Title

A Rapid-Mutation Approximation for Cell Population Dynamics

Journal

Bulletin of Mathematical Biology: A Journal Devoted to Research at the Junction of Computational, Theoretical and Experimental Biology Official Journal of The Society for Mathematical Biology, 72(2)

ISSN

1522-9602

Authors

Sachs, Rainer K.
Hlatky, Lynn

Publication Date

2010-02-01

DOI

10.1007/s11538-009-9450-6

Peer reviewed

Bulletin of Mathematical Biology (2010) 72: 359–374
DOI 10.1007/s11538-009-9450-6

ORIGINAL ARTICLE

A Rapid-Mutation Approximation for Cell Population Dynamics

Rainer K. Sachs[a,*], Lynn Hlatky[b]

[a] *Departments of Mathematics and Physics, University of California, 970 Evans Hall, MC 3840, Berkeley, CA 94720, USA*
[b] *Department of Medicine, Tufts School of Medicine, Boston, MA, USA*

Received: 23 February 2009 / Accepted: 14 August 2009 / Published online: 30 December 2009

Abstract Carcinogenesis and cancer progression are often modeled using population dynamics equations for a diverse somatic cell population undergoing mutations or other alterations that alter the fitness of a cell and its progeny. Usually it is then assumed, paralleling standard mathematical approaches to evolution, that such alterations are slow compared to selection, i.e., compared to subpopulation frequency changes induced by unequal subpopulation proliferation rates. However, the alterations can be rapid in some cases. For example, results in our lab on in vitro analogues of transformation and progression in carcinogenesis suggest there could be periods where rapid alterations triggered by horizontal intercellular transfer of genetic material occur and quickly result in marked changes of cell population structure.

We here initiate a mathematical study of situations where alterations are rapid compared to selection. A classic selection-mutation formalism is generalized to obtain a "proliferation-alteration" system of ordinary differential equations, which we analyze using a rapid-alteration approximation. A system-theoretical estimate of the total-population net growth rate emerges. This rate characterizes the diverse, interacting cell population acting as a single system; it is a weighted average of subpopulation rates, the weights being components of the Perron–Frobenius eigenvector for an ergodic Markov-process matrix that describes alterations by themselves. We give a detailed numerical example to illustrate the rapid-alteration approximation, suggest a possible interpretation of the fact that average aneuploidy during cancer progression often appears to be comparatively stable in time, and briefly discuss possible generalizations as well as weaknesses of our approach.

Keywords Carcinogenesis · Somatic cell population diversity · Proliferation · Fitness alterations · Evolutionary ecology · Replicator-mutator equations

*Corresponding author.
E-mail address: sachs@math.berkeley.edu (Rainer K. Sachs).

Abbreviation

"Alteration" refers to changes that can affect the fitness of a somatic cell and its progeny, for example, any of the following: point mutations in important genes, other comparatively small-scale DNA modifications, larger-scale DNA gains or losses, chromosome rearrangements such as translocations, changes in chromosome copy number, or persistent epigenetic changes.

1. Introduction

When analyzing carcinogenesis, one needs to consider various kinds of *alterations* in somatic cells that can be inherited by daughter cells and change fitness. In addition to point mutations in important genes, there are other comparatively small-scale DNA changes, larger scale structural chromosomal changes such as translocations and chromatin duplications or deletions (Mitelman et al., 2007), persistent epigenetic modifications (Chin and Gray, 2008; Vucic et al., 2008), changes in chromosome copy numbers (Duesberg et al., 2005; Chi and Jeang, 2007), horizontal transfer of genetic materials (Bjerkvig et al., 2005), etc.

Carcinogenesis and tumor progression are often modeled as evolutionary or ecological-evolutionary processes for a diverse somatic cell population whose subpopulation frequencies change due to *selection* (i.e., to heritably different net proliferation rates) or to random drift (reviewed in Gatenby and Vincent, 2003; Michor et al., 2003; Frank and Nowak, 2004; Nagy, 2004; Komarova, 2005; Merlo et al., 2006). For example, the classic two-stage clonal expansion carcinogenesis model and its generalizations assume one or more initiation alterations, each typically followed by a prolonged period of clonal expansion, then a further malignant-transformation alteration, and then tumor progression (reviewed in Moolgavkar and Luebeck, 2003; Little et al., 2008).

Usually these evolutionary approaches to carcinogenesis assume that selection is rapid compared to the time interval between fitness-altering alterations, paralleling analyses of species evolution where relevant mutations occur very infrequently compared to the lifetime of an organism. For example, one often starts by considering a zeroth-order limiting case where cell mutation rates are negligible and then estimates the effects of infrequent mutations perturbing this zeroth-order approximation (e.g. Michor et al., 2005). Even in analyses of "stochastic tunneling," where several alterations may occur before selection or random drift result in a complete takeover of a cell population niche by a subpopulation (Iwasa et al., 2004), the alteration rate is usually considered small compared to the proliferation rate.

However, in some situations, e.g., situations involving chromosomal instability (Yuen and Desai, 2008), time-dependent aneuploidy (Duesberg et al., 2007), horizontal transfer of genetic material or cell fusion (Bjerkvig et al., 2005), or mitochondrial mutations (Coller et al., 2001), the alterations can be rapid compared to selection. Rapid alterations call for starting from a zeroth-order approximation different from the no-alteration or slow-alteration approximations usually considered. Our goal is to devise a mathematical/computational approach to cell population dynamics that can systematically approximate situations where alterations are rapid.

To this end, we first introduce a proliferation-alteration system of ordinary differential equations which generalizes selection-mutation systems suggested earlier (Hofbauer and Sigmund, 1998; Marx et al., 2007). The proliferation-alteration system has the following three properties, which have not previously, as far as we are aware, been combined in a cell population dynamics approach.

- The equations involve cell subpopulation numbers rather than just subpopulation frequencies.
- Rather than just tracking independent subpopulations, the equations incorporate cell-cell interactions of a general type; such interactions are currently believed to play an important role in radiation-induced carcinogenesis and cancer progression (reviewed, e.g., in Sachs et al., 2005; Weinberg, 2007). Specifically, each subpopulation proliferation rate can depend nonlinearly on all the cell subpopulation numbers.
- In contrast to typical replicator-mutator equations (reviewed, e.g., in Page and Nowak, 2002; Komarova, 2005), our equations have alteration terms which are explicitly separated out from proliferation terms and are comparatively simple, allowing the implementation of a "rapid-alteration" approximation procedure.

