Taylor & Francis
Taylor & Francis Group

Multiscale estimation of cell kinetics

Larry W. Jean[ab], Martin T. Suchorolski[cd], Jihyoun Jeon[e] and E. Georg Luebeck[a]*

[a]*Program in Computational Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA;* [b]*Department of Applied Mathematics, University of Washington, Seattle, WA 98195-2420, USA;* [c]*Program in Human Biology, Fred Hutchinson Cancer Research Center, M2-B500, Seattle, WA 98109, USA;* [d]*Department of Molecular and Cell Biology, University of Washington, Seattle, WA 98195-7275, USA;* [e]*Program in Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA*

We introduce a methodology based on the Luria–Delbrück fluctuation model for estimating the cell kinetics of clonally expanding populations. In particular, this approach allows estimation of the net cell proliferation rate, the extinction coefficient and the initial (viable) population size. We present a systematic approach based on spatial partitioning, which captures the local fluctuations of the number and sizes of individual clones. However, partitioning introduces measurement error by inflating the number of clones, which is dependent on time and the degree of cell migration. We perform various *in silico* experiments to explore the properties of the estimators and we show that there exists a direct relationship between precision and observation time. We also explore the trade-off between the measurement error and the estimation accuracy. By exploring different scales of cellular fluctuations, from the entire population down to those of individual clones, we show that this methodology is useful for inferring important parameters in neoplastic progression.

Keywords: cell kinetics; Luria–Delbrück model; mutations; neoplastic progression

AMS Subject Classification: 60G50; 60J25; 44A10; 92C45; 46N60

1. Introduction

Mutations, clonal expansions and (epi)genetic heterogeneity are the hallmarks of many types of neoplasms [19]. Although Nowell's description of cancer as an evolutionary system [21] is widely accepted, key parameters of this description for measuring cell fitness remain generally undetermined. For example, what is the difference between cell replication and death rates (net cell proliferation rate)? What is the ratio between cell death and replication rates (extinction coefficient)? What is the actual number of (viable) cells that contribute to future neoplastic generations (initial population size)? Knowing these parameters will help measure, manage and further our understanding of neoplastic progression.

A number of experimental techniques have been devised to quantify the frequency of cell proliferation and cell death (apoptosis or necrosis), but have limitations [5,18]. In particular, most labelling methods provide only relative measures (or proportions) of dividing or dying cells, but do not provide absolute rates, i.e. the number of events (cell divisions or cell deaths) per unit time. Moreover, many labels used, such as the

mitotic index [9,11], tritiated thymidine labelling index [27,29] and 5-bromodeoxyuridine (BrdU) [12] have cytotoxic effects [4,24,28] and may lead to biased estimates. Furthermore, perhaps with the exception of continuous labelling methods, these assays only provide snapshots of cellular kinetics at the time the assay is performed.

The mathematical approach described here for quantifying cell kinetics has roots in the classical Luria–Delbrück fluctuation analysis [17], which was initially developed to study the natural selection of mutations. In an attempt to determine whether mutations arise spontaneously or from adaptation [30], Luria and Delbrück studied the origin of phage-resistant mutants when cultures of *Escherichia coli* bacteria are seeded in the presence of bacteriophage T1. Surprisingly, the number of observed T1-resistant mutants varied greatly from culture to culture, which led to the conclusion that mutation to phage resistance is the result of rare spontaneous events. Several extensions of the original Luria–Delbrück model were developed over the years [30], including the Lea–Coulson [13], discretized [3,13], Bartlett [3], Haldane [26,31] extensions and the generalized Luria–Delbrück formulation by Dewanji *et al.* [8].

While the use of Luria–Delbrück-type fluctuation models has been limited to the estimation of mutation rates, the technique remains essentially untested for the estimation of cell kinetics parameters such as the net cell proliferation rate, extinction coefficient and the initial (viable) population size in cell culture experiments. A mathematical framework for the quantitative analysis of the number of clones and their sizes has been developed previously by Dewanji *et al.* [7] and was subsequently used for the analysis of enzyme-altered liver foci generated in animal experiments for the estimation of both mutation rates and cell kinetics [14–16]. However, information on the number and sizes of clones are not always available in cell culture experiments.

The limitations of assays with no or imprecise information on the number and sizes of individual clones signal a need for a generally applicable method which attempts to capture the net cell proliferation rate, extinction coefficient and the initial population size. In this study, we introduce such a methodology based on the generalized Luria–Delbrück model developed by Dewanji *et al.* The strength of our approach is the capability of estimating cell kinetics in the absence of information on the number and sizes of individual clones.

We begin with an overview of the generalized Luria–Delbrück model followed by an introduction to the methodology, which discusses spatial partitioning of observation areas and the effects of measurement error on parameter estimation. We carry out a series of *in silico* experiments to explore the properties of the estimators, and show that they remain stable to a point where the measurement error introduced by the spatial partitions begins to dominate the characteristic size fluctuations of the clones. For each model parameter, this leads to a range of spatial partitions in which the bias in absolute value is minimized. We also examine the relationship between the estimation accuracy, precision and the measurement error due to partitioning of *in silico* cell cultures. This is done by observing changes in the estimator properties both in time and under variable spatial dispersion of the cells in a clone. Finally, the methodology is demonstrated on data obtained from a Barrett's Esophagus EPC-2 neoplastic cell line culture.

2. The generalized Luria–Delbrück model

The different variants of the Luria–Delbrück model are characterized according to the dynamics of the normal and mutant populations [30]. With the exception of the original formulation, the Luria–Delbrück models involve a stochastic component to either

describe the dynamics of the normal cells or mutants or both. Here, we adopt a framework that allows for stochastic cell kinetics in the mutants but not in normal cells, i.e. the size of the normal cell population is assumed to be large compared with the mutant population and is modelled deterministically. Specifically, let $X(t)$ and $Y(t)$ represent the population size of normal cells and mutated cells (mutants) at time t (in days), respectively. It is assumed that each normal cell acquires a mutation with rate $\nu(s)$ independent of other cells, and that once a cell has acquired a mutation at time s, it undergoes a clonal expansion in the form of a birth–death process with birth rate $\alpha(t, s)$ (per mutant cell per day) and death rate $\beta(t, s)$ (per mutant cell per day) to form a clone of size $Y(t, s)$ at time t (Figure 1).

