in Figures 445 5-6, respectively. Additionally we explore potential heterogeneity in diffusiv-446 ity and amplitude of the cell-to-cell interactions (Supplementary Material). 447 Overall, our results suggest that for certain cell size distributions, a simple 448 and computationally efficient single-species homogeneous model is preferable 449 over a thee-species heterogeneous model. 450

⁴⁵¹ There are number of ways this work can be extended which we leave for future
⁴⁵² analysis. All our simulations and analysis focus on treating the heterogeneity

in the population of cells by considering the total population to be composed of 453 three distinct subpopulations. For more extreme forms for heterogeneity, such 454 as multi-modal distributions, the results presented in this work could be ex-455 tended by considering additional subpopulations. Another simplification that 456 we invoke is to assume that the measured heterogeneity remains constant for 457 the duration of the experiment. Future studies could address the significantly 458 more complicated question of allowing the distributions to evolve in time on 459 the same time scale as the experiment to see if it is still possible to use a sim-460 pler homogeneous model in this more complicated scenario. Another avenue 461 for further exploration would be to consider heterogeneity in more than one 462 parameter at a time, whereas in this work we have taken the most fundamental 463 approach and examined heterogeneity in just one parameter in isolation from 464 the others. For both of these extensions, the modelling framework presented 465 in this study can be extended to explore these additional features, and we 466 leave such extensions for future consideration. Another option for extending 467 the work would be to consider further details in the mathematical models, 468 such as the effects of combined cell migration and combined cell proliferation. 469 Here we have not pursued this approach because our experimental data set 470 has been carefully prepared to exclude the effects of proliferation so that we 471 can focus just on cell migration and heterogeneity in cell migration alone. 472

473 Acknowledgements

This work is supported by the Australian Research Council (DP170100474).
Computational resources are provided by the High Performance Computing
and Research Support Group at QUT. REB is a Royal Society Wolfson Re-

bioRxiv preprint doi: https://doi.org/10.1101/705434; this version posted August 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

search 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 two anonymous referees.

481 References

- [1] Altschuler, S.J., Wu, L.F., 2010. Cellular heterogeneity: when do differences make
 a difference? Cell 141(4), 559-563.
- 484 [2] An, F.-Q., Matsuda, M., Fujii, H., Tang, R.-F., Amemiya, H., Dai, Y.-M.,
- 485 Matsumoto Y., 2001. Tumor heterogeneity in small hepatocellular carcinoma:
- 486 Analysis of tumor cell proliferation, expression and mutation of p53 and β -
- 487 Catenin. International Journal of Cancer 93, 468-474.
- ⁴⁸⁸ [3] Baker, R.E., Simpson, M.J., 2010. Correcting mean-field approximations for
 ⁴⁸⁹ birth-death-movement processes. Physical Review E 82, 041905.
- [4] Binder, B.J., Simpson, M.J., 2016. Cell density and cell size dynamics during *in vitro* tissue growth experiments: Implications for mathematical models of
 collective cell behaviour. Applied Mathematical Modelling 40, 3438-3446.
- ⁴⁹³ [5] Bobadilla, A.V.P., Arevalo. J., Sarro, E., Byrne, H.M., Maini, P.K., Carraro, T.,
- Balocco, S., Meseguer, A., Alarcon, T., 2019. In vitro cell migration quantification
 method for scratch assays. Journal of the Royal Society Interface 16, 20180709.
- [6] Cai, A.Q., Landman, K.A., Hughes, B.D., 2007. Multi-scale modeling of a wound-
- ⁴⁹⁷ healing cell migration assay. Journal of Theoretical Biology 245(3), 576-594.
- ⁴⁹⁸ [7] Callaghan, T., Khain, E., Sander, L.M., Ziff, R.M., 2006. A stochastic model for
 ⁴⁹⁹ wound healing. Journal of Statistical Physics 122(5), 909-924.
- [8] Edmondson, R., Broglie, J.J., Adcock, A.F., Yang, L., 2014. Three-dimensional
 cell culture systems and their applications in drug discovery and cell-based
 biosensors. ASSAY and Drug Development Technologies 12(4), 207-218.
- ⁵⁰³ [9] Essen BioScience: IncuCyte ZOOM live cell imaging.
- Available from: http://www.essenbioscience.com/essen-products/incucyte/ (Accessed: August 2019).

