Quantifying the Morphology and Mechanisms of Cancer Progression in 3D *in-vitro* environments: Integrating Experiments and Multiscale Models

Nikolaos M. Dimitriou[1*], **Salvador Flores-Torres**[1], **Joseph Matthew Kinsella**[1], **Georgios D. Mitsis**[1*]

[1]Department of Bioengineering, McGill University, QC, Canada

***For correspondence:** nikolaos.dimitriou@mail.mcgill.ca (NMD); georgios.mitsis@mcgill.ca (GDM)

Abstract Mathematical models of cancer growth have become increasingly more accurate both in the space and time domains. However, the limited amount of data typically available has resulted in a larger number of qualitative rather than quantitative studies. In the present study, we provide an integrated experimental-computational framework for the quantification of the morphological characteristics and the mechanistic modelling of cancer progression in 3D environments. The proposed framework allows for the calibration of multiscale, spatiotemporal models of cancer growth using state-of-the-art 3D cell culture data, and their validation based on the resulting experimental morphological patterns. Its implementation enabled us to pursue two goals; first, the quantitative description of the morphology of cancer progression in 3D cultures, and second, the relation of tumour morphology with underlying biophysical mechanisms that govern cancer growth and migration. We applied this framework to the study of the spatiotemporal progression of Triple Negative Breast Cancer cells cultured in Matrigel scaffolds, and validated the hypothesis of chemotactic migration using a multiscale Keller-Segel model. The results revealed transient, non-random spatial distributions of cancer cells that consist of clustered, and dispersion patterns. The proposed model was able to describe the general characteristics of the experimental observations and suggests that cancer cells exhibited chemotactic migration and accumulation, as well as random motion during the examined time period of development. To our knowledge, this is the first time that a multiscale model is used to quantify the relationship between the spatial patterns and the underlying mechanisms of cancer growth in 3D environments.

Introduction

Cancer progression is frequently accompanied by migration of cancer cells into the surrounding tissues, that eventually leads to metastasis. Both *in-vivo* and *in-vitro* studies of cancer cell migration have shown that cancers can exhibit several types of patterns including single cell migration, multi-cellular streaming, collective cell migration, as well as passive patterns, such as tissue folding, and expansive growth ***Friedl et al.*** (***2012***). Studies of breast cancer have shown that the tumour border is dominated by collective cell migration ***Friedl et al.*** (***2012***) forming small acinar structures with a central lumenal space, with cancer cells that maintain their epithelial morphology ***Tan et al.*** (***1999***). Evidence of multicellular streaming also exist from orthotopic breast cancer in xenograft mouse models ***Roussos et al.*** (***2011a***). Other clinical studies of the surface morphology of infiltrating duc-

tal adenocarcinoma have shown that the fractal dimension of cancerous tissue is larger compared to normal breast tissue ***Norton*** (***2005***). Similar observations of different fractality characteristics in different tumour stages have also been made in ovarian cancer ***Kikuchi et al.*** (***2002***). Though there is significant knowledge on the qualitative aspects of tumour morphology, the quantitative characterization of this morphology and the biophysical mechanisms that govern cancer growth and migration remain still elusive.

Significant insights into both morphological and mechanistic characteristics of cancer growth can be gained from the use of mathematical models. Spatiotemporal models of cancer growth can be distinguished in three general categories; discrete (e.g. agent based models), continuum (Partial Differential Equations, PDEs), and hybrid models ***Cristini and Lowengrub*** (***2010***). Each of these categories provides different information on the aspects of tumour growth. Specifically, discrete models can provide information on individual cell processes or tissue microarchitecture ***Anderson et al.*** (***2008***). Continuum models have been widely used, initially to describe qualitative aspects of tumour growth, albeit lacking experimental validation ***Greenspan*** (***1972***), and more recently, for a more detailed quantitative description of the macroscopic characteristics of spatiotemporal cancer growth and its response to therapy under both *in-vitro* ***Loessner et al.*** (***2013***); ***Stein et al.*** (***2007***); ***Warne et al.*** (***2019***); ***Jin et al.*** (***2016***) and *in-vivo* conditions ***Tunc et al.*** (***2021***); ***Hoehme et al.*** (***2018***); ***Lipková et al.*** (***2019***); ***Hormuth et al.*** (***2017***, 2018); ***Abler et al.*** (***2019***); ***Jarrett et al.*** (***2020***). Hybrid models attempt a multiscale description of cancer growth, by incorporating both continuous and discrete variables ***Rejniak and Anderson*** (***2011***); ***Sanga et al.*** (***2007***). Studies from Tweedy et al. ***Tweedy et al.*** (***2020***); ***Tweedy and Insall*** (***2020***) utilized experiments and hybrid discrete-continuum (HDC) models of chemotactic migration to investigate the role of self-generated chemotactic gradients in cancer migration. Even though there is a growing literature on spatiotemporal models of cancer, their validation using experimental data is important for quantitatively describing cancer ***Collis et al.*** (***2017***); ***Lima et al.*** (***2016***).

The validation of a model can be interpreted as the process of quantifying how the predictions accurately describe the experimental measurements ***Collis et al.*** (***2017***). Validation usually follows calibration of the model, which is usually defined as the process of inferring the model parameters that provide the best fit between model predictions and experimental data. Typically, model calibration is performed using a training dataset, and validation is used to assess the prediction of the calibrated model on a different dataset. In the work of Hawkins-Daarud et al. ***Hawkins-Daarud et al.*** (***2013***) a Bayesian framework was used for the validation of diffuse-interface models of tumour growth using synthetic data. Achilleos et al. ***Achilleos et al.*** (***2013***, 2014) utilized stochastic processes for the validation of a mixture-model using tumour measurements obtained from experiments in mice. Lima et al. ***Lima et al.*** (***2016***) employed the Occam Plausibility Algorithm (OPAL) ***Farrell et al.*** (***2015***) to validate various classes of PDE models using images of glioma tumours in mice. Recently, they ***Lima et al.*** (***2021***) performed calibration of hybrid models with *in-vitro* 2D culture data. Although these studies have yielded significant advances in model validation, several studies exclude model validation from their analyses.

A common reason for the absence of validation in tumour modelling studies is the lack of data availability. *In-vitro* studies usually include the use of 2D cultures ***Jin et al.*** (***2016***); ***Warne et al.*** (***2019***), resulting in a less realistic representation of cancer growth. *In-vivo* studies, both clinical and experimental, are more realistic; however, they also present limitations. On the one hand, caliper and microCT scan measurements of *in-vivo* tumours in mice do not typically provide information on tumour shape and invasiveness ***Loizides et al.*** (***2015***); ***Jensen et al.*** (***2008***). Intravital imaging is another common way of data collection for *in-vivo* models; however, this technique suffers from technical challenges, such as passive drift of cells or tissues, low penetration depth, tissue heating, and limitations on imaging intervals ***Friedl et al.*** (***2012***). On the other hand, clinical data can be limited in terms of time-points ***Lipková et al.*** (***2019***), resulting in model over-fit.

To this end, 3D cell culture models have become a promising experimental tool. The main reasons are their increased control of the experimental conditions and flexibility of data collection

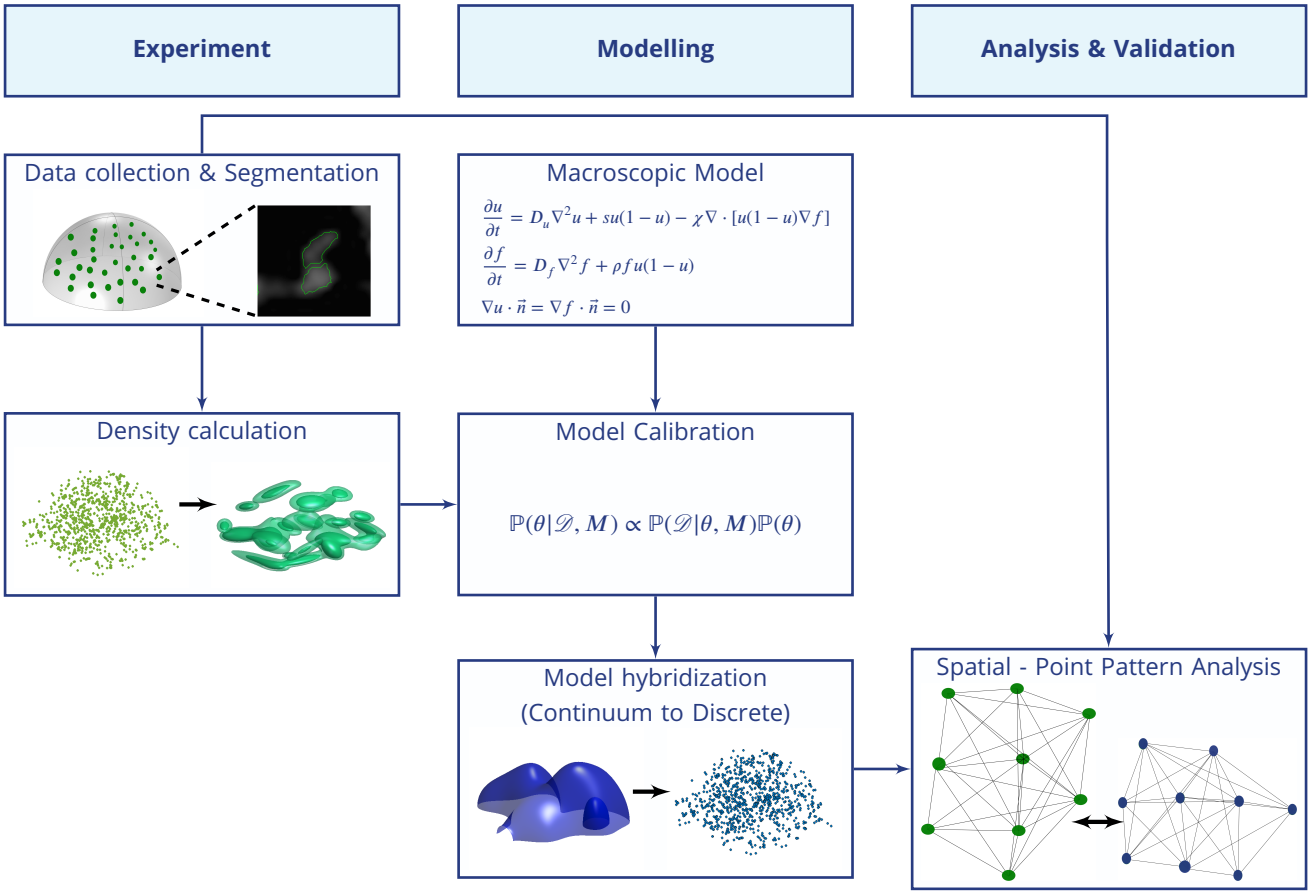
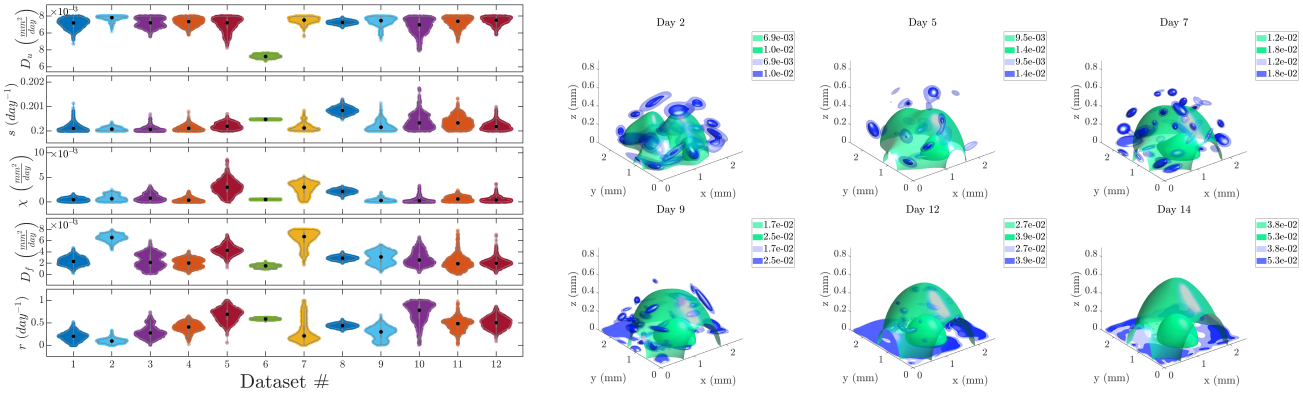


