Bayesian calibration of a stochastic, multiscale agent-based model for predicting *in vitro* tumor growth

Ernesto A. B. F. Lima[1,2☯], Danial Faghihi[3☯], Russell Philley[1], Jianchen Yang[4], John Virostko[5,6], Caleb M. Phillips[1], Thomas E. Yankeelov[1,4-8*]

1 Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, USA
2 Texas Advanced Computing Center, The University of Texas at Austin, Austin, TX, USA
3 Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY, USA
4 Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX, USA
5 Department of Diagnostic Medicine, The University of Texas at Austin, Austin, TX, USA
6 Department of Oncology, The University of Texas at Austin, Austin, TX, USA
7 Livestrong Cancer Institutes, Dell Medical School, The University of Texas at Austin, Austin, TX, USA
8 Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

☯These authors contributed equally to this work.
¤The Center of Computational Oncology, The University of Texas at Austin, Austin, TX, USA
* thomas.yankeelov@utexas.edu

Abstract

Hybrid multiscale agent-based models (ABMs) are unique in their ability to simulate individual cell interactions and microenvironmental dynamics. Unfortunately, the high computational cost of modeling individual cells, the inherent stochasticity due to probabilistic phenotypic transitions, and numerous model parameters that are difficult to measure directly are fundamental limitations of applying such models to predict tumor dynamics. To overcome these challenges, we have developed a coarse-grained two-scale ABM (cgABM) calibrated with a set of time-resolved microscopy measurements of cancer cells grown with different initial conditions. The multiscale model consists of a reaction-diffusion type model capturing the spatio-temporal evolution of glucose and growth factors in the tumor microenvironment (at tissue scale), coupled with a lattice-free ABM to simulate individual cell dynamics (at cellular scale). The experimental data consists of BT474 human breast carcinoma cells initialized with different glucose concentrations and tumor cell confluences. The confluence of live and dead cells was measured every three hours over four days. Given this model and data, we perform a global sensitivity analysis to identify the relative importance of the model parameters. The subsequent cgABM with a reduced parameter space is calibrated within a Bayesian framework to the experimental data to estimate model parameters, which are then used to predict the temporal evolution of the living and dead cell populations. To this end, a moment-based Bayesian inference is proposed to account for

the stochasticity of the cgABM while quantifying uncertainties in model parameters and observational data. The results indicate that the cgABM can reliably predict the spatiotemporal evolution of breast cancer cells observed by the microscopy data with an average error and standard deviation for live and dead cells being 7.61 ± 2.01 and 5.78 ± 1.13, respectively.

Author summary

The calibration of agent-based models of tumor growth to experimental data remains a challenge in computational oncology. Besides the computational cost of modeling thousands of agents, the model's intrinsic stochasticity demands numerous realizations of the simulations to accurately represent the statistical features of the model predictions. We developed a hybrid, multiscale, coarse-grain, agent-based model that captures the growth and decline of human breast carcinoma cells under different initial conditions. We determined the effects of coarse-graining the ABM on the multiscale model output, and the number of repetitions necessary to capture the stochastic transitions present in the model. We identified the most influential parameters on the model prediction through a sensitivity analysis and selected which parameters can be fixed and which ones should be calibrated. Using Bayesian calibration, we show that the model can accurately represent the experimental data. The validation step indicates that our model can reliably predict the *in vitro* data, depending on the choice of the training (calibration data) sets.

Introduction

Tumor growth and treatment response are governed by the complex interplay of numerous phenomena occurring at various spatial and temporal scales. Hybrid multiscale models of tumor development consist of discrete models of individual cell interactions and their phenotypic transitions coupled to continuum models of microenvironmental evolution. These models allow investigation of the complex mechanisms of tumor initiation and growth at the interface of cellular, microenvironmental, and tissue scales events. More specifically, hybrid multiscale agent-based models (ABMs) consist of a continuum model capturing the spatio-temporal evolution of nutrients and growth factors in the tumor microenvironment. At the cellular scale, agent-based models simulate individual cell division and growth, cell-cell and cell-microenvironment interactions, and phenotypic switches that follow a user-defined set of probabilistic rules. The discrete agent-based and continuum models are coupled such that cellular dynamics influence the continuum model through nutrient consumption, while the concentration of nutrient impacts the decision-making process of the individual agents in the cell-scale model [1–16]. The primary benefit of combining multiple models is the ability to simulate coupled, multiscale processes and mechanisms responsible for tumor growth and treatment response [17]. This provides an opportunity to computationally test a range of hypotheses on the underlying biological phenomena driving cancer development. In this regard, hybrid ABMs have become powerful computational tools to study the complex and multiscale processes of tumor development including, for example, proliferation [14, 18], migration [11, 19], invasion [3, 5], angiogenesis [20, 21] along with mechanical [3, 22] and biochemical [11, 18] cues. For example, Macklin *et al.* [1, 18] used a patient-calibrated agent-based cell of ductal carcinoma to estimate biophysical parameters that are challenging to observe experimentally, such as time duration of apoptosis and cell calcification. Additionally, Rocha *et al.* [3] developed a hybrid three scale model consisting of a reaction-diffusion

type continuum model of the tumor microenvironment (tissue scale), a lattice-free ABM of cell dynamics (cellular scale), and an inter- and intracellular signaling pathways model represented by a system of coupled nonlinear differential equations (sub-cellular scale). The model describes the major biological feature of avascular tumor dynamics, such as the proliferative, hypoxic, and necrotic regions. For a comprehensive review of discrete and hybrid tumor growth models and their applications, the interested reader is referred to [23], and the references cited therein.

In spite of these advances in agent-based models, they still possess fundamental limitations that restrict their ability to accurately predict the spatiotemporal evolution of a tumor given practical, experimental scenarios – a major goal of this developing field [24, 25]. In particular, the computational cost increases rapidly with the number of simulated agents since the evolution of the system relies upon the interactions of the individual cells with both each other and the surrounding milieu. This makes ABMs extraordinarily challenging to simulate large biological systems on practical time and length scales. Another challenge arises when attempting to calibrate hybrid ABMs to experimental data as the models typically require measurements that span the micro- to macroscopic scales. Obtaining such measurements may be cost-prohibitive or not technically feasible. Furthermore, the intrinsic stochasticity of ABM models (due to the probabilistic decision criteria describing, for example, phenotypic transitions) adds further computational complications [26, 27]. Thus, standard parameter estimation techniques cannot adequately characterize the errors in ABM parameter calibrations [28], and new approaches are needed. Despite these barriers, there have been several previous efforts attempting to calibrate ABMs using experimental data. For example, Jiang *et al.* [5] showed that an ABM successfully captures the growth of mouse mammary tumor spheroids observed from in vitro measurements. Macklin *et al.* [18] calibrated a hybrid ABM using x-ray mammographic measurements of ductal carcinoma *in situ* and demonstrated the ability of the model to depict tumor heterogeneity. However, there are pronounced uncertainties in the parameters estimated from calibrating ABMs to experimental data, which translates into uncertainties in model predictions.

A common approach to reducing the high computational cost associated with hybrid ABMs is representing clusters of cells within the biological system by individual agents [22, 29–31], rather than simulating each cell with its own agent. This approach is known as "coarse-graining" in particle simulations of chemical or physical processes whereby several physical particles are lumped into a single simulation agent (or bead) to substantially reduce the degrees of freedom (see, e.g., [32]). Additionally, improving the predictive utility of hybrid ABMs in cancer requires the integration of data and models through a systematic model calibration and validation scheme that rigorously handles uncertainties in data and parameter calibration [17, 33, 34] as well as parameter inference methods that cope with the inherent stochasticity of the model [35, 36].

In this contribution, we aim to overcome these challenges by developing a coarse-grained, two-scale ABM (cgABM) that can be calibrated with *in vitro*, time-resolved microscopy data of human breast carcinoma cells growing from a range of initial confluences and nutrient levels. A variance-based global sensitivity analysis is conducted to identify the relative importance of the parameters within the cgABM. We then perform a Bayesian calibration of the cgABM to the microscopy data, which allows for quantification of the uncertainties in both the model parameters and the observational data. The calibration is performed *via* a moment-based Bayesian inference that generalizes the likelihood function to account for the stochasticity of the cgABM in the inverse problem. Additionally, the Bayesian inferences are implemented using parallel codes with efficient use of high-performance computing resources that enable conducting the computationally expensive Bayesian calibration of the cgABM. Finally,

the validity of our approach is assessed by predicting a set of measurements outside the calibration data.

Materials and Methods

Hybrid two-scales agent-based model

The model we develop is a modification of a hybrid cell-tissue ABM we previously introduced [3] which links the tissue, cell, and sub-cell scales. Briefly, at the tissue level, the dispersion of nutrients and growth factors in the tumor microenvironment is modeled through reaction-diffusion equations. The ABM characterizes the cell level by describing normal and tumor cell dynamics, with cancer cells differentiated into proliferative, apoptotic, hypoxic, and necrotic states. Finally, the sub-cell scale integrates the epidermal growth factor receptor (EGFR) pathway as modeled by a system of coupled nonlinear differential equations. As our primary interest here is to calibrate this model to the time-resolved microscopy measurements of a growing tumor mass, we neglect the sub-cellular signaling pathway model, reduce the possible cell phenotypical states, and course-grain the discrete model so that each agent represents multiple cells with the same phenotypes. Additionally, to preserve the fidelity of the hybrid cgABM in depicting the observational measurements, the rule-based decisions of phenotypical transitions are enhanced compared to [3]. The details of the discrete cellular scale, the continuous tissue scale, and their coupling are summarized in the next subsections.

Discrete cellular scale model

The interactions between cells are captured by a discrete, lattice-free ABM in which the agents (i.e., a single tumor cell or a cluster of cells) are free to move throughout the domain unrestrained by a grid. As done in [3,18,20,37], we represent the cell as a circle, and track the cell radius over time. The geometry of the i^{th} cell, at time t and position $\boldsymbol{x}_i$, is defined by its radius R_i, with an incompressible nucleus of radius R_i^N. We also define an action radius R_i^A (with $R_i^N < R_i < R_i^A$) to specify short-range interactions capturing cell-cell adhesion and repulsion. With these definitions, cell movement is determined by the following three mechanisms [3,18]:

1. The cell-cell adhesive force, $\boldsymbol{F}_{cca}$, and cell-cell repulsive force, $\boldsymbol{F}_{ccr}$, between the i^{th} and j^{th} cells are defined as,

$$\boldsymbol{F}_{cca}^{ij} = -c_{cca}\boldsymbol{\nabla}\varphi\left(\boldsymbol{l}^{ij}; R_i^A + R_j^A\right), \tag{1}$$

$$\boldsymbol{F}_{ccr}^{ij} = -c_{ccr}\boldsymbol{\nabla}\psi\left(\boldsymbol{l}^{ij}; R_i^N + R_j^N, R_i + R_j\right), \tag{2}$$

where $\boldsymbol{l}^{ij} = \boldsymbol{x}_j - \boldsymbol{x}_i$ is the distance between the center of the i^{th} and j^{th} cells, and c_{cca} and c_{ccr} are the cell-cell adhesion and repulsion scale parameters, respectively. The interaction potentials for adhesion, φ, and repulsion, ψ, are given by,

$$\boldsymbol{\nabla}\varphi(\boldsymbol{l}, R^A) = \begin{cases} \left(\frac{|\boldsymbol{l}|}{R^A} - 1\right)^2 \frac{\boldsymbol{l}}{|\boldsymbol{l}|}, & 0 \le |\boldsymbol{l}| \le R^A; \\ \boldsymbol{0}, & \text{otherwise}; \end{cases} \tag{3}$$

$$\boldsymbol{\nabla}\psi(\boldsymbol{l}, R^N, R) = \begin{cases} -\left(\frac{R^N|\boldsymbol{l}|}{R^2} - \frac{2|\boldsymbol{l}|}{R} + 1\right)\frac{\boldsymbol{l}}{|\boldsymbol{l}|}, & 0 \le |\boldsymbol{l}| \le R^N; \\ -\left(\frac{|\boldsymbol{l}|^2}{R^2} - \frac{2|\boldsymbol{l}|}{R} + 1\right)\frac{\boldsymbol{l}}{|\boldsymbol{l}|}, & R^N \le |\boldsymbol{l}| \le R; \\ \boldsymbol{0}, & \text{otherwise}. \end{cases} \tag{4}$$

In Eq. (3), the effects of the adhesion between the cells begins when their action radius overlaps, with the adhesion intensity increasing as the distance between the cells decreases. This phenomenon is balanced by the repulsion force, Eq. (4), which begins to act when there is contact between two cells (i.e., the distance between the two cells is equal the sum of their radius). As we assume that cell nucleus is incompressible, the effects of the repulsion increases if the nucleus of the cells overlap.

2. The compression force, $\boldsymbol{F}_{ct}$, and the resistance to the compression force, $\boldsymbol{F}_{ccr}$, represent the effects of the boundary on the i^{th} cell as it grows and it is given by,

$$\boldsymbol{F}^i_{ct} = -c_{ct}\boldsymbol{\nabla}\varphi\left(\boldsymbol{l}^i; R^A_i\right), \tag{5}$$

$$\boldsymbol{F}^i_{rct} = -c_{rct}\boldsymbol{\nabla}\psi\left(\boldsymbol{l}^i; R^N_i, R_i\right), \tag{6}$$

where $\boldsymbol{l}^i$ is the distance between the i^{th} cell and the domain boundary, and c_{ct} and c_{rct} are the cell-boundary adhesion and repulsion scale parameters, respectively. The simulation domain represents the whole experimental well. Therefore, we consider a non-permeable incompressible boundary, such that the tumor cells cannot leave the domain.

3. Due to the low speed of interstitial flow, the linear drag force of interstitial fluid flow, $\boldsymbol{F}_{drag}$, is captured *via*,

$$\boldsymbol{F}^i_{drag} = -\nu\boldsymbol{v}_i, \tag{7}$$

where $\boldsymbol{v}_i$ is the velocity of the i^{th} cell, and the constant ν characterizes the fluid viscosity.

