Supplementary material for
"Quantitative analysis of tumour spheroid structure"

Alexander P Browning[†1,2], Jesse A Sharp[†1,2], Ryan J Murphy[1], Gency Gunasingh[4], Brodie Lawson[1,2], Kevin Burrage[1,2,3], Nikolas K Haass[‡4], and Matthew J Simpson[‡1*]

[1] *School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia*
[2] *ARC Centre of Excellence for Mathematical and Statistical Frontiers, QUT, Australia*
[3] *Department of Computer Science, University of Oxford, Oxford, UK*
[4] *The University of Queensland Diamantina Institute, The University of Queensland, Brisbane, QLD, Australia*

August 6, 2021

Contents

This supporting material document refers to code available on GitHub at
`https://github.com/ap-browning/Spheroids`

[*]Corresponding author. E-mail: matthew.simpson@qut.edu.au
[†]These authors contributed equally.
[‡]These authors contributed equally.

S1 Experimental Data

We collect data on enough spheroids to ensure at least 10 for each initial seeding density and observation day (20 for comparing steady-state structure) are available. We then randomly subsample to ensure a consistent number of spheroids are analysed for each initial seeding density and observation day. The total number of spheroids, and the number in the subset, are given in Table S1. Raw data (complete and the subset) are available on GitHub as supplementary material.

In Fig. S1 to Fig. S6 we show a random subset of 10 spheroids from the complete data set for each condition, from days 7 to day 21 (WM983b) and day 24 (WM793b).

	Condition		Day 3	4	5	7	10	12	14	16	18	21	24	Totals	
All	983b	2500	6	12	12	20	16	16	13	17	13	26	–	151	
	983b	5000	9	9	10	15	13	17	14	21	20	31	–	159	456
	983b	10000	6	10	9	18	15	21	13	19	19	16	–	146	
	793b	2500	–	5	10	22	28	27	19	18	20	15	23	187	
	793b	5000	–	12	11	23	25	20	21	19	22	14	21	188	538
	793b	10000	–	7	12	18	25	23	15	17	21	5	20	163	
Subset	983b	2500	6	10	10	10	10	10	10	10	10	20	–	106	
	983b	5000	9	9	10	10	10	10	10	10	20	10	–	108	318
	983b	10000	6	10	9	10	10	10	10	10	19	10	–	104	
	793b	2500	–	5	10	10	10	10	10	10	10	10	20	105	
	793b	5000	–	10	10	10	10	10	10	10	10	10	20	110	317
	793b	10000	–	7	10	10	10	10	10	10	10	5	20	102	

Table S1

S1.1 WM983b (seeded with 2500 cells)

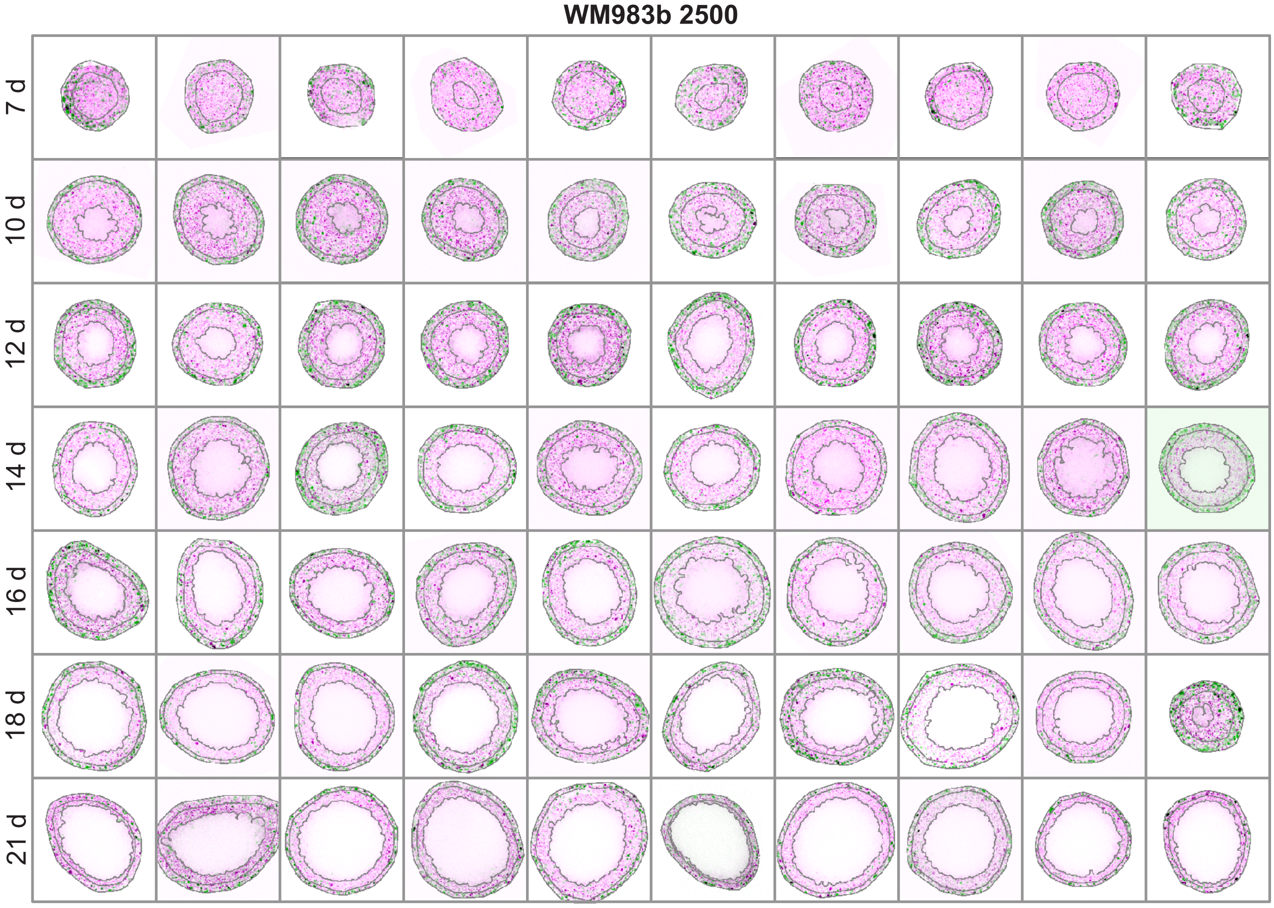

Figure S1

S1.2 WM983b (seeded with 5000 cells)