Figure 1. Proposed pipeline for the modelling, validation and analysis of cancer progression using *in-vitro* 3D experimental data.

compared to *in-vivo* experiments, as well as their more realistic representation of tumour progression compared to 2D cultures. Differences between 3D and 2D cultures have been observed in cancer growth and its related biochemical processes, such as the secretion of extracellular matrix (ECM) components, and cell-cell interaction components ***Hickman et al.*** (***2014***), while the histological and molecular features of *in-vitro* 3D spheroids exhibit more similarities to xenografts compared to 2D monolayers ***Hickman et al.*** (***2014***). Significant differences between 2D and 3D cultures have also been found in drug testing studies exhibiting alterations in the sensitivity of cytotoxic drugs ***Hickman et al.*** (***2014***). Another advantage of 3D cell culture models is their flexibility with regards to incorporating more than one cell populations, such as stromal cells, as well as on changing the stiffness of the ECM. The heterotypic intercellular interactions between cancer cells and stromal cells results in altered cancer cell proliferation and migration, as well as the formation of more compact spheroids compared to equivalent 3D cell mono-culture systems ***Hickman et al.*** (***2014***). The mechanical properties of the ECM also contribute to the spheroid formation, viability, invasiveness, and drug sensitivity of cancer cells ***Hickman et al.*** (***2014***). Additionally, the collection of imaging data for *in-vitro* 3D cell cultures is generally easier and more accurate than *in-vivo* models, and high resolution images can be obtained using confocal microscopy. Although 3D cell cultures cannot yet capture the full complexity of tumour growth in a living tissue, overall they yield significant potential for quantitatively describing cancer growth, as they even provide the opportunity to track even single cells.

The purpose of the present work is to introduce an integrated framework for the quantitative characterization of spatiotemporal progression of cancer, and its use for multiscale-spatiotemporal model validation for the study of cancer growth mechanisms. The framework presented in Fig. 1



(a) (b)

Figure 2. Inferred model parameters and simulation results (a) Violin plots of the marginalized posterior PDFs of the model parameters across the 12 datasets. The black dots represent the median values. (b) Isosurface plot of the experimental and simulated density profiles using the inferred parameters of the initial conditions of a representative dataset. The blue colour-map corresponds to the *in-vitro* cell density profiles and the green colour-map corresponds to the *in-silico* cell density profiles.

proposes a novel combination of experimental data from state-of-the-art 3D cell cultures, spatial statistical analysis techniques for the quantification of cancer morphology, and a multiscale HDC mathematical model for the quantitative description of the mechanisms underlying cancer progression. Given the spatial scales (μm up to mm) of the 3D cultures, the choice of HDC models instead of purely continuum or discrete models allows us to perform faster calibration on the continuum model component, albeit with a lower fidelity compared to the full model, and validation on the discrete component. In this work, we present a novel approach for model calibration and validation. Instead of splitting the datasets, we perform calibration and validation on the two different levels; calibration on the continuum, and validation on the discrete level. The introduction of the spatial pattern analysis not only enables us to validate the hybrid model, but also to interpret the observed patterns based on the underlying mechanisms. The rest of the article is organized in a Methods section, where we describe the experiments, data processing, the mathematical model, the calibration and validation techniques, followed by the Results where we present the calibrated model, the validity tests of the full model, as well as a description of the relation between morphology and the underlying mechanisms. Finally, we conclude with the Discussion and Conclusions, where we discuss our results compared to relevant literature, the advantages and limitations of our study, as well as possible extensions and improvements. The code and data of this work are available at https://nmdimitriou.github.io/HyMetaGrowth/.

Results

Estimation of the macroscopic model parameters

The continuum Keller-Segel model was used to generate simulation data. The resulting cell density profiles for a given parameter set were compared against the *in-vitro* estimated cell density profiles of a dataset. This process was applied to each of the 12 datasets separately. Approximately 14300 different sets of model parameters were assessed using the TMCMC method for each of the 12 datasets. The obtained manifold of the inferred PDFs for one dataset is presented in Appendix 2 Fig. 1a. The marginal distributions and the average values along with their corresponding standard deviations from the posterior PDFs of the model parameters of these datasets are presented in Fig. 2a, and in Appendix 2 Table 1. Most of the estimated model parameters exhibited low uncertainty compared to range of their respective prior PDFs. The growth rate s corresponded to a cell doubling time equal to 3.461 ± 0.013 days, (mean ± SEM). The diffusion constants for the cells, D_u, and chemotactic agents, D_f, suggest that diffusion was more dominant compared to advection,

especially closer to the end of the experiment (last 3 time-points). This occurs due to the fact that the chemotactic signals diffuse in space, hence their gradient towards the bottom becomes less steep (Appendix 2 Fig. 1b).

The parameter r was found to have the largest variation and uncertainty across and within the datasets, respectively, which implied that its contribution was relatively smaller. To test this, we performed global sensitivity analysis of the model parameters with response to the tumour volume in 3D space, and the tumour area at the bottom with density values greater than 10^{-3}. The resulting rank correlation matrix between the model parameters and the outputs confirmed that r contributed less than the rest of the parameters (Appendix 3).

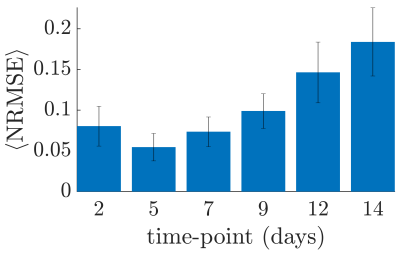


Figure 3. Average and standard deviation of Normalized Root Mean Squared error across all datasets across 6 time-points.

A visual representation of the *in-silico* cell density profiles, presented in Fig. 2b, using the calibrated parameters, shows that the model predictions reproduced the overall behaviour observed in the experiments, i.e. the biased movement of the cells towards the bottom. The Normalized Root Mean Squared Error (NRMSE) of the cell density evaluated at each spatial grid point per time point is presented in Fig. 3, excluding day 0, when simulation and experimental data were identical.

Spatial Analysis & HDC Model Validation

The estimated model parameters were subsequently used in the hybrid model (Fig. 7), separately for each dataset. The resulting *in-silico* cellular coordinates were analysed and compared to the corresponding *in-vitro* coordinates of the centroids from the segmented fluorescent nuclei of the cells. The quantitative characterization of the spatial distributions of the cells was performed using the IN, and NN Euclidean distance distributions. The IN distance distributions quantify the positioning of the cells relative to one another, while the NN distance distributions measure the distances between each cell and their nearest neighbouring cell. The resulting IN distance distributions, depicted in Fig. 4a, show that the distributions remained relatively stable across all samples and time, for both experiments and simulations, with a characteristic peak distance at ∼1 mm. The cosine similarity test yielded an average similarity value equal to 0.9896 ± 0.0109, suggesting high similarity between IN distance distributions from experiments and simulations. Their similarity remained high across all time-points, as shown in Fig. 4g. On the other hand, the NN distance distributions, presented in Fig. 4b, initially formed wide distributions that gradually tended to become narrower around lower neighbourhood radii values with respect to time, across all samples, with a characteristic peak at ∼15 μm for the experiments, and ∼10 μm for the simulations. These peaks can be interpreted based on the hybrid model hypotheses, specifically regarding the cell division where the daughter cells are placed next to each other, and the adhesion that prevents migration. The average cosine similarity between NN distance distributions from experiments and simulations was equal to 0.6184 ± 0.2226. The similarity between experimental and simulation NN distance distributions decreased as a function of time, as shown in Fig. 4g. The different characteristic peaks in the NN Distance Distributions contributed to the decreasing similarity values. We attribute these differences to grid size effects, from which lattice cellular automata typically suffer ***Rejniak and Anderson*** (***2011***). An increase of the grid size would correct this error. According to the definition of NN distance, it can be viewed as a special case of the IN distance. In turn, we would expect that the narrowing of the NN distance distributions would destabilize the IN distance distributions. However, the maintenance of their shape can be interpreted as a result of the organization of the cells into smaller clusters that maintained a relatively constant distance, the synchronized division of the cells, as well as their overall accumulation towards the glass bottom of the wells with respect to time.

