J. theor. Biol. (2002) **214,** 453–467
doi:10.1006/jtbi.2001.2471, available online at http://www.idealibrary.com on IDEAL®

The Different Effects of Apoptosis and DNA Repair on Tumorigenesis

JOSHUA B. PLOTKIN* AND MARTIN A. NOWAK

Institute for Advanced Study, Princeton, NJ 08540, *U.S.A.*

(*Received on* 14 *May* 2001, *Accepted in revised form on* 5 *October* 2001)

Complex multicellular organisms have evolved mechanisms to ensure that individual cells follow their proper developmental and somatic programs. Tumorigenesis, or uncontrolled cellular proliferation, is caused by somatic mutations to those genetic constraints that normally operate within a tissue. Genes involved in DNA repair and apoptosis are particularly instrumental in safeguarding cells against tumorigenesis. In this paper, we introduce a stochastic framework to analyse the somatic evolution of cancer initiation. Within this model, we study how apoptosis and DNA repair can maintain the transient stability of somatic cells and delay the onset of cancer. Focusing on individual cell lineages, we calculate the waiting time before tumorigenesis in the presence of varying degrees of apoptosis and DNA repair. We find that the loss of DNA repair or the loss of apoptosis both hasten tumorigenesis, but in characteristically different ways.

1. Introduction

Tumorigenesis is defined as the onset of unregulated cell proliferation. In humans and many other mammals, the process towards accelerated cellular growth is marked by the loss of important regulation genes which usually control cell-cycle functioning. These regulation mechanisms may be broadly characterized (Kinzler & Vogelstein, 1997; Vogelstein *et al*., 2000) as DNA repair genes, which repair mutations and DNA damage before further cell division, and tumor suppressor genes, which signal for cell-cycle arrest and induce apoptosis if substantial genomic damage is detected (Gottlieb & Oren, 1998).

The loss of DNA mismatch or DNA excision repair genes increases the effective mutation rate per cell division (Orr-Weaver & Weinberg, 1998). This phenomenon is often called the mutator phenotype. The loss of tumor suppressor genes, on the other hand, diminishes a cell's ability to recognize damage and induce apoptosis. Both increased mutation rates and deficient apoptotic fidelity accelerate tumorigenesis. But the relative importance of these two carcinogenic mechanisms is, *a priori*, unclear. The extent to which the mutator phenotype determines the timing of tumorigenesis is hotly contested in the literature. Some scientists (Loeb, 1991) have argued that an increased pre-malignant mutation rate is required for tumorigenesis to occur whatsoever. Many other scientists agree that the mutator phenotype plays an important, if not absolutely necessary, role in cancer development (Lengauer *et al*., 1998; Orr-Weaver & Weinberg, 1998; Murdoch & VanKirk, 1997). But others (Tomlinson *et al*., 1996; Tomlinson & Bodmer, 1999) contend that Darwinian selection on cellular proliferation rates dominates the process towards

* Author to whom correspondence should be addressed. E-mail: plotkin@ias.edu

0022–5193/02/030453 + 15 $35.00/0

carcinogenesis, superseding any effects of the mutator phenotype. Still others stress the importance of apoptosis in preventing tumorigenesis (Tomlinson & Bodmer, 1995; Hong *et al*., 2000; Harnois *et al*., 1997; Chang *et al*., 1998). Others yet contend that the immune system plays a central role in controlling tumors (Darnell, 1999).

The extent to which each of these factors—decreased apoptosis, increased mutation, increased proliferation, etc.—contribute towards cancer depends, no doubt, on the cancer type. Given the diversity of cancer types, it would be misguiding to debate *the* crucial carcinogenic processes. Nevertheless, a rigorous, qualitative understanding of the differences between these carcinogenic processes can inform the debate about their relative importance in various cancer settings.

In this paper, we develop a mathematical framework for investigating the effects of cell-cycle regulation genes. In particular, we use stochastic multistage models to investigate as to how DNA error repair and apoptosis stave off tumorigenesis. Our generic framework can be used to investigate qualitative patterns in the progression towards carcinogenesis. We do not initially specify a particular cell type or a particular cancer type. Some of our assumptions, however, constrain the applicability of our model to particular classes of cancer types. We do not address the progression of a malignant tumor through its various cancerous stages, angiogenesis, and eventually metastasis. Instead, we investigate the accumulation of mutations in a pre-malignant cell lineage.

Stochastic modelling of tumorigenesis was introduced in the 1950s (Armitage & Doll, 1954) for comparison with adult age-specific cancer incidence rates. Given the differences between spontaneous and inherited cancers, however, authors soon began to develop multiple stage models of tumorigenesis, starting with two-stage models (Knudson, 1971). Truly rigorous and complete analyses of the two-stage stochastic models soon followed (Moolgavkar & Venzon, 1979; Moolgavkar & Knudson, 1981), treating childhood and adult tumors separately. Multistage models have since been expanded (e.g. Mao *et al*., 1998). In all such models, mutation rates were generally assumed to be constant, and apoptosis and DNA repair were largely ignored. Outside of the general multistage framework, other authors have investigated the effects of apoptosis (Tomlinson & Bodmer, 1995) and mutation rates (Tomlinson *et al*., 1996). In addition to research into the genetic events that cause tumorigenesis, there is a large and detailed literature which models the physiological processes involved in tumor invasion (Chaplain, 1995; Perumpanani *et al*., 1996), growth (Byrne & Chaplain, 1996a, b), encapsulation (Sherratt, 2000), macrophage dynamics (Owen & Sherratt, 1999), and angiogenesis (Chaplain, 2000).

This paper is divided into nine sections. In Section 2, we discuss parameter values for mutation rates and approximations appropriate to modelling mutation. Section 3 presents our basic model of tumorigenesis. In Section 4, we analyse the simple case when apoptosis and increased mutation rates are both neglected. In Section 5, we compute the average waiting time before tumorigenesis in terms of the apoptotic rates and increased mutation rates. In Section 6, we compute the distribution of waiting times before tumorigenesis. Section 7 compares the effect of increased mutation with that of deficient apoptosis. Section 8 addresses intrinsic costs associated with elevated mutation rates. Concluding remarks are given in Section 9. The main text refers to Appendices A–C for mathematical details.

2. Target Genes and Mutation

Throughout our analysis, we assume that there are L genes involved, in some way, in regulating normal cell-cycle functioning. We call these L sites target genes because their removal can increase the chance of tumor initiation. For humans, L is rather large. As a rough approximation, given that there are over 150 genes involved in apoptosis alone (Aravind *et al*., 2001) and over 130 involved in DNA repair (Wood *et al*., 2001), we assume that $L \approx 500$. We imagine that the target genes—DNA repair genes and tumor suppressor genes—form a large network whose redundancy buffers the cell against tumorigenesis. We will assume that if any n of these target genes become defective in a cell, then the cell will start to proliferate causing the onset of cancer.

(In an alternative model, we can define tumor initiation as the mutation of a *particular* set of n genes from the L regulation genes.) The threshold n is usually small compared to the total number of target genes. Evidence suggests that as few as $n = 2$ defective genes can cause unregulated cell proliferation for certain cancer types, such as retinoblastoma (Knudson, 1971), while $n = 6$ or more are required for other cancer types (Loeb, 1991).

In reality, even when a single cell acquires n mutations, the immune system may yet prevent tumorigenesis by targeting the deviant cell (Darnell, 1999). The frequency and importance of this phenomenon, however, are hotly contested. We will therefore assume throughout that tumorigenesis is simply defined by n mutations, delaying our discussion of immune action until Section 9.

We assume that at every generation of cell division, each of the L regulation genes may undergo a debilitating mutation with probability μ. Once a gene is mutated, we can safely assume that it will never again back-mutate into a functional gene. Although the per-*base* mutation rate is well known for many organisms (Drake *et al.*, 1998), the per-*gene* mutation rate μ in non-germline cells is more difficult to measure in practice. Several authors have suggested that $\mu \approx 10^{-7}$ per cell division (Orr-Weaver & Weinberg, 1998), although the precision of this value is unclear.