The total number of mutants at time t is given by the filtered Poisson process [22]

$$Y(t) = \begin{cases} 0 & M(t) = 0, \\ \sum_{i=1}^{M(t)} Y(t, s_i) & M(t) \geq 1, \end{cases} \tag{1}$$

where $M(t)$ denotes the number of Poisson mutation events in the normal cell population at time t. According to Dewanji *et al.* [8], the distribution of $Y(t)$ takes the form

$$P_0(t) := \text{Prob}\{Y(t) = 0\} = \exp\left\{-\int_0^t \frac{\nu(s)X(s)}{G(t, s) + g(t, s)}\, ds\right\}, \tag{2}$$

$$P_k(t) := \text{Prob}\{Y(t) = k\} = \sum_{i=0}^{k-1} \left(\frac{k-i}{k}\right) P_i(t)h_{k-i}(t), \quad k = 1, 2, 3, \ldots, \tag{3}$$
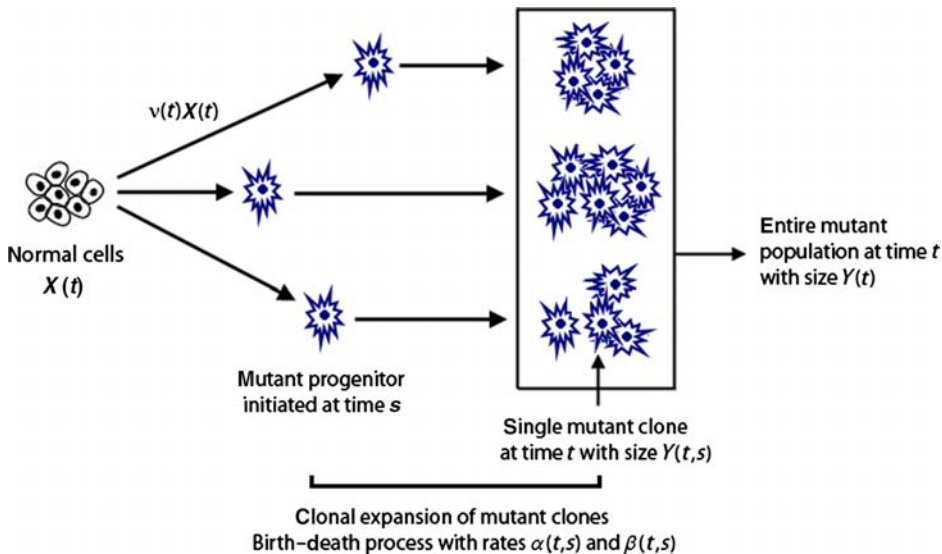

Figure 1. The generalized Luria–Delbrück model. $X(t)$ is the number of normal cells at time t (in days), $\nu(t)$ the rate of Poisson mutation events per normal cell, $Y(t, s)$ the size of a single mutant clone at time t whose progenitor cell is initiated at time s, $\alpha(t, s)$ and $\beta(t, s)$ the birth and death rates per mutant cell of such a clone per day, $Y(t)$ the sum of the sizes of all such clones and its distribution is a Luria–Delbrück distribution.

where

$$h_i(t) := \int_0^t \nu(s)X(s)\frac{g(t,s)}{G(t,s)^2}\left[\frac{G(t,s)}{G(t,s)+g(t,s)}\right]^{i+1} ds, \quad i = 1, 2, 3, \ldots \quad (4)$$

$$g(t,s) := \exp\left\{-\int_s^t [\alpha(u,s) - \beta(u,s)]\, du\right\}, \quad (5)$$

$$G(t,s) := \int_s^t \alpha(u,s)g(u,s)\, du. \quad (6)$$

3. Methodology

3.1 Parameter estimation

Here, we describe a general *in vitro* experiment to determine the rates of both the cell replication and cell death: x fluorescently labelled cells are seeded in each of y wells (petri dishes), cells in a specific medium are grown for t days, locations of labelled cells in a rectangular *field* are recorded and their spatial locations are digitized. In the case that individual clones can be distinguished, the final step is to obtain the total number of clones and their respective sizes; otherwise, count the total number of cells after t days. Some cells will not survive the initial seeding in such experiments due to damage during culturing or failure to adhere to medium. This set of procedures may be complemented with experimental methods of estimating net cell proliferation rate for validation of the estimates.

To keep the presentation simple, we assume constant birth and death rates, i.e. $\alpha(t,s) = \alpha$ and $\beta(t,s) = \beta$. When a normal cell is *initiated* (or 'mutated' in the Luria–Delbrück language), it is signalled to undergo clonal expansion. For the cell culture experiment described above, initiation can be considered as the initial seeding of cells that may undergo clonal expansion. This is expressed mathematically by defining $\nu(t)X(t) = \eta\delta(t)$, where η is the initial population size (initial number of viable cells) and $\delta(t)$ the Dirac delta function. This initial 'pulse' seeding of cells has been previously applied to the modelling of gestational mutations in the context of a multistage theory of carcinogenesis [20]. Within this framework, the cell kinetics parameters, we are interested in estimating, are defined as $\rho := \alpha - \beta$ (net cell proliferation rate), $\gamma := \beta/\alpha$ (extinction coefficient) and η (initial population size). Under the constant parameter assumption, the probability generating function for the total population, $Y(t)$, can be shown to be (see Appendix)

$$\Psi(z,t) := \mathbb{E}[z^{Y(t)}] = \sum_{i=0}^{\infty} \text{Prob}\{Y(t) = i\}z^i = \exp\left[\frac{(\gamma-1)\eta(z-1)}{(\gamma-z)\,e^{-\rho t} + (z-1)}\right]. \quad (7)$$

Due to the computational inefficiency of the recursive form of the distribution $P_k(t)$, $k = 1, 2, \ldots$ (Equation (3)), we resort to the inverse Laplace transform approximation of the distribution introduced by Abate *et al.* [1]. This method incorporates the Poisson summation formula [32], which uses a periodic function represented by its Fourier series to approximate the distribution. The formula is also useful for controlling the discretization error that comes from the numerical approximation. It can be shown that the exact forms of the distribution $P_k(t)$ for $k = 0$ and $k = 1$ are given by (see Appendix)

$$P_0(t) = e^{-\Lambda(t)} \quad (8)$$

and

$$P_1(t) = e^{-\Lambda(t)} \cdot \frac{\eta(1-\gamma)^2 e^{-\rho t}}{(1 - \gamma e^{-\rho t})^2}, \tag{9}$$

respectively, where

$$\Lambda(t) = \frac{(1-\gamma)\eta}{1 - \gamma e^{-\rho t}} \tag{10}$$

is the expected number of non-extinct clones at time t. According to Abate *et al.* [1], Equation (3) for $k = 2, 3, \ldots$ can be approximated by

$$P_k(t) \approx \frac{1}{2kr^k} \left\{ \Psi(r,t) + (-1)^k \Psi(-r,t) + 2\sum_{j=1}^{k-1} (-1)^j \operatorname{Re}\left[\Psi(re^{i\pi j/k}, t)\right] \right\}, \tag{11}$$

where an error bound of 10^{-h} may be achieved by choosing r such that $r \approx 10^{-h/2k}$. Furthermore, in the case where the total number of clones and their sizes are available, we use their respective distributions for the likelihood estimation of the model parameters ρ, γ and η. They are, respectively, given by

$$P_m^{(N)}(t) := \operatorname{Prob}\{M(t) = m\} = \frac{e^{-\Lambda(t)} \Lambda(t)^m}{m!}, \quad m = 0, 1, 2, \ldots \tag{12}$$

and

$$P_w^{(S)}(t,s) := \operatorname{Prob}\{Y(t,s) = w\} = \frac{(1-\gamma)e^{-\rho(t-s)}}{1 - \gamma e^{-\rho(t-s)}} \left[\frac{1 - e^{-\rho(t-s)}}{1 - \gamma e^{-\rho(t-s)}}\right]^{w-1}, \quad w = 1, 2, \ldots. \tag{13}$$

The superscripts N and S represent the *number* and *sizes* of individual clones, respectively. The derivation can be found in the Appendix. We point out that $P_w^{(S)}(t,s)$ is the size distribution for *non-extinct* clones.