As a zeroth-order approximation, we analyze the effects of the alteration terms alone. In the special case that will be emphasized here, this zeroth-order approximation reduces to the differential equations of an ergodic Markov process, so that many standard techniques, for example, the potent Perron–Frobenius theory of nonnegative matrices (Minc, 1988), can be brought to bear (Kijima, 1997). As first-order effects modulating the zeroth-order approximation, we then take into account heterogeneous subpopulation proliferation rates, using a "slowly varying amplitude" approach which has some similarities to the WKBJ approximation of quantum mechanics (Griffiths, 2004) and to approximations used in deriving geometric from physical optics (Born et al., 1999). Under surprisingly general conditions, this approach leads to a system-theoretical estimate for the net growth rate (positive or negative) of the diverse, interacting cell population considered as a whole, somewhat similar to the asymptotic overall Malthusian parameter for a heterogeneous population with asymptotically stable subpopulation frequencies at large times.

A specific example, illustrating the proliferation-alteration formalism and accuracy estimates for the rapid-alteration approximation will be given; the example grew out of a preliminary mathematical analysis of experiments studying aneuploidy during in vitro analogues of transformation and progression in our laboratory, but will here be treated simply as a computational example. We will also discuss briefly the following points:

- The rapid-alteration approximation could give new insights into other scenarios, especially for human solid tumors and in vitro analogues, where an unexpected degree of karyotype stability is often found despite highly heterogeneous aneuploidy at each instant (Nowell, 1976; Eshleman et al., 1998; Macville et al., 1999; Roschke et al., 2002; Jin et al., 2005; Li et al., 2009).
- The formalism can be generalized, for example, by considering a state space which is a direct sum of state spaces for each one of which our assumptions hold.
- The approach has weaknesses. For example, the case where proliferation rates are comparable to alteration rates, which is important in some cancer progression scenarios, is no more amenable to our treatment than it is to older treatments that assume selection is rapid compared to alteration.

2. Mathematical methods

2.1. Assumptions

We consider a diverse system of somatic cells, modeled as a population where the number of cells in the jth subpopulation at time t is represented by a smooth nonnegative function $y_j(t)$. We assume there are $N+1$ different cell subpopulations, "states," labeled by $j = 0, 1, 2, \ldots, N$ (letting j run from 0 to N rather than 1 to N keeps consistency with some of our main references); generalizations to infinite state-spaces are outlined in the Discussion section, but our application to data, and our formal developments here, will use finite N. Different cell subpopulations are thought of as being characterized by their own heritable, fitness-influencing genomic, and/or epigenetic configurations. The treatment will not attempt to take into account the fact that subpopulation numbers y_j are in some scenarios more realistically modeled as random functions of time which jump from nonnegative integer to nonnegative integer as the cell population, subject to demographic/stochastic fluctuations, develops in time. The underlying picture is thus deterministic rather than stochastic, despite our later use of Markov process theory as a mathematical tool. A partial justification for this deterministic approach is that the phenomenon of subpopulation extinction, which is often the main reason for replacing a deterministic by a more detailed stochastic analysis (e.g. Tan, 2002; Sachs et al., 2007), is not as important in the present context as in many other population dynamics models, because on the present picture rapid alterations can rapidly resurrect an extinct subpopulation.

We will assume the time-development of the cell population is governed by the following proliferation-alteration system of ordinary differential equations:

$$\frac{dy_k}{dt} = y_k f_k(y_0, y_1, \ldots, y_N) + \sum_{j=0}^{N} y_j Q_{jk}, \quad k = 0, 1, \ldots, N. \tag{1}$$

Here:

(1) f_k, modeling net proliferation rates, is a smooth function of the subpopulation numbers $y_j(t)$. For example, f_k might be a net proliferation rate constant λ_k, and if f_k is a given constant for each k then Eq. (1) is a linear system for the unknowns $y_k(t)$. More generally, one might have intercellular interactions, leading to density or other types of population-dependence. Then f_k could be a logistic term $\lambda_k(1 - y_k/K_k)$; or it could be a Lotka–Volterra sum, i.e., $f_k = [\lambda_k + \sum_{j=0}^{N} y_j a_{jk}]$ with a_{jk} real constants (Hofbauer and Sigmund, 1998); or it could be some more complicated function.

(2) Q_{jk} are real constants (independent of time and of subpopulation numbers) comprising the elements of the transition matrix $\mathbf{Q}$ of a continuous-time Markov process. That is, all off-diagonal elements of $\mathbf{Q}$ are nonnegative and $\sum_{k=0}^{N} Q_{jk} = 0$ for each j (Kijima, 1997).

The proliferation-alteration system, Eq. (1), is *time-homogeneous*, i.e., time does not appear explicitly. The intercellular interactions (if any) are contained in the dependence, if any, of f_k on the subpopulation numbers.

2.2. *Total population and subpopulation frequencies*

Instead of the dependent variables $y_j(t)$ we will usually use an equivalent set, consisting of the total population number, which we designate $P(t)$, and the subpopulation frequencies $x_k(t)$. Thus,

$$P(t) = \sum_{j=0}^{N} y_j(t) \quad \text{and} \quad x_k(t) = y_k(t)/P(t). \tag{2}$$

Each $x_k(t)$ is nonnegative and Eq. (2) implies that $\sum_{j=0}^{N} x_j(t) = 1$. Thus, we may regard the population frequencies as a discrete probability distribution: $x_k(t)$ is the probability that at time t a cell is in subpopulation k.

Substituting Eq. (2) into Eq. (1) gives, after a few manipulations, an equivalent equation set:

$$\begin{aligned} &\text{(a)} \quad dP/dt = \langle f \rangle P, \quad \text{and} \quad \text{(b)} \quad dx_k/dt = x_k\big[f_k - \langle f \rangle\big] + \sum_{j=0}^{N} x_j Q_{jk}; \\ &\text{where (c)} \quad \sum_{j=0}^{N} x_j(t) = 1, \quad \text{and} \quad \text{(d)} \quad \langle f \rangle \equiv \sum_{j=0}^{N} x_j f_j. \end{aligned} \tag{3}$$

Here, $\langle f \rangle$ is the population average proliferation rate at one instant; $f_k - \langle f \rangle$ is the instantaneous *relative fitness* of the kth subpopulation (Hofbauer and Sigmund, 1998). Summing the right-hand side of Eq. (3b) over k gives zero, as required for consistency with Eq. (3c).