When apoptosis and DNA repair do not occur, the stochastic process of accumulating cellular mutations is relatively simple. Consider the case of tumorigenesis defined by any n mutations. Each cell in a tissue of constant size is, on average, replicating and being replaced by its daughter cell at each generation. A cell is characterized by the number of mutant genes k, $0 \leqslant k \leqslant L$, it currently harbors. We say that a cell lineage becomes cancerous (i.e. has lost its regulation ability), if $k \geqslant n$. In the most simple case, the cell has a probability p of increasing its number of mutations by one, and probability $q = 1 - p$ of maintaining the same number of mutations at each replication/replacement event. We call this situation "neutral" because the cell behaves in the same manner regardless of its current mutational status.

Starting from a tissue containing N healthy cells, we will compute the expected time before at least one of the cells acquires any n mutations. We will also compute this time in the presence of apoptosis and repair genes. Before we analyse the waiting times before tumorigenesis, we make several preliminary remarks about modelling mutation. We are assuming an extremely simple mutational process. Regardless of the current number of mutations, k, we assume that in each somatic generation k either increases by one, with probability p, or remains constant, with probability $q = 1 - p$.

Even when we choose to ignore back mutation, this formulation is only approximately correct. The exact formulation, according to an independent forward mutation rate μ and backward mutation rate zero per gene, is given by

$$\mathbb{P}(k \to k') = \begin{cases} 0 & \text{for } k' < k, \\ \binom{L-k}{k'-k} \mu^{k'-k}(1-\mu)^{L-k'} & \text{for } k' \geqslant k. \end{cases} \quad (1)$$

To be exact, then, $\mathbb{P}(k \to k) = (1 - \mu)^{L-k}$. We approximate this exact equation by defining $q = 1 - \mu L \approx \mathbb{P}(k \to k)$ and $p = 1 - q$. This approximation is valid provided that $n \ll L$ and $L\mu \ll 1$, both of which are true for humans (Drake *et al.*, 1998).

3. A Model of Mutation, Apoptosis, and DNA Repair

We now formulate a stochastic, Markov-chain model of the mutational and apoptotic process. We imagine a large tissue of cells which, at all times preceding the onset of cancer, is almost completely mutation free. We keep track of the current number of mutations k harbored in a cell lineage. In our Markov model, any particular cell is replaced by its daughter cell in each generation. At each generation, the healthy cell ($k = 0$) either remains healthy, with probability q_0, or accumulates a single mutated gene, with probability $p_0 = 1 - q_0$.

A cell harboring $k \geqslant 1$ mutations, however, is subject to tumor-suppressor-induced apoptosis. In particular, at each generation, a cell harboring k mutated genes has probability α_k of destroying

itself and being replaced by another cell in the tissue, which we assume to be healthy. With probability $\beta_k = 1 - \alpha_k$, on the other hand, the mutated cell replicates—maintaining the same number of mutations with probability q_k, and acquiring an additional mutation with probability p_k. According to this formulation, a cell lineage is described by a Markov chain with the following $(n+1)\times(n+1)$ transition matrix:

$$P = \begin{pmatrix} q_0 & p_0 & 0 & \cdots & \cdots & 0 \\ \alpha_1 & \beta_1 q_1 & \beta_1 p_1 & 0 & \cdots & 0 \\ \alpha_2 & 0 & \beta_2 q_2 & \beta_2 p_2 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ \alpha_{n-1} & 0 & \cdots & 0 & \beta_{n-1} q_{n-1} & \beta_{n-1} p_{n-1} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}. \quad (2)$$

The state $k = n$, which we call tumorigenesis, is absorbing. Once the cell reaches this state, we say that the cell has escaped from proper cell regulation and begins malignant growth. We do not concern ourselves with the progression of cancer after this stage. Although active tumors certainly undergo mutation and apoptosis, the evolution of a tumor is a very different phase of cancer progression which we do not model here.

When apoptosis is ignored (i.e. all α_k are zero), then the formulation given by eqn (2) is equivalent (in discrete time) to the classical models of tumorigenesis formulated by Armitage & Doll (1954) and later developed by Moolgavkar (1978). Here we extend these models to include the action of apoptosis.

We will examine tumorigenesis for arbitrary apoptotic rates α_k. The simplest case occurs when α_k is constant. If we focus our attention on target genes which are related to apoptosis, then the apoptotic rates α_k should naturally be decreasing with k. This corresponds to the situation in which as tumor-suppressor genes become mutated, their efficiency at scanning for damage and inducing apoptosis is reduced. Hence, mutated cells gain a selective advantage by avoiding programmed cell death (Tomlinson & Bodmer, 1995). In this formulation, the road towards tumorigenesis ($k = n$) is a slippery slope; once a few suppressor genes have been damaged, it becomes increasingly hard for a cell to recognize mutations, and thereby to prevent further progression towards cancer (Kinzler & Vogelstein, 1997). On the other hand, we can also consider cases when the apoptotic rates α_k increase. This models the situation in which, say, progressively more oncogenes become mutated and signal for increased apoptosis.

We are primarily interested in the cases when α_k decreases, corresponding to the loss of tumor-suppressors and apoptotic fidelity. Similarly, we generally consider cases in which the effective mutation rates p_k are increasing. This assumption mirrors the progressive loss of DNA repair genes and the corresponding increase in effective mutation rate and genomic instability (Vogelstein *et al.*, 2000).

We are nominally modelling the behavior of a single-cell lineage. But whenever a cell undergoes apoptosis (with probability α_k), it is replaced by a cell from the surrounding tissue which we assume to be healthy. Thus, we are, in fact, modelling the behavior of a cell in a large tissue composed predominantly of healthy cells. Therefore, our cell "lineage" is not a strict lineage, but rather includes replacement by healthy cells upon apoptosis. Before the onset of cancer, we assume that almost all cells in the tissue are mutation free. This assumption, used here to simplify population dynamics, may be invalid for some cancer types. (Elsewhere we relax this assumption, and allow for apoptosis followed by replacement of nearby unhealthy cells, Nowak & Plotkin, in preparation) Colorectal cancer, for example, is often preceded by the invasion of numerous benign polyps which exhibit one or two mutated genes, often APC, throughout the tissue (Kinzler & Vogelstein, 1996). Other cancers—such as retinoblastoma—are not generally preceded by widespread polyps containing mutations, and are therefore more appropriately modelled by our formulation.

Our Markov model does not address differences in proliferative rates between cell lineages. Elsewhere (Nowak & Plotkin, in preparation), we develop deterministic models which treat inter-lineage population dynamics. Again, the one-dimensional Markov model is therefore less

appropriate for colorectal type cancers—which exhibit clonal expansion and competition through increased proliferation—than it is for retinoblastoma-type cancers (Cairns, 1998).

The Markov model keeps track of the current number of defective regulation genes, k. At each stage we do not know *which* k of the L target genes are mutated. In this sense, the reduced apoptotic fidelities α_k and the increased effective mutation rates p_k represent averages over all possible combinations of k mutated genes. In other words, the loss of some tumor suppressors, e.g. p53, can in reality be more deleterious than the loss of other ones; we average over these possibilities.

4. The Neutral Case

Before we analyse the full-blown Markov model, we consider the simple, neutral case in which apoptosis and DNA repair are both ignored. In other words, we assume for now that all α_k are zero and all p_k are equal. In Appendix A, we derive the expected number of cell divisions, T, required before at least one cell in a tissue of N cells, originally all healthy, acquires n mutations:

$$T = \frac{1}{\log(1/q)(n-1)!^N} \int_0^\infty \Gamma(n, a)^N \, \mathrm{d}a. \qquad (3)$$

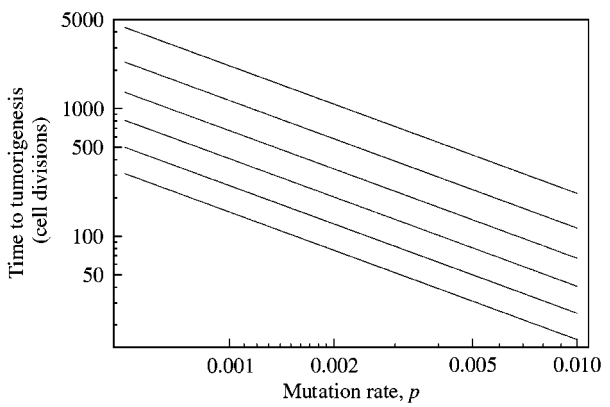

FIG. 1. The expected number of cell divisions before one cell in a tissue of size N acquires $n = 5$ mutations, ignoring apoptosis and DNA repair. The mutation rate p varies from $p = 500 \times 10^{-6}$ to 0.01 on the x-axis. Each curve represents a different tissue size ranging geometrically from $N = 10^1$, highest curve, to $N = 10^6$, lowest curve.