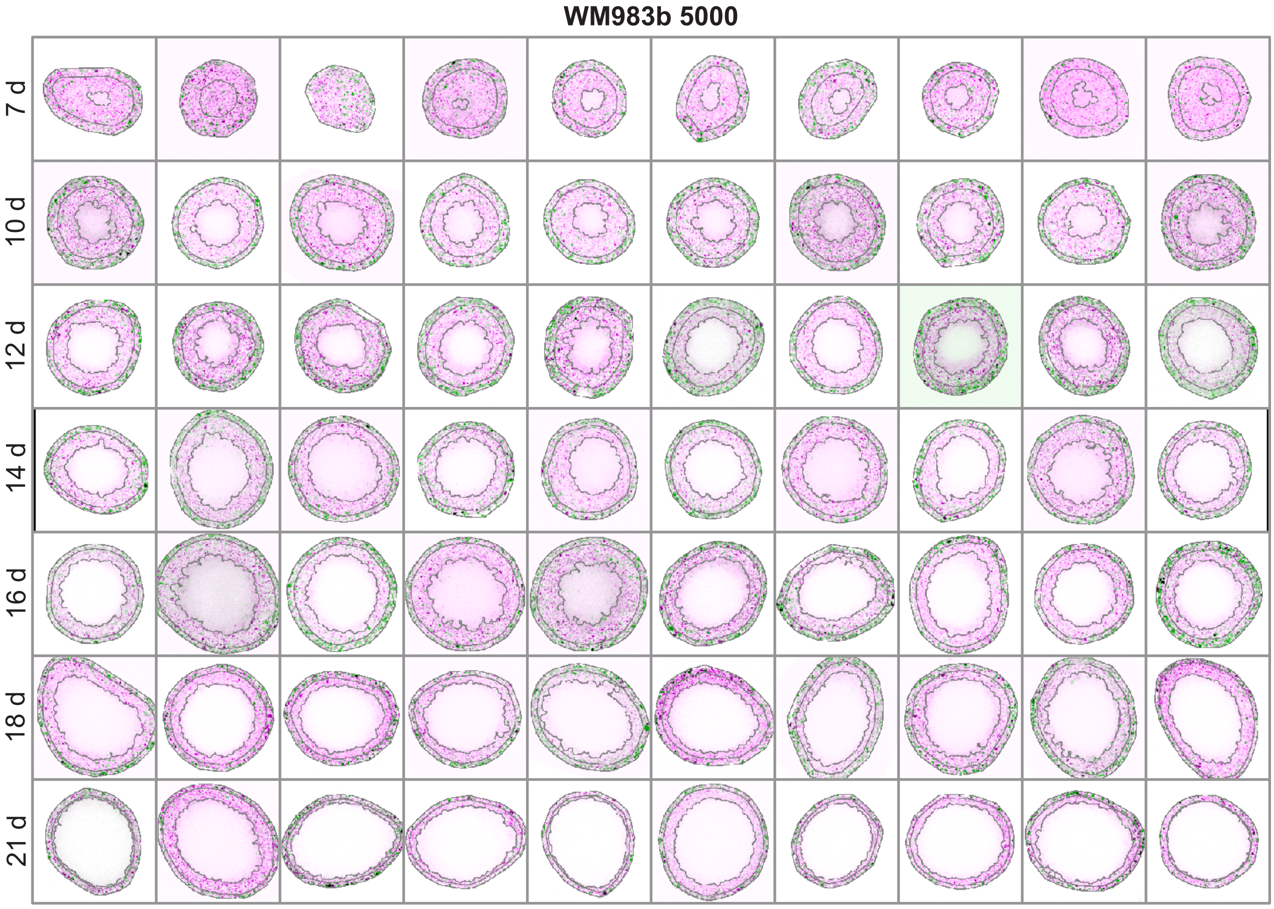

Figure S2

S1.3 WM983b (seeded with 10000 cells)