The balance of forces acting on the i^{th} cell of mass m_i is obtained by Newton's second law,

$$m_i\dot{\boldsymbol{v}}_i = \overbrace{\sum_{\substack{j=1\\ j\neq i}}^{N(t)}(\boldsymbol{F}^{ij}_{cca} + \boldsymbol{F}^{ij}_{ccr})}^{\text{cell-cell interaction}} + \underbrace{(\boldsymbol{F}^i_{drag} + \boldsymbol{F}^i_{ct} + \boldsymbol{F}^i_{rct})}_{\text{cell-microenvironment interaction}}, \tag{8}$$

where $N(t)$ is the total number of cells. Disregarding the inertial effects and substituting the drag force from Eq. (7) into Eq. (8) results in the velocity of the i^{th} cell as

$$\boldsymbol{v}_i = \frac{1}{\nu}\left(\sum_{\substack{j=1\\ j\neq i}}^{N(t)}(\mathbf{F}^{ij}_{cca} + \mathbf{F}^{ij}_{ccr}) + \mathbf{F}^i_{ct} + \mathbf{F}^i_{rct}\right). \tag{9}$$

Thus, the position of the cell at time t_{k+1} is given as:

$$\boldsymbol{x}_i(t_{k+1}) = \boldsymbol{x}_i(t_k) + \boldsymbol{v}_i\Delta t, \tag{10}$$

where $\Delta t = t_{k+1} - t_k$ indicates the time interval.

In our model, the possible cell phenotypes are quiescent, $\mathcal{Q}$, proliferative, $\mathcal{P}$, and dying cells, $\mathcal{D}$. Figure 1 provides a schematic illustration of rules for transitioning between these states. After a time $\tau_P - \tau_{G1}$, (i.e., the differences between the duration of the cell cycle, τ_P, and the duration of the growth phase, τ_{G1}) the cell undergoes mitosis in which two daughter cells, each with half the area of the parent cell, are created in a deterministic process. The daughter cells grow until they reach the area of the parent cell, and enter the quiescent state after time, τ_{G1}.

The transitions from the quiescent, $\mathcal{Q}$, to the proliferative, $\mathcal{P}$, and death, $\mathcal{D}$, phenotypes are stochastic processes [3, 18, 37] and governed by the following probabilities:

$$P(\mathcal{D}|\mathcal{Q}) = 1 - \exp(\alpha_D(\sigma)\Delta t), \tag{11}$$

$$P(\mathcal{P}|\mathcal{Q}) = 1 - \exp(\alpha_P(\sigma)\Delta t). \tag{12}$$

In Eqs. (11) and (12), the intensity factors α_D and α_P are functions of the nutrient concentration σ:

$$\alpha_D(\sigma) = \bar{\alpha}_D + \gamma_D \frac{1}{1 + \exp(-2k(\sigma_H - \sigma))}, \tag{13}$$

$$\alpha_P(\sigma) = \max\left(\bar{\alpha}_P \frac{\sigma - \sigma_H}{1 - \sigma_H}, 0\right), \tag{14}$$

where $\bar{\alpha}_D$ is the apoptosis rate, γ_D controls the increase in cell death due to the lack of nutrients, σ_H is the hypoxic threshold, and $\bar{\alpha}_P$ is the proliferation rate. The second term on the right-hand side in Eq. (13) is a smooth approximation to the step function, such that a larger value of k leads to a sharper transition at $\sigma = \sigma_H$. This term is incorporated to capture the effects of hypoxia, while its parameters can be well informed by the *in vitro* measurements of dead cell confluence. Eq. (14) simulates the regulatory effect of the nutrient on cell proliferation up to the threshold σ_H, below which the cell does not have enough nutrient to undergo mitosis.

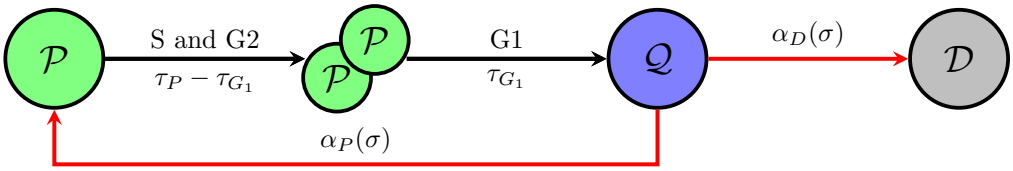

Fig 1. Schematic illustration of phenotypic transitions. The transitions between cell phenotypes can be deterministic (black arrows) or stochastic (red arrows). Proliferative ($\mathcal{P}$) tumor cells split into two cells with half of the original area after time interval $\tau_P - \tau_{G1}$, where τ_P is the duration of the cell cycle, and τ_{G1} is the duration of the growth phase G1. This time interval includes the S and G2 phases of cell cycle. The new cells enter the growth phase, reaching the same area as the original cell, and then return to the quiescent ($\mathcal{Q}$) state. The transitions from quiescent cells are controlled by the intensity functions α_P and α_D, which govern the transition to proliferative and dead ($\mathcal{D}$) cells, respectively. These transitions probabilities are proportional to the nutrient concentration, σ.

Continuum tissue scale model

The evolution of nutrient is modeled at the tissue scale in which the nutrient is assumed diffuse in the microenvironment and taken up by living cells. Nutrient concentration, σ, is governed by a mass conservation condition based on reaction-diffusion equations. Let Ω be the domain with a smooth boundary $\delta\Omega$ in which the development of living cells takes place due to the presence of nutrient, the governing equations at the macroscale is [3, 18, 37]:

$$\frac{\partial \sigma}{\partial t} = \boldsymbol{\nabla} \cdot (D \boldsymbol{\nabla} \sigma) - \Lambda(\boldsymbol{x}, t)\sigma, \qquad \text{in } \Omega, \tag{15}$$

$$\boldsymbol{n} \cdot \boldsymbol{\nabla} \sigma = 0, \qquad \text{on } \partial\Omega, \tag{16}$$

where D is the nutrient diffusion, $\boldsymbol{n}$ is a unit normal vector to $\partial\Omega$, and $\Lambda(\boldsymbol{x}, t)$ is a function describing the nutrient uptake rate. This function couples the discrete model with the continuum model (the cellular and tissue scales, respectively), and it is given as

$$\Lambda(\boldsymbol{x}, t) = \lambda\rho_t, \tag{17}$$

where λ is the nutrient consumption rate by living cells (i.e., quiescent plus proliferative cells), and ρ_t is the volume fraction occupied by live cells. The nutrient uptake rate function averages the microscale (cellular) events.

Numerical solution of the hybrid ABM

The ABM is implemented in C++ using an object-oriented approach to model each cell in an object that stores the position ($\boldsymbol{x}$), velocity ($\boldsymbol{v}$), and the forces acting on each cell. Additionally, the nucleus, cytoplasm, and the action radii are all saved to track each cell's growth. The continuum reaction-diffusion model is solved using the C++ finite-element library libMesh [38]. The code itself, as a well as a description of how to use it, is provided at `https://github.com/eabflima/avascular_abm`. In summary, the solution steps of the hybrid ABM are:

1. Initialize the nutrient field uniformly according to the value used in the *in vitro* experimental condition (we normalized the initial condition by the highest value used experimentally).
2. Initialize the discrete model by seeding the tumor agents positions randomly in the domain such as the model confluences match the experimental live and dead cells tumor confluences.
3. Solve the reaction-diffusion equation and update the microenviroment conditions.
4. Update the phenotypic states based on the nutrient concentration.
5. Compute the balance of the forces acting on each cell, and update cell positions.
6. Return to step 3 and solve the model until the desired length of simulation is reached.

Variance-based global sensitivity analysis

Global sensitivity analysis allows understanding and quantifying the impact of model parameters on the variation of the model outputs [39–41]. In time-dependent biological processes, such as those simulated by multiscale ABMs, sensitivity analysis determines the relative importance of each model parameter on the system responses during its evolution. In this regard, conducting sensitivity analysis aids in understanding the biological mechanisms that govern the system behavior. At the same time, it can guide the model calibration process by refining the estimations of the most effective parameters.

We employ a variance-based global sensitivity analysis method (also known as the Sobol' Indices), in which the sensitivity of the model output to input parameters is computed by the quantity of (conditional) variance in the output caused by that specific input [42–44]. This method allows analyzing numerous model parameters simultaneously as well as being sufficiently general to handle complex multiscale problems. We now summarize the application of this method to our problem.

In a non-additive model, such as an ABM with K uncertain parameters (input factors) $\{\theta_k\}_{k=1}^{K}$, the model output $\mathbf{d}(\boldsymbol{\theta}) = \mathbf{d}(\theta_1, \theta_2, \cdots, \theta_K)$ depends on the

interactions among the parameters. The model output variance, $\mathbb{V}(\mathbf{d})$, can be decomposed by conditioning with respect to all inputs except θ_k [45, 46],

$$\mathbb{V}(\mathbf{d}) = \mathbb{V}_{\boldsymbol{\theta}_{\sim k}}\left(\mathbb{E}_{\theta_k}(\mathbf{d}|\boldsymbol{\theta}_{\sim k})\right) + \mathbb{E}_{\boldsymbol{\theta}_{\sim k}}\left(\mathbb{V}_{\theta_k}(\mathbf{d}|\boldsymbol{\theta}_{\sim k})\right), \tag{18}$$

where θ_k is the k-th input factor, where $\boldsymbol{\theta}_{\sim k}$ indicates the matrix of all factors except θ_k, and $\mathbb{E}_{\boldsymbol{\theta}_{\sim k}}(\cdot)$ and $\mathbb{V}_{\boldsymbol{\theta}_{\sim k}}(\cdot)$ represents the mean and variance, respectively, taken over all possible values of $\boldsymbol{\theta}_{\sim k}$ while θ_k is fixed. Eq. (18) results in a sensitivity measure; the so-called *total effect index* [45, 46],

$$S_{T_k} = 1 - \frac{\mathbb{V}_{\boldsymbol{\theta}_{\sim k}}\left(\mathbb{E}_{\theta_k}(\mathbf{d}|\boldsymbol{\theta}_{\sim k})\right)}{\mathbb{V}(\mathbf{d})}. \tag{19}$$

In Eq. (19), $\mathbb{V}_{\boldsymbol{\theta}_{\sim k}}\left(\mathbb{E}_{\theta_k}(\mathbf{d}|\boldsymbol{\theta}_{\sim k})\right)$ denotes the expected reduction in variance if all values other than θ_k are fixed. Thus, the total effect, S_{T_k}, measures the contribution of the input θ_k to the model output variation. A "small" total effect index indicates that fixing θ_k at any value within the range of its uncertainty will not affect the model output significantly.

Monte-Carlo estimation of total effect sensitivity index

To compute the sensitivity index, we employ an efficient sampling method and an estimator proposed by Saltelli [40, 45, 47, 48]. Estimating S_{T_k} using this method consists of constructing two $N \times K$ matrices, $\mathbf{A}$ and $\mathbf{B}$, in which N random samples are drawn from a uniform distribution corresponding to the range of each parameters' uncertainty. Additionally, K matrices $\mathbf{A}_{\mathbf{B}}^{(k)}$, $k = 1, 2, \cdots, K$, are defined where all columns are from $\mathbf{A}$ except the k^{th} column, which comes from $\mathbf{B}$. The model outputs are then evaluated for each row of the matrices $\mathbf{A}$ and $\mathbf{A}_{\mathbf{B}}^{(k)}$ and the outputs are stored in the vectors $\boldsymbol{Y}_{\boldsymbol{A}}$ and $\boldsymbol{Y}_{\boldsymbol{AB}}^{(k)}$. The total-effect index for each parameter, $\{S_{T_k}\}_{k=1}^{K}$, can be approximated using the following estimator [48],

$$S_{T_k} \approx \frac{1}{2N}\sum_{j=1}^{N}\left(\left(\boldsymbol{Y}_{\boldsymbol{A}}\right)_j - \left(\boldsymbol{Y}_{\boldsymbol{AB}}^{(k)}\right)_j\right)^2. \tag{20}$$

This algorithm reduces the computational cost of estimating multi-dimensional integrals to $N(K+1)$ model evaluations. For a time-dependent process, the above steps can be repeated for each time instance to represent the temporally-varying importance of each model parameter. We present the time evolution of S_{T_k} for multiple ABM outputs (e.g., live and dead tumor cells with different nutrient and confluence initial conditions) in the results section.

Model calibration under uncertainty

Predictive modeling of biophysical systems requires characterizing the uncertainties in both the model parameters (due to simplifying assumptions made to develop the model) and the experimental data (due to noise and variability in measurements), as well as the uncertainty in the Quantity of Interest (QoI; i.e., the target of the prediction). Bayesian approaches to problems of statistical inference provide general frameworks for identifying the essential features of a predictive model, while also providing means to characterize uncertainty. The main feature of these approaches is that the model parameters, $\boldsymbol{\theta}$, and the observational data, $\mathbf{D}$, are random variables represented by probability density functions (PDFs), $\pi(\boldsymbol{\theta})$ and $\pi(\mathbf{D})$, respectively. In this section, we first summarize a Bayesian calibration and validation process, and then discuss the form of the likelihood function for Bayesian inference of our stochastic ABM along with the numerical solution using a multi-level sampling algorithm.

Bayesian statistical inference

To represent the uncertainties in both the data and the model parameters, we make use of a statistical inference method in which the probability density functions (PDF) of the calibrated parameters are given by Bayes' formula [49]:

$$\pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{D}) = \frac{\pi_{\text{like}}(\mathbf{D}|\boldsymbol{\theta}) \cdot \pi_{\text{prior}}(\boldsymbol{\theta})}{\pi_{\text{evid}}(\mathbf{D})}. \tag{21}$$

In Eq. (21), $\pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{D})$ is the posterior PDF defining the Bayesian update of the prior information represented by $\pi_{\text{prior}}(\boldsymbol{\theta})$, $\pi_{\text{like}}(\mathbf{D}|\boldsymbol{\theta})$ is the likelihood PDF, and $\pi_{\text{evid}}(\mathbf{D})$ is the evidence seen as a normalization factor (since $\int \pi_{post} = 1$),

$$\pi_{\text{evid}}(\mathbf{D}) = \int \pi_{\text{like}}(\mathbf{D}|\boldsymbol{\theta}) \cdot \pi_{\text{prior}}(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}. \tag{22}$$

One can use the principle of maximum entropy to construct the prior of the model parameters [50] based on their known features (e.g., bounds, mean, and variance). In the case that only the parameters' bounds are available, then a uniform distribution is used as a prior $\boldsymbol{\theta}$. Finally, to explain the posterior PDF, $\pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{D})$, with a point estimate, one can use a Maximum A Posteriori (MAP) defined as,

$$\boldsymbol{\theta}^{\text{MAP}} = \underset{\boldsymbol{\theta}}{\text{argmax}} \; \pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{D}). \tag{23}$$

Likelihood function for stochastic forward models

The form of the likelihood function reflects the way the discrepancy between the model output and the data are modeled. To account for uncertainties in computational models (i.e., model inadequacy) and measurement data (i.e., data noise), likelihood functions can be constructed by assigning a probability distribution, p_ϵ, to the error representing the difference between the observational data, $\mathbf{D}$, and the model output, $\mathbf{d}$. The hybrid, agent-based model involves inherent randomness due to the stochastic processes defining the transition between the quiescent ($\mathcal{Q}$), proliferative ($\mathcal{P}$), and dead ($\mathcal{D}$) states. Thus, the same set of parameter values, boundary conditions, and initial conditions will result in an ensemble of different outputs. The model output for the case of the stochastic model is represented by $\mathbf{d}(\boldsymbol{\theta}, \omega)$, where $\omega \in \Omega$ with Ω the set of possible outcomes. Under the additive noise assumption (see, e.g., [51, 52]), the total error is described as $\boldsymbol{\epsilon} = \boldsymbol{\eta} + \boldsymbol{\xi} = \mathbf{D} - \mathbf{d}(\boldsymbol{\theta}, \omega)$, in which $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ indicate data noise and model inadequacy, respectively. Then the likelihood function is the probability density function describing the total error and is written as

$$\pi_{\text{like}}(\mathbf{D}|\boldsymbol{\theta}) = p_\epsilon(\mathbf{D} - \mathbf{d}(\boldsymbol{\theta}, \omega)), \tag{24}$$

where p_ϵ is a probability distribution. Here we assume that the error in the data and the model are Gaussian random variables with zero mean,

$$\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{data}}^{-1}), \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{model}}^{-1}), \tag{25}$$

where $\boldsymbol{\Gamma}_{\text{data}}$ and $\boldsymbol{\Gamma}_{\text{model}}$ are the covariance matrices.