In the absence of information on the number ($M(t)$) and sizes ($Y(t,s)$) of individual clones, we would like to explore the size fluctuations on ever shorter length scales down to those of individual clones. Furthermore, the need for precision in the parameter estimation requires a sufficient number of observations. To accomplish this, we exploit the information hidden in local fluctuations of clone sizes and introduce the idea of *spatial partitioning*. Figure 2 shows a sample field spatially partitioned into N different ways, each time creating $n \times n$ *subfields* of identical dimension for $n = 1, 2, \ldots, N$.

In the case where the number and sizes of individual clones are not available, an *observation* is referred to as the total number of cells within a particular subfield. Hence, a field with $n \times n$ subfields will have $n \times n$ independent observations. In order to accurately obtain the cell count within any given subfield, it is necessary to identify all the cells within that subfield, which requires knowledge of their spatial coordinates. To accomplish this task, we digitize the cells and their locations using the ImageJ© software [2,23].

Figure 3 shows four identical *in silico* fields with spatial partitions of $n \times n = 4 \times 4$, 10×10, 20×20 and 30×30, respectively. In this computer-simulated experiment, the first progenitor cell of each clone is randomly seeded according to a spatial Poisson process, then the distance between any daughter cell and the progenitor cell from which it came is normally distributed with a variance proportional to clone size. Thus, clones with
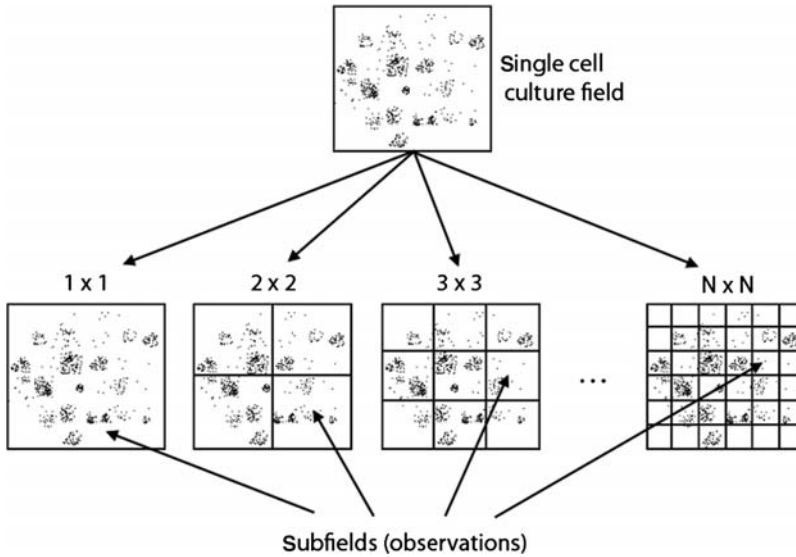
Figure 2. Schematic representation of a single field spatially partitioned into N different ways where for each $n = 1, 2, 3, \ldots, N$, $n \times n$ subfields with an identical dimension are created, each containing a fractional information of the entire field.
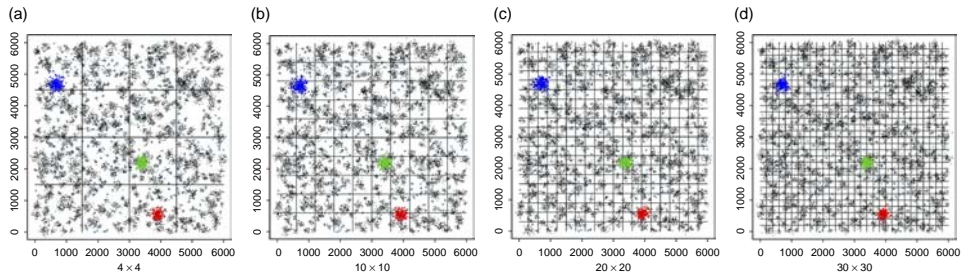

Figure 3. Single field of an *in silico* cell culture with spatial partitions (a) 4×4, (b) 10×10, (c) 20×20 and (d) 30×30. Each cell in an initial population size of $\eta = 1000$ undergoes a birth–death process with parameters $\alpha = 0.75$ and $\beta = 0.5$, respectively (hence $\rho = 0.25$ and $\gamma = 0.67$) and the field is obtained after $t = 1$ day. The blue, red and green regions represent the three largest clones in decreasing sizes. Increasing the number of spatial partitions increases the occurrence of partitioned clones.

a larger number of cells occupy a greater region. This is demonstrated in Figure 3, where blue, red and green represent the three largest clones in decreasing cell counts. Although increasing the number of spatial partitions by fine graining increases the number of observations, each subfield only contains a fraction of cell counts relative to the entire population. In addition, there is an increasing tendency for a clone to spread across multiple subfields for the finer partitions. We refer to these clones as *partitioned clones*. Thus, the question arises: how does excessive partitioning beyond the scale of local fluctuations of clone sizes affect parameter estimation? We address this question in more detail by simulations.
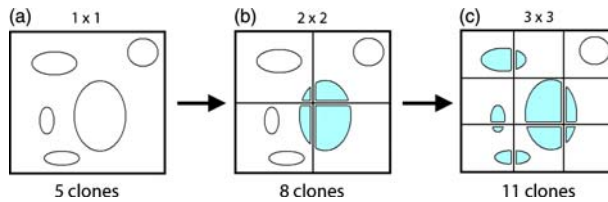
Figure 4. A closer look at partitioned clones and measurement error. A schematic field with (a) five clones is partitioned into (b) 2×2 and (c) 3×3 subfields. Partitioned clones (in blue) are created as clones spatially spread across multiple subfields, which artificially inflate the number of clones. The Luria–Delbrück theory regards the original 5 clones as 8 and 11 distinct clones under the 2×2 and 3×3 partitions cases, respectively, leading to biases in the parameters estimation.

3.2 Trade-offs associated with partitioning a single field

In the absence of experimental replicates and without spatial partitioning (i.e. using a single 1×1 partition), no estimation is possible since a single data point is insufficient for estimating three parameters. Spatial partitioning, on the other hand, introduces subfields (or multiple observations) and allows the estimation of all three model parameters in a single experiment. However, the process of spatial partitioning increases the frequency of partitioned clones (see Figure 4), thereby artificially inflating the number of clones and skewing their sizes towards smaller clones. We refer to this phenomenon simply as *measurement error*.

The trade-off associated with spatial partitioning suggests that for each parameter, there exists a range of partitions with acceptable estimation accuracy and precision. We refer to this set of partitions for each parameter as the optimal partition range. We will demonstrate the existence of such ranges via simulations and show how they assist in identifying the set of optimal estimates in an *in vitro* cell culture.

