



Published in final edited form as:

*Math Biosci.* 2005 September ; 197(1): 67–87.

## A comprehensive continuous-time model for the appearance of CGH signal due to chromosomal missegregations during mitosis

Richard Desper<sup>1,3</sup>, Michael J. Difilippantonio<sup>2</sup>, Thomas Ried<sup>2</sup>, and Alejandro A. Schäffer<sup>1,\*</sup>

<sup>1</sup> Computational Biology Branch, National Center for Biotechnology Information, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA

<sup>2</sup> Section of Cancer Genomics, Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA

<sup>3</sup> Present affiliation: Department of Biology, University College London, UK

### Abstract

Aneuploidy, the gain or loss of large regions of the genome, is a common feature in cancer cells. Irregularities in chromosomal copy number caused by missegregations of chromosomes during mitosis can be visualized by cytogenetic techniques including fluorescence in situ hybridization (FISH), spectral karyotyping (SKY) and comparative genomic hybridization (CGH). In the current work, we consider the propagation of irregular copy numbers throughout a cell population as the individual cells progress through ordinary mitotic cell cycles. We use an algebraic model to track the different copy numbers as states in a stochastic process, based on the model of chromosome instability of Gusev, Kagansky, and Dooley, and consider the average copy number of a particular chromosome within a cell population as a function of the cell division rate. We review a number of mathematical models for determining the length of the cell cycle, including the Smith-Martin transition probability model and the “sloppy size” model of Wheals, Tyson and Diekmann. The program MITOSIM simulates the growth of a population of cells using the aforementioned models of the cell cycle. MITOSIM allows the cell population to grow, with occasional resampling, until the average copy number of a given chromosome in the population reaches a preset threshold signifying a positive copy number alteration in this region. MITOSIM calculates the relationship between the missegregation rate and the growth rate of the cell population. This allows the user to test hypotheses regarding the effect chromosomal aberrations have upon the cell cycle, cell growth rates, and time to population dominance.

### Keywords

mitosis; comparative genomic hybridization; cancer cell lines; chromosome missegregation; aneuploidy in cancer; sloppy size models

### 1 Introduction

The vast majority of solid tumors of epithelial origin (carcinomas) often have genomic imbalances that range from defined chromosomal bands to entire chromosome arms and, in the most extreme cases, entire chromosomes [1]. This aneuploidy is specific for different tumor types, specific for certain tumor stages, and late tumor passages (such as in metastases or in established cell lines) still very faithfully maintain the cancer specific distribution of genomic

\*Address for Correspondence: Alejandro A. Schäffer, NCBI/NLM/NIH, Building 38A, Room 6S608, 8600 Rockville Pike, Bethesda, MD 20894 USA, E-mail:schaffer@helix.nih.gov, FAX: 301-480-2288.

imbalances seen in earlier stages [1]. These facts suggest that continuous selection pressure for the maintenance of these aberrations exists. Cytogenetic analyses have also shown that early chromosomal imbalances occur in clusters in a low percentage of cells, and that, at later stages of tumorigenesis, cells that maintain such aberrations constitute the majority of tumor cells [2]. Because chromosomal aneuploidy can result from errors during the segregation phase of mitosis, we present, in the current work, a comprehensive mathematical theory for the propagation of copy number abnormalities, combining missegregation models with a number of models for modeling intermitosis time. For the missegregation models, we start from a synchronized model of Gusev et al. [3,4] and show it can be modified to be made asynchronous, while enhancing the ability to analyze the model algebraically. We also present simulation software modeling the accumulation of copy number abnormalities to aid in interpretation of laboratory data on population dynamics in cancer cell lines.

Studies in population genetics have well established the idea that biological variants occur in all populations and at different ratios. Under certain conditions, particular alleles or genes can confer a selective advantage. Some examples of this are antibiotic resistance in bacteria [5,6], malaria resistance in people heterozygous for the sickle cell allele of hemoglobin [7–9] and the evolution of drug resistance in tumors [10–12]. Cells with the ability to repair DNA after ionizing radiation have much greater survival rates compared to cells lacking any number of these repair enzymes [13–16]. A faster growth rate or the capacity to overcome death or senescence will, all other things being equal, confer upon a variant cell the ability to eventually overtake and dominate a population. In fact, this is one of the fundamental principles of tumorigenesis.

Another well-established tenet of cancer research is that there is an increase in the number of genomic aberrations during the progression from a benign cell mass to a metastatic tumor [17]. These alterations of the cellular blueprint can take many forms. Small point mutagenic events, as in the case of mismatch repair deficient colorectal tumors [18–20], can accumulate in the genome and affect the function of individual proteins either by rendering them incapable of performing their biological function or by conferring upon them new functions. Amplifications or deletions modify the copy number of a given gene(s) thereby affecting protein expression levels. The affected genes are typically oncogenes or tumor suppressor genes, respectively, which ultimately alter cell division rates and/or the ability to arrest the cell cycle to allow repair of damaged DNA. A study by Visakorpi et al. [21] shows that amplification of the androgen receptor gene is selected for in tumors resistant to androgen deprivation therapy. Also, thousands of genes can be gained or lost simultaneously through the unequal partitioning of chromosomes during mitosis or non-reciprocal chromosome rearrangements. These gross chromosomal aneuploidies affect the expression of most of the genes on the involved chromosome [22–26].

Cytogenetic observations have provided ample evidence that the gain of chromosome 7 in colorectal polyps is often the earliest detectable genetic aberration [27–31]. It has been speculated that this specific chromosomal gain confers a growth advantage through an increase in the copy number of the epidermal growth factor receptor gene located at 7p13 [32–34]. Some of the mathematical explanation and all of the simulations below are based on the example of colorectal polyps and chromosome 7. Likewise, the loss of 17p in these same tumors may enable the cells to bypass the cell cycle arrest checkpoint as a result of the decreased copy number of the TP53 gene. Acquisition of extra copies of chromosome 3q in cervical carcinomas may allow the cells to escape senescence via an increase in the number of hTERT genes, since the encoded RNA is part of the machinery involved in the maintenance of telomere length [2]. These strictly conserved recurrent aberrations and genomic imbalances within or across tumor types are presumed by many to be instrumental for tumorigenesis, and are therefore

assumed to confer a selective advantage. However, they shed little light on how large the selective advantage is and how the advantage is achieved from one generation to the next.

One hypothesis for the advantage conferred upon cancer cells by means of chromosomal aneuploidy is that they can increase the rate of cell division. Thus, once a copy number change occurs in a few cells, these cells and their descendants could dominate the population. Suppose our protocol is to count a “gain” in a chromosome ( $i$ ) once its observed average copy number is at least  $2 + \delta$ , for some value of  $\delta > 0$ . Given a hypothesized rate  $p$  for chromosome missegregation and a function  $T_d$  that estimates the population doubling time as a function of the copy number  $i$ , one quantity of interest is the amount of time required to change the average copy number in our population of cells from 2 to  $2 + \delta$ . Several researchers have proposed that, in cancers of epithelial origin and derived cancer cell lines, chromosome missegregation and the resulting aneuploidy are dominant genetic aberrations [17,35–37].

Two laboratory techniques now in wide use to measure aberrations in cancer cells are spectral karyotyping (SKY)/multiplex FISH (M-FISH) [38,39] and comparative genomic hybridization (CGH) [40]. SKY shows chromosomal rearrangements qualitatively in color in metaphases from individual cells in a sample preparation. CGH, on the other hand, measures deviations from the normal chromosome copy number (2 for autosomes) quantitatively averaged over the entire sample. Since CGH/SKY/M-FISH studies can allow us to measure the average copy number, an estimate of the time required for a shift in the copy number of a given chromosomal region to appear in the population would enable us to decide if the assumed values of  $p$  and  $T_d(i)$  are biologically plausible.

To study the influence of chromosomal imbalances on cell division rates, we analyzed a mathematical model for cell growth, using parameters such as chromosome missegregation rates and population doubling times. Our goal for this model is to allow us to theoretically determine whether it would be feasible, and how long it would take, to shift the chromosome profile of a continuously growing cell population. This model could ultimately be used to refine the various parameters after in situ experimentation.