To develop the likelihood for stochastic models, at each time step, $i = 1, \cdots, N_t$, we denote a data point $\mathbf{D}_i^{(j)}$ as a sample from a distribution, $\mathbf{D}_i^{(j)} \sim p(\mathbf{D})$ with $j = 1, \cdots, N_D$. Similarly, to represent the randomness in the stochastic model, we consider $\mathbf{d}_i(\boldsymbol{\theta}, \omega_j) = \mathbf{d}_i^{(j)}$ as $j = 1, \cdots, N_r$ independent identically distributed realizations of the model output at $\{\omega_j\}_{j=1}^{N_r}$ as samples from a distribution $\mathbf{d}_i^{(j)} \sim p(\mathbf{d}|\boldsymbol{\theta})$. To represent the distance between the observational data $\mathbf{D}$ and the

model output $\mathbf{d}$ in the likelihood function, we take the first moments of $p(\mathbf{D})$ and $p(\mathbf{d}|\boldsymbol{\theta})$. The sample estimates of the mean of the data and model are, respectively,

$$\mu_i^d(\boldsymbol{\theta}) = \frac{1}{N_r}\sum_{j=1}^{N_r}\mathbf{d}_i^{(j)}(\boldsymbol{\theta}), \tag{26}$$

$$\mu_i^D = \frac{1}{N_D}\sum_{j=1}^{N_D}\mathbf{D}_i^{(j)}. \tag{27}$$

We note that using sample-based estimates of the statistical moments with a finite number of model evaluations introduces statistical uncertainties in computing the means. While one can account for such uncertainties by approximating the variances of the moment estimators using methods of moment-based inference [53, 54], we consider sufficiently large N_r to minimize the statistical error as shown in the Results section. The assumptions in Eq. (25) result in p_ϵ being a normal distribution $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}}^{-1})$, where $\boldsymbol{\Gamma}_{\text{noise}}$ is the covariance matrix representing the data noise and model error. Assuming $\boldsymbol{\Gamma}_{\text{data}} = (\sigma_i^D)^2\mathbf{I}$ and $\boldsymbol{\Gamma}_{\text{model}} = (\sigma_i^d(\boldsymbol{\theta}))^2\mathbf{I}$ in Eq. (25), one can write

$$\boldsymbol{\Gamma}_{\text{noise}} = (\sigma_i)^2\mathbf{I}, \quad (\sigma_i)^2 = (\sigma_i^D)^2 + (\sigma_i^d(\boldsymbol{\theta}))^2, \quad i = 1, 2, \cdots, N_t. \tag{28}$$

Following the above considerations, the proposed likelihood function for the stochastic ABM model can be written explicitly as

$$\ln(\pi_{\text{like}}(\mathbf{D}|\boldsymbol{\theta})) = \sum_{i=1}^{N_t}\left(\frac{1}{2}\ln(2\pi) - \ln(\sigma_i) - \frac{1}{2}\left(\frac{\mu_i^d(\boldsymbol{\theta}) - \mu_i^D}{\sigma_i}\right)^2\right). \tag{29}$$

Sampling method for Bayesian Inference

To conduct Bayesian inference, one needs to compute the posterior density $\pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{D})$ as the solution of the statistical inverse problem. Typically, Markov Chain Monte Carlo (MCMC) sampling methods are employed to characterize the posterior distribution as they guarantee asymptotically exact recovery of the posterior distribution as the number of samples increases (see, e.g., [55, 56]). The solution of the Bayesian problem is computationally expensive as the posterior distribution may be a complex object requiring a large number of model evaluations. For the Bayesian calibration and validation of the ABM, we make use of a parallel, adaptive, multilevel MCMC algorithm [57]. In this work, we employ the adaptive multilevel MCMC implemented in the C++ library QUESO (Quantification of Uncertainty for Estimation, Simulation, and Optimization) [58] and refer the interested reader to [57] for details on the computational implementation of this method.

Error metric and model validation

To access the quality of the calibrated model in matching experimental data, we propose a metric using the cumulative probability distribution functions in $L^1(\mathbb{R})$. If ϕ_α indicates cell confluence with $\alpha = L$ or D (i.e., live or dead confluences, respectively), then set $F_t(\phi_\alpha)$ and $S_t(\phi_\alpha)$ to be the cumulative distribution functions for the model output using $\pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{D})$ and the measured data at hour t, respectively. Then, the metric is given as

$$\mathbb{M}_t(F_t, S_t) = \frac{\int_{-\infty}^{\infty}|F_t(\phi_\alpha) - S_t(\phi_\alpha)|\,\mathrm{d}\phi_\alpha}{\Phi}, \tag{30}$$

where Φ is the mean total confluence from the data at time t. Eq. (30) can be considered as the relative error; however, here we take into account the uncertainties of

data and model prediction. Such error measure can be used to check whether a computational model is valid for predicting the quantities of interest. To this end, one must specify a tolerance (accepted) error for the model prediction ε_{tol}. If the prediction error is below such tolerance, say,

$$\frac{1}{N_t}\sum_{i=1}^{N_t}\mathbb{M}_t \leq \varepsilon_{\text{tol}}, \tag{31}$$

then the model is deemed to be valid and can be used for making predictions in scenarios well characterized by the calibration/validation data.

Experimental measurements

Cell lines and cell culture

BT474 human breast carcinoma cells were obtained from American Type Culture Collection (ATCC, Manassas, VA). The cells were grown in Dulbecco's modified eagle medium (DMEM, Thermo Fisher Scientific Inc., Waltham, MA) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich, St. Louis, MO) 1% L-glutamine (Thermo Fisher Scientific Inc.), and 1% Penicillin Streptomycin (Thermo Fisher Scientific Inc.) in 5% CO_2 and air at 37°C.

D-(+)-glucose solution (Sigma-Aldrich) was added to glucose free DMEM to yield media with glucose concentrations of 2, 5, and 10 mM. For each glucose level there were three initial seeding densities, each with four replicates. Cells were seeded at a density of 3.5×10^4, 5.0×10^4 and 6.0×10^4 cells/well on a 96-well tissue culture plate. We note that, in the present study, all references to "nutrient" in the model development presented above refer to glucose.

Time-resolved microscopy

Cells were incubated in the IncuCyte live-cell imaging system (Essen BioScience, USA). Multiple images were acquired with a 4× objective and automatically stitched together to obtain a whole well image for each well. The IncuCyte Cytotox Red reagent (Essen BioScience, USA), a highly sensitive cyanine nucleic acid dye, was added into the medium to quantify cell death. Once a cell's plasma membrane begins to lose integrity, the cytotox red enters the cell and yields a 100-1000-fold increase in fluorescence upon binding to deoxyribonucleic acid (DNA). Phase-contrast images and fluorescent images (Red channel, excitation wavelength: 585 nm and emission wavelength: 635 nm) were acquired every 3 hours.

Image segmentation to quantify confluence over time

The BT-474 cells within the phase-contrast images at each time point were segmented in Matlab (The Mathworks, Inc., Natick, MA). The first step was to define a mask corresponding to the size of a well in a 96-well-plate from the IncuCyte Software (Essen BioScience, Ann Arbor, MI). The mask was applied such that the region of interest only included the area within each well. The masked region was converted to grayscale and with the Matlab function "colfilt", we calculated the standard deviation of signal intensities within each 3-by-3 sliding block of the image to detect the edge of cell clusters. Following that, a Gaussian filter was applied to smooth the image which was then normalized to yield signal intensities between 0 and 1 (by dividing the value in each pixel by the highest signal intensity from each image). Next, with the Matlab function "imerode", we shrank the clusters size and enlarged the holes to avoid losing open space within clusters. The image was then binarized *via* the function "im2bw",

while the functions "imclose" and "imopen" were used to fill holes in the interior of cell clusters and to smooth object contours, respectively. Finally, small objects were removed from the image *via* the function "bwareaopen".

Results

In all simulations, a circular domain of radius 3192 μm is used, which corresponds to the *in vitro* experimental domain. The computational domain, for the continuum model, is discretized by 2413 triangular elements with no flux permitted through the boundary. This boundary condition mimics the tumor and glucose being contained by the walls of the culture plate's well.

Time-resolved microscopy data

Fig 2 displays a series of images showing tumor cell confluence over time in which the cells were seeded at a density of 5.0×10^4 cells/well with either 2 mM (row A) or 10 mM glucose (row B). It is observed that as time progresses, the tumor cells in row A rapidly consume the glucose, yielding an environment that becomes somewhat unfavorable for continued expansion as manifested by the increase in dead cells (red) decays after day 2. Conversely, the cells seeded within 10 mM of glucose (row B) are able to continue to expand with minimal cell death. (See figure caption for more details.)

Developing a coarse-grained ABM

The effects of the coarse-graining on the ABM simulations are restrained by the amount of relative error between the ABM and cgABM. Fig 3 shows the mean relative error and the 95% credible interval of live (panel A) and dead (panel B) cells confluence for different degrees of coarsening compared to the single-cell per agent. The simulations are conducted for 5 mM glucose concentration with the initial confluences of dead and live cells of 0.5 and 0.3, respectively. The values of the parameters used in these numerical experiments are defined in Table 1, and the tolerance for the desired mean relative absolute error is set at 5%. The highest degree of coarsening that satisfies this tolerance, both for live and dead cells, is the 100 cells/agent. It is readily apparent that increasing the degree of coarse-graining results in a higher error and, more importantly, higher variance in the cgABM simulations. Based on these results, all subsequent analyses are presented with a coarse-graining of 100 cells/agent.

Global sensitivity analysis

Variance-based global sensitivity analysis is conducted on the cgABM system to determine how each parameter contributes to the model outputs (i.e., the live and dead cell confluences). We perform the analysis for three different initial tumor confluences with the densities of 3.5×10^4 (low), 5.0×10^4 (medium), and 6.0×10^4 (high) cells/well, as well as three initial glucose concentrations (2, 5, and 10 mM), and for seven parameters ($K = 7$) that we posit control the cell population ($\bar{\alpha}_P$, $\bar{\alpha}_D$, D, λ, γ_D, k, and σ_H). The other model parameters (i.e., R, R_N, R_A, c_{ccr}, c_{cca}, τ_P, τ_{G1}, and τ_A) are kept constant to the values reported in [18] (see Table 1). Due to the inherent stochasticity of the cgABM, the sensitivity analysis is performed on the sample estimate of the mean of the model output; i.e., substituting $\mathbf{d}$ in Eq. (19) with μ^d from Eq. (26), leading to a computational cost of $N_t = N_r N(K + 1)$. For estimating the total-effect index for each parameter in Eq. (20), $N = 1000$ samples are drawn from the uniform distributions given in Table 1 with $N_r = 16$ model realizations per sample; therefore, the number of

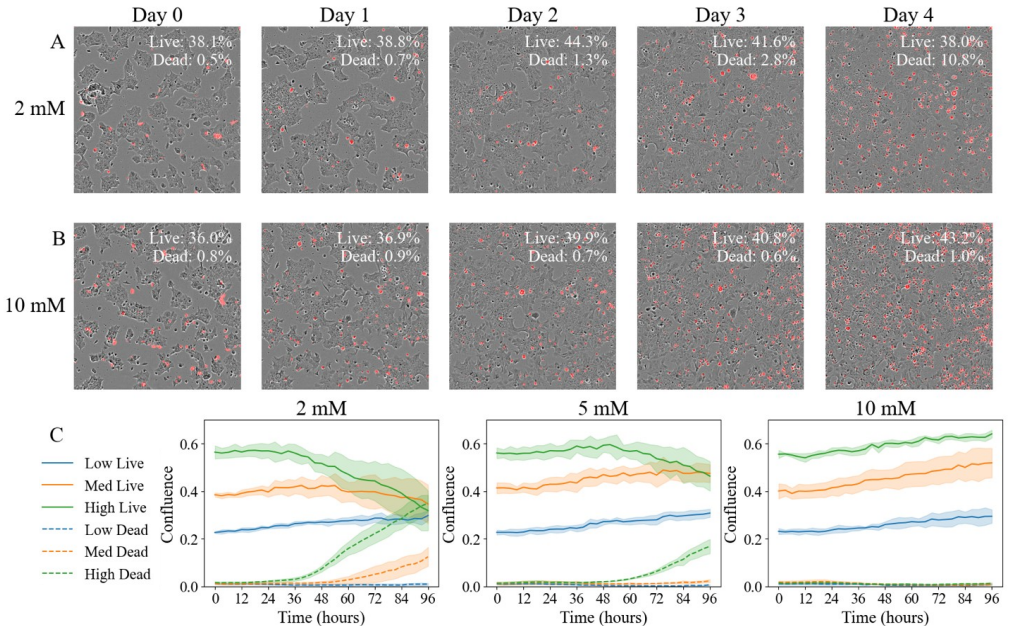

Fig 2. Time-resolved microscopy data. Row A presents example images from one well on days 0, 1, 2, 3, and 4 post seeding. While the whole-well image consists of 2240×2240 pixels, here we present a window of 400×400 pixels within the same area. These are merged images of phase contrast images and fluorescent images where dead cells were labeled in red by the Cytotox Red. Percent confluence of live and dead cells within the whole well are shown in the upper right corner of each panel. In this well, the cells were seeded at a density of 5.0×10^4 cells/well, supplied with culture medium containing 2 mM glucose. With these initial conditions, the available nutrient allowed the tumor cells to increase in confluence until day 2, after which the environment cannot sustain growth, resulting in cell death. Row B presents example images from another well on days 0, 1, 2, 3, and 4 post seeding, where cells were seeded at the same density, supplied with culture medium containing 10 mM glucose. With these initial conditions, the available nutrient allowed the tumor cells to increase in confluence in 4 days. Each panel in row C presents confluences as a function of time (sampled every three hours) with the initial glucose level shown in the subtitle. In each panel, the average confluence of live cells with low, medium, and high seeding density are shown in blue, green, and orange solid lines, respectively, with the 95% confidence interval shown with shaded regions. The average confluence of dead cells with low, medium, and high seeding density are shown in blue, green, and orange dashed lines, respectively, with the 95% confidence interval shown with shaded regions.