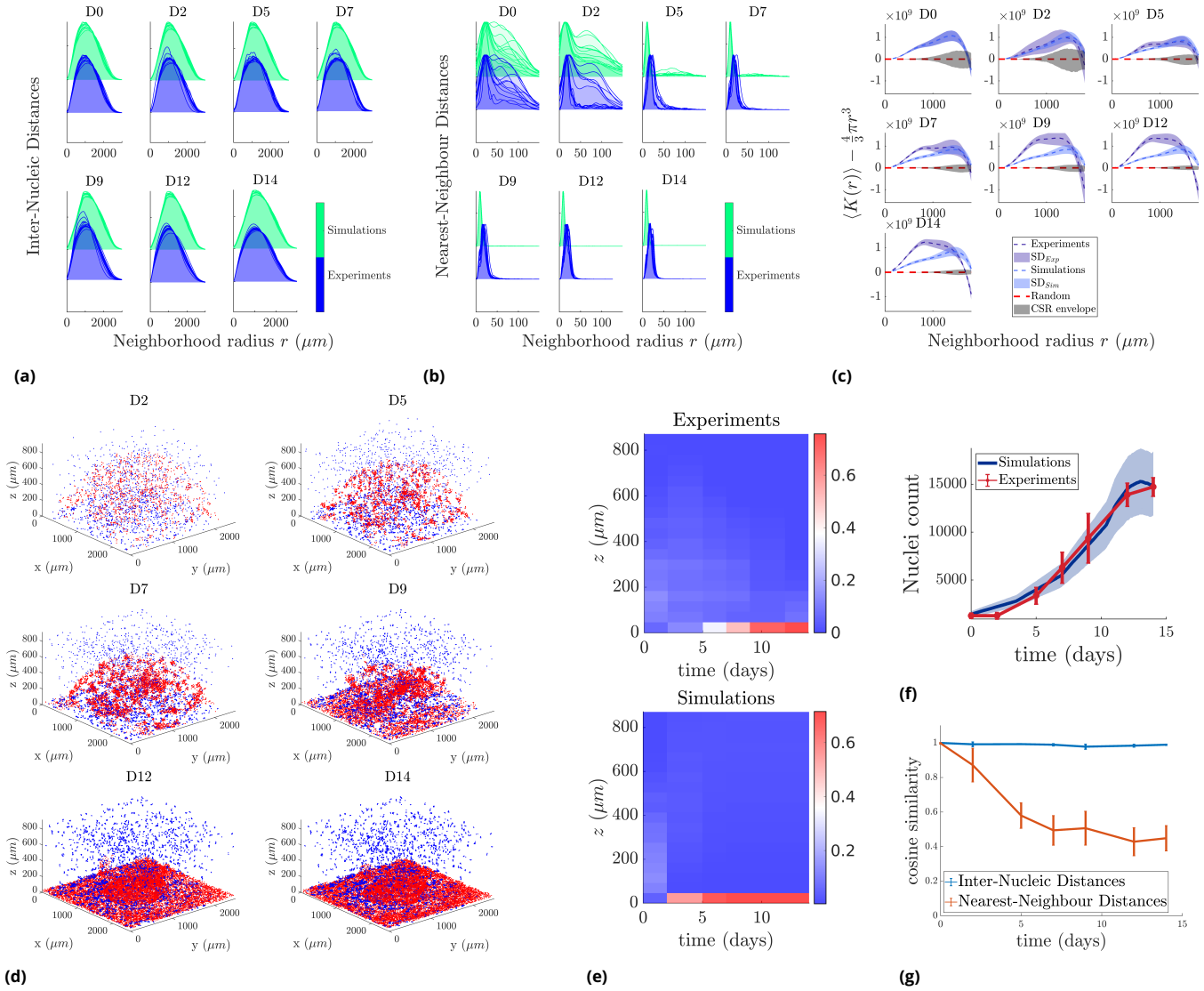


Figure 4. Spatial analysis and comparisons between experiments and simulations (a) Inter-Nucleic Euclidean distance distributions. The title (D#) denotes the time-point in days. (b) Nearest-Neighbour Euclidean distance distributions. (c) Complete Spatial Randomness test; average values of *K*-function across all samples and the corresponding standard error of mean (SEM). (d) Spatial distributions of cells from the cellular automaton (blue) and its corresponding experimental dataset (red) with respect to time. (e) Heatmaps of the normalized number of cells across the *z*-dimension, and across time. The normalization was performed across the z-dimension. (f) Average number of cells across all datasets with respect to time for simulations and experiments. (g) Cosine similarity test for IN and NN distance distributions between experimental and simulation results.

To investigate the spatial organization of the cells, we performed the CSR test, using Ripley's K-function ***Dixon*** (***2014***). Specifically, we examined whether the cells, represented by their nuclei centroids, were randomly distributed in space. The results depicted in Fig. 4c indicate substantial differences from a uniform random distribution for both experiments and simulations. For the experimental data, we observed clustering for a wide range of neighbourhood radii, as well as an increasing dispersion for longer distances across all samples, with respect to time. The results from the simulation data did not exhibit significant differences compared to the results from the experiments until day 2. Starting on day 9, we observed that the K-function of the experimental data indicates more pronounced clustering for smaller values of neighbourhood radii, and more pronounced dispersion patterns for longer distances, compared to the K-function of the simulation data.

The spatial distributions of the *in-silico* cells, and the corresponding experimental dataset are presented in Fig. 4d. We observe that both *in-silico* and *in-vitro* cells performed a biased movement towards the bottom, with the *in-silico* cells characterized also by a more pronounced random motion. Similarly, the numbers of cells across the z-dimension in Fig. 4e exhibit similarities between experiment and simulations, even though a relatively small number of cells appears to maintain elevated positions. The *in-vitro* and *in-silico* number of cells with respect to time, shown in Fig. 4f, are in coherence, exhibiting a logistic growth.

Discussion

We presented a novel framework that combines 3D cell culture experiments, multiscale models, parameter estimation, and spatial validation techniques to examine and quantify the morphology and mechanisms of cancer progression. We applied the proposed framework to 3D cultures of TNBC cells in Matrigel ECM, and we modeled this behaviour using a multiscale HDC model. The parameters of the continuum model were estimated using Bayesian inference and a TMCMC algorithm. Our results suggest an overall agreement between the calibrated model, and the experimental observations. The estimated parameters were used in the HDC model for a detailed simulation of the spatial distributions of the cells. The results of both experiments and simulations were analysed using spatial statistical analysis techniques to quantify the morphology of both *in-vitro* and *in-silico* cancer progression. In the following paragraphs, we discuss the relation between the observed spatial patterns, and the underlying mechanisms, particularly as related to the biased movement of the cells towards the bottom.

Relation between Morphological patterns and Biological mechanisms

The continuum KS model consists of diffusion, growth, and advection terms that represent the random motion, proliferation, and biased movement of the cells towards the bottom, respectively. The estimated diffusion constants (Fig. 2a) suggest that random motion played a significant role in the overall cell movement. Together with the unconstrained migration phase, these parameters affected the morphology of the cancer cells, which was reflected by the increased NN distance values on day 2 (Fig. 4b). The effect of advection, together with the constrained migration were more apparent after day 2 (Fig. 4d). These two parameters reflect the tendency of the cells to form clusters and their tendency to move towards the bottom. This effect was also observed in the NN distances between days 5 and 14 (Fig. 4b), as well as in the heatmaps of the number of cells across different z-values (Fig. 4e), which shows a comparable number of *in-silico* and *in-vitro* cells near the bottom. The visualization of the cells (Fig. 4d) shows that not all the *in-silico* cells tended to move towards the bottom. This is a result of the fast diffusion of the chemotactic signals, which led to a more uniform distribution across the z-dimension compared to day 0 (Appendix 2 Fig. 1b). The resulting differences in cell attachment to the bottom are also apparent in the K-function of the experiments and simulations. The *in-vitro* cell accumulation resulted in more pronounced clustering patterns for smaller neighbourhood radii with respect to time. The increase in the adhesion parameter with respect to time restricts migration to the cells that have not reached

the bottom, contributing to the resulting NN distances. This parameter contributes also to the increased clustering of the *in-silico* cells shown in Fig. 4c, even though the changes in *K*-function were very small.

Biased movement and cell sedimentation

Despite the fact that chemotactic migration has been studied in both biological and mathematical levels ***Roussos et al.*** (***2011b***), ***Bubba et al.*** (***2019***), there is very limited discussion on the observed behaviour of the cells to move towards the bottom of the culture ***Liu et al.*** (***2020***). In this study, we hypothesized that cell sedimentation is a result of active, chemotactic cell migration due to the fact that gravity is not sufficient to describe this behaviour. Similarly, ***Liu et al.*** (***2020***) showed that if cells are cultured on the interface of two hydrogel structures, then they tend to move across the interface instead of settling on the bottom. Thus, passive movement due to gravity may not be sufficient to explain this biased movement, and cell sedimentation due to chemotactic migration remains a probable hypothesis.

The selected mathematical model was able to reproduce this biased movement, and the overall framework allowed us to quantify the movement in terms of both spatial patterns and underlying mechanisms. The proposed computational part of the framework allowed us to investigate the mesoscopic scale (μm to mm) taking into account between 1000 and 18000 cells in 3D, exhibiting good performance in terms of processing times. The analysis showed that not all of the *in-silico* cells followed the chemotactic gradient. This phenomenon was also observed by ***Tweedy et al.*** (***2016***), but for a different reason. Their study showed that self-generated gradients may favour the leading wave of cells, because they break down chemoattractants; thus, the cells behind the front do not sense a gradient and move randomly. This phenomenon was not visible in our experiments, due to additional factors that contributed to the biased movement of the cells towards the bottom. These include the compression and degradation of Matrigel, as well as vibrations during the transfer of the samples to the microscope. These factors were not considered in the model; however, the proposed framework provides a promising tool for the study of models of higher complexity.

Future perspectives

The proposed framework enabled us to relate the underlying mechanisms of cancer progression with the observed morphological patterns. Future improvements may include incorporating a model term for the quantification of the effect of ECM degradation that may be responsible for the introduction of possible biases. The proposed framework can also be used to study the growth patterns of heterogeneous cell populations such as cancer cells and fibroblasts, as well as, study cancer progression in the presence of therapy. Importantly, potential differences in the morphological patterns in the presence and absence of therapy can be used to design therapeutic strategies that control not only the tumour size, but also their morphological patterns to minimize invasion. Overall, the presented framework yields great promise for a more complete quantitative understanding of the organization and progression of cancer.

Methods and Materials

Experiments

Cell preparation

Triple Negative Breast Cancer (TNBC) cells from the MDA-MB-231 cell line with nuclear GFP (histone transfection), were thawed and cultured at 5% CO2, 37 °C in DMEM (Gibco) at pH 7.2 supplemented with 10% fetal bovine serum (Wisent Bioproducts), 100 U/mL penicillin, 100 μg/mL streptomycin, and 0.25 μg/mL, and amphotericin B (Sigma) in T-75 flasks (Corning). The cells were passaged before reaching 85% confluence. Three passages were performed before the 3D cultures; cells were rinsed twice with DPBS and trypsin-EDTA (0.25%-1X, Gibco) was used to harvest them.

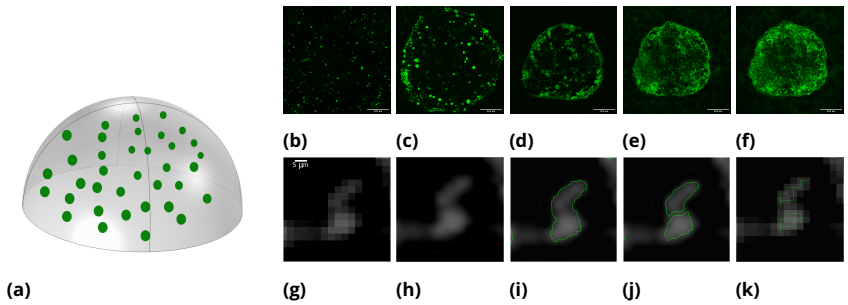


Figure 5. 3D cell cultures and nucleic segmentation. (a) Schematic representation of the cell/Matrigel geometry at day 0 of the experiment. Slices at $Z \approx 100\ \mu m$ from the bottom of the plate on days (b) 5, (c) 7, (d) 9, (e) 12, (f) 14. Scale bar: 500 µm. (g) Zoomed image of two cells on day 9. (h) Interpolation result resulting from (h). (i) Marker Controlled Watershed segmentation. (j) Nuclei splitting with Distance Based Watershed segmentation. (k) Rescaling back to original image size.