4. *In silico* experiments

Before applying the methodology to actual cell cultures, we explore the properties of the maximum likelihood estimators for ρ (net cell proliferation rate), γ (extinction coefficient) and η (initial population size), denoted as $\hat{\rho}_n$, $\hat{\gamma}_n$ and $\hat{\eta}_n$, respectively, where n represents the $n \times n$ partition case. In particular, we examine how the estimators behave in the following distinct *in silico* cell culture scenarios:

  (i) resolvable clones with no measurement error,
 (ii) non-resolvable clones with no measurement error,
(iii) non-resolvable clones with measurement error.

In this context, a cell culture with resolvable clones is one in which the number and sizes of individual clones are known, while *measurement error* refers to the artificial inflation of the number of clones due to partitioned clones (Figure 4). Hence, scenarios (i) and (ii) are less relevant to the real world, but are nevertheless important for understanding the properties of the estimators. For each scenario, we obtain population dynamics by randomly seeding cells and simulating clonal expansions based on arbitrarily chosen but fixed parameters. Doing so allows us to acquire the desired cell culture information (number of clones, their respective sizes or total cell counts) necessary to perform the maximum likelihood estimation. Finally, we study the effects of measurement error on parameters estimation.

4.1 Case (i): resolvable clones with no measurement error

In the first of three *in silico* cell cultures designed to explore the properties of the estimators, we restrict all daughter cells to be situated at the same location as the first progenitor cell from which they descended. Doing so uniquely assigns a clone to a single subfield, hence no measurement error is introduced. In addition, it is assumed that the number of clones and their sizes are known (*resolvable*) without measurement error, hence we can use the distributions given by Equations (12) and (13) rather than the Luria–Delbrück distribution for the estimation of the cell kinetics parameters from such observations. Consider a field with $n \times n$ subfields, let $\tilde{N}_i$ and $\tilde{S}_i = [\tilde{S}_{1i}, \tilde{S}_{2i}, \ldots, \tilde{S}_{\tilde{N}_i i}]$ be the observed number and sizes of clones, respectively, of the ith subfield for $i = 1, 2, \ldots, n \times n$ at $t = \tilde{T}$, then the likelihood is

$$\mathcal{L}(\rho, \gamma, \eta | \{\tilde{N}_i, \tilde{S}_i, i = 1, 2, \ldots, n \times n\}, \tilde{T}) = \prod_{i=1}^{n \times n} \left\{ P_{k=\tilde{N}_i}^{(N)}(t = \tilde{T}) \cdot \prod_{j=1}^{\tilde{N}_i} P_{k=\tilde{S}_{ji}}^{(S)}(t = \tilde{T}, s = 0) \right\}, \quad (14)$$

where $P_k^{(N)}(t)$ and $P_k^{(S)}(t)$ are given in Equations (12) and (13), respectively. For the case that the number and sizes of individual clones are known, it is immediately clear that different spatial partitions yield the same likelihood, and the estimators are independent of partition sizes. Indeed, simulations for different observation times reveal that the variances of the estimates decrease with increasing observation time (Figure 5). This observed effect may simply reflect the gain in efficiency of the estimators due to an increase in the number of stochastic events with time.

4.2 Case (ii): non-resolvable clones with no measurement error

A layer of complexity is added to the simulation so that for each cell, only the spatial location is given, with no information on which clone it belongs to or how many clones are observed. In this case, the Luria–Delbrück distribution plays a central role in the parameters estimation and we show that it introduces a bias on the estimates. As in case (i), measurement error as a result of partitioned clones is still assumed to be absent. For an $n \times n$ spatial partition, if $\tilde{Y}_i$ cells are observed at time $t = \tilde{T}$ in the ith subfield for
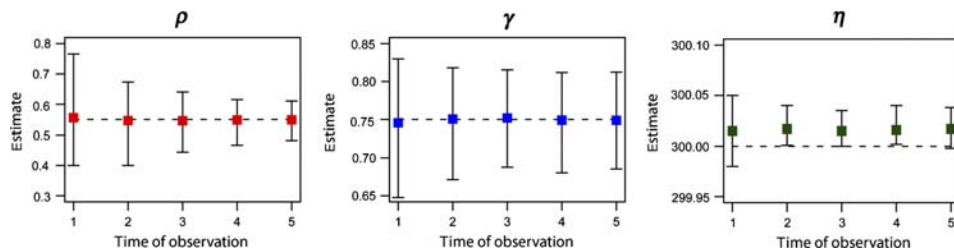

Figure 5. Analysis of resolvable clones with no measurement error. For each time point $t = 1-5$ days, 1000 independent fields of a cell culture with parameters $\rho = 0.55$ (net cell proliferation rate), $\gamma = 0.75$ (extinction coefficient) and $\eta = 300$ (initial population size) (dotted lines) are simulated and analyzed. For each parameter, the mean of the 1000 estimates and its 95% CI are plotted. Since different spatial partitions produce the identical likelihood function, only the results for the 10×10 partition are shown.

$i = 1, 2, \ldots, n \times n$, the likelihood function can be approximated by

$$\mathcal{L}(\rho, \gamma, \eta | \{\tilde{Y}_i, i = 1, 2, \ldots, n \times n\}, \tilde{T}) \approx \prod_{i=1}^{n \times n} P_{k=\tilde{Y}_i}(t = \tilde{T}), \tag{15}$$

where $P_k(t)$ for $k = 0$, $k = 1$ and $k \geq 2$ are given by Equations (8), (9) and (11), respectively. $\Psi(z, t)$ is given by Equation (7) and $r_i \approx 10^{-h/2\tilde{Y}_i}$ gives an accuracy of 10^{-h} for the ith subfield, where $i = 1, 2, \ldots, n \times n$. In comparison with case (i), we expect the accuracy of the estimates to diminish for some partitions due to the lack of information on the number and sizes of individual clones. However, accuracy should increase as the partition is fine-grained since ultimately every subfield contains at most a single clone, thus representing case (i).

Figure 6 shows the accuracy of the estimates as a function of spatial partition sizes. While the estimates for ρ and η show high accuracy independent of partition sizes, the estimates for the extinction coefficient γ exhibit a downward bias for partitions $n \leq 5$. An inspection of the distribution of the γ estimates reveals a bias in the mean for smaller sample sizes due to a lack of normality of the estimators for coarse-grained partitions (fewer subfields). For $n > 5$, the convergence can be attributed to the increasing number of informative subfields associated with the higher partitions, each capturing local fluctuations both in the number of clones and their sizes. Note, under this scenario, $\{\hat{\rho}_n | n = 2, 3, \ldots\}$, $\{\hat{\gamma}_n | n = 2, 3, \ldots\}$ and $\{\hat{\eta}_n | n = 2, 3, \ldots\}$ resemble statistically *consistent* sequences of estimators, as $n \rightarrow \infty$.

