Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Theoretical Biology 233 (2005) 15–23

Journal of Theoretical Biology

www.elsevier.com/locate/yjtbi

Population genetics of tumor suppressor genes

Yoh Iwasa[a,*], Franziska Michor[b], Natalia L. Komarova[c], Martin A. Nowak[b]

[a]*Department of Biology, Faculty of Sciences, Kyushu University, Hakozoki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan*
[b]*Department of Organismic and Evolutionary Biology, Department of Mathematics, Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138, USA*
[c]*Department of Mathematics, University of California at Irvine, Irvine, CA 92697, USA*

Received 29 April 2004; received in revised form 3 September 2004; accepted 9 September 2004
Available online 25 November 2004

Abstract

Cancer emerges when a single cell receives multiple mutations. For example, the inactivation of both alleles of a tumor suppressor gene (TSG) can imply a net reproductive advantage of the cell and might lead to clonal expansion. In this paper, we calculate the probability as a function of time that a population of cells has generated at least one cell with two inactivated alleles of a TSG. Different kinetic laws hold for small and large populations. The inactivation of the first allele can either be neutral or lead to a selective advantage or disadvantage. The inactivation of the first and of the second allele can occur at equal or different rates. Our calculations provide insights into basic aspects of population genetics determining cancer initiation and progression.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Tumor suppressor gene; Stochastic tunneling; Multi-state branching process; Somatic evolution of cancer

1. Introduction

Cancers result from an accumulation of mutations in gatekeepers, caretakers, and landscapers (Vogelstein and Kinzler, 2001). Gatekeepers such as oncogenes and tumor suppressor genes (TSGs) directly regulate cellular growth and differentiation pathways (Bishop, 1983; Weinberg, 1991). Oncogenes are activated by gain-of-function mutations that confer increased or novel function; TSGs, in contrast, are affected by loss-of-function mutations. Gatekeeper defects lead to abnormal cellular proliferation, differentiation, and apoptosis. Caretakers function in maintaining the genomic integrity of the cell and regulate DNA repair mechanisms, chromosome segregation, and cell cycle checkpoints (Rajagopalan et al., 2003). Caretaker defects lead to genetic instabilities that contribute to the accumulation of mutations in other genes that directly affect cell proliferation and survival (Lengauer et al., 1998). Landscaper defects do not directly affect cellular growth, but generate an abnormal stromal environment that contributes to the neoplastic transformation of cells.

The concept of a TSG emerged from a statistical analysis of retinoblastoma in children (Knudson, 1971). This study showed that familial cases of retinoblastoma have an earlier age of onset than sporadic cases and individuals are more likely to develop bilateral or multifocal disease. Based on these observations, Knudson developed a model hypothesizing that two hits or mutagenic events are necessary for retinoblastoma development in all cases. In individuals with the inherited form of retinoblastoma, one hit is already present in the germ line. However, inactivation of one allele of the susceptibility gene is insufficient for tumor formation, and the inactivation of the second allele emerges during somatic cell divisions. In the sporadic form of retinoblastoma, both hits emerge during somatic cell divisions. These observations and subsequent work led to the concept of a TSG (Moolgavkar and Knudson, 1981; Friend et al., 1986; Vogelstein and

*Corresponding author. Tel.: +81 92 642 2639; fax: +81 92 642 2645.
E-mail address: yiwasscb@mbox.nc.Kyushu-u.ac.jp (Y. Iwasa).

0022-5193/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.09.001

Kinzler, 2001). In the meanwhile, a large number of TSGs have been discovered that are involved in human cancers (Kinzler et al., 1991; Weinberg, 1991; Knudson, 1993; Levine, 1993) (Table 1).

A normal cell has two wild-type alleles of a TSG. The first hit can be neutral, disadvantageous, or advantageous. A cell with one inactivated allele correspondingly has a normal, decreased, or increased net reproductive rate. The first hit is neutral if the TSG is strictly recessive: the remaining wild-type allele has sufficient tumor suppressing function. The first hit is disadvantageous if the TSG is checked by apoptotic defense mechanisms: as soon as surveillance mechanisms discover an imbalance in the TSG product, apoptosis is triggered. The first hit is advantageous if the TSG is haploinsufficient: the remaining wild-type allele has insufficient tumor suppressing function. The second hit is typically advantageous, and a cell with two inactivated alleles has an increased net reproductive rate. In some cases, TSG inactivation might lead to a growth advantage only after yet another gene has been altered.

The first hit can be constituted by a point mutation, small insertion, deletion, structural change of the chromosome or chromosomal loss. The second hit can be constituted by all of these events plus mitotic recombination. Usually large deletions or chromosome loss do not account both for the first and second hit in one cell, because large homozygous deletions are often lethal.