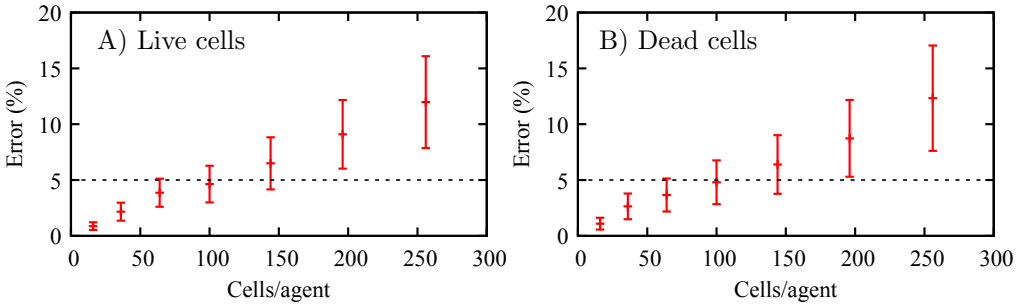

Fig 3. Effects of the coarse-graining on the ABM. Mean relative absolute error between the ABM and the cgABM, and its 95% credible interval (which is the Bayesian equivalent of the confidence interval), of live (A) and dead (B) cells. The simulations are executed for 16, 36, 64, 100, 144, 196, and 256 cells/agent. The horizontal dashed line indicates our 5% tolerance for the mean relative absolute error indicating that 100 cells/agent is the appropriate scenario for all subsequent analyses. The mean error and 95% credible interval between the coarse-grained model and the ABM, for the 100 cells/agent scenario, are $4.64 \pm 1.64\%$ and $4.80 \pm 1.96\%$ for live and dead cells, respectively.

Table 1. Model parameters.

Parameter	Physical meaning	Deterministic values coarse-graining studies	Distributions/Priors sensitivity analyses/calibration	How assigned
$\bar{\alpha}_P$	$\mathcal{Q} \to \mathcal{P}$ transition rate	$0.0493\ h^{-1}$	$\mathcal{U}(0,1)\ h^{-1}$	Calibrated
$\bar{\alpha}_D$	$\mathcal{Q} \to \mathcal{D}$ transition rate	$0.000408\ h^{-1}$	$\mathcal{U}(0,0.02)\ h^{-1}$	Calibrated
D	glucose diffusion coefficient	$50\ \mu m^2/h$	$\mathcal{U}(0,100)\ \mu m^2/h$	Calibrated
λ	glucose uptake rate	$0.0483\ h^{-1}$	$\mathcal{U}(0,1)\ h^{-1}$	Calibrated
γ_D	death rate increase	$0.0245\ h^{-1}$	$\mathcal{U}(0,0.05)\ h^{-1}$	Calibrated
k	smooth transition constant	50	$\mathcal{U}(0,100)$	Calibrated
σ_H	glucose threshold	0.0538	$\mathcal{U}(0,1)$	Calibrated
R	cell radius	$9.953\ \mu m$		Ref. [18]
R_N	cell nuclear radius	$5.295\ \mu m$		Ref. [18]
R_A	action radius	$1.214R$		Ref. [18]
c_{ccr}	cell-cell repulsion coefficient	$10\ \mu m/min$		Ref. [18]
c_{cca}	cell-cell adhesion coefficient	$0.0489\ \mu m/min$		Ref. [18]
τ_P	cell cycle time	$18\ h$		Ref. [18]
τ_{G1}	G_1 cell cycle phase time	$9\ h$		Ref. [18]
τ_A	apoptosis time	$8.6\ h$		Ref. [18]

Deterministic values of parameters used in the coarse-graining studies, parameter distributions for the sensitivity analyses, and priors for the Bayesian model calibration. The $\mathcal{U}(\cdot,\cdot)$ indicates the uniform probability distribution.

model evaluation required to obtain the total effect index (Eq. (20)) is $N_t = 128000$ (for each initial condition). Fig 4 shows the total effect index for each parameter over time, for live (panels A - C) and dead (panels D - F) cell confluences for the cases initially seeded with medium confluence (i.e., 5.0×10^4 cells). The simulation starts with every live tumor cell being in the quiescent state ($\mathcal{Q}$); after 24 hours, a heterogeneous tumor cell population has developed. The results of the sensitivity analyses in this figure indicate that the parameter γ_D (red line), which controls the rate of cell death due to lack of glucose, is the most critical parameter affecting the accumulation of dead cells during the simulation. Note that this parameter is also central to determining the time course of living cell confluence when the concentration of glucose is low. However, it is

the nutrient threshold, $_H$ (black line), that is the most important parameter for the live cell confluence when the initial nutrient concentration is 5 mM. The proliferation rate ($_P$, purple line) is the most influential parameter for the temporal development of live cells up to 36 hours, from 36 to 66 hours it is the glucose uptake rate (, blue line), while it is $_D$ (red line) that is the most influential parameter at longer times (for the 10 mM initial glucose concentration). In Fig 4, the total effect indices of parameters$k$ (orange line) and D (teal line) are consistently close to zero and never exceed 0.2. These results indicate that k and D do not have a significant effect on live and dead cell confluences during the 96 hours simulated. Thus, these parameters are assumed to be constant in the statistical calibrations described in the next section. The sensitivity analysis for low and high initial confluences are shown in Supplemental Figures S1 Fig and S2 Fig, respectively, and corroborate the conclusions obtained for the medium initial confluence. From Figs 4, S1 Fig, and S2 Fig, we conclude that the total effect indices of the parameters do not change for dead cells when the initial tumor confluence changes. However, for the live cells, these indices are affected by the initial tumor confluence. As the initial confluence increases, the importance of the nutrient uptake ( ) decreases, and the death rate due to lack of nutrients ($_D$) increases. As an example, the total effect index of reaches 0.6 for the live cells with 10 mM glucose and low initial confluence, and it drops to 0.2 as the initial confluence increases to high (i.e., $6{:}0 \quad 10^4$ cells/well). The opposite behavior is observed for the $_D$ total effect index, increasing from 0.2 to 0.6 as the initial confluence increases.

Fig 4. Medium confluence sensitivity analysis. Sensitivity analysis of proliferation rate ($_P$), death rate ($_D$), glucose diffusion (D), glucose uptake ( ), death rate increase due to lack of glucose ($_D$), smooth transition constant (k), and glucose threshold ($_H$) for live (top row) and dead (bottom row) cell phenotypes seeded with medium confluence. Panels A-F show the total effect index over time with Panels A, B, and C depicting live tumor cells, while Panels D, E, and F depict the dead tumor cells. The importance of the parameters is studied for three initial glucose concentrations: 2 mM (Panels A and D), 5 mM (Panels B and E), and 10 mM (Panels C and F). The glucose diffusion and the smooth transition constant have limited influence on the quantities of interest during the complete simulation (i.e., large changes in these parameters would yield small changes in tumor composition). Apart from these two parameters, the total effect index for every parameter is greater than 0.2 during the 96 hours simulated.

Scenario-specific calibrations

Guided by the results from the sensitivity analysis, we proceed with the Bayesian calibration of the cgABM, using the *in vitro* experimental data. In particular, we will calibrate the proliferation rate, death rate, nutrient uptake rate, death rate increase, and nutrient threshold, $\boldsymbol{\theta} = (\bar{\alpha}_P, \bar{\alpha}_D, \lambda, \gamma_D, \sigma_H)$, while holding the remaining parameters constant. The priors for the parameters to be calibrated and the fixed values for the constant parameters are defined in Table 1. The uniform priors were selected such that the bounds on the distribution include the ranges in [3, 18, 20], and live and dead cell confluences were able to reach 1.0 (i.e., the whole well covered by live and/or dead cells) within 96 hours. As indicated above, the calibration data D consists of the time evolution of live and dead cell confluences for nine scenarios, comprised of three different initial glucose concentrations (2 mM, 5 mM, and 10 mM) and three different initial tumor confluences with the densities of 3.5×10^4 (low), 5.0×10^4 (medium), and 6.0×10^4 (high) cells/well. As shown in Table 2, we utilize an abbreviation to refer to each scenario. For example, 2-M represent the scenario with 2 mM of initial glucose concentration and medium initial tumor cell confluence.

In this section, we present the "scenario-specific calibration" process, in which the cgABM is calibrated for each of the nine measurement scenarios individually, resulting in nine sets of calibration posteriors of the model parameters (see S2 Appendix for the computational details of the Bayesian implementation). Such a calibration process is performed to ensure the developed cgABM can simulate the experimentally observed responses of cell evolutions and to investigate the effect of the initial glucose concentrations and tumor confluences on the estimated parameters.

Figures 5, 6, and 7 compare the calibrated cgABM results with the *in vitro* data for 2 mM, 5 mM, and 10 mM initial glucose concentrations, respectively. The error bars show the data uncertainty due to the four replicates of the experimental measurements. The prediction results in this figure are obtained from 200 parameter samples drawn from the calibration posteriors and computing the mean of the cgABM simulations (solid line) and 95% credible interval (the green and red areas for live and dead cells respectively). Thus, the uncertainty in the model predictions is due to the stochasticity of the cgABM as well as the parameter uncertainty. Table 2 presents the error between the model outputs and the experimental measurements according to Eq. (30), and includes means and the 95% credible interval. Due to insufficient glucose for the higher tumor cell numbers, the scenarios 2-M, 2-H, and 5-H exhibit an increase of dead cells over time. As expected, the confluence of dead cells does not significantly increase with time in the scenarios with higher glucose concentration and lower initial tumor cell confluence. The scenario-specific calibration errors shown in Table 2 indicate that the model is able to capture the evolution of live and dead cells with an average error below 7%. Moreover, the model can successfully predict the increase of dead cells when glucose is consumed, as in scenarios 2-M, 2-H, and 5-H. The maximum discrepancy between data and model occurs in the 2-H scenario and the smallest mean error for the case of 5-H.

Figures 8, 9, and 10 present the kernel density estimation of the Bayesian calibration posteriors for all nine scenarios. Each figure summarizes scenarios with the same initial glucose concentration with low, medium, and high initial tumor confluences. For quantitative comparisons of the calibrated parameters in each scenario, we estimate the Maximum A Posteriori (MAP) points (see Eq. (23)) for the MCMC samples of the posterior. We next summarize observations from the posteriors of the parameters in the scenario-specific calibration process: the proliferation rate ($\bar{\alpha}_P$ - Panel A), death rate ($\bar{\alpha}_D$ - Panel B), glucose uptake rate (λ - Panel C), death rate increase (γ_D - Panel D), and glucose threshold (σ_H - Panel E).

As the initial glucose concentrations increase from 2 mM to 10 mM, the MAP of the proliferation rate ($\bar{\alpha}_P$) decreases. In particular, from 2 mM to 5 mM the average MAP

Table 2. Mean error and standard deviation.

Tumor initial condition	Glucose	Scenario name	Scenario-specific				Multi-scenario				Leave-one-out (Prediction)			
			Error live		Error dead		Error live		Error dead		Error live		Error dead	
High	10 mM	10-H	1.79	0.24	0.67	0.08	11.10	1.68	6.69	1.50	6.22	1.34	5.53	1.60
High	5 mM	5-H	4.44	0.68	1.73	0.77	6.32	1.12	2.46	0.47	7.28	1.89	4.74	2.27
High	2 mM	2-H	6.43	1.01	3.45	1.15	6.00	1.02	5.93	1.70	13.97	3.69	15.80	5.59
Medium	10 mM	10-M	5.90	0.43	1.71	0.15	8.14	0.72	6.45	1.37	13.93	2.77	7.32	1.55
Medium	5 mM	5-M	4.96	0.37	0.94	0.15	5.09	0.85	5.83	1.40	15.20	3.06	7.30	1.67
Medium	2 mM	2-M	6.97	1.24	2.11	0.64	10.29	1.83	5.09	0.84	20.56	2.85	12.86	1.74
Low	10 mM	10-L	3.68	0.52	1.84	0.20	10.01	1.26	6.44	1.25	7.61	2.01	5.78	1.13
Low	5 mM	5-L	3.41	0.33	1.98	0.33	7.28	1.75	7.27	1.38	9.08	2.00	4.99	0.93
Low	2 mM	2-L	2.18	0.31	2.10	0.18	18.04	3.47	9.37	2.10	21.49	4.27	9.56	2.39

Computed error for the scenario-specific and multi-scenario calibration, and for the leave-one-out prediction. The nomenclature of the scenarios refers to the glucose concentration (2 mM, 5 mM, and 10 mM) followed by the initial tumor confluence (low - L, medium - M, and high - H).

Fig 5. Calibration of scenarios with 2 mM initial glucose concentration.
Scenario-specific calibration of the cgABM to the time-resolved microscopy data for the 2 mM initial glucose concentration for scenarios 2-L (panel A), 2-M (panel B), and 2-H (panel C). The data mean and 95% credible interval for the live and dead cells are shown in blue and black, respectively. The mean of the simulation is represented as a solid line, and the area is the 95% credible interval for the live (green) and dead (red) cells. The low initial confluence in scenario 2-L assures that the initial glucose concentration can sustain tumor growth during the 96 hours.

of $_P$ decreases 68%, and from 5 mM to 10 mM it decreases 58%. This trend is consistent with the function describing the proliferation rate, which is a function of $_P$, $_H$ (Eq. (14)) and the current glucose concentration. The average MAP and standard deviation from 2 mM, 5 mM, and 10 mM are $6.74 \quad 10^{\ 2} \quad 1.13 \quad 10^{\ 2}\ h^{-1}$, $2.16 \quad 10^{\ 2} \quad 0.61 \quad 10^{\ 2}\ h^{-1}$, and $0.91 \quad 10^{\ 2} \quad 0.35 \quad 10^{\ 2}\ h^{-1}$, respectively. Note how the variance of $_P$ is smaller for the scenario with an initial glucose concentration of 10 mM (Fig 10). Once sufficient glucose is available for all proliferating cells (i.e., the

Fig 6. Calibration of scenarios with 5 mM initial glucose concentration.
Scenario-specific calibration of the cgABM to the time-resolved microscopy data for the 5 mM initial glucose concentration for scenarios 5-L (panel A), 5-M (panel B), and 5-H (panel C). The data mean and 95% credible interval for the live and dead cells are shown in blue and black, respectively. The mean of the simulation is represented as a solid line, and the area is the 95% credible interval for the live (green) and dead (red) cells. The high initial confluence in scenario 5-H is the only one for which the initial glucose concentration can not sustain tumor growth during the 96 hours, leading to an increase in the dead cell confluence.