In Section 2, we present our exponential model for cell division with missegregation errors. First, we consider a synchronized model for mitosis, with fixed intermitotic times, based on the work of Gusev et al. We then consider a continuous time model using an exponential distribution for intermitotic times. We show that the exponential model is amenable to methods from Markov analysis and linear algebra. Using a computer algebra system, and given a value of  $\delta > 0$ , it is possible to obtain an analytic estimate for the time  $T_\delta$  needed to change the average copy number from 2 to  $2 + \delta$ . In Section 3.1, we summarize a well-established mitosis model of Smith and Martin [41], the “transition probability model,” that includes a constant lag time before mitosis may occur in any daughter cell. The Smith-Martin model is, in principle, more complex than our exponential model because of an extra parameter (the lag). However, we point out that this parameter can be eliminated in the asymptotic, steady state by a change of variables. Thus, the derivations in Section 2 remain applicable with the addition of a constant lag. In Section 3.2, we summarize a different model of mitosis proposed by Wheals [42] and Tyson and Diekmann [43] called the “sloppy size” model. In Section 4, we present a computer program MITOSIM that simulates mitosis using either the exponential, lagged exponential or sloppy size models, and tracks the propagation of copy number aberrations as they randomly appear in the population. We show results of representative MITOSIM runs and demonstrate that they are in good agreement with analytical results for the exponential and lagged exponential models.

## 2 Mathematical models for mitosis

### 2.1 A discrete time model for tumor growth

To clarify the issues involved in modeling mitosis, we first consider an idealized scenario where all the cells in the population have been synchronized to undergo mitosis simultaneously, and we make a further assumption that intermitotic times are constant. This simple model is very similar to the model considered in [3] to study the propagation of chromosomal segregation errors over the long term. The general Gusev model [3,4] allows for modeling segregation errors over all chromosomal pairs simultaneously, but we prefer to focus on only one chromosomal pair at a time, as the explosion in the number of possible combinations of copy numbers makes computations unwieldy and seriously weakens the statistical power of any measurements.

Suppose we fix our attention on a specific chromosome, for example, chromosome 7 in a colorectal cancer cell line, whose copy number is allowed to range from 1 to  $k$ . Suppose also there is a constant missegregation probability  $p$ : If a cell with copy number  $i$  missegregates during mitosis, it produces two daughter cells, one with copy number  $i-1$  and one with copy number  $i+1$ . We enforce boundary conditions on the allowed copy numbers: if  $i=1$ , the cell with the smaller copy number dies, and if  $i=k$ , the cell with the larger copy number dies. If no chromosomal missegregation occurs, each of the two daughter cells has copy number  $i$ .

This process can be described by a transition matrix  $\mathbf{M}(p) = (m_{ij})$ , where  $m_{ij}$  is the expected number of cells in state  $i$  after mitosis, for each cell in state  $j$  before mitosis. Thus

- $m_{ii} = 2(1-p)$  for all  $i$
- $m_{ij} = p$  if  $|i-j|=1$ , with  $1 \leq i, j \leq k$ , and
- $m_{ij} = 0$  for  $i-j > 1$ .

Let  $\mathbf{v}(t)$  be the vector describing the distribution of cells in each state as a function of time,  $t$ , such that  $\mathbf{v}(0)$  is the initial state, and let  $\mathbf{M}^t$  be the  $t$ -fold matrix product of  $\mathbf{M}=\mathbf{M}(p)$  with itself. The expected distribution of cells at time  $t$  is  $\mathbf{v}(t,p)=\mathbf{M}^t\mathbf{v}(0)$ . While the exact number of cells in each state would be of interest, CGH technology only measures the average copy number in the cell population. To calculate this quantity, we use the vectors  $\mathbf{z}_k = (1, 2, \dots, k)$  and  $\mathbf{1}_k = (1, 1, \dots, 1)$  (a vector with  $k$  entries). The average copy number in the distribution  $\mathbf{v}(t)$  can be calculated as

$$r(t, p) = \frac{\mathbf{v}(t, p) \cdot \mathbf{z}_k}{\mathbf{v}(t, p) \cdot \mathbf{1}_k}.$$

To help determine the long-term behavior of  $r(t,p)$ , let  $\mathbf{u}(t, p) = \frac{\mathbf{v}(t, p)}{\mathbf{v}(t, p) \cdot \mathbf{1}_k}$ , the vector in the direction of  $\mathbf{v}(t)$  scaled so the sum of its entries equals 1.

Let us consider the case where  $k=5$ . Let  $r(t,p)$  denote the average copy number of chromosome 7 after  $t$  generations. We use the software package Mathematica [44] (calculations omitted) to calculate  $r(t, p)$  for  $k=5$ :

$$r(t, p) = \frac{3(\lambda_1^t - 3\lambda_3^t + (3 - 2\sqrt{3})\lambda_4^t + (3 + 2\sqrt{3})\lambda_5^t)}{((3 + 2\sqrt{3})\lambda_5^t + (3 - 2\sqrt{3})\lambda_4^t)}, \quad (2.1)$$

where  $\lambda_i$  is the  $i^{\text{th}}$  eigenvalue of the matrix  $\mathbf{M}(p)$ . Given any value for  $p$ , and any threshold  $r_{\min}$ , we can use Equation (2.1) to find the smallest integral value for  $t$  for which  $r(t, p) > r_{\min}$ . For example, if  $p=0.01$  and  $r_{\min} = 2.4$ , we note that  $r(2858, 0.01) = 2.39931$ , while  $r(2859, 0.01) = 2.40008$ .

## 2.2 Comparison with Gusev models

Gusev et al. [3] have presented two models for chromosomal missegregations. In [3], they use the following framework: During the segregation phase of mitosis, any chromosome can witness a segregation error (with probability  $p$ ) that sends both copies of that particular chromosome to one of the two resulting daughter cells. Thus, for example, a diploid cell will witness missegregations events of both copies of a given chromosome with probability  $p^2$ , and half the time both missegregations will result in a tetrasomic daughter cell (as well as a non-viable nullisomic daughter cell). Our model ignores this possibility as highly improbable and is nearly identical to the Gusev model for small values of  $p$ , as the matrix entries in the Gusev model corresponding to higher powers of  $p$  are dominated by the linear terms.

In [4], Gusev et al. consider the longer-term, but not asymptotic behavior of this system. Their simulations are constrained by memory limits, since they keep all cells that are used. They present simulations that have evolved for 200 generations, and claim to have studied simulations with more generations, though still with fewer than 1000 generations. Curiously, they find that the modal copy number in the cell population is 1, i.e. the cell population is dominated by monosomic cells, even when the system is seeded with one diploid cell. This phenomenon is a result of their decision to structure the underlying Markov chain such that monosomic cells are less likely (albeit only slightly less likely) to witness missegregations than cells with higher copy number. Thus, the stationary distribution of this system has highest weight on the monosomic state. In contrast, if the Markov chain were structured with identical missegregation rates for all cells, then the stationary distribution would necessarily be symmetrical about the mean of the copy numbers allowed ( $k/2$  in our nomenclature).

In practice, though chromosomal losses are commonly observed in tumor cells, it is far from desirable to have a mathematical framework that necessarily implies that monosomic cells must represent the dominant clone line. Similarly, it would be undesirable to require an expected average copy number as high as  $k/2$ . In the following work, we allow for a variety of possible stationary distributions by allowing variation of the mitosis rates as a function of the copy number of the chromosome in question. It would also be desirable to allow for variation of the segregation rates by copy number and, when possible, by chromosome type. The current work does not concern itself with the problem of tracking missegregations across different chromosomes, as this more difficult problem leads to computational unwieldiness. By restricting our attention to one chromosome we only need consider a  $k$ -state Markov chain, where  $k$  is the maximum allowed copy number. If we were to consider all of  $C$  chromosomes simultaneously, this would require a Markov chain with  $k^C$  states. Gusev et al. (4) use such a Markov chain, but to achieve computational feasibility are forced to use a biologically unrealistic fixed intermitotic time to study the long-term behavior of the system. Also, the Gusev Markov chain implicitly makes the biologically unrealistic assumption that all copy number aberrations should be equally likely, ignoring the selective advantage or disadvantage given by individual aberrations. Studies have shown (e.g., [17]) that copy number aberrations are not uniformly distributed, and, in fact, certain aberrations tend to be much more likely than others.

## 2.3 Continuous time modeling

While the discrete time model in Section 2.1 is useful for examining the long-term behavior of a synchronized population of cells with a constant mitosis rate, the assumptions used are

biologically unrealistic. Consider, for example, the role of the APC gene, a tumor suppressor that controls cell birth and death processes. It is believed that the inactivation or loss of both copies of APC leads to an increased cell birth to death ratio [45]. Thus, clone lines including this mutation would grow rapidly in the cell population. To address the problem of continuously varying mitosis rates in an asynchronous population, we need to shift from a discrete time model to a continuous time model for mitosis. The simulations of Gusev et al. [3] use two distributions for intermitotic time: a truncated normal distribution and a uniform distribution. Neither possibility has a biological justification, and both contradict studies modeling intermitotic times with real data [41–43,46–49]. We prefer an exponential distribution, for reasons that will become clear in Section 3.