3D cell cultures

A cell-Matrigel (Corning) suspension was created using 0.25 mL of Matrigel (4 °C) and 5×10^4 MDA-MB-231/GFP cells. Droplets of 5 µL cell-Matrigel mixture were manually deposited onto a high performance #1.5 glass bottom 6-well plate (0.170±0.005 mm) (Fisher Scientific) (Fig. 5a). In total, 12 datasets were produced with 7 samples on days 0, 2, 5, 7, 9, 12, 14 each. (Fig. 5b-5f).

Imaging and Data preparation

Data acquisition was performed every 2-3 days for a total of 15 days using a confocal microscope (Nikon A1R HD25) coupled with a cell-culture chamber. The dimensions of the 3D cultures were approximately 2.5×2.5×0.9 mm^3. Cell localization was made possible by the GFP fluorophore that was present in cell nuclei. The fluorescent nuclei were segmented using an image processing and segmentation pipeline ***Dimitriou et al.*** (***2021***). The preprocessing of the image stacks included: (i) image denoising using the Poisson Unbiased Risk Estimation-Linear Expansion of Thresholds (PURE-LET) technique ***Luisier et al.*** (***2010***), (ii) intensity attenuation correction across the z-dimension ***Biot et al.*** (***2008***), (iii) background subtraction using the rolling ball algorithm ***Sternberg*** (***1983***) and manual thresholding of low intensity values using High-Low Look Up Tables (HiLo LUTS), and (iv) cubic spline interpolation of the xy-planes of the image stacks. The segmentation of the nuclei was performed using Marker Controlled Watershed segmentation and a classic Distance Based Watershed segmentation to split fused nuclei (Fig. 5g-5k). The segmented nuclei were then mapped to a 3D Cartesian space by detecting their centroid locations using a 26-connected neighbourhood tracing algorithm implemented in MATLAB ***MATLAB*** (***2018***). The final step was the calculation of spatial density profiles of the cells represented by their centroids, using the Kernel Density estimation via the Diffusion method ***Botev et al.*** (***2010***). Density calculation was performed using a grid of size $167 \times 167 \times 61$ such that each cell approximately occupied one grid point. The density matrices were interpolated, using linear interpolation, to match the spatial grid size of the simulations $(480 \times 480 \times 176)$.

Multiscale HDC Model

Chemotactic hypothesis

Previous studies ***Liu et al.*** (***2020***) have shown that cells in 3D cultures using hydrogel matrices such as Collagen I or Matrigel tend to move towards the bottom of the space. We hypothesized that this behaviour occurs due to three main reasons; first, the MDA-MB-231 cells are naturally adherent cells, hence the cells tend to remain attached to each other to function properly; second, at the beginning and throughout the course of the experiment, the cells secrete chemotactic signals that enable cell migration and tend to bring the cells closer to each other; third, the cells that are closer to the glass bottom secrete signals at the beginning of the experiment creating a chemotactic gra-

dient decreasing from the bottom towards the top of the space. The rationale behind the third hypothesis is that the glass is a surface that favours cell attachment, hence the cells that are closer to this surface secrete these signals to indicate it as a site of preference. This hypothesis is supported by recent findings on self-generated chemotactic gradients ***Tweedy et al.*** (***2020***); ***Tweedy and Insall*** (***2020***); ***Tweedy et al.*** (***2016***); ***Susanto et al.*** (***2017***) with the difference that we assumed that the chemoattractants stem from the cells and they do not pre-exist in the 3D space.

Continuum model

To examine this hypothesis, we used a system of two Keller-Segel (KS) type equations for cancer cell density and chemotactic agent density respectively, which additionally takes into account random motion of cancer cells and chemotactic agents, logistic growth of cancer cells, as well as the increase of chemotactic agents depending on their current concentration in space and the presence of cancer cells. The spatiotemporal evolution of cancer cell, u, and chemotactic agent, f, densities are obtained by the following PDEs:

$$\frac{\partial u}{\partial t} = D_u \nabla^2 u + su(1-u) - \chi \nabla \cdot [u(1-u)\nabla f], \qquad \text{in } \Omega \tag{1}$$

$$\frac{\partial f}{\partial t} = D_f \nabla^2 f + \rho f u(1-u), \qquad \text{in } \Omega \tag{2}$$

$$\nabla u \cdot \vec{n} = \nabla f \cdot \vec{n} = 0, \qquad \text{in } \partial\Omega \tag{3}$$

where D_u, D_f are the diffusion constants, and s, ρ are the growth constants of the cell and signal densities, respectively, and χ is the advection constant of the cells. The right hand side of (1) consists of three terms; the diffusion term $D_u \nabla^2 u$ that represents the random motion and expansive growth of the cancer cells, the growth term $su(1-u)$ that increases the density of the tumour in a logistic manner, and the nonlinear advection term $-\chi \nabla \cdot [u(1-u)\nabla f]$ that represents the biased movement of the cells towards the direction where the gradient of the chemotactic signal density increases.

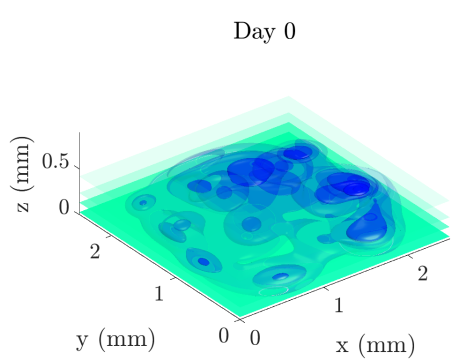


Figure 6. Initial conditions of the continuum model from one of the 12 datasets. The blue colour map represents the cell density profile, u, and it is directly obtained from the experimental data. The green colour map represents the chemotactic agents density profile, f and it is calculated using (4).

The $(1-u)$ factor in the advection term was added to avoid unwanted overcrowding of the cells that may lead to spikes of cell density ***Hillen and Painter*** (***2001***). In (2), the evolution of the signal density depends on the diffusion of the signal in 3D space, represented by $D_f \nabla^2 f$ and the production of signals depending on the current signal density and cell density in space, $\rho f u(1-u)$. Similarly, $(1-u)$ limits the signal when overcrowding takes place. The spatial domain Ω had the same size as the experimental data, 2.5×2.5×0.917 mm^3, and it was represented by 480×480×176 grid points. We considered no-flux Neumann boundary conditions (B.C.) in (3), where $\vec{n}$ is the outward unit normal to $\partial\Omega$.

The initial conditions (I.C.) for this problem were chosen based on the experimental data and the chemotactic hypothesis. Specifically, the initial cell density profiles of the simulations were chosen to be the spatial cell density profiles of day 0 of the experiment (Fig. 6). The I.C. for the chemotactic signals were based on the fact that the cells were, initially, uniformly distributed in the 3D space, and separated from each other. Cells attached to the bottom glass were assumed

to chemotactic secrete agents first, which in turn promoted the secretion of these agents by the above floating cells as described in (4).

$$f(x, y, z, t = 0) = e^{\left(-\frac{z}{0.26\ (\mathrm{mm})}\right)}\mathrm{I}(u),\ \mathrm{I}(u) = \begin{cases} 1, & \text{if } u > 0 \\ 0, & \text{if } u = 0 \end{cases} \tag{4}$$

Numerical methods

We used the operator-splitting technique to approximate the diffusion, advection, and reaction operators. The diffusion terms were approximated by the Alternating Direction Implicit (ADI) Douglas-Gunn (DG) method ***Douglas and Gunn*** (***1964***). The advection term was approximated by the explicit Lax-Wendroff (LxW) method ***Lax and Wendroff*** (***1960***), coupled with the Monotonic Upstream-Centered Scheme for Conservation Laws (MUSCL) flux limiter ***van Leer*** (***1979***). The integration in time was performed using the Strang splitting scheme ***Strang*** (***1968***). At every time-step, the Strang splitting scheme evolves the advection and reaction terms by $0.5\mathrm{d}t$, then the diffusion operator by $\mathrm{d}t$, and again the advection and reaction operators by $0.5\mathrm{d}t$. The accuracy of this scheme is second-order for both space and time. The proposed numerical scheme was implemented on GPUs using the CUDA/C programming language. Each simulation required approximately 1-5 minutes to complete in a V100-16GB Nvidia GPU (Appendix 1).

Hybrid model

We hybridized the KS model based on the technique presented in ***Anderson*** (***2003***, 2005). Specifically, we discretized (1) using the forward time central differences scheme (FTCS) using the approximations found in ***Franssen et al.*** (***2019***):

$$u^{n+1}_{i,j,k} = u^n_{i,j,k}P_0 + u^n_{i+1,j,k}P_1 + u^n_{i-1,j,k}P_2 + u^n_{i,j+1,k}P_3 + u^n_{i,j-1,k}P_4 + u^n_{i,j,k+1}P_5 + u^n_{i,j,k-1}P_6 \tag{5}$$

where the grouped terms $P_i,\ i = 0, ..., 6$ denote probabilities of the cells of remaining stationary (P_0) or moving back (P_1), front (P_2), left (P_3), right (P_4), down (P_5), up (P_6), defined as

$$\begin{aligned} P_0 &= 1 - \frac{6D_u\mathrm{d}t}{\mathrm{d}x^2} \\ P_{1,2} &= \frac{D_u\mathrm{d}t}{\mathrm{d}x^2} \mp \frac{\chi\mathrm{d}t}{4\mathrm{d}x^2}(f_{i+1,j,k} - f_{i-1,j,k}) \\ P_{3,4} &= \frac{D_u\mathrm{d}t}{\mathrm{d}x^2} \mp \frac{\chi\mathrm{d}t}{4\mathrm{d}x^2}(f_{i,j+1,k} - f_{i,j-1,k}) \\ P_{5,6} &= \frac{D_u\mathrm{d}t}{\mathrm{d}x^2} \mp \frac{\chi\mathrm{d}t}{4\mathrm{d}x^2}(f_{i,j,k+1} - f_{i,j,k-1}) \end{aligned} \tag{6}$$

Since the cells were approximately 15 μm in size and the spatial grid points had a distance of 5.2 μm between each other, we assumed that each cell occupied 3 grid points in each direction. To account for this, we modified (5) and (6) by changing the indices that point to a direction to two grid points instead of one, i.e. i±2 instead of i±1 etc. The moving probabilities were then passed to a cellular automaton that updated the position and state of each cell.