Here $\Gamma(\cdot, \cdot)$ denotes the incomplete Gamma function. Equation (3) demonstrates that in the neutral case the waiting time before tumorigenesis depends inversely on the logarithm of one replication fidelity (q). Figure 1 demonstrates the expected neutral waiting time before tumorigenesis for various mutation rates and tissue sizes. As is clear from the figure, increasing the tissue size N geometrically causes the waiting time T to decrease geometrically.

5. The Average Waiting Time before Tumorigenesis

Given arbitrary mutation rates p_k and apoptotic rates α_k, what is the mean time before carcinogenesis (state $k = n$)? In addition, what is the expected total time spent by a cell lineage in each of the various mutated classes? For example, do we expect to find cells with $k = 2$ mutations half as often as $k = 1$? From this section onwards, we answer these questions insofar as a single-cell lineage is concerned—an approach originally espoused by Armitage & Doll (1954) and later Moolgavkar (1978).

Define T_i as the expected number of generations before absorption into the cancerous state, assuming that the cell begins in state i at time zero. We are eventually interested in T_0, the mean time until tumorigenesis starting from a healthy cell. We might also be interested in T_1, if an individual inherits a defective gene. Define W_{ij} as the mean number of visits to state j prior to absorption, assuming that the cell starts in state i. As is clear from the Markovian property of our process, we have the following recursive relationships:

$$W_{ij} = \delta_{ij} + \sum_{m=0}^{n-1} P_{im} W_{mj} \quad \text{for } i = 0, \ldots, n-1, \qquad (4)$$

In Appendix B, we use this recursion to compute the mean time spent in each mutation class prior to tumorigenesis, starting from a healthy cell:

$$W_{0,k} = \frac{1}{p_{n-1}\beta_{n-1}} \prod_{x=k+1}^{n-1} \frac{p_x\beta_x + \alpha_x}{p_{x-1}\beta_{x-1}}$$

$$\text{for } 0 \leqslant k \leqslant n-1, \qquad (5)$$

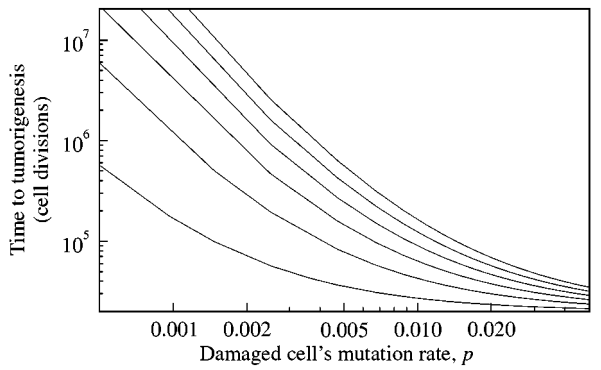

FIG. 2. A graph of the expected number of cell divisions before tumorigenesis ($n = 4$) in a cell lineage as a function of the damaged cell's mutation rate p. We assume a constant apoptotic rate [Eqn (9)]. The healthy cell's mutation rate is $p_0 = 500 \times 10^{-6}$. The damaged cell's mutation rate ranges 100-fold on the x-axis from $p = p_0$ to 0.05. Each of the six curves represents a different apoptotic rate, ranging linearly from $\alpha = 0.001$ (lowest curve) to $\alpha = 0.01$ (highest curve). Note that the time until tumorigenesis is less sensitive to the apoptotic fidelity whenever the damaged cell's mutation rate is high.

$$T_0 = \sum_{k=0}^{n-1} W_{0,k}, \tag{6}$$

where we have defined $\alpha_0 = 0$ for convenience. Note that a cell spends $(p_k\beta_k + \alpha_k)/(p_{k+1}\beta_{k+1})$ as much time with k mutations as it does with $k+1$ mutations, before tumorigenesis. In addition, the mean time a cell spends in mutational class k depends only upon the effective mutation rates and apoptotic rates of the classes $k' \geqslant k$. In other words, the selective pressure on k-th-order checkpoint genes does not depend upon the behavior of cells with fewer than k mutations.

In the case when all the apoptotic rates are constant, $\alpha_k = \alpha$, and when all the damaged mutation rates are constant ($p_1 = p_2 = \cdots = p_{n-1} = p$), eqns (5) and (6) simplify to the following:

$$W_{0,0} = \frac{1}{p_0}\left(\frac{p\beta + \alpha}{p\beta}\right)^{n-1}, \tag{7}$$

$$W_{0,k} = \frac{1}{p\beta}\left(\frac{p\beta + \alpha}{p\beta}\right)^{n-k-1} \quad \text{for } 1 \leqslant k \leqslant n-1, \tag{8}$$

$$T_0 = \frac{p\beta(\alpha + p_0)}{\alpha p_0(\alpha q + p)(1 - \alpha/(p + \alpha q))^n} - \frac{1}{\alpha}. \tag{9}$$

Assuming that apoptotic rates are constant, Fig. 2 shows as to how the timing of tumorigenesis depends upon the damaged cell's mutation rate. The loss of DNA repair genes in damaged cells can increase the effective mutation rate by a factor ranging from 10 to 10^4 (Tomlinson *et al.*, 1996). Note that the beneficial effect of apoptosis is minimized as the damaged cell's mutation rate increases. In other words, if there is a strong mutator phenotype, then the time before tumorigenesis is less significantly affected by the presence of functioning apoptotic machinery.

6. The Distribution of Waiting Times before Tumorigenesis

Equation (6) describes the expected time before tumorigenesis in a cell lineage, given the apoptotic and mutation rates. We are further interested in the complete distribution of waiting times until tumorigenesis. For instance, does the mutator phenotype increase the variance of the waiting time? In this section, we let T_0 denote the random time at which the Markov chain reaches the cancerous state, n. We will find expressions for the expected value $\mathbb{E}T_0$ and for the distribution function of T_0, in terms of the eigenvalues of P. As we will discuss below, the distribution of T_0 can be used to predict cancer age–incidence curves comparable to epidemiological census data.

Consider any $(n+1)$-state Markov chain $\hat{M}$ acting on the left. (In our case, take $\hat{M}$ to be the transpose of P.) Label the states 0 to n, and assume that state n is the only absorbing state. Consider the $n \times n$ submatrix M which excludes the absorbing state. Let $V_0 = (1, 0, \ldots, 0)^{\mathrm{T}}$ denote the initial condition: in state zero. Note that i-th coordinate of $M^t V_0$, denoted $[M^t V_0]_i$, is the probability that the chain is in state i at time t. We index the coordinates from 0 to $n-1$. Hence the following expression:

$$\sum_{i=0}^{n-1} [M^t V_0]_i = \|M^t V_0\|,$$

yields the probability that the cell is not yet in the absorbing state at time t. In other words, this sum is the complementary cumulative distribution

function of T_0. We use $\|\cdot\|$ to denote the sum of the components of a vector.

