Quantifying the Length and Variance of the Eukaryotic Cell Cycle Phases by a Stochastic Model and Dual Nucleoside Pulse Labelling

Tom Serge Weber[1,2], Irene Jaehnert[3], Christian Schichor[3], Michal Or-Guil[2]*, Jorge Carneiro[1]*

1 Instituto Gulbenkian de Ciência, Oeiras, Portugal, 2 Department of Biology, Humboldt-Universität zu Berlin, Berlin Germany and Research Center ImmunoSciences, Charité - Universitätsmedizin Berlin, Berlin, Germany, 3 Tumorbiological Laboratory, Neurosurgical Department, Ludwig-Maximilians-University Munich, Klinikum Grosshadern, Munich, Germany

Abstract

A fundamental property of cell populations is their growth rate as well as the time needed for cell division and its variance. The eukaryotic cell cycle progresses in an ordered sequence through the phases G_1, S, G_2, and M, and is regulated by environmental cues and by intracellular checkpoints. Reflecting this regulatory complexity, the length of each phase varies considerably in different kinds of cells but also among genetically and morphologically indistinguishable cells. This article addresses the question of how to describe and quantify the mean and variance of the cell cycle phase lengths. A phase-resolved cell cycle model is introduced assuming that phase completion times are distributed as delayed exponential functions, capturing the observations that each realization of a cycle phase is variable in length and requires a minimal time. In this model, the total cell cycle length is distributed as a delayed hypoexponential function that closely reproduces empirical distributions. Analytic solutions are derived for the proportions of cells in each cycle phase in a population growing under balanced growth and under specific non-stationary conditions. These solutions are then adapted to describe conventional cell cycle kinetic assays based on pulse labelling with nucleoside analogs. The model fits well to data obtained with two distinct proliferating cell lines labelled with a single bromodeoxiuridine pulse. However, whereas mean lengths are precisely estimated for all phases, the respective variances remain uncertain. To overcome this limitation, a redesigned experimental protocol is derived and validated in silico. The novelty is the timing of two consecutive pulses with distinct nucleosides that enables accurate and precise estimation of both the mean and the variance of the length of all phases. The proposed methodology to quantify the phase length distributions gives results potentially equivalent to those obtained with modern phase-specific biosensor-based fluorescent imaging.

Introduction

The cell cycle is one of the most fundamental processes in biology. Through this process, a parental cell transmits to its two daughter cells genetic and epigenetic information by accurately replicating its DNA and evenly apportioning all nuclear and extranuclear contents. The mechanism of cell cycle regulation is tailored to ensure accurate cellular content replication, but seems to be less constrained by how long it takes to complete this process successfully. Several check points exist that ensure that chromosomes are faithfully copied and that the parental cell has enough material in order to produce two viable isogenic daughter cells. Meeting the conditions of each of these check points takes variable time and delays the completion of the cell cycle. Yet, how long the cells take on average to complete the cell cycle is an important biological property. In unicellular organisms, the average inter-mitotic time is a direct measurement of the organism's fitness, while in multicellular organisms, the regulation of the rate of cell division is critical for development, stem cell maintenance, tissue or organ homeostasis, wound healing, and immunity. The temporal organization of the cell cycle is therefore under tight regulation, likely reflecting a fine balance between accuracy in information transmission and speed.

The average cell cycle time has been estimated at the population level by measuring the growth curve of exponentially proliferating cell cohorts, under conditions in which cells can be counted and cell death is negligible compared to the population wide growth rate. Under conditions in which cell counting is not possible or in which cell death rates cannot be neglected (e.g., homeostasis, immune reactions, cancer growth), indirect estimates for the average division time or the average death time are typically inferred e.g., through the rate of increase of cells arrested in mitosis after administration of colchicine, the fraction of labelled mitotic figures after pulse labelling (FLM method), and from long-

Author Summary

Among the important characteristics of dividing cell populations is the time necessary for cells to complete each of the cell cycle phases, that is, to increase the cell's mass, to duplicate and repair its genome, to properly segregate its chromosomes, and to make decisions whether to continue dividing or enter a quiescent state. The cycle phase times also determine the maximal rate at which a dividing cell population can grow in size. Cell cycle phase completion times largely differ between cell types, cellular environments as well as metabolic stages, and can thus be considered as part of the phenotype of a given cell. Our article advances the methods to quantitatively characterize this phenotype. We introduce a novel phase-resolved cell cycle progression model and use it to estimate the mean and variance of the cycle phase completion times based on nucleoside analog pulse labelling experiments. This classic workhorse of cell cycle kinetic studies is revamped by our approach to potentially rival in accuracy and precision with modern phase-specific biosensor-based fluorescent imaging, while superseding the latter in its application scope.

term labelling and delabelling time-series of deuterium or bromodeoxyuridine (BrdU) tracing experiments [1–3]. For growing cell populations these estimates depend on assumptions about the shape of the intermitotic time distribution [4]. The latter, when analyzed at a single-cell level, e.g., by time-lapse imaging, shows significant variability in otherwise seemingly homogeneous cell populations. This observation led more than forty years ago to the development of one of the first stochastic cell cycle models [5]. Smith and Martin proposed at that time that cell's life comprehends an **A** state and a **B** phase. Whereas the time cells spend in the **A** state was assumed to be exponentially distributed, the time cells spend in the **B** phase was, in this simplest scenario, a fixed delay. Experimental validation was provided by time-lapse imaging of growing cell cultures, measurements of fraction of labelled mitoses and fractions of sibling pairs with age difference greater than a specified value [6]. Even though later studies [7–10] have shown that the model assumptions do not exactly match experimental data, its simplicity and mathematical tractability makes the Smith-Martin model even today a popular theoretical model [6,11].

In the last ten years, 5-(and 6)-Carboxyfluorescein diacetate succinimidyl ester (CFSE) dilution assays in concert with a whole set of advanced modeling techniques [12–14] allowed to estimate the average duration, as well as inter-cellular variability in more complex scenarios with division time densities *in vitro* or *in vivo* after adoptive cell transfer. Especially generation structure, activation times and generation dependent cell death were included in these models and subsequently estimated in the context of lymphocyte proliferation. Inter-cellular variability not only of division times but also of death times were confirmed directly in long-term tracking of single HeLa cells [15] and B-lymphocytes [10]. The latter study provided extensive quantitative data on the shape of age-dependent division and death time distributions which are required to calibrate e.g., the Cyton [16] or similar models. A review on these, and alternative stochastic cell cycle models is given in [4].

At a higher temporal and functional resolution the eukaryotic cell cycle is structured into four distinct phases: 1) the **G₁** phase during which organelles are reorganized and chromatin is licensed for replication, 2) the **S** phase in which the chromosomes are duplicated by DNA replication, 3) the **G₂** phase which serves as a holding time for synthesis and accumulation of proteins needed in 4) the **M** phase, or mitosis, which is marked by chromatin condensation, nuclear envelope breakdown, chromosomal segregation, and finally cytokinesis, which completes the generation of two daughter cells in **G₁** phase [17].

Considering explicitly cell cycle phases in mathematical models of cell division probably dates back to the discovery that **DNA** is replicated mainly during a specific period of the cell cycle. Already in their seminal paper, Smith and Martin related the **A** state to the **G₁** phase and the **B** phase to the **S**, **G₂**, **M** and possibly to some part of the **G₁** phase. Subsequent studies that explored phase-resolved cell cycle models, majoritarely rooted in the field of oncology and cancer therapy, include [18–25]. As in the present work, most of these studies relied on flow cytometry (**FACS**) data generated by labelling selectively cells that are synthesizing **DNA** using nucleoside analogs (e.g., BrdU, iodo-deoxyuridine (IdU) or ethynyl-deoxyuridine (EdU)), together with a fluorescent intercalating agent to measure total DNA content (e.g., 4,6- diamidino-2-phenylindole (DAPI), and propidium iodide (PI)), in order to test the model assumptions and draw conclusions about the cells and conditions under consideration.

Here we present a simple stochastic cell cycle model that incorporates temporal variability at the level of individual cell cycle phases. More precisely, we extend the concept underlying the Smith-Martin model of delayed exponential waiting times to the cell cycle phases. We first demonstrate that the model is in good agreement with published experimental data on inter-mitotic division time distributions. We then show, based on stability analysis, that phase-specific variability remains largely undetermined when measurements are taken on cell populations under balanced growth (i.e., growth under asymptotic conditions in which the expected proportions of cells in each phase of the cycle are constant). We prove that by properly measuring proliferating cells under unbalanced growth, one can with at least three well placed support points, assuming noise-free conditions, uniquely identify the average and variance in the completion time of each of the cell cycle phases. When comparing our model with two experimental data sets obtained from conventional pulse-labelling experiments of distinct proliferating cell lines, we find that, while the kinetics extracted from these experiments are well approximated by the predictions of the proposed model, the information content is insufficient to determine accurately all the parameters. Finally we propose a modification of the prevailing experimental protocol, based on dual-pulse labelling with **BrdU** and, for example, EdU, that overcomes this shortcoming.

Results

Model definition

The eukaryotic cell cycle is defined as an orderly sequence of three phases distinguished by cellular DNA content, termed **G₁**, **S** and **G₂M**. A dividing cell is supposed to proceed, under this minimalist view, from one phase to another in a fixed order, until reaching the end of **G₂M** phase. Here it completes cytokinesis generating two genetically identical daughter cells that are by definition in **G₁** phase (Fig. 1 A). We assume that the completion time of any phase (i.e. the time lapse between the entry to and exit from that given phase) is a random variable τ, which is distributed according to a delayed (or shifted) exponential density function (Fig. 1 B),
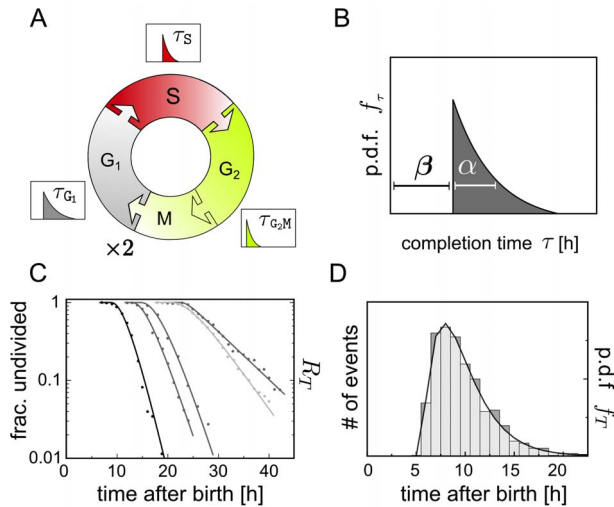
Figure 1. Stochastic cell cycle model. A: Scheme of the proposed cell cycle model with three phases G_1, S and G_2M. The dashed border between the G_2 and the M phase indicates that the G_2 and M phase are pooled into a single phase. The random time τ a cell needs to complete the processes associated to each of the phases, follows a delayed exponential distribution with specific parameters α and β for each phase. **B**: Delayed-exponential completion time distribution density f_τ with parameters α and β. **C**: Best fit of the complementary cumulative distribution R_T to the fraction of undivided cells after birth obtained by time lapse cinematography [5] of slow and fast dividing cell lines. **D**: Best fit of f_T defined by Eq. 4 (solid line) to inter-mitotic time distribution density measured by long-term video tracking of *in vitro* proliferating B-cells [10]. The data in C and D were read from the graphs in the original publications ([5] and [10] respectively).
doi:10.1371/journal.pcbi.1003616.g001

$$f_\tau(\tau) = \frac{1}{\alpha} e^{-\frac{1}{\alpha}(\tau-\beta)} H(\tau-\beta), \qquad (1)$$

where α is the reciprocal of the rate of the exponential (measured in time units) and β is the fixed delay (in time units), and H denotes the Heaviside step function whose value is zero for negative argument, i.e., for $\tau < \beta$, and one for positive argument. Notice that with a slight abuse of notation we denote here the random variable (subscript of density function f) and the value it assumes (the argument of the function f_τ) by the same symbol τ. This will allow us to denote the probability density function and the cumulative probability distribution of the random variable x by f_x and F_x respectively, and to define the complementary cumulative distribution $R_x = 1 - F_x$. The delay β in Eq. 1 'ensures' that a cell that enters a specific phase will remain therein for at least β time units (e.g. hours) before proceeding to the next phase. Besides this fixed minimal time β, additional less predictable effects that affect the completion of the processes associated to a phase are assumed to be exponentially distributed with both mean and standard deviation given by α. The phase specific mean completion time, denoted in the following by $\bar{\tau}$ is then $\alpha + \beta$ with standard deviation α and coefficient of variation $\alpha/\bar{\tau}$. The Laplace transform of Eq. 1 is given by

$$\mathcal{L}_\omega\{f_\tau(\tau)\} = \frac{e^{-\beta\omega}}{1+\alpha\omega}, \qquad (2)$$

where ω is the transformed variable corresponding to the time lapse τ. The temporal organization of the cell cycle is defined by the vector of phase-specific completion times, $\boldsymbol{\tau} = \{\tau_{G_1}, \tau_S, \tau_{G_2M}\}$, which in turn depend on the parameter vectors $\boldsymbol{\alpha} = \{\alpha_{G_1}, \alpha_S, \alpha_{G_2M}\}$ and $\boldsymbol{\beta} = \{\beta_{G_1}, \beta_S, \beta_{G_2M}\}$. The cell cycle length, understood as the time lapse between the entry into G_1 until exit out of G_2M, is the random variable $T = \tau_{G_1} + \tau_S + \tau_{G_2M}$. Its probability density function is the convolution of the three underlying delayed exponential distributions and corresponds to the delayed hypoexponential distribution. Explicit expressions can be computed using the inverse Laplace transform $\mathcal{L}^{-1}$ of the product of the Laplace transforms of the three densities given by Eq. 2, i.e.,

$$f_T(T) = \mathcal{L}_T^{-1}\left\{\prod_i \frac{e^{-\beta_i\omega}}{1+\alpha_i\omega}\right\} \text{ with } i = G_1, S, G_2M. \qquad (3)$$

In case that all entries in $\boldsymbol{\alpha}$ are distinct, we get

$$f_T(T) = \sum_i \left(\frac{\alpha_i e^{-\frac{(T-B)}{\alpha_i}}}{\prod_{j,\ i\neq j}(\alpha_i - \alpha_j)}\right) H(T-B), \qquad (4)$$

in which the indices i and j iterate over the three phases and B is the sum of the elements in $\boldsymbol{\beta}$.