The cellular automaton (CA) is presented in Fig. 7a. The CA takes into account three cellular states; alive, quiescent and dead. At every time step it checks if a cell can undergo spontaneous death based on the probabilities shown in Fig. 7b, and updates the age of the alive cells. The spontaneous death probability was increased after day 10. This hypothesis was based on increased cell crowding, which resulted in a potential shortage of nutrients or accumulation of metabolic waste products. The CA checks if any cell has reached the proliferation age that is determined based on the estimated parameter s (days)$^{-1}$ of the continuum model. We estimated the doubling time from the exponential phase of growth, e^{st}, and the resulting formula $t_{\text{double}} = \ln 2/s$. If a cell is ready to divide, the algorithm separates into two processes based on cell position in space. If the cell is attached on the glass and there is sufficient space, then the division will be performed on the

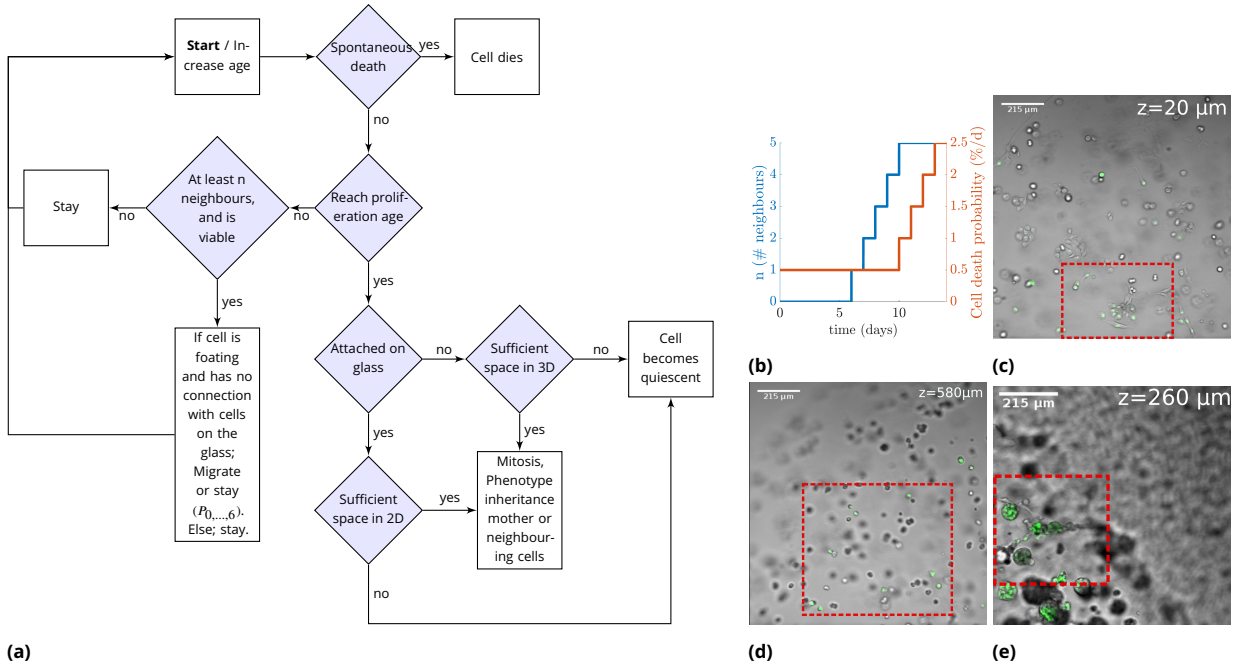


Figure 7. (a) Flowchart of the cellular automaton. (b) Migration, adhesion, and cell death probability parameters of the cellular automaton were changed over time. (c) Cells settled at the bottom (z=20 µm), and (d) cells floating at z=580 µm on day 2. Cells settled on the bottom had stellar shapes, while cells floating in the Matrigel had rounder shapes. (e) Cells floating at z=260 µm on day 7. Some floating cells changed to stellar shapes resembling those attached on the bottom, as shown in panel (c).

glass; otherwise, the cell will divide in any direction of the 3D space if there is sufficient space. On the other hand, if there is not sufficient space, the cell becomes quiescent. If the cell is not ready to divide, the CA turns to a migration program.

The first condition for migration considers an adhesion parameter, defined as the number of neighbours that surround a cell, and the second is the state of the cell. We hypothesized that the number of neighbours required for cell migration increases over time (Fig. 7b), due to the fact that the initial distribution of cells in the 3D space is sparse; hence they migrate, freely, to search for other cells to attach. However, as cell clustering occurs due to cell division or cell contact, migration becomes less frequent since the cells become more attached to each other. If a cell satisfies these conditions, the algorithm checks the position of the cell. If a cell is settled on the bottom of the space or is connected with a cell located on the bottom, it cannot migrate; otherwise, the cell can migrate in 3D space given the moving probabilities $P_0, ..., P_6$. These two, constrained and unconstrained, migration phenotypes resemble epithelial and mesenchymal phenotypes, respectively, and the transition between them can be found in the literature as mesenchymal to epithelial transition (MET) ***Gunasinghe et al.*** (***2012***). Indeed, changes in cellular morphology were observed between cells settled on the bottom and cells floating in the Matrigel. In Fig. 7c, 7d we observe floating cells during the early days of the experiment with round shapes. However, at later time-points (Fig. 7e), we observed stellar shapes for the floating cells probably, due to increased adhesion between them.

Bayesian Inference for calibration of the continuum model

The Keller-Segel model, M, (Eq (1)-(3)) includes a set of parameters $\theta = \{D_u, s, \chi, D_f, r\}$ that are considered unknown. We used their Probability Distribution Functions (PDF) and the calculated densities from the 3D cell culture data, $\mathscr{D}$, to assess the most probable parameter values according to Bayes' rule

$$\mathbb{P}(\theta|\mathscr{D}, M) \propto \mathbb{P}(\mathscr{D}|\theta, M)\mathbb{P}(\theta) \tag{7}$$

where $\mathbb{P}(\theta|\mathscr{D}, M)$ is the posterior PDF of the model parameters θ given the observed data $\mathscr{D}$ and the model M, $\mathbb{P}(\mathscr{D}|\theta, M)$ is the likelihood of the observed data $\mathscr{D}$ given the model M and the parameters θ, and $\mathbb{P}(\theta)$ is the prior PDF. We assume uninformative, uniform distributions for the model parameter prior PDFs. The experimental data consisted of 12 datasets and each of them had samples collected at 7 time-points. The datasets were assumed to be independent and the model was evaluated for each dataset separately. The likelihood was defined as

$$L(\theta; \mathbf{d}) = \prod_{i=1}^{n} \frac{1}{\sigma_d \sqrt{2\pi}} \exp\left(-\frac{(d_i - q_i(\theta))^2}{2\sigma_d^2}\right) \tag{8}$$

where n is the number of spatial grid points, $\mathbf{d}$ the density profile of the corresponding sample in a dataset, d_i, q_i the density values of the experimental sample and simulation result, respectively, at the grid point i, and σ_d the variance of the distribution of the likelihood.

We used a Transitional Markov Chain Monte Carlo (TMCMC) algorithm implemented in the Π4U package ***Hadjidoukas et al.*** (***2015***). The TMCMC algorithm iteratively constructs series of intermediate posterior PDFs

$$\mathbb{P}_j(\theta|\mathscr{D}, M) \propto \mathbb{P}(\mathscr{D}|\theta, M)^{\rho_j} \mathbb{P}(\theta) \tag{9}$$

where $j = 0, ..., m$ is the index of the Monte Carlo time series (generation index), and ρ_j controls the transition between the generations, and $0 < \rho_0 < \rho_1 < \cdots < \rho_m = 1$. The TMCMC method can utilize a large number of parallel chains that are evaluated in each Monte Carlo step to reach a result close to the true posterior PDF.

Since the ratio of model parameters to time-points is small (5:7) for the continuum model, we used all the time-points for the calibration of the continuum model. Validation was performed using the hybrid (discrete-continuum) model using the spatial statistical measures described below.

Spatial Analysis - HDC Model Validation

Complete Spatial Randomness Test of Spatial Cell Distributions

The Complete Spatial Randomness (CSR) test examines whether the observed spatial point patterns, in our case the centroids of the nuclei, can be described by a uniform random distribution ***de Back et al.*** (***2019***). The CSR test was performed using Ripley's *K*-function and the *spatstat* ***Baddeley et al.*** (***2004***) package of R ***R Core Team*** (***2020***). The K-function ***Dixon*** (***2014***) is defined as the ratio between the number of the events, i.e. locations of points, j within a distance t from the event i, over the total number of events N, in the studied volume V

$$K(t) = \frac{1}{\hat{\lambda}} \sum_i \sum_{j \neq i} I(d_{ij} < t), \; I(x) = \begin{cases} 1, \text{ if } x = \text{true} \\ 0, \text{ otherwise} \end{cases} \tag{10}$$

where $\hat{\lambda} = N/V$ denotes the average density of events, N, in the studied volume V, d_{ij} is the distance between events i and j, and t is the search radius. The K-function was calculated for all datasets and compared against complete spatial randomness following a Poisson process $K(t) = 4\pi t^3/3$ ***Dixon*** (***2014***) for three spatial dimensions. Isotropic edge correction was applied in the calculation of the K-function. The volume used for the calculation was the same with that used in the simulations, i.e. $2.5 \times 2.5 \times 0.917$ mm^3. To assess the uncertainty of the random variable K, we produced a CSR envelope by generating 100 random distributions and calculating the K-function for each of them. The envelope was created by keeping the minimum and maximum values of the resulting K values. A substantial upward separation of the observed K-function from the theoretical random K-function denotes clustered patterns, while a downward separation denotes dispersed patterns ***de Back et al.*** (***2019***). Both separation types suggest non-randomness of the examined spatial distribution.

Characterization of the Spatial Cell Distributions

The *Inter-Nucleic (IN) Distance Distribution* for a given sample was calculated by the pairwise Euclidean distances between all nuclei. Given two nuclei i and j with centroid positions $\mathbf{p_i} = (x_i, y_i, z_i)$ and $\mathbf{p_j} = (x_j, y_j, z_j)$ respectively, their pairwise Euclidean distance is given by

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2},\ i, j = 1...N,\ i \neq j \tag{11}$$

where N the total number of nuclei.

The *Nearest-Neighbour (NN) Distance Distribution* for a given sample was calculated using the distances between the nearest neighbours of the nuclei. The nearest neighbour distance for a given nucleus i is given by the minimum IN Distance between the nucleus i and all the other nuclei of the sample, i.e. $D^i_{NN} = \min_{i,j}\{D_{ij}\}$, $j \in [1, N]$, $j \neq i$.

The comparisons between the *in-vitro* and *in-silico* IN and NN distance distributions were performed using the cosine similarity test ***Han et al.*** (***2012***), in MATLAB ***MATLAB*** (***2018***) (Appendix 4).

Acknowledgments

This work was supported by the Cyprus Research and Innovation Foundation (Project: INTERNATIONAL/OTHER/0118/0018), the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 34362 awarded to GDM), and Compute Canada (www.computecanada.ca). N.M.D. thanks Ferenc Molnár (University of Notre Dame) for the discussions in GPU computing, Panagiotis Hadjidoukas (IBM Research-Zürich) and Panagiotis Angelikopoulos (D.E. Shaw Research) for the discussions on TMCMC, Remí Dagenais (McGill) for reviewing the article. N.M.D. thanks Stavros Niarchos Foundation (F237055R00), Werner Graupe (F202955R00) and McGill University (90025) for the scholarships. S.F.T. thanks McGill University for the McGill Engineering Doctoral Award (90025) and the FRQNT (291010) for the scholarships.

References

Abler D, Büchler P, Rockne RC. Towards Model-Based Characterization of Biomechanical Tumor Growth Phenotypes. In: Bebis G, Benos T, Chen K, Jahn K, Lima E, editors. *Mathematical and Computational Oncology* Cham: Springer International Publishing; 2019. p. 75–86.