Mathematical models have been developed to investigate many different aspects of cancers. Studies of the age-dependent incidence curves of human cancers (Nordling, 1953; Armitage and Doll, 1954, 1957; Fisher, 1958) led to the idea that multiple probabilistic events are required for the somatic evolution of cancer (Nunney, 1999; Tomlinson et al., 2002). Knudson's statistical analysis of the incidence of retinoblastoma in children (Knudson, 1971) was later extended to a two-stage stochastic model for the process of cancer initiation and progression (Moolgavkar and Knudson, 1981) and inspired much subsequent work (Grist et al., 1992; Luebeck and Moolgavkar, 2002; Gatenby and Vincent, 2003). Later on, specific theories were developed to explain drug resistance (Goldie and Coldman, 1979, 1983), angiogenesis (Anderson and Chaplain, 1998; Wodarz and Krakauer, 2001), immune responses against tumors (Owen and Sherratt, 1999), the age-specific acceleration of cancer (Frank, 2004), and genetic instabilities (Taddei et al., 1997; Strauss, 1998; Chang et al., 2001, 2003; Maser and DePinho, 2002; Nowak

Table 1
Tumor suppressor genes are involved in various human diseases

TSG	Location	Size (kb)	Function	Disorder
APC	5q21-q22	138.340	Metabolic function unknown	Periampullary Adenoma, Adenomatous Polyposis Coli, Colorectal Cancer, Desmoid Disease, Gardner Syndrome, Gastric Cancer, Turcot Syndrome
BRCA1	17q21	81.09	Transcriptional regulator, growth inhibitor	Breast Cancer, Ovarian Cancer, Proliferative Breast Disease, Papillary Serous Carcinoma of Peritoneum
BRCA2	13q12.3	84.188	Metabolic function unknown	Breast Cancer
CDH1 = E-cadherin	16q22.1	98.25	Cell–cell adhesion glycoprotein, tumor progression, invasion, metastasis	Gastric Cancer, Breast Cancer, Colorectal Cancer, Thyroid Cancer, Ovarian Cancer
CDKN1 = p21	6p21.1	10.794	Negative regulator of ras signal transduction	Neurofibromatosis, Juvenile Myelomonocytic Leukemia, Melanoma
CDKN2 = p16	9p21	30.585	Cell cycle G1 control, inhibitor of CDK4 kinase, stabilizer of p53	Melanoma, Nervous System Tumours, Pancreatic Cancer, Orolaryngeal Cancer, Cutaneous Malignant Melanoma, Bladder Cancer
NF1	17q11.2	279.538	Negative regulator of ras signal transduction	Neurofibromatosis, Juvenile Myelomonocytic Leukemia, Melanoma
p53	17p13.1	19.178	Activation of expression of genes that inhibit growth and/or invasion	Colorectal Cancer, Li-fraumeni Syndrome, Hepato-Cellular Carcinoma, Osteosarcoma, Histiocytoma, Thyr-oid Carcinoma, Nasopharyngeal Carcinoma, Pancreatic Cancer, Adrenal Cortical Carcinoma, Breast Cancer
Rb	13q14.2	178.234	Inhibits progression from G1 to S phase	Bladder Cancer, Osteosarcoma, Pinealoma with Bilateral Retinoblastoma, Retinoblastoma

The table shows the name, the genomic location, the size in kilobases (kb), the cellular function and the associated diseases of different tumor suppressor genes.

et al., 2002; Michor et al., 2003; Otsuka et al., 2003; Michor et al., 2004a, b; Komarova and Wodarz, 2003, 2004). All of these aspects of cancer include rare mutations, their spread, and the fixation in a random process. They are basically population genetics problems (Nowak et al., 2004).

In the present paper, we calculate the probability as a function of time that a population of cells has generated at least one cell with two inactivated alleles of a TSG. We study the dependence of this probability on the population size, the fitness values of cells with one or two inactivated alleles, and the rates of inactivation of the first and the second allele.

2. The mathematical model

Consider a population of N cells following the Moran model (Moran, 1962). Initially, the population consists of cells that are wild-type with respect to a specific TSG. A wild-type cell is denoted by a type 0 cell and has a relative fitness value of 1. At each time step, a cell is chosen for reproduction at random, but proportional to fitness. The chosen cell produces a daughter cell that replaces another randomly chosen cell that dies. The total number of cells remains strictly constant. The time unit is chosen such that the mean time of a cellular generation is 1. A type 0 cell is mutated with probability u_1 per cell division to give rise to a type 1 cell. A type 1 cell has one inactivated TSG allele and has a relative fitness value of r. The fitness value r can be less than, equal to or greater than 1. A type 1 cell is mutated with probability u_2 per cell division to give rise to a type 2 cell. A type 2 cell has two inactivated TSG alleles and has a large fitness advantage. Once a type 2 cell has been produced, it proliferates quickly and spreads within the population. We are interested in the probability distribution of the time of emergence of the first type 2 cell.