Equation (3b) is more general than a selection-mutation equation suggested earlier (Hofbauer and Sigmund, 1998, Eq. 20.4) inasmuch as f_k can be any smooth function of the subpopulation numbers $y_j = Px_j$. Moreover, Eq. (1) and thus Eq. (3) differ from replicator-mutator equations in the literature (e.g. Page and Nowak, 2002) in two main ways:

(1) They involve subpopulation numbers, not just subpopulation frequencies.
(2) In Eq. (1) and Eq. (3b), the alteration terms, involving **Q**, are explicitly separated from the proliferation terms and are linear in the dependent variables. The idea behind assuming linearity of the alteration terms is that alteration may be less sensitive to signals from other cells than proliferation, which is very responsive to such signals because for normal cells regulation by the local microenvironment is the key factor in organized tissue growth.

In the present preliminary analysis, we will unless explicitly stated to the contrary, always assume that **Q** is an irreducible matrix; generalizations are outlined in the Discussion section. Under our above assumptions about **Q**, irreducibility of **Q** is equivalent to the condition that the matrix $\exp[\mathbf{Q}] \equiv \sum_{n=0}^{\infty}(\mathbf{Q}^n/n!)$ has no zero elements (Kijima, 1997), and this irreducibility also implies a number of other useful properties, listed under Eq. (4) below.

2.3. Vector notation

In analyzing **Q,** it will sometimes be convenient to use vector notation. Because **Q** is not in general symmetric, we need to distinguish systematically between row-space and column-space. We will use the following conventions, based on the conventions of our main reference on Markov processes (Kijima, 1997). Boldface lower case letters **x**, **y**, **z**, etc. denote column vectors with $N+1$ components, e.g., $\mathbf{y} = (y_0, y_1, \ldots, y_N)^{\mathrm{T}}$, where T denotes the transpose. The column vector each of whose components is 1 is denoted by **1**, so that our condition $\sum_{k=0}^{N} Q_{jk} = 0$ above can be written $\mathbf{Q1} = \mathbf{0}$. In this notation, $\mathbf{y}^{\mathrm{T}}\mathbf{z} = \mathbf{z}^{\mathrm{T}}\mathbf{y}$ is the scalar sum $\sum_{j=0}^{N} y_j z_j$; for example, our above relation between subpopulation frequencies $\mathbf{x}(t)$ and subpopulation numbers $\mathbf{y}(t)$ can be written $\mathbf{x} = \mathbf{y}/(\mathbf{1}^{\mathrm{T}}\mathbf{y})$. However, $\mathbf{y}\mathbf{z}^{\mathrm{T}}$ is the tensor product, i.e., it is a matrix, with matrix elements $y_j z_k$.

2.4. Zeroth-order solutions

Corresponding to our rapid-alteration assumption we consider, as our zeroth-order approximation, the equation obtained by neglecting the proliferation terms in Eq. (1) or equivalently in Eq. (3). Setting $f_k = 0$ in Eq. (3b) and writing the zeroth-order approximation to **x** as **w** gives the following zeroth-order equations

$$d\mathbf{w}^{\mathrm{T}}/dt = \mathbf{w}^{\mathrm{T}}\mathbf{Q}, \quad \text{where } \mathbf{w}^{\mathrm{T}}\mathbf{1} = 1. \tag{4}$$

Equation (4) means that our zeroth-order analysis is formally identical to analyzing a continuous time, time-homogeneous Markov process with infinitesimal generator **Q** (Kijima, 1997). Under our default assumption that **Q** is irreducible, standard results about nonnegative matrices and Markov processes (Kijima, 1997) show the following properties of Eq. (4), of **Q**, and of the eigenvectors of **Q**.

(1) Given the initial condition $\mathbf{w}(0) = \mathbf{a}$ with **a** a nonnegative vector obeying $\mathbf{a}^{\mathrm{T}}\mathbf{1} = 1$, there is a unique solution of Eq. (4) for $t \geq 0$, namely $\mathbf{w}^{\mathrm{T}}(t) = \mathbf{a}^{\mathrm{T}} \exp[\mathbf{Q}t]$. For all $t > 0$, all matrix elements of $\exp[\mathbf{Q}t]$ and all components of $\mathbf{w}(t)$ are greater than zero.

(2) There is a unique vector $\boldsymbol{\pi}$ characterized by the following two conditions: $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{Q} = 0$ and $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{1} = 1$. $\boldsymbol{\pi}^{\mathrm{T}}$ is called the Perron–Frobenius left eigenvector of **Q**. Importantly, every component of $\boldsymbol{\pi}$ is real and positive. The $(N+1)$-dimensional space of row vectors is the direct sum of the 1-dimensional subspace spanned by $\boldsymbol{\pi}^{\mathrm{T}}$ and the N dimensional subspace V characterized by: $\mathbf{z}^{\mathrm{T}}\mathbf{1} = 0$ for all $\mathbf{z}^{\mathrm{T}}$ in V. Right multiplication by **Q** is an isomorphism of V onto itself. Any left eigenvector of **Q** not proportional to $\boldsymbol{\pi}^{\mathrm{T}}$ lies in V, and the corresponding eigenvalue has a negative real part.

(3) $\boldsymbol{\pi}^{\mathrm{T}}$ is a time-independent solution of Eq. (4). Moreover, for any initial vector $\mathbf{w}(0) = \mathbf{a}$ with $\mathbf{a}^{\mathrm{T}}\mathbf{1} = 1$, $\lim_{t\to\infty} \mathbf{w}(t) = \boldsymbol{\pi}$. The intuitive picture corresponding to $\boldsymbol{\pi}^{\mathrm{T}}$ is that of a dynamic steady state: rapid alterations take place, but in steady state, the alterations that move cells into any subpopulation are balanced by alterations which remove cells from that same subpopulation, so that the frequency of each subpopulation remains constant, the constant being the corresponding component of $\boldsymbol{\pi}$. Note also an ergodicity property: even if there is initially just a single subpopulation, each subpopulation frequency eventually approaches the positive value given by the corresponding

component of $\boldsymbol{\pi}$, such that the additional alterations (which then still occur) do not subsequently alter the subpopulation frequencies. Often it is reasonable to use $\boldsymbol{\pi}$ as an approximation for the *initial* value of the subpopulation frequencies, corresponding to a scenario where the subpopulation frequencies have more or less settled down before the first experimental observation is made.

(4) For t large,

$$\boldsymbol{\pi} - \mathbf{w} = O\left(e^{-\rho t}\right) \quad \text{component by component.} \tag{5}$$

Here, ρ is positive and $-\rho$ is the largest (i.e., least negative) real part for any nonzero eigenvalue of $\mathbf{Q}$. ρ is called the *decay parameter*, and $e^{-\rho}$ corresponds to the coefficient of ergodicity of an ergodic discrete-time Markov chain; various rather sophisticated approximations are available for estimating the decay parameter ρ (Kijima, 1997).