We can write $M = SDS^{-1}$ where D is the diagonal matrix of eigenvalues, $\lambda_0, \ldots, \lambda_{n-1}$, provided that they are all distinct. The columns of S are the corresponding eigenvectors. In this case, $M^t V_0 = SD^t S^{-1} V_0$. Hence, the cumulative distribution of T_0 is given by $\Phi(t) = 1 - \|SD^t S^{-1} V_0\|$. This in turn allows us to compute the expected time before tumorigenesis:

$$\mathbb{E}T_0 = \sum_{t=0}^{\infty} 1 - \Phi(t) = \left\| \sum_{t=0}^{\infty} SD^t S^{-1} V_0 \right\| = \|Q\|, \tag{10}$$

where

$$Q = S \begin{pmatrix} \frac{1}{1-\lambda_0} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{1-\lambda_{n-1}} \end{pmatrix} S^{-1} V_0. \tag{11}$$

Similarly, the i-th component of Q gives the expected value of W_i, the amount of time spent in class i before absorption. For any given parameter values, we can numerically compute the eigenvectors of P and use eqn (11) to find W_i. Amazingly, even when analytic expressions for the eigenvectors of P are difficult, we know that eqn (10) always simplifies to the exact solution given in eqn (6).

Finally, from our expression for the cumulative distribution Φ, we see that the probability density function $\phi(t) = \Phi(t) - \Phi(t-1)$ of T_0 is given by

$$\phi(t) = \left\| S \begin{pmatrix} \lambda_0^{t-1} - \lambda_0^t & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{n-1}^{t-1} - \lambda_{n-1}^t \end{pmatrix} S^{-1} V_0 \right\|. \tag{12}$$

In Appendix C, we compute the generating function of T_0 in the case of constant apoptosis. This provides an alternative analytic formula for the density function $\phi(t)$. Moreover, the generating function yields analytic expressions for all the moments of T_0.

There is great value in knowing the distribution, $\phi(t)$, of the waiting time before tumorigenesis in a single-cell lineage. As originally elucidated by Moolgavkar (1978), knowledge of $\phi(t)$ allows us to compute the hazard function of tumorigenesis in a large tissue of N cells. The probability that at least one tumor occurs in the tissue by time t is simply $F(t) = 1 - (1 - \Phi(t))^N$. Hence, the hazard function—that is to say, the instantaneous rate of cancer incidence—is given by $F'(t)/(1 - F(t)) = N\phi(t)/(1 - \Phi(t))$. The hazard function of tumorigenesis is directly comparable to cancer age-incidence data collected from a population (Armitage & Doll, 1954). In most cases, $\Phi(t)$ will be small within a human lifespan. Therefore, the rate of a cancer incidence will be approximated very well simply by $N\phi(t)$. Hence, by understanding the probabilistic behavior of single-cell lineage we can predict the cancer age-incidence rates within a human population (Armitage & Doll, 1954).

7. The Loss of Repair vs. the Loss of Apoptosis

Equations (5), (6) and (12) describe the precise influence of apoptosis and DNA repair as deterrents against cancer. For example, in Fig. 3 we plot the distribution of the waiting time before tumorigenesis in the simple two-hit case, $n = 2$. As expected, when compared with the neutral case, increased mutation rates (the loss of DNA repair genes) cause a decrease in mean time before cancer, while tumor-suppressor-induced apoptosis causes an increase. Interestingly, we also observe that apoptosis greatly increases the variance in the time before tumorigenesis—an observation that is confirmed, analytically, by eqn (23).

In Fig. 4, we compare the progressive loss of apoptotic activity to the progressive loss of DNA repair activity, in the case $n = 4$. If the apoptotic rates α_k and mutation rates p_k are constant, then the distribution of waiting times has a very large variance. We compare the model's behavior with constant α's and p's against two alternative scenarios. In the first alternative, we keep mutation

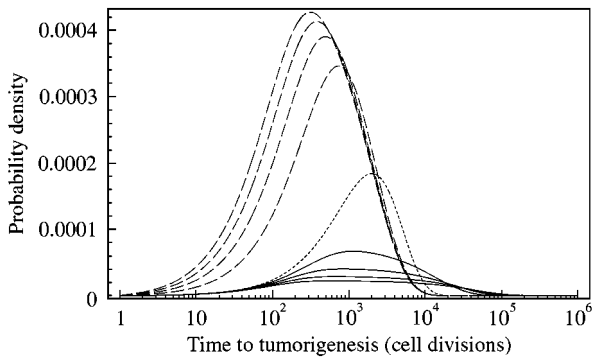

FIG. 3. The effect of increased mutation rates and apoptosis when $n = 2$ mutations are required for tumorigenesis. The figure shows the probability density function of the time before tumorigenesis, T_0, given by eqn (12). In the neutral case (-----), all states have the same mutation rate, $p_0 = p_1 = 500 \times 10^{-6}$, and there is no apoptosis, $\alpha_1 = 0$. The four curves drawn in (– – – –) show the effect of varying degrees of increased mutation, ranging linearly from $p_1 = 0.01$ (uppermost curve) to $p_1 = p_0$. The mutator phenotype decreases the mean time before tumorigenesis. The (——) curves show the effect of various amounts of apoptosis, ranging from $\alpha_1 = 0.01$ (lowermost curve) to $\alpha_1 = 0$. Apoptosis increases the mean waiting time before tumorigenesis. The mutator phenotype decreases the variance in the waiting time, while apoptosis greatly increases the variance.

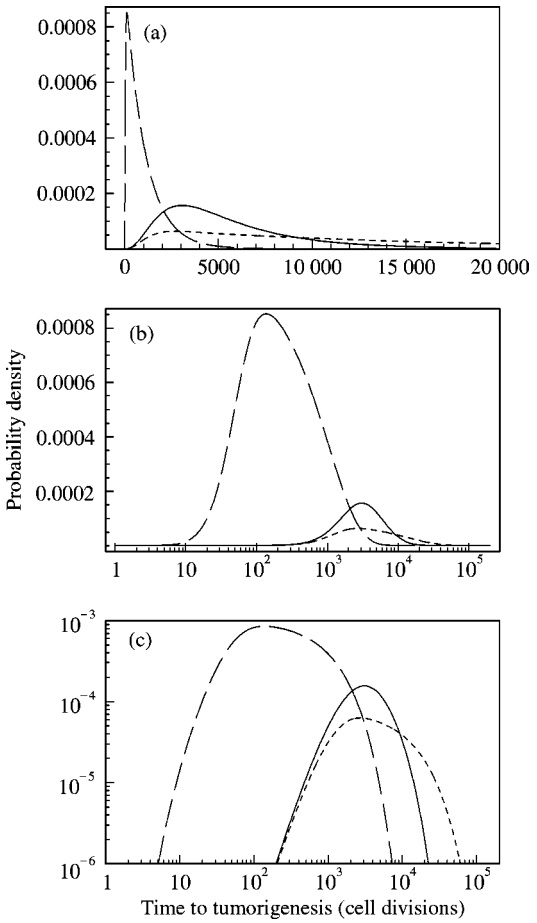

FIG. 4. The distribution of the waiting time before tumorigenesis [$n = 4$, using eqn (12)] shown on linear axes (top), log-linear axes (middle), and log–log axes (bottom). The (-----) curve corresponds to the case when $\alpha_1 = \alpha_2 = \alpha_3 = 0.001$ and $p_0 = p_1 = p_2 = p_3 = 500 \times 10^{-6}$. The (– – – –) curve corresponds to linearly increasing mutation rates, $p_3/p_0 = 100$. The (——) curve corresponds to linearly decreasing apoptotic rates, $\alpha_1/\alpha_3 = 100$. The mutator phenotype causes a dramatic decrease in the mean and variance of time before tumorigenesis. The loss of proper apoptotic function has a more modest effect on the waiting time.

constant and we let the apoptotic rates α_k decrease linearly with k. This models the loss of suppressor-induced apoptosis, resulting in a shorter waiting time before tumorigenesis. In the second alternative, we keep apoptosis constant and we let the mutation rates p_k increase with k. This models the mutator phenotype caused by the loss of DNA repair genes. In this case, tumorigenesis occurs much earlier on average and has a much smaller variance. In other words, the loss of repair has a stronger and more dramatic effect on tumorigenesis than the loss of proper apoptotic functioning, in the sense that an x-fold loss of repair decreases the time before tumor initiation far more than an x-fold loss of apoptotic fidelity (Fig. 4).

