



Mathematical modeling of multicellular tumor spheroids quantifies inter-patient and intra-tumor heterogeneity



Adam A. Malik^{1,7}✉, Kyle C. Nguyen^{2,3,7}, John T. Nardini^{1D}⁴, Cecilia C. Krona^{1D}⁵, Kevin B. Flores^{3,6} & Sven Nelander^{1D}⁵

In the study of brain tumors, patient-derived three-dimensional sphere cultures provide an important tool for studying emerging treatments. The growth of such spheroids depends on the combined effects of proliferation and migration of cells, but it is challenging to make accurate distinctions between increase in cell number versus the radial movement of cells. To address this, we formulate a novel model in the form of a system of two partial differential equations (PDEs) incorporating both migration and growth terms, and show that it more accurately fits our data compared to simpler PDE models. We show that traveling-wave speeds are strongly associated with population heterogeneity. Having fitted the model to our dataset we show that a subset of the cell lines are best described by a “Go-or-Grow”-type model, which constitutes a special case of our model. Finally, we investigate whether our fitted model parameters are correlated with patient age and survival.

Glioma grade IV, or glioblastoma multiforme (GBM) stands as the most prevalent diagnosis among all gliomas. It is notoriously malignant, with the majority of GBM patients succumbing to the disease within a year^{1,2}. This outcome can be attributed to pronounced inter- and intra-tumor heterogeneity, coupled with the infiltrative nature of the tumor itself³. Glioblastoma, often treated with surgery, radiotherapy, chemotherapy, or a combination, faces high recurrence rates¹. Surgical challenges arise due to the tumors’ diffuse nature, with cells capable of extensive migration through healthy tissue to infiltrate crucial regions of the brain². Research endeavors span multiple domains, encompassing imaging and surgical approaches⁴, advancements in radiotherapy techniques⁵, exploration of immunotherapeutic avenues⁶, pharmaceutical interventions⁷⁻⁹, as well as mathematical and computational modeling^{10,11}. These multifaceted efforts underscore the urgent need for innovative approaches to confront the formidable challenges posed by GBM. In this study we adopt a multidisciplinary approach towards GBM by combining large scale multicellular tumor spheroid experiments from patient-derived cells, with novel mathematical modeling and analysis.

A popular experimental model for GBM tumors are 3D cultured multicellular tumor spheroids (MCTSs). They replicate the *in vivo* environment more faithfully compared to 2D monolayer models¹². Unlike cells cultured on flat surfaces, those in 3D environments experience crucial cell-

cell and cell-extracellular matrix interactions that mimic the natural conditions within or in proximity to a tumor. Consequently, 3D tumor spheroid assays have emerged as invaluable tools in drug screens and therapeutic efficacy evaluation^{13,14}.

Many mathematical models have been introduced to describe the properties of GBM¹⁵⁻²⁴. For example, the Fisher Kolmogorov-Petrovsky-Piskunov (Fisher-KPP) equation, a type of reaction-diffusion equation,

$$\frac{\partial u}{\partial t} = \nabla \cdot (D \nabla u) + \rho u \left(1 - \frac{u}{K}\right),$$

is the simplest partial differential equation (PDE) model that can be used to describe the spatial and temporal *in vitro* or *in vivo* spreading of GBM¹⁵⁻²². With the aid of the Fisher-KPP equation, Baldoek et al.¹⁸ showed that patients with less invasive GBM tumors would receive more benefit from resection. While previous studies have shown that the Fisher-KPP equation is clinically relevant^{16-18,25}, the model assumes an intratumor homogeneous cellular behavior that might not be true in GBM populations²⁶. As a result, many PDE models have been introduced to account for the population heterogeneity in GBM populations^{24,26-34}.

MCTSs are marked by the presence of heterogeneity. The tumors are often described as having distinct regions, such as a core and an invasive

¹Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden. ²Biomathematics Graduate Program, North Carolina State University, Raleigh, NC, USA. ³Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC, USA. ⁴Department of Mathematics and Statistics, The College of New Jersey, Ewing, NJ, USA. ⁵Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. ⁶Department of Mathematics, North Carolina State University, Raleigh, NC, USA. ⁷These authors contributed equally: Adam A. Malik, Kyle C. Nguyen. ✉e-mail: maadam@chalmers.se

rim²⁶. A property that is often seen in MCTS spreading is that the core and the invasive rim expand at different velocities³⁵, a phenomenon that cannot be explained by the Fisher-KPP equation. Another property that has been the subject of debate, from both a biological³⁶⁻³⁹ as well as a mathematical^{26,33,40,41} viewpoint, is the “Go-or-Grow” hypothesis. The hypothesis postulates that cells are either migrating or proliferating, and that the two processes are mutually exclusive. This hypothesis requires a more intricate model than the Fisher-KPP equation.

Our novel model involves two distinct populations, whose governing differential equations include reaction, diffusion, and advection terms. This model provides a comprehensive framework to capture the aforementioned complexities of growing tumor spheroids. To validate the efficacy of our model, we compared it to three other PDE models: the Fisher-KPP model with and without advection, as well as a 2-system PDE without advection. Our results demonstrate the superior fit of our proposed model to the data, underscoring its capacity to more accurately represent the observed dynamics. Furthermore, we conducted extensive fitting of our model to a diverse set of patient-derived cell lines, revealing notable heterogeneity between different patients. This exploration not only enhances our understanding of the varying growth patterns but also positions our model as a valuable tool for personalized medicine strategies. Finally, our model’s ability to provide meaningful insights is exemplified by its correlation with patient age and survival. Certain model parameters were found to be indicative of spheroid growth patterns that are associated with patient outcomes, which highlights the potential clinical relevance of our mathematical framework and its capacity to contribute to a more nuanced understanding of growing tumor spheroids. To the best of our knowledge, this is the first time

PDE model parameters from MCTS experiments have been connected to patient survival.

The organization of this paper is outlined in Fig. 1 and proceeds as follows. The results are presented in the “Results” section (Fig. 1C), where we first present our mathematical model (see the “Mathematical models” section) and compare our model to existing models of the same type (see the “Population heterogeneity and advection improves model fit” section), analyze the fitted models in relationship to cell line heterogeneity (see the “Wave front speeds approximate population heterogeneity” section), determine which cell lines are exhibiting Go-or-Grow behavior (see the “A subset of cell lines exhibit Go-or-Grow behavior” section), as well as investigate the association between model parameters and patient survival (see the “Correlation between patient outcome and mathematical model parameters” section). The “Discussion” section is devoted to a discussion and concluding remarks. Our methodology is presented in the “Methods” section (Fig. 1A), where we describe the data acquisition process (see the “Spheroid invasion assay” section), followed by the data processing steps (see the “Data processing” section). We then describe the parameter estimation procedure (see the “Parameter estimation method” section) (Fig. 1B).

Results

Mathematical models

Previous work has shown the clinical relevancy of the Fisher-KPP equation’s parameters^{16,18,21,25}. Extending from this equation, we propose a new model while accounting for the “Go-or-Grow” hypothesis to describe populations of phenotypically different subpopulations. Our proposed model is a

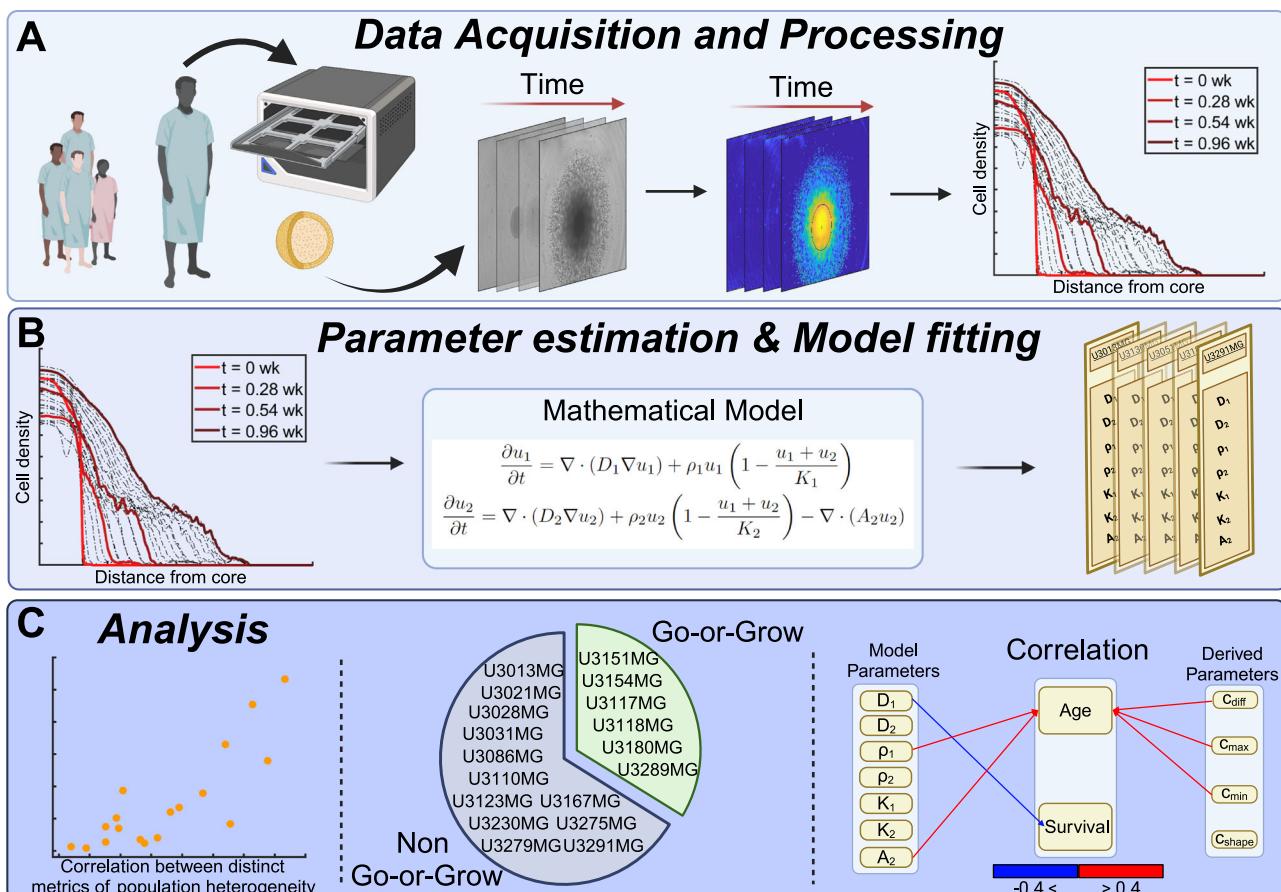


Fig. 1 | Paper outline and workflow. **A** Image sequences of multicellular tumor spheroids of patient-derived cell cultures are acquired. Images are processed and turned into density profiles describing the cell density as a function of radial distance from the tumor core. **B** Our mathematical model, a system of two PDEs, is fitted to the dataset and model parameters are obtained for each cell line. **C** Our analysis

demonstrates a number of insights, including the correlation between wave-parameters and population heterogeneity, identification of cell lines exhibiting Go-or-Grow behavior, as well as the correlation between model parameters and clinical data. This figure was created with BioRender.com.