- ⁵⁰⁶ [10] Frascoli, F., Hughes, B.D., Zaman, M.H., Landman, K.A., 2013. A
 ⁵⁰⁷ computational model for collective cellular motion in three dimensions: General
 ⁵⁰⁸ framework and case study for cell pair dynamics. PLoS ONE 8(3), e59249.
- [11] Galle, J., Loeffler, M., Drasdo D., 2005. Modeling thee effect of deregulated
 proliferation and apoptosis on the growth dynamics of epithelial cell populations
 in vitro. Biophysical Journal 88, 62-75.
- [12] Gerlee, P., 2013. The model muddle: In search of tumor growth laws. Cancer
 Research 73(8), 2407-2411.
- [13] George, M., Bullo F., Campas O., 2017. Connecting individual to collective cell
 migration. Scientific Reports 7, 9720.
- 516 [14] Hasenauer, J., Waldherr, S., Doszczak, M., Scheurich, P., Radde, N., Allgwer,
- F. 2011. Analysis of heterogeneous cell populations: A density-based modeling
 and identification framework. Journal of Process Control 21, 1417-1425.
- ⁵¹⁹ [15] Jeon J, Quaranta V, Cummings PT (2010). An off-lattice hybrid discrete-⁵²⁰ continuum model of tumor growth and invasion. Biophysical Journal 98 (1): 37-47.
- [16] Jin, W., Shah, E.T., Penington, C.J., McCue, S.W., Chopin, L.K., Simpson,
 M.J., 2016. Reproducibility of scratch assays is affected by the initial degree of
 confluence: Experiments, modelling and model selection. Journal of Theoretical
 Biology 390, 136-145.
- ⁵²⁵ [17] Jin, W., McCue, S.W., Simpson, M.J., 2018. Extended logistic growth model
 ⁵²⁶ for heterogeneous populations. Journal of Theoretical Biology 445, 51-61.
- [18] Johnston, S.T., Simpson, M.J., McElwain, D.L.S., 2014. How much information
 can be obtained from tracking the position of the leading edge in a scratch assay?
 Journal of the Royal Society Interface 11, 20140325.
- ⁵³⁰ [19] Kaighn, M.E., Narayan, K.S., Ohnuki, Y., Lechner, J.F., Jones, L.W., 1979.

531 Establishment and characterization of a human prostatic carcinoma cell line (PC-

- $_{532}$ 3). Investigative urology 17(1), 16-23.
- ⁵³³ [20] Kollimada, S.A., Kulkarni, A.H., Ravan, A., Gundiah, N., 2016. Advancing edge
- ⁵³⁴ speeds of epithelial monolayers depend on their initial confining geometry. PLoS
- 535 ONE 11(4), e0153471.
- ⁵³⁶ [21] Maini, P.K., McElwain, D.L.S., Leavesley, D., 2004. Travelling waves in a wound
- ⁵³⁷ healing assay. Applied Mathematics Letters 17(5), 575-580.
- [22] Maini, P.K., McElwain, D.L.S., Leavesley, D., 2004. Traveling wave model to
 interpret a wound-healing cell migration assay for human peritoneal mesothelial
 cells. Tissue Engineering 10(3/4), 475-482.
- ⁵⁴¹ [23] Matsiaka, O.M., Penington, C.J., Baker, R.E., Simpson, M.J., 2017. Continuum
- ⁵⁴² approximations for lattice-free multi-species models of collective cell migration.
 ⁵⁴³ Journal of Theoretical Biology 422, 1-11.
- [24] Matsiaka, O.M., Penington, C.J., Baker, R.E., Simpson, M.J., 2018. Discrete
 and continuum approximations for collective cell migration in a scratch assay with
 cell size dynamics. Bulletin of Mathematical Biology 80(4), 738-757.
- ⁵⁴⁷ [25] Matsiaka, O.M., Baker, R.E., Simpson, M.J., 2019. Mechanistic and
 ⁵⁴⁸ experimental models of cell migration reveal the importance of intercellular
 ⁵⁴⁹ interactions in cell invasion. Biomedical Physics and Engineering Express 5,
 ⁵⁵⁰ 045009.
- ⁵⁵¹ [26] Menon, T., Kumar, R., Nair, S., 2018. Cell size heterogeneity early in
 ⁵⁵² development is required for collective cell migration during gastrulation in
 ⁵⁵³ zebrafish. bioRxiv doi:10.1101/481325.
- ⁵⁵⁴ [27] Middleton, A.M., Fleck, C., Grima, R., 2014. A continuum approximation to
 ⁵⁵⁵ an off-lattice individual-cell based model of cell migration and adhesion. Journal
 ⁵⁵⁶ of Theoretical Biology 359, 220-232.