Figure 5 further elucidates the effect of apoptotic malfunction compared to increasing mutation rates. Again, we plot the density of the waiting time before tumorigenesis in various alternative scenarios. In the most simple scenario, all the apoptotic rates and mutation rates are constant: $\alpha_1 = \cdots = \alpha_{n-1}$; $p_0 = \cdots = p_{n-1}$. In one set of alternatives, the mutation rates p_k increase with k. The ratio p_{n-1}/p_0 measures the strength of the mutator phenotype. As the strength of the mutator phenotype increases, the mean and variance of the waiting time before tumorigenesis both decrease. In this model, the effect of the mutator phenotype does not saturate: the more the mutation rates increase, the earlier tumorigenesis occurs. In the other set of alternatives, the apoptotic rates α_k decrease with k. Although the loss of apoptosis also results in a shorter time before tumorigenesis, this effect eventually saturates.

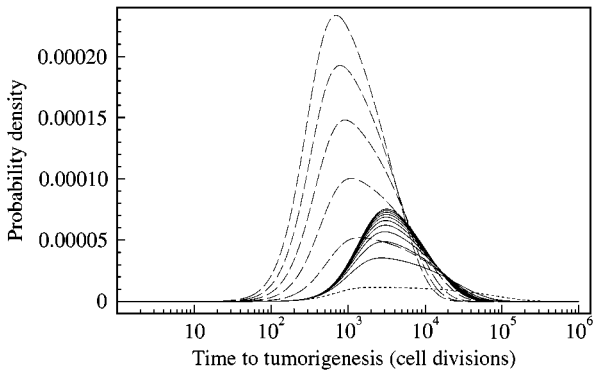

FIG. 5. The distribution of waiting times before tumorigenesis [$n = 4$, using eqn (12)] for various amounts of increased mutation and decreased apoptosis. The (-----) curve corresponds to the simple case of constant apoptotic rates and mutation rates, as in Fig. 4. The (– – – –) curves correspond to cases of increasing mutation rates, with p_3/p_0 ranging from 1 to 10 (uppermost curve). The (——) curves correspond to decreasing apoptotic fidelity with α_1/α_3 ranging from 1 to 20. Note that the effect of reduced apoptotic fidelity asymptotes, whereas the effect of the mutator phenotype does not saturate. This result remains true whether p_k and α_k vary geometrically or arithmetically with k.

Generally speaking, the distribution of the waiting time for cancer is nearly identical whenever α_1/α_3 exceeds 10, for $n \leqslant 6$. The exact effect of progressive loss of apoptotic fidelity depends upon the cancer threshold n and the base-line apoptotic fidelity α_1. But for any parameter values n and α_1, the distribution of T_0 asymptotes as α_1/α_{n-1} increases, whereas the distribution of T_0 does not asymptote with increasing p_{n-1}/p_1. This behavior is apparent in Fig. 5 and can also be seen, in expectation, from eqn (5) directly. This result highlights the importance of the mutator phenotype as a potential driving force behind early tumorigenesis. Conversely, this result indicates that the loss of apoptotic functioning hastens tumorigenesis, but that this effect has an intrinsic limitation.

8. Costs Associated with the Mutator Phenotype

So far, we have ignored any intrinsic costs associated with increased mutation rates. In reality, however, high mutation rates can lead either (i) to the loss of crucial genes needed for cellular function and/or (ii) to the triggering of programmed cell death. This phenomenon indicates that the mutator phenotype p_k and the apoptotic rates α_k should be correlated to some extent. In this section we will explore an extended model in which increased mutation is accompanied by increased cellular death.

We have assumed that there are L "target" genes involved in regulating cell-cycle functioning. Assume now that there are an additional M "critical" genes ($M \approx 10\,000$ in humans) involved in cellular functioning which do not cause cancer, but whose mutation causes cell death (even in a cancerous cell). For any mutation rate μ, we define

$$q_\mu = (1-\mu)^{M+L},$$

$$p_\mu = [1-(1-\mu)^L](1-\mu)^M,$$

$$s_\mu = 1 - q_\mu - p_\mu = [1-(1-\mu)^M].$$

Here q_μ represents the probability per cell division that no genes whatsoever are mutated; p_μ represents the probability that at least one target gene is mutated, but none of the critical M genes is mutated; and s_μ denotes the probability that at least one of the critical genes becomes mutated—which will cause immediate cell death.

Consider the following Markov chain model of tumorigenesis, which includes an intrinsic cost of higher mutation rates:

$$P = \begin{pmatrix} q_{\mu_0}+s_{\mu_0} & p_{\mu_0} & 0 & \cdots & \cdots & 0 \\ \alpha+\beta s_\mu & \beta q_\mu & \beta p_\mu & 0 & \cdots & 0 \\ \alpha+\beta s_\mu & 0 & \beta q_\mu & \beta p_\mu & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ \alpha+\beta s_\mu & 0 & \cdots & 0 & \beta q_\mu & \beta p_\mu \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}.$$

In this formulation, α denotes the apoptotic rate, which we assume to be constant. Here μ_0 denotes the intrinsic mutation rate per gene per replication event of a healthy cell and μ denotes the elevated mutation rate of a cell which lacks one or more regulation genes. Note that either apoptosis or the mutation of any "critical" gene causes a cell to die and be replaced by a surrounding healthy cell. Given that increased mutation now

carries a cost—namely the possibility of mutating a gene critical for cellular survival—we expect to find some optimum increased mutation rate μ which minimized the time to tumorigenesis.

As before, let $W_{0,k}$ denote the mean time spent in class k before absorption into class n, starting from a healthy cell. Using the methods described in Appendix B we deduce that

$$W_{0,k} = \frac{1}{1-\alpha-\beta(q_\mu+s)}\left(\frac{\beta p_\mu}{1-\beta q_\mu}\right)^{k+1-n}$$

$$\text{for } 1 \leqslant k < n-1, \tag{13}$$

$$W_{0,0} = \frac{1-\beta q_\mu}{p_{\mu_0}(1-\alpha-\beta(q_\mu+s))}\left(\frac{\beta p_\mu}{1-\beta q_\mu}\right)^{2-n}. \tag{14}$$

Therefore, the expected time to tumorigenesis is given by

$$T_0 = \sum_{k=0}^{n-1} W_{0,k} =$$

$$\frac{p_{\mu_0} + (p_\mu\beta)^{1-n}(1+q_\mu(\alpha-1))^{n-1}[\beta(p_\mu+q_\mu)-p_{\mu_0}-1]}{p_{\mu_0}(\beta(p_\mu+q_\mu)-1)}. \tag{15}$$

For any parameter values L, M, n, and μ_0, eqn (15) is always minimized at some value $\mu = \mu^*$, which is the optimal increased mutation rate (from the cancer's point of view). Figure 6(a) shows the expected time until tumorigenesis for varying degrees of the mutator phenotype, with associated costs. Figure 6(b) shows the optimum increased mutation rate μ^* as a function of the number of critical genes. The optimal mutation rate μ^* is always less than the mutation rate which simply maximizes the speed of forward mutation, p_μ. We find that the waiting time before tumorigenesis is strongly affected by the number of mutations required for tumorigenesis, n, but interestingly the value of the optimum mutation rate does not depend upon the tumor threshold n or the apoptotic rate α.