Table 1 | Summary of models and sets of parameters to be estimated

Model	Equations	Parameters
RD	$\frac{\partial u}{\partial t} = \nabla \cdot (D \nabla u) + \rho u (1 - \frac{u}{K})$	D, ρ, K
ARD	$\frac{\partial u}{\partial t} = \nabla \cdot (D \nabla u) + \rho u (1 - \frac{u}{K}) - \nabla \cdot (A u)$	D, ρ, K, A
RD-RD	$\frac{\partial u_1}{\partial t} = \nabla \cdot (D_1 \nabla u_1) + \rho_1 u_1 \left(1 - \frac{u_1 + u_2}{K_1}\right)$ $\frac{\partial u_2}{\partial t} = \nabla \cdot (D_2 \nabla u_2) + \rho_2 u_2 \left(1 - \frac{u_1 + u_2}{K_2}\right)$	$D_1, \rho_1, K_1, a,$ D_2, ρ_2, K_2
RD-ARD	$\frac{\partial u_1}{\partial t} = \nabla \cdot (D_1 \nabla u_1) + \rho_1 u_1 \left(1 - \frac{u_1 + u_2}{K_1}\right)$ $\frac{\partial u_2}{\partial t} = \nabla \cdot (D_2 \nabla u_2) + \rho_2 u_2 \left(1 - \frac{u_1 + u_2}{K_2}\right) - \nabla \cdot (A_2 u_2)$	$D_1, \rho_1, K_1, a,$ D_2, ρ_2, K_2, A_2

spatiotemporal system that consists of two PDEs describing cell densities of two populations of cells: reaction-diffusion equation (Eq. (3)) and advection-reaction-diffusion equation (Eq. (4)), we refer to this model as the RD-ARD model. The first population density, $u_1(\vec{x}, t)$, is governed by the reaction-diffusion equation while the second population density, $u_2(\vec{x}, t)$, is a more migrative population and is governed by an advection-reaction-diffusion equation. The model is described as follows:

$$\frac{\partial u_1}{\partial t} = \nabla \cdot (D_1 \nabla u_1) + \rho_1 u_1 \left(1 - \frac{u_1 + u_2}{K_1}\right), \quad (1)$$

$$\frac{\partial u_2}{\partial t} = \nabla \cdot (D_2 \nabla u_2) + \rho_2 u_2 \left(1 - \frac{u_1 + u_2}{K_2}\right) - \nabla \cdot (A_2 u_2), \quad (2)$$

where D_1 and D_2 are the diffusion coefficients, ρ_1 and ρ_2 are the proliferation rates, K_1 and K_2 are the carrying capacities for the first and second subpopulations, respectively, and $\vec{x} = (x, y, z)^T$. In addition, for the second subpopulation, we include an advection term with advection coefficient A_2 . We do not consider density-dependent or other non-linear effects in this work, instead we assume that all coefficients are positive real numbers. Cell migration due to diffusion influences both populations, however, only the second population is assumed to be influenced by extrinsic microenvironmental factors that result in the advection term. We want to note that the advection term has been previously used to model the cell migration in GBM^{23,26}. Although the sources of these factors are unknown, it has been suggested that it could be due to cells getting attracted by nutrients, oxygen, etc.²⁶, or other biases causing directed motion. The growth terms are governed by the logistic growth with the inclusion of competition between cells. The aggregated cell density is the sum of the two subpopulation densities: $u(r, t) = u_1(r, t) + u_2(r, t)$. Equations (1–2) describe the time evolution of the two cell subpopulations in three-dimensional space \vec{x} however, our processed data is the one-dimensional averaged cell density as a function of the radial distance from the spheroid center, r . To address this discrepancy, we simulate eqs (3–4) using radially-symmetric spherical coordinates

$$\frac{\partial u_1}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 D_1 \frac{\partial u_1}{\partial r} \right) + \rho_1 u_1 \left(1 - \frac{u_1 + u_2}{K_1}\right), \quad (3)$$

$$\frac{\partial u_2}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 D_2 \frac{\partial u_2}{\partial r} \right) + \rho_2 u_2 \left(1 - \frac{u_1 + u_2}{K_2}\right) - \frac{\partial}{\partial r} (A_2 u_2). \quad (4)$$

The initial conditions, $u_{10}(r)$ and $u_{20}(r)$, for the subpopulations are unknown, however, we observe the total initial cell population $u_0(r)$. Therefore, we assume that $u_{10}(r)$ and $u_{20}(r)$ are proportional to the initial total population: $u_{10}(r) = \alpha u_0(r)$ and $u_{20}(r) = (1 - \alpha) u_0(r)$, where α is a proportional constant parameter that can be estimated and $u_0(r)$ is the cell density data at the initial time point. Similar to previous studies^{24,42}, the boundary conditions are taken to be no-flux condition for each

subpopulation:

$$\begin{aligned} \frac{\partial u_1}{\partial r}(0, t) &= \frac{\partial u_1}{\partial r}(L, t) = 0, \\ \frac{\partial u_2}{\partial r}(0, t) &= \frac{\partial u_2}{\partial r}(L, t) = 0. \end{aligned}$$

Throughout this study, the spatial domain is chosen to be $0 \leq r \leq 5$ mm, and the temporal domain to be $0 \leq t \leq 1$ weeks, as is dictated by the multicellular tumor spheroid experiments conducted.

In addition to the RD-ARD model, Table 1 contains the other three models we test in this study. Note that we compute these three-dimensional models in spherical coordinates to facilitate their comparison to the one-dimensional data. The Reaction-diffusion (RD) model, which has been found to be clinically relevant^{16,18,21,25}, assumes that there is a single population of cells which undergoes diffusion and proliferation. The Advection-reaction-diffusion (ARD) model assumes that a single population of cells undergoes advection, diffusion, and proliferation. The reaction-diffusion and reaction-diffusion (RD-RD, or 2-population RD) model describes two populations of cells which both undergo diffusion proliferation; each population has their own parameters for these actions. The initial and boundary conditions for the RD-RD model are equivalent to those presented for the RD-ARD model. For the RD and ARD models, the initial condition is the initial total population density observed in the data, and the boundary condition is similar to that of the RD-ARD model (i.e., a no-flux condition for the homogeneous population).

Population heterogeneity and advection improves model fit

We estimated the parameters for all four models against a total of 136 datasets from 18 different patient-derived cell lines. The full description of the parameter estimation procedure can be found in the “Parameter estimation method” section. The best-fit simulations from all models are plotted against the first replicate of cell line U3013MG (Fig. 2, see Supplementary Figures 1–4 for additional representative results for other cell lines). This figure shows that the RD-ARD model best captures the dynamics in the data, especially the observed “hump”, i.e., the invasive front of cells at cell densities around 0.1. On the other hand, other models, especially the RD and ARD models, fail to capture this. Similar observations can be found for other replicates of cell line U3013MG (see GitHub repository provided in the Code availability statement). We summarized the performance of all four models in describing each dataset by computing their sum of squares error (SSE) and Akaike information criterion (AIC) scores (Fig. 3). The RD-ARD model achieves the lowest SSE and AIC values in most cases compared to the other models. We thus find that the RD-ARD model outperforms the other models, as is demonstrated by having a lower AIC score despite being penalized for having more parameters than the other models.

Wave front speeds approximate population heterogeneity

A property of GBM spheroids is that the core and the invasive region do not expand at equal speeds. This has been observed in previous studies²⁶, and in our dataset as well. This behavior is not described by the RD model, but is indeed captured by the RD-ARD model we propose. In order to study this behavior we propose a method (complete description in the “Traveling wave estimation” section) for calculating wave speed values at different cell densities, which in turn characterizes the spatial heterogeneity in the cell population. In short, we numerically calculate two different wave propagation speeds at two distinct cell densities. The corresponding wave speeds are denoted by c_{\min} and c_{\max} . As we will show, these together with the difference $c_{\text{diff}} = c_{\max} - c_{\min}$ provide insights into the heterogeneity in cell population, or in other words, the deviation away from a homogeneous single population Fisher-KPP model. We also define the parameter c_{shape} to be the slope of the line passing through the densities corresponding to c_{\max} and c_{\min} .

We first sought to validate our proposed method for calculating wave speed values to characterize spatial heterogeneity using simulated

Fig. 2 | A visualization of fitting results using 4 models. **a** RD, **b** ARD, **c** RD-RD (or 2-population RD), and **d** RD-ARD models for the first replicate of cell line U3013MG. We plot the total cell density data against the location for different time point as dotted color curves. The solid color curves represents the model fittings. Time is measured in weeks.

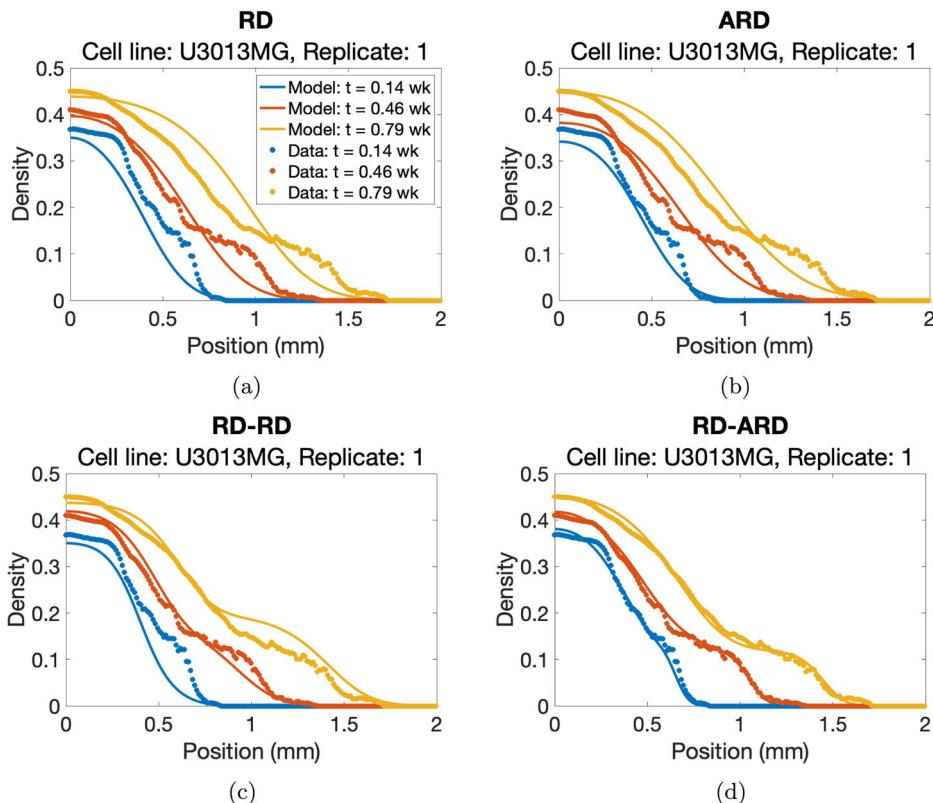
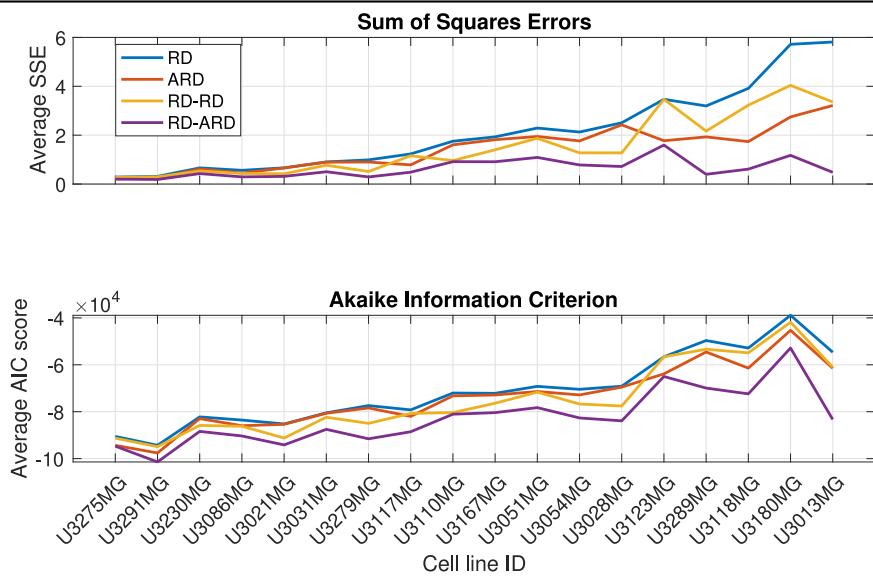