In Fig. 1 B we plot the shape of the phase specific completion time distribution f_τ defined by Eq. 1, which illustrates that the probability for a cell to complete a given phase in less than β time units is zero under this model. A graphical representation of the cell cycle model is provided in Fig. 1 A. Notice that each phase can have distinct parameter values α and β for the completion time distribution.

As a first validation, we compared the empirical frequency of undivided cells as a function of time after 'birth' (reported by [5]) with the respective probability according to the model $1 - F_T$, which we denote as R_T (Fig. 1 C). As a second test, we fitted the cell cycle length density f_T given by Eq. 4 to data extracted from video-tracking of *in vitro* proliferating B cells [10]. The delayed hypoexponential distribution f_T (shown in Fig. 1 D), but also the delayed log-normal and the delayed gamma distribution (not shown) with parameter values proposed in [10], reproduce closely the measured division time histogram. While the two latter depend on three parameters each, the hypoexponential distribution depends on six parameters, that remain largely undetermined given this kind of data.

Balanced growth

A proliferating cell population that obeys the probability model specified in the previous section can be represented by a non-Markov multidimensional random process, whose evolution depends on its history. There exist an infinite number of possible histories or realizations of the population size dynamics $N(t)$. We focus here on a specific important subset, namely those under balanced growth. Under balanced growth a cell population grows exponentially $E[N(t)] \stackrel{\text{def}}{=} E[N_{G_1}(t) + N_S(t) + N_{G_2M}(t)] \propto e^{\mu t}$ with mean growth rate μ and constant mean proportions of cells in the three phases $\boldsymbol{n} = \{n_{G_1}, n_S, n_{G_2M}\}$, where e.g., $n_{G_1} \stackrel{\text{def}}{=} E[N_{G_1}(t)/(N_{G_1}(t) + N_S(t) + N_{G_2M}(t))]$. The expectation operator $E[]$ is defined over all possible realizations of the process.

We will now derive explicit expressions for n_{G_1}, n_S and n_{G_2M} and a transcendental equation that defines μ, the growth rate. A

first step in obtaining the constant frequencies of cells in each of the phases consist in computing the ratio between the cells that complete a given phase and the total number of cells inside the same phase at time t. This phase-specific quantity, denoted here by γ, represents the asymptotic efflux rate constant, which will be useful, as we will see, to construct a transition probability matrix $\mathcal{Q}$. The latter will enable us to employ methods from linear algebra to solve the steady state condition.

Suppose for example that a cohort of cells entered a given phase at time t_{in}. Then the density of cells leaving this phase at time t_{out} will be $f_\tau(t_{out} - t_{in})$. Similarly if a cohort of cells entered this phase at time t_{in}, then a proportion $R_\tau(t - t_{in})$ will remain in it until time t.

Recalling that the influx of cells into a given phase is proportional to $e^{\mu t}$ and that $R_\tau(t)$ is the complementary cumulative distribution of τ, $(1 - F_\tau(t))$, which is Laplace transformed to $\mathcal{L}_\omega\{1 - F_\tau(t)\} = (1 - \mathcal{L}_\omega\{f_\tau(t)\})/\omega$, we integrate over all past entries and finally take the ratio to obtain

$$\gamma = \frac{\int_{-\infty}^{t} e^{\mu x} f_\tau(t-x) dx}{\int_{-\infty}^{t} e^{\mu x} R_\tau(t-x) dx} \qquad (5)$$

$$= \frac{\mu \mathcal{L}_\mu\{f_\tau(\tau)\}}{1 - \mathcal{L}_\mu\{f_\tau(\tau)\}} \qquad (6)$$

$$= \frac{\mu}{(\alpha\mu+1)e^{\beta\mu} - 1}. \qquad (7)$$

While the second equality is a consequence of the definition of the Laplace transform, the third equality follows by substituting $\mathcal{L}_\mu\{f_\tau(\tau)\}$ using Eq. 2. For a phase without a delay, i.e., $\beta = 0$, the last expression simplifies to the familiar mass action principle, where the transition probability is directly proportional to the decay rate $1/\alpha$. Assuming that cells are immortal and recalling that division occurs as cells proceed from G_2M to G_1 we build up the transition probability matrix as follows

$$\mathcal{Q} = \begin{bmatrix} -\gamma_{G_1} & 0 & 2\gamma_{G_2M} \\ \gamma_{G_1} & -\gamma_S & 0 \\ 0 & \gamma_S & -\gamma_{G_2M} \end{bmatrix}. \qquad (8)$$

The balanced growth condition can now be formulated in matrix form

$$\mathcal{Q} \begin{bmatrix} n_{G_1} \\ n_S \\ n_{G_2M} \end{bmatrix} = \mu \begin{bmatrix} n_{G_1} \\ n_S \\ n_{G_2M} \end{bmatrix}, \qquad (9)$$

where the growth rate μ is an eigenvalue of $\mathcal{Q}$ and the proportions vector $\boldsymbol{n} = \{n_{G_1}, n_S, n_{G_2M}\}$ is the corresponding eigenvector. It can be shown that there exists a single dominating real positive eigenvalue for $\mathcal{Q}$ (see Materials and Methods) whose associated normalized eigenvector is

$$\begin{bmatrix} n_{G_1} \\ n_S \\ n_{G_2M} \end{bmatrix} = \begin{bmatrix} 2 \times (1 - 1/(e^{\beta_{G_1}\mu}(\alpha_{G_1}\mu+1))) \\ 1 - n_{G_1} - n_{G_2M} \\ -1 \times (1 - e^{\beta_{G_2M}\mu}(\alpha_{G_2M}\mu+1)) \end{bmatrix}. \qquad (10)$$

The uniqueness and existence of a dominating positive real root ultimately motivates our focus on balanced exponential growth, as any immortal proliferating cell population with sufficient nutrients and space will eventually enter this stationary phase. The time it takes, either starting with a single cell or a synchronized cell cohort to enter this state depends on the cell cycle parameters. The exponential growth rate μ is the unique real positive root of the characteristic equation $\det(\mathcal{Q} - \mu\mathbf{1}) = 0$ which writes as

$$\frac{\mu^3(2 - \prod_i e^{\beta_i\mu}(1+\alpha_i\mu))}{\prod_i(e^{\beta_i\mu}(1+\alpha_i\mu) - 1)} = 0. \qquad (11)$$

It is easy to see that the denominator in Eq. 11 is always positive. To determine a non-trivial μ it remains to solve the transcendental equation in the numerator

$$2 - \prod_i e^{\beta_i\mu}(1+\alpha_i\mu) = 0. \qquad (12)$$

Numerical solutions to this equation can be computed using e.g., the Newton-Raphson root finding algorithm, with fast convergence if the initial value is set to $\mu_0 = \log(2)/\overline{T}$, where $\overline{T}$ is the average cell cycle length, i.e., the sum of the elements in $\bar{\boldsymbol{\tau}} = \{\bar{\tau}_{G_1}, \bar{\tau}_S, \bar{\tau}_{G_2M}\}$. This first guess is a naive estimate for μ assuming that cells divide according to a deterministic division time identical to the average of the hypoexponential density defined in Eq. 3.

Learning from cell frequencies measured under balanced growth

The predicted fractions of cells in each of the phases can be compared to frequencies extracted experimentally from bivariate analysis of cell populations transiently exposed to nucleoside analogs and subsequently examined both for the intensities of the signals due to incorporated nucleoside analog and total DNA content [26] (e.g. the so called BrdU-DAPI staining dot plot). The question that we want to address in this section is: What can potentially be learned about the parameters of the model, given this type of experimental data? By definition, the measured frequencies will sum to one, and therefore we have for three populations effectively only two equations but six model parameters. This makes it impossible to identify all the parameter values, irrespective of the number of samples we take. It is however possible to derive analytical expressions for the upper and lower bounds for both the parameters and the average completion time of each phase.

Consider the experimentally determined frequencies, denoted by $\tilde{\boldsymbol{n}} = \{\tilde{n}_{G_1}, \tilde{n}_S, \tilde{n}_{G_2M}\}$. Substituting the vector $\boldsymbol{n}$ by $\tilde{\boldsymbol{n}}$ in Eq. 10 and solving for each phase specific parameter α, we obtain

$$\alpha = (\kappa e^{-\beta\mu} - 1)/\mu, \qquad (13)$$

where κ is a phase specific element of the vector

$$\boldsymbol{\kappa} = \left\{ \kappa_{G_1}, \kappa_S, \kappa_{G_2M} \right\} = \left\{ \frac{2}{2 - \tilde{n}_{G_1}}, \ \frac{2 - \tilde{n}_{G_1}}{1 + \tilde{n}_{G_2M}}, \ 1 + \tilde{n}_{G2M} \right\}. \quad (14)$$

The phase specific parameters α and β, respectively the reciprocal rate and delay, are by definition greater or equal to zero. These conditions propagate into Eq. 13 which allows us to specify boundaries for α and β. First notice that α is, for each phase, a monotonically decreasing function of β with a maximum $(\kappa - 1)/\mu$ at $\beta = 0$ and a zero crossing at $\beta = \ln(\kappa)/\mu$. The maximum and the root represent the upper bounds for α and β respectively, while the lower bounds are zero for both. We thus have for each phase

$$\alpha \in [0, \ (\kappa - 1)/\mu] \quad \text{and} \quad \beta \in [0, \ \ln(\kappa)/\mu]. \quad (15)$$

The mean phase-specific completion time, $\bar{\tau}$, the sum of the reciprocal rate α and the delay β, is also bounded, with an interval given by

$$\bar{\tau} = (\alpha + \beta) \ \in \ [\ln(\kappa)/\mu, \ (\kappa - 1)/\mu]. \quad (16)$$

This result is derived from the fact that $(\alpha + \beta)$ is concave having its unique minimum at $\beta = ln(\kappa)/\mu$, which follows from setting the derivative $\partial_\beta (\alpha + \beta) = 1 - \kappa e^{-\beta \mu}$ to zero. This implies that $(\boldsymbol{\alpha + \beta})$ is a monotonically decreasing function in the interval $\boldsymbol{\beta} \in [0, \ \ln(\boldsymbol{k})/\mu]$ with the corresponding extrema specified above. It is important to note that the intervals defined by Eqs 13–16 depend on the average growth rate μ which is in general not known. Formally if one specific pair of parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ explains the measured frequencies with growth rate μ, the scaled parameter vectors $c\boldsymbol{\alpha}$ and $c\boldsymbol{\beta}$ mimic equally well the same data for arbitrary positive c, however with a reduced growth rate μ/c. This can be easily verified by substituting these expressions in Eq. 10 and Eq. 12. The direct consequence is that μ remains undefined. However for the relative average time a cells spends e.g., in G_1 phase $(\alpha_{G1} + \beta_{G1})/\overline{T}$ the growth rate cancels out.

Using the fact that $\kappa \in]1,2[$ and the appropriate series expansion for the natural logarithm, the widths of the intervals bounding α, β and $(\alpha + \beta)$ for each phase can be written as:

$$w_\alpha = \frac{1}{\mu}(\kappa - 1) = \frac{1}{\mu} \sum_{i=1}^{1} \frac{(-1)^{i+1}}{i}(\kappa - 1)^i,$$

$$w_\beta = \frac{1}{\mu} \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i}(\kappa - 1)^n, \quad (17)$$

$$w_{\alpha + \beta} = \frac{1}{\mu} \sum_{i=2}^{\infty} \frac{(-1)^i}{i}(\kappa - 1)^n.$$

From this it is straight-forward to show that $w_\alpha > w_\beta > w_{\alpha + \beta}$. This implies that by using measurements of the phase-specific stationary cell frequencies to infer the phase-specific completion times τ results in estimates of the mean value $\alpha + \beta$ that are more precise than the estimates of the standard deviation α. Notice that the width of the intervals can be interpreted as a naive lower bound for the uncertainty about the respective parameter values. For the two data sets analyzed in this article (see details in next section), we computed the intervals for the phase-specific standard deviations w_α that were on average ~ 10 times wider than the intervals for the expected phase-specific completion times $w_{\alpha + \beta}$.

Transient unbalanced growth

Balanced growth analysis does not allow to distinguish between fixed $(\alpha = 0)$ and purely exponentially distributed $(\beta = 0)$ completion times even if μ is known. This follows from Eq. 15 because possible values for the standard deviation α include 0 and $(\kappa - 1)/\mu$, and the latter requires, according to Eq. 16, the delay β to be null.