Let $v(t)$ be the probability that the first type 2 cell has emerged before time t, given that the population consists only of type 0 cells at time $t=0$. Let $f(t)$ be the probability that the first type 2 cell has emerged before time t in a lineage that started with a single type 1 cell at time $t=0$. To relate $v(t)$ with $f(t)$, consider $1-v(t+\Delta t)=[1-v(t)][1-Nu_1\Delta t\cdot f(t)]$. Here terms smaller than Δt are neglected. This is the probability that no type 2 cell arises in a time interval of length $t+\Delta t$. In a time interval of length Δt, a type 1 cell is produced with probability $Nu_1\Delta t$ and leads with probability $f(t)$ to the appearance of a type 2 cell. In the limit $\Delta t\to 0$, the above formula becomes $dv/dt=Nu_1 f(t)[1-v(t)]$. With $v(0)=0$, we have

$$v(t)=1-\exp\left[-Nu_1\int_0^t f(s)\,ds\right]. \tag{1}$$

Now, consider the events occurring in a time interval of length Δt. We have

$$f(t+\Delta t)=\Delta t\cdot 0+r\Delta t\left[(1-u_2)\left(1-[1-f(t)]^2\right)+u_2\cdot 1\right] + [1-(1+r)\Delta t]f(t)$$

as the probability that a type 2 cell arises in the time interval of length $t+\Delta t$ within a lineage that started from a single type 1 cell. The probability is decomposed to three terms describing the events that occur within Δt. The first term on the right-hand side accounts for the extinction of the type 1 cell, the second term accounts for cell division, and the third for the absence of these events. Each cell division has the probability that a type 2 cell is produced; this probability is u_2. Each cell division has the probability that a second type 1 cell is produced; this probability is $1-u_2$. If a second type 1 cell emerges, then the probability that a type 2 cell arises can be expressed in terms of the probability that there is only one type 1 cell. Here, we use the assumption of the branching process theory, i.e. the assumption of the independence of two lineages starting from two different cells. This is a good approximation if the total population size is large. With $\Delta t=0$, we have

$$\frac{df}{dt}=ru_2-(1-r+2ru_2)f-r(1-u_2)f^2. \tag{2}$$

By setting $df/dt=0$, we have a quadratic equation with the solutions $f=a$ and $-b$, where

$$a=\frac{1}{2(1-u_2)}\left[-\left(\frac{1-r}{r}+2u_2\right)+\sqrt{\left(\frac{1-r}{r}+2u_2\right)^2+4u_2}\right], \tag{3a}$$

$$b=\frac{1}{2(1-u_2)}\left[\frac{1-r}{r}+2u_2+\sqrt{\left(\frac{1-r}{r}+2u_2\right)^2+4u_2}\right]. \tag{3b}$$

With $f(0)=0$, integration of Eq. (2) gives

$$f(t)=\frac{\exp[r(1-u_2)(a+b)t]-1}{(1/a)\exp[r(1-u_2)(a+b)t]+1/b}. \tag{4}$$

Eqs. (1) and (4) give the probability that a type 2 cell emerges before time t as

$$v(t)=1-\exp\left[-Nu_1\left\{\frac{a+b}{c}\ln\frac{e^{ct}+a/b}{1+a/b}-bt\right\}\right], \tag{5}$$

where $c=r(1-u_2)(a+b)$.

2.1. Approximations

We can derive simple approximations for $v(t)$ that have a clearer parameter dependence than Eq. (5).

The function *f*(*t*) increases smoothly from 0 to a positive constant (Eq. (4)), and we approximate it in two limits:

$$f(t) = ru_2t \quad \text{for small } t, \tag{6a}$$

$$f(t) = a \quad \text{for large } t. \tag{6b}$$

The limit $f(\infty) = a$ is the maximum probability that a type 2 cell emerges from a lineage that started from a single type 1 cell at $t = 0$. The two limits correspond to two approximations,

$$v(t) = 1 - \exp\left[-\frac{1}{2}Nru_1u_2t^2\right] \tag{7a}$$

and

$$v(t) = 1 - \exp[-Nu_1at]. \tag{7b}$$

Note that both approximations overestimate the exact probability $v(t)$ as given in Eq. (5), because Eqs. (6a) and (6b) exceed the exact probability *f*(*t*) as given in Eq. (4).

2.1.1. Half-time

The half time, $T_{1/2}$, is defined as the time when the probability that a type 2 cell emerged is $\frac{1}{2}$; i.e. the time satisfying $v(T_{1/2}) = 1/2$. From Eqs. (7a) and (7b), we derive

$$T_{1/2} = \sqrt{\frac{2 \ln 2}{Nru_1u_2}} \tag{8a}$$

and

$$T_{1/2} = \frac{\ln 2}{Nu_1a}. \tag{8b}$$

2.1.2. Fitness

The limit $f(\infty) = a$ strongly depends on the fitness of type 1 cells. From Eq. (3a), we have

$$a = \sqrt{u_2} \quad \text{when } 1 - \sqrt{u_2} < r < 1 + \sqrt{u_2}\,\text{(neutral)}, \tag{9a}$$

$$a = \frac{ru_2}{1 - r} \quad \text{when } r < 1 - \sqrt{u_2}\,\text{(disadvantageous)}, \tag{9b}$$

$$a = 1 - \frac{1}{r} \quad \text{when } r > 1 + \sqrt{u_2}\,\text{(advantageous)}. \tag{9c}$$