Fig. 3 | Model performance comparison. Model performance comparison using sum of squares error (top) and Akaike Information Criterion (bottom).



data. We compared the approximated wave speeds at two different cell densities, as representations for densities at c_{\max} and c_{\min} , for various cases from the RD-ARD model: (a) a spatially heterogeneous population with $D_2 = 10D_1$, $A_2 = 0.4$; (b) a spatially heterogeneous population with $D_2 = 10D_1$, $A_2 = 0$; (c) a spatially homogeneous population with $D_2 = D_1$, $A_2 = 0$; and (d) a homogeneous population (Fig. 4). The full details for the parameter values for all cases are described in Table 2. For the spatially homogeneous population with $D_2 = D_1$, $A_2 = 0$ in Fig. 5c, while both subpopulations have homogeneous migrating parameters (i.e., diffusion and advection coefficients are the same), their growth parameters are different. Thus, we refer to it as a spatially homogeneous population to distinguish it from the spatially heterogeneous populations in Fig. 4a, b

and the homogeneous population in Fig. 4d, where there is only one population. The results demonstrate substantial differences in the computed wave speeds at high density compared to low density for spatially heterogeneous populations with distinct migration parameter values. For instance, in the case of $D_2 = 10D_1$ and $A_2 = 0.4$, the approximated wave speed at density 0.4 is 0.2284, compared to 0.9985 at density 0.1. Similarly, for the spatially heterogeneous population with $D_2 = 10D_1$ and $A_2 = 0$, the approximated wave speeds are 0.2605 and 0.6137 at the respective densities. In contrast, for a spatially homogeneous population with identical migration parameter values ($D_2 = D_1$ and $A_2 = 0$), the approximated wave speeds are relatively close, as seen in Fig. 4c. The homogeneous population in Fig. 4d exhibits similar trends. The

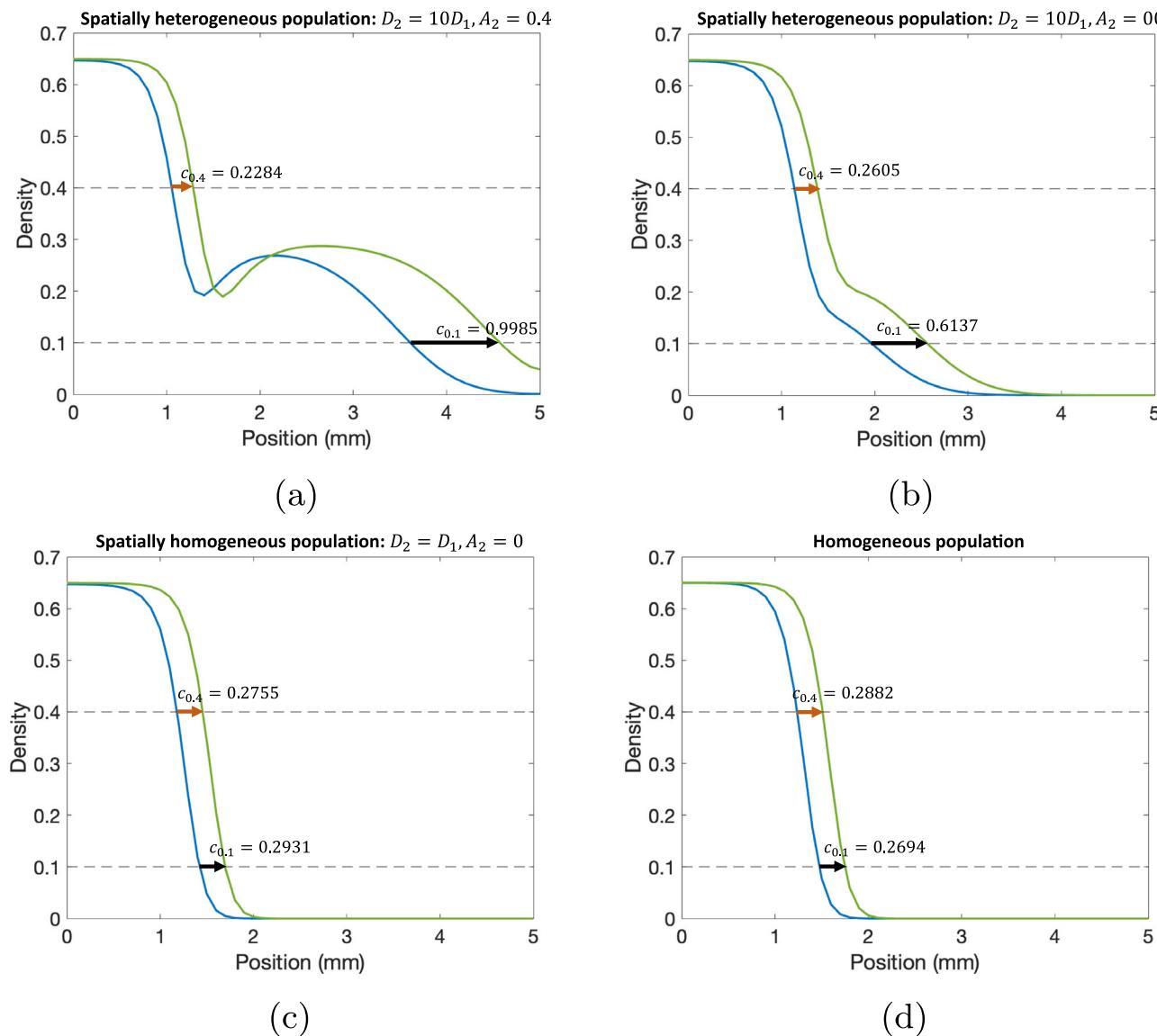


Fig. 4 | Visualization for wave front speeds at different densities for different scenarios. **a** A spatially heterogeneous population with $D_2 = 10D_1$ and $A_2 = 0.4$, **b** a spatially heterogeneous population with $D_2 = 10D_1$ and $A_2 = 0$, **c** a spatially homogeneous population with $D_2 = D_1$ and $A_2 = 0$, and **d** a homogeneous

population. In these figures, we compare the wave front speeds at two different time points: (blue) $t = 4.03$ wk and (green) $t = 5.00$ wk. All parameter values for each simulation are presented in Table 2.

computed c_{diff} values are: (a) $c_{\text{diff}} = 0.7701$, (b) $c_{\text{diff}} = 0.3532$, (c) $c_{\text{diff}} = 0.0176$, and (d) $c_{\text{diff}} = 0.0188$. These findings demonstrate that our proposed method, utilizing wave speed, effectively characterizes population spatial heterogeneity, particularly when subpopulations have distinct migration parameter values, i.e., D_1 , D_2 , and A_2 .

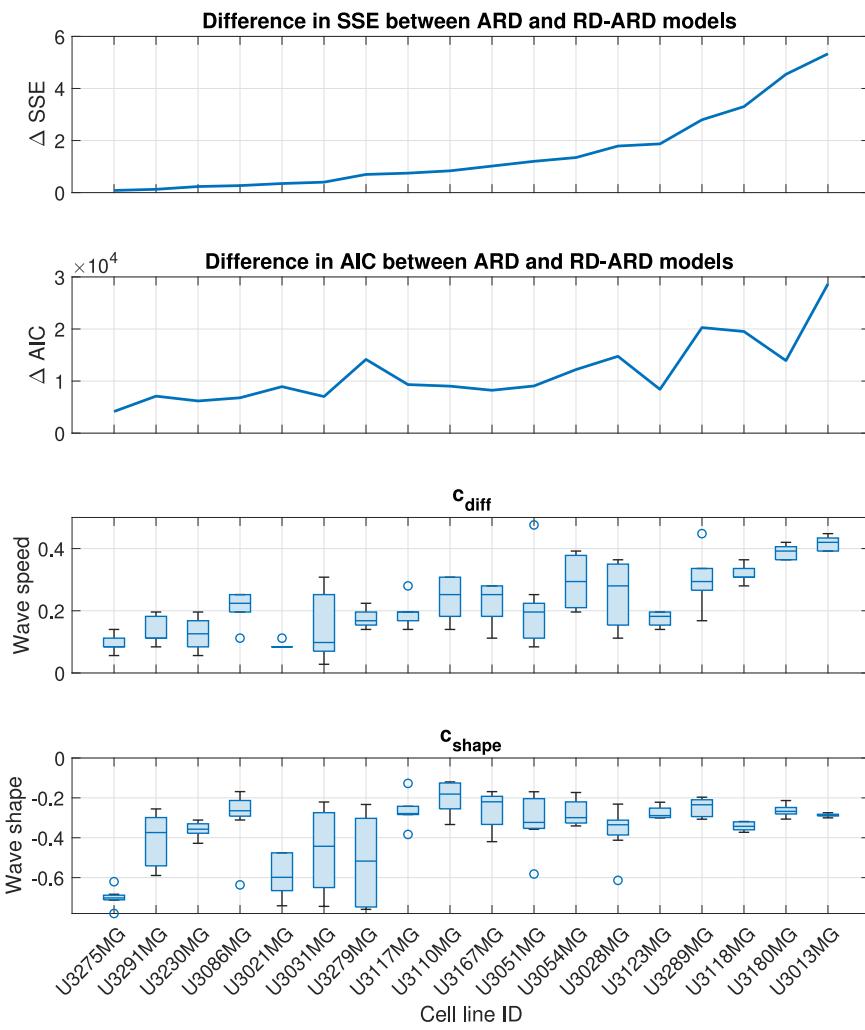
We applied our wave speed heterogeneity characterization approach to the best-fit simulations from each multicellular tumor spheroids experiment. Recall that we wish to determine whether c_{diff} is a good proxy for population spatial heterogeneity. To this end, we calculate the difference in SSE and AIC values between the ARD model and the RD-ARD model, which serve as representative models for spatially homogeneous and heterogeneous GBM populations, respectively. We use the difference in SSE and AIC values to represent spatial heterogeneity because these values quantify the added benefit of using a heterogeneous model to describe the data instead of a homogeneous model. We chose to use the ARD model, rather than the RD model, as the representative homogeneous model due to the lower SSE and AIC values of the ARD model compared to the RD model (refer to Fig. 3). In Fig. 5c, we present a boxplot illustrating c_{diff} . The general trend observed in Fig. 5