2.5. First-order and exact solutions

First-order modulations of the solutions to the zeroth-order approximation discussed above were calculated using perturbation theory, as described in the next section. The steady state solution $\boldsymbol{\pi}$, other zeroth-order solutions, solutions of the first-order equations, and corresponding solutions of the full proliferation-alteration system, Eq. (1), were obtained numerically in a specific example; these zeroth and first-order approximations were compared to the full solutions. Throughout, Fortran and Mathematica were used in computational implementations.

3. First-order approximation

In accordance with our rapid-alteration assumption, we assume that to zeroth-order the subpopulation frequencies are given, apart from a possible brief initial transient not essential to the overall process, by the Perron–Frobenius eigenvector, $\mathbf{x} = \boldsymbol{\pi}$, independent of time. Then we implement a standard perturbation scheme as follows. Introduce a formal perturbation parameter ε, regarded intuitively as "small" but to be set equal to 1 at the end of the calculation; replace the proliferation rate f_k by εf_k everywhere in our basic proliferation-alteration system, Eq. (3), to make explicit the assumption that f_k is small; expand the vector $\mathbf{x}$ of subpopulation frequencies in a formal power series $\mathbf{x} = \boldsymbol{\pi} + \varepsilon\mathbf{b}(t) + \varepsilon^2\mathbf{c}(t) + \cdots$, where $\mathbf{b}^{\mathrm{T}}\mathbf{1} = 0 = \mathbf{c}^{\mathrm{T}}\mathbf{1}$, etc.; expand Eq. (3b) systematically as a formal power series in ε and require the coefficients of ε^n to match, to obtain an equation for the nth approximation; in Eq. (3a) retain only terms up through order ε^n in the nth approximation. The procedure is similar to the WKBJ approximation in quantum mechanics (Griffiths, 2004). We shall in this paper consider only the first-order modulation of the zeroth-order solutions. The intuitive picture is that comparatively slow (though cumulative) changes in total cell number $P(t)$ occur due to net proliferation, and slight deviations $\mathbf{b}(t)$ of subpopulation frequencies from the steady state frequencies given by $\boldsymbol{\pi}$ occur due to heterogeneous proliferation rates, but rapid alterations hold the subpopulation frequencies close to $\boldsymbol{\pi}$.

Carrying out the perturbation calculation outlined above gives for Eq. (3a) to first-order:

$$dP_1/dt = P_1\langle f\rangle_1, \quad \text{where } \langle f\rangle_1 = \sum_{k=0}^{N} \pi_k f_{k1} \text{ and}$$
$$f_{k1} = f_k(P_1\pi_0, P_1\pi_1, \ldots, P_1\pi_N). \tag{6}$$

Here, $\langle f\rangle_1$ is the average fitness to first-order and P_1 denotes the first-order approximation to total population number. The only unknown in this equation is P_1 (**b** has dropped out), so we regard the equation as determining $P_1(t)$ from its initial value $P_1(0)$. The intuitive picture for Eq. (6) is that when alterations are rapid the population, though diverse, tends to grow or decrease as a unit. For example, if one cell subpopulation has a comparatively large net proliferation rate, the result is not that the frequency of that subpopulation in the total population increases markedly. Rather, that subpopulation also loses many cells through alterations; the altered cells increase the cell numbers of other cell subpopulations, and thus the high fitness of one cell subpopulation contributes to the overall growth of the entire population rather than just the growth of the highly fit subpopulation. According to Eq. (6), when the population grows as a unit, its overall Malthusian parameter is $\langle f\rangle_1$ and is obtained as a weighted average of the subpopulation growth rates, the weights being the components of the Perron–Frobenius eigenvector $\boldsymbol{\pi}$.

For Eq. (3b), the calculation gives to first-order

$$db_k/dt = \pi_k\big[f_{k1} - \langle f\rangle_1\big] + \sum_{j=0}^{N} b_j Q_{jk} \equiv g_k + \sum_{j=0}^{N} b_j Q_{jk}. \tag{7}$$

In general, we must regard the vector **g**, with components $\pi_k[f_{k1} - \langle f\rangle_1]$, as a function of time, $\mathbf{g}(t)$, due to the possible dependence of f_k on the numbers in the various subpopulations which in turn depend on time through $P_1(t)$ (compare Eq. (6)). We regard Eq. (7) as a system of equations for the unknowns $b_k(t)$, $\boldsymbol{\pi}$ being known from the zeroth-order calculation and P_1 being known from solving Eq. (6). In general, Eq. (6) is nonlinear; however, the system (7) is linear in the unknowns, so we can obtain an explicit solution, namely

$$\mathbf{b}^{\mathrm{T}}(t) = \mathbf{b}^{\mathrm{T}}(0)\exp[\mathbf{Q}t] + \int_0^t ds\,\mathbf{g}^{\mathrm{T}}(s)\exp\big[\mathbf{Q}(t-s)\big]. \tag{8}$$

Equation (8) can be interpreted by standard methods for linear differential equations, in terms of influence operators (i.e., Green's functions), as follows: (a) the initial value $\mathbf{b}^{\mathrm{T}}(0)$ is carried forward to time t by the influence operator $\exp[\mathbf{Q}t]$ that tracks alterations; and (b), for the term in Eq. (8) involving the integral that same influence-operator carries forward from s to t the small increment $ds\,P(s)\mathbf{g}^{\mathrm{T}}(s)$ added to $\mathbf{b}^{\mathrm{T}}$ by proliferation during the short time interval ds near s.