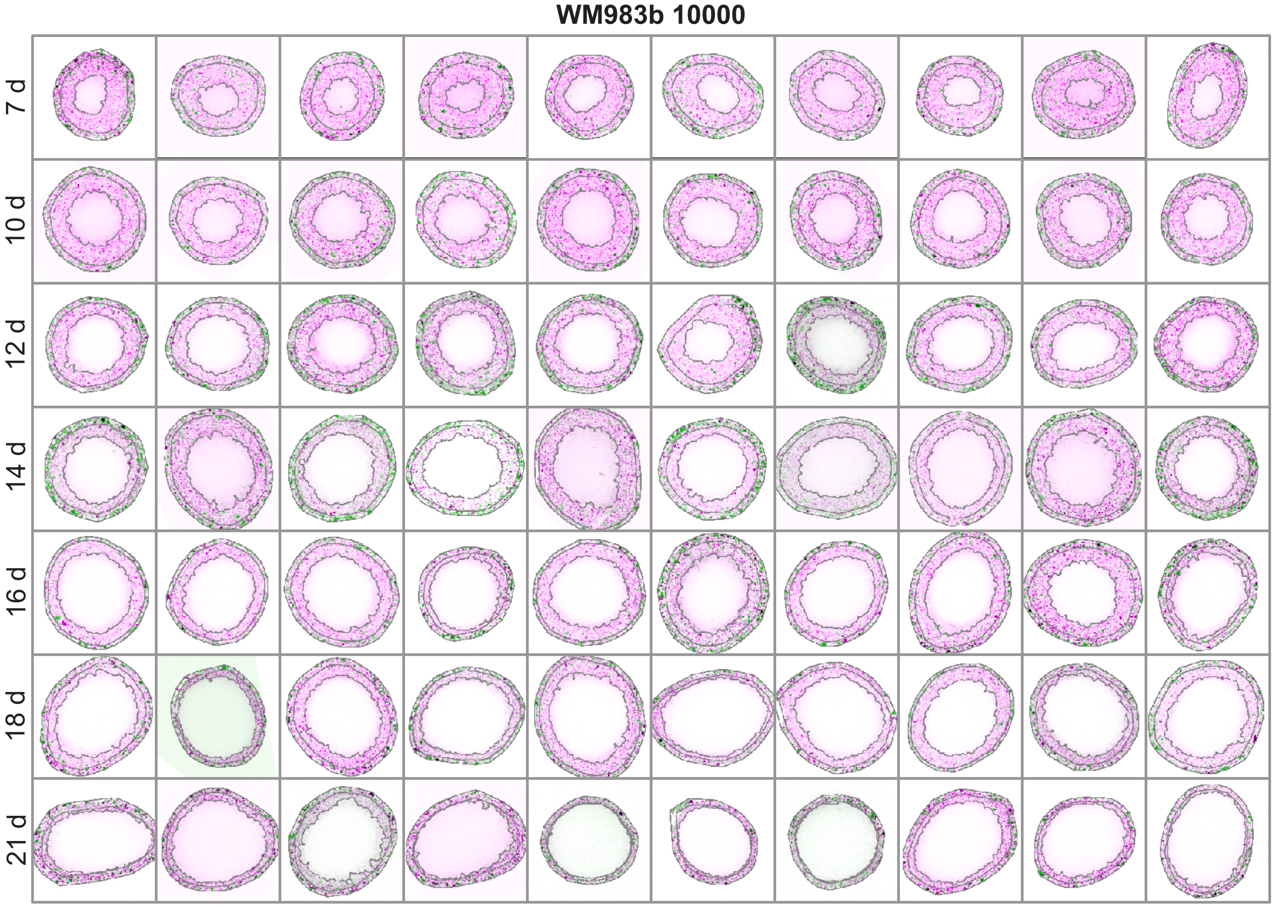

Figure S3

S1.4 WM793b (seeded with 2500 cells)

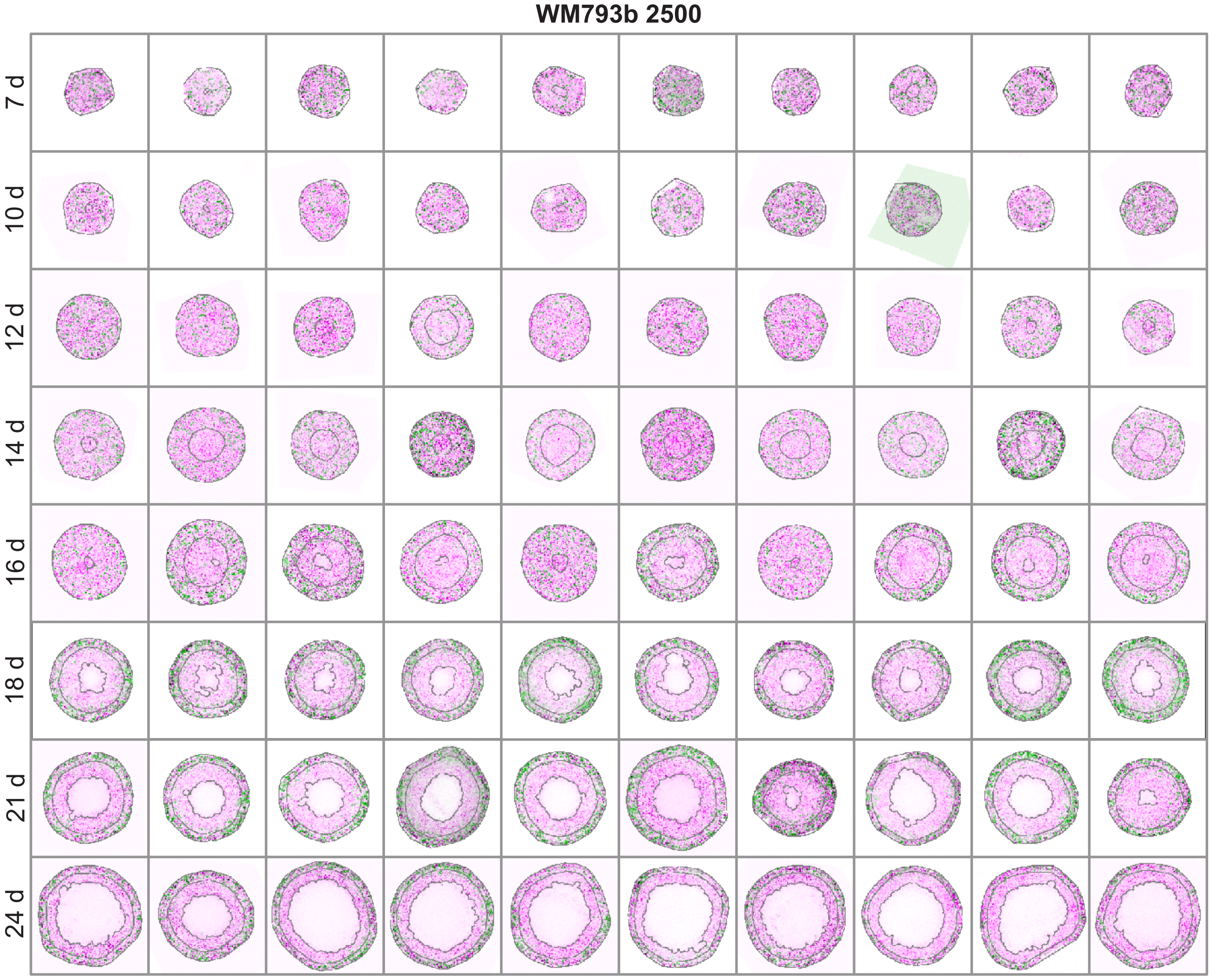

Figure S4

S1.5 WM793b (seeded with 5000 cells)