Table 2 | Parameter values used for simulation to validate wave speed as proxy for spatial heterogeneity

Parameter	Figure 4a	Figure 4b	Figure 4c	Figure 4d
D_1	0.007	0.007	0.007	0.007
D_2	0.07	0.07	0.007	0
ρ_1	2.5	2.5	2.5	2.5
ρ_2	1.5	1.5	1.5	0
K_1	0.65	0.65	0.65	0.65
K_2	0.4	0.4	0.4	0
A_2	0.4	0	0	0
α	0.5	0.5	0.5	1

across all three plots indicates an increase in spatial heterogeneity. This suggests that the difference in numerical wave speeds at different cell densities can effectively characterize spatial heterogeneity in GBM populations.

Fig. 5 | Visualization for the trends in ΔSSE , ΔAIC , c_{diff} and c_{shape} . The cell lines are in an increasing order of ΔSSE .



A subset of cell lines exhibit Go-or-Grow behavior

In order to better understand the similarities and differences between the cell lines, we performed hierarchical clustering using MATLAB (MATLAB's built-in function `linkage` with “average” method and “Standardized Euclidean distance” metric) on the learned model parameters for each cell line. The resulting dendrogram with 4 clusters is shown in Fig. 6: high-activity, “Go”, low-activity, and “Go-or-Grow” (see Supplementary Figures 1–4 for examples). We can see that in the first split, U3013MG constitutes the high-activity cluster, while all other cell lines make up the second group. This distinction arises from the unique nature of U3013MG within our experimental dataset. Within this cell line, the first population exhibits high diffusion and growth rates, while the second population has both low diffusion and growth rates but a high advection rate leading to the “hump” characteristic in Fig. 2. This sets the U3013MG cell line apart from the rest of the cell lines in our dataset. In the second group, we observe the separation between U3110MG (the “Go” cluster) and the two larger clusters. The U3110MG cell line is another unique case, where both subpopulations mainly focus on migrating without proliferating. This is because the cluster is characterized by high and similar diffusion rates and low and similar growth rates. Additionally, the advection rate for the second subpopulation is also low compared to others. This explains why the SSE and AIC curves overlap in Fig. 3 at the U3110MG cell line, and the difference in SSE between the ARD and RD-ARD models is relatively low (see Fig. 5). When looking at the next split that occurs when four clusters are present, we can see that the previous large cluster is split in such a way that U3180MG, U3289MG, U3054MG, U3118MG and U3051MG constitute the fourth cluster. By manually inspecting the learned model parameters (complete set of learner

parameters are provided in Supplementary Table 2) we notice two things. The first is that most cell lines in cluster three exhibit low activities with low diffusion, proliferation and advection. The second thing we notice is that the cell lines of cluster four have the property that one subpopulation is dominated by migration (low growth rate and high diffusion and advection rates) and the other subpopulation by growth (high growth rate, low diffusion rate), resulting in the “hump” characteristic. We therefore form the hypothesis that this cluster can be characterized as a “Go-or-Grow” cluster. We test our hypothesis by applying our Go-or-Grow characterization method (see the section “Go-or-Grow classification”) to the best-fit parameters from each cell line (Fig. 7). We found that a total of 6 of the 18 cell lines are classified as Go-or-Grow (U3051MG, U3054MG, U3117MG, U3118MG, U3180MG, U3289MG). Five of these exactly match those making up cluster four obtained from the hierarchical clustering. The only cell line we classified as Go-or-Grow that is missing from cluster four is U3117MG. These results support the notion that a subset of cell lines can be described as exhibiting Go-or-Grow behavior. We can see that the cell lines which were not classified as Go-or-Grow have other characteristic behaviors. For example, some cell lines have a “weak” Go-or-Grow phenotype, but the relationship is not as strong (i.e., the Go-or-Grow criteria would be satisfied for smaller k). This is the case for U3021MG, U3031MG, U3123MG, and U3167MG. In one case we saw both migration and proliferation of comparable magnitude in both subpopulations (U3110MG). An outlier to these classes of behavior is the cell line U3013MG. The first subpopulation is both diffusing and proliferating, while the second subpopulation is dominated by the advective term. This behavior is not seen in any of the other cell lines. It is worth noticing that this cell line is one where

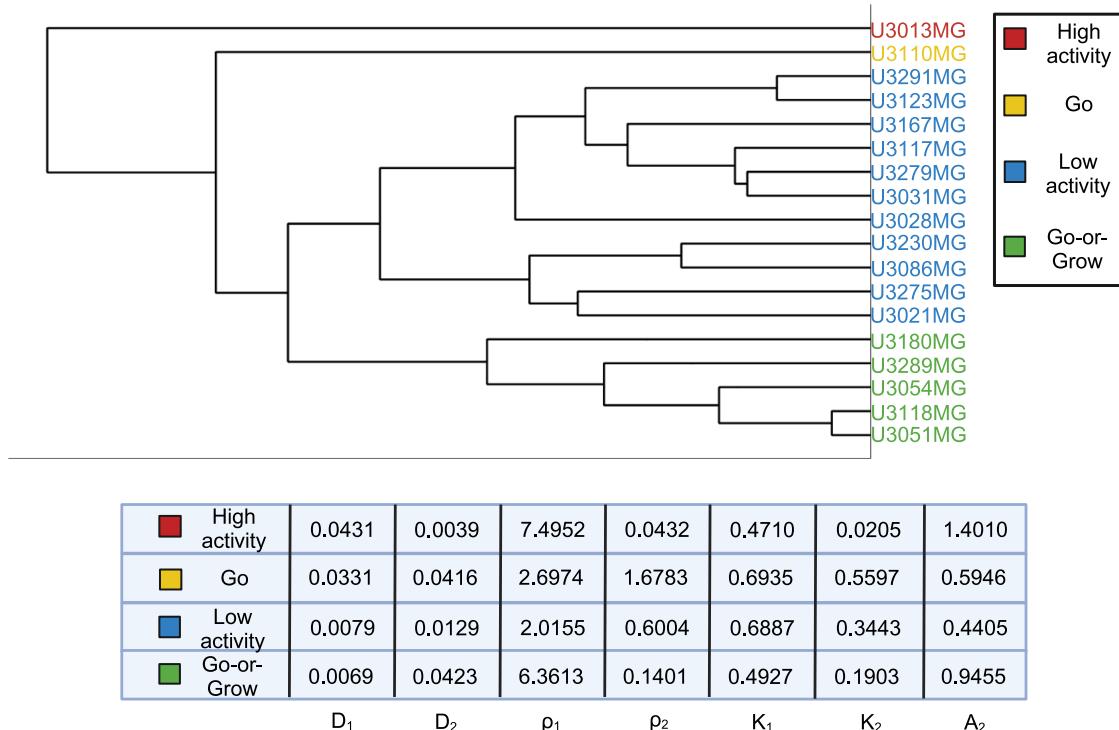


Fig. 6 | Dendrogram visualizing the outcome of the hierarchical clustering. Coloring based on 4 clusters. The names are descriptive and chosen based on the properties of the cell lines within each cluster. The table shows the parameter averages for each cluster.

Parameter table with Go-or-Grow Classification					
	D ₁	D ₂	ρ_1	ρ_2	A ₂
U3013MG	2.60	0.24	3.69	0.02	2.18
U3021MG	0.41	1.58	1.41	0.30	0.63
U3028MG	0.12	1.23	0.48	0.60	1.15
U3031MG	0.58	1.50	0.68	0.05	1.13
U3051MG	0.30	2.23	2.06	0.09	1.20
U3054MG	0.32	3.22	3.03	0.11	1.38
U3086MG	1.49	0.55	0.69	0.71	0.52
U3110MG	2.00	2.51	1.33	0.83	0.93
U3117MG	0.35	1.09	0.93	0.08	0.87
U3118MG	0.22	2.08	2.36	0.05	1.33
U3123MG	0.36	0.53	1.64	0.03	0.62
U3167MG	0.05	0.31	1.29	0.30	0.88
U3180MG	0.61	1.46	5.10	0.07	1.77
U3230MG	0.59	0.20	1.52	0.64	0.25
U3275MG	0.73	0.63	1.02	0.22	0.38
U3279MG	0.09	0.83	0.42	0.21	0.85
U3289MG	0.64	3.77	3.09	0.02	1.67
U3291MG	0.46	0.14	0.82	0.10	0.28

Fig. 7 | Parameter table with Go-or-Grow classification. Table of the fitted parameters after normalizing with respect to population averages, and the Go-or-Grow classification obtained from the criteria in (10) and (11).

we notice the largest improvement in model fit compared to a single population model, as can be seen in Fig. 3.

Correlation between patient outcome and mathematical model parameters

We proceed to investigate if the best-fit RD-ARD model parameters correlate with patient age and survival. Of the 18 patient-derived cell lines we used in our work, two (U3123MG, U3291MG) were excluded on the basis that the patients were still alive at the time of creating the Human Glioma Cell Culture (HGCC) biobank, and hence have no associated actual survival time.

We calculate the correlation between the parameters and patient age at diagnosis (row 1, Fig. 8), as well as between parameters and patient survival

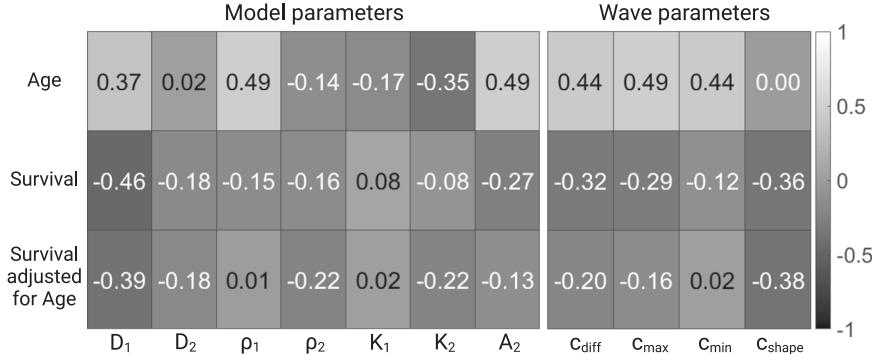
(row 2, Fig. 8). Since it has previously been established that survival is influenced by age at diagnosis^{43–45}, we also calculated the partial correlation between model parameters and survival, adjusting for age (row 3, Fig. 8). For three variables, A, B, and C, the correlation between A and B, controlling for C is calculated via:

$$\rho_{A,B,C} = \frac{\rho_{A,B} - \rho_{A,C}\rho_{B,C}}{\sqrt{1 - \rho_{A,C}^2}\sqrt{1 - \rho_{B,C}^2}}.$$

where $\rho_{A,B}$ is the correlation between A and B.