Achilleos A, Loizides C, Stylianopoulos T, Mitsis GD. Multi-process dynamic modeling of tumor-specific evolution. In: *13th IEEE International Conference on BioInformatics and BioEngineering*; 2013. p. 1–4. doi: 10.1109/BIBE.2013.6701614.

Achilleos A, Loizides C, Hadjiandreou M, Stylianopoulos T, Mitsis GD. Multiprocess Dynamic Modeling of Tumor Evolution with Bayesian Tumor-Specific Predictions. Annals of Biomedical Engineering. 2014; 42(5):1095–1111. https://doi.org/10.1007/s10439-014-0975-y, doi: 10.1007/s10439-014-0975-y.

Adams, B M , Bohnhoff, W J , Dalbey, K R , Ebeida, M S , Eddy, J P , Eldred, M S , Frye, J R , Geraci, G , Hooper, R W , Hough, P D , Hu, K T , Jakeman, J D , Khalil, M , Maupin, K A , Monschke, J A , Ridgway, E M , Rushdi, A A , Stephens, J A , Swiler, L JG. Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.8 Reference Manual. Sandia Technical Report SAND2014-5015; 2014.

Anderson ARA, Chaplain MAJ, Rejniak KA, Fozard JA. Single-cell-based models in biology and medicine. Mathematical Medicine and Biology: A Journal of the IMA. 2008; 25(2):185–186. https://doi.org/10.1093/imammb/dqn008, doi: 10.1093/imammb/dqn008.

Anderson ARA. In: Alt W, Chaplain M, Griebel M, Lenz J, editors. A Hybrid Discrete-continuum Technique for Individual-based Migration Models Basel: Birkhäuser Basel; 2003. p. 251–259. https://doi.org/10.1007/978-3-0348-8043-5{_}19, doi: 10.1007/978-3-0348-8043-5_19.

Anderson ARA. A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. Mathematical Medicine and Biology: A Journal of the IMA. 2005; 22(2):163–186. http://dx.doi.org/10.1093/imammb/dqi005.

de Back W, Zerjatke T, Roeder I. In: Klein G, Wuchter P, editors. Statistical and Mathematical Modeling of Spatiotemporal Dynamics of Stem Cells New York, NY: Springer New York; 2019. p. 219–243. https://doi.org/10.1007/978-1-4939-9574-5{_}17, doi: 10.1007/978-1-4939-9574-5_17.

Baddeley AJ, Turner R, Others, Spatstat: An R package for analyzing spatial point pattens. University of Western Australia. Department of Mathematics and Statistics; 2004.

Biot E, Crowell E, Hofte H, Maurin Y, Vernhettes S, Andrey P. A new filter for spot extraction in N-dimensional biological imaging. In: *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* IEEE; 2008. p. 975–978.

Botev ZI, Grotowski JF, Kroese DP, Others. Kernel density estimation via diffusion. The annals of Statistics. 2010; 38(5):2916–2957.

Bubba F, Pouchol C, Ferrand N, Vidal G, Almeida L, Perthame B, Sabbah M. A chemotaxis-based explanation of spheroid formation in 3D cultures of breast cancer cells. Journal of Theoretical Biology. 2019; 479:73–80. https://www.sciencedirect.com/science/article/pii/S0022519319302796, doi: https://doi.org/10.1016/j.jtbi.2019.07.002.

Collis J, Connor AJ, Paczkowski M, Kannan P, Pitt-Francis J, Byrne HM, Hubbard ME. Bayesian Calibration, Validation and Uncertainty Quantification for Predictive Modelling of Tumour Growth: A Tutorial. Bulletin of Mathematical Biology. 2017; 79(4):939–974. https://doi.org/10.1007/s11538-017-0258-5, doi: 10.1007/s11538-017-0258-5.

Cristini V, Lowengrub JS. Multiscale modeling of cancer: an integrated experimental and mathematical modeling approach. United Kingdom: Cambridge University Press; 2010.

Dimitriou NM, Flores-Torres S, Kinsella JM, Mitsis GD. Detection and Spatiotemporal analysis of in-vitro 3D migratory Triple-Negative Breast cancer cells. bioRxiv. 2021; https://www.biorxiv.org/content/early/2021/07/30/2021.07.29.454312, doi: 10.1101/2021.07.29.454312.

Dixon PM. Ripley's K Function. Wiley StatsRef: Statistics Reference Online. 2014; 3:1796–1803.

Douglas J, Gunn JE. A general formulation of alternating direction methods. Numerische Mathematik. 1964; 6(1):428–453. https://doi.org/10.1007/BF01386093, doi: 10.1007/BF01386093.

Farrell K, Oden JT, Faghihi D. A Bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems. Journal of Computational Physics. 2015; 295:189–208. https://www.sciencedirect.com/science/article/pii/S0021999115002430, doi: https://doi.org/10.1016/j.jcp.2015.03.071.

Franssen LC, Lorenzi T, Burgess AEF, Chaplain MAJ. A Mathematical Framework for Modelling the Metastatic Spread of Cancer. Bulletin of Mathematical Biology. 2019; 81(6):1965–2010. https://doi.org/10.1007/s11538-019-00597-x, doi: 10.1007/s11538-019-00597-x.

Friedl P, Locker J, Sahai E, Segall JE. Classifying collective cancer cell invasion. Nature Cell Biology. 2012; 14(8):777–783. https://doi.org/10.1038/ncb2548, doi: 10.1038/ncb2548.

Greenspan HP. Models for the Growth of a Solid Tumor by Diffusion. Studies in Applied Mathematics. 1972; 51(4):317–340. https://onlinelibrary.wiley.com/doi/abs/10.1002/sapm1972514317, doi: 10.1002/sapm1972514317.

Gunasinghe NPAD, Wells A, Thompson EW, Hugo HJ. Mesenchymal–epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. Cancer and Metastasis Reviews. 2012; 31(3):469–478. https://doi.org/10.1007/s10555-012-9377-5, doi: 10.1007/s10555-012-9377-5.

Hadjidoukas PE, Angelikopoulos P, Papadimitriou C, Koumoutsakos P. Π4U: A high performance computing framework for Bayesian uncertainty quantification of complex models. Journal of Computational Physics. 2015; 284:1–21. http://www.sciencedirect.com/science/article/pii/S0021999114008134, doi: https://doi.org/10.1016/j.jcp.2014.12.006.

Han J, Kamber M, Pei J. 2 - Getting to Know Your Data. In: Han J, Kamber M, Pei JBTDMTE, editors. *The Morgan Kaufmann Series in Data Management Systems* Boston: Morgan Kaufmann; 2012.p. 39–82. http://www.sciencedirect.com/science/article/pii/B9780123814791000022, doi: https://doi.org/10.1016/B978-0-12-381479-1.00002-2.

Hawkins-Daarud A, Prudhomme S, van der Zee KG, Oden JT. Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumor growth. Journal of Mathematical Biology. 2013; 67(6):1457–1485. https://doi.org/10.1007/s00285-012-0595-9, doi: 10.1007/s00285-012-0595-9.

Hickman JA, Graeser R, de Hoogt R, Vidic S, Brito C, Gutekunst M, van der Kuip H, Consortium IMIP. Three-dimensional models of cancer for pharmacology and cancer cell biology: Capturing tumor complexity in vitro/ex vivo. Biotechnology Journal. 2014; 9(9):1115–1128. https://onlinelibrary.wiley.com/doi/abs/10.1002/biot.201300492, doi: 10.1002/biot.201300492.

Hillen T, Painter K. Global Existence for a Parabolic Chemotaxis Model with Prevention of Overcrowding. Advances in Applied Mathematics. 2001; 26(4):280–301. https://www.sciencedirect.com/science/article/pii/S0196885801907213, doi: https://doi.org/10.1006/aama.2001.0721.

Hoehme S, Bertaux F, Weens W, Grasl-Kraupp B, Hengstler JG, Drasdo D. Model Prediction and Validation of an Order Mechanism Controlling the Spatiotemporal Phenotype of Early Hepatocellular Carcinoma. Bulletin of Mathematical Biology. 2018; 80(5):1134–1171. https://doi.org/10.1007/s11538-017-0375-1, doi: 10.1007/s11538-017-0375-1.

Hormuth DA, Weis JA, Barnes SL, Miga MI, Rericha EC, Quaranta V, Yankeelov TE. A mechanically coupled reaction–diffusion model that incorporates intra-tumoural heterogeneity to predict in vivo glioma growth. Journal of The Royal Society Interface. 2017; 14(128):20161010. https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2016.1010, doi: 10.1098/rsif.2016.1010.

Hormuth DAI, Weis JA, Barnes SL, Miga MI, Quaranta V, Yankeelov TE. Biophysical Modeling of In Vivo Glioma Response After Whole-Brain Radiation Therapy in a Murine Model of Brain Cancer. International journal of radiation oncology, biology, physics. 2018 apr; 100(5):1270–1279. doi: 10.1016/j.ijrobp.2017.12.004.

Jarrett AM, Hormuth DA, Adhikarla V, Sahoo P, Abler D, Tumyan L, Schmolze D, Mortimer J, Rockne RC, Yankeelov TE. Towards integration of 64Cu-DOTA-trastuzumab PET-CT and MRI with mathematical modeling to predict response to neoadjuvant therapy in HER2 + breast cancer. Scientific Reports. 2020; 10(1):20518. https://doi.org/10.1038/s41598-020-77397-0, doi: 10.1038/s41598-020-77397-0.

Jensen MM, Jørgensen JT, Binderup T, Kjær A. Tumor volume in subcutaneous mouse xenografts measured by microCT is more accurate and reproducible than determined by 18 F-FDG-microPET or external caliper. BMC Medical Imaging. 2008; 8(1):16. https://doi.org/10.1186/1471-2342-8-16, doi: 10.1186/1471-2342-8-16.

Jin W, Shah ET, Penington CJ, McCue SW, Chopin LK, Simpson MJ. Reproducibility of scratch assays is affected by the initial degree of confluence: Experiments, modelling and model selection. Journal of Theoretical Biology. 2016; 390:136–145. http://www.sciencedirect.com/science/article/pii/S0022519315005676, doi: https://doi.org/10.1016/j.jtbi.2015.10.040.

Kikuchi A, Kozuma S, Sakamaki K, Saito M, Marumo G, Yasugi T, Taketani Y. Fractal tumor growth of ovarian cancer: Sonographic evaluation. Gynecologic Oncology. 2002; 87(3):295–302. doi: 10.1006/gyno.2002.6842.

Lax P, Wendroff B. Systems of conservation laws. Communications on Pure and Applied Mathematics. 1960; 13(2):217–237. https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160130205, doi: https://doi.org/10.1002/cpa.3160130205.

van Leer B. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. Journal of Computational Physics. 1979; 32(1):101–136. https://www.sciencedirect.com/science/article/pii/0021999179901451, doi: https://doi.org/10.1016/0021-9991(79)90145-1.

Lima EABF, Oden JT, Hormuth DA, Yankeelov TE, Almeida RC. Selection, calibration, and validation of models of tumor growth. Mathematical Models and Methods in Applied Sciences. 2016; 26(12):2341–2368. https://doi.org/10.1142/S021820251650055X, doi: 10.1142/S021820251650055X.