The incapacity to resolve the values of α and β is overcome if one selects and follows a subpopulation within which the proportions of cells in each phase are transiently different from the balanced growth proportions. Consider a simple *thought* experiment that consists in taking a population under balanced growth and labelling all the cells that are in a specific phase, say o, which can be either G_1, S or G_2M. Initially all the cells are in the same phase o, but as time passes by the labelled cells progress through the cell cycle and eventually distribute over the three phases. The labelled cell subpopulation which is initially not balanced will return asymptotically to balanced growth conditions, restoring the corresponding proportions of cells in the three phases. We refer to this transient dynamics of a selected subpopulation as transient unbalanced growth. It turns out that measuring the transient dynamics of this subpopulation yields information that potentially allows to distinguish between a fixed and a purely exponentially distributed phase completion time. More specifically, a mathematical proof will show that taking samples at three well chosen time points (support points) permits under ideal conditions accurate estimation of the average and the variability in the time required to complete the phase o.

The initial average fraction of cells in phase o which are selectively labelled at time t_0 is determined by Eq. 10. To predict when the labelled cells will have completed o, we need to specify when they entered this phase. For the time before labelling the average influx into o is proportional to $e^{\mu t}$. For the time after the labelling, because by definition all labelled cells entered phase o before t_0 (otherwise they would not be labelled 'as being in phase o'), the entry of cells is zero. Hence, the average influx to the labelled subpopulation is proportional to $H(t_0 - t)e^{\mu t}$, where H denotes the Heaviside step function. Let us assume that within the subpopulation of labelled cells and their progeny one could identify how many phases a cell or a cohort of cells went through since the labelling event, and let P count the number of phases since labelling.

In close analogy to expression Eq. 5 we compute the time-dependent exit-rate density distribution for cells with $P = 0$ as

$$\begin{aligned} \gamma_{0 \to 1}^o(t) &= \frac{\int_{-\infty}^{t} H(t_0 - x)e^{\mu x} f_{\tau_o}(t - x)dx}{\int_{-\infty}^{t_0} H(t_0 - x)e^{\mu x} R_{\tau_o}(t_0 - x)dx} \\ &= \frac{\int_{-\infty}^{t_0} e^{\mu x} f_{\tau_o}(t - x)dx}{\int_{-\infty}^{t_0} e^{\mu x} R_{\tau_o}(t_0 - x)dx} \\ &= \frac{\mu \mathcal{L}_\mu \{f_{\tau_o}(\tau + t)\}}{1 - \mathcal{L}_\mu \{f_{\tau_o}(\tau)\}} \quad \text{for} \quad t_0 = 0. \end{aligned} \quad (18)$$

where, for convenience, we interpreted and will interpret in the following o both as a phase and a phase index. As before, the third row follows from the definition of the Laplace transform setting $t_0 = 0$.. On the left-hand side, the arrow from 0 to 1 represents the transition from the initial phase o ($P = 0$) to the next phase ($P = 1$), corresponding to the completion of the initial phase o. In contrast to Eq. 5, the denominator accounts for the cells that entered or initiated phase o sometime in the past, and did not complete this phase until the instant of labelling t_0 (and not at time t as in Eq. 5), while the numerator, except for the altered average influx, remains unchanged.

After computing $\mathcal{L}_\mu\{f_\tau(\tau + t)\}$ and substituting $\mathcal{L}_\mu\{f_\tau(\tau)\}$ using Eq. 2, Eq. 18 yields for $t > t_0$

$$\gamma^o_{0\to1}(t) = \begin{cases} \dfrac{\mu e^{\mu(t-t_0)}}{(\alpha_o\mu+1)e^{\beta_o\mu}-1} & t < t_0 + \beta_o \\[3mm] \dfrac{\mu e^{\frac{\beta_o + \alpha_o\beta_o\mu - t + t_0}{\alpha_o}}}{(\alpha_o\mu+1)e^{\beta_o\mu}-1} & t \geq t_0 + \beta_o \end{cases}. \qquad (19)$$

It follows that the accumulated average cell flux that at time t has completed o and progressed to the next phase is given by

$$\Gamma^o_{0\to1}(t) = \int_{t_0}^{t} \gamma^o_{0\to1}(x)\,dx, \qquad (20)$$

which for $t \to \infty$ approaches one, reflecting the fact that all cells will eventually complete o.

The Laplace transform of Eq. 20 writes as

$$\mathcal{L}_\omega\left\{\Gamma^o_{0\to1}(t)\right\} = \frac{\mu + \alpha_o\omega\mu - \mu e^{\beta_o(\mu-\omega)}(1+\alpha_o\mu)}{\omega(e^{\beta_o\mu}(1+\alpha_o\mu)-1)(\omega-\mu)(1+\alpha_o\omega)},$$

where ω is, as before, the transformed variable corresponding to t.

Within a cohort of cells isolated for instance in $\mathbf{S}$ phase, i.e., $o = \mathbf{S}$, the accumulated average cell flux out of the subsequent $\mathbf{G_2M}$ phase can then be derived recalling Eq. 2 and using the properties of the inverse Laplace transform as

$$\Gamma^{\mathbf{S}}_{1\to2}(t) = \mathcal{L}_t^{-1}\left\{\mathcal{L}_\omega\left\{\Gamma^{\mathbf{S}}_{0\to1}(t)\right\} \times \frac{e^{-\beta_{\mathbf{G_2M}}\omega}}{1+\alpha_{\mathbf{G_2M}}\omega}\right\}. \qquad (21)$$

For an arbitrary cell cohort originally in o, the accumulated average flux, completing P phases and entering the $(P+1)^{th}$ phase since isolation, can be written in general as

$$\Gamma^o_{P\to P+1}(t) = \mathcal{L}_t^{-1}\left\{\mathcal{L}_\omega\left\{\Gamma^o_{0\to1}(t)\right\} \times \prod_{p=1}^{P} \frac{e^{-\beta_{q(o,p)}\omega}}{1+\alpha_{q(o,p)}\omega}\right\}, \qquad (22)$$

in which $q(o,p)$ denotes a function which returns an appropriate phase index. For $p \in \mathbb{N}$ and $o \in \{\mathbf{G_1}, \mathbf{S}, \mathbf{G_2M}\}$ it is defined as

$$q(o,p) \overset{\text{def}}{=} \begin{cases} \phi_{0+(p\bmod 3)} & \text{if } o = \mathbf{G_1} \\ \phi_{1+(p\bmod 3)} & \text{if } o = \mathbf{S} \\ \phi_{2+(p\bmod 3)} & \text{if } o = \mathbf{G_2M} \end{cases},$$

where $\bmod$ is the modulo operation, and $\boldsymbol{\phi} = \{\mathbf{G_1}, \mathbf{S}, \mathbf{G_2M}\}$ is a vector of cell cycle phase indices. The function $q(o,p)$ thus returns, for increasing p, in a cyclical fashion, the cell cycle phase indices, starting with o for $p = 0$. Notice that Eq. 21 corresponds to Eq. 22 for $o = S$ and $P = 1$.

Analytical expression for Eq. 22, although solved relatively easily with modern algebra software, can become quite cumbersome for values of P larger than six. In our case, deriving the expressions for p up to a value of five was sufficient to simulate the experiments.

Because we want to compare the model predictions with experimentally measured cell frequencies, more interesting than the accumulated fluxes are the expected proportions of cells inside each phase over time. These can be computed using Eqs 20–22, closely following the methodology outlined in [11,12]. For the fraction of cells initially in phase o we have

$$n^o_0(t) = \frac{n^o}{e^{\mu t}} \times (1 - \Gamma^o_{0\to1}(t)), \qquad (23)$$

where the lower index 0 in $n^o_0(t)$ indicates that this expression describes cells which completed zero phases since t_0. The first term on the right hand side corresponds to the fraction of cells in phase o at t_0 divided by $e^{\mu t}$, which accounts for the total population growth during the same interval. The second term stands for the fraction of cells that remained in phase o up to time t relative to the initial number of cells in this phase. By evaluating the integral in Eq. 20, substituting in Eq. 23 and letting as before, without loss of generality, the time of partition t_0 be zero, we get for $t > 0$

$$n^o_0(t) = \frac{n^o}{e^{\mu t}} \times \begin{cases} 1 + \dfrac{1-e^{\mu t}}{(1+\alpha_o\mu)e^{\beta_o\mu}-1} & t < \beta_o \\[4mm] \dfrac{\alpha_o\mu e^{\frac{\beta_o + \alpha_o\beta_o\mu - t}{\alpha_o}}}{(1+\alpha_o\mu)e^{\beta_o\mu}-1} & t \geq \beta_o \end{cases}. \qquad (24)$$

Expressions for cells initially in $\mathbf{S}$, $\mathbf{G_1}$ or $\mathbf{G_2M}$ phase can be obtained by substituting o by the respective phase.

If there were no cell division (i.e., $\mu = 0$) we could readily obtain the average fraction of cells that completed P phases at time t as the difference between the cells that entered the Pth phase, i.e., $\Gamma_{P-1\to P}(t)$, and those that left it, i.e., $\Gamma_{P\to P+1}(t)$, divided by $e^{\mu t}$. To account for cell division, we need to multiply this difference by an additional term λ^o_P, which increases by a factor 2 each time cell cohorts make a transition from $\mathbf{G_2M} \to \mathbf{G_1}$. This term is defined, for each case, as follows: $\lambda^{\mathbf{G_1}}_P = 2^{\lfloor\frac{P}{3}\rfloor}$, $\lambda^{\mathbf{S}}_P = 2^{\lfloor\frac{P+1}{3}\rfloor}$ and $\lambda^{\mathbf{G_2M}}_P = 2^{\lfloor\frac{P+2}{3}\rfloor}$, where the brackets in the exponent represent the floor operator.

In general we get for all consecutive phases for cells initially in phase o the relatively manageable expression

$$n^o_p(t) = \frac{\lambda^o_p \times n^o}{e^{\mu t}} \times (\Gamma^o_{(P-1)\to P}(t) - \Gamma^o_{P\to(P+1)}(t)). \qquad (25)$$

As for Eq. 24, the resulting solutions are defined as piecewise-continuous functions in time. Also notice that most expressions in this section can be written in more compact, but less intuitive, vector form, by dropping the initial phase index o and using bold vector notation as before.

Learning from cell frequencies measured in transiently unbalanced growing subpopulations

In this section we will show that data from the transient kinetics generated by our *thought* experiment allows to accurately estimate the average and the variability in the individual completion times. The proof is based on the analytical expressions derived in the previous section, and also on the assumption that the kinetics are acquired under the ideal conditions of large population sizes and no measurement errors. The latter condition, although clearly unrealistic, can always be approached in practice by increasing the number of samples at each support point.

For the sake of generality, consider a subpopulation of cells that are in an arbitrary phase and are labelled at $t_0 = 0$. Assuming that the 'label' does not in any way affect the cell cycle of the cells, the parameters α and β of the labelled subpopulation are the same as those of the full population under balanced growth. Under these conditions, we can obtain α using Eq. 13 and Eq. 14 with the fractions $\tilde{\mathbf{n}}$ of the full population observed at time t_0. Substituting α in the upper row of Eq. 24 and solving for μ to find

$$\mu = \frac{\log(\kappa\tilde{n}) - \log(\tilde{n} + (\kappa-1)\tilde{n}_0(t_{]0,\beta[}))}{t_{]0,\beta[}}, \qquad (26)$$

where $t_{]0,\beta[}$ denotes an arbitrary time point that lies in the interval $]0,\beta[$, $\tilde{n} = \tilde{n}_0(0)$ and $\tilde{n}_0(t)$ is the experimentally determined equivalent of Eq. 24. This shows that the balanced growth rate μ is fully determined by only two support points, one

immediately after the partition at $t = 0$ and a second at an arbitrary $t_{]0,\beta[}$. This also makes clear that placing more support points in the interval $t_{]0,\beta[}$ does not increase knowledge about μ nor the parameter values, under ideal conditions. Importantly the uncertainty about the phase-specific variability discussed in previous sections remains.

By replacing the same expression for α in the second row of the right-hand side of Eq. 24 we get

$$\tilde{n}_0(t_{[\beta,\infty]}) = \frac{\tilde{n}\, e^{\left(\frac{\kappa\mu t_{[\beta,\infty]} - \beta\mu\exp(\beta\mu)}{\exp(\beta\mu) - \kappa}\right)}}{(\kappa-1)/(\kappa - \exp(\beta\mu))}. \qquad (27)$$

After experimentally acquiring μ and the phase specific κ and $\tilde{n}_0(t_{[\beta,\infty]})$, this expression will depend on a single unknown β. One can show that Eq. 27 is solved by a unique β. This follows from the fact that the right hand side of Eq. 27 is a monotonically decreasing function in $\beta \in [0, \ln(\kappa)/\mu]$ with corresponding values lying in the interval $[\tilde{n}\exp(-\frac{\kappa\mu t_{[\beta,\infty]}}{\kappa} - 1), 0]$ while the left hand side is positive by definition. Substituting the solution for β into Eq. 13 yields the remaining parameter vector α.

Taken together this proves that in theory samples of the three cell cohorts G_1, S and G_2M taken at three support points, a first at $t = 0$, a second at $0 < t < \min(\beta)$ and a third at $t > \max(\beta)$ are sufficient to determine all the parameters of the model.