A useful point is that the entire time development of **b** refers to an N-dimensional subspace V, characterized by $\mathbf{z}^{\mathrm{T}}\mathbf{1} = 0$ and discussed above, of $(N+1)$-dimensional row-space. V is invariant under right multiplication by **Q**, only the restriction of **Q** to this subspace is relevant to the time development of **b**, and the eigenvalues of that restriction

all have negative real parts. To make these points explicit, one can introduce the projection matrix $\mathbf{S} = \mathbf{I} - \mathbf{1}\boldsymbol{\pi}^{\mathrm{T}}$, where $\mathbf{I}$ is the identity matrix. Right multiplication of row vectors by $\mathbf{S}$ projects row-space onto V. Moreover, $\mathbf{S}^2 = \mathbf{S}$, $\mathbf{SQ} = \mathbf{QS} = \mathbf{Q}$, $\mathbf{b}^{\mathrm{T}}\mathbf{S} = \mathbf{b}^{\mathrm{T}}$, and $\mathbf{g}^{\mathrm{T}}\mathbf{S} = \mathbf{g}^{\mathrm{T}}$. Multiplying Eq. (8) on the right by $\mathbf{S}$ and using these identities gives an equation explicitly restricted to V. Thus, for $t > 0$, the influence operator $\exp[\mathbf{Q}t]$ acting on V should be regarded conceptually as having an overall damping term $\exp[-\rho t]$, where ρ is the decay constant for $\mathbf{Q}$ (see Eq. (5)); similarly the influence operator under the integral corresponds to decay at least as rapidly as $\exp[-\rho(t-s)]$ with $(t-s) \geq 0$. Intuitively speaking, Eq. (8) thus indicates an interplay between proliferation and alteration: when proliferation causes an incremental change in $\mathbf{b}$ the cell subpopulation frequencies are pulled away from $\boldsymbol{\pi}$ somewhat, but alterations are constantly tending to force the subpopulation frequencies back toward $\boldsymbol{\pi}$, i.e., tending to decrease $\mathbf{b}$ to zero.

4. Example

We illustrate the formalism with an example. The example, including some of the parameters chosen, was motivated by experiments in our laboratory which will be reported on elsewhere. Here, we will present a purely computational analysis, comparing solutions of our basic proliferation-alteration Eq. (1) with the zeroth-order solution $\boldsymbol{\pi}$ for subpopulation frequencies and with solutions of the first-order Eqs. (6) and (7). The results illustrate in a specific case how the accuracy of our zeroth- and first-order approximations depends on the rapidity of the alterations.

4.1. Mathematical formulation of the example

Mice have 21 chromosome types, namely autosomes 1–19, X, and Y. Consider a mouse cell population whose cells sometimes fuse with other cells, thereby increasing the chromosome number in the fused cell, and sometimes lose some of their chromosomes, resulting overall in heterogeneous aneuploidy. This aneuploidy situation in vitro is regarded as being a biological model for mouse and human solid cancers, which typically also show heterogeneous aneuploidy (Mitelman et al., 2007).

The number of chromosomes of each type in a given metaphase cell can be determined. Focus attention on one particular autosome type. For the sake of our example, assume viable cells have at most 8 chromosomes of this type, compared to the normal chromosome copy number of 2, and divide the cell population into nine subpopulations $y_j(t)$ $(j = 0, 1, \ldots, 8)$ according to the copy number in that cell. Suppose the proliferation-alteration equations (1) above hold with the proliferation terms $y_k f_k$ determined by a logistic form $f_k = [1 - (P/\zeta)]\mu_k$, where $P(t)$ is the total population number as before, ζ is a positive constant interpreted as an over-all carrying capacity, and each μ_k is a real constant giving the net proliferation rate for the kth subpopulation when $P/\zeta \ll 1$; thus, $\boldsymbol{\mu}$ tracks how copy number influences net cell proliferation. In this mathematical example, we do not attempt to track the influence of copy numbers for the other 20 chromosome types.

The alteration matrix $\mathbf{Q}$ for our example can be written in the form $\mathbf{Q} = \gamma\mathbf{R}$, where γ is a dimensionless overall scale factor and $\mathbf{R}$ is a matrix whose nonzero elements R_{kj} are given by Table 1.

Here, for $\gamma = 1$, the parameters have the following interpretations:

Table 1 The positive parameters α, β, and ψ are rate constants: α for the addition of two chromosomes by fusion; β and ψ for chromosome loss. For each k the kth row sum of **R** is zero, as required

k	$R_{k,k-1}$	$-R_{k,k}$	$R_{k,k+2}$
0	•	α	α
1–4	$k\beta$	$\alpha + k\beta$	α
5, 6	$k[\beta + (k-4)\psi]$	$\alpha + k[\beta + (k-4)\psi]$	α
7, 8	$k[\beta + (k-4)\psi]$	$k[\beta + (k-4)\psi]$	•

- α is a rate constant for a cell in the population to fuse with a cell of an external, ambient, normal cell population, presumed to have the same cell density throughout the entire process. The result of such a fusion is to increase the copy number by 2.
- β is the rate at which chromosomes copy number decreases by 1, e.g., due to a lagging chromosome entering a micronucleus (Thompson and Compton, 2008) and subsequently getting lost.
- ψ describes accelerated chromosome loss for copy number greater than 4.

To recapitulate, our example is defined as follows:

(1) $P(t) \equiv \sum_{j=0}^{8} y_j(t) \equiv \mathbf{y}^{\mathrm{T}}\mathbf{1}$ is the total cell population number.
(2) The parameters are the scale factor γ, the carrying capacity ζ, three aneusomy rate parameters (α, β, ψ), and the proliferation rate constants μ_k.
(3) Equation (1) determines the time dependence for the vector $\boldsymbol{y}(t)$ of subpopulation numbers, with $f_k = [1 - (P/\zeta)]\mu_k$, **Q** given by $\mathbf{Q} = \gamma\ \mathbf{R}$, and **R** given by Table 1.
(4) $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{Q} = 0$ and $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{1} = 1$ determine the zeroth-order approximation $\boldsymbol{\pi}$ to the subpopulation frequencies $\mathbf{x} \equiv \mathbf{y}/\mathbf{y}^{\mathrm{T}}\mathbf{1}$.
(5) Equation (6) gives the first-order approximation $P_1(t)$ to total population number.
(6) Equation (7), which is equivalent to Eq. (8), gives the first-order approximation, $\mathbf{b}(t) = \mathbf{x}(t) - \boldsymbol{\pi}$, to the deviation of subpopulation frequencies from $\boldsymbol{\pi}$.

The example has two convenient scaling properties. First, if every parameter except the scale factor γ and the carrying capacity ζ (i.e., every parameter with dimensions of inverse time) is multiplied by a positive factor and the time scale is multiplied by the same factor, the equations remain invariant. We can and shall take advantage of this scaling property to set $\alpha = 1$, without essential loss of generality; this convention is equivalent to using the dimensionless time $\tau = \alpha t$. Second, if **y** and the carrying capacity are multiplied by the same positive factor, the equations remain invariant. We can and shall take advantage of this second scaling property by setting the initial value of P to 1.