In addition, the confidence intervals for ρ and γ decrease with spatial partitions but increase for the initial population size η. This can be explained as follows: consider a field with $n \times n$ subfields, let $\hat{\xi}_n$ be the estimated initial number of viable cells for each *subfield*, hence $\hat{\eta}_n = n^2 \hat{\xi}_n$ is the estimated initial number of viable cells for the entire field. Then the variance (hence the confidence interval), $\mathrm{var}[\hat{\eta}_n] = n^4 \, \mathrm{var}[\hat{\xi}_n]$, is amplified by a factor of n^4 for increasing partition size n even though $\mathrm{var}[\hat{\xi}_n]$ decreases with increasing spatial partitions (not shown). Despite the lack of information on the number and sizes of individual clones, the information stored in the local fluctuations can be 'recaptured' with fine-grained partitions in the absence of measurement error, since in the limit the observation of the number of clones is either one or zero in each subfield.

4.3 Case (iii): non-resolvable clones with measurement error

Experimental evidence suggests that cells migrate inside the wells upon seeding. To represent the spatial dispersion of cells in our simulations which, upon partitioning of an observation field, leads to measurement error, we assume that all daughter cells disperse randomly within a radial distance $\omega = \sqrt{A_{\mathrm{clone}}/\pi}$ away from the first progenitor cell. $A_{\mathrm{clone}} = S_{\mathrm{clone}} \pi \sigma^2$ is the spatial area occupied by a clone having S_{clone} cells. Here, 2σ is the average pairwise distance between the centre of masses of any two cells, and we refer to σ as the *dispersion* factor. The spatial location of any daughter cell is relative to that of the progenitor. In particular, the displacement (x, y) of a daughter cell from the progenitor is sampled from $\mathcal{N}(0, \omega^2) \times \mathcal{N}(0, \omega^2)$.

To better understand the effects of measurement error caused by partitioning on parameter estimation, we simulated a cell culture 1000 times for four different values of the cellular dispersion factor σ. Figure 7 shows this *in silico* 3-day culture partitioned into 10×10 subfields for the cellular dispersion factors $\sigma = 5, 10, 15$ and 20. We see that the
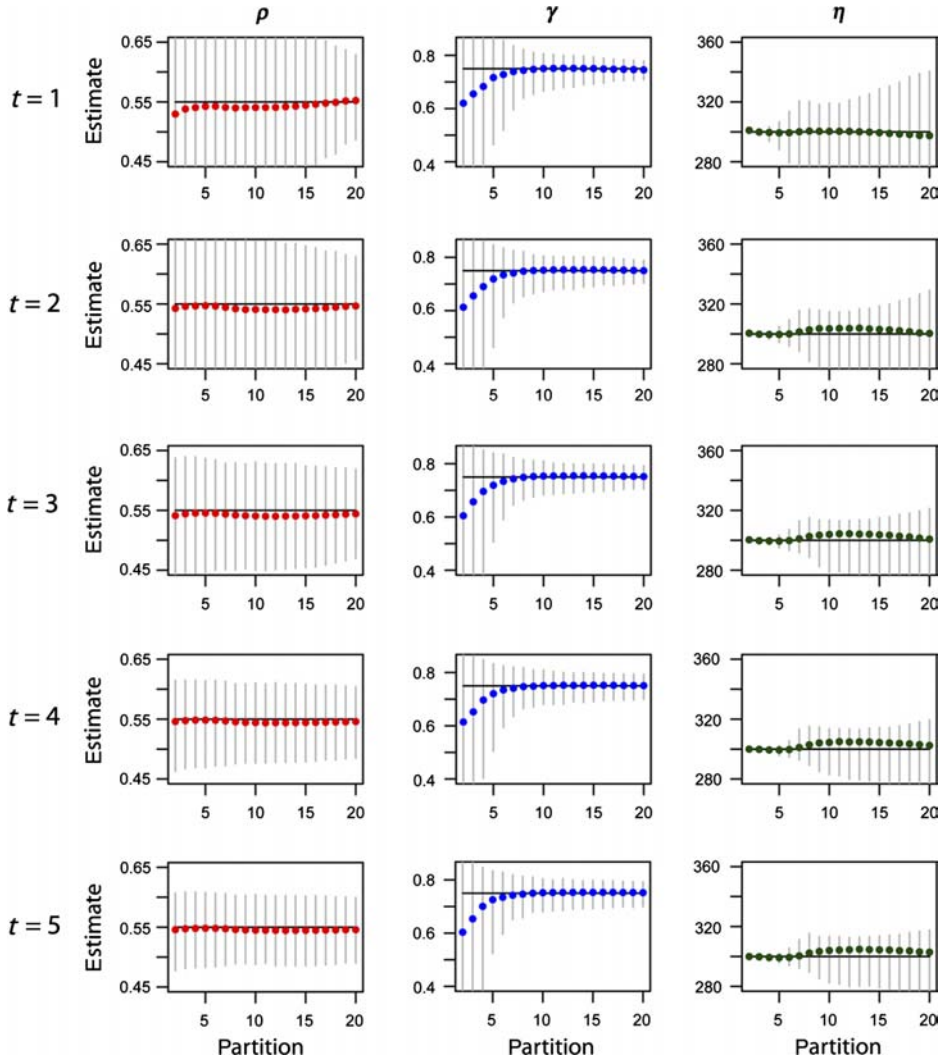
Figure 6. Non-resolvable clones with no measurement error. For each observation time of $t = 1-5$ days (rows), 1000 individual fields from a single cell culture with parameters $\rho = 0.55$ (net cell proliferation rate), $\gamma = 0.75$ (extinction coefficient) and $\eta = 300$ (initial population size) (solid black lines) are simulated, each time obtaining an estimate. For each parameter, the mean of the 1000 estimates and its 95% CI are plotted versus the spatial partitions for $n \times n = 2 \times 2$, $3 \times 3, \ldots, 20 \times 20$.

frequency of occurrence of partitioned clones increases as the dispersion factor, hence the measurement error increases.