Lima EABF, Faghihi D, Philley R, Yang J, Virostko J, Phillips CM, Yankeelov TE. Bayesian calibration of a stochastic, multiscale agent-based model for predicting in vitro tumor growth. PLoS Computational Biology. 2021; 17(11). doi: 10.1371/journal.pcbi.1008845.

Lipková J, Angelikopoulos P, Wu S, Alberts E, Wiestler B, Diehl C, Preibisch C, Pyka T, Combs SE, Hadjidoukas P, Van Leemput K, Koumoutsakos P, Lowengrub J, Menze B. Personalized Radiotherapy Design for Glioblastoma: Integrating Mathematical Tumor Models, Multimodal Scans, and Bayesian Inference. IEEE Transactions on Medical Imaging. 2019 aug; 38(8):1875–1884. doi: 10.1109/TMI.2019.2902044.

Liu H, Lu T, Kremers GJ, Seynhaeve ALB, ten Hagen TLM. A microcarrier-based spheroid 3D invasion assay to monitor dynamic cell movement in extracellular matrix. Biological Procedures Online. 2020; 22(1):3. https://doi.org/10.1186/s12575-019-0114-0, doi: 10.1186/s12575-019-0114-0.

Loessner D, Flegg JA, Byrne HM, Clements JA, Hutmacher DW. Growth of confined cancer spheroids: a combined experimental and mathematical modelling approach. Integrative Biology. 2013; 5(3):597–605. https://doi.org/10.1039/c3ib20252f, doi: 10.1039/c3ib20252f.

Loizides C, Iacovides D, Hadjiandreou MM, Rizki G, Achilleos A, Strati K, Mitsis GD. Model-Based Tumor Growth Dynamics and Therapy Response in a Mouse Model of De Novo Carcinogenesis. PLOS ONE. 2015 dec; 10(12):e0143840. https://doi.org/10.1371/journal.pone.0143840.

Luisier F, Vonesch C, Blu T, Unser M. Fast interscale wavelet denoising of Poisson-corrupted images. Signal Processing. 2010; 90(2):415–427. http://www.sciencedirect.com/science/article/pii/S0165168409003016, doi: https://doi.org/10.1016/j.sigpro.2009.07.009.

MATLAB. 9.7.0.1190202 (R2019b). Natick, Massachusetts: The MathWorks Inc.; 2018.

Norton L. Conceptual and Practical Implications of Breast Tissue Geometry: Toward a More Effective, Less Toxic Therapy. The Oncologist. 2005; 10(6):370–381. http://theoncologist.alphamedpress.org/content/10/6/370.abstract, doi: 10.1634/theoncologist.10-6-370.

NVIDIA, CUDA C++ Programming Guide; 2021. https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html{#}abstract.

R Core Team, R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. https://www.r-project.org/.

Rejniak KA, Anderson ARA. Hybrid models of tumor growth. Wiley Interdisciplinary Reviews: Systems Biology and Medicine. 2011; 3(1):115–125. https://onlinelibrary.wiley.com/doi/abs/10.1002/wsbm.102, doi: 10.1002/wsbm.102.

Roussos ET, Balsamo M, Alford SK, Wyckoff JB, Gligorijevic B, Wang Y, Pozzuto M, Stobezki R, Goswami S, Segall JE, Lauffenburger DA, Bresnick AR, Gertler FB, Condeelis JS. Mena invasive (MenaINV) promotes multicellular streaming motility and transendothelial migration in a mouse model of breast cancer. Journal of Cell Science. 2011; 124(13):2120–2131. https://jcs.biologists.org/content/124/13/2120, doi: 10.1242/jcs.086231.

Roussos ET, Condeelis JS, Patsialou A. Chemotaxis in cancer. Nature Reviews Cancer. 2011; 11(8):573–587. https://doi.org/10.1038/nrc3078, doi: 10.1038/nrc3078.

Sanga S, Frieboes HB, Zheng X, Gatenby R, Bearer EL, Cristini V. Predictive oncology: A review of multidisciplinary, multiscale in silico modeling linking phenotype, morphology and growth. NeuroImage. 2007; 37:S120–S134. https://www.sciencedirect.com/science/article/pii/S1053811907004946, doi: https://doi.org/10.1016/j.neuroimage.2007.05.043.

Stein AM, Demuth T, Mobley D, Berens M, Sander LM. A Mathematical Model of Glioblastoma Tumor Spheroid Invasion in a Three-Dimensional In Vitro Experiment. Biophysical Journal. 2007; 92(1):356–365. http://www.sciencedirect.com/science/article/pii/S0006349507708330, doi: https://doi.org/10.1529/biophysj.106.093468.

Sternberg SR. Biomedical image processing. Computer. 1983; (1):22–34.

Strang G. On the construction and comparison of difference schemes. SIAM journal on numerical analysis. 1968; 5(3):506–517.

Susanto O, Koh YWH, Morrice N, Tumanov S, Thomason PA, Nielson M, Tweedy L, Muinonen-Martin AJ, Kamphorst JJ, Mackay GM, Insall RH. LPP3 mediates self-generation of chemotactic LPA gradients by melanoma cells. Journal of Cell Science. 2017; 130(20):3455–3466. https://doi.org/10.1242/jcs.207514, doi: 10.1242/jcs.207514.

Tan DSP, Potts HWW, Leong ACK, Gillett CE, Skilton D, Harris WH, Liebmann RD, Hanby AM. The biological and prognostic significance of cell polarity and E-cadherin in grade I infiltrating ductal carcinoma of the breast. The Journal of Pathology. 1999; 189(1):20–27. https://doi.org/10.1002/(SICI)1096-9896(199909)189:1{%}3C20::AID-PATH394{%}3E3.0.COhttp://2-2, doi: 10.1002/(SICI)1096-9896(199909)189:1<20::AID-PATH394>3.0.CO;2-2.

Tunc B, Hormuth D, Biros G, Yankeelov TE. Modeling of Glioma Growth with Mass Effect by Longitudinal Magnetic Resonance Imaging. IEEE Transactions on Biomedical Engineering. 2021; .

Tweedy L, Insall RH. Self-Generated Gradients Yield Exceptionally Robust Steering Cues. Frontiers in Cell and Developmental Biology. 2020; 8:133. https://www.frontiersin.org/article/10.3389/fcell.2020.00133, doi: 10.3389/fcell.2020.00133.

Tweedy L, Knecht DA, Mackay GM, Insall RH. Self-Generated Chemoattractant Gradients: Attractant Depletion Extends the Range and Robustness of Chemotaxis. PLOS Biology. 2016 mar; 14(3):e1002404. https://doi.org/10.1371/journal.pbio.1002404.

Tweedy L, Thomason PA, Paschke PI, Martin K, Machesky LM, Zagnoni M, Insall RH. Seeing around corners: Cells solve mazes and respond at a distance using attractant breakdown. Science. 2020; 369(6507). https://science.sciencemag.org/content/369/6507/eaay9792, doi: 10.1126/science.aay9792.

Warne DJ, Baker RE, Simpson MJ. Using Experimental Data and Information Criteria to Guide Model Selection for Reaction–Diffusion Problems in Mathematical Biology. Bulletin of Mathematical Biology. 2019 jun; 81(6):1760–1804. https://doi.org/10.1007/s11538-019-00589-x, doi: 10.1007/s11538-019-00589-x.

Appendix 1

Numerical methods

ADI Douglas-Gunn (DG) for the diffusion operator

The ADI-DG scheme ***Douglas and Gunn*** (***1964***) is a multi-step method and can be applied to the diffusion term as follows

$$\left(1-\frac{1}{2}\upsilon\delta_x^2\right)u^{n,*} = (1+\frac{1}{2}\upsilon\delta_x^2+\upsilon\delta_y^2+\upsilon\delta_z^2)u^n \tag{12}$$

$$\left(1-\frac{1}{2}\upsilon\delta_y^2\right)u^{n,**} = u^{n,*}-\frac{1}{2}\upsilon\delta_y^2 u^n \tag{13}$$

$$\left(1-\frac{1}{2}\upsilon\delta_z^2\right)u^{n+1} = u^{n,**}-\frac{1}{2}\upsilon\delta_z^2 u^n \tag{14}$$

where, $\upsilon = \frac{D_u \mathrm{d}t}{2h^2}$, h is the spatial grid step, assuming $h = \mathrm{d}x = \mathrm{d}y = \mathrm{d}z$, $\delta^2_{x,\ y,\ z}$ are the central difference operators for the second derivatives in $x,\ y,\ z$ respectively, and $u^{n,*},\ u^{n,**}$ are the intermediate values of u.

Explicit Lax-Wendrof (LxW) with Monotonic Upstream-Centered Scheme for Conservation Laws (MUSCL) flux limiter for the advection operator

The explicit LxW-MUSCL ***Lax and Wendroff*** (***1960***); ***van Leer*** (***1979***) method applied to the advection term can be written as follows

$$u^{n+1}_{i,j,k} = u^n_{i,j,k} + \chi\frac{\mathrm{d}t}{h}(F_{i-1/2}-F_{i+1/2}+F_{j-1/2}-F_{j+1/2}+F_{k-1/2}-F_{k+1/2}) \tag{15}$$

Here, $F_{i\pm1/2}$ are defined as follows

$$F_{i-1/2} = (u\nabla f)_{i-1} + \phi_-\frac{1}{2}\mathrm{sign}((\nabla f)_i)(1-c)[u_i(\nabla f)_i - u_{i-1}(\nabla f)_{i-1}] \tag{16}$$

$$F_{i+1/2} = (u\nabla f)_i + \phi_+\frac{1}{2}\mathrm{sign}((\nabla f)_i)(1-c)[u_{i+1}(\nabla f)_{i+1} - u_i(\nabla f)_i] \tag{17}$$

where $c = \chi\frac{\mathrm{d}t}{h}$,

$$(u\nabla f)_i = u_i\max(0,(\nabla f)_i) - u_{i+1}\max(0,-(\nabla f)_{i+1}) \tag{18}$$

$$(u\nabla f)_{i-1} = u_{i-1}\max(0,(\nabla f)_{i-1}) - u_i\max(0,-(\nabla f)_i) \tag{19}$$

and

$$(\nabla f)_i = \frac{f_i - f_{i-1}}{h},\quad (\nabla f)_{i-1} = \frac{f_{i-1}-f_{i-2}}{h},\quad (\nabla f)_{i+1} = \frac{f_{i+1}-f_i}{h} \tag{20}$$

The $\phi_\pm$ are the flux limiter variables and are defined as follows

$$\phi_\pm = \phi(r_{i\pm1/2}) = \max(0,\min(2r_{i\pm1/2},\frac{1}{2}(r_{i\pm1/2}+1),2)) \tag{21}$$

where,

$$r_{i-1/2} = \frac{u_I - u_{I-1}}{u_i - u_{i-1}},\ r_{i+1/2} = \frac{u_{I+1}-u_I}{u_{i+1}-u_i} \tag{22}$$

and $I = i - \mathrm{sign}((\nabla f)_i)$. The same procedure is repeated for $F_{j\pm1/2}$ and $F_{k\pm1/2}$. The LxW-MUSCL method is conditionally stable. Hence, the size of the time-step should satisfy the Courant-Friedrichs-Lewy (CFL) condition, $\mathrm{d}t \le \frac{1}{3}\frac{h}{\chi\max((\nabla f)_{i,j,k})}$.