To illustrate the formalism, we used the following default values for the parameters: carrying capacity $\zeta = 10$ (i.e., ten times as large as the initial cell population size), proliferation rates $\boldsymbol{\mu} = \{-0.01, 0.03, 0.03, 0.2, 0.03, 0.03, 0.02, 0.01, -0.01\}^{\mathrm{T}}$, and aneusomy rates ($\alpha = 1, \beta = 0.2, \psi = 0.45$). With these aneusomy rates, the Perron–Frobenius eigenvector $\boldsymbol{\pi}$ has components $\boldsymbol{\pi} \approx \{0.0062, 0.031, 0.093, 0.21, 0.37, 0.18, 0.084, 0.024, 0.0052\}$, roughly equal to observed ploidy distributions. The values of $\boldsymbol{\mu}$ specify a hypothetical growth advantage (large relative fitness) for trisomy, and growth disadvantages especially for nullisomy or copy number higher than 5. Note that here the smallest nonzero entry of **R** equals $\max_k |\mu_k|$. With this condition, which involves no essential loss of generality, γ acquires an intuitive interpretation as a "rapidity parameter," a measure of how

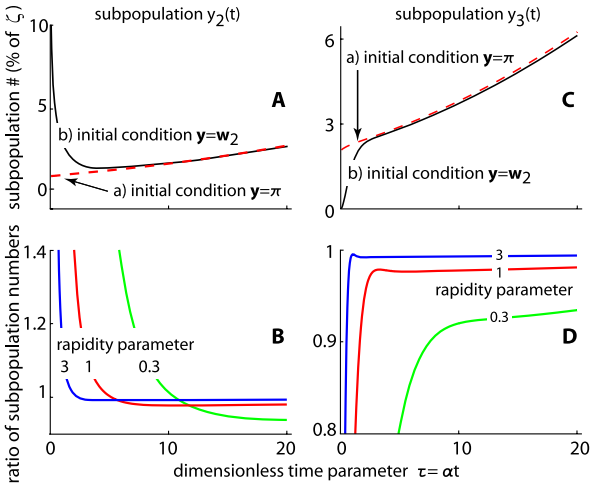

Fig. 1 Effects of transients on the solutions of Eq. (1). Panels A and C show the time dependence of subpopulation numbers, expressed as percentages of the overall carrying capacity ζ, for rapidity parameter $\gamma = 1$ and for the following two different initial conditions: (a) $\mathbf{y}(0)$ is the Perron–Frobenius eigenvector; or (b) $\mathbf{y}(0)$ is for a purely diploid population, i.e., $y_2(0) = 1$ and $y_k(0) = 0$ if $k \neq 2$. After the initial transient has decayed away, there is little difference in the curves. Panel A shows, as one example, the results for y_2 and panel C is for y_3 as one additional example. Panel B shows the ratio of the two curves in panel A, and shows corresponding ratios for other values of the rapidity parameter. Similarly panel D shows the ratio of the curves in panel C and corresponding ratios for the other values of γ. The transients decay more quickly, and the curves are thereafter more nearly identical, for larger rapidity parameters. The decay parameter for $\mathbf{R}$ is $\rho \approx 0.755$, i.e., -0.755 is the least negative real part for any nonzero eigenvalue of $\mathbf{R}$; since $\mathbf{Q} = \gamma\mathbf{R}$, $\gamma\rho$ governs the rate of approach of subpopulation frequencies to $\boldsymbol{\pi}$ (Eq. (5)). The effect of the initial conditions on all other components of $\mathbf{y}$, and on the total cell population number $P(t)$, is even smaller than in the two examples shown.

big the rate constants in the alteration matrix $\mathbf{Q}$ are compared to the rate constants for proliferation.

With these parameter values, we analyzed how the solutions of our equation depend on rapidity parameter γ and on time. Figure 1 examines transients; Figs. 2 and 3 give insights into the accuracy of the rapid alteration approximations.

4.2. Dependence of solution on initial conditions

In order to specify a unique solution of the differential equation system (1), one need only specify the subpopulation frequencies at time $t = 0$, the initial value of the total population P being set to 1 without essential loss of generality by our above conventions. One reasonable assumption is that the initial subpopulation frequencies are the components of $\boldsymbol{\pi}$, appropriate for a situation where transients have damped out before we first observe the population at $t = 0$. In that case the appropriate assumption for the initial value of the frequency deviation $\mathbf{b}$ is $\mathbf{b}(0) = 0$. For any other initial condition there are initial transients which, however, die out rapidly if the rapidity parameter γ is at least 1, as indicated in Fig. 1. The transients are not really relevant to our argument and henceforth we shall always use the initial condition $\mathbf{x}(0) = \boldsymbol{\pi}$.

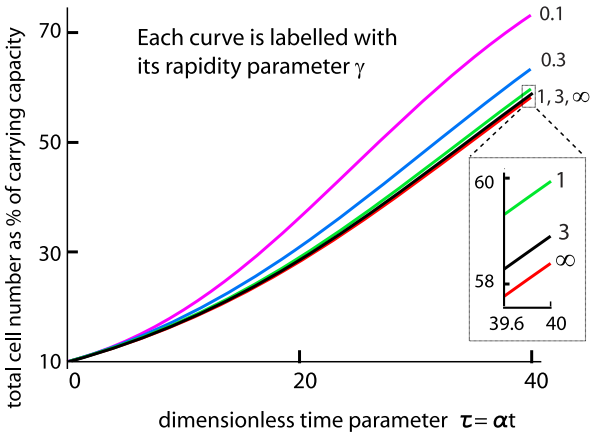

Fig. 2 Total cell number. The figure shows $P(t)$ obtained from numerical integration of the proliferation-alteration Eq. (1) using our default parameters described above and using values of the rapidity parameter given on the figure, from $\gamma = 0.1$ to large values (indicated by the limiting curve $\gamma = \infty$). The insert shows a magnified view of part of the curves for $\gamma = 1$, $\gamma = 3$, and $\gamma = \infty$. The first-order approximation $P_1(t)$ is independent of rapidity parameter and coincides with the $\gamma = \infty$ curve. For rapidity parameter ≥ 3, the first-order approximation is accurate within a few percent over the full range of times, as indicated by the figure.

4.3. First-order approximation

For the total cell number $P(t)$, the full result, without rapid-alteration approximations, is obtained by integrating Eq. (1) numerically and summing the components of $\mathbf{y}$. The first-order approximation, $P_1(t)$, is obtained by integrating Eq. (6). Equation (6) happens in this case to reduce to the logistic equation and thus have an explicit analytic solution, namely $P_1(t) = \zeta \exp[\nu t]/(\zeta + \exp[\nu t] - 1)$ where $\nu = \boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{\mu}$. Thus, $P_1(t)$ is independent of the rapidity parameter γ. As illustrated in Fig. 2, if $\gamma \geq 1$, then $P_1(t)$ is close to $P(t)$ for all times t; the difference goes to zero pointwise as $\gamma \to \infty$.