Suppose that the time  $T$  for a cell  $c$  with copy number  $i$  to undergo mitosis can be described as an exponentially distributed random variable with parameter  $\lambda_i$ , and let  $X_i(t)$  denote the number of cells with copy number  $i$ , for  $i = 1, 2, \dots, k$ . We can describe the overall behavior of this system using a system of differential equations:

$$X_i'(t) = (1 - \lambda_i)X_i(t) + 2\lambda_i(1 - p)X_i(t) + r_{i-1}pX_{i-1}(t) + \lambda_{i+1}pX_{i+1}(t).$$

The expected behavior of the system is described by  $\mathbf{M}(t) = \exp(\mathbf{Q}t)$ , where

$$\mathbf{Q} = \begin{pmatrix} 1 + \lambda - 2\lambda_1 p & \lambda_2 p & 0 & \dots & \dots & \dots & 0 \\ \lambda_1 p & \ddots & & & & & \vdots \\ \vdots & & \ddots & \lambda_i p & 0 & & \vdots \\ \vdots & & \lambda_{i-1} p & 1 + \lambda_i - 2\lambda_i p & \lambda_{i+1} p & & \vdots \\ \vdots & & 0 & \lambda_j p & \ddots & & 0 \\ \vdots & & & & & \ddots & \lambda_k p \\ 0 & \dots & \dots & \dots & 0 & \lambda_{k-1} p & 1 + \lambda_k - 2\lambda_k p \end{pmatrix}.$$

i.e.  $q_{ii} = 1 + \lambda_i - 2\lambda_i p$ ,  $q_{i-1i} = q_{i+1i} = \lambda_i p$ , and  $q_{ij} = 0$  for  $|i - j| > 1$ . Given known values for the mitosis rates ( $\lambda_i$ ), we can express the entries of  $\mathbf{M}(t)$  in closed form as a function of  $p$  (although the exact formulae can be quite complicated). Given the decomposition of  $\mathbf{Q}$

as  $\mathbf{Q} = \mathbf{H}\mathbf{D}\mathbf{H}^{-1}$ , where  $\mathbf{D} = \begin{pmatrix} d_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d_k \end{pmatrix}$  is a diagonal matrix, note

that  $\exp(\mathbf{D}t) = \begin{pmatrix} \exp(d_1 t) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \exp(d_k t) \end{pmatrix}$ . We can then express  $\mathbf{M}(t)$  as

$$\mathbf{M}(t) = \mathbf{H}\exp(\mathbf{D}t)\mathbf{H}^{-1}$$

Consider a cell population growing at an exponential rate. The relationship between the doubling time  $T_d$  of the population and the exponential parameter  $\lambda$  is given by  $\lambda = \frac{\log 2}{T_d}$ . As an example, consider the case where the doubling time of a population of unaltered cells is 18 hours,  $k = 5$ , and cells with three or more copies of chromosome log 7 have an increased rate of growth with a doubling time of 17 hours; i.e.,  $\lambda_1 = \lambda_2 = \frac{\log 2}{18}$ , and  $\lambda_3 = \lambda_4 = \lambda_5 = \frac{\log 2}{17}$ . Let



us consider a missegregation rate of  $p = .01$  (suggested by Lengauer et al. [35]) and examine the behavior of  $\mathbf{M}(t)$ . In this case,

$$\mathbf{Q} = \begin{pmatrix} 1.03774 & 0.000385 & 0 & 0 & 0 \\ 0.000385 & 1.03774 & 0.0004077 & 0 & 0 \\ 0 & 0.000385 & 1.03996 & 0.0004077 & 0 \\ 0 & 0 & 0.0004077 & 1.03996 & 0.0004077 \\ 0 & 0 & 0 & 0.0004077 & 1.03996 \end{pmatrix}.$$

Let us assume we start with one cell with copy number 2. At time  $t$ , the expected state of the system can be described as

$$\mathbf{v}(t) = \mathbf{M}(t)\mathbf{v}(0) = \exp(\mathbf{Q}t) \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} v_1(t) \\ v_2(t) \\ v_3(t) \\ v_4(t) \\ v_5(t) \end{pmatrix}.$$

In this formulation, each function  $v_i(t)$  is a linear combination of the non-zero entries of  $\exp(\mathbf{Q}t)$ , i.e. a linear sum of five exponential functions. We are interested in the value of  $t$  for which the average copy number of a cell in the population will be at least 2.4, as this is the threshold for detection of a copy number aberration using the CGH methodology as explained below. (See du Manoir et al [50] for an extended discussion of threshold choices in CGH studies.) The average copy number in a sample can be found by considering the ratio

$$r(t) = \frac{\mathbf{z}_5 \cdot \mathbf{v}(t)}{\mathbf{1}_5 \cdot \mathbf{v}(t)},$$

where, as in Section 2.1,  $\mathbf{z}_5 = (1, 2, 3, 4, 5)$  and  $\mathbf{1}_5 = (1, 1, 1, 1, 1)$ . The numerator of this ratio counts the total number of chromosomes of interest (e.g., chromosome 7 in a colorectal polyp), while the denominator counts the total number of cells. With the parameters chosen,  $r(809.374) = 2.4000005$ . Thus with the parameters chosen, one would expect a CGH signal that would be interpreted as a gain sometime between 809 and 810 hours, i.e. after 33 days and 17–18 hours. This is considerably faster than the estimate from Section 2.1, a result attributable to the increased mitosis rate hypothesized for cells with larger copy numbers.

In the CGH literature, a gain is considered to be present when the fluorescence ratio comparing a tumor cell to a healthy cell is 1.2 or greater. In principle, the fluorescence ratio of 1.2 corresponds to an average copy number in the tumor cells of at least  $2 \cdot 1.2 = 2.4$ . In experimental practice, when the fluorescence ratio is 1.2, the average copy number may be slightly higher or lower than 2.4 due to imprecision.

### 3 Modeling mitosis: a historical perspective

#### 3.1 Transition probability model

In Section 2, we considered some simple models for mitosis, to facilitate a simultaneous consideration of mitosis and mutation. In this section, we consider several historical models for mitosis.

Smith and Martin provided one of the first coherent mathematical models of the cell cycle [41]. The cell cycle can be broken into four phases: the  $G_1$  phase before DNA synthesis, the  $S$

phase during which DNA is synthesized, the G<sub>2</sub> phase after DNA synthesis and before the M phase, when mitosis occurs. Whereas the G<sub>2</sub>, S, and M phases are typically of fixed duration in most cell types, the length of the G<sub>1</sub> phase can vary greatly, even within a single tissue sample. Smith and Martin proposed modeling the process as consisting of a state of indeterminate length (state A), contained in the G<sub>1</sub> phase, and a state of fixed length encompassing the other three phases (state B).

This process is modeled as follows: let  $T$  be a random variable describing the intermitotic time of a cell  $c$ . Then  $T = T_A + T_B$ , where  $T_A$  is a random variable describing the time spent in state A and  $T_B$  is a constant describing the time spent in state B. Consideration of observed interphase data led Smith and Martin to conclude that  $T$  is exponentially distributed.

The assumption that  $T$  is exponentially distributed has been widely used in practical flow cytometry (see e.g., Sahar et al [51] and Bertuzzi et al [52]), with or without the additional assumptions of the transition probability model. The assumption of an exponential distribution was used implicitly by Gray [53] to generate simulated data that was later used by Dean [54] to design and test the popular software SFIT for estimating DNA distributions in flow cytometry. Smith and Martin showed that the transition probability model together with the conclusion that  $T$  is exponentially distributed imply that  $T_A$  is exponentially distributed (see Appendix). Let  $\lambda_A$  be the parameter of this distribution, such that, for values of  $t$  greater than  $T_B$  and small values of  $\Delta t$ ,  $P[T \in (t, t + \Delta t) | T > t] \approx \lambda_A \Delta t$ .

Suppose a population of cells is growing according to the Smith-Martin model. Let  $A(t)$  denote the set of cells in state A at time  $t$ ,  $B(t)$  denote the set of cells in phase B at time  $t$  and let  $N(t) = |A(t) \cup B(t)|$ . Cell lines usually have a doubling time that is experimentally repeatable if there is enough space to grow and culture conditions (such as the amount and type of nutrients in the media) are approximately the same. This observation implies that, if one ignores cell loss, then the behavior of the system as a whole can be modeled as exponential growth, such that  $N(t) = N(0) \exp(\lambda t)$  for some constant  $\lambda$ .