Since the number of patients is relatively small, it is difficult to interpret the results in a definite manner. However, since the correlation between age

Fig. 8 | Correlation between model parameters and patient age and survival. Correlation between model parameters and age (row 1), between model parameters and survival (row 2) and between model parameters and age-adjusted survival (row 3).



at diagnosis and survival is well established for GBM patients, we calculated it in our dataset, and found it to be -0.46 . This can be used as a reference to compare the other correlation coefficients against.

Our analysis reveals that the model parameters are generally most strongly correlated with patient age ($D_1, p_1, A_2, c_{\text{diff}}, c_{\text{max}}, c_{\text{min}}$). When looking at how the parameters are correlated with survival, both with and without adjusting for age, we see that D_1 is the most strongly correlated with survival, followed by c_{shape} .

Discussion

This study combined tumor spheroid experiments, microscopy imaging, mathematical modeling, and data analysis methods to develop and analyze a novel data-driven PDE model for in vitro tumor spheroid growth and invasion. We performed 136 total tumor spheroid experiments from 18 separate cell lines and quantified the radial spheroid density of each experiment over time using signal processing methods. To better interpret this data, we compared four mathematical models of tumor spheroid growth: the RD, ARD, RD-RD (or 2-population RD), and RD-ARD models. In the majority of cases the RD-ARD model outperformed the three other PDE models (quantified with the Akaike Information Criterion under the framework of least square) in describing data from the 18 cell lines. A traveling wave analysis characterized the spatial heterogeneity of the RD-ARD model and can be used to determine parameter values where this model behaves differently from the homogeneous models. A clustering analysis on the estimated parameters from the RD-ARD model allowed us to group together cell lines with similar parameter estimate values.

Population heterogeneity and advection are assumed in the RD-ARD model, allowing it to capture the “hump” characteristic observed in some of the cell lines, as shown in Fig. 2d. Figures 2 and 3 have demonstrated the importance of including not only the population heterogeneity but also the advection term when modeling multicellular tumor spheroids. A key feature of the RD-ARD model is that a portion of the spheroid population migrates by advection while the entire population also migrates by diffusion. Advection refers to the directed movement of cells; in our model, cells are moving away from the spheroid center. Many previous studies have proposed that spheroid cells perform chemotaxis, or migration towards increasing chemoattractant chemical gradient levels. Possible chemoattractants for GBM cells include oxygen, glucose, CXCL8, CXCL12, transforming growth factor, epidermal growth factor, etc. A chemoattractant gradient away from the spheroid center forms in response to cell consumption. The authors of ref. 46 used a one-compartment PDE with density-dependent diffusion and chemoattractant-dependent chemotaxis to model glioma growth and invasion. This model agreed with previous experiments (and those in our study) that the outer invasive zone grows faster than the inner proliferative zone of the spheroid. In future work, it would be interesting to compare the performances of the RD-ARD model and the model from ref. 46 in fitting the spheroid data from our study.

Previous studies on the traveling wave speeds of tumor growth and invasion relied on the PDE modeling framework with a single population, using the RD model, i.e., the Fisher-KPP equation. This assumption of a

single population allows for the analytical computation of the minimum traveling wave speed constant and the wave shape-related ratio. While these analytically computed wave parameters have been found to be clinically relevant, computing them for models with heterogeneous populations poses a challenge. Instead, we numerically computed the wave parameters and proposed that they can be used to characterize spatial heterogeneity in the population. We observed a general trend where the approximated wave parameters increase as the SSE and AIC values increase, indicating that these approximated wave parameters can potentially be used to characterize spatial heterogeneity.

The RD-ARD model assumes that one population migrates and proliferates, whereas the second population, in addition also exhibits motion due to advection. An advantage of this model is that it is very flexible and, as such, can be used to describe many different population behaviors. For example, it is possible that both subpopulations act similarly (e.g., both invade and grow fast or slowly) or distinct from each other (e.g., one primarily invades while the other primarily grows). Having defined a criteria for Go-or-Grow, we proposed that 6 of the 18 cell lines appear to exhibit the Grow-or-Go behavior, in which one subpopulation primarily grows while the other predominantly invades. Other cell lines exhibited a weaker Grow-or-Go phenotype or similar behavior between the two subpopulations. These vast differences in parameter estimates between cell lines highlights the wide variation in MCTS dynamics from our data. This indicates the need for precision medicine tools to determine the most effective treatment strategies for each type of population behavior found with our modeling framework. The parameter c_{shape} quantifies the shape of the invasive front. A smaller number indicates a more diffuse spheroid, and a larger value a spheroid with a sharp boundary. This parameter is obtained numerically and plays a role similar to that of the ratio D/ρ from the Fisher-KPP equation.

Our correlation analysis showed that a number of model parameters are correlated with patient age at diagnosis, and D_1 and to a lesser extent c_{shape} , with patient survival. Although one always wishes for stronger correlations, we find the results encouraging. The fact that we found a correlation between D_1 and survival, similar in magnitude to that of age and survival, is encouraging and warrants further research with a larger number of patient-derived cell lines. The reason for why D_1 shows the strongest association with survival is a matter we can only speculate around. One possible hypothesis is that since it is the second subpopulation that is most often driven by diffusion, having a large D_1 would correspond to a strongly invasive tumor spheroid, invading since both subpopulations (the entire tumor) have high motility.

In future work, we will investigate whether combining the modeling framework for MCTSs described here can provide an additional informative dimension to predicting the outcome of drug screens or interpreting gene expression data. For example, the estimated parameters from the RD-ARD model, simple metrics such as spheroid diameter as a proxy for growth^{14,47}, and information from in vitro assays probing cytotoxicity⁴⁸ or apoptosis⁴⁹, can be concatenated into a feature vector for each patient-derived tumor cell line in order to predict drug efficacy. Alternatively, the estimated parameters

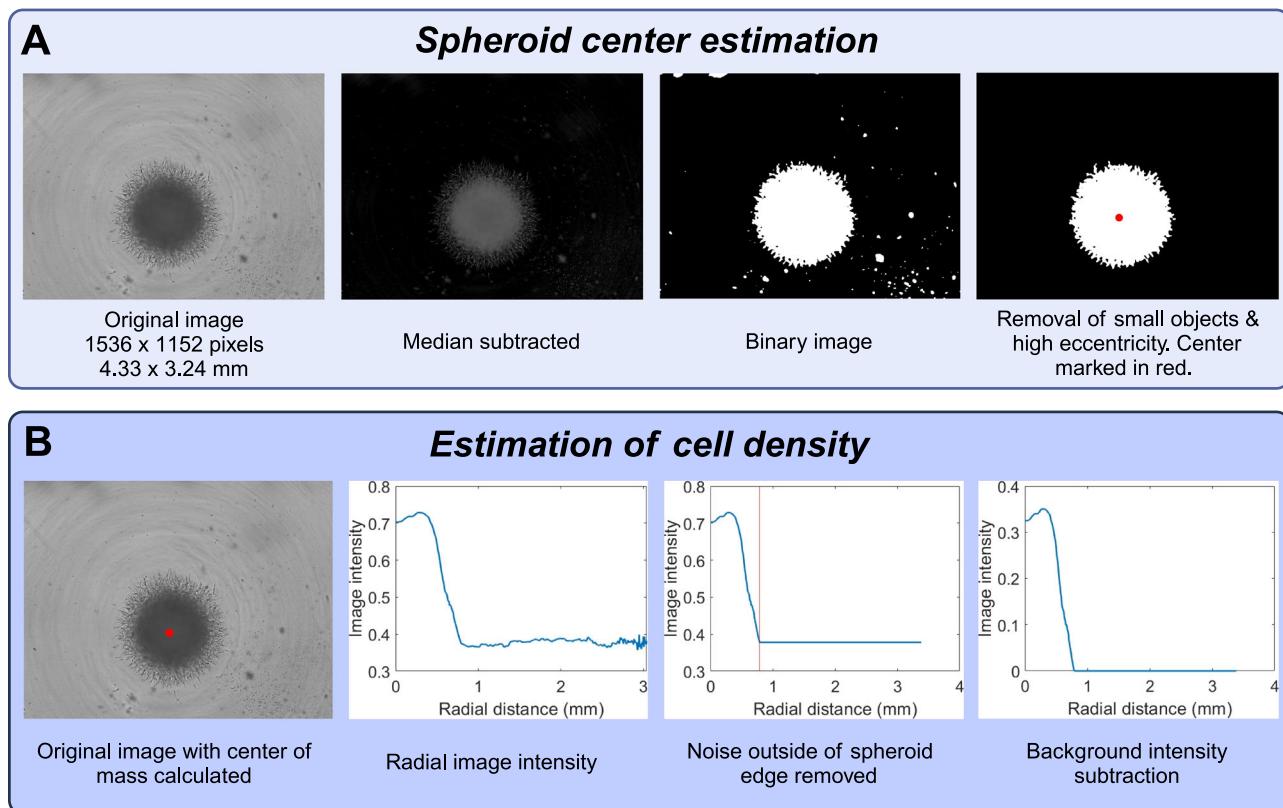


Fig. 9 | Data processing pipeline. The first step (A) is to estimate the location of the spheroid center. The second step (B) is to estimate cell density as a function of radial distance from the spheroid center. This figure was created with BioRender.com.

from the RD-ARD model for a large set of patient-derived cell lines could be correlated with gene expression differences to provide hypotheses about which intracellular signaling pathways regulate phenotypes that affect spheroid dynamics, e.g., growth, diffusion, and advection. Other possible directions for future work include utilizing deep learning for data de-noising or automated model selection. In this study, we chose to use simulations generated by the optimized model parameters to compute the wave parameters due to the noise in the experimental data. For the future work, we propose employing neural networks for the data de-noising task. Neural networks have demonstrated superior performance in the presence of biologically realistic observation noise for the Fisher-KPP model⁵⁰. This could potentially enable the computation of wave parameters without the need for parameter estimation and model simulation. Moreover, equation learning methods relying on neural networks have been implemented for automated model discovery in spatial-temporal data of reaction-diffusion equations^{51,52} and mean-field approximations of agent-based models for cell populations incorporating birth, death, and migration processes²². Despite the RD-ARD model exhibiting the best performance among the models discussed in this study, we observed that there still exists some model discrepancy for some cell lines. Equation learning could potentially be used to correct model discrepancy in these instances, enabling the ability to discover novel biophysical mechanisms governing the underlying dynamics of longitudinal MCTSs data.