Figure 3 shows that a similar situation holds for the subpopulation frequencies. For rapidity parameter at least 1, the deviations from the zeroth-order values $\boldsymbol{\pi}$ are small, and in addition the first-order approximation to these small differences is highly accurate.

5. Discussion

The implication of our main result, Eqs. (6) and (7) is that during very rapid alteration phases cancer should be modeled as a diverse "ecological" system of interacting cells, rather than as one increasingly malignant clone. For example, suppose that one subpopulation proliferates rapidly, that another is slow growing but invasive, that the first can alter rapidly into the second, and that the two are synergistic. Then the two subpopulations grow in step, and it is their interrelation rather than the individual properties of either which is the key to carcinogenesis.

This scenario of interacting populations gives a mechanism for maintaining the clonal heterogeneity that is typical of late (progression) stages of solid tumors. Cancers and

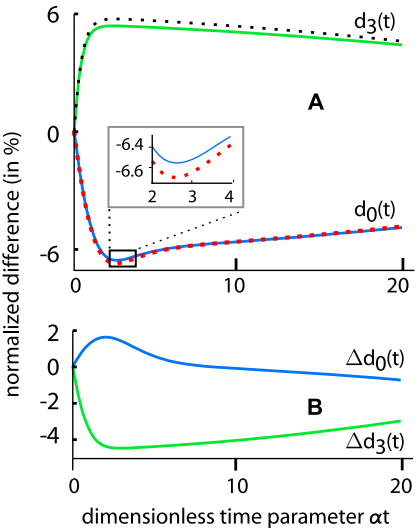

Fig. 3 Comparing exact and approximate calculations for individual subpopulations. The parameters used for the figure are the default parameters together with rapidity parameter $\gamma = 1$. Panel A shows components $d_k(t) = [x_k(t) - \pi_k]/\pi_k$, i.e., the differences of the subpopulation frequencies from those given by the Perron–Frobenius vector $\boldsymbol{\pi}$, normalized by dividing by the same component for $\boldsymbol{\pi}$. Values d_k obtained from solutions of the proliferation-alteration equation itself, Eq. (1) or equivalently Eq. (3), are shown as dotted lines; values d_{k1} obtained as first-order approximations from Eqs. (6) and (7) are shown as solid lines. The insert shows some details on the curves for d_0 and d_{01}. It is seen that even for subpopulation 3, which has the highest net proliferation rate, and subpopulation 0, which has the lowest (indeed negative) net proliferation rate, relative differences from the zeroth-order, proliferation-independent solutions are at most about 6%. Panel B shows values of $\Delta d_k \equiv (d_{k1} - d_k)/\sqrt{d_k^2 + (d_{k1})^2}$, a normalized difference of differences. It is seen from panel B that the difference between approximate and exact differences is only a few percent of the differences, which are themselves small (panel A); correspondingly our first-order approximation $x_0(t) \approx \pi_0 + d_{01}(t)$ approximates $x_0(t)$ to a relative accuracy of better than 0.003 for all values of t. For other components x_k or for rapidity parameter $\gamma > 1$ the accuracy is even higher.

in vitro cancer cell lines, are typically aneuploid, with highly heterogeneous populations, and with individual cell lineages observed to undergo rapid chromosome copy number alterations (Duesberg et al., 2005; Chi and Jeang, 2007). However, various authors have commented on the surprising temporal stability ("stability within instability," Gusev et al., 2001) of average total chromosome number, given the heterogeneity of the population (Nowell, 1976; Eshleman et al., 1998; Macville et al., 1999; Roschke et al., 2002; Jin et al., 2005). It could be that this stability within instability is the result of rapid alterations that determine a stable subpopulation frequency distribution even while the overall cell number is increasing.

Our formalism has various drawbacks. Like all present carcinogenesis formalisms, this one is unduly simplistic. Moreover, it is not envisaged that Eq. (1) with fixed parameters and fixed state space remains a useful approximation throughout cancer progression. Rather there should be periods where an equation of the form (1) is useful, interspersed with alterations to brand new states; after such an alteration a different set of parameters and a larger state space would be required. Perhaps such a pattern could be approximated

by using a $\mathbf{Q}$ in Eq. (1) corresponding to a reducible Markov process, or a leaky one, or by using an infinite state space.

Formally, it is clear that the basic proliferation-alteration system can in fact be generalized considerably to try to overcome some drawbacks without vitiating the rapid-alteration approximation approach. Here are the simplest examples.

- If state space is denumerably infinite, but $\mathbf{Q}$ is still ergodic, there is still a unique vector $\boldsymbol{\pi}$ obeying $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{Q} = 0$ and $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{1} = 1$, and each component of $\boldsymbol{\pi}$ is positive (Kijima, 1997). Then the entire formalism goes through as before provided appropriate restrictions are placed on f_k for k large, e.g., that the sum $\langle f \rangle_1 = \sum_{k=0}^{\infty} \pi_k f_{k1}$ converges.
- Suppose state space is the direct sum of a finite number of subspaces, with each subspace invariant under right multiplication by $\mathbf{Q}$ and the restriction of $\mathbf{Q}$ to each ergodic. Then with appropriate, essentially bookkeeping, modifications, the entire formalism remains applicable. This direct sum construct is relevant, for example, if we consider aneusomy of all 21 mouse chromosome types, with the alterations of each type governed by different parameters but cell fitness governed jointly by all 21 copy number values.
- Suppose $\sum_{j=0}^{N} y_j Q_{jk}$ in Eq. (1) is replaced by a nonlinear function $Q_k(Px_0, Px_1, \ldots, Px_N)$ and suppose that for any fixed positive constant P the equations $dx_j/dt = Q_k$ have a unique asymptotically stable solution $\boldsymbol{\pi}(P)$, depending parametrically on P, that obeys $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{1} = 1$ and has nonnegative components. Then the rapid-alteration approximation remains applicable in essentially the same form as analyzed here.

To summarize, we have here presented a population dynamical approach for analyzing an interacting, diverse, proliferating, rapidly altering somatic cell population. The approach is plausibly applicable to restricted time intervals during some cancer progression processes and can readily be generalized somewhat. It predicts that the overall growth rate is an emergent parameter of the entire cell system, given by an appropriate average.

Acknowledgement

Research supported by DOE in the Low Dose Radiation Research Program (LH) and NASA NSCOR04-0014-0017 NNJ04HJ12G/NNJ06HA28G (RKS).