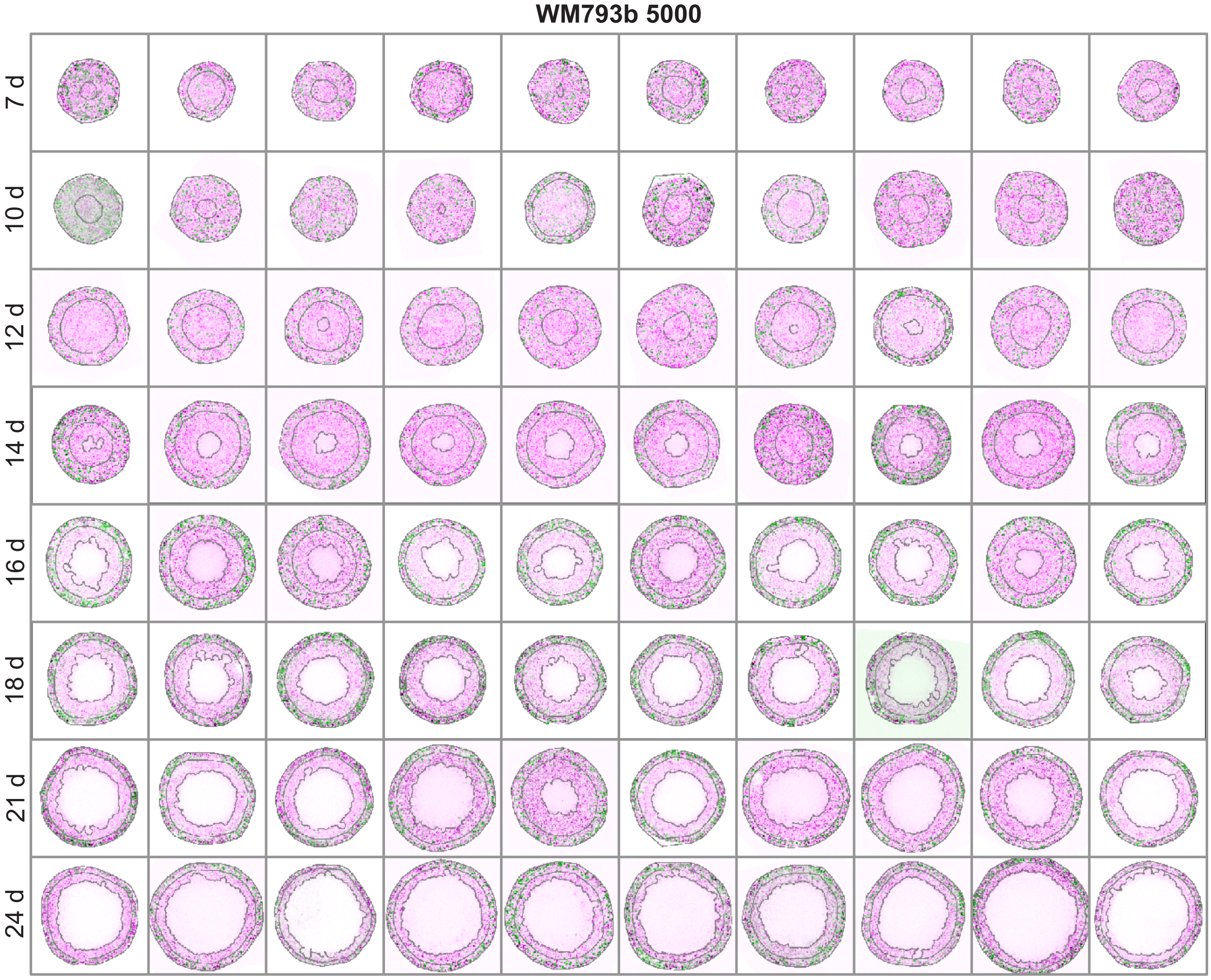

Figure S5

S1.6 WM793b (seeded with 10000 cells)