2.1.2.1. Neutrality:.

If type 1 cells are neutral, $1 - \sqrt{u_2} < r < 1 + \sqrt{u_2}$, we have $f(\infty) = a \approx \sqrt{u_2}$ (Eq. (9a)). Two approximations of the half-time hold for different limits:

$$T_{1/2} = \sqrt{\frac{2 \ln 2}{Nu_1u_2}} \quad \text{if } N \text{ is very large}, \tag{10a}$$

$$T_{1/2} \approx \frac{\ln 2}{Nu_1\sqrt{u_2}} \quad \text{if } N \text{ is intermediate}. \tag{10b}$$

Eq. (10a) shows that the half-time, $T_{1/2}$, is inversely proportional to the square root of the population size, *N*. On a $\log N$-$\log T_{1/2}$ plane, the half-time appears as a straight line with slope $-\frac{1}{2}$. Eq. (10b) shows that the half-time, $T_{1/2}$, is inversely proportional to the population size, *N*. On a $\log N$-$\log T_{1/2}$ plane, the half-time appears as a straight line with slope -1. These two lines cross at the critical population size $N_c = \ln 2/2u_1$.

In Fig. 1, the solid curve labeled B shows the half-time calculated from probability $v(t)$ given by Eq. (5), for $r = 1$. Broken curves labeled L and I are its approximations for large and intermediate *N*, as given by Eqs. (10a) and (10b), respectively. The solid circles represent the exact computer simulation of the stochastic process as discussed in Section 2.1.3.

The solid curve representing Eq. (5) is very accurate except for small population sizes ($N < 10$). We derive a formula for small *N* in Section 2.1.4, and a formula valid both for small *N* and for intermediate *N* in Section 2.1.5. Eq. (5) is closely approximated by the two straight lines representing Eqs. (10a) and (10b). Eq. (10a) is accurate for very large population sizes ($N > 10^5$), and Eq. (10b) is accurate for intermediate population sizes ($10^2 < N < 10^4$).

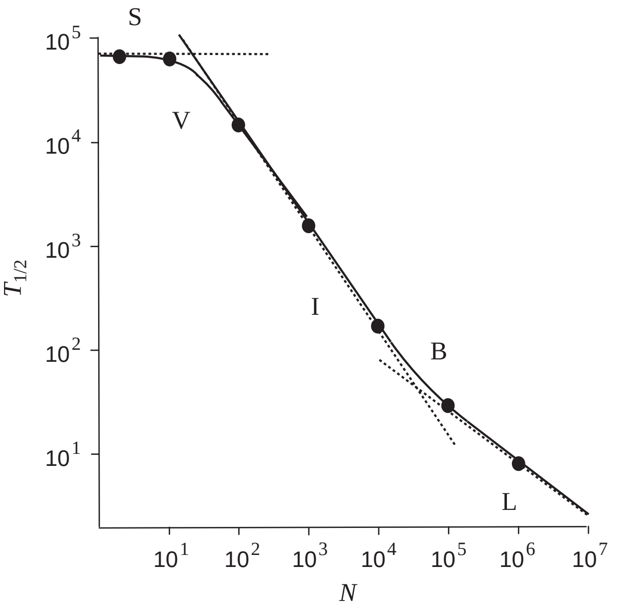

Fig. 1. Half-time of a TSG. The half-time is defined as the time when the probability of having inactivated both alleles of a TSG is $\frac{1}{2}$. A cell with one inactivated allele is neutral, $r = 1$. The population size, *N*, is shown on the horizontal axis. The circles depict the data points of the exact computer simulation. The solid curve labeled B is the branching process formula Eq. (5), and the one labeled V is from Eq. (14). Three broken curves are simplified formulas: those labeled L, I, and S are curves for large *N* (Eq. (10a)), intermediate *N* (Eq. (10b)), and small *N* (Eq. (13)), respectively.

2.1.2.2. Fitness disadvantage:. If type 1 cells are disadvantageous, $r < 1 - \sqrt{u_2}$, we have $f(\infty) = a \approx ru_2/(1-r)$ (Eq. (9b)). Two approximations of the half-time hold for different limits:

$$T_{1/2} = \sqrt{\frac{2 \ln 2}{Nru_1u_2}} \quad \text{if } N \text{ is very large,} \tag{11a}$$

$$T_{1/2} = \frac{(1-r)\ln 2}{Nu_1u_2r} \quad \text{if } N \text{ is intermediate.} \tag{11b}$$

The two approximations cross at the critical population size $N_c = (1-r)^2 \ln 2/2ru_1u_2$. Eq. (11a) is valid for $N \gg N_c$, whereas Eq. (11b) is valid for $N \ll N_c$. The critical population size for disadvantageous type 1 cells is much larger than the critical size for neutral type 1 cells ($N_c = \ln 2/2u_1$), because $u_2 \ll 1$.