Methods

Spheroid invasion assay

Patient-derived glioma cell cultures (GCCs) were chosen from the Human Glioma Cell Culture (HGCC) collection⁵³. Each cell line was seeded into eight wells each of 96-well PrimeSurface S-BIO round bottom ultra-low attachment plates (MoBiTec), in defined serum-free neural stem cell (NSC) medium, supplemented with B-27, N2, EGF, and FGF as previously described⁵⁴. A 50% matrigel (Corning, 356232, ≥8 mg/ml) overlay in defined

NSC medium was added on day three and the spheres were allowed to invade into the extracellular matrix for six additional days. Starting from day three, images were acquired in the phase contrast and brightfield channels with the 4x objective every six hours, on an IncuCyte®S3 live-cell imaging system (Sartorius) using the spheroid software module version 2019B. The subtype classification and genomic profiles of all cell cultures used in this study have been previously described⁵⁵. A total of 25 cell lines were originally included in the experiment, however, six were excluded because of lack of growth, and an additional cell line was excluded due to large amounts of noise when calculating the density profiles. The 18 cell lines, and the corresponding number of replicates for each can be found in Supplementary Table 1.

Data processing

In order to estimate cell density profiles from the brightfield images (of size 1536 x 1152 pixels, and 4.33 x 3.24 mm), we developed custom software in MATLAB. The goal is to convert 2D images of spheroids to radial density profiles. The data processing therefore consists of two distinct problems: (i) finding the center of mass of the spheroid, (ii) converting image intensity to cell density. The major steps are shown in Fig. 9.

The process of estimating the center of mass is illustrated in Fig. 9 and consists of the following steps:

1. Background subtraction,
2. Binarization, and
3. Removal of small objects.

We then proceed to estimate cell density as a function of radial distance to the spheroid center. This process consists of the following steps.

1. Calculation of image intensity as a function of radial distance to the spheroid core,
2. Removing noise outside the spheroid boundary, and

Table 3 | Parameter descriptions and ranges

Parameter	Description	Range	Unit
D_1	Population 1 diffusion coefficient	0–0.2	mm^2/week
D_2	Population 2 diffusion coefficient	0–0.2	mm^2/week
ρ_1	Population 1 growth rate	0–15	week^{-1}
ρ_2	Population 2 growth rate	0–15	week^{-1}
K_1	Population 1 carrying capacity	0–1	cells/mm^3
K_2	Population 2 carrying capacity	0–1	cells/mm^3
A_2	Population 2 advection rate	0–3	mm/week
α	Proportional of initial condition belonging to population 1	0–1	Unitless

3. Background subtraction.

Note that the intensity profiles obtained after processing are referred to as cell densities. These are not normalized to have values in any particular range. We found that they typically range between 0 and 0.5. The best case would be to be able to obtain the actual cell number, but this is a challenging problem, and to the best of our knowledge, is not possible to do from 2D projected images of 3D cultures.

Algorithm 1. Pseudocode for estimation of spheroid center of mass

```

Input image  $im_0$ ;
 $m = \text{median}(im_0)$ ;
 $im_{\text{subtracted}} = im_0 - m$ ;
 $im_{\text{smoothed}} = \text{imgaussfilt}(im_{\text{subtracted}}, 4)$ ;
 $im_{\text{binary}} = \text{im2bw}(im_{\text{smoothed}}, 0.08)$ ;
 $ConnComp = \text{bwconncomp}(im_{\text{binary}})$ ;
for  $object$  in  $ConnComp$  do
| If  $object$  has area < 4000 pixels2, remove it from  $im_{\text{binary}}$ ;
end
Delete all objects in  $im_{\text{binary}}$  except the one with the smallest
eccentricity (most round shape);
Find all rows  $r$  and columns  $c$  where  $im_{\text{binary}} = 1$ ;
Center of mass = ( $\text{mean}(r)$ ,  $\text{mean}(c)$ );

```

The function names written in bold are built-in MATLAB functions. We first subtract the median intensity from the original image, as a form of background subtraction. The function $\text{imgaussfilt}(im_{\text{intensity}}, 4)$ applies a Gaussian filter with standard deviation 4. The function $\text{im2bw}(im_{\text{smoothed}}, 0.08)$ binarizes the image based on the threshold 0.08. The function $\text{bwconncomp}(im_{\text{binary}})$ finds the connected components (i.e., distinct cohesive objects) in the binary image. We then remove objects smaller than 4000 pixels². This corresponds to an area of around 32,000 μm^2 . For reference, this is half the size of a typical spheroid before it starts growing. Having deleted small objects, we continue to delete all remaining objects except for the one with lowest eccentricity (most rounded shape). This step is performed due to certain images having large objects from the background noise remaining after the previous steps. However, they are typically not rounded, and can hence be removed by this step. Having a single rounded object remaining in the image (the spheroid) we compute the center of mass by taking the mean of the non-zero row indices and the non-zero column indices.

Algorithm 2. Pseudocode for estimation of radial cell density

```

Input image  $im_0$  and center of mass of the spheroid;
 $im_{\text{intensity}} = 1 - im_0$ ;
for each pixel in  $im_{\text{intensity}}$  do
| Calculate distance between  $pixel$  and spheroid center of mass;
end

```

Discretize the space into 600 bins of size 2 pixels. Put all pixels into their respective bins. Calculate average intensity for each bin. We obtain an array y_i , $i = 1, \dots, 600$, of length 600, with values between 0 and 1;

Calculate simple moving average y_{avg} of the density y ;

Estimate the boundary of the spheroid by finding the first point i^* where the derivative of the moving average changes sign; Set the intensity to be constant after i . $y_j = y_{i^*}$, for $j \geq i^*$; Remove background intensity by subtracting $y = y - y_{i^*}$.

The discretization into bins of size 2 pixels helps to reduce noise. The maximum distance of 1200 pixels (600 bins) was chosen to guarantee that the entire spheroid was captured. In order to detect the spheroid boundary we found the first inflection point where the derivative changed from negative to positive. This can be challenging when the data is noisy, and for that reason we took two additional measures to ensure its accuracy. First we used the simple moving average (100 bins on either side of the given point), which handles noise much better. Second, we look for points to the right of bin 100, which ensures that we ignore points inside the tumor core. The final steps of the processing is to set the intensity to be constant from the spheroid boundary and outward, and then to subtract it from the density, so that the surrounding region has density 0. The resulting image intensities are the ones we use as a proxy for cell density throughout our paper. They typically attain values between 0 and 0.5, but are not normalized to lie in any particular range.

Parameter estimation method

In order to perform parameter estimation, we first need to numerically solve a PDE model to approximate the true model solution. In this work, we implement central differencing to approximate the first-order and second-order derivatives in space. We then use the MATLAB built-in function `ode15s` to numerically integrate the model. During the parameter estimation step, the algorithm minimizes the sum of squares error (SSE) to find the most optimal set of parameters, $\hat{\mathbf{q}}$. The optimized parameters are determined from

$$\hat{\mathbf{q}} = \arg \min_{\mathbf{q} \in \mathcal{Q}} \sum_{i,j}^{N_t, N_r} \left[u_{i,j}^o - \hat{u}(t_i, r_j; \mathbf{q}) \right]^2, \quad (5)$$

where $u_{i,j}^o$ and $\hat{u}(t_i, r_j; \mathbf{q})$ are the observed and simulated total cell density at the i th temporal point and j th spatial radial location. In addition, N_t and N_r are the total number of temporal points and spatial locations in the data. \mathcal{Q} denotes the set of admissible values for the model parameters with the parameter bounds being shown in Table 3 along with model parameters information. We select the parameter ranges for the diffusion coefficient and growth rate based on the patient-reported values in the previous study¹⁶. We converted Wang et al. estimated values¹⁶ from mm^2/year and year^{-1} to mm^2/week and week^{-1} . We then choose the appropriate upper bounds to be within an order of magnitude of their estimated values. Similar values of D and ρ in mm^2/year and year^{-1} also have been used to learn equations of simulated data generated using the Fisher-KPP equation with one dimensional spatial data⁵².

As a result of the non-linearity in the proposed model, using a local optimization algorithm might not guarantee the most optimal results. To perform parameter estimation, we first implement a global optimization algorithm called the DIRECT (DIviding RECTangles) algorithm⁵⁶. DIRECT is a derivative-free global optimization algorithm that works by dividing all hyper-rectangles of parameter space into smaller hyper-rectangles to find the optimal hyper-rectangle, that is, the hyper-rectangle with the smallest error at the center. While DIRECT is fast to approach the area surrounding the global minimum, it can be slow to converge to the global minimum with high accuracy⁵⁷. In this work, we use the DIRECT algorithm as the initial step for parameter estimation. In the second step, we use the optimized parameters estimated by the DIRECT algorithm as the initial starting values for a local optimization algorithm. The DIRECT algorithm is implemented using a MATLAB package provided in ref. 58. For the local algorithm, we use `fmincon`, a MATLAB built-in function, with `interior-point` as the algorithm.

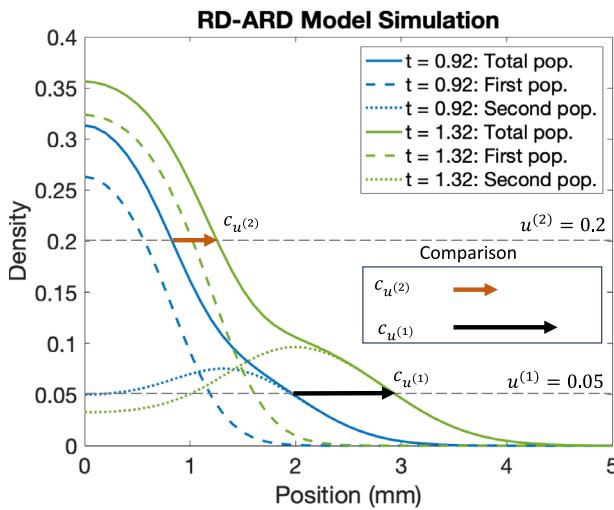


Fig. 10 | A representative figure for comparing wave speeds at different densities. The dashed blue and green curves are the cell density for the first population at two different time points, t_i (blue) and t_f (green) with $t_i < t_f$. Similarly, the dotted blue and green curves are the cell density for the second population. The solid blue and green curves are the total cell density. The orange and black arrows represent the wave speed vectors at $u^{(1)} = 0.05$ and $u^{(2)} = 0.2$. As being shown, the front density at 0.2 travels at slower speed compared to the front density 0.05. This shows that the speed of the density profile depends on the chosen density threshold, $u^{(i)}$, utilized in calculating the wave speed.