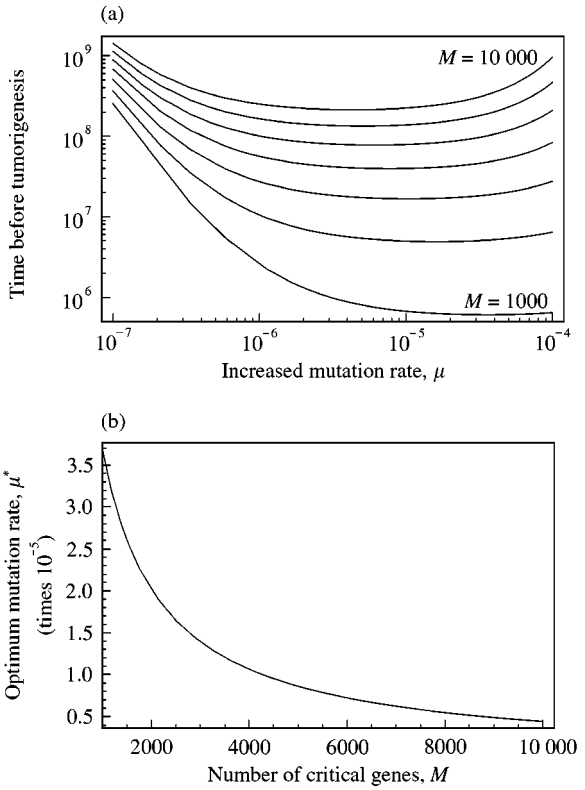

FIG. 6. (a) The expected time until tumorigenesis as a function of the increased mutation rate $\mu > \mu_0 = 10^{-7}$, in the presence of mutation-associated costs [eqn (15)]. There are $L = 500$ target genes, and the apoptotic rate is $\alpha = 0.001$. Six curves are shown, depending upon the number of critical genes, M. If there are more critical genes, then the time to tumorigenesis is increased. For each set of parameters, there is an optimum mutation rate $\mu = \mu^*$ at which the time to tumorigenesis is minimized. (b) The dependence of μ^* on the number of critical genes. If there are more critical genes, then the optimum mutation rate is decreased, and the effect of a strong mutator phenotype can be detrimental (from the cancer's viewpoint). The optimum mutation rate does not depend upon the apoptotic rate α or the cancer threshold n.

9. Conclusions

We have introduced a stochastic framework for investigating the relations between apoptosis, the mutator phenotype, and carcinogenesis. As we have seen, the loss of apoptosis and the loss of DNA repair both hasten the onset of cancer, but their specific effects on the pre-cancer waiting time are dramatically different. The mutator phenotype can cause a dramatic decrease in the mean and variance of the time before cancer, while the loss of apoptosis reduces the mean but not the variance. In addition, stronger mutator phenotypes lead to increasingly rapid

tumorigenesis, while the effect of apoptotic malfunction asymptotes.

We have seen that the effect of mutator phenotype can also asymptote, however, if we include costs associated with increased mutation rates. If a certain set of genes is critical for cellular replication—without which even cancer cells cannot survive—then there exists an optimal value of the increased mutation rate (from the cancer's viewpoint) which minimizes time before tumorigenesis. This optimal mutation rate does not depend upon the amount of apoptosis or the number of hits required for tumorigenesis.

This paper extends multistage models for cancer initiation (Armitage & Doll, 1954; Moolgavkar, 1978) to include the effects of apoptosis and mutator phenotypes. In general, this approach allows us to determine the waiting time before tumorigenesis and the cancer age–incidence curve of a population in terms of the probabilistic behavior of a cell lineage. Because we have not considered the progression of a tumor through its various cancerous stages, the therapeutic significance of our results is limited to cancer prevention. Our results would indicate that gene therapies preserving efficient DNA repair may be more important than those preserving the apoptotic fidelity of cells.

We hasten to discuss some important simplifications used in the modelling framework. The mechanisms and connections between apoptosis and repair are, no doubt, much more complicated in reality than in our model. Although our stochastic framework allows for coordinated behavior, we have largely treated apoptosis (α_k) and DNA repair (p_k) as independent processes. In some sense, we have treated tumor suppressor genes and DNA repair genes separately. In reality, however, some tumor suppressors such as p53 (Vogelstein *et al*., 2000) often induce cell-cycle arrest (Laronga *et al*., 2000) and thereby allow DNA repair genes to remove DNA damage (Tanaka *et al*., 2000). Hence, reduced apoptotic rates α_k may be partially correlated with increased mutation rates p_k.

Similarly, we have assumed that, at each time step, a cell divides and is replaced by a daughter cell. Although this process occurs on average, sometimes in reality an individual cell may divide and produce two viable cells, possibly hastening the spread of cell lineages which have undergone somatic mutations. Similarly, some mutations may increase the rate of cell division, which we have assumed to be constant. Nevertheless, the simplifying assumption of direct daughter-cell replacement has long been considered to be a reasonable model in the cancer literature (Armitage & Doll, 1954; Moolgavkar, 1978). Preliminary work (Nowak & Plotkin, in preparation) indicates that the timing of tumorigenesis is largely unchanged even when this simplifying assumption is relaxed.

Our model also ignores the effect of the immune system and tumor immunity. We have treated each cell lineage as an independent, autonomous process. In reality, the immune system certainly plays a role in some cancer types (Burnet, 1970), while the extent of this role is hotly contested. Recent developments (Albert *et al*., 1998a, b) indicate that apoptosis of tumor cells may be followed by uptake into dendritic cells thereby training T cells to recognize and target the tumor. This process suggests that apoptosis may play a much larger albeit indirect role —by sensitizing the immune system—in tumor control. Our model investigates the patterns of tumorigenesis in the absence of immune activity —which could, by failure to match empirical patterns, provide evidence for immune activity. In fact, the observed limitations of apoptotic malfunction as a direct cause of tumorigenesis in our model, together with the empirical observation that tumor suppressors like p53 are usually mutated in many tumors, may suggest that apoptosis does indeed have an important indirect effect in controlling cancer.

J. B. P. gratefully acknowledges support from the National Science Foundation and the Burroughs Wellcome Fund. J. B. P. and M. A. N. also thank the Alfred P. Sloan Foundation, the Ambrose Monell Foundation, the Florence Gould Foundation, and the J. Seward Johnson Trust. Both authors also appreciate helpful discussions with J. Dushoff, R. Darnell, and M. Albert.

REFERENCES

ALBERT, M. L., DARNELL, J. C., BENDER, A., FRANCISCO, L. M., BHARDWAJ, N. & DARNELL, R. B. (1998a). Tumor-specific killer cells in paraneoplastic cerebellar degeneration. *Nat. Med.* **4,** 1321–1324.

ALBERT, M. L., SAUTER, B. & BHARDWAJ, N. (1998b). Dendritic cells acquire antigen from apoptotic cells and induce class i restricted ctls. *Nature* **392,** 86–89.

ARAVIND, L., DIXIT, V. M. & KOONIN, E. V. (2001). Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. *Science* **291,** 1279–1280.

ARMITAGE, P. & DOLL, R. (1954). The age distribution of cancer and multistage theory of carcinogenesis. *Br. J. Cancer* **8,** 1–12.

BURNET, F. M. (1970). The concept of immunological surveillance. *Progr. Exp. Tumor Res.* **13,** 1–23.

BYRNE, H. M. & CHAPLAIN, M. A. J. (1996a). Modelling the role of cell–cell adhesion in the growth and development of carcinoma. *Math. Comput. Modelling* **24,** 1–17.

BYRNE, H. M. & CHAPLAIN, M. A. J. (1996b). Growth of necrotic tumors in the presence and absence of inhibitors. *Math. Biosci.* **135,** 187–216.

CAIRNS, J. (1998). Mutation and cancer: the antecedents to our studies of adaptive mutation. *Genetics* **148,** 1433–1440.

CHANG, W. C. L., CHAPKIN, R. S. & LUPTON, J. R. (1998). Fish oil blocks azoxymethane-induced rat colon tumorigenesis by increasing cell differentiation and apoptosis rather than decreasing cell proliferation. *J. Nutr.* **128,** 491–497.

CHAPLAIN, M. A. J. (1995). The mathematical modelling of tumour angiogenesis and invasion. *Acta Biotheor.* **43,** 387–402.

CHAPLAIN, M. A. J. (2000). Mathematical modelling of angiogenesis. *J. Neuro-Oncol.* **50,** 37–51.

DARNELL, R. B. (1999). The importance of defining the paraneoplastic neurologic disorders. *N. Engl. J. Med.* **340,** 1831–1833.

DRAKE, J. W., CHARLESWORTH, B., CHARLESWORTH, D. & CROW, J. F. (1998). Rates of spontaneous mutation. *Genetics* **148,** 1667–1686.