10 mM case), the proliferation rate is relatively constant with changes in initial tumor cell confluence. Additionally, the largest variance is observed in the posterior of $_P$ for the 2-H scenario (Fig 8). The higher uncertainty in the estimate of $_P$ is due to the substantial live cells' rate of change of confluence observed in Fig 5, as well as the competition between cell death and growth.

As the confluence of the dead cells is evident only in three scenarios, 2-M, 2-H, and 5-H (Figs 5 and 6), the average MAP and standard deviation of the death rate, $_D$, is $2{:}32 \quad 10^{5} \quad 1{:}47 \quad 10^{5}$ h^{-1} when disregarding these scenarios. For scenarios 2-M, 2-H, and 5-H, the MAP point is $1{:}53 \quad 10^{4}$ h^{-1}, $6{:}52 \quad 10^{4}$ h^{-1}, and $5{:}07 \quad 10^{4}$ h^{-1}, respectively. The higher MAP values of $_D$ are due to the limited availability of glucose for these three cases, in which a greater number of proliferative cells cause a faster consumption of the glucose and subsequent increase of dead cells due to glucose depletion. Additionally, for the scenarios with 10 mM of initial glucose, $_D$ remains the same with respect to the initial tumor confluences shown in Fig 10. The constant values of $_D$ are because the tumor cells are in favorable glucose environments in scenarios 10-L, 10-M, and 10-H, and no increase in dead cells is observed (see Fig 7). Furthermore, similar to the $_P$, the 2-H scenario represents the largest variance in the $_D$ posterior, due to the distinct increase of dead cells in the observational data of Fig 5.

The glucose uptake rate ( ), death rate increase ($_D$), and glucose threshold ($_H$) parameter are strongly correlated due to their indirect relations to the glucose field. The parameter in the continuum model (see Eq. (17)) controls how fast the glucose concentration decreases. The parameter $_D$ controls the increase in cell death due to

Fig 7. Calibration of scenarios with 10 mM initial glucose concentration. Scenario-specific calibration of the cgABM to the time-resolved microscopy data for the 10 mM initial glucose concentration for scenarios 10-L (panel A), 10-M (panel B), and 10-H (panel C). The data mean and 95% credible interval for the live and dead cells are shown in blue and black, respectively. The mean of the simulation is represented as a solid line, and the area is the 95% credible interval for the live (green) and dead (red) cells. The 10 mM glucose condition is sufficient to sustain tumor growth during the 96 hours for all seeding density tested.

the effects of hypoxia. Finally, $_H$ is the hypoxic threshold below which cells do not have enough glucose to undergo mitosis (see Eq. (14)). In all scenarios, the glucose uptake rate is proportional to the initial tumor cell confluence, in which a higher number of tumor cells leads to an increase in glucose consumption rate. The average MAP value of decreases as the glucose concentration increases, with a 20% decrease in the average MAP from 2 mM to 5 mM, and 32% from 5 mM to 10 mM. Only in the scenarios 2-M, 2-H, and 5-H, do the glucose levels drop below the threshold $_H$, leading to activation of the cell death mechanism due to the effects of hypoxia (the second term on the right-hand side of Eq. (13)). This observation indicates that $_D$ does not affect the evolution of the tumor cells in the other six scenarios (i.e., 2-L, 5-L, 5-M, 10-L, 10-M, and 10-H).

Fig 11 shows snapshots of one realization of the cgABM simulation for the 2-H scenario (i.e., high initial tumor confluence and 2 mM of initial glucose concentration). For this simulation, we use the MAP estimates of the parameter posteriors from scenario 2-H (Fig 8), which are $^{MAP}_{P} = 5{:}50 \quad 10^{-3}\ h^{-1}$, $^{MAP}_{D} = 6{:}52 \quad 10^{-4}\ h^{-1}$, $^{MAP} = 6{:}37 \quad 10^{-3}\ h^{-1}$, $^{MAP}_{D} = 2{:}18 \quad 10^{-3}\ h^{-1}$, and $^{MAP}_{H} = 3{:}63 \quad 10^{-3}$. The cgABM simulation in Fig 11 begins with confluences of 0.56 and 0.02 for the quiescent and dead cells, respectively, and a uniform glucose distribution of 2 mM. With the available glucose, the tumor cells proliferate, leading to the presence of growing daughter cells on day 1. The combination of an initial low glucose concentration, and high tumor confluence, yield a rapid decrease of glucose below the level required to maintain cell viability. By day 2, dying cells appear in the cgABM simulation, in agreement with the increase in dying cells observed in the experimental data (Fig 5).

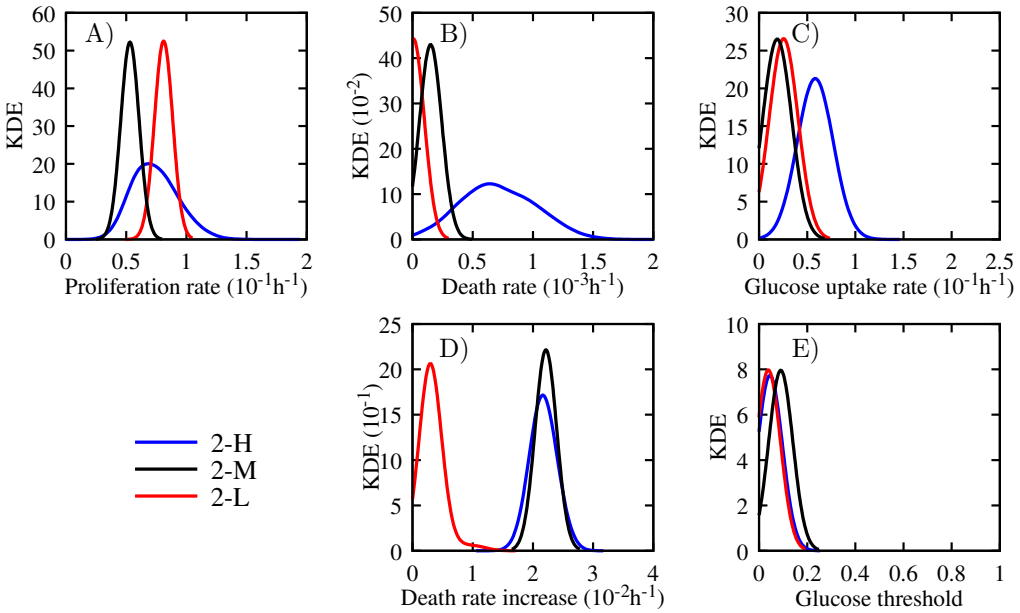

Fig 8. Posterior kernel density estimation (KDE) of scenarios with 2 mM initial glucose concentration. The KDE obtained during the scenario-specific calibration of the 2 mM initial glucose concentration and low (red), medium (black), and high (blue) tumor initial condition for the following parameters: $\bar{\alpha}_P$ (proliferation rate - panel A), $\bar{\alpha}_D$ (death rate - panel B), λ (glucose uptake rate - panel C), γ_D (death rate increase - panel D), and σ_H (glucose threshold - panel E). For this initial glucose concentration, the parameter γ_D does not influence the dynamics of the model and can not be calibrated in the 2-L scenario as the glucose concentration is enough to sustain tumor growth and avoid the increase of the confluence of dead cells.

Fig 12 shows snapshots of one realization of the cgABM simulation for the 10-L scenario (i.e., low initial tumor confluence and 10 mM of initial glucose concentration). For this simulation, we use the MAP estimates of the parameter posteriors from scenario 10-L (Fig 10), which are $\bar{\alpha}_P^{\text{MAP}} = 7.31 \times 10^{-3}\ h^{-1}$, $\bar{\alpha}_D^{\text{MAP}} = 1.73 \times 10^{-4}\ h^{-1}$, $\lambda^{\text{MAP}} = 3.44 \times 10^{-3}\ h^{-1}$, $\gamma_D^{\text{MAP}} = 8.45 \times 10^{-3}\ h^{-1}$, and $\sigma_H^{\text{MAP}} = 1.94 \times 10^{-1}$. The cgABM simulation in Fig 12 begins with confluences of 0.23 and 0.01 for the quiescent and dead cells, respectively, and a uniform glucose distribution of 10 mM. The final confluence for the quiescent and dead cells, for this realization of this stochastic simulation, is 0.28 and 0.01, respectively. With this low tumor cell confluence, the glucose never drops below σ_H, such that there is no increase in cell death due to the lack of glucose.

Multi-scenario calibration and prediction

In the previous section, we made use of the scenario-specific calibration to investigate the capability of the cgABM to capture the experimentally observed evolutions of live and dead cells. In this section, we study the ability of the calibrated model to *predict* the evolution of live and dead cells in a range of initial conditions. To this end, we implement two parameter identification strategies; namely, the "multi-scenario calibration" and "leave-one-out calibration". The purpose of multi-scenario calibration is to make use of Bayesian inference to determine one set of cgABM parameters that represent live and dead cell evolution of all scenarios with different initial tumor confluences and glucose concentrations. For the multi-scenario calibration, the reported

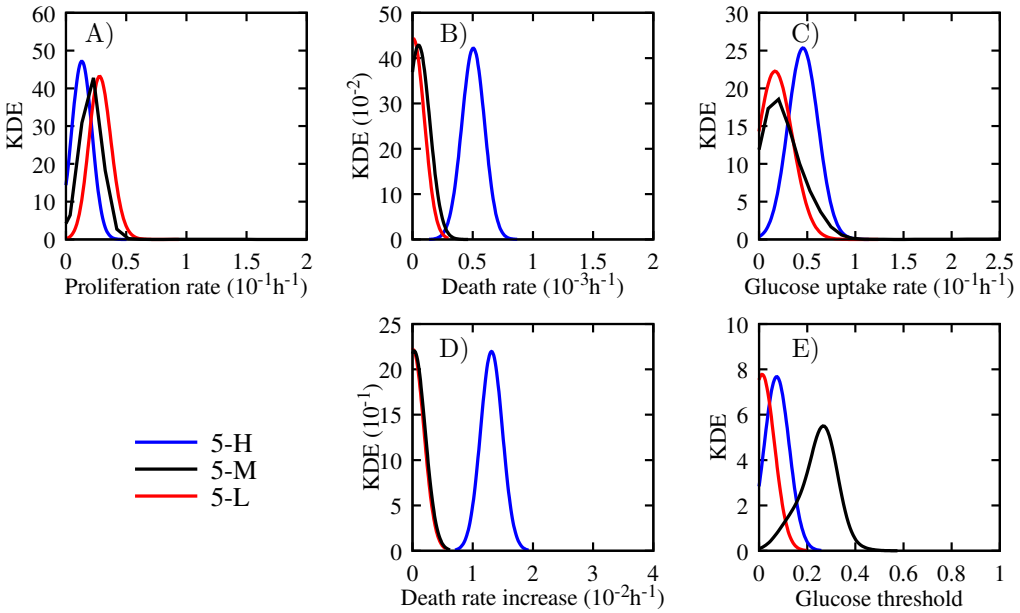

Fig 9. Posterior kernel density estimation (KDE) of scenarios with 5 mM initial glucose concentration. The KDE obtained during the scenario-specific calibration of the 5 mM initial glucose concentration and low (red), medium (black), and high (blue) tumor initial condition for the following parameters: $\bar{\alpha}_P$ (proliferation rate - panel A), $\bar{\alpha}_D$ (death rate - panel B), λ (glucose uptake rate - panel C), γ_D (death rate increase - panel D), and σ_H (glucose threshold - panel E). For this initial glucose concentration, the parameters $\bar{\alpha}_D$ and γ_D are different than zero just in the 5-H scenario. In this initial condition, the high initial confluence demands more glucose than it is available, leading to an increase of the dead cells confluence. This dynamic does not happen in scenarios 5-L and 5-M, where the initial glucose concentration is enough to sustain tumor growth.

error is the comparison among the cgABM model output and each experimental scenario. In the leave-one-out approach we calibrate the cgABM parameters using the data of eight scenarios, excluding one data set (prediction scenario) from the calibration process. We then compare the calibrated cgABM simulation with measured data in the prediction scenario to investigate the model's ability to forecast tumor responses in scenarios not included in the calibration data. Table 2 shows the error between the model output and the experimental measurements according to Eq. (30). For the leave-one-out, the reported means and the 95% credible intervals in Table 2 show the discrepancies between model and data in the prediction scenarios; i.e., the data sets left out of the calibration data.

