Supplementary material for:

Cell Growth and Size Homeostasis in Proliferating Animal Cells

Amit Tzur; Ran Kafri; Valerie S. LeBleu; Galit Lahav; Marc, W. Kirschner

Table of Contents

Introduction

The following supplementary material consists of three sections; first, a careful description of both derivation and meaning of Eq. 2 from the article main text. Second, a description of the experimental approaches used in this study, and third, a critical statistical analysis and justification of the methods used in the study.

Tables of notation

Table 1 – Notation for main text

Symbol used	**description**	**Range of values**
s	Volume (expressed in fl)	500 fl to 3000 fl (see Fig. 1A)
s_n^i	The size of an individual cell, *i*, at a single time $t = n$.	500 fl to 3000 fl (see Fig. 1A)
s_n	The average cell size of the synchronized population at time *n*	500 fl to 3000 fl (see Fig. 3A)
β_n^i	The linear growth constant of a single cell, *i*, at time point *n*.	10 to 200 fl/hour (see Fig. 2A)
N_n	The total number of cells in the population at time $t = n$	
$\phi_n(\beta)$	The probability distribution of β_n^i values at time *n*	0 to 1 (hour/fl)
β_n	The average value of the linear growth constant within the synchronized population at time *n*, i.e. $\beta_n = \frac{1}{N_n}\sum_{i=1}^{N_t}\beta_n^i$ or $\beta_n = \int \beta\Phi_n(\beta)d\beta$	10 to 200 fl/hour (see Fig. 2A)
k_n	The average value of the exponential growth constant within the synchronized population at time *n*.	0.01 to 0.1 hour $^{-1}$ (see Fig. 3E)

c_n^i	The linear cumulative growth constant for cell *i*. Note that in the main text this variable is described as the difference between a cells' size at time $t = n$ to its' size at birth, i.e. $c_n^i = s_n^i - s_0^i$. These two descriptions are interchangeable because $\Delta t = 1$ and thus $s_n^i = s_0^i + \sum_j \beta_j^i \Delta t$ $= s_0^i + \sum_j \beta_j^i = s_0^i + c_n^i$	
c_n	The linear cumulative growth constant. $c_n = \sum_{j=1}^{n} \beta_j$ And since $\Delta t = 1$ $c_n = \sum \beta_j \Delta t = s_n - s_0$	10 to 1000 *fl/hour*
$\varphi_n(c)$	The probability distribution of cumulative growth constants, *c*.	0 to 1 (hour/fl)
v	Growth rate (in fl/hour)	10 to 200 fl/hour (see Fig. 2A)
τ	Cell age. For an ideally synchronized population $\tau =$ *time*. For asynchronous populations different cells correspond to different values of τ at any single time point.	0 to 12 hours
α	The frequency of cell divisions in the asynchronous population.	0.053 *hour* -1
Δ	A random variable describing the size difference between sister cells corresponding to a random mitosis event.	-50 fl to 50 fl (see Fig. 1E)
f_a	The size probability distribution of the asynchronous steady state population.	0 to 1 fl^{-1} (see Fig. 1A)
F_a	The cumulative size probability distribution of the asynchronous steady state population.	0 to 1 (unitless)
F_m	The cumulative size probability distribution of the mitotic subpopulation.	0 to 1 (unitless)
F_0	The cumulative size probability distribution of the newborn subpopulation	0 to 1 (unitless)
f_t	The size probability distribution of the synchronized population at time *t*.	0 to 1 fl^{-1}

δ	The probability distribution of size differences between daughter cells exiting mitosis. Note that differences can be arbitrarily negative or positive, depending on the "directionality" of the subtraction. This distribution is thus, by definition, symmetric around $\mu = 0$.	0 to 1 fl^{-1} (see Fig. 1E)
$\psi(\tau, s)$	The probability distribution function for cell division as a function of cell age, τ, and size, s.	0 to 1 (unitless)

Table 2 –Additional notation for the supplementary online material:

Symbol used	**description**	**Range of values**
$p_n(k)$	The probability distribution of exponential growth constants, k_n^i, in the synchronized population at time *n*.	0 to 1 (unitless)
k_n^i	The exponential growth constant of a single cell, *i*, at time point *n*.	0.01 to 0.1 $hour^{-1}$
$\Omega_{n\beta}(s, \beta)$	The probability of having cells with size s and linear growth constant β at time *n*.	0 to 1 (unitless)
$\Omega_{nk}(s, k)$	The probability of having cells with size s and an exponential growth constant *k* at time *n*.	0 to 1 (unitless)
$\widehat{f}_t$	An approximation for the probability distribution of the synchronized population at time *t* obtained by use of Eq. 2. When $\widehat{f}_t$ is equal to the measured size distribution, f_t, from time *t*, it means that we have approximated the correct functional form of $\varphi(c)$ (see definition below).	0 to 1 (units, fl^{-1})
$\widetilde{c}_n$	The exponential cumulative growth constant. $\widetilde{c}_n = \sum_{j=1}^{n} k_j$	0.001 to 1 $hour^{-1}$
$G_{k}^{t_1 \to t_2}$	A function mapping the size of a cell at time t_1 to its size at time t_2. Note that no relationship exists between times t_1 and t_2 in the sense that t_1 may be either greater or smaller than t_2.	

A note on notation and mathematical terminology

In the text below we have made effort to appeal to a wide audience from a range of disciplines. To achieve this we have made compromises, being highly formal and over-simplifying in some sections and yet highly non-formal and intuitive explanations in other sections. Additionally we provide a Mathematical Appendix section to explain methods that are used in this study but are not very frequently encountered in biological research. Examples of the latter include a brief explanation on how to understand the method of convolution (in the context of our work) and the method of kernel density estimation. Apart from the methodology, one consequence of the multidisciplinary nature of this study is that it requires a use of language that borrows technical jargon from both statistical mathematics and cell biology. Unfortunately, words may sometimes mean different things in these different disciplines. The most crucial example in our case is the term 'distribution function' which most audiences would understand as the probability distribution of some variable. Yet, in probability theory 'distribution function' is typically understood as the 'cumulative probability function' (i.e., the probability that a variable takes on a value *less than or equal* to some value). To avoid such confusions we clarify below some of the potentially more confusing technical formalities:

Probability distributions (cumulative and non-cumulative)

We will use the term '*pobability distribution*' to refer to the '*probability distribution function'* or *density function*. Care should be taken not to confuse this with the term '*distribution function*' which is usually (but not in our text) used to describe the cumulative probabilities. When referring to the cumulative probability function we will specifically use the term 'cumulative distribution'. Further, when describing probability distributions we will distinguish between cumulative to non-cumulative frequencies by use of small caps vs. capital letters. Thus, F(x) would describe the cumulative distribution of x (i.e. the chance of getting a value less than or equal to x) while f(x) would describe the non-cumulative probability distribution (i.e. the chance of getting exactly x).

Parentheses – triangular, curved and curled.

We employ the conventional use of parentheses nomenclature. Specifically, we use curved parentheses, i.e. (x1, x2), denote intervals or ordered lists. For example (t_1,t_2)

should be interpreted as the time interval between t_1 and t_2. In contrast, curled parentheses, {x1,x2, x3}, would correspond to a set of objects in cases where the order of objects does not matter. Examples in our case would be a set of cells, {$Cell_1$, $Cell_2$, $Cell_3$} or a set of cell sizes {s_1, s_2, s_3}. In these latter cases the fact that s_3 comes after s_1 is arbitrary. Last, we use triangular parentheses, i.e. $\langle x \rangle$, to denote the average. For example, $\langle s^i \rangle$ represents the average cell size in the population.