GOTTLIEB, T. M. & OREN, M. (1998). p53 and apoptosis. *Semin. Cancer Biol.* **8,** 359–368.

HARNOIS, D. M., QUE, F. G., CELLI, A., LARUSSO, N. F. & GORES, G. J. (1997). Bcl-2 is overexpressed and alters the threshold for apoptosis in a cholangiocarcinoma cell line. *Hepatology* **26,** 884–890.

HONG, M. Y., LUPTON, J. R., MORRIS, J. S., WANG, N. Y., CARROLL, R. J., DAVIDSON, L. A., ELDER, R. H. & CHAPKIN, R. S. (2000). Dietary fish oil reduces o6-methylguanine dna adduct levels in rat colon in part by increasing apoptosis during tumor initiation. *Cancer Epidemiol. Biomarkers Prevent.* **9,** 819–826.

KINZLER, K. W. & VOGELSTEIN, B. (1996). Lessons from hereditary colorectal cancer. *Cell* **87,** 159–170.

KINZLER, K. W. & VOGELSTEIN, B. (1997). Cancer-susceptibility genes—gatekeepers and caretakers. *Nature* **386,** 761–762.

KNUDSON, A. G. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl Acad. Sci. U.S.A.* **68,** 820–823.

LARONGA, C., YANG, H. Y., NEAL, C. & LEE, M. H. (2000). Association of the cyclin-dependent kinases and 14-3-3 sigma negatively regulates cell cycle progression. *J. Biol. Chem.* **275,** 23106–23112.

LENGAUER, C., KINZLER, K. W. & VOGELSTEIN, B. (1998). Genetic instabilities in human cancers. *Nature* **396,** 643–649.

LOEB, L. A. (1991). Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51,** 3075–3079.

MAO, J. H., LINDSAY, K. A., BALMAIN, A. & WHELDON, T. E. (1998). Stochastic modelling of tumorigenesis in p53 deficient mice. *Br. J. Cancer* **77,** 243–252.

MOOLGAVKAR, S. (1978). The multistage theory of carcinogenesis and the age distribution of cancer in man. *J. Natl. Cancer Inst.* **61,** 49–52.

MOOLGAVKAR, S. & VENZON, D. (1979). Two-event models for carcinogenesis: incidence curves for childhood and adult tumors. *Math. Biosci.* **47,** 55–77.

MOOLGAVKAR, S. H. & KNUDSON, A. G. (1981). Mutation and cancer—a model for human carcinogenesis. *J. Natl. Cancer Inst.* **66,** 1037–1052.

MURDOCH, W. J. & VANKIRK, E. A. (1997). Oestradiol inhibits spontaneous and cisplatin-induced apoptosis in epithelial ovarian cancer cells: relationship to DNA repair capacity. *Apoptosis* **2,** 478–484.

ORR-WEAVER, T. L. & WEINBERG, R. A. (1998). Tumorigenesis—a checkpoint on the road to cancer. *Nature* **392,** 223–224.

OWEN, M. R. & SHERRATT, J. A. (1999). Mathematical modelling of macrophage dynamics in tumours. *Math. Models Methods Appl. Sci.* **9,** 513–539.

PERUMPANANI, A. J., SHERRATT, J. A., NORBURY, J. & BYRNE, H. M. (1996). Biological inferences from a mathematical model for malignant invasion. *Invasion Metastasis* **16,** 209–221.

SHERRATT, J. A. (2000). Traveling wave solutions of a mathematical model for tumor encapsulation. *SIAM J. Appl. Math.* **60,** 392–407.

TANAKA, H., ARAKAWA, H., YAMAGUCHI, T., SHIRAISHI, K., FUKUDA, S., MATSUI, K., TAKEI, Y. & NAKAMURA, Y. (2000). A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature* **404,** 42–49.

TOMLINSON, I. & BODMER, W. (1999). Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nat. Med.* **5,** 11–12.

TOMLINSON, I. P. M. & BODMER, W. F. (1995). Failure of programmed cell-death and differentiation as causes of tumors—some simple mathematical-models. *Proc. Natl Acad. Sci. U.S.A.* **92,** 11130–11134.

TOMLINSON, I. P. M., NOVELLI, M. R. & BODMER, W. F. (1996). The mutation rate and cancer. *Proc. Natl Acad. Sci. U.S.A.* **93,** 14 800–14 803.

VOGELSTEIN, B., LANE, D. & LEVINE, A. J. (2000). Surfing the p53 network. *Nature* **408,** 307–310.

WOOD, R. D., MITCHELL, M., SGOUROS, J. & LINDAHL, T. (2001). Human DNA repair genes. *Science* **291,** 1284–1285.

APPENDIX A

The Neutral Case

In this appendix, we analyse the neutral case when apoptosis and repair are ignored. We will calculate the expected time until one of the N cells has at least n mutations, in terms of the genic mutation rate μ. Let us first consider

a single cell. We define the stopping time T as the smallest number of divisions required before a cell has n mutations, given that it starts with $k = 0$ mutations. Clearly, then, the time to tumorigenesis T equals or exceeds n. The probability that the cell lineage becomes cancerous at generation $n + x$ is given by

$$\mathbb{P}(T = n + x) = p^n q^x \binom{n + x - 1}{x}, \quad \text{(A.1)}$$

where, as discussed earlier, $q = (1 - \mu)^L$, and $p = 1 - q$. Note that, for a single cell, the expected value of T is simply $\sum_{x=0}^{\infty}(n + x)\, p^n q^x \binom{n+x-1}{x} = n/p$.

We are interested, however, in the expected time at which any one of the N cells comprising the tissue becomes cancerous. In other words, we want the expected value of $Y = \min\{T_1, T_2, \ldots, T_N\}$ where the T_i are independent random variables each distributed as T. We know that for $a \geqslant n$

$$\mathbb{P}(T \geqslant a) = \sum_{i=a}^{\infty} p^n q^{i-n} \binom{i-1}{i-n}$$

$$= q^{a-1} \sum_{i=0}^{n-1} (p/q)^i \binom{a-1}{i}.$$

The complementary cumulative distribution of Y is given by $\mathbb{P}(Y \geqslant a) = \mathbb{P}(T \geqslant a)^N$. This, in turn, allows us to compute the expected value of Y:

$$\mathbb{E}Y = \sum_{a=1}^{\infty} \mathbb{P}(Y \geqslant a)$$

$$= n + \sum_{a=n+1}^{\infty} \left[q^{a-1} \sum_{i=0}^{n-1} (p/q)^i \binom{a-1}{i} \right]^N. \quad \text{(A.2)}$$

The exact, discrete-time solution [eqn (A.2)] is unwieldy, but it can be approximated very well using a continuous-time, discrete-state model (Poisson process). In this approximation, each cell experiences an exponentially distributed waiting time before accumulating another mutation. The time until a healthy cell acquires one mutation is distributed according to the density function $\phi_1(t) = \lambda e^{-\lambda t}$. We choose $\lambda = -\log(1 - p)$ so that the cell has probability p of acquiring one mutation within one time unit. The distribution $\phi_i(t)$ for the time of i-th mutation satisfies the recursion

$$\phi_i(t) = \int_{s=0}^{t} \phi_{i-1}(s)\, \phi_1(t - s)\, ds,$$

whence we find that the waiting time before tumorigenesis has the Gamma distribution $\phi_n(t) = 1/(n-1)!\lambda^n t^{n-1} e^{-\lambda t}$. This in turn yields the complementary CDF of T: $\mathbb{P}(T \geqslant a) = \int_a^{\infty} \phi_n(t)\, dt = 1/(n-1)!\Gamma(n, \lambda a)$, where Γ is the incomplete Gamma function. We can now compute the mean of Y:

$$\mathbb{E}Y = \int_0^{\infty} \mathbb{P}(T \geqslant a)^N da = \int_0^{\infty} \left(\frac{\Gamma(n, \lambda a)}{(n-1)!} \right)^N da$$

$$= \frac{1}{\lambda (n-1)!^N} \int_0^{\infty} \Gamma(n, a)^N da. \quad \text{(A.3)}$$

Equation (A.3), which is straightforward to compute numerically, provides an accurate approximation of eqn (A.2).