Fig 13 compares the computational prediction of cgABM with the unseen experimental data for two representative cases of the leave-one-out calibration. In panel A, we illustrate predicting scenario 2-H when the cgABM is calibrated with the other eight scenarios. The 2-H (high initial tumor confluence and 2 mM initial glucose) scenario was not included in the calibration process, and is the one with the highest dead cell confluence (see Fig 5) among the nine scenarios. The prediction errors of the leave-one-out calibration in Table 2 demonstrate the influence of using observational data of the 2-H scenario in informing the model parameters. That is, by excluding 2-H from the calibration data, the cgABM is unable to precisely capture the dynamics of the dead cells. Thus, this case results in an invalid cgABM for predicting the *in vitro* tumor

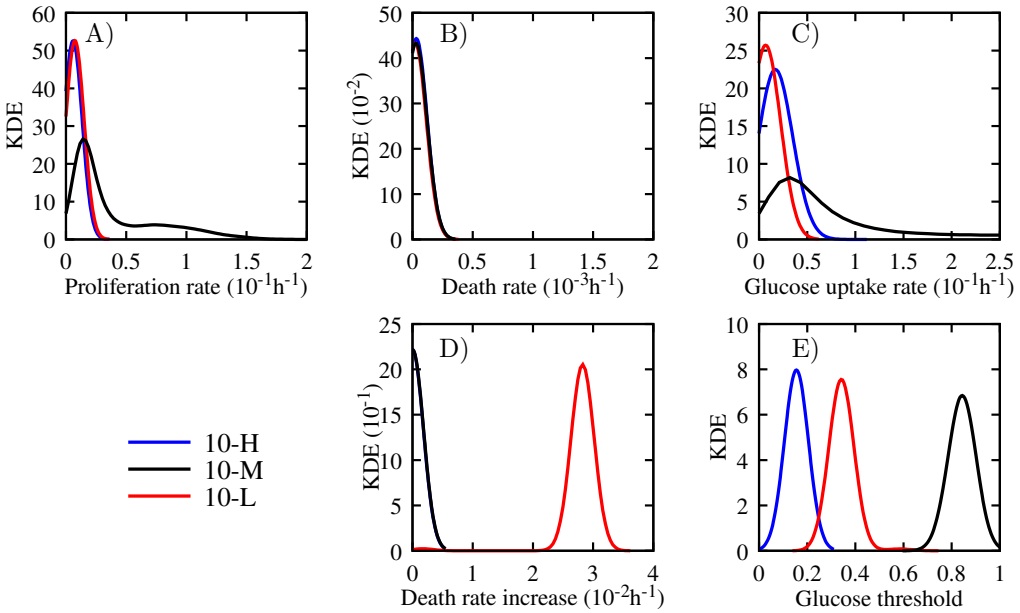

Fig 10. Posterior kernel density estimation (KDE) of scenarios with 10 mM initial glucose concentration. The KDE obtained during the scenario-specific calibration of the 10 mM initial glucose concentration and low (red), medium (black), and high (blue) tumor initial condition for the following parameters: $\bar{\alpha}_P$ (proliferation rate - panel A), $\bar{\alpha}_D$ (death rate - panel B), λ (glucose uptake rate - panel C), γ_D (death rate increase - panel D), and σ_H (glucose threshold - panel E). This initial glucose concentration is enough to sustain tumor growth to all initial tumor confluences, such as that no increase in the confluence of dead cells is noticed. Due to the lack of the dead cells, and the fact that the glucose uptake rate is low, the glucose concentration never drops below σ_H, and the parameter γ_D does not play a role on the tumor dynamics.

cell behavior.

Panel B of Fig 13 shows predicting scenario 10-L when the cgABM is calibrated with the other eight scenarios. The 10-L (low initial tumor confluence and 10 mM initial glucose) scenario, the confluence of the dead cells is constant over time, and the growth rate of the tumor cells is negligible (see Fig 7). From the prediction errors in Table 2, one can conclude that the 10-L scenario does not provide new information related to the tumor cells dynamics for calibrating the model parameters. That is, the other eight scenarios are sufficient to inform the dynamics of live and dead cells, resulting in a valid cgABM for computational prediction of the tumor cell behavior in a wide range of initial conditions.

Fig 14 shows the kernel density estimate (KDE) of the parameter posteriors obtained from the multi-scenario calibration process. This figure also shows the parameter posteriors for the two representative results of the leave-one-out calibrations from the calibration data (i.e., those presented in 13). Rigorous parameter estimation using different measurements must account for the data uncertainty in each training set. Characterization of the degree of confidence in the estimated parameter is critical in assessing the reliability of model prediction. The Bayesian method used in this work provides a suitable framework for quantifying the uncertainty in each measurement scenario. Comparing the parameter posteriors in Fig 14 with those obtained from scenario-specific calibration (Figs 8, 9, and 10) indicate the posterior distributions of the multi-scenario are affected more by the data with lower uncertainty (see error bars

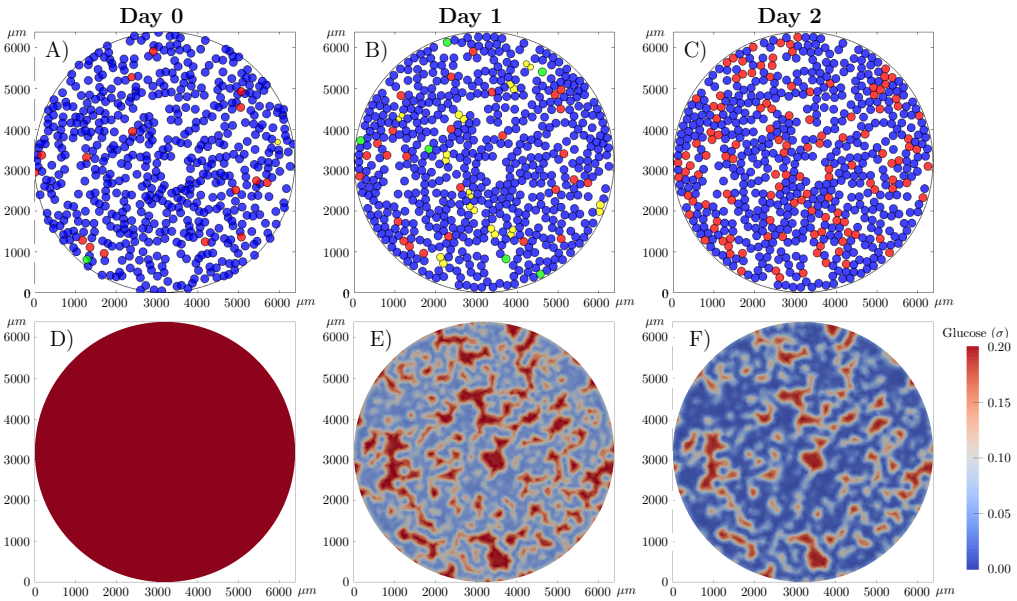

Fig 11. ABM simulation for the scenario 2-H. Snapshots of one realization of the cgABM simulation for the scenario with high initial tumor cell confluence and 2 mM initial glucose (2-H). Spatial-temporal development of tumor cells (panels A, B, and C) and glucose concentration (D, E, and F) over three days. Panels A and D display the initial conditions, while panels B and E present the model at day 1, and panels C and F display the model at day 2. The numerical simulation was performed using the MAP points of the calibrated parameters from the 2-H scenario ($\bar{\alpha}_P^{\text{MAP}} = 5.50 \times 10^{-3}\ h^{-1}$, $\bar{\alpha}_D^{\text{MAP}} = 6.52 \times 10^{-4}\ h^{-1}$, $\lambda^{\text{MAP}} = 6.37 \times 10^{-3}\ h^{-1}$, $\gamma_D^{\text{MAP}} = 2.18 \times 10^{-3}\ h^{-1}$, and $\sigma_H^{\text{MAP}} = 3.63 \times 10^{-3}$). In panels A, B, and C, the quiescent tumor cells are blue, dead cells are red, proliferative cells are green, and the daughter cells are in yellow. As glucose is consumed by the tumor cells, its concentration drops below the threshold needed for proliferation, leading to an increase in dead cells by day 2.

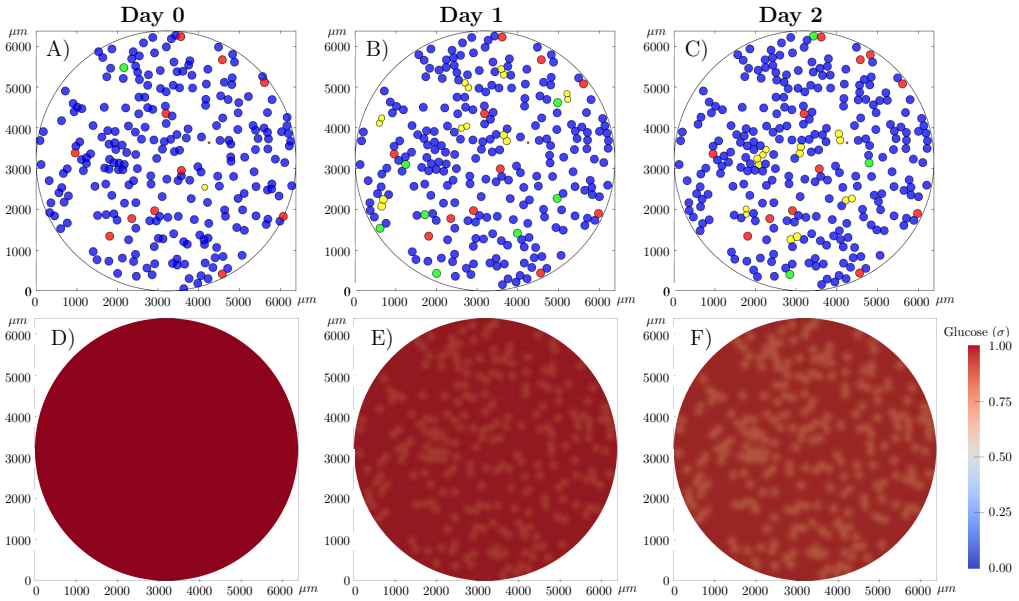

Fig 12. ABM simulation for the scenario 10-L. Snapshots of one realization of the cgABM simulation for the scenario with low initial tumor cell confluence and 10 mM initial glucose (10-L). Spatial-temporal development of tumor cells (panels A, B, and C) and glucose concentration (D, E, and F) over three days. Panels A and D display the initial conditions, while panels B and E present the model at day 1, and panels C and F display the model at day 2. The numerical simulation was performed using the MAP points of the calibrated parameters from the 10-L scenario ($\bar{\alpha}_P^{\text{MAP}} = 7.31 \times 10^{-3}\ h^{-1}$, $\bar{\alpha}_D^{\text{MAP}} = 1.73 \times 10^{-4}\ h^{-1}$, $\lambda^{\text{MAP}} = 3.44 \times 10^{-3}\ h^{-1}$, $\gamma_D^{\text{MAP}} = 8.45 \times 10^{-3}\ h^{-1}$, and $\sigma_H^{\text{MAP}} = 1.94 \times 10^{-1}$). In panels A, B, and C, the quiescent tumor cells are blue, dead cells are red, proliferative cells are green, and the daughter cells are in yellow. The low tumor cell confluence does not reduce the glucose to values below σ_H. Thus, there is no cell death due to the lack of glucose.

Fig 13. Model predictions of scenarios 2-H and 10-L. Leave-one-out prediction of scenarios 2-H (A) and 10-L (B). The cgABM was calibrated using the time-resolved microscopy data for all scenarios, excluding 2-H (panel A) and 10-L (panel B). The data mean and 95% credible interval for the live and dead cells are shown in blue and black, respectively. The mean of the simulation is represented as a solid line, and the area is the 95% credible interval for the live (green) and dead (red) cells. The average error and standard deviation for live and dead cells in scenario 2-H are 1.97 3.69% and 15.80 5.59%, respectively. While that for scenario 10-M are 7.61 2.01% and 5.78 1.13%, respectively.

in Figs 5, 6, 7). That is, poor characterization of measurement error during calibration leads to bias in model prediction.

Table 3 presents the MAP estimated of the posterior distributions from Fig 14. The MAP estimates of Table 3 demonstrate that the difference between the parameters of the calibration excluding 10-L data and the ones obtained when calibrating the nine scenarios together is smaller than the discrepancy between the posteriors of the multi-scenario $_{post}(\ jD)$ and the calibration excluding 2-H $_{post}(\ jD_{\ 2H})$. These remarks confirm the importance of including the 2-H experimental data into the calibration process to adequately inform the model parameters and enhance the predictive capability of the cgABM.

Table 3. Maximum A Posteriori (MAP) estimates.

Parameter	Maximum A Posteriori (MAP)		
	Multi-scenario $_{post}(\ jD)$	Leave-one-out $_{post}(\ jD_{\ 2H})$	Leave-one-out, $_{post}(\ jD_{\ 10L})$
Proliferation rate (h^{-1})	2.65 $10^{\ 2}$	1.03 $10^{\ 1}$	9.05 $10^{\ 2}$
Death rate (h^{-1})	2.47 $10^{\ 3}$	6.29 $10^{\ 4}$	1.77 $10^{\ 3}$
Glucose uptake rate (h^{1})	3.88 $10^{\ 2}$	6.90 $10^{\ 1}$	5.92 $10^{\ 2}$
Death rate increase (h^{1})	2.09 $10^{\ 2}$	1.34 $10^{\ 4}$	2.35 $10^{\ 2}$
Glucose threshold	4.68 $10^{\ 2}$	2.04 $10^{\ 2}$	2.00 $10^{\ 2}$

MAP estimates of the parameter posteriors for the multi-scenario calibration and the two cases (excluding scenarios 2-H and 10-L) of the leave-one-out calibration (see Fig 14).

Discussion

We have presented a hybrid stochastic agent-based model to simulate the interaction among tumor cells and glucose consumption. In this model, the tumor cell movement, growth, and phenotypic transitions are represented by a discrete cellular-scale model, while a continuum tissue-scale model governs the glucose evolution. In our model,

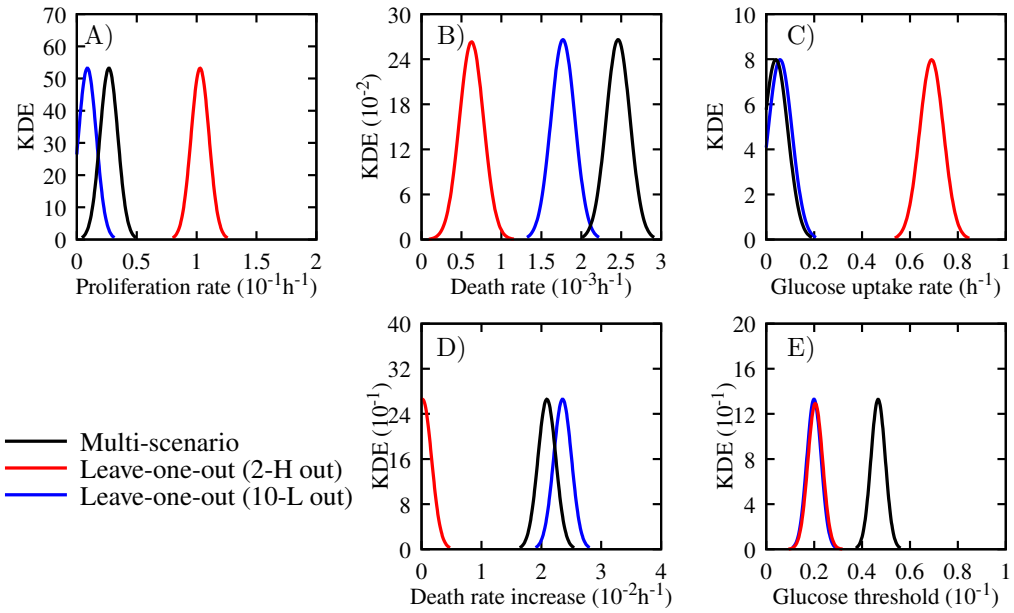

Fig 14. Posterior kernel density estimation (KDE) of the multi-scenario and the leave-one-out experiments. The KDE obtained during the calibration of the multi-scenario calibration $\pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{D})$ (black), the high tumor initial condition and 2 mM glucose condition left out $\pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{D}_{\sim 2H})$ (red), and the low tumor initial condition and 10 mM glucose condition left out $\pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{D}_{\sim 10L})$ (blue) for the following parameters: $\bar{\alpha}_P$ (proliferation rate - panel A), $\bar{\alpha}_D$ (death rate - panel B), λ (glucose uptake rate - panel C), γ_D (death rate increase - panel D), and σ_H (glucose threshold - panel E). Leaving scenario 2-H out of the calibration causes a stronger shift in the parameter distributions than leaving out scenario 10-L. This indicates that 2-H is a scenario that cannot be represented adequately by the other experiments.

tumor cell proliferation and death are stochastic events that are proportional to the available glucose. The tumor cell proliferation reduces as the glucose decreases, and the chances of cell death increase if the glucose is below a threshold.