Random variables and the distinction between capital 'S' to small case 's' in representing cell size.

As customary in texts of probability theory, we will use capital vs. small script letters to distinguish random variables from regular variables. In our case, this applies to use of the symbol '*s*' to denote cell volume while '*S*' is a *random variable* describing the volume of a particular cell *drawn at random* from an underlying distribution (see 'Cell size as a random variable' in the 'Mathematical Appendix'). Note that the distinction between 's' and 'S' may not be necessary for a clear intuitive understanding of the manuscript, yet we felt that it was necessary for the formal equations. For example, the notation $P(S = s)$ is the probability that a randomly selected cell from the population would have a size s. For a more formal explanation see Appendix on 'Random Variables'.

Explanation of the Collins-Richmond Equation

The Collins-Richmond equation (Eq. S1) is a conservation of mass equation that states that the number of cells that grow beyond any particular size threshold (e.g. s_0 in Fig S1) must be balanced by the number of new cells generated by cell divisions.

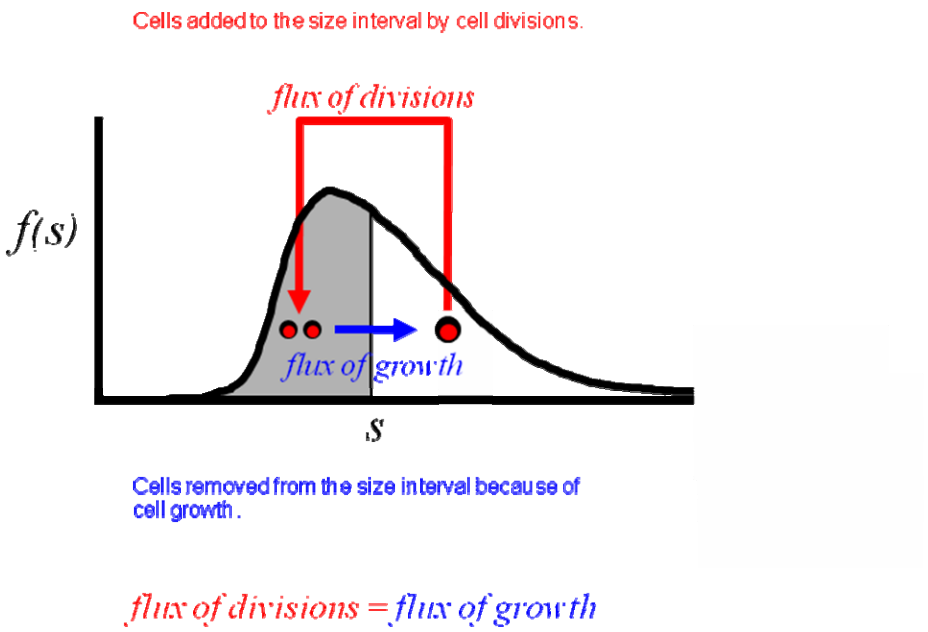


Figure S1: The Collins-Richmond conservation law. The figure depicts the distribution of cell size in an asynchronous steady-state population. To illustrate the Collins-Richmond calculation, consider the proportion of cells that are smaller than or equal to a given size value, *s* (gray region under the curve). At steady-state the flux of newborn cells born with a size smaller then *s* is balanced by the rate at which cells grow past the size threshold (*s*). Some cells in the gray zone (size smaller *s*) may also divide and produce newborns with smaller size and this is taken into account in the calculation. Setting the flux of growth equal to the flux of division and integrating over the whole distribution allows the calculation of growth rate as a function of size.

The Equation can be formulated in terms of fluxes of cells as

$$\underbrace{N_t \alpha F_a(s)}_{\text{total population increase}} = \underbrace{2N_t \alpha F_0(s) - N_t \alpha F_m}_{\text{cell divisions}} \underbrace{- v(s) N_t f_a(s)}_{\text{cell growth}} \qquad \text{(S1)}$$

Where N_t is the total number of cells in the population at time *t*.

$N_t \alpha F_a(s)$ – is the total increase in cell number per unit time of the population of cells that are smaller than (or equal to) *s*.

$2N_t \alpha F_0(s)$ – is the total number of newborn cells, per unit time, that

emerge from cell division with a size of that is smaller than or equal to *s*. The factor '2' is added because there are 2 newborns per cell division.

$N_t \alpha F_m$ – is the total number of cells, with a size of that is smaller than or equal to *s*, that divide (per unit time). Note that $N_t \alpha F_m = N_t \left(F_0 * \delta\right)(s)$, as explained later in this document.

$2N_t \alpha F_0(s) - N_t \alpha F_m$ – is the rate at which the population of cells that are smaller than (or equal to) *s* increases due to cell divisions (blue arrow, Fig S1).

$v(s) N_t f_a(s)$ – is the total number of cells that cross (as a result of growth) the '*s*' size barrier per unit time (orange arrow in Fig. S1).

To obtain Eq. 1 (in the main text) from Eq. S1 one must cancel out the N_t terms in Eq. S1 and rearrange to solve for the growth rate *v*.

Deriving daughter cell size correlation

We first collected microscope images of cells exiting mitosis where the furrow had progressed almost completely and two distinguishable round daughter cells were observed (96 examples). Fortunately the L1210 cells are nearly spherical, with an average axial ratio for a cross section of 1.015 ± 0.010 (N=192), allowing simple segmentation and volume approximation. To ensure that shape was not distorted, we imaged only mitotic cells that did not touch the plastic surface (floating cells). Segmentation was judged by using a GFP reporter localized to the membrane. Fig. 1C shows a typical pair of segmented daughter cells. (See Experimental approaches and Statistics pp, 19-22 for segmentation algorithm, statistics and justification of the spherical assumption). By measuring relative differences between daughters, rather than absolute sizes, we effectively cancel out systematic errors that may arise due to various lens distortions.

Analysis of MOLT4 growth using the Collins-Richmond equation.

We also applied the Collins-Richmond equation on MOLT4 human lymphoblasts to calculate their growth rate as function of size. We generated F_α and F_0 for MOLT4; however we used the $\delta(\Delta)$ obtained from the L1210 to calculate F_m (the shape of MOLT4 cells is more eccentric than L1210 and hence it was more difficult to measure their cross sectional area). The justification for assuming the same daughter cell size correlation for L1210 and for MOLT4 is that the two cell lines are comparable in both the variability of newborn sizes (C.V. = 25%, L1210 and C.V. = 29%, MOLT4) and variability of growth rates (C.V.= 49% L1210 and C.V.= 47% for MOLT4). Furthermore the propagated error in the growth curve due to errors in the daughter cell ratios should be very small. However, the errors in the MOLT4 plot are somewhat greater then the errors in the L1210 plot, which we believe results from a higher contamination of unsynchronized cells in the newborn population.

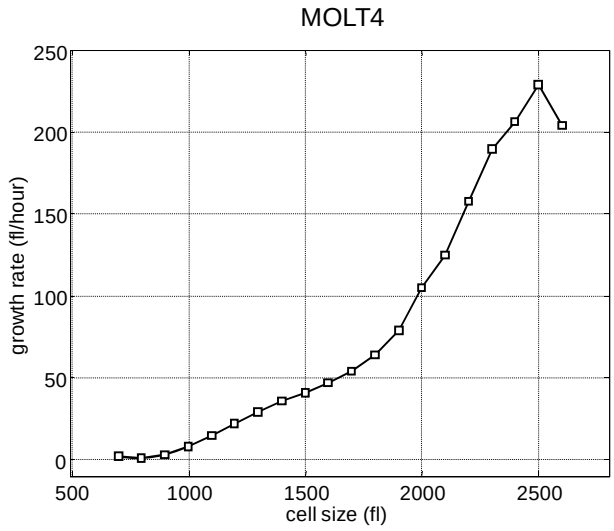


Figure S2: Growth rate as a function of cell size. Mean growth rate (fl/hour) is shown as a function of cell size (fl) for the MOLT4 cell line, using the Collins-Richmond plot. Curve was calculated from the Coulter Counter® measurements of asynchronous size distribution (10^6 cells), the size distribution of newborns (10^5 cells) and the daughter cell size correlation measured for L1210 using the Collins-Richmond method.