- ⁵⁵⁷ [28] Murray, P.J., Edwards, C.M., Tindall, M.J., Maini, P.K., 2009. From a discrete
- to a continuum model of cell dynamics in one dimension. Physical Review E 80(3),
 031912.
- ⁵⁶⁰ [29] Nardini, J.T., Chapnick, D.A., Liu, X., Bortz, D.M., 2016 Modeling keratinocyte
- wound healing dynamics: Cell-cell adhesion promotes sustained collective
 migration. Journal of Theoretical Biology 400, 103-117.
- [30] Newman, T.J., Grima, R., 2004. Many-body theory of chemotactic cell-cell
 interactions. Physical Review E 70, 051916.
- ⁵⁶⁵ [31] O'Dea, R.D., King, J.R., 2012. Continuum limits of pattern formation in
 ⁵⁶⁶ hexagonal-cell monolayers. Journal of Mathematical Biology 64, 579-610.
- ⁵⁶⁷ [32] Osborne, J.M., Fletcher, A.G., Pitt-Francis, J.M., Maini, P.K., Gavaghan, D.J.,
- ⁵⁶⁸ 2017. Comparing individual-based approaches to modelling the self-organization
- of multicellular tissues. PLOS Computational Biology 13(2), e1005387.
- ⁵⁷⁰ [33] Painter, K.J., Sherratt, J.A., 2003. Modelling the movement of interacting cell
 ⁵⁷¹ populations. Journal of Theoretical Biology 225(3), 327-339.
- ⁵⁷² [34] Painter, K.J., Armstrong, N.J., Sherratt, J.A., 2010. The impact of adhesion
 ⁵⁷³ on cellular invasion processes in cancer and development. Journal of Theoretical
 ⁵⁷⁴ Biology 264, 1057-1067.
- [35] Pozzobon, V., Perre, P., 2018. Han's model parameters for microalgae grown
 under intermittent illumination: Determined using particle swarm optimization.
 Journal of Theoretical Biology 437, 29-35.
- [36] Rutter, E.M., Banks, H.T., Flores, K.B., 2018. Estimating intratumoral
 heterogeneity from spatiotemporal data. Journal of Mathematical Biology 77:
 1999-2022.
- 581 [37] Sadeghi, H.M., Seitz, B., Hayashi, S., LaBree, L., McDonnell, P.J., 1998. In

vitro effects of Mitomycin-C on human keratocytes. Journal of Refractive Surgery
14, 534-40.