In Fig. 2, the solid curve labeled B shows the half-time calculated from probability $v(t)$ given by Eq. (5), for $r = 0.9$. Broken curves labeled L and I are its approximations for large and intermediate N, as given by Eqs. (11a) and (11b), respectively. The solid circles represent the exact computer simulation of the stochastic process as discussed in Section 2.1.3. Eqs. (11a) and (11b) are parallel to but located above Eqs. (10a) and (10b), respectively. While Eq. (11b) is shifted considerably from Eq. (10b), Eq. (11a) is very close to (10a).

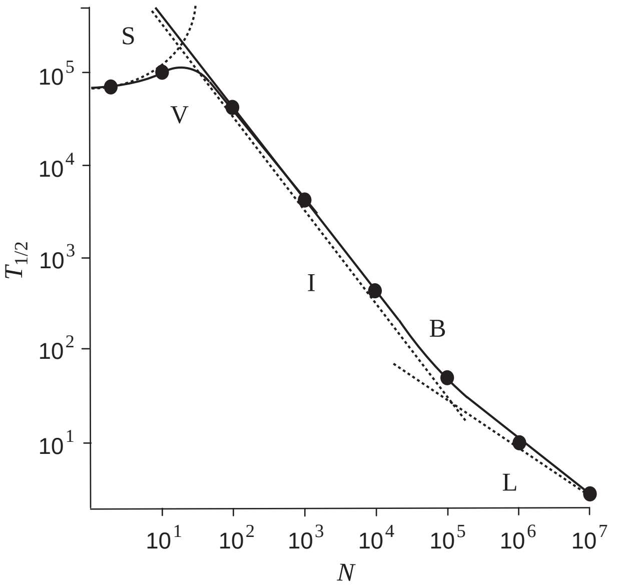

Fig. 2. Half-time of a TSG. A cell with one inactivated allele is deleterious, $r = 0.9$. The population size, N, is shown on the horizontal axis. The circles depict the data points of the exact computer simulation. The solid curve labeled B is the branching process formula Eq. (5), and the curve labeled V represents Eq. (14). Three broken curves are simplified formulas: those labeled L, I, and S are curves for large N (Eq. (11a)), intermediate N (Eq. (11b)), and small N (Eq. (13)), respectively.

2.1.2.3. Fitness advantage:. If type 1 cells are advantageous, $r > 1 + \sqrt{u_2}$, we have $f(\infty) = a \approx 1 - 1/r$ (Eq. (9c)). Two approximations of the half-time hold for different limits:

$$T_{1/2} = \sqrt{\frac{2 \ln 2}{Nru_1u_2}} \quad \text{if } N \text{ is very large,} \tag{12a}$$

$$T_{1/2} = \frac{r \ln 2}{Nu_1(r-1)} \quad \text{if } N \text{ is intermediate.} \tag{12b}$$

The two approximations cross at the critical populations size $N_c = u_2r^3 \ln 2/2u_1(r-1)^2$. Eq. (12a) is valid for $N \gg N_c$, whereas Eq. (12b) is valid for $N \ll N_c$. The critical population size for an advantageous type 1 cell is quite small.

In Fig. 3, the solid curve labeled B shows the half-time calculated from probability $v(t)$ given by Eq. (5), for $r = 1.1$. Broken curves labeled L and I are its approximations for large and intermediate N, as given by Eqs. (12a) and (12b), respectively. The solid circles represent the exact computer simulation of the stochastic process as discussed in the following.

2.1.3. Exact computer simulations

In Figs. 1–3, we compare our analytical results with direct computer simulations of the Moran process. At each elementary step of the Moran process, one cell is chosen for reproduction at random, but proportional to fitness. If there are i mutated cells with fitness r in a population of N cells, then the probability that a

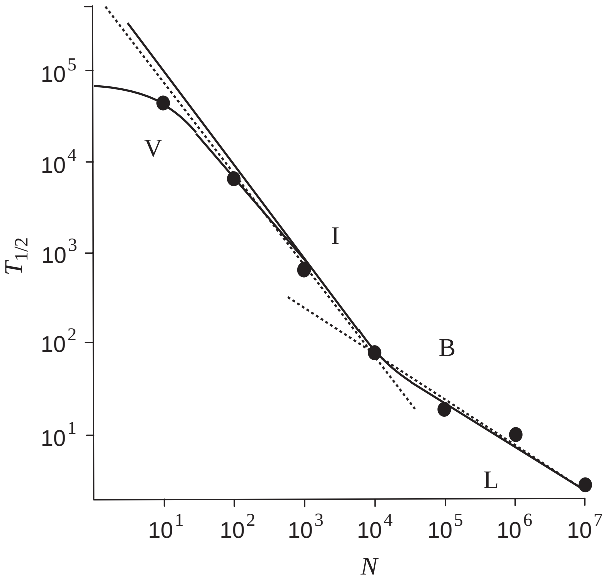

Fig. 3. Half-time of a TSG. A cell with one inactivated allele is advantageous, $r = 1.1$. The circles depict the data points of the exact computer simulation. The solid curve labeled B is the branching process formula Eq. (5), and the one labeled V is given by Eq. (14), which overlaps with the curve for small N (Eq. (13)). Two broken curves labeled L and I are for large N (Eq. (12a)) and intermediate N (Eq. (12b)), respectively.