Distortions of the Collins-Richmond plot caused by cell heterogeneity

While the Collins-Richmond method is very useful and accurately gains important information, interpreting this information may not be so simple. Mainly, this is due to the growth rate heterogeneity in the population. For example, suppose that cells that grow more slowly during the cell cycle reach smaller volumes when they divide; lower growth rates would then be underrepresented among larger cells and could produce Collins-Richmond plots like those we observe. In Fig S3 and S4 we demonstrate these problems.

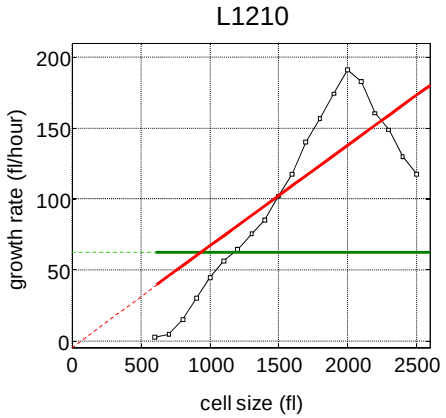


Figure S3: There are major deviations of the observed Collins-Richmond distributions from both linear and exponential cell growth kinetics. The observed growth rates vs. cell size data from Figure 3A (black) is compared to 'purely' linear or exponential growth laws. For this comparison we plot growth curves of a population undergoing ideal exponential growth with, k=0.07 hr^{-1} (red) and linear growth (green). In these idealized growth curves we assumed that there is no intrinsic cell-to-cell variability in growth rates.

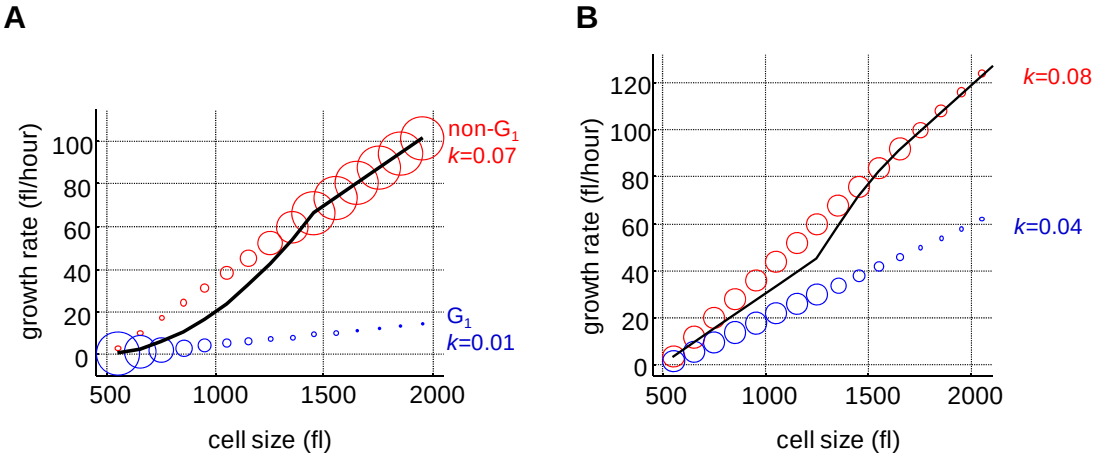


Figure S4: Distortions of the Collins-Richmond plot caused by cell heterogeneity. (A) Schematic demonstration of the effect of the observed G1 growth rate repression on the Collins-Richmond plot using simulated data. Here, a population of cells is considered to either be in G_1 (blue) with a repressed growth rate (k=0.01) or in later stages of cell cycle (red), having non-repressed growth rates (k=0.07). Sizes of the circles indicate the proportion of cells in either of the two categories. As cells increase in size, the proportion of cells in G_1 decreases, as indicated by the smaller sizes of the blue circles. The mean population growth rate (black), which is the weighted average of the two, approaches the rate of the fast growing cells. (**B)**. The effect of a cell-to-cell variation in growth rates on the Collins-Richmond plot. In this simple model, cells in the population are considered to have one of two growth rates: a fast growth rate (red, k=0.08) and a slow growth rate (blue, k=0.04). As shown in Fig 5B, slow growing cells are expected to divide at smaller volumes. Therefore, as indicated by the circles, the proportion of slow growers will decrease with cell size. This decrease causes the mean population growth rate (black) to shift upwards in this plot of growth rate vs. cell size. Although for sake of clarity we considered in this example two discrete growth rate values, in reality a continuous range of growth rates probably exists in the population.

Derivation and explanation of Eq. 2; time dependency of cell growth

In the following text we explain how growth rates are calculated as a function of cell cycle progression (time). The section is provided to clarify the mathematical method employed by Eq. 2 in the main text. In general, our method follows two principles: first, within sufficiently short time intervals growth can be estimated as linear or exponential, or by any other function regardless of the underlying complexity of the 'real' growth function. Second, cell-to-cell variability can be accounted for by integrating over intrinsic growth rates in the population, thus obtaining the mean growth rate at any time point. Despite the apparent redundancy, we will derive formulas and address both the exponential and linear estimates as they rely on sharply contrasting simplifications and are, thus, complementary rather than redundant approaches. In fact, the observation that both the linear and the exponential estimates result in the same growth law demonstrates the power of the method and its' relative insensitivity to the underlying simplifications, and the validity of the results.

The segmental estimates

As noted above, within sufficiently short time intervals cell growth can be estimated by linear, $s(t) = s_0 + \beta\Delta t$, or exponential, $s(t) = s_0 e^{k\Delta t}$ functions, where β and k are the linear or exponential growth constants respectively. For example, with the exponential estimate we say that the growth function between times $t_0 = 0$ and t_1 obeys $s(t) = s(0)e^{k_0 t}$ and between times t_2 and t_3 growth obeys $s(t) = s(t_2)e^{k_2(t-t_2)}$. Note that we assume a different growth constant, k_n, for every time interval (k_0 for (t_0, t_1) and k_1 for (t_1, t_2)).

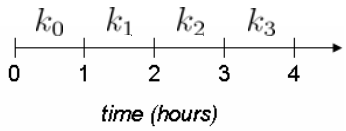


Fig S5: Size distributions were sampled from the synchronized population at times 1, 2, 3, 4 hours after birth. Between each two time points we obtained the growth constant k_j that best estimates growth rates in that interval.