Conventional single pulse-labelling assays

The *thought* experiment analyzed so far, although conceptually simple, poses a series of experimental challenges, that make a one-to-one realization difficult. The technical difficulties lie mostly in initially separating the cells according to their phase and in following these cells as they enter the subsequent phases. A widely used technique, namely DNA-nucleoside-analog pulse-chase labelling experiments, generates nevertheless to a certain extent
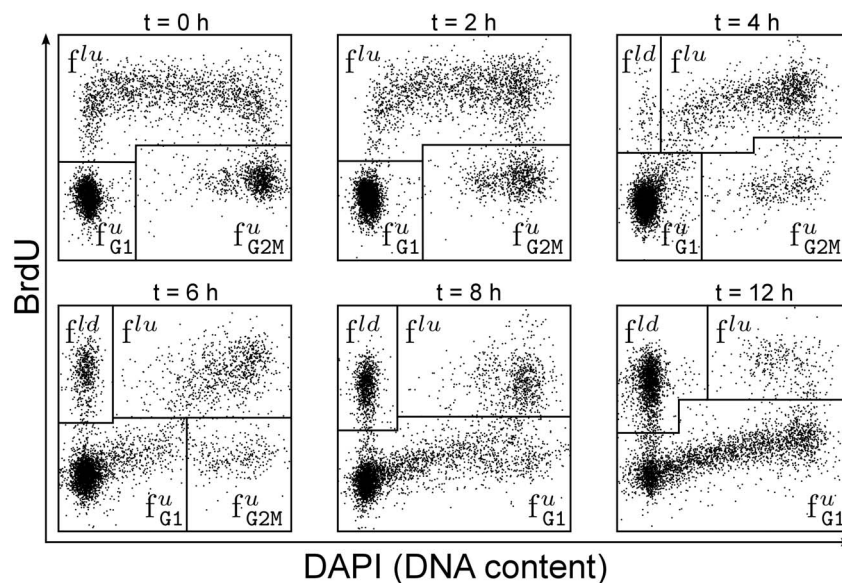

Figure 2. DAPI-BrdU pulse-chase labelling FACS data. Samples taken at several time points after pulse labelling proliferating U87 human glioblastoma cells with BrdU. The four gated populations are f^{lu}, $f^u_{G_2M}$, $f^u_{G_1}$ and f^{ld} which are defined precisely in the main text. Briefly, the subscript indicates the phase at the instant of labelling, while the superscripts 'u', 'lu' and 'ld' refers to cells 'unlabelled', 'labelled and undivided' and 'labelled and divided', respectively. The data was generated as described in the Experimental Methods section.
doi:10.1371/journal.pcbi.1003616.g002

comparable data. The latter achieves the initial phase-specific partitioning by exposing during a short time window proliferating cells with a nucleoside analog (e.g., BrdU, IdU or EdU) that gets selectively incorporated into the DNA of cells that are actively replicating their genome. Measuring subsequently by FACS simultaneously the DNA content and the amount of incorporated nucleoside analog per cell permits to discern the three phases G_1, S and G_2M immediately after the pulse. In addition, due to the permanent staining property of the nucleoside analogs, it is possible to follow, up to a certain degree, the labelled and unlabelled cell cohorts over time. Several dies, such as Hoechst 33342, the dihydroanthraquinone analog DRAQ5, DAPI, and PI are commonly available to stain DNA content in cells [27], and can be used in combination with nucleotide analogs.

In theory, this method would largely correspond to the hypothetical experiment that we analyzed so far. In practice however, the overlap of the subpopulations in the FACS scatter plots prevents the exact determination of the frequencies of cells described by Eq. 24 and Eq. 25. For example labelled cells that have completed the S phase but remain in G_2M phase are indistinguishable from those that did not complete the initial S phase yet. As has been reported previously, only four different subpopulations can be identified with reasonable accuracy [26]. These are:

- f^{lu}: labelled undivided cells which at time of labelling (t_0) were in S phase ($n_0^S + n_1^S$)
- $f_{G_2M}^u$: unlabelled cells that were in G_2M phase at t_0 ($n_0^{G_2M}$)
- f^{ld}: first generation progeny of labelled cells which were initially in S phase ($\sum_{p=2}^{4} n_p^S$)
- $f_{G_1}^u$: unlabelled cells and progeny of cells that were in G_1 at t_0 accompanied by the progeny of $f_{G_2M}^u$ and f^{ld} ($1 - f^{lu} - f_{G_2M}^u - f^{ld}$)

where the corresponding populations in our *thought* experiment are indicated in brackets. This shows that computing Eq. 25 up to $p = 4$ is sufficient to describe a complete *in silico* BrdU pulse labelling experiment. The reason is that, using current protocols, fluorescence of labelled cells becomes indistinguishable from background as soon as the cells divide a second time. In other words, cells that leave population f^{ld} by dividing a second time join population $f_{G_1}^u$ (see Fig. 2). For the experimental data, analyzed in the next section, the fraction of labelled cells that completed two cell divisions during the 12 hours time frame of the experiment is negligible.

The population $f_{G_2M}^u$ is the only sub-population that matches directly the type of data considered before and its temporal evolution follows as such Eq. 24. The remaining three populations in contrast represent mixtures of cell cohorts whose kinetics could be described individually by Eqs 24–25.

Learning from single pulse-labelling data

By analyzing two data sets from samples of **BrdU** single pulse-labelling experiments, we tested the model and the effect of population intermixing on the identification of the model parameter values. The two cell lines considered were *in vitro* cultured U87 human glioblastoma cancer cells (for details see Materials and Methods) and *in vitro* cultured V79 Chinese hamster cells (courtesy G. Wilson). We will refer to these data as the U87 and the V79 data sets. Both data sets consist of samples taken from asynchronously dividing cell populations at several time points after a single BrdU pulse, with sample sizes ranging
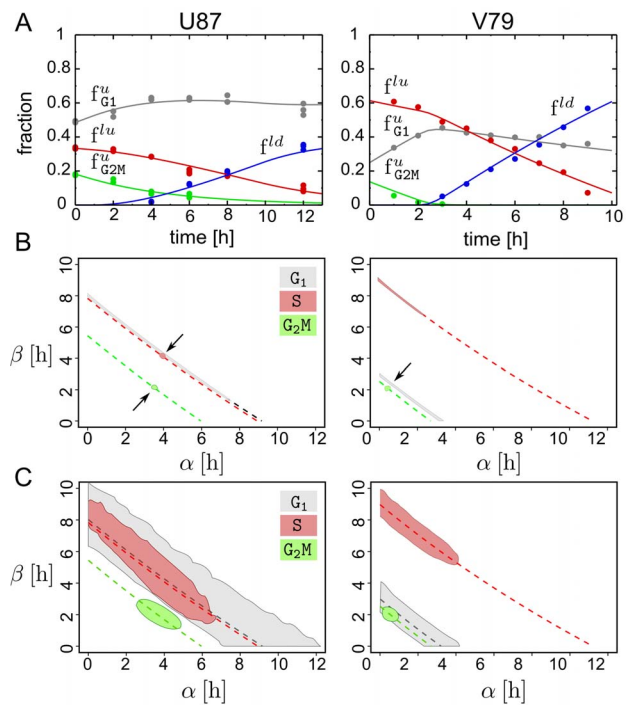

Figure 3. Model based parameter estimation. A: Best fit of the model predictions (lines) to experimentally determined cell fractions after BrdU pulse labelling (dots). U87: In vitro cultured U87 human glioblastoma cancer cell line (three replicates). V79: In vitro cultured V79 Chinese hamster cells (single replicate) (courtesy G. Wilson). Best fit parameter values used to compute model predictions (U87: $\alpha = \{3.2, 3.9, 3.4\}, \beta = \{5.7, 4.1, 2.1\}$, V79: $\alpha = \{1.6, 1.1, 0.5\}, \beta = \{1.4, 7.8, 1.9\}$, units are hours). **B**: Approximate ML regions for the parameters α and β associated to each phase (gray: G_1, red: S, green: G_2M). **C**: Bayesian bi-variate 99%-credibility regions for the parameters α and β for each phase. Arrows indicate point estimates and the dashed lines delineate the information that could have been gained in our *thought* experiment under noise-free conditions from two support points, one at $t = 0$ and a second at $t < \min(\beta)$. The U87 data set was generated as described in the Experimental Methods section. The V79 data set was a kind gift of G. Wilson.
doi:10.1371/journal.pcbi.1003616.g003

from 5000 to 50000 cells each. Data points represent simultaneous measurements of BrdU as well as DAPI or PI (DNA content) in a single cell by fluorescent activated cell sorting.

As a preliminary test we minimized the residual sum of squares (**RSS**), i.e., least-squares fitting, of adequate mixtures of Eq. 24 and Eq. 25 to extracted frequencies at different time points after the pulse. We found that, for properly chosen parameter values, both data sets were reasonably well approximated by the model predictions (Fig. 3 A).

While this indicated that the model captured some of the relevant temporal characteristics of cell cycle progression, a subsequent analysis revealed that an infinite number of different parameter combinations fitted the measured frequencies with the same minimal **RSS** (not shown). This implies that there exist, given the available data, no single best-fit parameter combination, but a whole region in parameter space that can explain the data equally well.

When we then interrogated the same data by approximate maximum likelihood (ML) estimation, using a simple *ad hoc* likelihood function (see Materials and Methods), we found again

Table 1. Bayesian summary statistics.

	G_1			S			G_2M			$\bar{\bar{T}}$
	α	β	$\bar{\tau}$	α	β	$\bar{\tau}$	α	β	$\bar{\tau}$	
U87	4.8	4.1	8.9	3.3	4.8	8.2	3.6	2.0	5.6	22.9
	0.0:11.2	0.0:9.7	6.7:11.7	0.0:5.9	2.0:8.8	7.1:9.5	2.7:4.6	1.3:2.7	5.1:6.2	19.4:26.4
V79	1.6	1.5	3.1	1.3	7.6	9.0	0.5	2.0	2.5	14.7
	0.0:3.6	0.0:3.4	2.5:3.8	0.3:3.6	5.5:9.6	8.4:9.7	0.2:0.7	1.7:2.3	2.3:2.7	13.7:15.9

Bayesian summary statistics (mean, 99%-credibility intervals) for individual cell cycle parameters, average durations $\bar{\tau} = \alpha + \beta$ and the total cell cycle length $\bar{\bar{T}}$. The intervals for the average durations are narrow compared to those for the individual parameters α and β. All values are given in hours.
doi:10.1371/journal.pcbi.1003616.t001

that relative large regions in parameter space mapped to the same ML (see Fig. 3 B). It turned out that these regions were entirely superimposed onto the lines defined by Eq. 13 and Eq. 26 (dashed lines). These lines define what could have potentially been learned in our *thought* experiment with only two support points, one at $t = 0$ and a second at $t < \min(\beta)$. In both experiments, ML parameters associated with the G_1 phase were spread out almost everywhere along these lines (Fig. 3 B, gray regions). Parameters related to the S phase were more concentrated but still in the case of the V79 data a substantial region of ML estimates were observed. Finally the region for the G_2M phase parameters approached that of a point estimate for both data sets.

The spread of the ML estimates suggests that even in the ideal case of large population size and noise-free data, the specific choice of the support points in these experiments does not allow to determine uniquely neither the delay nor the standard deviation for all the phases. In contrast the average completion time for each phase and the total division time can be estimated with relatively high precision.

To better quantify the uncertainty of these estimates, Bayesian 99% credibility regions (CR) were computed by the Markov chain
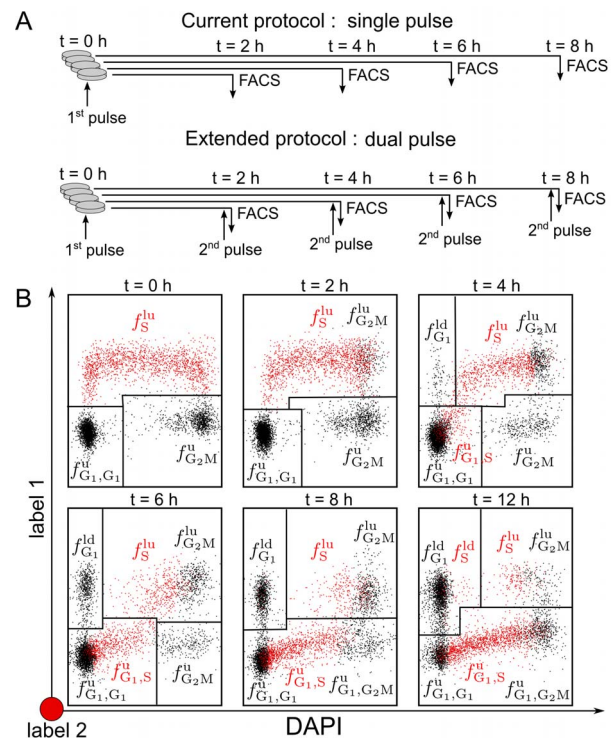

Figure 4. Dual pulse protocol. A: Simplified schematic representations of the protocols corresponding to a conventional single pulse labelling with one nucleoside analog (e.g., BrdU) and a dual pulse labelling experiment with two different nucleoside analogs (e.g., BrdU together IdU or EdU). **B**: Artificial staining of single-pulse labelling data (for original data see Fig. 2), showing eight of the nine subpopulations that could potentially be identified with double-pulse labelling. Notice that the four population f^{lu}, f^{ld}, $f^u_{G_1}$ and $f^u_{G_2M}$ that can be followed by the conventional protocol, have each been subdivided according to the cell cycle phases. The naming convention for the populations is as follows: the superscript (lu = 'labelled undivided', ld = 'labelled divided', u = 'unlabelled') indicates whether the population is labelled and whether it has divided since the time of the first pulse; the first and the second subscript (G_1, S, G_2M) stand for the phase in which the population was at the time of the first and the second pulse respectively. Double subscripts are used only when necessary.
doi:10.1371/journal.pcbi.1003616.g004

Monte Carlo method (MCMC) using the same likelihood function as before (Fig 3 C). CRs followed mainly the same trends as the regions observed in the ML estimates, covered however as expected a larger volume. An exception was the 'blown up' CR of the S phase parameter for the U87 cell line, for which the ML estimates wrongly insinuated a well defined point estimate.