We consider the growth rate as follows: let  $c$  be a cell chosen uniformly from  $X(t) = A(t) \cup B(t)$ , and let  $T$  be the amount of time that passes before  $c$  undergoes mitosis. An exponential distribution  $P$  is based on the proposition that there is a constant

$$\lambda = \lim_{\Delta t \rightarrow 0} P[T < \tau + \Delta t | T > \tau]$$

that is independent of  $\tau$ . Smith and Martin provided a formula relating  $\lambda$ ,  $\lambda_A$  and  $T_B$ . We present the formula and its derivation in the Appendix.

### 3.2 Sloppy size models

The Smith-Martin model for mitosis described in Section 3.1 provides a good fit for observed statistics of the cell cycle [55], and has been widely used and highly cited. The essential hypothesis that time to mitosis is exponentially distributed was used in developing the SFIT software, as noted above. However, its simplicity has been criticized for example, by Murphy et al [46] for failing to explain the correlations in generation times of sister cells, by Tyson [56] for failing to match some experimental data comparing the size and age of cells at mitosis, and by Koch [57] for lacking a biological basis. The exponential distribution, which Smith and Martin use to describe the lag time the cell experiences in the A state, is used often in the physical sciences but rarely in biological sciences. Whereas the exponential distribution requires a “memoryless” system, biological systems tend to be too complex to be accurately described as memoryless. Thus, the Smith-Martin model does not allow for any correlation between cell mass and the probability of mitosis, nor does it allow for any correlation between the intermitotic times of sister cells.



In contrast to the transition probability model, which expresses the probability of mitosis as a (fixed) function of time, the sloppy size model first proposed by Wheals [42] and then formalized by Tyson and Diekmann [43], expresses the probability of mitosis as a function of cell size (where ‘size’ can mean mass, volume, length, or some other measure of the cell). In later papers [47,58], which benefited from increased understanding of the molecular mechanisms of mitosis, size was modeled as the number of molecules of one or more proteins in the mitosis-associated CDC protein family.

In sloppy size modeling, it is typical to scale cell sizes to lie in the interval (0,1). The sloppy size model also presumes that there is a constant  $a$ , with  $0.5 < a < 1$ , such that cell mitosis only occurs in those cells whose size is at least  $a$ , but must occur in any cell before it reaches a size of 1 (on this scale). The lower bound serves the mathematical purpose of preventing the possibility of a new daughter cell from a mitosis immediately redividing a second time. Some biological justification for the lower bound on  $a$  can be found in the experimental data of Wheals [42], although that study concerns itself solely with asymmetric cell division in yeast.

Each daughter cell resulting from a mitosis inherits approximately half of the size  $s$  of the parent cell (and thus has size in the interval  $(a/2, 1/2)$ ). To allow some deviation in the sizes of the daughter cells, we divide the cell mass into two parts of size  $rx$  and  $(1-r)x$ , where  $r$  is normally distributed about  $\mu=0.5$  with a standard deviation of  $\sigma=0.016$ . These values of  $\mu$  and  $\sigma$  have been estimated by Sveiczer et al [48]. Letting  $x(t)$  represent the size of a cell at time  $t$ , we model cell growth using a growth function  $V(x)$  according to the differential equation

$$\frac{dx}{dt} = V(x).$$

Tyson and Diekmann [43] suggest a number of candidates for  $V(x)$ , two of which we discuss below.

In the sloppy size model, mitoses occur according to a probability distribution  $b(x)$  on the size variable  $x$ ; i.e., if  $\Delta t$  is the length of a small interval, the probability of witnessing a mitosis of a cell of size  $x$  is  $b(x)\Delta t$ .

Consider an individual cell  $c$ . Let  $x(0)$  be the size of  $c$  at the time of its separation from a parent cell. Then  $c$  will grow to at least the size  $a$  before mitosis is possible. Letting  $T_0$  denote the amount of time required for this initial stage of growth, we observe that

$$T_0 = \int_{x(0)}^a [V(x)]^{-1} dx.$$

For  $x > a$ , let  $P(x)$  be the probability that  $c$  will grow to at least size  $x$  before dividing. Then

$$P(x) = \exp\left(-\int_a^x \frac{b(\xi)}{V(\xi)} d\xi\right).$$

With the constraint that no cell grows greater than 1, the choices of  $b$  and  $V$  must result in the equality  $P(1) = 0$ , i.e.

$$\lim_{x \rightarrow 1^-} \int_a^x \frac{b(\xi)}{V(\xi)} d\xi = \infty.$$

To model exponential growth, we let  $V(x) = kx$  for some constant  $k$ . With this choice of  $V$ , the distribution  $b$  must have a singularity near  $x=1$  and must approach zero as  $x$  approaches  $a$  from

above, since  $b(x) = 0$  for  $x \leq a$ . To ensure continuity on the interval  $(0,1)$ , Tyson and Diekmann suggested the following choice for  $b$ :

$$b(x) = \begin{cases} 0, & 0 < x < a, \\ \frac{\hat{b}(x-a)^2}{(1-x)}, & a < x < 1, \end{cases}$$

where  $\hat{b}$  is a constant such that  $\int_0^1 b(\xi) d\xi = 1$ . Alternatively, rather than forcing the mitosis probability distribution to have a singularity as  $x$  approaches 1, we could use a growth function that leads to decelerating growth as  $x$  approaches 1. The logistic growth function  $V(x) = kx(1-x)$  allows the usage of any bounded distribution  $b$  with the constraint that  $b = 0$  outside the interval  $(a,1)$ . Following [43], we have used

$$b(x) = \begin{cases} 0, & 0 < x < a, \\ \hat{b}(x-a)^2, & a < x < 1, \end{cases}$$

again with  $\hat{b}$  chosen such that  $\int_0^1 b(\xi) d\xi = 1$ .

## 4 Simulation software

### 4.1 Software description

We have developed the software package MITOSIM, implementing the model of chromosomal missegregation described in Section 2 and the models of mitosis described in Section 3. MITOSIM maintains a queue of cells, ordered by the time to mitosis. The basic function is `sampleTime`: this function selects the cell at the front of the queue, creates two daughter cells, randomly determines whether the chromosomes divide equally to the two daughter cells, and calculates the subsequent division times for each daughter cell. Division time can be calculated according to any of the following distributions: normal, exponential, exponential with user-provided lag, discrete, constant, and sloppy size (with either exponential or logistic cell growth).

MITOSIM tests for chromosomal missegregation with each simulated mitosis. If no missegregation occurs, a cell with copy number  $i$  for the chromosome under consideration yields two daughter cells with copy number  $i$ , but a missegregation produces one daughter cell with copy number  $i-1$  and one with copy number  $i+1$  for the chromosome in question. Only cells with copy numbers in the range  $[1,k]$  are considered viable: if  $i-1 = 0$ , the cell has lost all copies of the genes residing on that chromosome (i.e. is nullisomic), is not considered to be viable and is removed from the simulation. Similarly, if  $i = k$ , the cell with copy number  $i+1$  is considered not to be viable and is removed from the simulation.

After each mitosis, the program adjusts the total chromosome count:  $\sum_c \# c$  where  $c$  ranges over the cells in the queue, and  $\# c$  is the copy number of the selected chromosome in cell  $c$ . The total copy number is then divided by the number of cells to calculate the average copy number. If the average copy number exceeds a user-provided threshold, the program notes the time elapsed and terminates, returning statistics including total time elapsed, total number of mitoses, total number of missegregations, total number of cells eliminated due to a zero or an excessively high copy number, and the average doubling time witnessed as the population grew.

In the laboratory setting, the cell lines may be resampled after a predetermined period of time, to maintain a limit to the size of the cell population. This resampling is modeled by MITOSIM by including each cell uniformly with a constant probability when a certain amount of time has passed. In theory, the resampling process should have a negligible effect on the average copy number, presuming the minimum sample size is sufficiently large (the default setting uses  $n=1000$ ). Also, using similar logic, this modeling ignores the possibility of regular cell death, as random cell deaths should have a negligible effect on the average copy number of the population.

#### 4.2 Parameter settings

To test MITOSIM, we modeled the growth of intestinal cells, with a specific eye toward the presence or absence of chromosome 7. We tested the rate of signal appearance based on two possibilities: a null hypothesis that missegregation of chromosome 7 has no effect on mitosis rates, and the hypothesis that cells with extra copies of chromosome 7 divide at an accelerated rate. The increased rate of cell division was calculated based on the following information.