Strang splitting

The Strang splitting technique ***Strang*** (***1968***) to preserve second-order accuracy in time is the following; For each time-step:

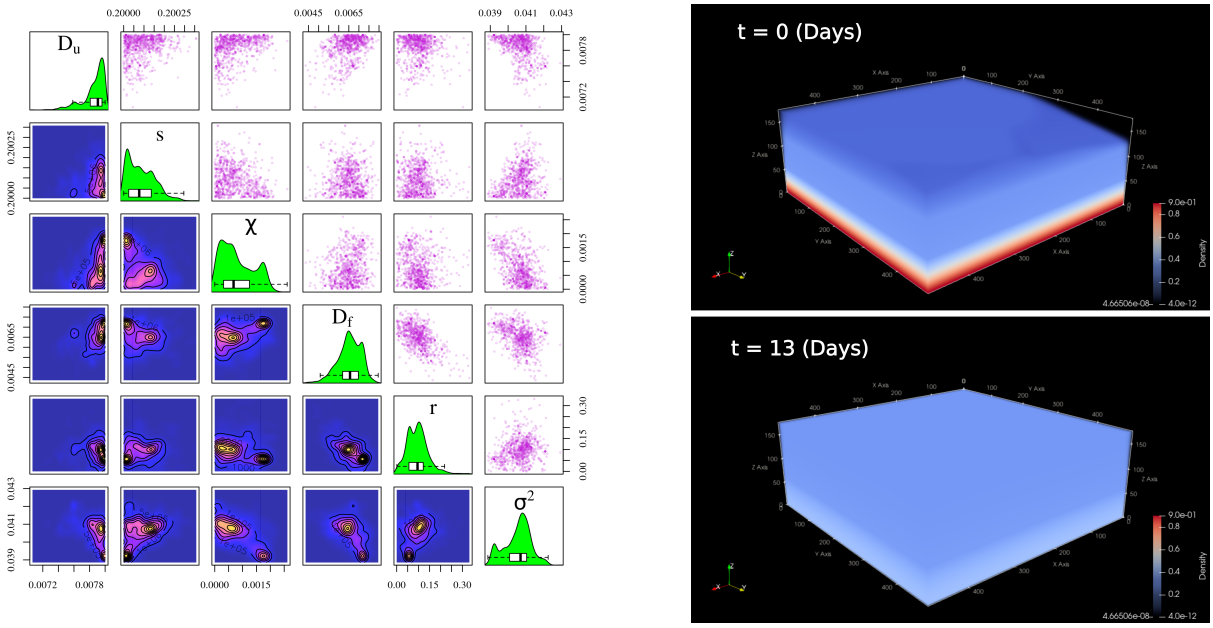
1. evolve the explicit term (advection) for a time-step $\mathrm{d}t/2$,
2. evolve the implicit term (diffusion) for a time-step $\mathrm{d}t$,
3. evolve the explicit term (advection) for a time-step $\mathrm{d}t/2$.

Since we use two numerical methods; an implicit and an explicit, we can use two different time steps. The LxW-MUSCL method is conditionally stable and we choose a time-step based on the CFL condition $\mathrm{d}t \leq \frac{h}{\chi \max((\nabla f)_{i,j,k})}$. The time-step is adaptive and has to be re-evaluated in every time-step, because it depends on $\max((\nabla f)_{i,j,k})$, which is also also dynamic. On the other hand, the ADI-DG is unconditionally stable, hence a larger time-step can provide sufficiently smooth solutions. We selected the time-step for the ADI-DG as $h^2/\max(D_u, D_f)$, a value that would violate the Neumann stability condition in the case of an explicit central differences scheme. The two different time-steps require a modification of the Strang splitting to the following form; For each implicit time-step:

1. evolve the explicit term for j steps until $\sum_j \mathrm{d}t_{\text{exp}} = \mathrm{d}t_{\text{imp}}/2$,
2. evolve the implicit term for $\mathrm{d}t_{\text{imp}}$,
3. evolve the explicit term for j steps until $\sum_j \mathrm{d}t_{\text{exp}} = \mathrm{d}t_{\text{imp}}/2$.

where $\mathrm{d}t_{\text{exp}}$ and $\mathrm{d}t_{\text{imp}}$ are the time-steps for the LxW-MUSCL and ADI-DG methods, respectively.

The implementation of the numerical schemes was performed in CUDA/C language ***NVIDIA*** (***2021***). More information on implementation can be found on the documentation of the code available at https://nmdimitriou.github.io/HyMetaGrowth/.



(a) **(b)**

Appendix 2 Figure 1. Model Calibration (a) Model parameter inference for a representative dataset using the TMCMC method. Above diagonal: Projected TMCMC samples of the posterior distribution in 2D space. Diagonal: Marginals of the joint posterior obtained via kernel densities. Boxplot whiskers demarcate the 95% percentiles. Below diagonal: 2D projected densities of the posterior obtained using 2D kernel densities. (b) Changes in the gradient of chemotactic signals over time. The gradient became less steep, suggesting that chemotactic migration diminished with respect to time.

Appendix 2

Continuum model calibration

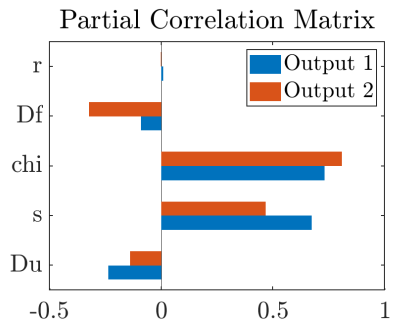
The obtained manifold of the inferred PDFs of the model parameters for one dataset using the TMCMC method is presented in Appendix 2 Fig. 1a. In Appendix 2 Fig. 1b we observe the changes of the chemotactic signalling gradient at two different. The gradient became less steep with respect to time, thus making chemotactic migration less pronounced at later time-points, compared to earlier time-points. The average and standard deviation of the inferred model parameters across the 12 datasets are presented in Appendix 2 Table 1.

Appendix 2 Table 1. Average and standard deviation of the inferred parameter values of the continuum KS model across all datasets.

Dataset	$D_u \in [10^{-3}, 8] \times 10^{-3}$ $(mm^2 d^{-1})$	$s \in [2,5] \times 10^{-1} (d^{-1})$	$\chi \in [0,5] \times 10^{-2} (mm^2 d^{-1})$	$D_f \in [10^{-3}, 8] \times 10^{-3}$ $(mm^2 d^{-1})$	$r \in [0, 0.5] \times 10^{-1} (d^{-1})$
1	7.55 ± 0.28	2.001 ± 0.001	0.050 ± 0.031	2.36 ± 0.69	1.96 ± 0.95
2	7.86 ± 0.15	2.001 ± 0.001	0.081 ± 0.058	6.54 ± 0.57	0.96 ± 0.49
3	7.61 ± 0.27	2.001 ± 0.001	0.096 ± 0.071	2.36 ± 1.22	2.88 ± 1.31
4	7.65 ± 0.23	2.001 ± 0.001	0.046 ± 0.042	2.00 ± 0.73	4.06 ± 1.02
5	7.50 ± 0.40	2.002 ± 0.001	0.300 ± 0.144	4.31 ± 0.84	6.76 ± 1.43
6	5.63 ± 0.09	2.005 ± 0.000	0.051 ± 0.001	1.47 ± 0.24	5.89 ± 0.14
7	7.75 ± 0.17	2.002 ± 0.002	0.276 ± 0.124	6.34 ± 1.40	3.12 ± 2.45
8	7.63 ± 0.10	2.008 ± 0.001	0.214 ± 0.038	2.91 ± 0.31	4.37 ± 0.40
9	7.68 ± 0.26	2.002 ± 0.002	0.040 ± 0.032	2.97 ± 1.03	2.76 ± 1.36
10	7.42 ± 0.45	2.004 ± 0.003	0.044 ± 0.049	2.86 ± 1.20	7.25 ± 2.13
11	7.63 ± 0.31	2.004 ± 0.002	0.066 ± 0.045	2.09 ± 1.20	4.63 ± 1.32
12	7.72 ± 0.21	2.002 ± 0.001	0.052 ± 0.045	2.07 ± 0.76	5.02 ± 1.25

Appendix 3

Global sensitivity analysis



Appendix 3 Figure 1. Global sensitivity analysis of the model parameters with respect to the tumour volume with density values greater than 10^{-3} (Output 1), and area of tumour found at the bottom xy−plane with density values greater than 10^{-3} (Output 2). The partial correlation matrix shows that the chemotactic signal production rate, r, was less correlated to the output. Hence it exhibited less contribution compared to the rest of the parameters.

The results obtained from the calibration of the continuum spatiotemporal model against the experimental data using TMCMC algorithm indicated increased variability of the chemotactic signal production rate, r, across the samples. This may indicate decreased sensitivity of the parameter r to the final output. To examine this hypothesis, we performed global sensitivity analysis for the model parameters with respect to the model output. The sensitivity analysis was performed using the Latin Hypercube Sampling (LHS) method ***Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Ebeida, M.S., Eddy, J.P., Eldred, M.S., Frye, J.R., Geraci, G., Hooper, R.W., Hough, P.D., Hu, K.T., Jakeman, J.D., Khalil, M., Maupin, K.A., Monschke, J.A., Ridgway, E.M., Rushdi, A.A., Stephens, J.A., Swiler, L*** (***2014***). The model parameters (input) were examined with respect to the tumour volume with density values greater than 10^{-3} (Output 1), as well as the area of tumour found at the bottom of the space with density values greater than 10^{-3} (Output 2). The sampled parameter space remained the same as the one used for the calibration. In total, 512 samples were produced, and the relation of the model response to the input was calculated using the Partial Rank Correlation Matrix between input and outputs 1 and 2. The results depicted in Apendix 3 Fig. 1 denote that the chemotactic signal production rate, r was less correlated to both outputs, making r less sensitive compared to the rest of the parameters. Additionally, we observed that the advective constant, χ, and the tumour growth rate, s, were positively correlated to both outputs, hence they contributed to tumour clustering. In contrast, the diffusion constants D_f, D_u were negatively correlated with both outputs, and they contributed to the dispersion of the tumour. Both observations are consistent with respect to the function of the model.

Appendix 4

Cosine similarity test

The similarity between *in-vitro* and *in-silico* IN Distance Distributions was estimated using the cosine similarity measure ***Han et al.*** (***2012***). The cosine similarity measure emerges from the Euclidean dot product, whereby the similarity of two given vectors $\vec{a}$, $\vec{b}$, is defined as $\text{sim}(\vec{a}, \vec{b}) := \cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{||\vec{x}||_2 ||\vec{y}||_2}$, where $||\cdot||_2$ the Euclidean norm. The similarity measure can take values between -1 and 1 indicating exactly opposite and identical vectors respectively. A similarity value of zero denotes orthogonal vectors. Increasing values between 0 and 1 denote low, intermediate and high similarity.