In Table 1 we summarized the obtained Bayesian summary statistics. One can see that the intervals for the average duration of each phase $\alpha+\beta$ are narrow compared to those for the individual parameters α and β. In both cases the data allows for a deterministic S phase ($\alpha=0$), while for the U87 data set variability in G_2M is a necessary characteristic to reproduce accurately the data. Notably, when contrasting the two cell lines, are the short G_1 phase of Chinese hamster cells and the approximately two times more extended G_2M phase of the human glioblastoma cell line. It is out of the scope of this paper to interpret or relate these differences to cell line specific conditions. More importantly in this context is the fact that the information of the analyzed data is too sparse to narrow down all the parameter values even under noise-free conditions.

Redesigned dual pulse-labelling assay

The information extracted from the U87 and V79 data sets is apparently insufficient to pinpoint all six parameters related to the three phases of our simple cell cycle model. This is disappointing especially because the number of support points largely exceeds the three ideally required, and the support points seem to include at least for the U87 data set one at $t=0$, a second at $0<t<\min(\beta)$ and a third at $t>\max(\beta)$.

A potential explanation for this poor resolution in the estimates is the previously mentioned intermixing of the cell population clusters in the BrdU versus DAPI scatter plots compared to ideal conditions discussed earlier. The cluster overlap in the data makes it impossible to measure directly the frequencies of most of the populations, including the cell cohorts described by Eq. 24.

In order to approach the conditions assumed in the *thought* experiment by avoiding the loss of information caused by the intermixing, we devised an extension of the current single pulse protocol, which places a second pulse immediately before measuring or fixing each sample (see Fig. 4, top). The second pulse is expected to expose the cells with a further nucleoside analog that can be distinguished from the first one by **FACS**. Depending on the cell cycle kinetics and the length of the measuring period, the additional pulse increases the number of classifiable populations from four up to nine distinct populations.

To appreciate the additional populations identified by double pulse labelling, data from a single pulse-chase labelling experiment was artificially colored, to mimic the expected FACS output from proliferating cells labelled according to the protocol described before. In Fig. 4, besides the gates defining the populations f^{lu}, f^{ld}, $f_{G_1}^u$ and $f_{G_2M}^u$, cells that have incorporated the second label are drawn in red. For the time immediately after the pulse (i.e., $t=0$), no extra information is gained by the second pulse. However, already two hours later, one additional population can be discerned. Twelve hours after the first pulse, seven population, instead of three, can be recognized. Thus by resolving the four initial population according to the cell cycle phases, it is possible to measure the kinetics of nine subpopulations ($f^{lu}\rightarrow\{f_S^{lu},f_{G_2M}^{lu}\}$, $f^{ld}\rightarrow\{f_{G_1}^{ld}, f_S^{ld}, f_{G_2M}^{ld}\}$, $f_{G_1}^u\rightarrow\{f_{G1,G1}^u, f_{G1,S}^u, f_{G1,G2M}^u\}$, and $f_{G_2M}^u\rightarrow\{f_{G_2M}^u\}$). Because all these kinetics depend on the cell cycle parameters, each of them can in principle tell us something about the phase completions times. However some information is redundant. For example if f_S^{lu} and f_S^{ld} are measured, then $f_{G1,S}^u$ is
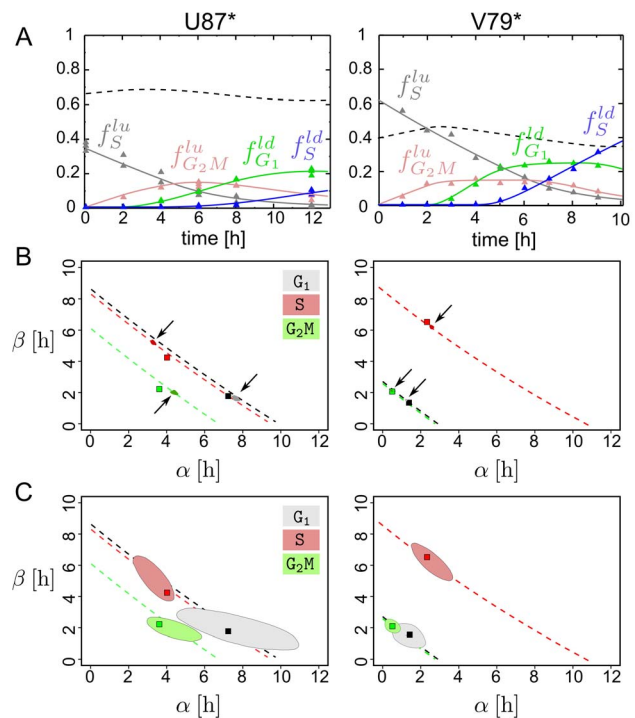

Figure 5. Analysis of simulated dual pulse labelling data. A: Average kinetics of unlabelled (dashed line) and labelled cell cohorts (colored lines) were computed from Eq. 25, using ML parameter estimates from the U87 and the V79 data sets (U87: $\alpha=\{7.1,3.9,3.6\}$, $\beta=\{1.9,4.1,2.1\}$, V79: $\alpha=\{1.4,2.3,0.5\}$, $\beta=\{1.4,6.5,2.0\}$, units are hours). Support points and repeats were chosen according to the real experiments. Multinomial noise was added, mimicking the residuals found in the original data sets (see the Computational Methods section for more details). Finally, model solutions (lines) were fitted to the synthetic data sets (triangles). Best fit parameters (U87: $\alpha=\{7.5,3.2,4.5\}$, $\beta=\{1.7,5.3,2.0\}$, V79: $\alpha=\{1.2,2.5,0.4\}$, $\beta=\{1.4,6.2,2.1\}$, units are hours) **B**: ML parameter estimates from simulated data. All ML regions converge to point estimates (arrows). Squares indicate parameters used for generating the data (see A). **C**: Bayesian bi-variate 99%-credibility regions for the parameters α and β for each phase, based on the artificial data. doi:10.1371/journal.pcbi.1003616.g005

defined by the total fraction of cells in **S** phase, because $n_S=f_S^{lu}+f_{G1,S}^{lu}+f_{G1,S}^u$. Similarly from $f_{G_2M}^{lu}$, $f_{G_2M}^{ld}$ one can deduce $f_{G_1,G_2M}^u+f_{G_2M}^u$, by knowing the frequency of cells in G_2M phase.

Double-label experiments using pairs of nucleoside analogs like BrdU, IdU and EdU, also in combination with radioactive tritiated thymidine ($[^3H]$), have been explored in several cancer cell proliferation studies [19,28–31]. In recent years, dual pulse experiments using BrdU in combination with EdU have become more common. Studies relying on this method estimated changes in **DNA** replication, inferred mitochondrial DNA bio-genesis and stained proliferating cells in the bone marrow in vivo [32–34], in general with the aim to increase the statistical power of the conventional methods.

To assess if the latter method would allow quantifying more accurately and precisely the parameters of the model, we generated *in silico* data mimicking the output of a hypothetical dual pulse experiment using Eq. 24 and Eq. 25 (see Fig. 5 A). We found that by employing the redesigned protocol with the same replicates and time points as in the corresponding data sets, we could reduce the regions corresponding to the ML up to point
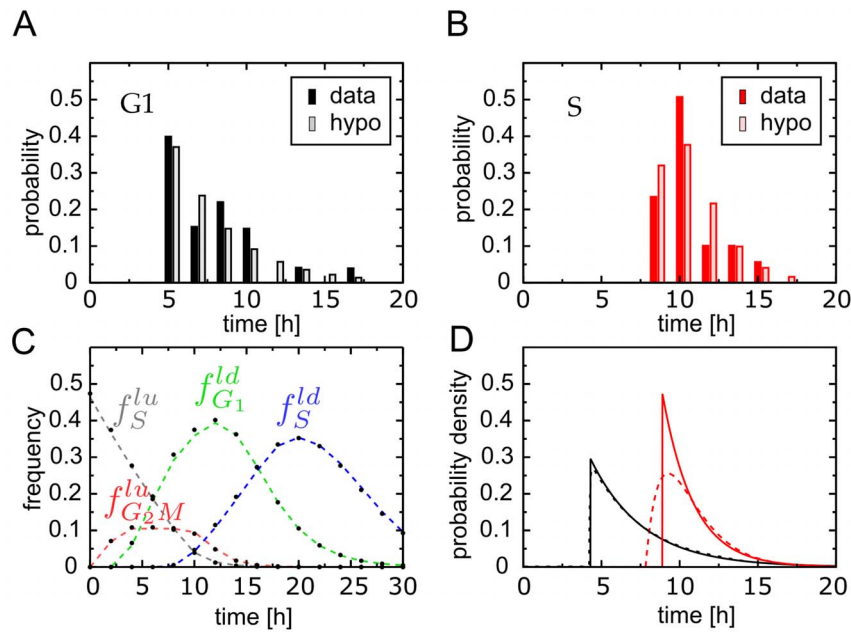
Figure 6. Robustness of parameter estimates to empirical phase duration distributions that are not delayed exponential functions.
A-B: Least-squares fitting of histograms predicted from a hypoexponential distribution with two decay and one delay parameter $\{\alpha_0, \alpha_1, \beta\}$ to measurements of phase durations using fluorescent biosensors [35]. The number of cells that were tracked in the original study was around 15 cells.
C: Best fit of the cell cycle model with delayed exponential completion time distribution densities to synthetic data generated from a model with hypoexponential completion time distribution densities for the G_1 and S phase with parameters as in A and B. **D**: Recovery of the initial distribution densities (solid lines) using the delayed exponential model (dashed line). Both the average and the variability in the S phase completion time distribution (original average: 10.70 h, estimated average: 10.88 h; original std: 2.03 h, estimated std: 1.99 h) were estimated accurately. The data shown in A-B was read from the graphs in the original publication ([35]).
doi:10.1371/journal.pcbi.1003616.g006

estimates (Fig. 5 B). Furthermore, the uncertainties due to noise became also significantly smaller (Fig. 5 C). Pooling this artificial data according to the output expected from a single pulse experiment, reproduced again the uncertainties seen in Fig. 3 C (not shown). Together this indicates that the redesigned dual pulse protocol provides parameter estimates with higher accuracy and precision. Real dual pulse labelling experiments will however be needed to confirm these theoretical predictions.

Robustness of the estimates to other probability distributions of the phase completion times and to concurrent cell loss

The cell cycle model introduced here is deliberately simple and neglects cell loss. In this section, we ask whether the estimates of its parameters are reasonable when some of the simplifying assumptions of the model do not hold. Specifically, we ask how accurate are the mean and standard deviation of the phase completion times estimated using this simple model if the true completion times were not distributed as a delayed exponential function or if there was concurrent phase-specific cell loss.

Empirical measurements [35] indicate that the cycle phase time for the S phase is distributed closer to a delayed hypoexponential or a delayed gamma distribution (see below) rather than the caricatural delayed exponential. Therefore, an important question which arises is how much do the estimates of the average and standard deviation in phase durations obtained with this simple model depend on the true underlying distribution? While many different scenarios could be tested we opted to fit a delayed hypoexponential density with two decay and one delay parameter to direct *in vitro* measurements of G_1 and S phase durations

employing fluorescent biosensors (Fig. 6 A-B, [35]). Using the obtained best-fit estimates, we then performed *in silico* dual-pulse labelling experiments, in which the phase durations were drawn in the case of the S and G_1 phase from delayed hypoexponential density functions (Fig. 6 C). Finally we fitted the simple model, i.e., Eq. 24 and Eq. 25, which is based on delayed exponential distributions, to this data, to see if we could recover the original averages and standard deviations despite using the 'wrong' caricatural model. Both summary statistics (i.e., mean, standard deviation) of phase durations could successfully be re-estimated (Fig. 6 D). Although generalizing this finding lies out of the scope of this article, it suggests that even if the true underlying distribution is not a delayed-exponential function, important quantities like the average and standard deviation of the phase durations may still be estimated with the simple model developed herein. It also indicates that BrdU labelling experiments with a realistic number of samples are unlikely to have the power to discriminate between delayed exponential and more complex density distributions.

We now turn to the issue of how much the presence of phase-specific cell death (or loss in general), which is unaccounted for in our model, affects the accuracy of the estimates of the mean and standard deviation of the phase durations. To this end, we will first introduce the extensions necessary to describe cell death in the model. We rely on the fact that if the probability of death per cell cycle is less that 50%, the average population size will asymptotically grow exponentially with an effective growth rate v, where $0 < v < \mu$. This implies that the arguments used to analyze exponential growth without death remain valid for a model that allows moderate levels of cell death.

To consider death, we assume that cells have two possible fates per phase, either they progress to the next phase or they die. Let $f_\tau(t)$, as before, be the phase completion time density, conditioned however on the cell being alive at time t. And let $f_\delta(t)$ be the phase-specific time to death density conditioned on the cell having not progressed to the next phase. Then, as e.g., in [36], assuming that both events compete with each other (i.e., whatever fate happens first, prevents the other), the resulting density $f_{(\tau,\delta)}(t)$ becomes

$$f_{(\tau,\delta)}(t) = f_\tau R_\delta(t) + f_\delta R_\tau(t). \qquad (28)$$

Consider now a scenario of an exponentially growing population, in which cell death occurs exclusively during phase o. Let us assume further that the $o-$phase specific time to death density $f_{\delta_o}(t)$ is a simple exponential density with mean ρ. Using straightforward probabilistic arguments, we can compute analytically, for this simple scenario two important quantities, namely the probability to die in this phase (p_{δ_o}), and the expected value of the effective completion time, distributed as $f_{(\tau,\delta)}(t)$. We get

$$p_{\delta_o} = \int_0^\infty f_{\delta_o}(x) R_{\tau_o}(x) dx = 1 - \left(\frac{\rho}{\rho + \alpha_o}\right) e^{-\beta_o/\rho}, \quad \mathbb{E}[f_{(\tau,\delta)}(t)] = \rho p_{\delta_o}.$$

Note that, in this simple case, p_{δ_o} is also the probability to die per division cycle.