Model comparison

To appropriately assess model performance, the use of the SSE for comparison might not be sufficient, as it fails to consider the inherent complexity of the models. Since the RD-ARD model is the most complex model, i.e., having more parameters than the other models, it is expected to outperform the other models with the SSE metric. To account for the complexity of the model, we rely on the Akaike Information Criteria (AIC)^{59,60}, which penalizes complex models with more parameters. The AIC under the framework of least squares problem was developed by Banks and Joyner⁶⁰ and is described as follows:

$$AIC = N_t N_r [\ln(2\pi) + 1] + N_t N_r \ln\left(\frac{SSE}{N_t N_r}\right) + 2(\kappa_q + 1), \quad (6)$$

where N_t and N_r are the total number of temporal and spatial points, SSE is the sum of squares error described as a sum in Eq. (5), and κ_q is the total number of parameters being estimated.

Traveling wave estimation

Many previous PDE frameworks utilized single compartment PDE models, such as the Fisher-KPP equation, to describe tumor spheroid growth and invasion. A common pattern that arises in these models is a traveling wave, or a stationary spatial profile that moves at a fixed speed (referred to as the wave speed). For example, the diffusion coefficient D and the growth rate ρ can be approximated by identifying the minimum traveling wave speed constant, $2\sqrt{D\rho}$, and the wave shape related ratio (or relative invasiveness ratio), D/ρ ^{16–18,21}. These parameters are then used to classify patient tumors into different categories: slow (low D and low ρ), diffuse (high D and low ρ), nodular (low D and high ρ), and fast (high D and high ρ)^{16–18,21,52}. While this approach has been found to be clinically relevant^{16–18}, it is based on the assumption of intratumoral homogeneity. In addition, MCTSs comprised of GBM cells often have a core region and an invasive region where cells appear to exhibit different behavior. In particular, the core region expands at a slower speed compared to the invasive region²⁶. Single compartment models which admits a traveling wave solution fail to capture this behavior. On the other hand, two-compartment models lacking traveling wave

solutions can be analyzed by tracking the propagation speed at different cell densities, which is what we do in this study.

We choose two total cell density values $u^{(1)}$ and $u^{(2)}$, and numerically (following refs. 23,24) calculate the speed with which both values propagate. Throughout the paper we will refer to these as wave speeds, even though they are not technically traveling waves. In models admitting traveling waves, the two wave speeds are equal, whereas in spatially heterogeneous two-compartment models they generally are not. We propose that the difference in propagation speeds at different cell densities is a good proxy for intra-tumor spatial phenotypical heterogeneity.

Algorithm 3. Numerically approximated wave speed computation algorithm

```

Choose the density value for  $u^*$  ;
Choose the starting time index  $j_{\min}$  to compute the wave speed ;
for  $j = j_{\min}, \dots, N_t$  do
    Find the location,  $r^*$  where the density is approximately equal to  $u^*$ :
    
$$r^* = \arg \min |u^* - u(j, :)| ;$$

     $r_{\text{tracker}}(j) = r^* ;$ 
end
Compute the slope of the curve,  $r_{\text{tracker}}$ :
 $c = \text{polyfit}(t(j_{\min}, \dots, N_t), r_{\text{tracker}}) ;$ 

```

An example simulation of a spatially heterogeneous population is shown in Fig. 10, and illustrates the idea behind our proposed approach, for $u^{(1)} = 0.05$ and $u^{(2)} = 0.2$. In Fig. 10, we observe that the wave speed for $u^{(1)}$ is greater than the wave speed for $u^{(2)}$. The difference in wave speed is due to how the higher density profile is primarily affected by the first population, whereas the lower density profile is influenced by both the first and second populations. This indicates that, the front of density moves faster where the density is low compared to where it is high. In other words, in a spatially heterogeneous population, the speed of the wavefront changes depending on the specific density threshold, $u^{(i)}$, being used to compute the wave speed.

While in Fig. 10, we choose to visualize the difference in wave speeds at $u^{(1)} = 0.05$ and $u^{(2)} = 0.2$, in practice, it is challenging to predetermine the exact density values for comparison, because cell lines exhibit different density profiles. This means a cell lines might have different maximum density values and densities value at the “hump” characteristic. Therefore, we compute the wave speed at different cell densities and compare the difference between the maximum and minimum computed wave speeds. Let $\{u^{(1)}, u^{(2)}, \dots, u^{(M-1)}, u^{(M)}\}$ be the set of predetermined cell densities where we compute the wave speeds. We then define

$$\begin{aligned} c_{\max} &= \max\{c_1, \dots, c_M\}, \\ c_{\min} &= \min\{c_1, \dots, c_M\}, \\ c_{\text{diff}} &= c_{\max} - c_{\min}, \end{aligned}$$

where c_1, \dots, c_M are the wave speeds at cell densities $u^{(1)}, \dots, u^{(M)}$, respectively. In addition, another challenge that could arise is to decide which the time points for which we compute the wave speeds. We choose to compute the wave speed using a linear regression-based method over all time points, as suggested in previous studies^{23,24}. For each density value $u^{(i)}$, we track the location x where the density is approximately equal to $u^{(i)}$ over time. The approximated wave speed at total density $u^{(i)}$ is the slope of the curve formed by the tracked r values over time. We use the MATLAB function `polyfit` to compute the slope of the best fit line between time and the location of $u^{(i)}$ over time. In other words, for each total density threshold $u^{(i)}$, we track the list of $t^{(i)}$ and $r^{(i)}$ such that $u(t^{(i)}, r^{(i)}) = u^{(i)}$. The curve formed by the tracked points $(t^{(i)}, r^{(i)})$ is then fitted using a linear line with the form $r = c_i t + b$, where c_i is the approximated wave speed at $u^{(i)}$. The pseudocode for the numerically approximated wave speed computation is described in

Algorithm 3. We also define the parameter c_{shape} to be the slope of the line passing through the densities corresponding to c_{\max} and c_{\min} at the final time point. We will refer to these four parameters as wave parameters throughout this paper.

In a study conducted by Stepien et al., a numerical method was utilized to calculate the wave speed for a model that admits a stable traveling wave solution²³. The stability of the traveling wave solution is reflected by a wave profile that maintains a constant speed once the tumor core has stabilized at its maximum cell density. Consequently, after reaching this maximum density, the wave speed can be determined numerically from the solution. Because the collected data did not reach the maximum cell density during the data collection process, we simulated the RD-ARD model with the estimated parameters from each dataset and utilized the cell density data from $t = 0.75t_f$ to $t = t_f$ where t_f is the final time value to calculate the wave speed for each cell line. We computed the wave speeds at 100 cell density values ranging from 0.02 to $0.8u_{\max}$ for each data set, with u_{\max} being the maximum density value in the data set. The maximum and minimum of the approximated wave speeds were then used for comparison. The reason to compute the wave parameters simulations was driven by the experimental data's inherent noise, which could significantly influence computed results. Simulations enable us to denoise the data before proceeding with the computation of wave parameters.

Go-or-Grow classification

Our RD-ARD model can be regarded as a Go-or-Grow type model, whenever one subpopulation is predominantly migrating and the other proliferating. In a model of RD-RD (or 2-population RD) type where no advection is present, one can check if the diffusion coefficient of one subpopulation is significantly larger than that of the other subpopulation, while the roles of the proliferation terms are reversed. For our model we need to take the advection term into account, since it can contribute to the motility of the second subpopulation. There is no straightforward way to compare the motility resulting from advection to that resulting from diffusion. We chose to do it by normalizing each coefficient with the population-average, rendering each coefficient dimensionless. We first calculate the population averaged coefficients:

$$\begin{aligned}\bar{D} &= \sum_{i=1}^{18} \frac{D_1^i + D_2^i}{2}, \\ \bar{\rho} &= \sum_{i=1}^{18} \frac{\rho_1^i + \rho_2^i}{2}, \\ \bar{A} &= \sum_{i=1}^{18} A_2^i.\end{aligned}$$

For cell line i , with fitted model parameters $\{D_1^i, D_2^i, \rho_1^i, \rho_2^i, K_1^i, K_2^i, A_2^i\}$ we next normalize the fitted coefficients with the population averages:

$$\bar{D}_1^i = \frac{D_1^i}{\bar{D}}, \quad \bar{D}_2^i = \frac{D_2^i}{\bar{D}}, \quad (7)$$

$$\bar{\rho}_1^i = \frac{\rho_1^i}{\bar{\rho}}, \quad \bar{\rho}_2^i = \frac{\rho_2^i}{\bar{\rho}}, \quad (8)$$

$$\bar{A}_2^i = \frac{A_2^i}{\bar{A}}. \quad (9)$$

A model is classified as being of Go-or-Grow type if it satisfies one of the conditions:

$$\bar{D}_1 > k \cdot (\bar{D}_2 + \bar{A}_2) \text{ and } \bar{\rho}_2 > k \cdot \bar{\rho}_1, \quad (10)$$

$$(\bar{D}_2 + \bar{A}_2) > k \cdot \bar{D}_1 \text{ and } \bar{\rho}_1 > k \cdot \bar{\rho}_2, \quad (11)$$

where k is a multiplicative factor that governs how much larger the migration and proliferation of one population must be, relative to the other population. The first condition ensures that the first subpopulation is mainly migrating and that the second subpopulation is mainly proliferating, and the second condition reverses the roles for the two subpopulations. We use $k = 5$ as a criteria, meaning that the rescaled migration and proliferation is at least five times larger in one subpopulation compared to the other.

Data availability

The raw data (cell density profiles) can be made available upon request.

Code availability

The code is publicly available at: <https://github.com/kcnguyen3191/Modeling-Tumor-Spheroids>.