We expect the mean of the estimates in this *in silico* cell culture to deviate further from the true parameters as σ increases due to the increasing measurement error on the number of clones and their sizes. This is demonstrated in Figure 8, which shows that for smaller σ, the estimates come closer to the true values in general. As in case (ii), we see a reduction in estimation accuracy for coarser spatial partitions. However, we now find that the biases for all three estimators increase for spatial partitions $n = 7$. The biases occur because fine graining increases the tendency for partitions to intersect the clones, thereby artificially
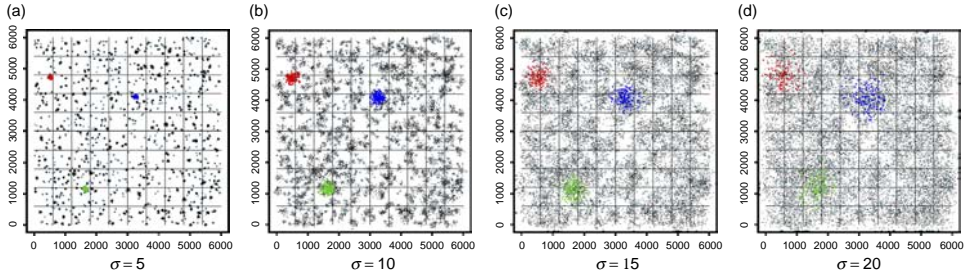
Figure 7. An *in silico* cell culture with parameters $\rho = 0.55$, $\gamma = 0.75$ and $\eta = 300$ observed after time $t = 3$ days post-initial seeding. The effect of measurement error is quantified by the cellular dispersion factor σ. The four plots are for (a) $\sigma = 5$, (b) $\sigma = 10$, (c) $\sigma = 15$ and (d) $\sigma = 20$, and the blue, red and green regions in each case represent the three largest clones in decreasing sizes.
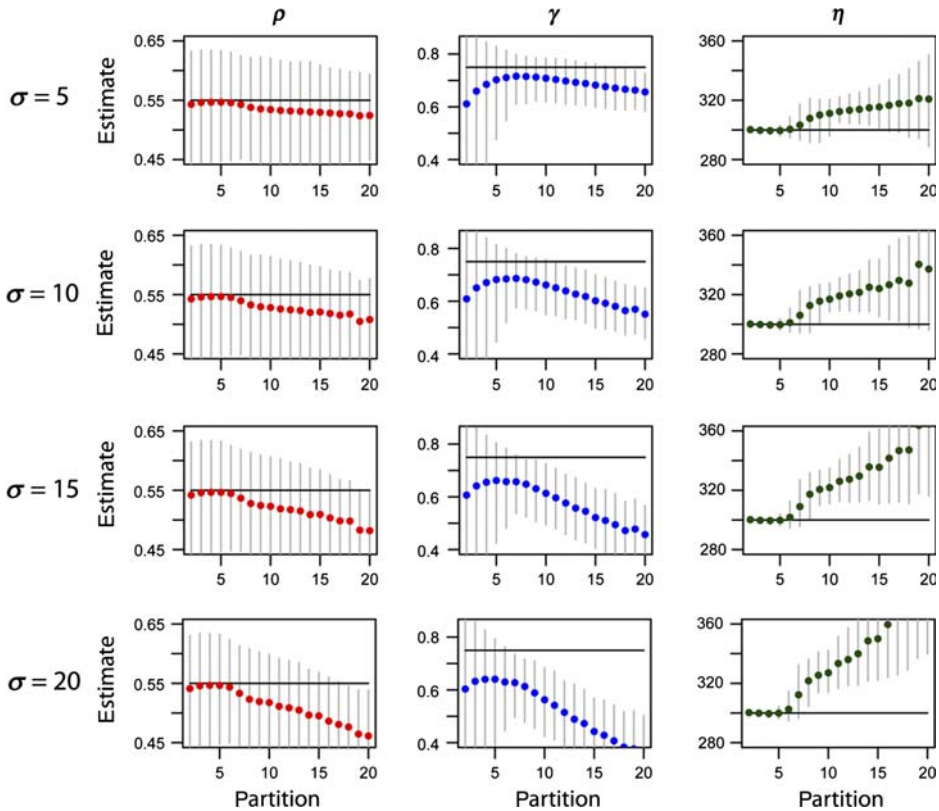

Figure 8. Effects of measurement error. For each dispersion factor of $\sigma = 5$, 10, 15 and 20 (rows), 1000 independent *in silico* experiments of a cell culture with parameters $\rho = 0.55$ (net cell proliferation rate), $\gamma = 0.75$ (extinction coefficient) and $\eta = 300$ (initial population size) (solid black lines) are simulated, obtaining an estimate after $t = 3$ days. For each parameter, the mean of the 1000 estimates and its 95% CI are plotted versus the spatial partitions for $n \times n = 2 \times 2$, 3×3, ..., 20×20.

inflating their number. Hence more clones are being analyzed via the Luria–Delbrück assumption than what are actually observed (measurement error).

For the extinction coefficient (γ), the measurement error increases the bias of the estimators as the dispersion factor (σ) increases. A similar behaviour is observed (not shown) as the time of observation is increased for fixed σ. Furthermore, the confidence interval does not seem to be affected by the change in dispersion factors, suggesting that the variance of the estimators is independent of the measurement error. The bias due to measurement error seems to have a greater effect on γ and η than it does on ρ, indicating that the estimators for the net cell proliferation rate may be more robust under measurement error than those of the other two parameters.

For *in silico* cell cultures of this type, the optimal estimates for each parameter exist and may come from different partitions. We demonstrate next how these optimal estimates may be identified in an *in vitro* cell culture using data from a cell culture experiment.

5. Experimental results

We demonstrate the methodology using data collected from the Barrett's Esophagus EPC-2 cell line. A field with 5772 cells was observed 21.167 h following an initial seeding of $\approx 12{,}000$ cells. The field was spatially partitioned into sizes ranging from $2 \times 2, 3 \times 3, \dots,$ 20×20, where the coordinates of the cell positions were obtained via the ImageJ© software [2,23]. For each $n \times n$ partition, we obtained n^2 total subfields (observations), which were used in the likelihood analysis to produce one estimate. We used the Luria–Delbrück form of the likelihood (Equation (15)) for analysis, since the number of clones and their sizes in the field were unknown.

As Figure 9 shows, the estimates for ρ and η display a similar behaviour as the mean of the estimates for the *in silico* cell culture scenario (iii) in Section 4.3 and appear stationary over a narrow range of partitions. In contrast, the estimates for γ do not exhibit a clear plateau or shoulder as seen in Figure 8, but with increasing cellular dispersion factor (σ), the estimates for γ are increasingly biased downward as the partition size increases, overwhelming the small sample bias on the left. Thus, we conclude that the true value of γ is larger than the observed maximum value of 0.96. On the other hand, the estimates using partitions less than 5 appear to suffer from small sample biases and tend to be unreliable. The lack of a stationarity in the estimates for γ further indicates that the observation time was not long enough and that cell death immediately after seeding (<24 h) is more extensive before reaching a constant rate in the surviving cell population. It is also possible that due to dense seeding of the progenitor cells, the resulting clones can no longer
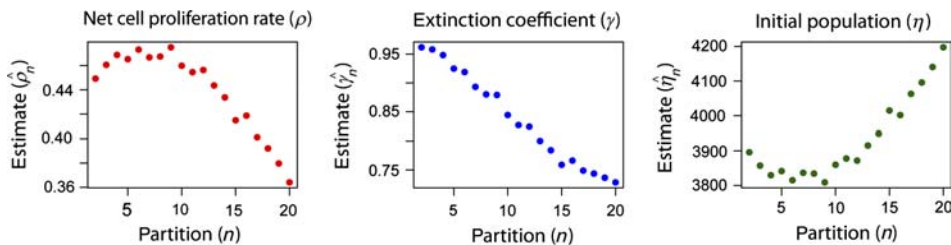

Figure 9. A single cell culture field from Barrett's Esophagus EPC-2 cell line data. The cell culture images were reproduced *in silico* after cellular coordinates were obtained via ImageJ©, and partitioned from 2×2, 3×3, up to 20×20. The number and sizes of individual clones were unknown, hence our methodology was used to estimate the cell kinetics parameters ρ, γ and η. For each parameter, one estimate was obtained for each partition.

be considered as growing independently, which violates a basic assumption of our model. However, the qualitative agreement between experimental and *in silico* results suggests that our method is applicable, especially for cell counts from single wells at a single time point. Multiple experiments and multiple time points will likely improve the estimates.