References

Bjerkvig, R., Tysnes, B.B., Aboody, K.S., Najbauer, J., Terzis, A.J., 2005. Opinion: the origin of the cancer stem cell: current controversies and new insights. Nat. Rev. Cancer 5, 899–904.

Born, M., Wolf, E., Bhatia, A.B., 1999. Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. Cambridge University Press, Cambridge.

Chi, Y.H., Jeang, K.T., 2007. Aneuploidy and cancer. J. Cell. Biochem. 102, 531–8.

Chin, L., Gray, J.W., 2008. Translating insights from the cancer genome into clinical practice. Nature 452, 553–63.

Coller, H.A., Khrapko, K., Bodyak, N.D., Nekhaeva, E., Herrero-Jimenez, P., Thilly, W.G., 2001. High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. Nat. Genet. 28, 147–50.

Duesberg, P., Li, R., Fabarius, A., Hehlmann, R., 2005. The chromosomal basis of cancer. Cell. Oncol. 27, 293–318.

Duesberg, P., Li, R., Sachs, R., Fabarius, A., Upender, M.B., Hehlmann, R., 2007. Cancer drug resistance: The central role of the karyotype. Drug Resist. Updat.

Eshleman, J.R., Casey, G., Kochera, M.E., Sedwick, W.D., Swinler, S.E., Veigl, M.L., Willson, J.K., Schwartz, S., Markowitz, S.D., 1998. Chromosome number and structure both are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53. Oncogene 17, 719–25.

Frank, S.A., Nowak, M.A., 2004. Problems of somatic mutation and cancer. Bioessays 26, 291–9.

Gatenby, R.A., Vincent, T.L., 2003. An evolutionary model of carcinogenesis. Cancer Res. 63, 6212–20.

Griffiths, D.J., 2004. Introduction to Quantum Mechanics. Pearson/Prentice Hall, Upper Saddle River.

Gusev, Y., Kagansky, V., Dooley, W.C., 2001. Long-term dynamics of chromosomal instability in cancer: a transition probability model. Math. Comput. Model. 33, 1253–1273.

Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge.

Iwasa, Y., Michor, F., Nowak, M.A., 2004. Stochastic tunnels in evolutionary dynamics. Genetics 166, 1571–9.

Jin, Y., Jin, C., Lv, M., Tsao, S.W., Zhu, J., Wennerberg, J., Mertens, F., Kwong, Y.L., 2005. Karyotypic evolution and tumor progression in head and neck squamous cell carcinomas. Cancer Genet. Cytogenet. 156, 1–7.

Kijima, M., 1997. Markov Processes for Stochastic Modeling. Cambridge University Press, Cambridge.

Komarova, N.L., 2005. Mathematical modeling of tumorigenesis: mission possible. Curr. Opin. Oncol. 17, 39–43.

Li, L., McCormack, A.A., Nicholson, J.M., Fabarius, A., Hehlmann, R., Sachs, R.K., Duesberg, P.H., 2009. Cancer-causing karyotypes: chromosomal equilibria between destabilizing aneuploidy and stabilizing selection for oncogenic function. Cancer Genet. Cytogenet. 188, 1–25.

Little, M.P., Vineis, P., Li, G., 2008. A stochastic carcinogenesis model incorporating multiple types of genomic instability fitted to colon cancer data. J. Theor. Biol. 254, 229–38.

Macville, M., Schrock, E., Padilla-Nash, H., Keck, C., Ghadimi, B.M., Zimonjic, D., Popescu, N., Ried, T., 1999. Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. Cancer Res. 59, 141–50.

Marx, C., Posch, H.A., Thirring, W., 2007. Emergence of order in selection-mutation dynamics. Phys. Rev. E 75, 061109.

Merlo, L.M., Pepper, J.W., Reid, B.J., Maley, C.C., 2006. Cancer as an evolutionary and ecological process. Nat. Rev. Cancer 6, 924–35.

Michor, F., Frank, S.A., May, R.M., Iwasa, Y., Nowak, M.A., 2003. Somatic selection for and against cancer. J. Theor. Biol. 225, 377–82.

Michor, F., Hughes, T.P., Iwasa, Y., Branford, S., Shah, N.P., Sawyers, C.L., Nowak, M.A., 2005. Dynamics of chronic myeloid leukaemia. Nature 435, 1267–70.

Minc, H., 1988. Nonnegative Matrices. Wiley, New York.

Mitelman, F., Johansson, B., Mertens, F., 2007. The impact of translocations and gene fusions on cancer causation. Nat. Rev. Cancer 7, 233–45.

Moolgavkar, S.H., Luebeck, E.G., 2003. Multistage carcinogenesis and the incidence of human cancer. Genes Chromosomes Cancer 38, 302–6.

Nagy, J.D., 2004. Competition and natural selection in a mathematical model of cancer. Bull. Math. Biol. 66, 663–87.

Nowell, P.C., 1976. The clonal evolution of tumor cell populations. Science 194, 23–8.

Page, K.M., Nowak, M.A., 2002. Unifying evolutionary dynamics. J. Theor. Biol. 219, 93–8.

Roschke, A.V., Stover, K., Tonon, G., Schaffer, A.A., Kirsch, I.R., 2002. Stable karyotypes in epithelial cancer cell lines despite high rates of ongoing structural and numerical chromosomal instability. Neoplasia 4, 19–31.

Sachs, R.K., Chan, M., Hlatky, L., Hahnfeldt, P., 2005. Modeling intercellular interactions during carcinogenesis. Radiat. Res. 164, 324–31.

Sachs, R.K., Shuryak, I., Brenner, D., Fakir, H., Hlatky, L., Hahnfeldt, P., 2007. Second cancers after fractionated radiotherapy: Stochastic population dynamics effects. J. Theor. Biol. 249, 518–531.

Tan, W., 2002. Stochastic Models with Applications to Genetics, Cancers, AIDS, and Other Biomedical Systems. World Scientific, Singapore.

Thompson, S.L., Compton, D.A., 2008. Examining the link between chromosomal instability and aneuploidy in human cells. J. Cell. Biol. 180, 665–72.
Vucic, E.A., Brown, C.J., Lam, W.L., 2008. Epigenetics of cancer progression. Pharmacogenomics 9, 215–34.
Weinberg, R.A., 2007. The Biology of Cancer. Garland Science, New York. Chaps. 13 and 15.
Yuen, K.W., Desai, A., 2008. The wages of CIN. J. Cell. Biol. 180, 661–3.