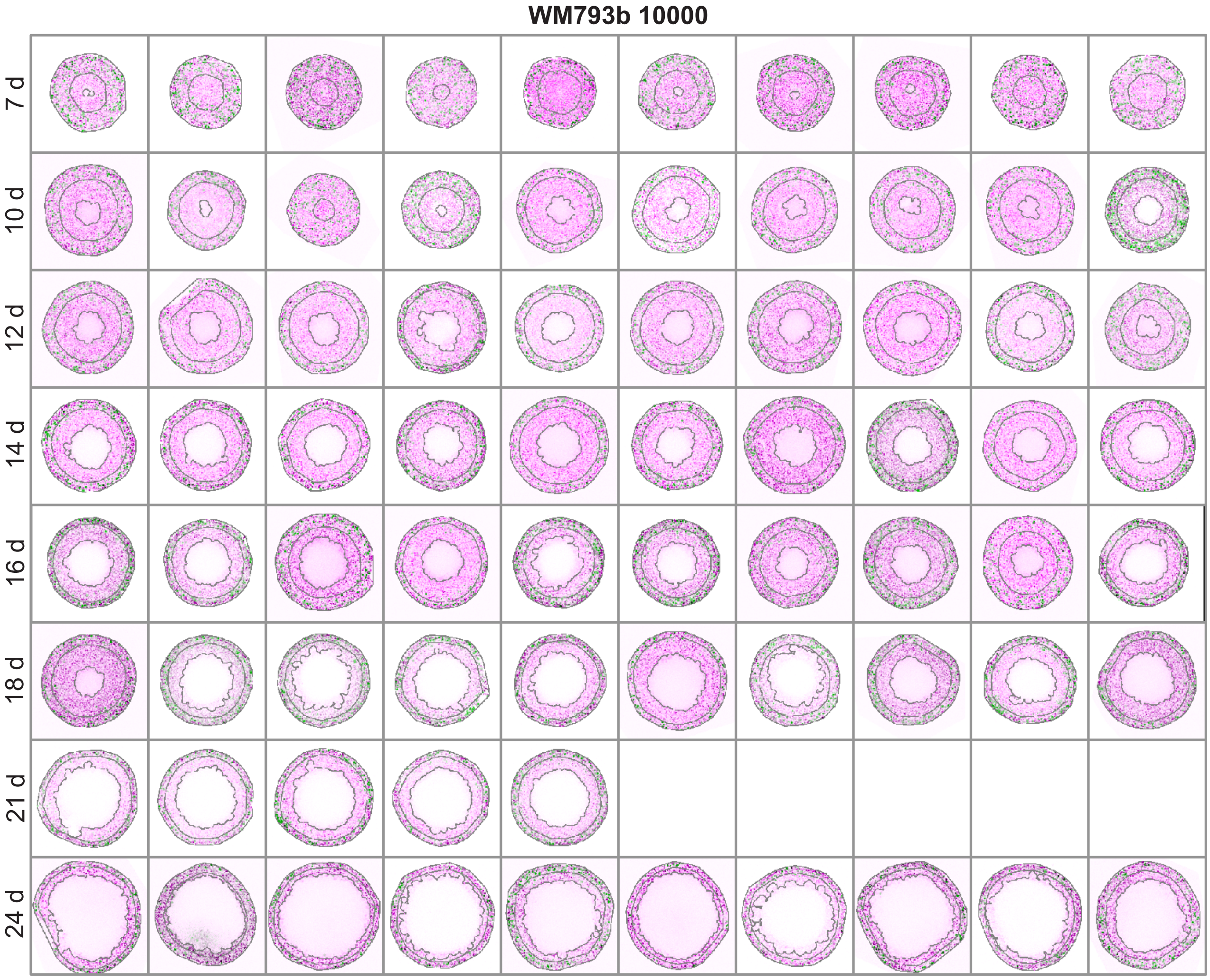

Figure S6

S2 Steady state model solution

The steady-state, denoted $(\bar{R}, \bar{\phi}, \bar{\eta})$ is given by setting $\mathrm{d}R/\mathrm{d}t = 0$ (Eq. (3) in the main document) yielding the non-linear system of equations

$$
\begin{aligned}
0 &= 1 - \bar{\phi}^3 - \gamma\bar{\eta}^3, \\
0 &= 2\bar{R}^2\bar{\eta}^3 - 3\bar{R}^2\bar{\eta}^2 + \bar{R}^2 - \bar{R}_{\mathrm{c}}^2, \\
0 &= \bar{R}^2\bar{\phi}^3 + \big(Q^2\bar{R}_{\mathrm{c}}^2 - \bar{R}^2(1 + 2\bar{\eta}^3)\big)\bar{\phi} + 2\bar{\eta}^3\bar{R}^2.
\end{aligned} \tag{S1}
$$

Applying the substitution $\rho = \bar{\eta}/\bar{\phi}$, where $0 \le \rho \le 1$, and algebraic manipulation (see the Mathematica workbook `SteadyStateSolution.nb`) allows the solution to Eqs. S1 to be expressed as the root of $f(\rho; Q, \gamma)$, where

$$
f(\rho; Q, \gamma) = \sum_{m=0}^{12} c_m \rho^m, \tag{S2}
$$

and where

$$
\begin{aligned}
c_0 &= 3Q^2 - 3Q^4 + Q^6, \\
c_1 &= 0, \\
c_2 &= -9Q^2, \\
c_3 &= 18Q^2 - 18Q^4 + 6Q^6 - 2\gamma + 9Q^2\gamma - 9Q^4\gamma + 3Q^6\gamma, \\
c_4 &= 27Q^4, \\
c_5 &= -36Q^2 - 9Q^2\gamma, \\
c_6 &= 36Q^2 - 36Q^4 - 15Q^6 - 6\gamma + 36Q^2\gamma - 36Q^4\gamma \\
&\quad + 12Q^6\gamma - 3\gamma^2 + 9Q^2\gamma^2 - 9Q^4\gamma^2 + 3Q^6\gamma^2, \\
c_7 &= 54Q^4 + 27Q^4\gamma, \\
c_8 &= -36Q^2 - 36Q^2\gamma, \\
c_9 &= 24Q^2 - 24Q^4 + 8Q^6 + 36Q^2\gamma - 36Q^4\gamma - 15Q^6\gamma - 6\gamma^2 + 18Q^2\gamma^2 \\
&\quad - 18Q^4\gamma^2 + 6Q^6\gamma^2 - \gamma^3 + 3Q^2\gamma^3 - 3Q^4\gamma^3 + Q^6\gamma^3, \\
c_{10} &= 54Q^4\gamma, \\
c_{11} &= -36Q^2\gamma, \\
c_{12} &= 8\gamma.
\end{aligned}
$$

Since ρ is subject to the constraint $0 \le \rho \le 1$, we solve $0 = f(\rho; Q, \gamma)$ using bisection[1], which is guaranteed to converge provided there exists only one root in the interval $0 \le \rho \le 1$. In Fig. S7a, we demonstrate that in the parameter region of interest ($0 < Q < 1$, $\gamma > 0$) there exists only a single solution to Eq. S2. We do this by finding all 12 roots of Eq. S2[2] and counting

[1] Implemented to within machine precision using `Roots.jl`: `https://github.com/JuliaMath/Roots.jl`

[2] Implemented by finding the eigenvalues of the characteristic matrix using `Polynomials.jl`: `https:`

the number of real roots where $0 \le \rho \le 1$.

The solution to Eq. S1 is then given by

$$\bar{R} = f_R(\rho, \bar{\phi}, \boldsymbol{\theta}) = \frac{R_c}{(1-\rho\phi)\sqrt{1+2\rho\phi}} \tag{S3a}$$

$$\bar{\phi} = f_\phi(\rho, \boldsymbol{\theta}) = \frac{1}{(1+\gamma\rho^3)^{1/3}}, \tag{S3b}$$

$$\bar{\eta} = f_\eta(\rho, \bar{\phi}) = \rho\phi, \tag{S3c}$$

where $\boldsymbol{\theta} = (Q, R_c, \gamma)$.

In Fig. S7b, we compare a numerical solution to the transient model to the semi-analytical solution for the steady state showing an excellent match. All algorithms used to produce the results relating to the mathematical model are available in `Module/Greenspan.jl`.