Received: 2 February 2024; Accepted: 10 January 2025;

Published online: 15 February 2025

References

1. Anjum, K. et al. Current status and future therapeutic perspectives of glioblastoma multiforme (GBM) therapy: a review. *Biomed. Pharmacother.* **92**, 681–689 (2017).
2. Holland, E. C. Glioblastoma multiforme: the terminator. *Proc. Natl. Acad. Sci. USA* **97**, 6242–6244 (2000).
3. Perrin, S. L. et al. Glioblastoma heterogeneity and the tumour microenvironment: implications for preclinical research and development of new treatments. *Biochem. Soc. Trans.* **47**, 625–638 (2019).
4. Bush, N. A. O., Hervey-Jumper, S. L. & Berger, M. S. Management of glioblastoma, present and future. *World Neurosurg.* **131**, 328–338 (2019).
5. Mann, J., Ramakrishna, R., Magge, R. & Wernicke, A. G. Advances in radiotherapy for glioblastoma. *Front. Neurol.* **8**, 748 (2018).
6. Yu, M. W. & Quail, D. F. Immunotherapy for glioblastoma: current progress and challenge. *Front. Immunol.* **12**, 1637 (2021).
7. Mittal, S., Pradhan, S. & Srivastava, T. Recent advances in targeted therapy for glioblastoma. *Expert Rev. Neurother.* **15**, 935–946 (2015).
8. dos Santos Fernandes, G. F., Fernandes, B. C., Valente, V. & Dos Santos, J. L. Recent advances in the discovery of small molecules targeting glioblastoma. *Eur. J. Med. Chem.* **164**, 8–26 (2019).
9. Ghosh, D., Nandi, S. & Bhattacharjee, S. Combination therapy to checkmate glioblastoma: clinical challenges and advances. *Clin. Transl. Med.* **7**, 1–12 (2018).
10. Alfonso, J. C. L. et al. The biology and mathematical modelling of glioma invasion: a review. *J. R. Soc. Interface* **14**, 20170490 (2017).
11. Everett, R. et al. A tutorial review of mathematical techniques for quantifying tumor heterogeneity. *Math. Biosci. Eng.* **17**, 3660–3709 (2020).
12. Han, S. J., Kwon, S. & Kim, K. S. Challenges of applying multicellular tumor spheroids in preclinical phase. *Cancer Cell Int.* **21**, 1–19 (2021).
13. Tofani, L. B., Abriata, J. P., Luiz, M. T., Marchetti, J. M. & Swiech, K. Establishment and characterization of an in vitro 3D ovarian cancer model for drug screening assays. *Biotechnol. Prog.* **36**, e3034 (2020).
14. Zhang, X. et al. Development of an in vitro multicellular tumor spheroid model using microencapsulation and its application in anticancer drug screening and testing. *Biotechnol. Prog.* **21**, 1289–1296 (2005).
15. Swanson, K. R., Alvord, E. C. & Murray, J. D. A quantitative model for differential motility of gliomas in grey and white matter. *Cell Prolif.* **33**, 317–329 (2000).
16. Wang, C. H. et al. Prognostic significance of growth kinetics in newly diagnosed glioblastomas revealed by combining serial imaging with a novel biomathematical model. *Cancer Res.* **69**, 9133–9140 (2009).
17. Neal, M. L. et al. Response classification based on a minimal model of glioblastoma growth is prognostic for clinical outcomes and

distinguishes progression from pseudoprogression. *Cancer Res.* **73**, 2976–2986 (2013).

18. Baldock, A. L. et al. Patient-specific metrics of invasiveness reveal significant prognostic benefit of resection in a predictable subset of gliomas. *PLoS ONE* **9**, e99057 (2014).
19. Le, M. et al. MRI based Bayesian personalization of a tumor growth model. *IEEE Trans. Med. Imaging* **35**, 2329–2339 (2016).
20. Rutter, E. M. et al. Mathematical analysis of glioma growth in a murine model. *Sci. Rep.* **7**, 2508 (2017).
21. Hawkins-Daarud, A., Johnston, S. K. & Swanson, K. R. Quantifying uncertainty and robustness in a biomathematical model-based patient-specific response metric for glioblastoma. *JCO Clin. Cancer Inform.* **3**, 1–8 (2019).
22. Nardini, J. T., Baker, R. E., Simpson, M. J. & Flores, K. B. Learning differential equation models from stochastic agent-based model simulations. *J. R. Soc. Interface* **18**, 20200987 (2021).
23. Stepien, T. L., Rutter, E. M. & Kuang, Y. A data-motivated density-dependent diffusion model of in vitro glioblastoma growth. *Math. Biosci. Eng.* **12**, 1157–1172 (2015).
24. Stepien, T. L., Rutter, E. M. & Kuang, Y. Traveling waves of a go-or-grow model of glioma growth. *SIAM J. Appl. Math.* **78**, 1778–1801 (2018).
25. Whitmire, P. et al. Sex-specific impact of patterns of imageable tumor growth on survival of primary glioblastoma patients. *BMC Cancer* **20**, 447 (2020).
26. Stein, A. M., Demuth, T., Mobley, D., Berens, M. & Sander, L. M. A mathematical model of glioblastoma tumor spheroid invasion in a three-dimensional in vitro experiment. *Biophys. J.* **92**, 356–365 (2007).
27. Kim, Y., Lawler, S., Nowicki, M. O., Chiocca, E. A. & Friedman, A. A mathematical model for pattern formation of glioma cells outside the tumor spheroid core. *J. Theor. Biol.* **260**, 359–371 (2009).
28. Martínez-González, A., Calvo, G. F., Pérez Romasanta, L. A. & Pérez-García, V. M. Hypoxic cell waves around necrotic cores in glioblastoma: a biomathematical model and its therapeutic implications. *Bull. Math. Biol.* **74**, 2875–2896 (2012).
29. Pham, K. et al. Density-dependent quiescence in glioma invasion: instability in a simple reaction-diffusion model for the migration/proliferation dichotomy. *J. Biol. Dyn.* **6 Suppl 1**, 54–71 (2012).
30. Alfonso, J. C. L. et al. Why one-size-fits-all vaso-modulatory interventions fail to control glioma invasion: in silico insights. *Sci. Rep.* **6**, 37283 (2016).
31. Han, L. et al. Patient-specific parameter estimates of glioblastoma multiforme growth dynamics from a model with explicit birth and death rates. *Math. Biosci. Eng.* **16**, 5307–5323 (2019).
32. Rutter, E. M., Banks, H. T. & Flores, K. B. Estimating intratumoral heterogeneity from spatiotemporal data. *J. Math. Biol.* **77**, 1999–2022 (2018).
33. Tursynkozha, A., Kashkynbayev, A., Shupeyeva, B., Rutter, E. M. & Kuang, Y. Traveling wave speed and profile of a “go or grow” glioblastoma multiforme model. *Commun. Nonlinear Sci. Numer. Simul.* **118**, 107008 (2023).
34. Nguyen, K., Rutter, E. M. & Flores, K. B. Estimation of parameter distributions for reaction-diffusion equations with competition using aggregate spatiotemporal data. *Bull. Math. Biol.* **85**, 62 (2023).
35. Tektonidis, M. et al. Identification of intrinsic in vitro cellular mechanisms for glioma invasion. *J. Theor. Biol.* **287**, 131–147 (2011).
36. Giese, A., Bjerkvig, R., Berens, M. & Westphal, M. Cost of migration: invasion of malignant gliomas and implications for treatment. *J. Clin. Oncol.* **21**, 1624–1636 (2003).
37. Dhruv, H. D. et al. Reciprocal activation of transcription factors underlies the dichotomy between proliferation and invasion of glioma cells. *PLoS ONE* **8**, e72134 (2013).
38. Wang, S. D. et al. Ephb2 receptor controls proliferation/migration dichotomy of glioblastoma by interacting with focal adhesion kinase. *Oncogene* **31**, 5132–5143 (2012).
39. Hara, A. et al. Treatment strategies based on histological targets against invasive and resistant glioblastoma. *J. Oncol.* **2019**, 2964783 (2019).
40. Gerlee, P. & Nelander, S. The impact of phenotypic switching on glioblastoma growth and invasion. *PLoS Comput. Biol.* **8**, e1002556 (2012).
41. Hatzikirou, H., Basanta, D., Simon, M., Schaller, K. & Deutsch, A. ‘go or grow’: the key to the emergence of invasion in tumour progression? *Math. Med. Biol.* **29**, 49–65 (2012).
42. Hogaia, C., Davatzikos, C. & Biros, G. An image-driven parameter estimation problem for a reaction–diffusion glioma growth model with mass effects. *J. Math. Biol.* **56**, 793–825 (2008).
43. Simmons, M. L. et al. Analysis of complex relationships between age, p53, epidermal growth factor receptor, and survival in glioblastoma patients. *Cancer Res.* **61**, 1122–1128 (2001).
44. Tykocki, T. & Eltayeb, M. Ten-year survival in glioblastoma. A systematic review. *J. Clin. Neurosci.* **54**, 7–13 (2018).
45. Stark, A. M., van de Bergh, J., Hedderich, J., Mehdorn, H. M. & Nabavi, A. Glioblastoma: clinical characteristics, prognostic factors and survival in 492 patients. *Clin. Neurol. Neurosurg.* **114**, 840–845 (2012).
46. Khain, E., Sander, L. M. & Stein, A. M. A model for glioma growth. *Complexity* **11**, 53–57 (2005).
47. Gong, X. et al. Generation of multicellular tumor spheroids with microwell-based agarose scaffolds for drug testing. *PLoS ONE* **10**, e0130348 (2015).
48. Fayad, W. Identification of a novel anticancer compound through screening of a drug library on multicellular spheroids. *Egypt. Pharm. J.* **18**, 188–200 (2019).
49. Herrmann, R., Fayad, W., Schwarz, S., Berndtsson, M. & Linder, S. Screening for compounds that induce apoptosis of cancer cells grown as multicellular spheroids. *J. Biomol. Screen.* **13**, 1–8 (2008).
50. Lagergren, J. H., Nardini, J. T., Michael Lavigne, G., Rutter, E. M. & Flores, K. B. Learning partial differential equations for biological transport models from noisy spatio-temporal data. *Proc. R. Soc. A: Math., Phys. Eng. Sci.* **476**, 20190800 (2020).
51. Lagergren, J. H., Nardini, J. T., Baker, R. E., Simpson, M. J. & Flores, K. B. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. *PLOS Comput. Biol.* **16**, e1008462 (2020).
52. Nardini, J. T. et al. Learning equations from biological data with limited time samples. *Bull. Math. Biol.* **82**, 119 (2020).
53. Xie, Y. et al. The human glioblastoma cell culture resource: validated cell models representing all molecular subtypes. *EBioMedicine* **2**, 1351–1363 (2015).
54. Pollard, S. M. et al. Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. *Cell Stem Cell* **4**, 568–580 (2009).
55. Johansson, P. et al. A patient-derived cell atlas informs precision targeting of glioblastoma. *Cell Rep.* **32**, 107897 (2020).
56. Jones, D. R., Perttunen, C. D. & Stuckman, B. E. Lipschitzian optimization without the Lipschitz constant. *J. Optim. Theory Appl.* **79**, 157–181 (1993).
57. Jones, D. R. & Martins, J. R. R. A. The DIRECT algorithm: 25 years later. *J. Glob. Optim.* **79**, 521–566 (2021).
58. Finkel, D. E. Direct optimization algorithm user guide. https://ctk.math.ncsu.edu/Finkel_Direct/ (2003).
59. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
60. Banks, H. T. & Joyner, M. L. AIC under the framework of least squares estimation. *Appl. Math. Lett.* **74**, 33–45 (2017).

Acknowledgements

Figures 1 and 2 were created with BioRender.com. K.N. was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-2137100. A.M. was supported by Knut and Alice Wallenberg

foundation and WASP-DDLS. The authors acknowledge use of the ELSA high performance computing cluster at The College of New Jersey for conducting the research reported in this paper. This cluster is funded in part by the National Science Foundation under grant numbers OAC-1826915 and OAC-1828163. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding sponsors.

Author contributions

Supervision: K.F. and S.N. Data acquisition: C.K. Model development: A.M., K.N., J.N., and K.F. Simulation and analysis: A.M. and K.N. Manuscript writing: A.M., K.N., J.N., and K.F. Finalising and editing of manuscript: All authors provided critical feedback and discussion and assisted in the revision of the manuscript.

Funding

Open access funding provided by Chalmers University of Technology.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-025-00492-3>.

Correspondence and requests for materials should be addressed to Adam A. Malik.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025