The characteristic growth constants

Thus, the challenge we face is to obtain the growth constants, $k_0, k_1 \ldots k_n$ (or $\beta_0, \beta_1 \ldots \beta_n$ for the linear case) from our measured data, where k_n is the growth constant between times t_n and t_{n+1}. To obtain these values we relied on size distributions sampled from the synchronized population of L1210 cells at different times, $t_1 \cdots t_n$, after birth. Sampling times (times at which distributions were measured), were separated by 1 hour intervals ($t_{n+1} - t_n = \Delta t = 1$ hour). In other words, our data consisted of the probability distribution functions, $f_0, f_1 \ldots f_n$, where f_n is the probability distribution of cell size n hours after birth. Throughout this text we well use t_n to describe time $t = n\Delta t$ where Δt is a one hour interval. n will be used to describe the time interval Δt corresponding to (t_n, t_{n+1}). One question that may be asked at this juncture is whether the time interval of $\Delta t = 1$ hour is a sufficiently short to model growth with the linear or exponential estimates. In principle, the size of the interval that would be sufficient for this purpose

depends mostly on the curvature of the growth function we are trying to obtain. A large curvature would require shorter time intervals. Results in Fig. 4 in the main text suggest that for our case $\Delta t = 1 \text{ hour}$ is sufficiently small.

Cell to cell variation - the characteristic growth constants as random variables

While the linear and exponential estimates given above may describe the kinetics of a single cell, different cells in a population are expected to have different growth rates, i.e. different values of β or k. To account for this cell to cell variability we consider a distribution of values for both β and k. For example, consider the size distribution of cells at birth, f_0, to correspond to the measured set of sizes $\{s_0^i\}$, where s_0^i is the size of a single cell i at birth (time $t = 0$). We will use the notation k_n^i to denote the growth constant of cell i at time interval n. Using the exponential estimates, our goal is to seek the set of values $\{k_0^i\}$ such that $\{s_0^i e^{k_0^i}\}$ will distribute according to the measured probability distribution function f_1. Note that we can write $s_0^i e^{k_0^i}$ instead of $s_0^i e^{\Delta t k_0^i}$ since $\Delta t = 1 \text{ hour}$. Following the same framework, using the linear growth estimates we seek the set of values $\{\beta_0^i\}$ such that $\{s_0^i + \beta_0^i\}$ is best described by the measured f_1.

Following the above description it can be said that for any time interval, n, there exists a distribution of growth constants described by the probability distribution functions $\phi_n(\beta)$ for the linear case, or $p_n(k)$ for the exponential case. Further, for any time interval, n, we have a mean growth constant $k_n = \langle k_n^i \rangle = \frac{1}{N} \sum_{j=1}^{N} k_n^j$ or $k_n = \langle k_n^i \rangle = \int k p_n(k) dk$ as well as the variation of growth constant values. For simplicity in notation we will use the k_n to refer to $\langle k_n^i \rangle$ and β_n to refer to $\langle \beta_n^i \rangle$. Thus k_n (for the exponential case), or β_n (for the linear case), is the average growth constant in the population at time *n*. We will use $\phi_n(\beta)$ and $p_n(k)$ to describe the probability distribution of *k* and β values in time interval *n*.

The assumptions of size independent growth constants and the difference between the linear and exponential estimates

Our model relies on an underlying assumption that growth constants, whether k (exponential) or β (linear), or *c* (for any general growth constant which is not limited by a specific function) are independent of a cells' size at birth.

Formally, this assumption could be written as:

$$\Omega_{n\beta}(s,\beta) = f_n(s)\phi_n(\beta) \quad \text{(assumption for the linear estimate)} \tag{S2}$$

$$\Omega_{nk}(s,k) = f_n(s)p_n(k) \quad \text{(assumption for the exponential estimate)} \tag{S3}$$

where $\Omega_{nc}(s,c)$ is the probability for cells to have size s and growth constant c at time n where $c = k$ using the exponential estimates and $c = \beta$ for the linear estimate.

Yet, this assumption is controlled for by the fact that our results yield a single conclusion regardless of whether the exponential or linear simplification is applied (see below).

While both the exponential and linear estimates may seem equally justified for the estimating growth kinetics in short time intervals, their implementation involves *important and contradictory assumptions*. Specifically, in both cases we apply the simplifying assumption that, in the newborn population, growth constants are not dependent on cell size. In other words, our simplification is that the size of a cell at birth does not correlate with its' growth constant. Note, however, that this assumption *means* different things in the linear and exponential models. In the linear model, described by $s(t) = s_0 + \beta t$, the growth constants, *β*, are *exactly equal to* the growth rates, v, as inferred from $v = \frac{ds(t)}{dt} = \beta$. So, using the linear estimates, our simplification means that *growth rates* are independent of birth (post-division) size. By contrast, in the exponential case, i.e. $s(t) = s_0 e^{kt}$, the growth constant, *k*, *does not* equal growth rate but is rather

obeys the relationship, $v = \frac{ds(t)}{dt} = ks(t)$. Thus, in the exponential case the growth rates will be *dependent* on cell size (growth rate = cell size × growth constant) even if the growth constants are not.

The consequence of these assumptions for *'size independent growth constants'* is that, only for newborns, we are allowed to randomly associate each cell with an arbitrary growth constant sampled from the distributions $\phi_n(\beta)$ and $p_n(k)$, regardless of its' size.

It is important to stress that this simplification of *'size independent growth constants'* was by no means applied to cells at later time points; this would clearly be an invalid assumption and, is not required by our method. In fact, in later times a there is a clear dependency between a cells' size and the value of its' growth constant, as cells with larger growth constants will increase their size more rapidly and, consequently, obtain larger sizes. So, at a later time points in the synchronized population, we expect the larger cells to be the ones that grew faster and hence those that have larger growth constants.

Relating interval growth constants to cumulative growth constants.

Intuitively, it may be expected that the mean growth constant k_n may be obtained from comparison of size distributions f_n and f_{n+1}. Yet implementing this is non trivial since, as explained in the above section, we do not know how growth constants at these times are associated with cell size. To circumvent this problem we calculate *cumulative growth constants* $\widetilde{c}_n$ (or c_n with the linear estimate) by comparing distributions, f_n (for any given n), exclusively with the probability distribution of the newborn population, f_0, where growth constants are assumed to be size-independent. We then obtain the growth constants, Δt and β_n, for the separate time intervals from their cumulative values, $\widetilde{c}_n$ or c_n (Fig S6). To understand this calculation, consider a single cell with size s_0 at birth ($t = 0$). Using the exponential estimate, we search for a multiplier, e^{k_0}, such that $s(\Delta t) = s_0 e^{k_0}$. At time $t = 2\Delta t$ the size of this cell would be:

$$s(2\Delta t) = s(\Delta t)e^{k_1}$$

$= \left(s_0 e^{k_0}\right) e^{k_1} = s_0 e^{k_0+k_1}$

And at time $t = n\Delta t$,

$s(n\Delta t) = s_0 e^{\tilde{c}_n}$

Where $\tilde{c}_n$ is the cumulative growth constant defined by:

$$\tilde{c}_n \equiv \sum_{j=0}^{n} k_j$$

Leading to:

$$\tilde{c}_n = k_n + \tilde{c}_{n-1} \tag{S4}$$

In statistical terms we can say that if $\{s_0^i\}$ are the measured cell sizes at birth and are described by the probability distribution, f_0, then we seek the set of multipliers $\{e^{\tilde{c}_n^i}\}$ such that the distribution of $\{s_0^i e^{\tilde{c}_n^i}\}$ is best described by f_n. Now, assume we can calculate the values of $\tilde{c}_{n-1}$ (or c_{n-1} for the linear model) from the observed probability distributions f_0 and f_{n-1} and $\tilde{c}_n$ (or c_n) from the observed f_0 and f_n (details of how this calculation is performed are given later). Having these values we can then calculate the growth constant, k_n, for the time interval (t_n, t_{n+1}) by $k_n = \tilde{c}_n - \tilde{c}_{n-1}$ (see Eq. S3). While the above relationship was derived using the exponential model, it is straight forward to prove its correctness for the linear model as well.