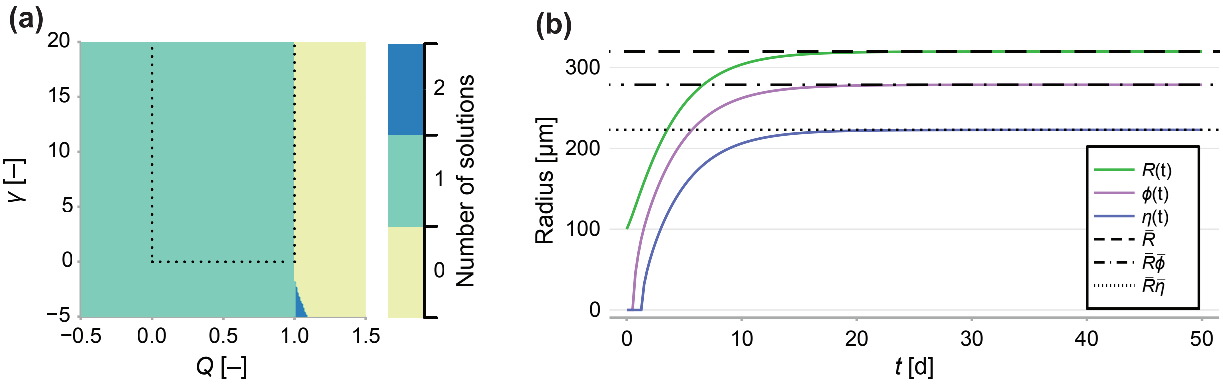

Figure S7. (a) Number of solutions to Eq. S2 subject to the constraint $0 \le \rho \le 1$. Dashed line indicates the region of interest, where $\gamma > 0$ and $0 < Q < 1$. (b) Comparison between a long-term solution to the transient model and the semi-analytical solution to the steady state, where $Q = 0.8$, $\gamma = 1$, $R_c = 150$, $s = 1$ and $R_0 = 100$.

`//github.com/JuliaMath/Polynomials.jl`

S2.1 Jacobian of the steady state model

In the main document, we denote the solution to Eq. S1 as $\mathbf{m}(\boldsymbol{\theta})$. Here, we demonstrate how given a value $(\bar{R}, \bar{\phi}, \bar{\eta}) = \mathbf{m}(\boldsymbol{\theta})$, we can obtain an analytical expression for the model Jacobian,

$$\mathbf{J}_{\mathbf{m}}(\boldsymbol{\theta}) = \frac{\partial \mathbf{m}}{\partial \boldsymbol{\theta}}. \tag{S4}$$

Given ρ, we can form an analytical expression for Eq. S4. Noting that the coefficients of Eq. S2 are functions of $\boldsymbol{\theta}$, we consider

$$\begin{aligned}
\frac{\partial}{\partial c_i}(0) = 0 = \sum_{m=0}^{12} \frac{\partial}{\partial c_i}\big(c_m \rho^m\big) &= \frac{\partial}{\partial c_i}\big(c_i \rho^i\big) + \sum_{\substack{m=0\\ m\neq i}}^{12} \frac{\partial}{\partial c_i}\big(c_m \rho^m\big), \\
&= \rho^i + c_i i \rho^{i-1} \frac{\partial \rho}{\partial c_i} + \sum_{\substack{m=0\\ m\neq i}}^{12} c_m m \rho^{m-1} \frac{\partial \rho}{\partial c_i}, \\
&= \rho^i + \frac{\partial \rho}{\partial c_i} \sum_{m=0}^{12} c_m m \rho^{m-1},
\end{aligned}$$

which yields

$$\frac{\partial \rho}{\partial c_i} = \frac{-\rho^i}{\sum_{m=0}^{12} m c_m \rho^{m-1}} = -\rho^i \left(\frac{\partial f}{\partial \rho}\right)^{-1} \tag{S5}$$

Therefore,

$$\frac{\mathrm{d}\rho}{\mathrm{d}\boldsymbol{\theta}} = \frac{\partial \rho}{\partial \mathbf{c}} \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}},$$

where $\mathbf{c} = (c_0, c_1, ..., c_{12})$; $\partial \rho / \partial \mathbf{c} = (\partial \rho / \partial c_0, ..., \partial \rho / \partial c_{12})$ and $\partial \mathbf{c} / \partial \boldsymbol{\theta}$ is the Jacobian of $\mathbf{c}$ with respect to $\boldsymbol{\theta}$.

Therefore, we have that

$$\frac{\mathrm{d}\bar{\phi}}{\mathrm{d}\boldsymbol{\theta}} = \frac{\partial f_\phi}{\partial \bar{\phi}} + \frac{\partial f_\phi}{\partial \rho} \frac{\mathrm{d}\rho}{\mathrm{d}\boldsymbol{\theta}}. \tag{S6}$$

and it follows that

$$\frac{\mathrm{d}\bar{\eta}}{\mathrm{d}\boldsymbol{\theta}} = \frac{\partial f_\eta}{\partial \bar{\phi}} \frac{\mathrm{d}\bar{\phi}}{\mathrm{d}\boldsymbol{\theta}} + \frac{\partial f_\eta}{\partial \rho} \frac{\mathrm{d}\rho}{\mathrm{d}\boldsymbol{\theta}}, \tag{S7}$$

$$\frac{\mathrm{d}\bar{R}}{\mathrm{d}\boldsymbol{\theta}} = \frac{\partial f_R}{\partial \bar{\phi}} \frac{\mathrm{d}\bar{\phi}}{\mathrm{d}\boldsymbol{\theta}} + \frac{\partial f_R}{\partial \rho} \frac{\mathrm{d}\rho}{\mathrm{d}\boldsymbol{\theta}} + \frac{\partial f_R}{\partial \boldsymbol{\theta}}. \tag{S8}$$

Therefore, an analytical expression for $\mathbf{J}_{\mathbf{m}}(\boldsymbol{\theta})$ (Eq. S4) is given by

$$\mathbf{J}_{\mathbf{m}}(\boldsymbol{\theta}) = \left(\frac{\mathrm{d}\bar{R}}{\mathrm{d}\boldsymbol{\theta}}, \frac{\mathrm{d}\bar{\phi}}{\mathrm{d}\boldsymbol{\theta}}, \frac{\mathrm{d}\bar{\eta}}{\mathrm{d}\boldsymbol{\theta}}\right). \tag{S9}$$