mutated cell is chosen for reproduction is $ri/(ri+N-i)$. The chosen cell produces a daughter cell, possibly with mutation, that replaces another randomly chosen cell that dies. The total number of cells remains strictly constant. For each parameter choice, we compute many independent runs of the stochastic process. Then we calculate the half-time, $T_{1/2}$, until the probability of emergence of at least one type 2 cell is $\frac{1}{2}$. The results of the exact computer simulations are depicted by solid circles in Figs. 1–3.

In Figs. 1–3, the solid curves labeled B indicates the half-time, $T_{1/2}$, calculated from the exact probability that at least one type 2 cell emerges (Eq. (5)), which is based on the branching process calculation. Broken curves labeled with L and I are approximations of curve B for very large population sizes (Eq. (10a) for neutral type 1 cells, Eq. (11a) for disadvantageous type 1 cells, and Eq. (12a) for advantageous type 1 cells), and for intermediate population sizes (Eq. (10b) for neutral type 1 cells, Eq. (11b) for disadvantageous type 1 cells, and Eq. (12b) for advantageous type 1 cells).

For large and intermediate N, the simulation results indicated by solid circles are very well predicted by the branching process formula (curve B). Simplified formulas are also very accurate within the respective range of N.

2.1.4. Fixation of type 1 cells

Figs. 1–3 show that neither of the approximations nor the exact probability (Eq. (5)) is accurate for small N. Eq. (5) is based on the assumption of the branching process theory that lineages starting from different cells behave independently. In small populations, however, type 1 cells are frequently fixed before the first type 2 cell emerges. When a lineage descending from a single type 1 cell is fixed in the population, all lineages starting from other type 1 cells that exist at the same time die out. Hence, if the fixation of type 1 cells occurs quickly, the assumption of independent lineages does not hold.

If the fixation of type 1 cells always precedes the emergence of the first type 2 cell, then we can calculate the waiting time as follows. The time until the fixation of type 1 cells is a stochastic variable with an exponential distribution and mean $1/Nu_1\rho(r)$. The fixation probability of a single cell with fitness r is $\rho(r) = (1-1/r)/(1-1/r^N)$ in the Moran process (Komarova et al., 2003). The expression $1/Nu_1\rho(r)$ is accurate if the waiting time for the appearance of the first successful type 1 cell is much longer than the time required for its lineage to be fixed, i.e. for small population sizes. In this case, the waiting time for the appearance of the first type 2 cell follows an exponential distribution with mean $1/Nu_2$. The probability $v(t)$ is a convolution of two exponential distributions. The fixation of type 1 cells takes much longer than the appearance of a type 2 cell, because the fixation probability of a neutral cell is $\rho(1) = 1/N$ and the fixation probability of a disadvantageous cell is $\rho(r) < 1/N$. We have $1/Nu_1\rho(r) \gg 1/Nu_2$, and therefore $v(t) = 1-\exp[-Nu_1\rho(r)t]$. The half-time is given by

$$T_{1/2} = \frac{\ln 2}{Nu_1\rho(r)} \quad \text{if } N \text{ is small.} \tag{13}$$

In Figs. 1–3, the half-time calculated by Eq. (13) is indicated by broken curves labeled S. If type 1 cells are neutral, then $\rho(1) = 1/N$ and the half-time is independent of N and appears as a horizontal line in Fig. 1. If type 1 cells are disadvantageous, the half-time increases with N for small N, reaches a maximum at the boundary of small and intermediate N, and decreases with N for intermediate N (Fig. 2). If type 1 cells are advantageous, then the half-time decreases with N for all N (Fig. 3). The formula of Eq. (13) is smoothly connected to Eq. (12b) for intermediate N.

2.1.5. Between small and intermediate population sizes

When type 1 is either neutral or deleterious, at the boundary between small and intermediate N, neither Eq. (11b), Eq. (12b), nor Eq. (13) is accurate. Note that a type 2 cell can emerge before or after the fixation of type 1 cells. The appearance of a type 2 cell after the fixation of type 1 cells can be described as a two-step process (see Section 3.). This process is slow if type 1 cells are disadvantageous and the population size is large. Alternatively, the first type 2 cell can emerge before the fixation of type 1 cells. This process is called 'stochastic tunneling' (Iwasa et al., 2004). The total rate of the appearance of a type 2 cell is given by the sum of the two possibilities. The half-time is given by

$$T_{1/2} = \frac{\ln 2}{Nu_1[1-V_1]}. \tag{14}$$

The quantity V_1 is calculated from an iterative procedure. The boundary conditions $V_0 = 1$ and $V_N = 0$ and the initial condition $V_i = 1 - i/N$ for $i = 0, \ldots, N-1$ specify V_i' as

$$V_i' = \frac{rV_{i+1} + V_{i-1}}{ru_2(ir+(N-i))/(N-i)+r+1} \quad i = 1, 2, 3, \ldots, \; N-1. \tag{15}$$

The converging value of the iteration, V_1, is used in Eq. (14) (for derivation, see Iwasa et al., 2004). The solid curves labeled V in Figs. 1–3 represent Eq. (14). Eq. (14) is very accurate for small and intermediate N.