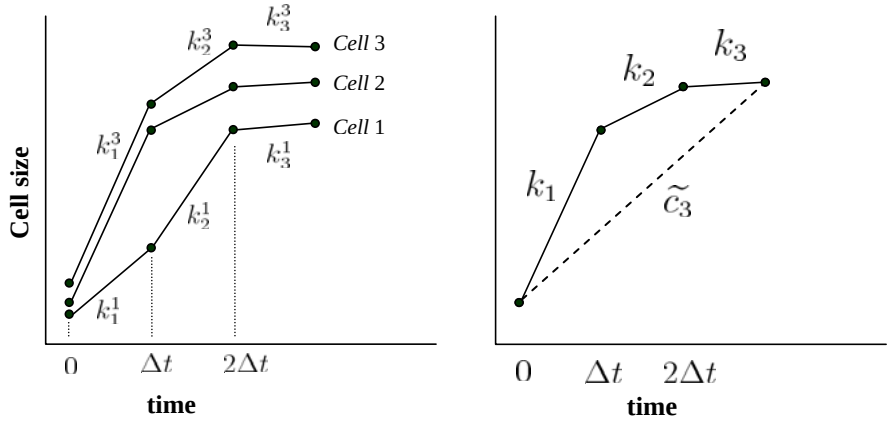


Figure S6: Relationship between growth constants, k_n^i, to the cumulative growth constants, $\widetilde{c_n^i}$. Relationship is shown in trajectories of single cells. **A.** Cell size as a function of time for a population of three cells. Cell-to-cell variation exists in initial cell size and growth constants. **B.** Comparison of cumulative growth constant, $\widetilde{c_3^1}$, with the respective interval growth constants.

A general derivation of the time dependent growth rates for all possible functions

Eq. 2A in the main text and the explanations that precede it describe how mean growth rates are independently calculated for each of the time intervals. As mentioned in the main text, Eq. 2A has been derived for the case of the linear segmental estimates and does not apply to other possible models such as the exponential estimates which we have also used. In what follows we will derive a completely general formula (Eq. S8) for obtaining the time dependent growth rates. We then present the linear (S9) and exponential (S10) formulas as specific cases of Eq S8. Note that Eq. S9 is identical to Eq. 2A from the main text, but was derived more generally. As a reminder, Eq. 2B from the main text is described in this SOM by Eq. S3. As is often the case, the very general equations that follow require a higher abstraction and, consequently, require more effort on the part of the reader.

To obtain generality and simplify writing we will use the notation c_n to describe the cumulative growth constant, regardless of whether the linear, exponential or any other

estimates are used. We will use $\varphi_n(C=c)$, or $\varphi_n(c)$, to describe the distribution of $\{c_n^i\}$ values.

To begin, we define a function, $\underset{n\to 0}{G_c}$ mapping the size of cells with age $t=n\Delta t$ and $c_n=c$ to their size at birth ($t=0$). In other words, $\underset{n\to 0}{G_c}\,(s_n^c)=s_0^c$ where s_n^c is the size of a cell that at time $t=n\Delta t$ has a cumulative growth constant, c. Specific forms of G_c are:

$$\underset{n\to 0}{G_c}\,(s_n^c)=s_0^c=s_n^c-c_n \qquad \text{(linear estimate)} \qquad \text{(S5)}$$

$$\underset{n\to 0}{G_c}\,(s_n^c)=s_0^c=s_n^c e^{-c_n} \qquad \text{(exponential estimate)} \qquad \text{(S6)}$$

To understand the implementation of $\underset{n\to 0}{G_c}$ consider all cells in the population that are characterized by a single cumulative growth constant, $c_n=c$. For *this subpopulation* of cells we can obtain the probability distribution function of size at time $t=n\Delta t$ from their size probability distribution at birth by using standard transformation of variables:

$$f_n^c(s_n^c)=f_0^c\left(\underset{n\to 0}{G_c}\,(s_n^c)\right)\left|\frac{\partial\,\underset{n\to 0}{G_c}}{\partial s_n^c}\right| \qquad \text{(S7)}$$

Where f_n^c is the size distribution of cells with cumulative growth constant $c_n=c$ at age $t=n\Delta t$.

Relying on the assumptions, described earlier, that growth constants are independent of size at birth we have:

$$f_0^c=f_0\varphi_n(c) \qquad \text{(S8)}$$

Introducing Eq. S7 into Eq. S6 we

$$f_t^c(s) = f_0\left(\underset{0\to n}{G_c}(s)\right)\left|\frac{\partial\,\underset{0\to n}{G_c}}{\partial s}\right|\varphi_n(c)$$

And integrating over all values of c we get:

$$f_t(s) = \int_c f_0\left(\underset{0\to n}{G_c}(s)\right)\left|\frac{\partial\,\underset{0\to n}{G_c}}{\partial s}\right|\varphi_n(c)dc \qquad \text{(S9)}$$

Eq. S8, given explicitly for both the exponential and linear estimates is:

$$f_n(s) = \int_{c=0}^{\infty} f_0\left(s-c\right)\varphi_n(c)dc \qquad \text{(linear estimate)} \qquad \text{(S10)}$$

$$f_n(s) = \int_{c=1}^{\infty} f_0\left(se^{-c}\right)e^{-c}\varphi_n(c)dc \qquad \text{(exponential estimate)} \qquad \text{(S11)}$$

We solve Eq. S8 by assuming a Gaussian form of φ_n, with a mean μ and a standard deviation σ.

$$\varphi_n(c;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}exp\left(-\frac{(c-\mu)^2}{2\sigma^2}\right)$$

Introducing $\varphi_n(c;\mu,\sigma)$ to Eq. S8 results in

$$\widehat{f}_n(s;\mu,\sigma) = \int_c f_0\left(\underset{n\to 0}{G_c}(s)\right)\left|\frac{\partial\,\underset{n\to 0}{G_c}}{\partial s}\right|\varphi_n(c;\mu,\sigma)dc \qquad \text{(S12)}$$

Note that $\widehat{f}_n$ represents an estimate for the real probability distribution function f_n assuming that we have the correct distribution function for the cumulative cell growth constants, $\varphi_n(c)$. Our goal is, then, to find the combination of μ and σ that would minimize the difference between our estimate, $\widehat{f}_n$, and the true measured probability distribution, f_n. To quantify this difference we use the function $L = \sum_s \left(\widehat{f}_n(s) - f_n(s)\right)^2$. Since $f_n(s)$ is measured data and $\widehat{f}_n(s)$ is dependent on measured data and on the selected values of μ and σ, L is a function of only μ and σ. i.e. $L(\mu,\sigma) = \sum_s \left(\widehat{f}_n(s;\mu,\sigma) - f_n(s)\right)^2$. Thus, by performing standard computer search algorithms we obtain a unique pair of values, μ and σ, that minimizes L for any given time interval. By performing this calculation on all time points we obtain the mean and variance for any c_n.

Validity and consequence of the Gaussian approximation.

Note that in the case that G is linear or exponential (as is the case in Eq. S10, Eq. S11 and Eq. 2 from the main text) the Gaussian approximation described above is not required and φ_n can be solved for by any method of numerical deconvlution, e.g. the Weiner deconvolution. In fact, to control for the validity of the use of the Gaussian approximation in our text we have additionally solved for φ_n with the Weiner deconvolution and obtained essentially identical results.

An additional test for the validity of the Gaussian approximation is to ask whether the values obtained for L are satisfactory. In other words, whether $\widehat{f}_n(S)$ is a good estimate for $f_n(S)$. Fig S7 shows $f_5(S)$ (blue bars) along with its estimate $\widehat{f}_5(S)$ (red line). It is clear from the figure that obtained estimate is very close to the measured data. More generally, Fig S7 shows the fits obtained from the Gaussian estimate for all time points from $t = 1$ to $t = 8$. At time $t = 9$ and above cells in the population undergo division generating bimodal size distribution of cells before and after division, and our growth calculation no longer describes the measured size distributions.