S3 Results for WM793b

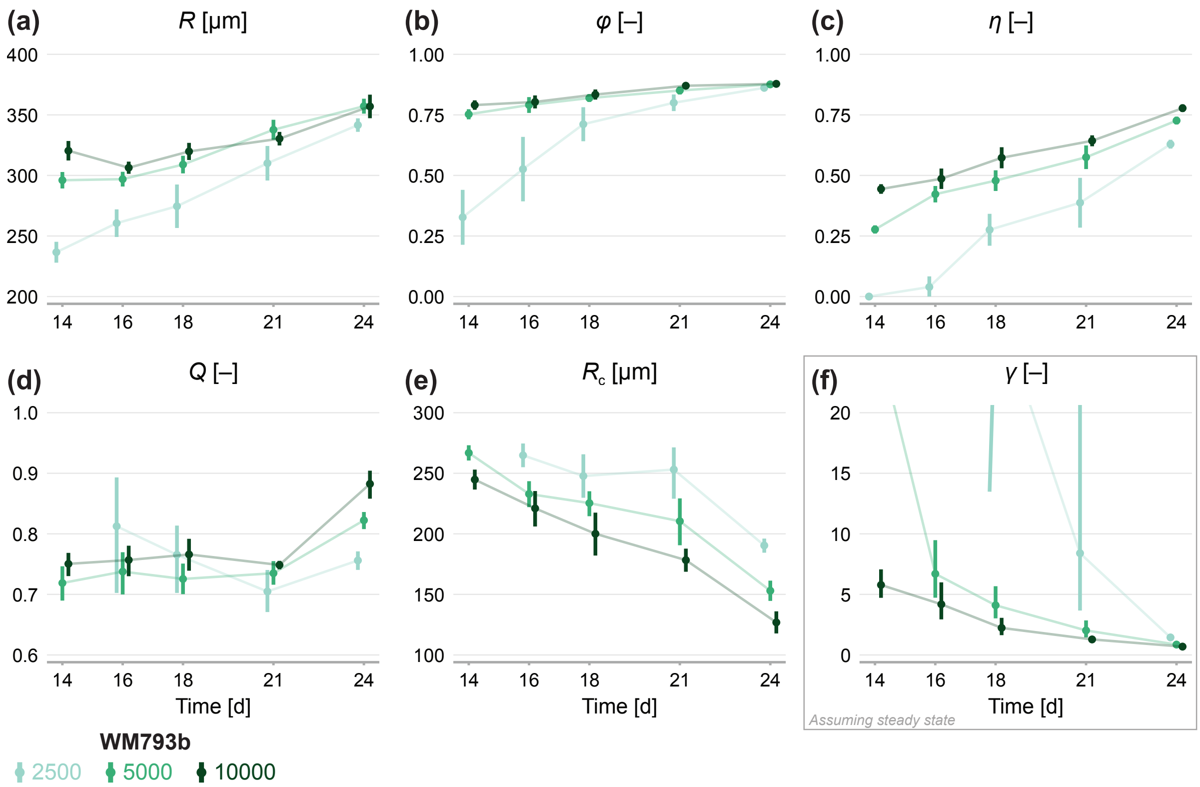

Figure S8. Estimates of parameters using the structural model with data from various time points. (a–c) Parameters are the mean of each observation. (d–f) Parameters in the mathematical model. As estimates Q and R_c can be derived from the structural model, which applies at any time during phase 3, we expect to see consistent estimates across observation times. Given that WM793b spheroids initiated with 2500 cells do not reach phase 3 until day 14, we exclude day 12 for these spheroids from the mathematical analysis. As estimates of γ can only be derived from the steady-state model, which assumes the outer radius is no longer increasing, we only expect consistency for later observation days. Bars indicate an approximate 95% confidence interval.

S4 Total squares regression

In typical least-squares estimation we fit a model of the form

$$y_i = a + bx_i + \varepsilon_{y,i}, \tag{S10}$$

where $\varepsilon_{y,i} \sim \mathcal{N}(0, \sigma_y)$ is assumed to be a normally distributed error component in y component [1], and (a, b) are model parameters. Least-squares and maximum likelihood estimates $(\hat{a}, \hat{b})$ can then be found by minimising the sum-square error

$$(\hat{a}, \hat{b}) = \underset{(a,b)}{\operatorname{argmin}} \sum_i (y_i - (a + bx_i))^2. \tag{S11}$$

We demonstrate this in Fig. S9. In typical least squares estimation, we minimise the *vertical distance* between the data points and the regression line (blue dashed).

In the main document, we fit a linear model to data of the form (R, ϕ, η), where each component contains an error term. In two-dimensions, this is akin to a model of the form

$$y_i = a + bx_i + \varepsilon_{y,i} + b\varepsilon_{x,i}. \tag{S12}$$

where we have included an additional error term $\varepsilon_{x,i} \sim \mathcal{N}(0, \sigma_x)$, assumed to be a normally distributed error component in x_i. In this case, the least squares estimate is given by minimising the total *perpendicular distance* between the data points and the regression line (Fig. S9, blue solid) [1].

In the main paper, we fit a linear model of the form

$$(R(\tau), \phi(\tau), \eta(\tau)) = (R_\mathrm{c}, \phi_\mathrm{c}, 0) + \tau\hat{\mathbf{q}}, \tag{S13}$$

parameterised by $R_\mathrm{c}, \phi_\mathrm{c}$ and a unit vector $\hat{\mathbf{q}}$.

If we denote $\mathbf{y}_0 = (R_\mathrm{c}, \phi_\mathrm{c}, 0)$ and $\mathbf{y}_1 = (R_\mathrm{c}, \phi_\mathrm{c}, 0) + \hat{\mathbf{q}}$, then the shortest distance between observation $\mathbf{x}_i = (R_i, \phi_i, \eta_i)$ is given by

$$d(\mathbf{x}_i; R_\mathrm{c}, \phi_\mathrm{c}, \hat{\mathbf{q}}) = \frac{\|(\mathbf{x}_i - \mathbf{y}_0) \times (\mathbf{x}_i - \mathbf{y}_1)\|}{\|\mathbf{y}_0 - \mathbf{y}_1\|}, \tag{S14}$$

where $\|\cdot\|$ denotes the Frobenius norm, and $\times$ denotes the vector cross product.