- [38] Sarapata, E.A., de Pillis, L.G., 2010. A comparison and catalog of intrinsic
 tumor growth models. Bulletin of Mathematical Biology 76, 2010-2024.
- [39] Savla, U., Olson, L.E., Waters, C.M., 2004. Mathematical modeling of airway
 epithelial wound closure during cyclic mechanical strain. Journal of Applied
 Physiology 96, 566-574.
- [40] Sengers, B.A., Please, C.P., Oreffo, R.O.C., 2007. Experimental characterization
 and computational modelling of two-dimensional cell spreading for skeletal
 regeneration. Journal of the Royal Society Interface 4, 1107-1117.
- ⁵⁹² [41] Sepulveda, N., Petitjean, L., Cochet, O., Grasland-Mongrain, E., Silberzan, P.,
- Hakim, V., 2016. Collective cell motion in an epithelial sheet can be quantitatively
 described by a stochastic interacting particle model. PLOS Computational Biology
 9(3), e1002944
- ⁵⁹⁶ [42] Shah, E.T., Upadhyaya, A., Philp, L.K., Tang, T., Skalamera, D., Gunter,
 ⁵⁹⁷ J., Nelson, C.C., Williams, E.D., Hollier, B.G., 2016. Repositioning "old" drugs
 ⁵⁹⁸ for new causes: identifying new inhibitors of prostate cancer cell migration and
 ⁵⁹⁹ invasion. Clinical & Experimental Metastasis 33, 385-399.
- [43] Sheardown, H., Cheng, Y.-L., 1995. Mechanisms of corneal epithelial wound
 healing. Chemical Engineering Science 51(19), 4517-4529.
- [44] Sherratt, J.A., Murray, J.D., 1990. Models of epidermal wound healing.
 Proceedings of the Royal Society B 241, 29-36.
- [45] Sundstrom, A., Bar-Sagi, D., Mishra, B., 2016. Simulating heterogeneous tumor
 cell populations. PLoS ONE 11(12), e0168984.
- [46] Simpson, M.J., Treloar, K.K., Binder, B.J., Haridas, P., Manton, K.J.,
 Leavesley, D.I., McElwain, D.L.S., Baker, R.E., 2013. Quantifying the roles of

bioRxiv preprint doi: https://doi.org/10.1101/705434; this version posted August 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

- cell motility and cell proliferation in a circular barrier assay. Journal of the Royal
 Society Interface 10, 20130007.
- [47] Simpson, M.J., Haridas P., McElwain D.L.S., 2014. Do pioneer cells exist? PLoS
 ONE 9(1), e85488.
- 612 [48] Simpson M.J., Jin, W., Vittadello, S.T., Tambyah, T.A., Ryan, J.M.,
- Gunasingh, G., Haass, N.K., McCue, S.W., 2018. Stochastic models of cell invasion
- with fluorescent cell cycle indicators. Physica A 510, 375-386.
- ⁶¹⁵ [49] Tersoff, J., 1988. New empirical approach for the structure and energy of ⁶¹⁶ covalent systems. Physical Review B 37(12), 6991-7000.
- 617 [50] ThermoFisher Scientific. Available from: 618 https://www.thermofisher.com/au/en/home/life-science/cell-
- culture/mammalian-cell-culture/reagents/trypsin/tryple-express.html (Accessed:
- ⁶²⁰ August, 2019).
- [51] Treloar, K.K., Simpson, M.J. 2013. Sensitivity of edge detection methods for
 quantifying cell migration assays. PLoS ONE 8(6), e67389.
- ⁶²³ [52] Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B.D., Stevens, G.W., Landman,
- K.A., O'Connor, A.J., 2009. Cell migration and proliferation during monolayer
 formation and wound healing. Chemical Engineering Science 64, 247-253.
- [53] Vo, B.N., Drovandi, C.C., Pettitt, Simpson, M.J., 2015. Quantifying uncertainty
 in parameter estimates for stochastic models of collective cell spreading using
 approximate Bayesian computation. Mathematical Biosciences 263, 133-142.
- [54] Warne, D.J., Baker, R.E., Simpson, M.J., 2019. Simulation and inference
 algorithms for stochastic biochemical reaction networks: from basic concepts to
 state-of-the-art. Journal of the Royal Society Interface 16, 20180943.

bioRxiv preprint doi: https://doi.org/10.1101/705434; this version posted August 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

- 632 [55] Wise, S.M., Lowengrub, J.S., Frieboes, H.B., Cristini, V., 2008. Three-
- dimensional multispecies nonlinear tumor growth–I Model and numerical method.
- Journal of Theoretical Biology 253(3), 524-543.