We investigated the validity of the hybrid multiscale ABM in predicting *in vitro* experimental data of human breast carcinoma cells in nine scenarios, with combinations of different initial glucose concentration and tumor confluence. To this end, we addressed several challenges in model calibration and predictive hybrid multiscale ABMs. To overcome the high computational cost of the ABM, we coarse-grained the discrete model such that one agent represents a cluster of cells. By controlling the coarse-graining error below 5%, the developed cgABM enables simulating the entire domain of the *in vitro* experiment plate in a realistic time scale. The higher variations of the hybrid cgABM with a larger coarse-graining degree indicates that although coarse-graining of ABMs leads to a computationally tractable model, parameter inference of cgABMs demands methods that cope with significant stochasticity of the model. We also note that while our computational capabilities allow for full-scale ABM simulations and parameter inference, studying the predictive capabilities of cgABMs are critically needed for unrestricted use of these models in the cancer community, given their moderate computational costs.

The time-dependent variance-based sensitivity analysis method was employed to identify how each parameter contributes to the model outputs (live and dead cell confluences) during the system development. The results of the sensitivity analysis showed that the most influential parameter in the multiscale cgABM is the death rate

increase due to the lack of glucose (γ_D), followed by the glucose threshold (σ_H), and the glucose consumption rate (λ). The total effect sensitivity indices of the proliferation and death intensity reach values above 0.2 during the tumor development. The sensitivity analysis results allow us to determine which parameters must be accurately calibrated to represent the data. This analysis also helps design future experiments that can help to improve the calibration of model parameters. Two examples of experimental measurements that would significantly improve the cgABM calibration and its computational prediction are 1) temporal measurement of glucose concentration; 2) the use of hypoxia markers to track hypoxic cell confluence. Such experimental measurements would enable a more rigorous investigation of the glucose consumption rate (λ) and the glucose threshold (σ_H) that increases death and affects proliferation.

Directed by the sensitivity analysis, a series of statistical calibrations of the cgABM using the *in vitro* experimental data were conducted. A generalized likelihood function was proposed within the Bayesian inference to account for the intrinsic stochasticity of the cgABM in the statistical inverse problem. To perform the computationally expensive sampling algorithms for the Bayesian inferences of the stochastic cgABM, we implemented a computational infrastructure with efficient use of high-performance computing resources. We presented three model identification strategies. First, the results of scenario-specific calibration of cgABM to the time-resolved microscopy measurements indicate that the developed hybrid multiscale model can predict the experimentally observed responses of human breast carcinoma cells in a wide range of initial glucose concentrations and tumor confluences. Comparing the calibrated parameters for all nine scenarios shows that the cgABM can recapitulate the main features and the underlying tumor development mechanisms in the avascular stage, with an average error of less than 8%. These processes include cell proliferation and mitosis in favorable glucose environments, as well as cell death due to lack of glucose and through apoptosis and hypoxia. Furthermore, the Bayesian inference method used for parameter calibration allows for assessing the associated uncertainties in model parameters while handling the inherent stochasticity of the cgABM. Second, a multi-scenario calibration was performed to identify the cgABM parameters from the in vitro data of all nine experimental data sets. The model calibration using multiple data sets with different data noise levels indicates the requirement of rigorous characterization of uncertainty in model parameters. Thus, employing Bayesian inference for statistical calibration of the ABMs is essential for characterizing the level of confidence in their computational prediction. Third, we designed an extensive calibration process, based on the leave-one-out approach, to challenge the cgABM by testing its validity in predicting unseen data. To this end, the cgABM is calibrated using the data of eight scenarios, excluding one data set (prediction scenario) from the calibration process. The results of leave-one-out calibration indicate the developed cgABM can predict the *in vitro* experiments with an error ranging from 5% to 21%. In particular, if the calibration data include the scenarios with distinct dead cell dynamics, the calibrated cgABM can precisely capture the accumulation of the dead cells, and it is a valid model for predicting the *in vitro* tumor cell behavior according to our validation criteria.

Recently, there are similar efforts in the literature on the calibration of agent-based models to tumor cells [18, 59, 60]. Jagiella *et al.* [59] calibrated a single cell-based mathematical model for multi-cellular tumor spheroids using non-small cell lung cancer (NSCLC) data. Similar to our model, the phenotypic states in their model allowed for proliferative, quiescent, and dead agents. However, the main difference between their model and the one presented here, besides their model being 3D and lattice-based, is the number of environmental factors considered in their model. In [59], the glucose, oxygen, lactate, extracellular matrix, and waste material released by dying cells are also modeled. According to the authors, this was the simplest model to capture their data,

consisting of the growth kinetics and the corresponding spatial staining patterns for nuclei, different cell states, and cell environments. Although our model is simpler compared to [59], it is able to simulate our experimental observations accurately. A similar result between our model and the on in [59] is that apoptosis does not play an essential role in dead cell dynamics. As demonstrated by our sensitivity analysis, the effects of limited glucose, which translates into an increase in cell mortality (γ_D), is the most important factor to determine the dynamics of dead cells. In [60], a hybrid ABM is developed to model the growth of multi-cellular tumor spheroids. The possible cell phenotypes are quiescent, proliferative, hypoxic, dead, and necrotic (dead cells that have undergone cytolysis). The environmental factors modeled are oxygen, glucose, and the concentration of a hypoxia-activated prodrug (SN30000). To capture the radiation's effects, the linear-quadratic model was employed by [61]. The data used for the calibration comes from *in vitro* experiments with human colon cancer cells treated with radiation, in a range of dose rates (0.2-1 Gy/min) and SN30000. Among the measurements used are the temporal changes of the spheroid diameter, number of cells, cell viability, hypoxic fraction, and S-phase cell fraction. In their model, apoptosis is not considered, with the death of the tumor cells being due to oxygen or glucose deprivation and the treatment effects. This assumption is aligned with the results found in our calibration, where γ_D is at least ten times higher than apoptosis in every scenario. One main difference between our model and the ones in [59, 60], is that we developed a lattice-free model, which is generally more computationally expensive (considering the same number of model constituents). The methodology employed here for model coarsening allows us to perform statistical calibration considering the model stochasticity.

While the proposed hybrid two-scale ABM can represent the tumor growth and decline observed in the experimental measurements, there remain several areas that can be further developed in future studies. In the current effort, the spatial distribution of the tumor cells was not taken into account during the model calibration (i.e., only the confluence data was used in the model calibration). Thus, we are not considering the mobility of the tumor, as previously done in [3]. This effect would be necessary to characterize, for example, angiogenesis [20]. The position of the tumor cells influences the growth pattern of the new vasculature. Another direction for investigation is to calibrate the ABM using 3D tumor platforms [62–65]. However, a limiting factor in our 3D simulations would be the computational cost characteristic of discrete models with multiple agents. The 2D cgABM simulation simulates the observational data within a 5% error compared to the one cell per agent ABM. However, we expect the coarse-graining error would be more substantial in 3D simulations. Finally, we showed that the cgABM was able to capture the evolution of human breast carcinoma cells, in three different initial confluences and three different glucose availability, with an average error of less than 8%. One aspect that could improve the simulation is a better characterization of the initial distribution of the cell cycle within the tumor cells. The current model considers all cells initially in a quiescent state. Additionally, labeling the cell phenotypes in the *in vitro* experiments, see, e.g., [66] could provide better observational data to portray the initial conditions.

Conclusions

We have developed a hybrid, two-scale, stochastic agent-based model of tumor development and systematically investigated its ability to simulate and predict *in vitro* experimental observations of live and dead cell number over time given the initial conditions. A notable feature is the rigorous characterization of uncertainty in the prediction provided by the stochastic ABM, using novel time-dependent sensitivity

analysis and Bayesian inference methods along with efficient use of high-performance computing resources. The study indicates that the developed multiscale ABM can reliably predict the spatiotemporal evolution of tumors subjected to a wide range of initial conditions.

Supporting information

S1 Fig. Low confluence sensitivity analysis. Sensitivity analysis of the proliferation rate ($_P$), death rate ($_D$), glucose diffusion (D), glucose uptake ( ), death rate increase due to lack of glucose (δ), transition contrast (k), and glucose threshold ($_H$) for live (top row) and dead (bottom row) cell phenotypes seeded with low confluence. Panels A-F show the total effect index over time with Panels A, B, and C depicting live tumor cells, while Panels D, E, and F depict the dead tumor cells. The importance of the parameters is studied for three initial glucose concentrations: 2 mM (Panels A and D), 5 mM (Panels B and E), and 10 mM (Panels C and F). The glucose diffusion and the smooth transition constant have limited influence on the quantities of interest during the complete simulation (i.e., large changes in these parameters would yield small changes in tumor composition). Apart from these two parameters, the total effect index for every parameter is greater than 0.2 during the 96 hours simulated.

S2 Fig. High confluence sensitivity analysis. Sensitivity analysis of the proliferation rate ($_P$), death rate ($_D$), glucose diffusion (D), glucose uptake ( ), death rate increase due to lack of glucose ($_0$), transition contrast (k), and glucose threshold ($_H$) for live (top row) and dead (bottom row) cell phenotypes seeded with high confluence. Panels A-F show the total effect index over time with Panels A, B, and C depicting live tumor cells, while Panels D, E, and F depict the dead tumor cells. The importance of the parameters is studied for three initial glucose concentrations: 2 mM (Panels A and D), 5 mM (Panels B and E), and 10 mM (Panels C and F). The glucose diffusion and the smooth transition constant have limited influence on the quantities of interest during the complete simulation (i.e., large changes in these parameters would yield small changes in tumor composition). Apart from these two parameters, the total effect index for every parameter is greater than 0.2 during the 96 hours simulated.

S1 Appendix. Parameter Identifiability Analysis.

As the hybrid multiscale ABM is quite complete with many model parameters, it is natural to determine (before model calibration against experimental observations) if it is feasible to infer the parameters of the model from data. Unidentifiable parameters are the model parameters that cannot be learned from a set of \perfect" data resulting in the same model predictions for more than one value of the parameters [67]. In a Bayesian setting, different combinations of unidentifiable parameters lead to the same likelihood. Additionally, due to strong correlations among the posterior probability distribution of the parameters, the existence of unidentifiable model parameters slows down the convergence rate of MCMC algorithms [68].

For these reasons, before conducting the Bayesian calibration of the cgABM against the in vitro experimental data, we perform a parameter identifiability analyses. To this end, we make use of the cgABM with a set of \true" parameter values to generate *in silico* data. We then calibrate the cgABM parameters against the *in silico* data using Bayesian inference. To check the parameter identifiability, we compare the true parameter values with the MAP estimates of the calibration posteriors.

Table 4 shows the true values of the cgABM parameters used to generate the set of *in silico* data. The true parameters are chosen such that the data show the initial increase in live cell confluence, followed by decrease of live cell confluence due to the lack of glucose in the microenvironment, and, consequentially, an increase in dead cells. The domain size, degree of coarse-graining, and other simulation features are the same as those used for the sensitivity analyses and calibration in the main manuscript. The MAP estimates of the calibrated parameters inferred from the data are shown in Table 4. We observe that the inferred parameters are close to their true values with a

Table 4. Parameter identifiability.

Parameter	True value	Maximum A Posteriori
Proliferation intensity	4.90e-02	4.96e-02
Death intensity	4.10e-04	5.08e-04
glucose uptake rate	4.80e-02	5.73e-02
Death rate increase	2.40e-02	2.83e-02
glucose threshold	5.40e-02	4.06e-02

True parameters value used to generate the *in silico* data and the values of MAP estimates obtained from calibration posteriors.

maximum discrepancy of 25% in the glucose threshold. Fig 15 presents the *in silico* data and the cgABM model outputs using the calibration posterior. The comparison between data and model in this figure indicates that the model can capture the main features of the data with an average error below 0.21%.

The results of the parameter identifiability analysis in Table 4 and Fig 15 indicate that the cgABM calibration parameters are identifiable. They also verify our Bayesian calibration approach and computational infrastructure to be used for calibrating the cgABM against *in vitro* experiments.

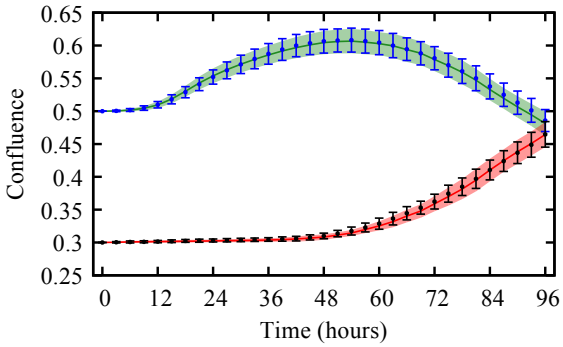

Fig 15. Spatiotemporal evolution of the *in silico* data and the calibrated model. Comparison between *in silico* data (points with 95% error bars) and model calibration (mean as a continuous line and 95% confidence interval as shadow area). The mean relative error for the living (blue and green) and dead (black and red) cells were $0.22 \pm 0.06\%$ and $0.19 \pm 0.05\%$, respectively.

S2 Appendix. Bayesian calibration computation.

For all calibrations, the Bayesian inference is conducted using 16 chains within the adaptive multi-level Monte Carlo algorithm, along with computing the means of $N_r = 17$ realizations of the model per sample. As each forward model is computed in serial, the total number of processors per simulation is the number of chains times the number of realizations (i.e., 272 processors in the scenario-specific calibration). For the multi-scenario and the leave-one-out calibration, we multiply this number of processors by the number of scenarios used in the calibration. In Table 5, we present the computational time for all the calibration experiments presented in this work.