Evaluating Eq. 18 using $f_{(\tau_o, \delta_o)}$ instead of f_{τ_o}, we obtain for Eq. 24,

$$n_{0,\delta}^o(t) = \frac{n_\delta^o}{e^{\mu_\delta^- t}} \times$$

$$\left\{ \begin{array}{ll} 1 + e^{-\frac{t}{\rho_o}}\left(1 - e^{\frac{t}{\rho_o}} - \dfrac{\rho_o\left(e^{(\mu + \frac{1}{\rho_o})t} - 1\right)}{(\alpha_o + \rho_o + \mu_\delta \alpha_o \rho_o) e^{\beta_o(\mu_\delta + \frac{1}{\rho_o})} - \rho_o}\right) & t < \beta_o \\[6mm] \dfrac{\alpha_o(1 + \mu_\delta \rho_o) e^{\mu_\delta \beta_o + \frac{(\alpha_o + \rho_o)(\beta_o - t)}{\alpha_o \rho_o}}}{(\alpha_o + \rho_o + \mu_\delta \alpha_o \rho_o) e^{\beta_o(\mu_\delta + \frac{1}{\rho_o})} - \rho_o} & t \geq \beta_o \end{array} \right. , \qquad (29)$$

where μ_δ and n_δ^o represent the equivalents of μ and n^o, we had previously defined for the case of no cell loss. The former quantities, which now depend on ρ_o, are derived applying to Eq. 5 the same substitution as above. Expressions equivalent to Eq. 10 and Eq. 11 are obtained along the same lines. These become however rather lengthy and are therefore omitted here. Eq. 29 reproduces accurately f_S^{lu} in simulated BrdU pulse labelling experiments, if death occurs, as specified above (see Fig. 7 A for an example with $o = S$ and $p_{\delta_S} \in \{0.0, 0.3\}$). The differences between the analytical predictions for f_S^{lu} with 30% death and without death (denoted by Δf_S^{lu}) are, for the parameter sets that we tested, relatively small, and vanish as expected, as p_{δ_S} tends to zero (see Fig. 7 B for Δf_S^{lu} computed at one specific time point ($t = 4$ h) for different values of p_{δ_S}).

To further test, how much both cell death and a completion time with a shape distinct from a delayed exponential may jointly affect parameter estimates, we simulated BrdU pulse labelling experiments, where two major assumptions underlying Eq. 24

were simultaneously violated. First, we assumed a delayed gamma distribution (with shape parameter of two) for the completion time of each phase. Second, we considered cell death during S phase, and adjusted ρ_S, such that p_{δ_S} was either zero or 0.3. The population size (starting with five cells) took about twice as much time to grow to a similar size for $p_{\delta_S} = 0.3$ compared to a the scenario without death (see Fig. 7 C, middle column, for five independent simulations). In addition, the variability in the population sizes between the simulations appeared higher for increased death rates. In contrast, when estimating the mean and variance of f_{τ_S} by non-linear least squares fitting using Eq. 24, the marked changes seen in the population kinetics where not paralleled by changes in the estimates. Both the mean and the variance were accurately determined in both cases (see Fig. 7 C, right column). Taken together, this suggests, that the estimates for the mean and variance of f_{τ_S} using Eq. 24, at least for the reasonable parameter values that we tested, are relatively robust to simultaneous changes in the shape of the completion time, and moderate levels of cell death.

Discussion

In this article, we propose a simple stochastic model that aims at approximating the time it takes for a cell to accomplish the sequential phases of the cell cycle, by defining the completion time in each phase as a delayed exponential density distribution. At first sight this might seem a gross oversimplification of all the processes involved. However, when compared with experimental data, this simplistic model performs surprisingly well.

While the observation that the model reproduces closely the experimental time series has to be interpreted with care, we think its success can be explained by the fact that the probability rule captures simultaneously two important regimes of complex biochemical processes that qualitatively differ in their completion time distribution. As was shown recently by Bel et al. [37] the completion time for a large class of complex theoretical biochemical systems, including models for DNA synthesis and repair, protein translation and molecular transport, simplify either to deterministic or to exponentially distributed completion times, with a very narrow transition between the two regimes depending on the rate parameters. These are precisely the 'ingredients' of the delayed exponential distribution. Under this light our model could be naively interpreted as a sensor that measures approximately the relative contribution of delay and decay processes in each of the cell cycle phases. However, whereas delays connected in series form again a delay, this is not true for decays. Sequentially coupled decays form a process with hypoexponential distributed completion times with a shape similar to the frequency distribution of cell cycle phase completion time reported in [35]. Thus a more flexible model for the completion time of each phase could be a hypoexponential distribution of the family that we are currently using to model the total cell cycle length distribution (i.e., Eq. 3). If instead, processes are not connected in an ordered series but rather concurrent, the times for all the processes to complete is dominated by the largest delay or the smallest decay parameter.

It is tempting to interpret the relative weight of constant delay and exponential decay (i.e., the coefficient of variation) as a measure of the precision of the processes regulating each phase, which in turn might reflect a selective pressure on timing. Tighter pressure might reduce the coefficient of variation, as our results suggest for the S phase when compared to the remaining phases. Yet, this might also reflect the conjunction of many parallel and independent process such as replication forks whose number is expected to increase the timing precision by the law of large
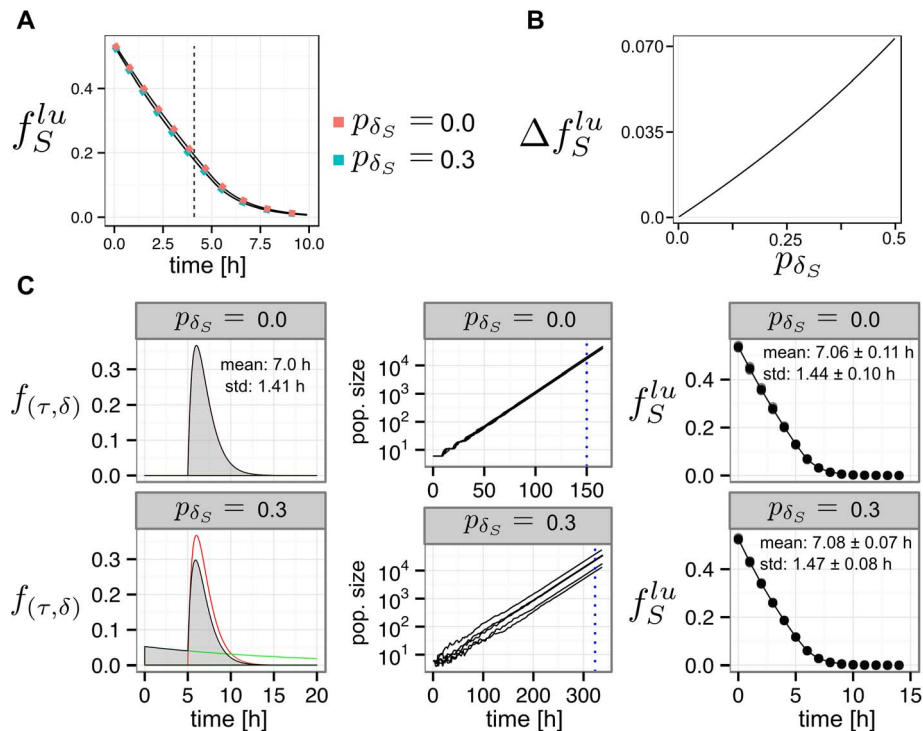
Figure 7. Effect of cell death and completion time distribution on parameter estimates. A: Comparison of analytical predictions (lines, Eq. 29) with simulated BrdU labelling experiment (squares). Cell death is assumed to occur exclusively during S phase with probability 0 (red) and 0.3 (blue) respectively. Only the f_S^{lu} population is considered. Parameters: $\alpha = \{2.5, 2.0, 1.5\}$, $\beta = \{1.0, 5.0, 0.5\}$, units are hours. **B**: Difference between Eq. 29 (accounting for cell death) and Eq. 24 (neglecting cell death) at time $t = 4$ h (see dashed line in A), as a function of p_{δ_S}. **C**: BrdU labelling experiments were simulated assuming gamma distributed phase completion times (red curve, graphs on left column) and cell death during S phase with probability $p_{\delta_S} = 0$ and $p_{\delta_S} = 0.3$ (green curve, graphs on left column). The effective completion time $f_{(\tau, \delta)}$ (gray density plot, left column), the population growth (middle column) and the estimation of the mean and the standard deviation of f_τ are shown for both cases. Approximate confidence intervals for the estimates are computed as 1.96 times the standard error. Even though $f_{(\tau, \delta)}$ and the population growth are strongly influenced by the value of p_{δ_S}, both f_S^{lu} and the estimates extracted from f_S^{lu} are barely affected. The dashed lines in the middle column indicate the time of the first pulse, which was chosen such that the average population was similar in both scenarios. Parameters for gamma distributed completion time distribution of the three phases: shape: $k = \{2.0, 2.0, 2.0\}$, scale: $\theta = \{1.25, 1.0, 0.75\}$, delay: $\beta = \{1.0, 5.0, 0.5\}$.
doi:10.1371/journal.pcbi.1003616.g007

numbers. In fact, the mean time and the variance of the **S** phase are shorter in the early phase of the embryo when cells display a higher number of replication forks in which the DNA polymerase progresses at the same rate [38].

An important simplification of our model consists in the assumption that cell loss by death, differentiation or immigration is small compared to population wide division rates, such that we can neglect it when fitting the model to experimental data. The main reason to adopt this approach was simplicity and the fact that the available data sets did hardly permit the determination of the possibly large number of additional parameters. While for the U87 LIFE/DEAD discrimination was performed, the markers used for gating are specific for late stages of apoptosis or necrosis typically after membrane integrity is lost and therefore do not necessarily reflect the true fraction of dying cells. The fraction of dead cells identified and excluded by this method was typically low. In case that experimental conditions would however suggest substantial cell loss, the model is flexible enough to be adapted without major technical difficulties, along the lines of Eq. 28 and Eq. 29. For instance, when the number of new-born cells equals the number of dying cells, solving the model analytically turns out to be easier, because $\mu = 0$. And given that the apoptotic state (e.g., defined by Annexin-V staining) would be measured simultaneously with nucleoside incorporation and DNA content, this could open up the possibility to assess the duration of apoptosis *in vivo*. These

potential extensions not withstanding, it is reassuring that considering concurrent phase-specific cell death of up 30% may not change the estimates of the mean and standard deviation of the phase completion time obtained using a caricatural model that neglects cell death, as our results indicate.

Another fundamental abstraction of our model is that the completion times for the cell cycle phases of a given cell are uncorrelated, which also implies uncorrelated division times of parental cells and siblings. Even though positive correlation in division times between parental and daughter cells [10] and between siblings [36] has been observed recently *in vitro* by direct long-term microscopy of activated proliferating B cells, Schultze *et al.* reported many years ago for *in vivo* murine crypt epithelial cells the lack of correlation of completion times of a cell through successive phases [31]. It remains to be shown experimentally how much of the correlation or lack of correlation is due to cell type or environment. In any case, it would be interesting to extend the present model to include correlation in phase completion times.

The live cell biosensor-based fluorescent imaging strategy exploited in [35] allows for direct quantification of the stochastic timing of the cell cycle phases. It is worth comparing the estimates of cell cycle phase-specific completion times obtained with this direct method with those provided by the indirect pulse labelling method. The mean S phase completion time was reported for the lines NCI-H292 and HeLa cell line to be 8.2 and 8.4 hours

respectively with standard deviation of 0.5 and 2.9 hours (extracted from Fig.2 in [35]), which lie in the range of the estimates we obtained, despite the different human cell lines that have been analyzed (Table 1). In principle, pulse labelling with nucleoside analogs can be used *in vivo* to quantify the stochasticity of the cell cycle in anatomical places that are currently not feasible to visualize by multiphoton microscopy, given that a sufficiently large (over 1000) and representative sample of cells can be harvested. Our method therefore provides, concerning the G_1, S and G_2M phases, very similar information as these imaging methods, yet it has a much wider application scope.

In comparison with the Smith and Martin cell cycle model, that assumes a single variable phase [5], we have proposed a more complex model with three variable phases. A question can be raised whether a less complex model with variability in only one or two of the three phases would reproduce equally well our BrdU pulse labelling data. This could simplify the analysis and reduce the issue of parameter identification. One might, for example, consider a scenario, similar to the double transition probability model analyzed in [39], in which the G_1 and the G_2 phase have delayed exponentially distributed durations, while the durations of S and M phase are fixed. It is easy to see that such a less complex model is embedded into our model, as it suffices to set $\alpha_S = 0$, while assuming that the variability in the duration of the G_2M phase is generated entirely during the G_2 phase. Clearly, from a data fitting perspective, and especially for the V79 data set, the simpler embedded model and the larger model would perform equally well. This can be read directly from Fig. 3, as the set of approximate ML estimates for α_S includes values that are equal or close to zero. However, the interpretation of the V79 data set based on these two models would be fundamentally different. For instance, by relying on the deterministic model, one would be lead to conclude the S phase duration is for every cell about 9 hours. By allowing however for $\alpha_S \geq 0$, possible interpretations of the data encompass the latter case, but in addition include scenarios in which some cells complete their S phase in about 7 hours, while other cells may take far longer. Even though the original data does not permit to discriminate between these models, simulated dual pulse labelling experiments indicate that this is in principle feasible. Finally, in view of the experimental data provided by Hahn *et al.* ([35], used in Fig. 6 B), the scenario of variable S-phase duration with $\alpha_S \geq 0$ is well justified.