On average, the lining of the colon is replaced (through a combination of shedding and apoptosis) every 3–5 days in mammals [59]. Thus, a proliferating crypt cell must divide around once every 96 hours in order to keep pace with the rate of cell loss. A colon polyp of  $1\text{ cm}^3$  containing  $1 \times 10^9$  cells takes about 7 years to develop, and would require 30 population doublings (i.e.  $2^{30}$ ). The healthy colon tissue growing at a standard exponential rate would go through approximately 640 generations over a seven year period. Thus, the appearance of a tumor reflecting an additional 30 population doublings could be accomplished by an increase in the growth rate of approximately 5%. We have modeled this increase in the growth rate by assuming that cells with at least one extra copy of chromosome 7 witness an increase in their mitosis rates of approximately 5% per extra copy. These calculations assume that an increase in cell proliferation, and not a marked reduction in apoptosis, is the dominant result of a genetic growth advantage. This is supported by the fact that Ki67 protein levels, a marker of cell proliferation, are increased [60].

Most CGH studies declare that a gain of a chromosomal region has occurred when the fluorescence ratio exceeds 1.2, which corresponds to 2.4 chromosomes per test cell against 2 chromosomes per reference cell. Thus, we have enforced a stopping point when the average copy number for cells in the population of the chromosome under consideration exceeded 2.4. (The value of 2.4 is provided as input by the user; analogous experiments could be done using any threshold.) CGH is the typical research tool used to assess gains and losses of chromosome material in tumor samples since it can be performed without needing the tumor to grow in culture, and also can be used (retrospectively) on formalin-fixed archived material. Nullisomic cells (i.e. cells missing all copies of a given chromosome) were eliminated from the simulation population because such cells are usually not observed *in vivo*.

#### 4.3 Comparison of various distributions

In this section, we consider the time required for the stop threshold to be reached under various conditions. We tested each of the following distributions: exponential, exponential + 4 hour lag, exponential + 8 hour lag, normal, sloppy size with an exponential growth rate, and sloppy size with a logistic growth rate. To test the null hypothesis, we calibrated the parameters of each distribution to have a doubling time of 18 hours. The 18 hour estimate was based on the behavior of colorectal cancer cell lines HCT116 (American Type Culture Collection cat # CCL-247), p53HCT116 [60], and DLD-1 (American Type Culture Collection cat # CCL-221) in our laboratory. The calibration was done analytically for the various exponential distributions, and by simulation for the sloppy size distributions. (The normal distribution was simply set to have a mean mitosis time of 18 hours, as the expected intermitotic time equals

the expected doubling time in this case.) In contrast to the 5-state model used in prior sections, these simulations were performed using a 7-state model. Results of these simulations are shown in Figure 1.

Figure 1 also includes a curve labeled “algebraic” that shows the values expected according to Mathematica calculations performed as described in Section 2.2 (but using 7 states instead of 5). As we see, the plot of signal time vs. missegregation rate is nearly identical regardless of which distribution is used to model the cell cycle when the missegregation rate is high.

To test the effect of an increased division rate, we then performed the same test, but scaled the rate of the cell cycle as a function of the number of chromosomes in a cell. Using the rough approximation of a 5% increase in the cell division rate for tumor cells, we scaled the mitosis rate parameters such that doubling times were decreased by 1 hour for each extra copy of chromosome 7 (and increased the doubling time of cells with copy number 1 by one hour). This hypothesis led to considerably faster growth of aberrant subpopulations, and to considerably faster achievement of the stop threshold. Results are shown in Figure 2.

Again, we see that the shape of the curve does not depend on the choice of the distribution used to simulate the cell cycle. The exponential curve suggests that, regardless of the distribution used for intermitotic times at the cellular level, an asynchronous population will observe mitoses at an exponential rate.

#### 4.4 Calculating p-values

Ultimately, the intended uses of MITOSIM are to test hypotheses about the missegregation ratio, the growth advantage for cells with extra chromosomes, and other parameters of the model. To compare two parameter settings  $P_1$  and  $P_2$  in a hypothesis testing framework, it is desirable to know to what extent the output distributions for time to CGH signal (and other outputs) under  $P_1$  and  $P_2$  overlap, if at all. Suppose that parameter setting  $P_1$  results in generally shorter times to reach the CGH ratio of 1.2. Then one could say that the time distribution for  $P_1$  differs from that of  $P_2$  at P-value (confidence level)  $\alpha$ , if the lower  $(1 - \alpha)$  fraction of the times for  $P_1$  do not overlap with the times for  $P_2$  or if the upper  $(1 - \alpha)$  fraction of the times for  $P_2$  do not overlap with the times for  $P_1$ ; which rule is applicable depends on which of  $P_1$  or  $P_2$  corresponds to the “null hypothesis”. The standard value of  $\alpha$  is 0.05. To facilitate such tests, MITOSIM prints out the 0.05 and 0.95 level for time to CGH signal and other quantities of interest.

Consider two examples based on the simulations summarized in Figures 1 and 2. First compare  $P_1 = 5\%$  mitotic advantage for an extra copy to  $P_2 =$  no advantage, with other settings fixed at  $\text{exp} + 8\text{hr}$  lag for the time to mitosis and 0.01 as the missegregation rate [35]. In this case, we consider  $P_2$  as the null hypothesis. The 0.95 level for time to CGH signal under  $P_1$  is 818.73 hours, while the minimum time under  $P_2$  is 1440.02 hours, so these parameter settings lead to statistically distinguishable outcomes. Second, compare  $P_1 = 0.05$  missegregation ratio,  $P_2 = 0.01$  missegregation ratio, and  $P_3 = 0.005$  missegregation ratio with all other settings fixed at 5% mitotic advantage and  $\text{exp} + 8\text{hr}$  lag time for mitosis time. Again we consider  $P_2$  as the null hypothesis, since 0.01 was the missegregation ratio measured in [35] for a colorectal cancer cell line. Under the  $P_1$  hypothesis, 95% of the simulations required less than 363.01 hours, while none of the 1000 simulations under the  $P_2$  hypothesis took less than 506.3 hours, which indicates that  $P_1$  and  $P_2$  can be statistically distinguished. In contrast, under the  $P_3$  hypothesis, 95% of the simulations required at least 774.97 hours, a threshold also achieved by 166 of the  $P_2$  simulations, indicating that the  $P_2$  and  $P_3$  distributions have a non-trivial overlap.

## 5 Discussion

We have presented a matrix algebra model, and accompanying simulation software MITOSIM, to consider the effects of chromosome copy number changes in cancer cell lines on mitosis times. Our simulations with MITOSIM agree with estimates of the matrix algebra model calculated using the software package Mathematica [44]. This modeling can be useful in testing hypotheses about why aberrant cells eventually dominate the population in a tumor. The presence of chromosome copy number changes can be measured by CGH. A much simpler modeling and simulation method has been used by Roschke et al. [61] to test predictions about chromosome rearrangements in cancer cell lines as measured by SKY.

The work of Gusev et al. [3,4] introduced the possibility of an algebraic model for chromosomal missegregations. Nowak et al. [62] and Komarova et al. [63] have also used mathematical modeling to investigate the onset of cancer and the role played by chromosomal instability during this process. The Nowak/Komarova models, though mathematically sophisticated, are more concerned with isolated genes than aneuploidy, and are not suitable for modeling the appearance of CGH signal in a population of cells.

There are numerous ways in which our models are more useful and relevant to real cell line data than the models of Gusev et al. First, the previous models did not allow for variability in the growth rates of cells with different copy numbers, which is the hypothesis we wish to test. Second, the Gusev model, by treating missegregations of each chromosome as an independent event with equal likelihood (as opposed to considering missegregations by chromosome *type*), creates a situation where monosomic cells are necessarily found in the most stable cell lines. While some chromosomal losses are believed to lead to a selective advantage for the respective clone line, it is undesirable for a mathematical model to have this as a necessary relationship. Third, our non-synchronized model using the exponential time distribution, with or without lag, can be solved algebraically to find the time to CGH signal for any amount of time; in contrast, the Gusev et al. method requires repeated multiplication of a non-sparse matrix. Fourth, the predictions of our models are expressed in terms of average copy number, which can be directly estimated in the laboratory by CGH. The Gusev et al. model does not lead to predictions easily testable by CGH or SKY. Fifth, we introduced in MITOSIM the technique of sampling, which overcomes the limitation on simulation time that Gusev et al. encountered. Sampling in the simulation is realistic because cell lines are sampled every few population doublings when they are grown in the laboratory.

The software MITOSIM offers two classes of cell cycle models, the transition probability of Smith and Martin [41] and the sloppy size control model as formulated by Tyson and Diekmann [43]. When viewed in the aggregate, a population of cells growing according to the complex process known as the cell cycle lends itself to analysis using the tools of linear algebra. The key to the current analysis is the usefulness of the exponential distribution that arises in the Smith-Martin model. The idea that the Smith-Martin model is a good starting point to make analytic estimates is also emphasized by Cain and Chau [64] and Baker et al [65], for example.