Broken lines labeled I in Figs. 1–3, and formulas Eqs. (8b), (10b), and (11b) are all for the tunneling rate. In contrast, the broken line labeled S (representing Eq. (13)) accounts for the situation in which type 1 cells become fixed before a type 2 cell emerges. Both of these are covered by Eq. (14) (see solid curves labeled V).

However, Eq. (14) does not hold for large N, nor any of its approximations. The mathematical analysis leading to Eqs. (14) and (15) assumes that the time until the appearance of the first successful type 1 cell (i.e., the lineage which eventually generates the first type 2 cell) is much longer than the time between the appearance of the first successful type 1 cell and the generation of the first type 2 cell within its lineage (Iwasa et al., 2004). This assumption holds if mutations from type 0 to type 1 occur infrequently, $Nu_1 \ll 1$. If this inequality does not hold, the time for the descendents of the first successful type 1 cells to spread in the population becomes a major part of the half-time. This causes a deviation from Eq. (14) (curve V in Figs. 1–3) for very large N, in which the formula based on branching process (Eq. (5)) works very well (see curve B in the figures).

3. Step number

The threshold population size separating intermediate and large populations is $N_c = \ln\, 2/2u_1$ if type 1 cells are neutral, and $N_c = (1-r)\ln\, 2/2ru_1u_2$ if type 1 cells are disadvantageous. The second expression is much larger than the first because $u_2 \ll 1$. The formula for intermediate populations holds for much larger populations if type 1 cells are disadvantageous than if type 1 cells are neutral. For populations sizes $1/u_1 \ll N \ll 1/u_1u_2$, the formula for large populations holds if type 1 cells are neutral, but the formula for intermediate populations holds if type 1 cells are disadvantageous.

The difference between disadvantageous and neutral type 1 cells is also observed in the increase of the probability $v(t)$ with time. The probability of having the first type 2 cell before time t can be written as $v(t) = 1 - \exp[-R(t)]$, where $R(t) = Nu_1 \int_0^t f(t')\,\mathrm{d}t'$. The function $R(t)$ is proportional to t if Eq. (7b) holds and to t^2 if Eq. (7a) holds. Hence a single-step transition describes the process in intermediate populations, but a two-step transition in large populations. Since $f(t)$ converges to an asymptote $f(\infty) = a$, $R(t)$ has a linear dependence on t for a very large t:

$$R(t) = Nu_1 \left\{ \frac{a+b}{c} \ln \frac{\mathrm{e}^{ct} + a/b}{1 + a/b} - bt \right\} \approx Nu_1 a(t - \delta).$$

The time delay is $\delta = \left((a+b)/ac\right) \ln\left(1 + a/b\right)$. If type 1 cells are disadvantageous, $r < 1 - \sqrt{u_2}$, we have $a \approx ru_2/(1-r)$, $b \approx (1-r)/r$, and $c \approx 1 - r$. Then the time delay becomes $\delta = 1/(1-r)$, which is of the order of 1. If type 1 cells are neutral, $1 - \sqrt{u_2} < r < 1 + \sqrt{u_2}$, we have $a \approx \sqrt{u_2}$, $b \approx \sqrt{u_2}$, and $c \approx 2\sqrt{u_2}$. Then the time delay becomes $\delta = \ln 2\big/\sqrt{u_2}$. This is of the order of $1\big/\sqrt{u_2}$, which is fairly long because $u_2 \ll 1$. Hence, the function $R(t)$ increases linearly with time with a very short delay if type 1 cells are disadvantageous, but nonlinearly (quadratically) with time for a much longer time if type 1 cells are neutral.

Even if multiple mutations are needed for a phenotypic effect, the incidence curve may not clearly reflect a multi-step process if intermediate mutants are disadvantageous. There are two situations in which a clear two-step incidence curve is observed: first, if type 1 cells are neutral, and second, if type 1 cells reach fixation. Type 1 cells reach fixation because the population is small or because type 1 cells are advantageous. Traditional models of cancer genetics assume a multi-state branching process neglecting fixation. The model is fitted to the observed age-dependent incidence curves of cancer. The neglected possibility of fixation, however, is misleading because it incorrectly suggests the neutrality of intermediate mutants. A recent model of multi-step progression of cancer (Luebeck and Moolgavkar, 2002) assumes an initial population size of 1—the stem cell—that undergoes clonal expansion when mutated. The expansion is described by a branching process. Including the possibility of fixation in models that are fitted to age-dependent incidence data is crucial for understanding the population genetics of cancer.