APPENDIX B

The Mean Time before Tumorigenesis

In this appendix, we calculate the mean time before tumorigenesis in the Markov model with arbitrary apoptotic and mutation rates. We start by calculating $W_{0,n-1}$ directly. Then we compute the ratios $W_{0,k}/W_{0,k+1}$. We eventually combine these results into expressions for $W_{0,k}$, yielding $T_0 = \sum_{k=0}^{n-1} W_{0,k}$.

In order to compute $W_{0,n-1}$, we imagine a reduced Markov chain in which states zero through $n - 2$ are collapsed into a single class. This chain has the following transition matrix:

$$M = \begin{pmatrix} \varepsilon & 1 - \varepsilon & 0 \\ \alpha_{n-1} & \beta_{n-1} q_{n-1} & \beta_{n-1} p_{n-1} \\ 0 & 0 & 1 \end{pmatrix},$$

where $\varepsilon \in (0, 1)$ is some value determined by the apoptotic rates, mutation rates, and n. We need not actually calculate the value of ε. We denote

the three states of this process by s, $n-1$, and n (s stands for starting). We can apply eqn (4) to this reduced process in order to calculate $W_{s,n-1}$, which equals $W_{0,n-1}$ in the full Markov chain. The resulting two-dimensional linear system has the unique solution $W_{0,n-1} = 1/p_{n-1}\beta_{n-1}$, independent of ε.

Next, we show that $W_{0,k}/W_{0,k-1} = p_{k-1}\beta_{k-1}/(p_k\beta_k + \alpha_k)$, for $k = 2, 3, \ldots, n-1$. We consider another reduced Markov chain. This time, we collapse states zero through $k-2$ into a single state, and states $k+1$ to $n-1$ into a single state. This yields the following transition matrix with five states, called s, $k-1$, k, f, and n:

$$M = \begin{pmatrix} \varepsilon & 1-\varepsilon & 0 & 0 & 0 \\ \alpha_{k-1} & \beta_{k-1}q_{k-1} & \beta_{k-1}p_{k-1} & 0 & 0 \\ \alpha_k & 0 & \beta_k q_k & \beta_{k-1}p_k & 0 \\ \gamma & 0 & 0 & \delta & 1-\gamma-\delta \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

where, again, ε, γ, and δ are constants which we need not compute. Note that $\gamma + \delta < 1$, because the only absorbing state is n. Although we cannot use this reduced system to compute $W_{0,k}$ or $W_{0,k-1}$ directly, we can compute the ratio of these quantities. We apply eqn (4) twice, once with $j = k$ and once with $j = k-1$. For $j = k$, we obtain a four-dimensional system with unique solution $W_{s,k} = (1-\delta)/(p_k\beta_k(1-\gamma-\delta))$. When $j = k-1$, eqn (4) gives another four-dimensional system with solution $W_{s,k-1} = (1-\delta)(p_k + q_k\alpha_k)/(p_k p_{k-1}\beta_k\beta_{k-1}(1-\gamma-\delta))$. On dividing, we find that $W_{s,k}/W_{s,k-1} = W_{0,k}/W_{0,k-1} = p_{k-1}\beta_{k-1}/(p_k\beta_k + \alpha_k)$, as desired. A similar argument for the boundary value $k = 1$ implies that $W_{01}/W_{00} = p_0/(p_1\beta_1 + \alpha_1)$. By combining these results, we obtain the exact expressions given by eqns (5) and (6).

APPENDIX C

The Generating Function of T_0

In this appendix, we compute the generating function of T_0 in the case when all damaged cells have the same mutation rates $p_1 = \cdots = p_{n-1} = p > p_0$ and apoptotic rates $\alpha_1 = \cdots = \alpha_{n-1} = \alpha$. Let $u_{k,i} = \mathbb{P}(T_k = i)$ denote the probability of absorption at time i, starting from state k. According to the Markov chain in eqn (2), we have the following recursive relations:

$$u_{k,i+1} = \beta q u_{k,i} + \beta p u_{k+1,i} + \alpha u_{0,i} \quad \text{for } 1 \leqslant k < n, \tag{C.1}$$

$$u_{0,i+1} = q_0 u_{0,i} + p_0 u_{1,i} \tag{C.2}$$

subject to the boundary condition $u_{n,i} = \delta_{n,0}$. We define the generating function of T_k as the formal sum

$$U_k(s) = \sum_{i=0}^{\infty} u_{k,i} s^i.$$

On multiplying eqns (C.1) and (C.2) by s^{i+1} and summing over i, we find that

$$U_k(s) = s\beta q U_k(s) + s\beta p U_{k+1}(s) + s\alpha U_0 \quad \text{for } 1 \leqslant k < n, \tag{C.3}$$

$$U_0(s) = s q_0 U_0(s) + s p_0 U_1(s) \tag{C.4}$$

subject to the boundary condition $U_n(s) = 1$. In order to solve these recursions, we may treat the formal variable s as a constant. In general, the recursion

$$\phi(x) = a\phi(x) + b\phi(x+1) + c\phi(1)$$

has the solution

$$\phi(x) = \frac{\Delta b^{-x}(b^2(1-a)^x + b^x c(a-1) + b(a+c-1)(1-a)^x)}{(b+c)(a+b-1)},$$

where Δ is a constant determined by a boundary condition. In our case, we first solve eqn (C.4) for $U_0(s)$ in terms of $U_1(s)$. Treating s as a constant and substituting into eqn (C.3) we find that a recursion of the appropriate form whose solution,

along with boundary condition $U_n(s) = 1$, is

$$U_i(s) = \frac{\alpha p_0 s(\beta q s - 1) + (\beta p(s-1)(1 + s\alpha + s p_0 - s))(1 - 1/p + 1/(ps\beta))^i}{\alpha p_0 s(\beta q s - 1) + (\beta p(s-1)(1 + s\alpha + s p_0 - s))(1 - 1/p + 1/(ps\beta))^n},$$

valid for $1 \leqslant i \leqslant n$. Using eqn (C.4) again, we find that

$$U_0(s) = \frac{p_0(1 + (\alpha - 1)s)(\beta q s - 1)}{\alpha p_0 s(\beta q s - 1) + (\beta p(s-1)(1 + s\alpha + s p_0 - s))(1 - 1/p + 1/(ps\beta))^n},$$

which gives the generating function of T_0. With $U_0(s)$ in hand, we can easily calculate the probability distribution of T_0, or of any T_k:

$$\mathbb{P}(T_0 = i) = \frac{1}{i!}\frac{\mathrm{d}U_0(s)}{\mathrm{d}s^i}\bigg|_{s=0}.$$

We also obtain analytic formulae for all the moments of T_0. For instance,

$$\mathbb{E}T_0 = \frac{\mathrm{d}U_0(s)}{\mathrm{d}s}\bigg|_{s=1}$$

$$= \frac{p\beta(\alpha + p_0)}{\alpha p_0(\alpha q + p)(1 - \alpha/(p + \alpha q))^n} - \frac{1}{\alpha},$$

in agreement with eqn (9). Similarly,

$$\mathrm{Var}\, T_0 = \mathbb{E}(T_0^2) - (\mathbb{E}T_0)^2 \tag{C.5}$$

$$= \frac{\mathrm{d}U_0(s)}{\mathrm{d}s^2}\bigg|_{s=1} + \mathbb{E}T_0 - (\mathbb{E}T_0)^2 \tag{C.6}$$

$$= \frac{X}{\alpha^2}\left[1 + \frac{pZ(pXZ(\alpha + p_0)^2 + \alpha p_0(\alpha^2(p-1) + 2p_0(n-1) + p(p_0+2) + W))}{(\alpha + p\beta)^2 p_0^2}\right], \tag{C.7}$$

where $X = \alpha - 1$, $Z = (1 + \alpha/(p\beta))^n$, and $W = \alpha(p_0 q - p + 2n)$.