6. Summary

We introduced a multiscale image processing technique in conjunction with a Luria–Delbrück distribution for estimating cell kinetics parameters in the absence of information on the number and sizes of individual clones. We find that it is possible to estimate cell kinetics parameters via optimal spatial partitioning of the field of observation. More specifically, the estimated net cell proliferation rate, extinction coefficient and initial population size are given by optimal spatial partitions, which may differ from parameter to parameter. We show that for *in silico* cell cultures, there exists a trade-off between estimation accuracy versus bias introduced by measurement errors due to the inability to assign clones uniquely to a particular subfield of a partition. For later observation times and for increasing measurement error due to cellular migration, the bias increases due to a higher occurrence of partitioned clones, thereby decreasing the accuracy of the estimates, i.e. increasing the bias of the estimators.

Several aspects of the methodology presented can be improved. Properties of the estimators should also be explored when observations from multiple time points are available. This may lead to more accurate estimates since each time point provides a snapshot of a growing random population undergoing size fluctuations that are determined by the cell kinetics. Furthermore, spatial clustering algorithms might be helpful in improving the estimates and in identifying spatial inhomogeneities in the initial distribution of seeded cells that violate the underlying homogeneous Poisson assumption. It would also be useful to validate estimates by independent measurements of cell kinetics using experimental techniques, such as BrdU for the measurement of the cell replication rates. The methodology described here for analysis of *in vitro* data may also be applied to cross-sectional images of a neoplastic population *in vivo*, although stereological methods [6,10,25] may need to be invoked to transform the cross-sectional data into three dimensions.

The net cell proliferation rate, extinction coefficient and initial (viable) population size are key parameters in understanding the evolutionary biology in neoplastic progression. We have shown that by exploring different scales of cellular fluctuations from the entire population down to those of individual clones, this approach can be useful for estimating these parameters in a clonally expanding population. Ultimately, measuring proliferation and death rates in directly competing cell lines simultaneously could further our understanding of their relative fitness and could potentially provide quantitative information that better characterizes the clonal evolution of neoplastic populations.

References

[1] J. Abate and W. Whitt, *The Fourier-series method for inverting transforms of probability distributions*, Queueing Syst. 10(1) (1992), pp. 5–88.
[2] M.D. Abramoff, P.J. Magelhaes, and S.J. Ram, *Image processing with ImageJ*, Biophotonics Int. 11(7) (2004), pp. 36–42.

[3] P. Armitage, *The statistical theory of bacterial populations subject to mutations*, J. Royal Statist. Soc. B 14 (1952), p. 41.

[4] E. Asher, C.M. Payne, and C. Bernstein, *Evaluation of cell death in EBV-transformed lymphocytes using agarose gel electrophoresis, light microscopy and electron microscopy. II. Induction of non-classic apoptosis ('para-apoptosis') by tritiated thymidine*, Leuk. Lymphoma Res. 19 (1995), pp. 107–119.

[5] M.J. Beresford, G.D. Wilson, and A. Makris, *Measuring proliferation in breast cancer: Practicalities and applications*, Breast Cancer Res. 8 (2006), p. 216.

[6] H.A. Campbell, H.C. Pitot, V.R. Potter, and B.A. Laishes, *Application of quantitative stereology to the evaluation of enzyme altered foci in rat liver*, Cancer Res. 42 (1982), pp. 465–472.

[7] A. Dewanji, D.J. Venzon, and S.H. Moolgavkar, *A stochastic two-stage model for cancer risk assessment. II. The number and size of premalignant clones*, Risk Anal. 9(2) (1989), pp. 179–187.

[8] A. Dewanji, E.G. Luebeck, and S.H. Moolgavkar, *A generalized Luria–Delbrück model*, Math. Biosci. 197 (2005), pp. 140–152.

[9] C.W. Elston and I.O. Ellis, *Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up*, Histopathology 19 (1991), pp. 403–410.

[10] R.L. Fullman, *Measurement of particle sizes in opaque bodies*, J. Metals 5 (1953), pp. 447–452.

[11] P. Genest and A. Villeneuve, *Lithium and mitotic index*, Lancet 2(7737) (1971), p. 1325.

[12] H.G. Gratzner, *Monoclonal antibody to 5-bromo- and 5-iododeoxyuridine: A new reagent for detection of DNA replication*, Science 218 (1982), pp. 474–475.

[13] D.E. Lea and C.A. Coulson, *The distribution of the numbers of mutants in bacterial populations*, J. Genet. 49 (1949), pp. 264–285.

[14] E.G. Luebeck and S.H. Moolgavkar, *Stochastic analysis of intermediate lesions in carcinogenesis experiments*, Risk Anal. 11(1) (1991), pp. 149–157.

[15] E.G. Luebeck, S.H. Moolgavkar, A. Buchmann, and M. Schwarz, *Effects of polychlorinated biphenyls in rat liver: Quantitative analysis of enzyme-altered foci*, Toxicol. Appl. Pharmacol. 111(3) (1991), pp. 469–484.

[16] E.G. Luebeck, A. Buchmann, S. Stinchcombe, S.H. Moolgavkar, and M. Schwarz, *Effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on initiation and promotion of GST-P-positive foci in rat liver: A quantitative analysis of experimental data using a stochastic model*, Toxicol. Appl. Pharmacol. 167(1) (2000), pp. 63–73.

[17] S.E. Luria and M. Delbrück, *Mutations of bacteria from virus sensitivity to virus resistance*, Genetics 28 (1943), pp. 491–511.

[18] D.C. Macallan, C.A. Fullerton, R.A. Neese, K. Haddock, S.S. Part, and M.K. Hellerstein, *Measurement of cell proliferation by labeling of DNA with stable isotope-labeled glucose: Studies* in vitro*, in animals, and in humans*, Proc. Natl Acad. Sci. USA 95 (1998), pp. 708–713.

[19] L.M.F. Merlo, J.W. Pepper, B.J. Reid, and C.C. Maley, *Cancer as an evolutionary and ecological process*, Nature 6 (2006), pp. 924–935.

[20] R. Meza, E.G. Luebeck, and S.H. Moolgavkar, *Gestational mutations and carcinogenesis*, Math. Biosci. 197(2) (2005), p. 188.

[21] P.C. Nowell, *The clonal evolution of tumor cell populations*, Science 194 (1976), pp. 23–28.

[22] E. Parzen, *Stochastic Processes*, Holden-day Series in Probability and Statistics, Holden-Day Inc, San Francisco, 1962.

[23] W.S. Rasband, ImageJ, US National Institutes of Health, Bethesda, MD, USA. Available at http://rsb.info.nih.gov/ij/, 1997–2008.