Therefore, least-squares estimates of the parameters can then be found by minimising the sum-square error

$$\min_{(R_\mathrm{c}, \phi_\mathrm{c}, \hat{\mathbf{q}})} \sum_i d(\mathbf{x}_i; R_\mathrm{c}, \phi_\mathrm{c}, \hat{\mathbf{q}}). \tag{S15}$$

S4.1 Approximating the likelihood

To implement a log-likelihood-ratio based hypothesis test, we must approximate the likelihood at the parameter estimates. To do this, we note that the total square error, denoted ε_i^2, is of the form

$$\varepsilon_i^2 = c_1\varepsilon_{x,i}^2 + c_2\varepsilon_{y,i}^2 + c_3\varepsilon_{z,i}^2, \tag{S16}$$

where $\varepsilon_{x,i}$, $\varepsilon_{y,i}$, and $\varepsilon_{z,i}$ are normally distributed with variances σ_x^2, σ_y^2 and σ_z^2, respectively. If $\sigma_x^2 = \sigma_y^2 = \sigma_z^2$, ε_i^2 would have an approximate chi-squared distribution by the Welch-Satterthwaite equation [2], a special case of the gamma distribution. Therefore, we approximate the distribution of ε_i^2 by fitting a gamma-distribution to the observed square error when a total squares estimate is fit to the combined data (Fig. S9b).

Therefore, the approximate log-likelihood is given by

$$\ell(R_{\mathrm{c}}, \phi_{\mathrm{c}}, \hat{\mathbf{q}}) = \sum_i \log f_\Gamma \left(d^2(\mathbf{x}_i; R_{\mathrm{c}}, \phi_{\mathrm{c}}, \hat{\mathbf{q}}) \right), \tag{S17}$$

where $f_\Gamma(\cdot)$ is the probability density function of the fitted gamma function.

S4.2 Log-likelihood-ratio based test

We denote $\hat{\boldsymbol{\theta}}_0 = (R_{\mathrm{c}}, \phi_{\mathrm{c}}, \hat{\mathbf{q}})$ the maximum likelihood estimate when the data from all initial conditions is pooled, and $\hat{\boldsymbol{\theta}}_N = (R_{\mathrm{c}}, \phi_{\mathrm{c}}, \hat{\mathbf{q}})$ the estimates from initial condition $N \in \{2500, 5000, 10000\}$. As the models must be nested for the likelihood-ratio test, we estimate the noise model, $f_\Gamma(\cdot)$, using the estimates from the pooled data.

The test-statistic is given by

$$\lambda = \ell(\hat{\boldsymbol{\theta}}_{2500}) + \ell(\hat{\boldsymbol{\theta}}_{5000}) + \ell(\hat{\boldsymbol{\theta}}_{10000}) - \ell(\hat{\boldsymbol{\theta}}_0), \tag{S18}$$

where $\lambda \sim \chi^2_\nu$, and

$$\nu = \dim(\hat{\boldsymbol{\theta}}_{2500}) + \dim(\hat{\boldsymbol{\theta}}_{5000}) + \dim(\hat{\boldsymbol{\theta}}_{10000}) - \dim(\hat{\boldsymbol{\theta}}_0) = 8. \tag{S19}$$

Our implementation of this test is provided on GitHub in `Module/Inference` in the function `lm_orthogonal_test`.

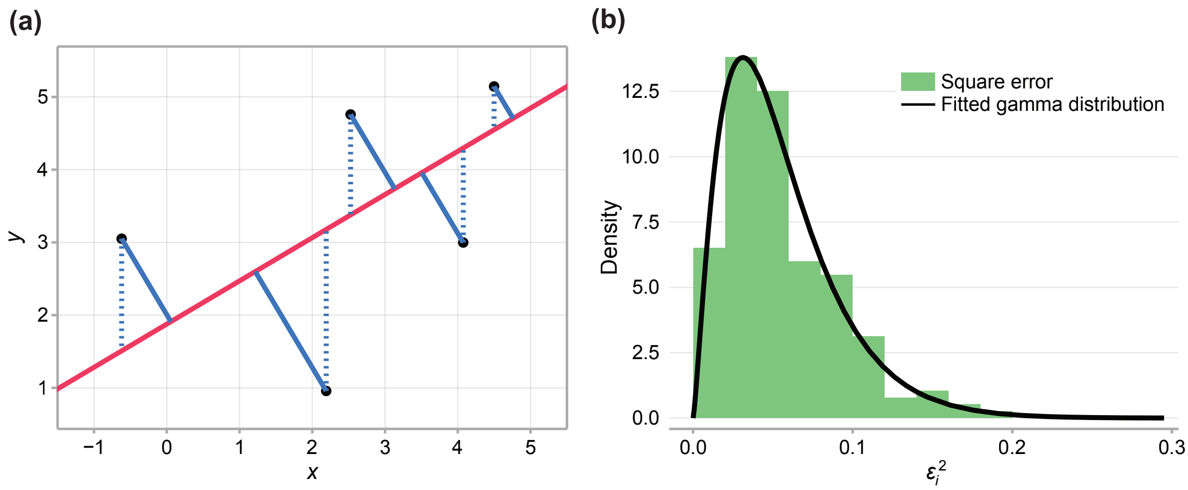

Figure S9. (a) Comparison between typical least-squares error (blue dashed), and total-least-squares error (blue solid). (b) Square error observed in the data and fitted gamma distribution

References

[1] Markovsky I, Huffel SV. 2007. Overview of total least-squares methods. *Signal Processing* **87**:2283–2302. doi:10.1016/j.sigpro.2007.04.004.

[2] Welch BL. 1947. The Generalization of 'Student's' Problem when Several Different Population Variances are Involved. *Biometrika* **34**:28. doi:10.2307/2332510.