Another important factor regarding the justification of the Gaussian assumption is the following question: if cumulative growth constants do not normally distribute, how will this affect interpretation of our results. Specifically, consider the possibility that the true distribution of c_n is not φ_n but rather some alternate form $\widetilde{\varphi_n}$. In such case, the relevant question is how would the obtained values μ differ from the values $\widetilde{\mu}$? Since our biological inferences are based only on the mean values we may expect that different distribution functions may lead to similar means, when fitted to the same data. We thus conclude that although the Gaussian form of φ_n is an assumption, results obtained from this method are valuable for inference regardless of whether this assumption is correct.

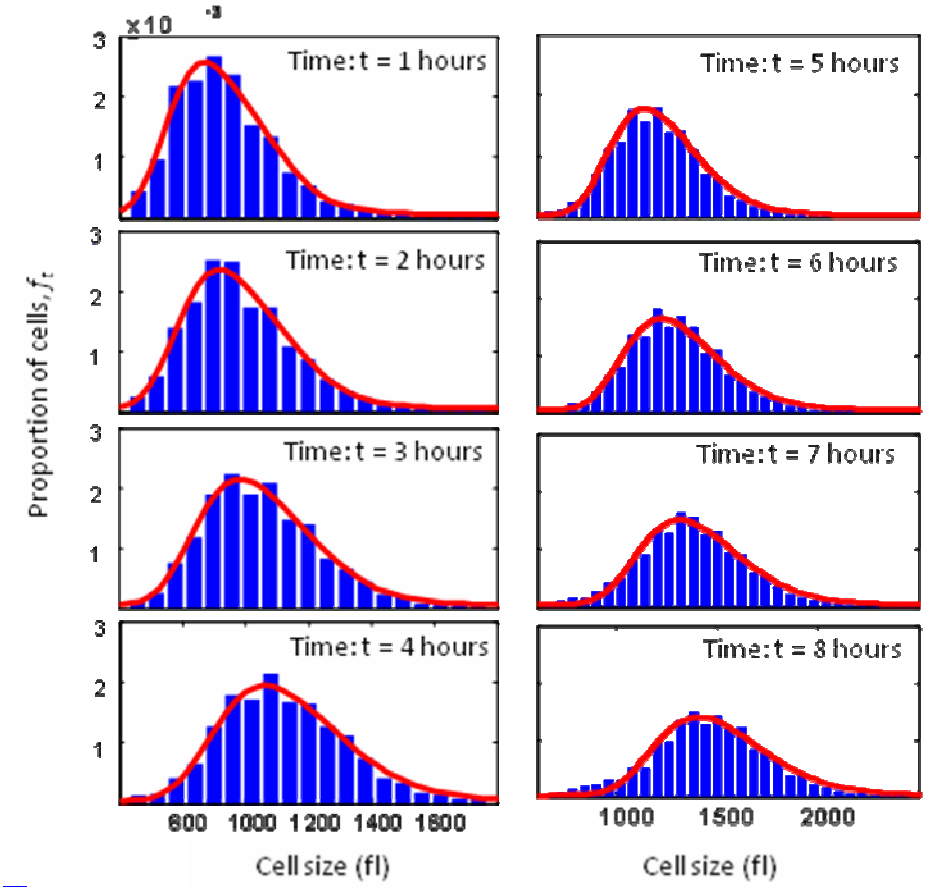


Figure S7: $f_t(S)$ (blue bars) along with its estimate $\widehat{f_t}(S)$ (red line) for t = 5 to 8.

Experimental approaches

Tissue culture and cell synchronization

Mouse L1210 and human MOLT4 lymphoblastoid cells were a gift from Charles Helmstetter. Cells were grown in Leibovitz's L-15 CO2 independent media (Invitrogen) supplemented with 10% FBS (Invitrogen), 1g/L D-(+)-glucose solution (Sigma-Aldrich) and 1% 100X penicillin-streptomycin solution (Gemini).

Newborn cells were generated from a continuously dividing population by mechanical synchronization. Instrument design and preparation, and experimental techniques described in great details in LeBleu *et al* 2006 (see main text, reference *22*). In brief, cell populations synchronized at cell division were generated by growing normally on a coated nitrocellulose membrane and constantly washed in a closed system (for coating we used concanavalin A [Sigma-Aldrich] for L1210 and Poly-D-Lysine [Sigma-Aldrich] for MOLT4). As cells divide, one of two daughters detaches from the membrane, and the population of newborns is centralized by rotation and washed out into a cultured flask. Throughout this procedure cells are maintained at 37°C. The size distribution of both newborn and asynchronous populations was measured by Coulter Counter® (Beckman-Coulter). The quality of synchronization was validated by DNA analysis (propidium iodide) using FACSCalibur® flow cytometer (BD Biosciences).

Cell labeling, image analysis and segmentation

The PH domain of PLC-delta1 tagged with GFP was cloned into the retroviral vector pLNCX2. The viral supernatant was kindly given by Guillaume Charras (UCL, London, UK) and was used for infection of L1210 cells. Single cells expressing GFP were sorted by FACSAria® (BD Biosciences) and maintained in conditioned medium containing 0.4 mg/ml G 418 sulfate (Calbiochem) for two weeks and later in regular culture medium under the same antibiotic selection. For experiments involving live cell imaging we picked a colony of cells with clear membrane labeling and cell cycle and cell size profiles indistinguishable from that of unlabeled cells. During image acquisition, cells were maintained in 32mm uncoated MatTek® dishes at 37°C. We looked for ellipsoidal or dumbbell-shaped cells (in anaphase or telophase) and generated z-stack images (0.5 μm steps, 100X magnification) every 2-3 minutes until the furrow had

progressed almost completely and two distinguishable round daughter cells were observed. We analyzed only cells that remain floating throughout the time-lapse imaging to minimize distortion due to contact with the surface. Images were taken in the Nikon Imaging Centre using Nikon TE2000E Inverted Fluorescence Microscope.

Statistics

Segmentation of round cells:

We used a multistage process to determine cell volumes. From grayscale thresholding (MATLAB®) we obtained rough approximations for the circle enclosing the cell image and for the circle center (see yellow dot in Fig. S8). We then extended rays (see white lines) of fixed width (5 pixels) from the approximate center of the cell and collected the average image intensity within the lines. These lines corresponded to profiles that had a maximum at the cell membrane. By repeating this step for different angles we obtained N=30 points along the cell membrane (see green dots) and used least squares to fit a circle to these N circumference points (see red circle). This method resulted with a mean radius for every cell image and for every focal plane. Cell volume was estimated based on largest mean radius. We defined 'sphericality' as the variance in the length of the rays emanating from the cell center.

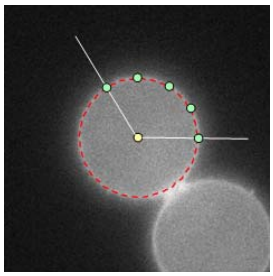


Figure S8: Example of the L1210 segmentation algorithm.

Note that all errors or inaccuracies that may be associated with this microscopy-based method are contained in a single number, namely the value of σ^2. Since the contribution of independent variables to a variance is additive, σ^2 inevitably contains a component of

true daughter cell variation together with a component of experimental error and, thus, represents an upper bound to the true daughter cell asymmetry. Nevertheless, due to the small asymmetry in division, errors in this measurement make an insignificant statistical contribution to the calculated growth curve obtained from the Collins-Richmond equation (see Fig. 3).