[24] B. Rocha, G. Penit, C. Baron, F. Vasseur, N. Dautigney, and A.A. Freitas, *Accumulation of bromodeoxyuridine-labeled cells in central and peripheral lymphoid organs: Minimal estimates of production and turnover rates of mature lymphocytes*, Eur. J. Immunol. 20 (1990), pp. 1697–1708.

[25] S.A. Saltykov, *The determination of the size distribution of particles in an opaque material from measurement of the size distribution of their sections*, in *Stereology* (Proc. Second Int. Cong. for Stereology), Springer-Verlag, New York, 1967, p. 163.

[26] S. Sarkar, *Haldane's solution of the Luria–Delbrück distribution*, J. Genet. 127 (1991), p. 257.

[27] R.J. Sklarew, J. Hoffman, and J. Post, *A rapid* in vitro *method for measuring cell proliferation in human breast cancer*, Cancer 40 (1977), pp. 2299–2302.

[28] P. Taupin, *BrdU immunohistochemistry for studying adult neurogenesis: Paradigms, pitfalls, limitations, and validation*, Brain Res. Rev. 53(1) (2007), pp. 198–214.

[29] F.M. Waldman, K. Chew, B.M. Ljung, W. Goodson, J. Hom, L.A. Duarte, H.S. Smith, and B. Mayall, *A comparison between bromodeoxyuridine and 3H thymidine labeling in human breast tumors*, Mod. Pathol. 4 (1991), pp. 718–722.

[30] Q. Zheng, *Progress of half century in the study of the Luria–Delbrück distribution*, Math. Biosci. 162 (1999), pp. 1–32.

[31] Q. Zheng, *On Haldane's formulation of Luria and Delbrück's mutation model*, Math. Biosci. 209 (2007), pp. 500–513.

[32] A. Zygmund, *Trigonometric Series*, 2nd ed., Cambridge University Press, Cambridge, 1988.

Appendix

A.1 Derivation of Equation (7)

The probability generating function for a single clone with cell count $Y(t, s)$ initiated at time s is given by Luebeck *et al.* [14] to be

$$\phi(z; t, s) = \mathbb{E}[z^{Y(t,s)}] = 1 - \frac{z - 1}{(z - 1)G(t, s) - g(t, s)}, \tag{16}$$

where under constant birth and death rates, i.e. $\alpha(t, s) = \alpha$ and $\beta(t, s) = \beta$,

$$g(t, s) = \exp\left\{-\int_s^t (\alpha(u, s) - \beta(u, s))\, du\right\} = e^{-(\alpha - \beta)(t - s)}, \tag{17}$$

$$G(t, s) = \int_s^t \alpha(u, s) g(u, s)\, du = \int_s^t \alpha e^{-(\alpha - \beta)(u - s)}\, du = \frac{\alpha}{\alpha - \beta}[1 - e^{-(\alpha - \beta)(t - s)}]. \tag{18}$$

By assuming $\rho = \alpha - \beta$, $\gamma = \beta / \alpha$ and $\nu(t)X(t) = \eta\delta(t)$, the probability generating function (PGF) for $Y(t)$ can be shown to be

$$\begin{aligned}
\Psi(z; t) = \mathbb{E}[z^{Y(t)}] &= \exp\left\{\int_0^t \nu(s)X(s)[\phi(z; t, s) - 1]\, ds\right\} \\
&= \exp\left\{\int_0^t \eta\delta(s)[\phi(z; t, s) - 1]\, ds\right\} \\
&= \exp\{\eta[\phi(z; t, 0) - 1]\} \\
&= \exp\left[-\eta\frac{z - 1}{(z - 1)(1/(1 - \gamma))(1 - e^{-\rho t}) - e^{-\rho t}}\right] \\
&= \exp\left[\frac{(\gamma - 1)\eta(z - 1)}{(\gamma - z)e^{-\rho t} + (z - 1)}\right].
\end{aligned} \tag{19}$$

A.2 Derivation of Equations (8)–(10)

According to Luebeck *et al.* [14], $M(t)$ is a Poisson process with rate

$$\Lambda(t) = \int_0^t \frac{\nu(s)X(s)}{G(t, s) + g(t, s)}\, ds = \frac{\eta}{(1/(1 - \gamma))(1 - e^{-\rho t}) + e^{-\rho t}} = \frac{(1 - \gamma)\eta}{1 - \gamma e^{-\rho t}}. \tag{20}$$

Equation (8) is therefore derived as

$$P_0(t) = \exp\left\{\int_0^t \nu(s)X(s)[\phi(0; t, s) - 1]\, ds\right\} = \exp\left\{-\int_0^t \frac{\nu(s)X(s)}{G(t, s) + g(t, s)}\, ds\right\} = e^{-\Lambda(t)} \tag{21}$$

and Equation (9) follows immediately from the PGF $\Psi(z; t)$,

$$
\begin{aligned}
P_1(t) &= \left.\frac{\partial \Psi(z; t)}{\partial z}\right|_{z=0} = \Psi(0; t) \cdot \left.\frac{\eta(1-\gamma)\{(z-1)(1-e^{-\rho t}) - [(\gamma-z)e^{-\rho t} + (z-1)]\}}{[(\gamma-z)e^{-\rho t} + (z-1)]^2}\right|_{z=0} \\
&= e^{-\Lambda(t)} \cdot \frac{\eta(1-\gamma)^2 e^{-\rho t}}{(1-\gamma e^{-\rho t})^2}.
\end{aligned}
\tag{22}
$$

A.3 Derivation of Equation (13)

According to Luebeck *et al.* [14],

$$
\begin{aligned}
P_k^{(S)}(t, s) &= \text{Prob}\{Y(t, s) = k\} = \left.\frac{1}{k!}\frac{d^k}{dz^k}\mathbb{E}[z^{Y(t,s)}]\right|_{z=0} \\
&= \frac{g(t, s)}{G(t, s)}\left[\frac{G(t, s)}{G(t, s) + g(t, s)}\right]^k \\
&= \frac{e^{-\rho(t-s)}}{(1/(1-\gamma))[1 - e^{-\rho(t-s)}]}\left\{\frac{(1/(1-\gamma))[1 - e^{-\rho(t-s)}]}{(1/(1-\gamma))[1 - e^{-\rho(t-s)}] + e^{-\rho(t-s)}}\right\}^k \\
&= \frac{e^{-\rho(t-s)}}{(1/(1-\gamma))[1 - e^{-\rho(t-s)}] + e^{-\rho(t-s)}}\left[\frac{1 - e^{-\rho(t-s)}}{1 - e^{-\rho(t-s)} + (1-\gamma)e^{-\rho(t-s)}}\right]^{k-1} \\
&= \frac{(1-\gamma)e^{-\rho(t-s)}}{1 - \gamma e^{-\rho(t-s)}}\left[\frac{1 - e^{-\rho(t-s)}}{1 - \gamma e^{-\rho(t-s)}}\right]^{k-1}.
\end{aligned}
\tag{23}
$$