On the other hand, we distinguished only three cell cycle phases, although the cell cycle is typically structured into at least four biologically distinct phases. This simplification stems from the fact that quantification of DNA content by flow cytometry cannot discriminate between cells in the G_2 and M phase. Additional biomarkers, such as pS780 reported by Jaccoberger *et al* [40], could be used together with DNA content dyes and nucleoside analogs in extended labelling protocols to identify the four main cell cycle phases. Extending the model to distinguish accordingly a fourth phase would be rather straightforward mathematically. Despite restricting the model to three phases, it is worth noticing that we are extending the work of Cain and Chau [39,41], who studied both balanced and non-balanced growth conditions, assuming one and two random transitions, mapped respectively to part of G_1 and the remaining cell cycle phases. Also, we extend the work of Larsson *et* al. [42] who were able to infer the variation in the completion times of S and G_2 based on the histograms of DNA content.

Long-term labelling with BrdU has been used *in vivo* to study disease progression of infected rhesus macaques with the simian immunodeficiency virus [1,43] and due to toxicity more rarely in HIV-1 [44]. These studies typically targeted turnover rates of T lymphocytes subpopulation over a time period of several weeks and provided average birth and death rate estimates. In contrast, the method outlined here measures cell proliferation at a much short timescale ($12-24$ hours) and has the potential to yield phase specific estimates of both the average and the variability of completion times. We anticipate that valuable complementary information about SIV and HIV infection could be gained using the redesigned protocol proposed here, especially in the light of the known modulation of the cell cycle checkpoints by accessory viral proteins [45].

Recently, in a computational 'tour de force', Falcetta et al. [25] used a stochastic model of cell cycle progression with discrete age-structure to derive qualitative conclusions about the mechanism of action of several anti-cancer therapies. This model was able to mimic (in their wording 'rendering') quantitative data on single BrdU pulse labelling assay and time-lapse imaging. The empirical distribution cell cycle lengths they reported is akin to the hypoexponential family in our model, however, the distributions of phase lengths remain implicit in their simulation framework, in which time is discrete and the parameters are transition probabilities per time step. This prevents knowing how uncertain are the estimates of the phase length variances based on single pulse labelling using their approach.

Dual pulse labelling with a pair of thymidine analogs has been used before to study cell cycle kinetics [19,28,31]. What is common to those studies is scheduling the two consecutive pulses by fixing the time lapse between the pulses, irrespectively of the time at which cell samples are collected for cell cycle phase analysis. It is worth stressing that, according to the present study, specially when the second pulse is timed according to each individual sample (i.e. adjusting accordingly the interval between pulses) one can harness the potential of the model to quantify the mean and variance of the phase-specific time. Making the second pulse at a fixed minimal time before collecting cells for analysis allows to resolve cellular cohorts, which would otherwise be confounded.

New technologies like the one developed by Hahn et al. [35] but also the ubiquitination-based cell cycle indicator, termed 'Fucci' [46] will greatly increase our understanding of phase resolved cell cycle progression and unveil its epigenetic and stochastic variability in isogenic cell populations. To translate this knowledge gained mainly from *in vitro* cell cultures into an *in vivo* context, long term (greater than 12 hours) and continuous multi-photon imaging may be required. This however is technically very demanding, and may remain prohibitive for cells deep inside tissues despite major technological advances in the field. The methodology presented here allows to measure phase specific cell cycle progression variability *in vivo* by relatively simple technical means. Even though nucleoside analogs are potentially carcinogenic, the adverse effects of low dose pulse labelling remain typically undetectable. Determining accurately cell cycle progression variability in mouse models of cancer might become a crucial step in understanding the high variability in susceptibility to cell cycle specific anti-cancer drugs.

Materials and Methods

Stability analysis

Here we will show that a cell population that follows the stochastic model specified before will eventually enter a stationary exponential growth phase. The requirement for such an asymptotic behavior is, recalling Eq. 11, that the complex valued function

$$Q(\mu) = 2 - \prod_i (e^{\beta_i \mu} + \alpha_i \mu e^{\beta_i \mu}) \qquad (30)$$

has for positive valued elements of the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ a unique positive real root which represents the upper bound of the real part of any of its other potentially infinite number of roots. The complex numbers μ that solve Eq. 30 correspond, according to our model, to the stationary phase growth rate of the proliferating cell population. In case that μ is real, the population is growing exponentially, while if μ is purely imaginary growth is oscillating. In general, roots have both non-zero real and imaginary parts, which leads to oscillations with growing or decaying amplitude. If for real x and y we write $\mu = x + iy$ the real and imaginary part of $Q(\mu)$ are computed as

$$\text{Re}(Q) = 2 - e^{Bx}(\psi_1 \cos(By) - \psi_2 \sin(By)),$$

$$\text{Im}(Q) = -e^{Bx}(\psi_2 \cos(By) + \psi_1 \sin(By)), \qquad (31)$$

where

$$B = \sum_i \beta_i,$$

$$\psi_1 = \prod_i (1 + \alpha_i x) - y^2 \left(\sum_i \alpha_i^{-1} + 3x \right) \prod_i \alpha_i, \qquad (32)$$

$$\psi_2 = y \left(\sum_i \alpha_i + (2 \sum_i \alpha_i^{-1} + 3x^2 - y^2) \prod_i \alpha_i \right).$$

For μ to be a root of Q, both real and imaginary part have to vanish. We restrict our analysis to the positive complex half plane, i.e. $x \geq 0$, since we are interested in growing and not contracting cell populations. Due to the symmetries in the trigonometric functions $\sin(-y) = -\sin(y)$ and $\cos(-y) = \cos(y)$ and $\psi_1(-y) = \psi_1(y)$ and $\psi_2(-y) = -\psi_2(y)$ one can easily see that if $\mu = x + iy$ is a root, its complement $\mu^* = x - iy$ is also a root. We can thus reduce the analysis even further to values with positive imaginary parts. If for fixed x we plot Q in the complex plane as a parametric function of $y \in [0, \infty]$ we get a spiral with the distance from a center point $c = 2 + i\,0$ given by

$$r = \sqrt{(\text{Re}(Q) - 2)^2 + \text{Im}(Q)^2},$$

$$= e^{Bx} \sqrt{\prod_i ((1 + \alpha_i x)^2 + \alpha_i^2 y^2)}. \qquad (33)$$

Crucially, as r is a monotone increasing function of y, the spiral never crosses itself. For $y = 0$ the imaginary part of Q vanishes as expected because $\lim_{y \to 0} \sin(wy) = 0$ and $\lim_{y \to 0} \psi_2 = 0$. For this
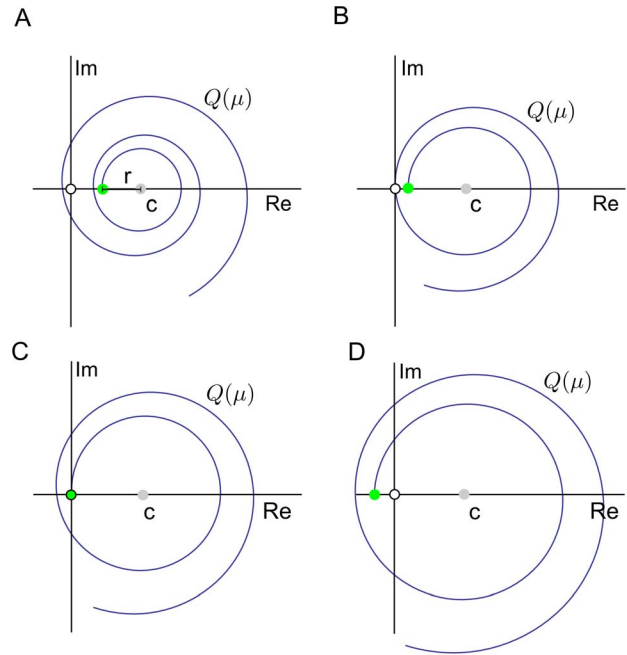
Figure 8. Stability analysis. $Q(x + iy)$ as a function of $y \geq 0$ for fixed values of $x \geq 0$. For $y = 0$ (green circle) the real part of Q takes, depending on $x \in [0, \infty]$, a value in the interval $[1, -\infty]$. The values for x are increasing from A-D, while α and β remain unchanged. For relatively low values of x (A-B) the real part $\text{Re}(Q)$ is positive for $y = 0$. After one or several turns, i.e by increasing y the spiral can potentially cross the origin only once (empty circle). In A the spiral misses the origin, while in B the spiral crosses the origin after one turn. Crossing of the origin means that the corresponding complex number $\mu = x + iy$ is a root of Q. In C the spiral starts at the origin. This represents the only real positive root of Q. For initially negative values of $\text{Re}(Q)$ (D) the spiral can never cross the origin because the distance to the center point (gray circle) is already in the beginning for $y = 0$ larger than the distance between the latter and the origin. By increasing y this distance will even grow further according to Eq. 33.
doi:10.1371/journal.pcbi.1003616.g008

special case $\text{Re}(Q) = 2 - \prod_i (e^{\beta_i \mu} + \alpha_i \mu e^{\beta_i \mu})$ is obviously monotone decreasing with x and restricted to the interval $[1, -\infty]$. This means that the spiral can only 'start' in the interval between one and minus infinity. Taken together, this implies that if for $y = 0$ and fixed x, the real part of Q is positive, then there exist a single 'opportunity' to cross the origin, while if negative there exists none. At the border where the real part is zero (Fig. 8 C), the corresponding value of x is the only positive real root. Due to the monotonicity of $\text{Re}(Q)$ any value of x greater than the positive real root will result for $y = 0$ in $\text{Re}(Q) < 0$ which does not admit for any solution. The different possible scenarios are exemplified in Fig. 8.

Experimental methods

Cell culture. Human astrocytoma cells U-87 MG (ATCC-LGC) were routinely cultured with Dulbecco's modified Eagles medium (DMEM, Biochrom AG) supplemented with non-essential amino acids (NEAA, Invitrogen GmbH), heat-inactivated fetal bovine serum (FBS, 10%, Biochrom AG) and additives (penicillin-streptomycin-glutamine, Invitrogen GmbH) in plastic flasks (TPP AG) at 37°C in 5% CO_2-humified incubators and were passaged twice a week using Dulbecco's PBS (DPBS, Apotheke Innenstadt Uni Mnchen) and Trypsin/EDTA (Biochrom AG) before reaching confluence.

Treatment with BrdU. For cell cycle analysis cells $(2.0 \times 10^4 \text{ cm}^{-2})$ were seeded in 75 cm^2 culture flasks and incubated for 24 h followed by the BrdU pulse. For this purpose, medium was replaced by medium supplemented with BrdU (10 μM, Bromodeoxyuridine, Becton Dickinson GmbH), cells were incubated for 30 min at 37°C followed by washing away of BrdU for two times with fresh medium. Cells were then again incubated at 37°C for a designated period of time (0 h, 2 h, 4 h, 6 h, 8 h, 12 h) to measure proliferation over 12 h.

Preparation of samples. Collecting of cells was performed by trypsinization using DPBS, Trypsin/EDTA and medium followed by washing of cells in DPBS. To exclude dead cells from the analysis staining of dead cells was performed. For this purpose cells were incubated for 30 min with fluorescent dye (LIVE/ DEAD Fixable Green Dead Cell Stain Kit, Invitrogen) according to the manufacturers instructions followed by washing with DPBS. Consequent steps of sample preparation were processed using the APC BrdU Flow Kit (Becton Dickinson GmbH). Cells were washed once with Perm/Wash Buffer and fixed for 30 min on ice with Cytofix/Cytoperm Buffer. After washing with BD Perm/ Wash Buffer cells were resuspended in Cytoperm Plus Buffer and incubated on ice for 10 min followed by washing with Perm/Wash Buffer and incubation in Cytofix/Cytoperm Buffer for 5 min on ice. Cells were then washed with Perm/Wash Buffer and incubated with 2 M HCl-Triton (1%) for 30 min at room temperature followed by washing twice with Perm/Wash Buffer. For detection of incorporated BrdU cells were incubated with diluted (1:50) fluorochrome-conjugated anti-BrdU antibody for 20 min at room temperature. Cells were then washed with BD Perm/Wash Buffer and further incubated with DAPI (0.5 μg/ml in staining buffer: 100 mM Tris, pH 7.4, 150 mM NaCl, 1 mM CaCl$_2$ 0.5 mM MgCl$_2$ 0.1% Nonidet P-40) for 30 min at room temperature. All samples have subsequently been stored on ice until acquisition.

Acquisition and analysis. Acquisition of data was performed by measuring fluorescence intensity using a BD LSR II Cytometer at the excitation wavelength of 660 nm for APC and 450 nm for DAPI and the software BD FACSDiva.

Computational methods

Modeling and simulations. Anti-derivatives, equations as well as eigenvalue problems were solved with the help of Mathematica. Stochastic simulations, Markov chain Monte Carlo and optimization algorithms (e.g. least square fitting and Newton-Raphson root finding) were implemented in C++. To fit the parameters of the model to the data we relied on the population based covariance matrix adaptation evolution strategy provided by the C++ library SHARK [47].