Aside from the transition probability model and the sloppy size model, there have been a number of previous models for the distributions of intermitotic times or the growth of cell populations. Some studies have focused on the statistical properties of intermitotic times, as did Smith and Martin. In [46,66], the Eyring-Stover survival theory was shown to better explain the observed positive correlation between sister cells than the Smith-Martin model (which implies no such correlation). Cain and Chau [64] extended the Smith-Martin model to consider variables such as cell death, a post-mitotic state of constant length, and also introduced substrate-dependent unbalanced growth using variable cell maturity velocity and/or an extra quiescent state during the B phase that growing cells can randomly enter or depart [67].

Tomasovic et al [68] used matrix algebra to estimate and compare parameters of the cell cycle under different conditions. This work attempted to estimate delays in the cell cycle caused by either x-rays or a drug, using a state space corresponding to time periods in the cell cycle. In theory, this type of approach can be used for any further refinement of time, though it would only be useful when distinct observations are available for each of the time periods in question. A set of coupled differential equations structurally similar to those we used for continuous time modeling was proposed by Gray [53]. Like Tomasovic et al., Gray used the states to mean subintervals of the cell cycle. He used the differential equations to drive a computer simulation and parameter fitting for the cell cycle, but did not use matrix algebra to obtain closed-form solutions.

Other models of the cell cycle have focused less on the intermitotic times than on either the cell sizes or the accumulation of mitosis initiator enzymes, just as the sloppy size model does. Kimmel et al [49] modeled mitosis probabilities purely as a function of the accumulation of the unequal division of nucleic acids. The 4-parameter continuum model of Cooper [69] is another model for cell mitosis where occurrence of mitosis happens as a function of cell size or of the amount of mitosis initiator in each cell. The continuum model suggests that variable lengths of the  $G_1$ -phase are a consequence, rather than a cause, of varying intermitotic times. Since the lengths of the M,  $G_2$ , and S phases are essentially constant, the variation in intermitotic times, regardless of its cause, leads to a variation in the length of the  $G_1$  phase. This model can explain the lack of the  $G_1$  phase in prokaryotic cells, a lack generally ignored by most models.

There are several other models of the cell cycle, which might be considered as alternatives. For most of the models mentioned here, it is unclear how to represent the simple hypothesis that an extra copy of a chromosome leads to faster mitosis within the numerous model parameters. One model, the “tandem model” proposed by Tyson and Hanngsen [70], is intermediate between the two models we have considered. The tandem model uses two states like the Smith-Martin model but uses rigid size control instead of sloppy size control.

Koch has considered the historical evolution of the transition probability model, and has criticized it for being primarily phenomenological, with insufficient biological motivation [57]. But for our purposes, namely considering the large-scale growth behavior of a cell population while trying to reduce the number of variables considered, the transition probability model appears to be adequate. All of the models we considered, including the more complex sloppy size models, yielded nearly identical exponential decay curves (Figures 1, 2) describing the appearance of CGH signal as a function of missegregation rate. One can see some differences in the estimated time for low missegregation rates in Figure 1 when the expected mitosis time is the same for all copy numbers, but the ability to discriminate between small advantage for extra copies (Figure 2) and no advantage (Figure 1) is virtually identical. Thus, though the sloppy size model (and presumably other models) may represent a more comprehensive biological understanding, the increased complexity of these models does not appear to add to our understanding of the growth of larger populations. Indeed, the sloppy size model requires the usage of density functions that we have defined arbitrarily, and whose direct observation would be problematic given typical data set sizes. In conclusion, given the easy transition from the simple transition probability model to the matrix algebraic formulation of the problem, we prefer the TP model based on a consideration of the limit of the amount of data readily available, and a general preference for simpler models over more complex models, when both models appear to be adequate for a given task.

#### Acknowledgements

We send our thanks to Prof. Marc Lalande for pointing out the work of Dean and Gray and the software SFIT. Also, thanks to the anonymous referee for helpful comments



## References

1. Ghadimi BM, Schröck E, Walker RL, Wangsa D, Jauho A, Meltzer PS, Ried T. Specific chromosomal aberrations and amplification of the *AIB1* nuclear receptor coactivator gene in pancreatic carcinomas. *Am J Pathol* 1999;154:525–536. [PubMed: 10027410]
2. Heselmeyer-Haddad K, Janz V, Castle PE, Chaudhri N, White N, Wilber K, Morrison LE, Auer G, Burroughs FH, Sherman ME, Ried T. Detection of genomic amplification of the human telomerase gene (*TERC*) in cytologic specimens as a genetic test for the diagnosis of cervical dysplasia. *Am J Pathol* 2003;163:1405–1416. [PubMed: 14507648]
3. Gusev Y, Kagansky V, Dooley WC. A Stochastic Model for Chromosome Segregation Errors with Reference to Cancer Cells. *Mathematics and Computer Modelling* 2000;32:97–111.
4. Gusev Y, Kagansky V, Dooley WC. Long-term dynamics of chromosomal instability in cancer: a transition probability model. *Mathematics and Computer Modelling* 2001;33:1253–1273.
5. Hayward CMM, Griffin GE. Antibiotic resistance: the current position and the molecular mechanisms involved. *Br J Hosp Med* 1994;52:473–478. [PubMed: 7874363]
6. Olsen JE. Antibiotic resistance: Genetic mechanisms and mobility. *Acta Vet Scand Suppl* 1999;92:15–22. [PubMed: 10783713]
7. Friedman MJ. Erythrocytic mechanism of sickle cell resistance to malaria. *Proc Natl Acad Sci USA* 1978;75:1994–1997. [PubMed: 347452]
8. Aluoch JR. Higher resistance to *Plasmodium falciparum* infection in patients with homozygous sickle cell disease in western Kenya. *Trop Med Int Health* 1997;2:568–571. [PubMed: 9236824]
9. Lell B, May J, Schmidt-Ott RJ, Lehman LG, Luckner D, Greve B, Matousek P, Schmid D, Herbich K, Mockenhaupt FP, Meyer CG, Bienzle U, Kreamsner PG. The role of red blood cell polymorphisms in resistance and susceptibility to malaria. *Clin Inf Dis* 1999;28:794–799.
10. Pastan I, Gottesman M. Multiple-drug resistance in human cancer. *N Engl J Med* 1987;316:1388–1393. [PubMed: 3553950]
11. Wasenius VM, Jekunen A, Monni O, Joensuu H, Aebi S, Howell SB, Knuutila S. Comparative genomic hybridization analysis of chromosomal changes occurring during development of acquired resistance to cisplatin in human ovarian carcinoma cells. *Genes Chromosomes Cancer* 1997;18:286–291. [PubMed: 9087568]
12. Knutsen T, Mickley LA, Ried T, Green ED, du Manoir S, Schröck E, Macville M, Ning Y, Robey R, Polymeropoulos M, Torres R, Fojo T. Cytogenetic and molecular characterization of random chromosomal rearrangements activating the drug resistance gene, *MDR1/P-glycoprotein*, in drug-selected cell lines and patients with drug refractory ALL. *Genes Chromosomes Cancer* 1998;23:44–54. [PubMed: 9713996]
13. Thacker J, Ganesh AN, Stretch A, Benjamin DM, Zahalsky AJ, Hendrickson EA. Gene mutation and V(D)J recombination in the radiosensitive *irs* lines. *Mutagenesis* 1994;9:163–168. [PubMed: 8201950]
14. Weaver DT. What to do at an end: DNA double-strand-break repair. *Trends Genet* 1995;11:388–392. [PubMed: 7482764]
15. Nussenzweig A, Sokol K, Burgman P, Li L, Li GC. Hypersensitivity of *Ku80*-deficient cell lines and mice to DNA damage: The effects of ionizing radiation on growth, survival, and development. *Proc Natl Acad Sci USA* 1997;94:13588–13593. [PubMed: 9391070]
16. Abbott DW, Freeman ML, Holt JT. Double-strand break repair deficiency and radiation sensitivity in BRCA2 mutant cancer cells. *J Natl Cancer Inst* 1998;90:978–985. [PubMed: 9665145]
17. Ried T, Heselmeyer-Haddad K, Blegen H, Schröck E, Auer G. Genomic changes describing the genesis, progression, and malignancy potential in solid human tumors: A phenotype/genotype correlation. *Genes. Chromosomes and Cancer* 1999;25:195–204.
18. Müller A, Korabiowska M, Brinck U. Review. DNA-mismatch repair and hereditary nonpolyposis colorectal cancer syndrome. *In Vivo* 2003;17:55–59. [PubMed: 12655791]
19. Jacob S, Praz F. DNA mismatch repair defects: role in colorectal carcinogenesis. *Biochimie* 2002;84:27–47. [PubMed: 11900875]
20. Fishel R. The selection for mismatch repair defects in hereditary nonpolyposis colorectal cancer: revising the mutator hypothesis. *Cancer Res* 2001;61:7369–7374. [PubMed: 11606363]

21. Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinänen R, Palmberg C, Palotie A, Tammela T, Isola J, Kallioniemi OP. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* 1995;9:401–406. [PubMed: 7795646]
22. Phillips JL, Hayward SW, Wang Y, Vasselli J, Pavlovich C, Padilla-Nash H, Pezullo JR, Ghadimi BM, Grossfeld GD, Rivera A, Linehan WM, Cunha GR, Ried T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Res* 2001;61:8143–8149. [PubMed: 11719443]
23. Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J, Nosrati A, Willson JKV, Mack D, Ried T, Markowitz S. Silence of chromosomal amplifications in colon cancer. *Cancer Res* 2002;62:1134–1138. [PubMed: 11861394]
24. Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousset S, Rozenblum E, Ringnér M, Sauter G, Monni O, Elkahoul A, Kallioniemi OP, Kallioniemi A. Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res* 2002;62:6240–6245. [PubMed: 12414653]
25. Pollack JR, Sørlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Børresen-Dale AL, Brown PO. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA* 2002;99:12963–12968. [PubMed: 12297621]
26. Upender MB, Habermann JK, McShane LM, Korn EL, Barrett JC, Difilippantonio MJ, Ried T. Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. *Cancer Res* 2004;64:6941–6949. [PubMed: 15466185]
27. S. Heim, F. Mitelman, *Cancer Cytogenetics*, 2<sup>nd</sup> edition, Wiley-Liss, New York, 1995.
28. Bardi G, Johansson B, Pandis N, Heim S, Mandahl N, Andréén-Sandberg Å, Hägerstrand I, Mitelman F. Trisomy 7 in short-term cultures of colorectal adenocarcinomas. *Genes Chromosomes Cancer* 1991;3:149–152. [PubMed: 2069912]
29. Gebhart E, Rau D, Neubauer S, Dingermann T. Trisomy 7 in short-term cultures of colorectal adenocarcinomas. *Genes Chromosomes Cancer* 1992;4:104–105. [PubMed: 1377004]
30. Ried T, Knutzen R, Steinbeck R, Blegen H, Schröck E, Heselmeyer K, du Manoir S, Auer G. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* 1996;15:234–245. [PubMed: 8703849]
31. Bomme L, Lothe RA, Bardi G, Fenger C, Kronborg O, Heim S. Assessments of clonal composition of colorectal adenomas by FISH analysis of chromosomes 1, 7, 13 and 20. *Int J Cancer* 2001;92:816–823. [PubMed: 11351301]
32. Nicholson RI, Gee JMW, Harper ME. EGFR and cancer prognosis. *Eur J Cancer* 2001;37(Suppl 4):S9–15. [PubMed: 11597399]
33. Roberts RB, Min L, Washington MK, Olsen SJ, Settle SH, Coffey RJ, Threadgill DW. Importance of epidermal growth factor receptor signaling in establishment of adenomas and maintenance of carcinomas during intestinal tumorigenesis. *Proc Natl Acad Sci USA* 2002;99:1521–1526. [PubMed: 11818567]
34. Cohen RB. Epidermal growth factor receptor as a therapeutic target in colorectal cancer. *Clin Colorectal Cancer* 2003;2:246–251. [PubMed: 12620146]
35. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. *Nature* 1997;386:623–627. [PubMed: 9121588]
36. Pihan GA, Doxsey SJ. The mitotic machinery as a source of genetic instability in cancer. *Semin Cancer Biol* 1999;9:289–302. [PubMed: 10448116]
37. Li R, Sonik A, Stindl R, Rasnick D, Duesberg P. Aneuploidy vs. gene mutation hypothesis of cancer: recent study claims mutation but is found to support aneuploidy. *Proc Natl Acad Sci USA* 2000;97:3236–3241. [PubMed: 10725343]
38. Schröck E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-An I, Soenksen D, Ried T. Multicolor spectral karyotyping of human chromosomes. *Science* 1996;273:494–497. [PubMed: 8662537]
39. Speicher MR, Gwyn Ballard S, Ward DC. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genet* 1996;12:368–375. [PubMed: 8630489]

40. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992;258:818–821. [PubMed: 1359641]
41. Smith JA, Martin L. Do cells cycle? *Proc Natl Acad Sci USA* 1973;70:1263–1267. [PubMed: 4515625]
42. Wheals AE. Size control models of *Saccharomyces cerevisiae* cell proliferation. *Mol Cell Biol* 1982;2:361–368. [PubMed: 7050671]
43. Tyson JJ, Diekmann O. Sloppy size control of the cell division cycle. *J Theor Biol* 1986;118:405–426. [PubMed: 3520151]
44. S. Wolfram, *The Mathematica Book*, 4<sup>th</sup> edition, Cambridge University Press, Cambridge, UK, 1999.
45. Komarova NL, Lengauer C, Vogelstein B, Nowak MA. Dynamics of genetic instability in sporadic and familial colorectal cancer. *Cancer Biol Ther* 2002;1:685–692. [PubMed: 12642695]
46. Murphy JS, Landsberger FR, Kikuchi T, Tamm I. Occurrence of cell division is not exponentially distributed: Differences in the generation times of sister cells can be derived from the theory of survival of populations. *Proc Natl Acad Sci USA* 1984;81:2379–2383. [PubMed: 6585804]
47. Tyson JJ. Effects of asymmetric division on a stochastic model of the cell division cycle. *Math Biosci* 1989;96:165–184. [PubMed: 2520196]
48. Svecizer A, Tyson JJ, Novak B. A stochastic, molecular model of the fission yeast cell cycle: role of the nucleocytoplasmic ratio in cycle time regulation. *Biophysical Chem* 2001;92:1–15.
49. Kimmel M, Darzynkiewicz Z, Arino O, Traganos F. Analysis of a cell cycle model based on unequal division of metabolic constituents to daughter cells during cytokinesis. *J Theor Biol* 1984;110:637–664. [PubMed: 6084152]
50. du Manoir S, Schröck E, Bentz M, Speicher MR, Joos S, Ried T, Lichter P, Cremer T. Quantitative analysis of comparative genomic hybridization. *Cytometry* 1995;19:27–41. [PubMed: 7705182]
51. Sahar E, Wage ML, Latt SA. Maturation rates and transition probabilities of cycling cells. *Cytometry* 1983;4:202–210. [PubMed: 6661987]
52. Bertuzzi A, Gandolfi A, Germani A, Spanò M, Starace G, Vitelli R. Analysis of DNA synthesis rate of cultured cells from flow cytometric data. *Cytometry* 1984;5:619–628. [PubMed: 6518937]
53. Gray JW. Cell-cycle analysis of perturbed cell populations: Computer simulation of sequential DNA distributions. *Cell Tissue Kinet* 1976;9:499–516. [PubMed: 187341]
54. Dean PN. A simplified method of DNA distribution analysis. *Cell Tissue Kinet* 1980;13:299–308. [PubMed: 6989507]
55. Minor PD, Smith JA. Explanation of degree of correlation of sibling generation times in animal cells. *Nature* 1974;248:241–243. [PubMed: 4206605]
56. Tyson JJ. The coordination of cell growth and division – intentional or incidental? *BioEssays* 1989;2:72–77.
57. Koch A. The re-incarnation, re-interpretation and re-demise of the transition probability model. *J Biotech* 1999;71:143–156.
58. Alt W, Tyson JJ. A stochastic model of cell division (with applications to fission yeast). *Math Biosci* 1987;84:159–187.
59. van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Battle E, Coudreuse D, Haramis AP, Tjon-Pon-Fong M, Moerer P, van den Born M, Soete G, Pals S, Eilers M, Medema R, Clevers H. The  $\beta$ -catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 2002;111:241–250. [PubMed: 12408868]
60. Bunz F, Dutriaux A, Lengauer C, Waldman T, Zhou S, Brown JP, Sedivy JM, Kinzler KW, Vogelstein B. Requirement for p53 and p21 to sustain G<sub>2</sub> arrest after DNA damage. *Science* 1998;282:1497–1501. [PubMed: 9822382](p53HCT116 cell line)
61. Roschke AV, Stover K, Tonon G, Schäffer AA, Kirsch IR. Stable karyotypes in epithelial cancer cell lines despite high rates of ongoing structural and numerical chromosomal instability. *Neoplasia* 2002;4:19–31. [PubMed: 11922387]
62. Komarova NL, Sengupta A, Nowak MA. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *J Theor Biol* 2003;223:433–450. [PubMed: 12875822]