Table 5. Model parameters.

Scenario	Scenario-specific	Leave-one-out
2-L	04 h 33 min and 35 s	26 h 38 min and 45 seconds
2-M	04 h 46 min and 01 s	33 h 59 min and 18 seconds
2-H	04 h 47 min and 31 s	28 h 00 min and 34 seconds
5-L	03 h 38 min and 54 s	24 h 09 min and 30 seconds
5-M	03 h 53 min and 50 s	16 h 57 min and 24 seconds
5-H	05 h 48 min and 24 s	34 h 29 min and 08 seconds
10-L	03 h 57 min and 30 s	37 h 40 min and 34 seconds
10-M	05 h 54 min and 16 s	15 h 50 min and 40 seconds
10-H	12 h 05 min and 05 s	13 h 40 min and 10 seconds

Computational time for the scenario-specific and the leave-one-out calibrations. The middle and right columns indicate the scenario-specific calibrated and the scenario left out during the calibration for the leave-one-out experiments, respectively. The computational time for the multi-scenario calibration was 45 h 23 min and 25 s.

Acknowledgments

We thank the National Cancer Institute for funding *via* U01CA174706, R01CA186193, U01CA253540, and U24 CA226110. We also thank the Cancer Prevention and Research Institute of Texas (CPRIT) for funding via RR160005. T.E.Y. is a CPRIT Scholar in Cancer Research. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported within this paper.

References

1. Macklin P, Edgerton ME. Agent-based cell modeling: application to breast cancer. Cambridge University Press; 2010.
2. Byrne H, Drasdo D. Individual-based and continuum models of growing cell populations: a comparison. Journal of mathematical biology. 2009;58(4):657–687.
3. Rocha HL, Almeida RC, Lima EABF, Resende ACM, Oden JT, Yankeelov TE. A hybrid three-scale model of tumor growth. Mathematical Models and Methods in Applied Sciences. 2018;.
4. Szabó A, Merks RM. Cellular potts modeling of tumor growth, tumor invasion, and tumor evolution. Frontiers in oncology. 2013;3:87.
5. Jiang Y, Pjesivac-Grbovic J, Cantrell C, Freyer JP. A multiscale model for avascular tumor growth. Biophysical journal. 2005;89(6):3884–3894.
6. Sun X, Zhang L, Tan H, Bao J, Strouthos C, Zhou X. Multi-scale agent-based brain cancer modeling and prediction of TKI treatment response: Incorporating EGFR signaling pathway and angiogenesis. BMC Bioinformatics. 2012;13.
7. Olsen MM, Siegelmann HT. Multiscale Agent-based Model of Tumor Angiogenesis. Procedia Computer Science. 2013;18:1016–1025.
8. Anderson AR, Chaplain M. Continuous and discrete mathematical models of tumor-induced angiogenesis. Bulletin of mathematical biology. 1998;60(5):857–899.
9. Rens EG, Merks RM. Cell shape and durotaxis explained from cell-extracellular matrix forces and focal adhesion dynamics. Iscience. 2020;23(9):101488.
10. Pourhasanzade F, Sabzpoushan S, Alizadeh AM, Esmati E. An agent-based model of avascular tumor growth: Immune response tendency to prevent cancer development. Simulation. 2017;93(8):641–657.
11. Ghaffarizadeh A, Heiland R, Friedman SH, Mumenthaler SM, Macklin P. PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLoS computational biology. 2018;14(2):e1005991.
12. Letort G, Montagud A, Stoll G, Heiland R, Barillot E, Macklin P, et al. PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. Bioinformatics. 2018;35(7):1188–1196.
13. Gong C, Milberg O, Wang B, Vicini P, Narwal R, Roskos L, et al. A computational multiscale agent-based model for simulating spatio-temporal tumour immune response to PD1 and PDL1 inhibition. Journal of The Royal Society Interface. 2017;14(134):20170320.
14. Boghaert E, Radisky DC, Nelson CM. Lattice-based model of ductal carcinoma in situ suggests rules for breast cancer progression to an invasive state. PLoS computational biology. 2014;10(12):e1003997.
15. Kim PS, Lee PP. Modeling protective anti-tumor immunity via preventative cancer vaccines using a hybrid agent-based and delay differential equation approach. PLoS computational biology. 2012;8(10):e1002742.

16. Mirams GR, Arthurs CJ, Bernabeu MO, Bordas R, Cooper J, Corrias A, et al. Chaste: an open source C++ library for computational physiology and biology. PLoS computational biology. 2013;9(3):e1002970.

17. Oden JT, Lima EABF, Almeida RC, Feng Y, Rylander MN, Fuentes D, et al. Toward predictive multiscale modeling of vascular tumor growth. Archives of Computational Methods in Engineering. 2016;23(4):735–779.

18. Macklin P, Edgerton ME, Thompson AM, Cristini V. Patient-calibrated agent-based modelling of ductal carcinoma in situ (DCIS): from microscopic measurements to macroscopic predictions of clinical progression. Journal of theoretical biology. 2012;301:122–140.

19. Norton KA, Gong C, Jamalian S, Popel AS. Multiscale agent-based and hybrid modeling of the tumor immune microenvironment. Processes. 2019;7(1):37.

20. Phillips CM, Lima EABF, Woodall RT, Brock A, Yankeelov TE. A hybrid model of tumor growth and angiogenesis: In silico experiments. Plos one. 2020;15(4):e0231137.

21. Lyu J, Cao J, Zhang P, Liu Y, Cheng H. Coupled Hybrid Continuum-Discrete Model of Tumor Angiogenesis and Growth. PloS one. 2016;11(10):e0163173.

22. Van Liedekerke P, Neitsch J, Johann T, Alessandri K, Nassoy P, Drasdo D. Quantitative cell-based model predicts mechanical stress response of growing tumor spheroids over various growth conditions and cell lines. PLOS Computational Biology. 2019;15(3):1–40.

23. Metzcar J, Wang Y, Heiland R, Macklin P. A Review of Cell-Based Computational Modeling in Cancer Biology. JCO Clinical Cancer Informatics. 2019;(3):1–13.

24. Yankeelov TE, Quaranta V, Evans KJ, Rericha EC. Toward a science of tumor forecasting for clinical oncology. Cancer research. 2015;75(6):918–923.

25. Yankeelov TE, Atuegwu N, Hormuth D, Weis JA, Barnes SL, Miga MI, et al. Clinically relevant modeling of tumor growth and treatment response. Science translational medicine. 2013;5(187):187ps9–187ps9.

26. Fröhlich F, Thomas P, Kazeroonian A, Theis FJ, Grima R, Hasenauer J. Inference for stochastic chemical kinetics using moment equations and system size expansion. PLoS computational biology. 2016;12(7):e1005030.

27. Chen X, Li X, Zhao W, Li T, Ouyang Q. Parameter sensitivity analysis for a stochastic model of mitochondrial apoptosis pathway. PloS one. 2018;13(6):e0198579.

28. Glen CM, Kemp ML, Voit EO. Agent-based modeling of morphogenetic systems: Advantages and challenges. PLOS Computational Biology. 2019;15(3):1–31.

29. Leung MC, Hutson MS, Seifert AW, Spencer RM, Knudsen TB. Computational modeling and simulation of genital tubercle development. Reproductive Toxicology. 2016;64:151–161.

30. Buske P, Galle J, Barker N, Aust G, Clevers H, Loeffler M. A comprehensive model of the spatio-temporal stem cell and tissue organisation in the intestinal crypt. PLoS computational biology. 2011;7(1):e1001045.

31. Setty Y, Cohen IR, Dor Y, Harel D. Four-dimensional realistic modeling of pancreatic organogenesis. Proceedings of the National Academy of Sciences. 2008;105(51):20374–20379.

32. Noid W, Chu JW, Ayton GS, Krishna V, Izvekov S, Voth GA, et al. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. The Journal of chemical physics. 2008;128(24):244114.

33. Oden JT, Babuška I, Faghihi D. Predictive computational science: Computer predictions in the presence of uncertainty. Encyclopedia of Computational Mechanics Second Edition. 2017; p. 1–26.

34. Kazerouni AS, Gadde M, Gardner A, Hormuth II DA, Jarrett AM, Johnson KE, et al. Integrating quantitative assays with biologically-based mathematical modeling for predictive oncology. Iscience. 2020; p. 101807.

35. Gallaher J, Hawkins-Daarud A, Massey SC, Swanson K, Anderson A. Hybrid approach for parameter estimation in agent-based models. BioRxiv. 2017; p. 175661.

36. Lambert B, MacLean AL, Fletcher AG, Combes AN, Little MH, Byrne HM. Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. Journal of mathematical biology. 2018;76(7):1673–1697.

37. D'Antonio G, Macklin P, Preziosi L. An agent-based model for elasto-plastic mechanical interactions between cells, basement membrane and extracellular matrix. Math Biosci Eng. 2013;10(1):75–101.

38. Kirk BS, Peterson JW, Stogner RH, Carey GF. `libMesh`: A C++ Library for Parallel Adaptive Mesh Refinement/Coarsening Simulations. Engineering with Computers. 2006;22(3–4):237–254.

39. Saltelli A, Aleksankina K, Becker W, Fennell P, Ferretti F, Holst N, et al. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environmental Modelling & Software. 2019;114:29–39.

40. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. Global Sensitivity Analysis: The Primer. Wiley; 2008.

41. Saltelli A, Chan K, Scott EM. Sensitivity Analysis. No. no. 2008 in Wiley paperback series. Wiley; 2009.

42. Sobol' IM. Sensitivity estimates for nonlinear mathematical models. Matematicheskoe Modelirovanie. 1990;2:112–118.

43. Sobol' IM. Sensitivity analysis for non-linear mathematical models. Mathematical Modeling and Computational Experiment. 1993;1:407–414.

44. Sobol' IM. Global sensitivity analysis indices for the investigation of nonlinear mathematical models,. Matematicheskoe Modelirovanie. 2007;19:23–24.

45. Homma T, Saltelli A. Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering and System Safety. 1996;52(1):1 – 17.

46. Saltelli A, Tarantola S. On the Relative Importance of Input Factors in Mathematical Models. Journal of the American Statistical Association. 2002;97(459):702–709.

47. Saltelli A. Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications. 2002;145(2):280 – 297.

48. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications. 2010;181(2):259 – 270.

49. Jaynes ET. Probability theory: the logic of science. Cambridge university press; 2003.

50. Jaynes E. The relation of Bayesian and maximum entropy methods. In: Maximum-entropy and Bayesian methods in science and engineering. Springer; 1988. p. 25–29.

51. Faghihi D, Sarkar S, Naderi M, Rankin JE, Hackel L, Iyyer N. A Probabilistic Design Method for Fatigue Life of Metallic Component. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering. 2018;4(3):031005.

52. Prudencio E, Bauman P, Faghihi D, Ravi-Chandar K, Oden J. A computational framework for dynamic data-driven material damage control, based on Bayesian inference and model selection. International Journal for Numerical Methods in Engineering. 2015;102(3-4):379–403.

53. Ruess J, Lygeros J. Moment-based methods for parameter inference and experiment design for stochastic biochemical reaction networks. ACM Transactions on Modeling and Computer Simulation (TOMACS). 2015;25(2):8.

54. Zechner C, Ruess J, Krenn P, Pelet S, Peter M, Lygeros J, et al. Moment-based inference predicts bimodality in transient gene expression. Proceedings of the National Academy of Sciences. 2012;109(21):8340–8345.

55. Roberts GO, Rosenthal JS, et al. General state space Markov chains and MCMC algorithms. Probability Surveys. 2004;1:20–71.

56. Gelman A, Rubin DB, et al. Inference from iterative simulation using multiple sequences. Statistical science. 1992;7(4):457–472.

57. Prudencio E, Cheung SH. Parallel adaptive multilevel sampling algorithms for the Bayesian analysis of mathematical models. International Journal for Uncertainty Quantification. 2012;2(3).

58. Prudencio EE, Schulz KW. The parallel C++ statistical library 'QUESO': Quantification of Uncertainty for Estimation, Simulation and Optimization. In: Euro-Par 2011: Parallel Processing Workshops. Springer; 2012. p. 398–407.

59. Jagiella N, Müller B, Müller M, Vignon-Clementel IE, Drasdo D. Inferring growth control mechanisms in growing multi-cellular spheroids of NSCLC cells from spatial-temporal image data. PLoS computational biology. 2016;12(2):e1004412.

60. Mao X, McManaway S, Jaiswal JK, Patel PB, Wilson WR, Hicks KO, et al. An agent-based model for drug-radiation interactions in the tumour microenvironment: Hypoxia-activated prodrug SN30000 in multicellular tumour spheroids. PLoS computational biology. 2018;14(10):e1006469.

61. Kellerer AM, Rossi HH. The theory of dual radiation action. Current Topics in Radiation Research Quarterly. 1974; p. 85–158.

62. Szot CS, Buchanan CF, Freeman JW, Rylander MN. 3D in vitro bioengineered tumors based on collagen I hydrogels. Biomaterials. 2011;32(31):7905–7912.

63. Ozkan A, Ghousifam N, Hoopes PJ, Yankeelov TE, Rylander MN. In Vitro vascularized liver and tumor tissue microenvironments on a chip for dynamic determination of nanoparticle transport and toxicity. Biotechnology and bioengineering. 2019;116(5):1201–1219.

64. Buchanan CF, Voigt E, Szot CS, Freeman JW, Vlachos PP, Rylander MN. Three-Dimensional Microfluidic Collagen Hydrogels for Investigating Flow-Mediated Tumor-Endothelial Signaling and Vascular Organization. Tissue Engineering Part C: Methods. 2013;20(1):64–75.

65. Antoine E, Vlachos P, Rylander MN. Tunable Collagen I Hydrogels for Engineered Physiological Tissue Micro-environments. PloS one. 2014;10(3):e0122500.

66. Tyson DR, Garbett SP, Frick PL, Quaranta V. Fractional proliferation: a method to deconvolve cell population dynamics from single-cell data. Nature methods. 2012;9(9):923–928.

67. Rateitschak K, Winter F, Lange F, Jaster R, Wolkenhauer O. Parameter identifiability and sensitivity analysis predict targets for enhancement of STAT1 activity in pancreatic cancer and stellate cells. PLoS computational biology. 2012;8(12).

68. Rannala B. Identifiability of parameters in MCMC Bayesian inference of phylogeny. Systematic Biology. 2002;51(5):754–760.