In silico data. In order to anticipate and compare the information content in data sets that could potentially be acquired according to the dual-pulse protocol, *in silico* data was generated. The simulated data consisted of frequencies computed according to our model using ML parameter estimates. Noise was added to the frequencies by simulating a sampling process with replacement with frequencies given by the model and a population size of 300 and 600 for the U87 and V79 data set respectively. This reproduced approximately the variability observed in the original data sets. To make comparison with available data reasonable support points were taken to be the same as in the respective data set.

Bayesian inference. When estimating, by FACS analysis, frequencies of cells in different phases of the cell cycle, measurement noise becomes unavoidable. Potential sources of noise include variability in experimental conditions, gating errors,

stochasticity in cell division, FACS measurement errors, and many more. Here we describe an attempt to account, in a simple way, for the observed experimental noise by taking a Bayesian approach. This provides us not only with maximum likelihood estimate regions of the model parameter, but in addition will give us an idea about the uncertainty that we have about the parameter values.

Even though considering all potential sources of noise would be most consistent, the resulting probability model would become far more complex than our initial cell cycle model. To avoid this overload we assume that a relatively simple *ad hoc* multivariate probability density function approximates reasonably well the average and the noise in the observed frequencies at a single time point. This probability density function, which corresponds to the likelihood $\mathcal{P}_i$ of a single measurement event $\tilde{n}_i$, is defined by

$$\mathcal{P}_i(\tilde{\boldsymbol{n}}_i | \boldsymbol{\alpha}, \boldsymbol{\beta}, N; t_i) = \Gamma(N) \prod_{j=1}^{k} \frac{\tilde{n}_{i,j}^{N \times n_{i,j} - 1}}{\Gamma(N \times n_{i,j})}, \qquad (34)$$

where Γ is the Euler gamma function. The right-hand side of Eq. 34 corresponds to a continuous approximation of a scaled multinomial distribution with support $x_j \in [0,1]$ and $\sum x_j = 1$ [48]. The parameter N, which determines the spread of the distribution, can be interpreted as an effective population size. Taking e.g., a sample of size N from a population of cells containing k sub-populations with proportions given by n_i yields frequencies with a probability density approximately distributed accordingly. If N is small the density distribution is broad, while if N becomes large the density distribution becomes narrow.

Following in general terms the notation in the main text, the $\tilde{\boldsymbol{n}}_{i,j}$ denote the k measured population frequencies from experiment i and the $n_{i,j}$ stand for the corresponding frequencies predicted by the cell cycle model. The latter obviously depend on the parameter vector $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and the time t_i.

Having defined the likelihood $\mathcal{P}_i$ for an outcome of a single pulse labelling experiment, the likelihood for the outcomes of a set of m experiments is the product $\mathcal{P} = \prod_{i=1}^{m} \mathcal{P}_i$ under the reasonable assumption that noise in a specific experiment is independent of all the other experiments. By numerically inverting $\mathcal{P}$, using Bayes theorem, one can obtain the posterior and subsequently the uncertainty over the model parameter given the data, the model and prior knowledge.

To estimate the maximum likelihood regions, the posteriors and the uncertainties in the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for the U87 and V79 data sets, we implemented in C++ the adaptive Markov-Chain-Monte-Carlo algorithm proposed in [49]. The estimates for the maximum likelihood regions are obtained by fixing N to a very large value (e.g., 10^5). For Bayesian inference, N was considered as an additional parameter. For simplicity, improper priors uniformly distributed over the positive real number were assumed for all parameter. The first 10^6 steps of the initially 10^7 step-long chains were discarded, and of the remaining chains every $1000'$th step was included in the subsequent analysis. The credibility regions were computed from the resulting MCMC chains using the 'HPDregionplot' routine in the R package 'emdbook', and convergence of the chains were confirmed using the Gelman convergence test.

Author Contributions

Performed the experiments: TSW IJ CS. Analyzed the data: TSW MOG JC. Wrote the paper: TSW MOG JC. Conceived and designed the mathematical modelling: TSW MOG JC. Performed the modelling and simulations: TSW. Wrote the simulation code: TSW.

References

1. Mohri H, Bonhoeffer S, Monard S, Perelson AS, Ho DD (1998) Rapid turnover of T lymphocytes in SIV-infected rhesus macaques. Science 279: 1223–1227.
2. Ribeiro RM (2007) Dynamics of CD4(+) T cells in HIV-1 infection. Immunology and Cell Biology 85: 287–294.
3. De Boer RJ, Mohri H, Ho DD, Perelson AS (2003) Estimating average cellular turnover from 5-bromo-2′-deoxyuridine (BrdU) measurements. Proceedings Biological Sciences 270: 849–858.
4. Zilman A, Ganusov VV, Perelson AS (2010) Stochastic models of lymphocyte proliferation and death. PloS One 5: e12775+.
5. Smith JA, Martin L (1973) Do Cells Cycle? Proceedings of the National Academy of Sciences 70: 1263–1267.
6. Cain SJ, Chau PC (1997) A transition probability cell cycle model simulation of bivariate DNA/bromodeoxyuridine distributions. Cytometry 27: 239–249.
7. Smith JA, Laurence DJ, Rudland PS (1981) Limitations of cell kinetics in distinguishing cell cycle models. Nature 293: 648–650.
8. Nelson S, Green PJ (1981) The random transition model of the cell cycle. A critical review. Cancer Chemotherapy and Pharmacology 6: 11–18.
9. Tyrcha J (2001) Age-dependent cell cycle models. Journal of Theoretical Biology 213: 89–101.
10. Hawkins ED, Markham JF, McGuinness LP, Hodgkin PD (2009) A single-cell pedigree analysis of alternative stochastic lymphocyte fates. Proceedings of the National Academy of Sciences of the United States of America 106: 13457–13462.
11. Lee HYY, Perelson AS (2008) Modeling T cell proliferation and death in vitro based on labelling data: generalizations of the Smith-Martin cell cycle model. Bulletin of mathematical biology 70: 21–44.
12. León K, Faro J, Carneiro J (2004) A general mathematical framework to model generation structure in a population of asynchronously dividing cells. Journal of Theoretical Biology 229: 455–476.
13. Ganusov VV, Pilyugin SS, de Boer RJ, Murali-Krishna K, Ahmed R, et al. (2005) Quantifying cell turnover using CFSE data. Journal of Immunological Methods 298: 183–200.
14. De Boer RJ, Ganusov VV, Milutinović D, Hodgkin PD, Perelson AS (2006) Estimating lymphocyte division and death rates from CFSE data. Bulletin of Mathematical Biology 68: 1011–1031.
15. Goldstein JC, Kluck RM, Green DR (2000) A single cell analysis of apoptosis. Ordering the apoptotic phenotype. Annals of the New York Academy of Sciences 926: 132–141.
16. Gett AV, Hodgkin PD (2000) A cellular calculus for signal integration by T cells. Nature immunology 1: 239–244.
17. Lodish H, Berk A, Kaiser CA, Krieger M, Scott MP, et al. (2007) Molecular Cell Biology. W. H. Freeman, 6th edition.
18. Yanagisawa M, Dolbeare F, Todoroki T, Gray JW (1985) Cell cycle analysis using numerical simulation of bivariate DNA/bromodeoxyuridine distributions. Cytometry 6: 550–562.
19. Ritter MA, Fowler JF, Kim YJ, Gilchrist KW, Morrissey LW, et al. (1994) Tumor cell kinetics using two labels and flow cytometry. Cytometry 16: 49–58.
20. Baisch H, Otto U, Hatje U, Fack H (1995) Heterogeneous cell kinetics in tumors analyzed with a simulation model for bromodeoxyuridine single and multiple labelling. Cytometry 21: 52–61.
21. Liu YHH, Bi JXX, Zeng APP, Yuan JQQ (2007) A population balance model describing the cell cycle dynamics of myeloma cell cultivation. Biotechnology Progress 23: 1198–1209.
22. Sherer E, Tocce E, Hannemann RE, Rundell AE, Ramkrishna D (2008) Identification of age-structured models: Cell cycle phase transitions. Biotechnol Bioeng 99: 960–974.
23. Altinok A, Lévi F, Goldbeter A (2009) Identifying mechanisms of chronotolerance and chronoefficacy for the anticancer drugs 5-fluorouracil and oxaliplatin by computational modeling. European Journal of Pharmaceutical Sciences 36: 20–38.
24. Bernard S, Čajavec Bernard B, Lévi F, Herzel H (2010) Tumor Growth Rate Determines the Timing of Optimal Chronomodulated Treatment Schedules. PLoS Comput Biol 6: e1000712+.
25. Falcetta F, Lupi M, Colombo V, Ubezio P (2013) Dynamics rendering of the heterogeneous cell response to anticancer treatments PLoS Computational Biology 9: e1003293.
26. White RA, Terry NH, Meistrich ML (1990) New methods for calculating kinetic properties of cells in vitro using pulse labelling with bromodeoxyuridine. Cell and Tissue Kinetics 23: 561–573.
27. Darzynkiewicz Z (2011) Critical Aspects in Analysis of Cellular DNA Content. Current Protocols in Cytometry Chapter 7, Unit 7.2.
28. Hyatt GA, Beebe DC (1992) Use of a double-label method to detect rapid changes in the rate of cell proliferation. The Journal of Histochemistry and Cytochemistry 40: 619–627.
29. Bokhari SA, Raza A (1992) Novel derivation of total cell cycle time in malignant cells using two DNA-specific labels. Cytometry 13: 144–148.
30. Pollack A, Terry NH, Van NT, Meistrich ML (1993) Flow cytometric analysis of two incorporated halogenated thymidine analogues and DNA in a mouse mammary tumor grown in vivo. Cytometry 14: 168–172.
31. Schultze B, Kellerer AM, Maurer W (1979) Transit times through the cycle phases of the mouse. Analysis in terms of the mean values and the variances. Cell and Tissue Kinetics 12: 347–359.
32. Bradford JA, Clarke ST (2011) Dual-Pulse labelling Using 5-Ethynyl-2-Deoxyuridine (EdU) and 5-Bromo-2-Deoxyuridine (BrdU) in Flow Cytometry. Current Protocols in Cytometry Chapter 7, Unit 7.38.
33. Cappella P, Gasparri F, Pulici M, Moll J (2008) A novel method based on click chemistry, which overcomes limitations of cell cycle analysis by classical determination of BrdU incorporation, allowing multiplex antibody staining. Cytometry Part A: the journal of the International Society for Analytical Cytology 73: 626–636.
34. Lentz SI, Edwards JL, Backus C, McLean LL, Haines KM, et al. (2010) Mitochondrial DNA (mtDNA) Biogenesis: Visualization and Duel Incorporation of BrdU and EdU Into Newly Synthe-sized mtDNA In Vitro. The journal of histochemistry and cytochemistry 58: 207–218.
35. Hahn AT, Jones JT, Meyer T (2009) Quantitative analysis of cell cycle phase durations and PC12 differentiation using fluorescent biosensors. Cell Cycle 8: 1044–1052.
36. Duffy KR, Wellard CJ, Markham JF, Zhou JHS, Holmberg R, et al. (2012) Activation-Induced B Cell Fates Are Selected by Intracellular Stochastic Competition. Science 335: 338–341.
37. Bel G, Munsky B, Nemenman I (2009) The simplicity of completion time distributions for common complex biochemical processes. Physical Biology 7: 016003+.
38. Nordman J, Orr-Weaver TL (2012) Regulation of DNA replication during development. Development 139: 455–464.
39. Cain SJ, Chau PC (1997) Transition Probability Cell Cycle Model. Part I Balanced Growth. Journal Theoretical Biology 185: 55–67.
40. Jaccoberger JW, Frisa PS, Sramkoski RM, Stefan T, Shults KE, Soni DV (2008) A new biomarker for mitotic cells. Cytometry Part A 73A: 5–15.
41. Cain SJ, Chau PC (1997) Transition Probability Cell Cycle Model. Part II Non-balanced Growth. Journal Theoretical Biology 185: 69–79.
42. Larsson S, Ryden T, Holst U, Oredsson S, Johansson M (1997) Estimating the distribution of the G2 phase duration from flow cytometric histograms. Mathematical Biosciences 211: 1–17.
43. Rosenzweig M, DeMaria MA, Harper DM, Friedrich S, Jain RK, et al. (1998) Increased rates of CD4(+) and CD8(+) T lymphocyte turnover in simian immunodeficiency virus-infected macaques. Proceedings of the National Academy of Sciences of the United States of America 95: 6388–6393.
44. Kovacs JA, Lempicki RA, Sidorov IA, Adelsberger JW, Herpin B, et al. (2001) Identification of dynamically distinct subpopulations of T lymphocytes that are differentially affected by HIV. The Journal of Experimental Medicine 194: 1731–1741.
45. Chang F, Re F, Sebastian S, Sazer S, Luban J (2004) HIV-1 Vpr induces defects in mitosis, cytokinesis, nuclear structure, and centrosomes. Molecular Biology of the Cell 15: 1793–1801.
46. Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, et al. (2008) Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell 132: 487–498.
47. Igel C, Heidrich-Meisner V, Glasmachers T (2008) Shark. Journal of Machine Learning Research 9: 993–996.
48. Johnson NL (1960) An Approximation to the Multinomial Distribution: Some Properties and Applications. Biometrika 47: 93+.
49. Roberts GO, Rosenthal JS (2009) Examples of adaptive MCMC. Journal of Computational and Graphical Statistics 18: 349–367.