63. Nowak MA, Komarova NL, Sengupta A, Jallepalli PV, Shih IM, Vogelstein B, Lengauer C. The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci USA* 2002;99:16226–16231. [PubMed: 12446840]
64. Cain SJ, Chau PC. Transition probability cell cycle model part I – balanced growth. *J Theor Biol* 1997;185:55–67. [PubMed: 9093553]
65. Baker CTH, Bocharov GA, Paul CAH, Rihan FA. Modelling and analysis of time-lags in some basic patterns of cell proliferation. *J Math Biol* 1998;37:341–371. [PubMed: 9819894]
66. Murphy JS, D’Alisa R, Gershey EL, Landsberger FR. Kinetics of desynchronization and distribution of generation times in synchronized cell populations. *Proc Natl Acad Sci USA* 1978;75:4404–4407. [PubMed: 279924]
67. Cain SJ, Chau PC. Transition probability cell cycle model part II – non-balanced growth. *J Theor Biol* 1997;185:69–79. [PubMed: 9093554]
68. Tomasovic SP, Roti Roti JL, Dethlefsen LA. Matrix algebraic simulation of mitotic cell selection experiments. *Cell Tissue Kinet* 1980;13:117–33. [PubMed: 7363299]
69. Cooper S. The continuum model: statistical implications. *J Theor Biol* 1982;94:783–800. [PubMed: 7078225]
70. Tyson JJ, Hannngsen KB. The distributions of cell size and generation time in a model of the cell cycle incorporating size control and random transitions. *J Theor Biol* 1985;113:29–62. [PubMed: 3999773]

## Appendix

In this appendix, we review the relationship between the expected intermitotic time of a single cell  $T$  and the expected doubling time  $T_d$  of a population of cells. We also consider how the doubling time varies when considering the Smith-Martin distribution where the intermitotic time  $T$  is generated as the sum of an exponentially distributed variable  $T_A$  and a constant  $T_B$ . The derivations are from Smith and Martin [41] under their assumptions that the set of cells have a doubling time  $T_d$ , and that one can ignore cell loss because the cells are sampled to start a new culture every few days. The derivations are presented here for the sake of completeness.

In the simplest case,  $T$  is a constant over all cells, and  $E[T] = E[T_d] = T$ ; i.e. the doubling time equals the expected intermitotic time. But in general, this equality need not hold. Consider the case where  $T$  is exponentially distributed with parameter  $\lambda$ . In this case,  $E[T] = \frac{1}{\lambda}$ , but

$E[T_d] = \frac{\ln 2}{\lambda}$ . The doubling time is less than the expected mitosis time because quickly dividing cells decrease the doubling time more than slowly dividing cells increase it. In practice, we let  $\hat{T}_d$  be the observed doubling time, and solve for our estimate for  $\lambda$ ,  $\hat{\lambda} = \frac{\ln 2}{\hat{T}_d}$ .

Now consider the Smith-Martin distribution, where  $T = T_A + T_B$ ,  $T_A$  is random and  $T_B$  is a constant. The goal is to show, given that  $T$  is exponentially distributed (with parameter  $\lambda$ , that  $T_A$  is itself exponentially distributed, and to derive a closed form for its parameter  $\lambda_A$  as a function of  $\lambda$  and  $T_B$ . Let  $A(t)$  and  $B(t)$  denote the sets of cells in state A and B respectively at time  $t$ , and let  $N(t) = |A(t) \cup B(t)|$ , the total number of cells at time  $t$ .

Let  $c$  be a cell in the population at time  $t$ . The cell  $c$  undergoes mitosis at time  $t$  if and only if  $c$  leaves state A at time  $t - T_B$ . Since the former event is memoryless event whose passage time can be described by an exponentially distributed variable  $T$ , the passage time of the latter event can also be described by an exponentially distributed variable. It follows that  $T_A$  is also exponentially distributed, and we can define its parameter  $\lambda_A$  using the equation:

$$\begin{aligned} \lambda \Delta t &= P[c \text{ divides in interval } (t, t + \Delta t)] \\ &= P[c \text{ leaves state A in interval } (t - T_B, t - T_B + \Delta t)] \\ &= P[c \in A(t - T_B)] \lambda_A \Delta t. \end{aligned}$$

Let  $N_A(t)$  and  $N_B(t)$  denote the number of cells in the population in state A or state B, respectively, at time  $t$ , such that  $N(t) = N_A(t) + N_B(t)$ . Let  $t_0$  be an arbitrary real number, and let  $t_1 = t_0 + T_B$ . The number of new cells created in an interval of length  $T_B$  is precisely equal to the number of cells in state B at the beginning of the interval. Thus

$$N(t_1) - N(t_0) = N_B(t_0).$$

Also, presuming exponential growth of the population, the behavior of  $N(t)$  can be described by the equation

$$N(t) = N(0)\exp(\lambda t). \quad (\text{A.1})$$

Thus

$$N(t_0)\exp(\lambda T_B) - N(t_0) = N_B(t_0). \quad (\text{A.2})$$

Adding  $N_A(t_0)$  to (A.2) yields

$$N(t_0) = [N(t_0)\exp(\lambda T_B) - N(t_0)] + N_A(t_0),$$

which leads to (after dividing by  $N(t_0)$ )

$$\frac{N_A(t_0)}{N(t_0)} = 2 - \exp(\lambda T_B),$$

so  $N_A/N$  is invariant as a function of  $t$ .

Therefore, one introduces a “rate constant”  $\lambda_A$  such that the expected number of cells undergoing mitosis at time  $t + T_B$  is

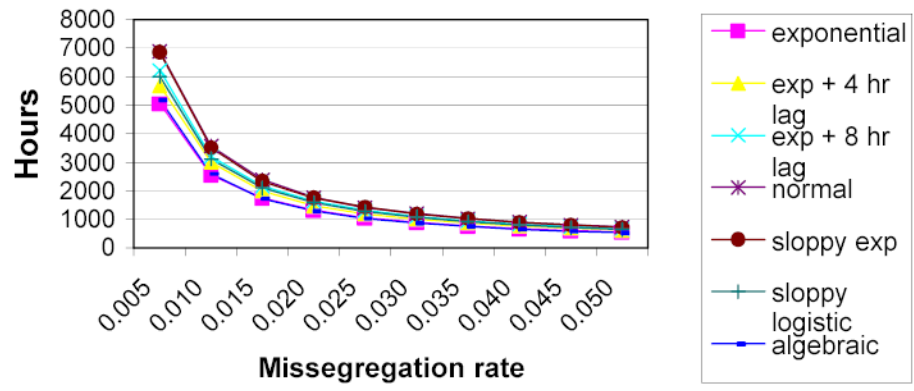
$$\lambda N(t + T_B) = \lambda_A N_A(t). \quad (\text{A.3})$$

The equation relates the rate of mitosis to the rate at which cells leave state A. Since the time to mitosis is exponential, the set of cells in state B is decaying exponentially with parameter  $\lambda$ , and hence the set of cells in state A at time  $t$  will also decay exponentially with parameter  $\lambda_A$ . We can use Equation (A.1) to transform the left-hand side of (A.3):

$$\lambda N(t + T_B) = \lambda N(t)\exp(\lambda T_B),$$

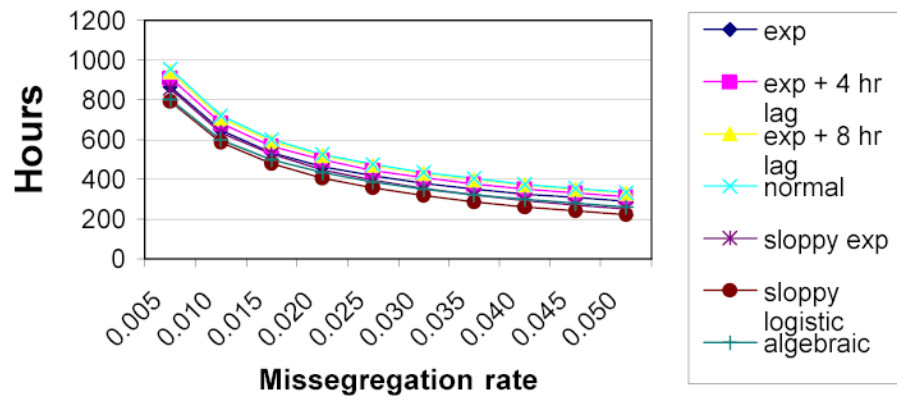
and thus

$$\lambda_A = \frac{\lambda \exp(\lambda T_B) N(t)}{N_A(t)} = \frac{\lambda \exp(\lambda T_B)}{2 - \exp(\lambda T_B)}.$$



**Figure 1.**  
Signal times for constant rate mitosis models





**Figure 2.**  
Signal times for varying rate mitosis models