Spherical assumption:

As described in the main article text, imaged cells were maintained "floating", thus their geometry was not perturbed by the surface of the plate within which they were imaged. To estimate the deviation from 'sphericality' we calculated the variance of the distances between the identified circle center to the *N* circumference points. The square root of this variance (the standard deviation) was found to be in the order of magnitude of 1% of the circle radius, indicating the high quality of the segmentation algorithm and the justification for the spherical assumption.

Independence between newborn cell size, S_0, and the differences, Δ, between newborns cells after division (justification for calculation of mitotic size distribution)

To calculate the size distribution of mitotic cells we used the convolution equation, $F_m(s) = (F_0 * \delta)(s)$, where $F_0(s)$ is the cumulative distribution of newborn sizes and δ is the distribution of size differences between daughter cells emerging from common mitosis (daughter cells). To understand this intuitively, what we do is randomly associate each newborn with a given Δ. In other words, we randomly sample a newborn size, S_0, from the measured distribution f_0 and add it to a size difference, Δ, randomly sampled from the measured distribution δ to obtain $S_m = S_0 + \Delta$. Affectively, iterating this random sampling and pairing procedure multiple times would result in a set of S_m values that distribute according to $F_m(s) = (F_0 * \delta)(s)$. Note, however, that for this procedure to be valid there must be no correlation between S_0 and Δ otherwise the random pairing is not possible. Fig S9 shows the correlation between S_0 and Δ, for the dataset resulting

from image segmentation. The correlation is weak (0.13), justifying use of the convolution theorem.

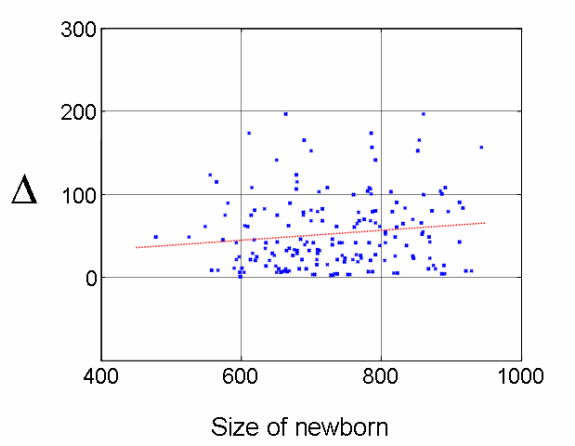


Figure S9: Independence between newborn size and differences between daughter cells. Sizes and size differences are shown in fl. Correlation (Pearson) between the two variables (S_0, Δ) is 0.13 (p value 0.08)

Error analysis of the Collin-Richmond method.

Calculation of growth rates based on the Collins-Richmond method was performed based on the Eq:

$$v(s) = 2\alpha \frac{F_0(s)}{f_a(s)} - \alpha \frac{(F_0 * \delta)(s)}{f_a(s)} - \alpha \frac{F_a(s)}{f_a(s)} \qquad \text{(S13)}$$

Where f_a is the non-cumulative probability distribution in the asynchronous population; F_a is the cumulative probability distribution in the asynchronous population; F_0 is the cumulative probability distribution of the newborn population; $\delta(\Delta)$ is the probability distribution of the size differences between daughter cells after division and α is the frequency of cell divisions in the asynchronous population

Size measurement data for the distributions of f_a, F_a and F_0 were obtained by Coulter Counter®. To approximate the probability distribution function describing size measurements we relied on kernel estimation with a Gaussian kernel, w. Briefly, kernel

density estimation, or the Parzen window method, is a method which, like histograms, provides a means to estimate an underlying probability distribution from measured data samples (see Appendix).

Calculation of error in the probability estimation

To estimate errors on the calculated distribution estimates (f_a, F_a, F_0, δ) we used $f \pm E$ where f is the obtained probability distribution estimate and E is the error or confidence interval:

$$E = 2\sqrt{f}\sqrt{\frac{\epsilon}{Nh}} + \frac{\epsilon}{Nh} \quad \text{(S14)}$$

Eq. S14 was taken from (*1*) page 30. In Eq. S14, ε is a number corresponding to the width of the kernel and is given by $\epsilon = \int w^2(s)ds$. N is the sample size and h is the kernel width as described above.

Note that in ϵ ranges from 0.1 to 0.2 depending on sample size, h ranges from 10 to 20 while total sample size, *N*, ranges from 5×10^5 for the newborn population to 10^6 for the asynchronous population. Thus, errors in frequency estimation are in the order of magnitude of 10^{-5} to 10^{-6}. Example of these confidence intervals for the newborn size distributions is shown in Fig. S10. Note that we have plotted 10 or 100 width confidence intervals, emphasizing the high confidence of the probability distribution function estimates.

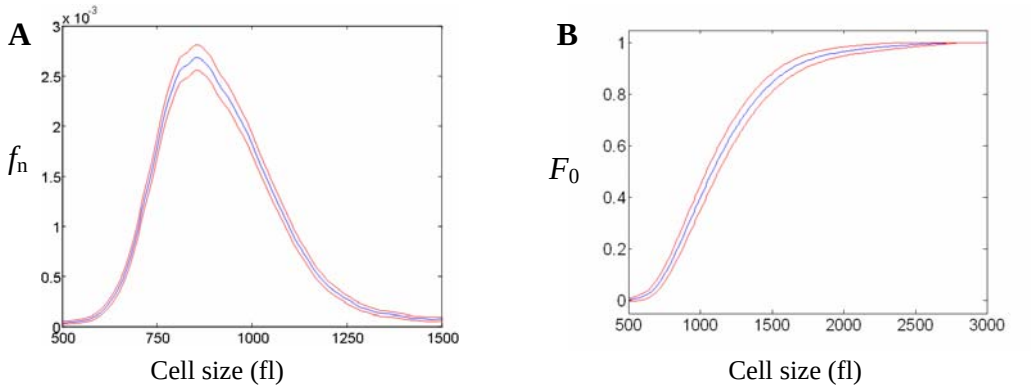


Figure S10. **C**onfidence interval bounds for the cumulative (B) and non-cumulative (A) size distribution of L1210 newborns. Importantly, to emphasize the errors we used 10 confidence interval widths for the non-cumulative density and 100 confidence interval widths for the cumulative density. Thus, the confidence intervals displayed represent exaggerations for the purpose of illustration. For cumulative distributions, error was approximated as binomial and estimated by $E = \sqrt{\frac{F(1-F)}{N}}$.

Effect of errors in probability estimation on calculated growth values

We used the "propagation of errors" method to calculate how measurement errors in the individual measurements affect the growth rate values given by Eq. 1. Specifically, the total error in v is given by:

$$\Delta v = \left|\frac{\partial v}{\partial F_0}\right| \Delta F_0 + \left|\frac{\partial v}{\partial F_a}\right| \Delta F_a + \left|\frac{\partial v}{\partial \delta}\right| \Delta \delta + \left|\frac{\partial v}{\partial f_a}\right| \Delta f_a + \left|\frac{\partial v}{\partial \alpha}\right| \Delta \alpha \qquad \text{(S15)}$$

Where ΔX is the error associated with X. Note that Δ in this context does NOT represent differences between daughter cells, as in the main text. The terms in Eq. 3 can be simply calculated from the Collins-Richmond equation (Eq. 1). Here we simplify by assuming that F_a and f_a are independent measurements (since this assumption can only increase errors, its use in this context is justified). We perform these calculations on the three size measurements obtained from the Coulter Counter® (F_0, F_a, f_a).

$$\left|\frac{\partial v}{\partial F_0}\right| = \frac{\alpha}{f_a}$$

$$\left|\frac{\partial v}{\partial F_a}\right| = \frac{\alpha}{f_a}$$

$$\left|\frac{\partial v}{\partial f_a}\right| = \frac{v}{f_a} = \frac{\alpha}{f_a^2}\left(2F_0 - (F_0 * \delta) - F_a\right)$$

Total error in growth rate calculation.

As seen from Eq 1, α multiplies every term in the equation. Therefore, it is evident that errors in the value of α will not affect the shape of the growth curve we obtained, but only scale it to higher or lower values. Thus to a good approximation errors in growth rate calculation are given by:

$$\Delta v \approx \frac{1}{f_a}\left(\alpha\Delta F_0 + \alpha\Delta F_a + v\Delta f_a\right) \qquad \text{(S16)}$$

Where maximal value of ΔF_0 is 1.5×10^{-3}, the maximal value of ΔF_a is 10^{-3}, the maximal value of Δf_a is 3×10^{-6} and α= 0.05 (for L1210). Thus Eq. 4 can be used to approximate the affective range error which comes to be: $\Delta v = \frac{10^{-4}}{f_a} + \frac{v \times 10^{-6}}{f_a}$. The most obvious insight from this analysis is that, since v is in the range of 10 to 200 (fl/hour), the error in calculated growth rate is limited almost exclusively by the value of the probability distribution function in the asynchronous population. Specifically, to get an error in the range of ± 1 (fl/hour) we need to rely on probabilities as low as $f_a = 10^{-5}$. From comparison with Fig 2 from the main text we see that these probabilities correspond to a cell size greater than 3000 fl, a size that is not considered in our study. Thus, we conclude that error in the growth rate calculation by use of Eq. 1 is limited to under 1 fl/hour, which is insignificant for our purposes.

Mathematical Appendix

Kernel density (or probability distribution) estimation

Kernel density estimation is a method that, based on a sample of measurements, estimates the underlying distribution. For example, imagine the volume of 1,000 cells from a given population. We want, based on these measurements, to estimate the distribution of cell size. Two options are a hand. One is to draw a histogram and the other option is the kernel method, also known as the 'Parzen window method'. Intuitively, the basic principle behind the kernel density estimation method is to 'draw a Gaussian' around each data point and then sum up these Gaussians to build the estimate (*1*).

Cell size as a random variable (small caps s vs. capital S)

A random variable, represented in this text by capital letters, is a function that assigns a probability value (a real number) to every possible result of a random experiment ((*2*) pp. 155). For example, $P(S = s)$ is the probability that a randomly selected cell from the population would have a size s.

What is a convolution and why it is used in the present study?

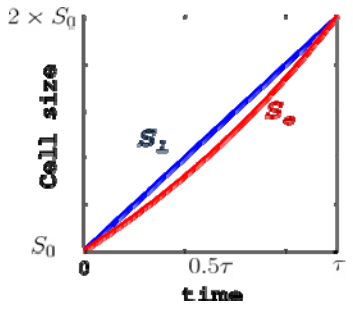
Suppose $f_x(x)$ is the probability distribution of the random variable X and $f_y(y)$ is the probability distribution of *Y*. Now suppose we are interested in the probability distribution, $f_z(z)$ of $Z = X + Y$. In the case that X and Y are independent variables (there is no correlation between x and y) than the answer to our question is $f_z(z) = \int_{a=-\infty}^{\infty} f_x(a) f_y(z-a) da$. The expression $\int f_x(a) f_y(z-a) da$ is called the *convolution* of f_x with f_y and by convention is written with the symbol '*' as $f_z(z) = \int f_x(a) f_y(z-a) da = \Big(f_x * f_y\Big)(z)$ or in short, $f_z = f_x * f_y$. In our study convolution is applied to calculate the mitotic size distribution from the distribution of differences between daughter cells and the distribution of newborns (note: $2 \times S_0 + \Delta = S_m$). Deconvolution denotes the inversion of a convolution. In our case,

Eq. S8 and Eq. 2 in the main text describe the convolution $f_n = f_0 * \varphi_n$. Since f_n and f_0 are known from measurement we seek to 'deconvolute' f_0 from f_n to obtain φ_n. Since there is no simple method for deconvolution, we assume a Gaussian φ_n and minimize as described above.

Calculation of the <6% difference in cell size of the linear vs. exponential growth models

As described in the text, to experimentally distinguish exponential from linear growth requires a measurement resolution sufficient to detect size differences less than 6% of cell size. Intuitively, it may seem unreasonable that linear and exponential growth models differ by such a small extent. In what follows we calculate the maximal size difference predicted between the two models and show that it is 5.63%. Due to the nature of the question, we will use notation that is somewhat simplified from that which has been used throughout this supplement. We, therefore, add a small separate table of notation below.

Figure S11



Linear growth model: $S_l(t) = S_0 + \beta t$

Exponential growth model: $S_e(t) = S_0 e^{kt}$

t - time

τ - length of cell cycle (time)

S_0 - size of cell at birth

$2 \times S_0$ - size of cell at division

Linear Growth	**Exponential Growth**

$S_l(t) = S_0 + \beta t$ Imposing two-fold* $S_0 + \beta\tau = 2S_0$ $\beta = \frac{S_0}{\tau}$ We can thus write the growth equation in terms of $\omega \equiv \frac{t}{\tau}$ $S_l(t) = S_0 + S_0\omega$	$S_e(t) = S_0 e^{kt}$ Imposing two-fold* $S_0 e^{k\tau} = 2S_0$ $k = \frac{ln(2)}{\tau}$ We can thus write the growth equation in terms of $\omega \equiv \frac{t}{\tau}$ $S_e(t) = S_0 e^{\omega ln(2)} = S_0 2^{\omega}$

* The cell size range between birth and division is two fold

The volume difference between the exponential and linear models is:

$$D \equiv S_l(t) - S_e(t)$$
$$= S_0\left(1 + \omega - e^{\omega ln(2)}\right)$$

The maximum difference between exponential and linear models occurs at time $\tau_{0.5}$ such that

$$\left(\frac{dD}{dt}\right)_{t=\tau_{0.5}} = 0$$

$$\frac{dD}{dt} = \frac{dD}{d\omega}\frac{d\omega}{dt} = \frac{1}{\tau}\frac{dD}{d\omega} = \frac{S_0}{\tau}\left(1 - ln(2)2^{\omega}\right)$$

To find $\tau_{0.5}$ we impose

$$\frac{S_0}{\tau}\left(1 - ln(2)2^{\omega}\right) = 0$$

Thus,

$$\omega' = log_2\left(\frac{1}{ln(2)}\right) = 0.529$$

Now to calculate maximum percentage difference

$$\frac{S_l(\tau_{0.5}) - S_e(\tau_{0.5})}{S_l(\tau_{0.5})} = \frac{S_0\left(1 + \omega' - e^{\omega' ln(2)}\right)}{S_0(1 + \omega')} = \frac{1 + 0.529 - 2^{0.529}}{1 + 0.529} = 0.0563$$

Thus the exponential and linear model are differentiated by (at max) 5.63% volume differences.

References

1. A. W. Bowman, A. Azzalini, *Applied smoothing techniques for data analysis : the kernel approach with S-Plus illustrations* (Clarendon Press, Oxford, 1997), pp. xi, 193 p.
2. H. Frank, S. C. Althoen, *Statistics : concepts and applications* (Cambridge University Press, Cambridge [England] ; New York, NY, USA, 1994), pp. xxvi, 853 p.