4. Discussion

In this paper, we study the following question: what is the probability that a single cell with two inactivated TSG alleles has arisen by time t in a population of N cells? Interestingly, the answer depends on the population size, N, as compared with the mutation rates that constitute the first and second hit, u_1 and u_2. There are three different cases. We illustrate this concept by considering the case in which cells with one inactivated TSG allele are neutral.

First, in small populations, $N < 1\big/\sqrt{u_2}$, a cell with one inactivated allele reaches fixation in the population before a cell with two inactivated alleles arises. The probability that at least one cell with two hits emerges before time t is

$$P(t) = 1 - \frac{Nu_2\, \exp(-u_1 t) - u_1\, \exp(-Nu_2 t)}{Nu_2 - u_1}.$$

For very short times, $t < 1/Nu_2$, we can approximate $P(t) \approx Nu_1u_2t^2\big/2$. Therefore, this probability accumulates as second order of time: it takes two rate limiting hits to inactivate a TSG in a small population of cells.

Second, in populations of intermediate size, $1\big/\sqrt{u_2} < N < 1/u_1$, a cell with two inactivated alleles emerges before a cell clone with one inactivated allele has taken over the population. The population 'tunnels' from wild type directly to the second hit without ever having fixed the first hit. The probability that at least one cell with two hits has arisen before time t is

$$P(t) = 1 - \exp\left(-Nu_1\sqrt{u_2}t\right).$$

This probability accumulates as a first order of time: it takes only one rate limiting hit to inactivate a TSG in a population of intermediate size.

Third, in very large populations, $N > 1/u_1$, cells with one inactivated allele arise immediately and the waiting time for a cell with two inactivated alleles dominates the dynamics. The probability that at least one cell with two hits has arisen before time t is

$$P(t) = 1 - \exp\left(-Nu_1u_2t^2/2\right).$$

This probability again accumulates as a second order of time. Eliminating a TSG in a large population of cells is, however, not rate limiting for the overall process of tumorigenesis. Due to the large population size, mutated cells are constantly being produced, and the inactivation of a TSG is not rate limiting.

4.1. Half-time

In the present paper, we have examined the half-time, defined as the time when the probability of having inactivated both TSG alleles in a population of N cells is $\frac{1}{2}$. Three different approximations describe the half-time for small, intermediate, and large populations and appear as distinct curves in Figs. 1–3.

In small populations, type 1 cells reach fixation before a type 2 cell emerges. The waiting time for the first type 2 cell is expressed as a convolution integral of two random variables: the time for emergence and fixation of type 1 cells, with mean $1/Nu_1\rho(r)$, and the waiting time for the emergence of the first type 2 cell, with mean $1/Nu_2$. The latter event occurs at a much faster rate than the former unless type 1 cells are advantageous. Consequently, the overall waiting time can be approximated by a single-step transition with rate $Nu_1\rho(r)$ (broken lines labeled S in the figures).

In intermediate and large populations, a type 2 cell emerges before type 1 cells reach fixation. This phenomenon is called 'stochastic tunneling' and occurs as a single step transition (Nowak et al., 2002; Komarova et al., 2003; Iwasa et al., 2004). Type 1 cells are continuously produced until one cell succeeds to give rise to a type 2 cell. The formulas are very accurate (see broken lines labeled I).

For small and intermediate populations, the waiting time for the first successful type 1 mutant cell to arise is much longer than the time between the appearance of the first successful type 1 cell and that of the first type 2 cell within the lineage starting from it. However, this assumption does not hold if the population size is very large. If the population size is very large, the time of the spreading of type 1 cells is similar or even longer than the waiting time for the first type 1 cell whose lineage will eventually gives rise to a type 2 mutant. If the spreading time is much longer, the formulas for large population sizes are valid (broken lines labeled L in the figures).

The branching process formula (Eq. (5)) gives an accurate prediction for both intermediate and large populations (see solid curves labeled B). For small populations, however, this formula is inaccurate because the branching process calculation neglects the fixation of type 1 cells. We have an alternative formula Eq. (14) based on recursion, derived by Iwasa et al. (2004), which is valid for small and intermediate populations, (see solid curves labeled V).

Thus, the kinetic laws of cancer initiation and progression strongly depend on the population size, the fitness values of mutant cells, and mutation rates.

Acknowledgments

This work was done during Y.I.'s stay at the Institute for Advanced Study, Princeton, in 2002–2003 and also during his visit to the Program for Evolutionary Dynamics, Harvard University, in 2003. The Program for Evolutionary Dynamics in supported by Jeffrey E. Epstein.

References

Anderson, A.R., Chaplain, M.A., 1998. Continuous and discrete mathematical models of tumor-induced angiogenesis. Bull. Math. Biol. 60, 857–899.

Armitage, P., Doll, R., 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br. J. Cancer 8, 1–12.

Armitage, P., Doll, R., 1957. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. Br. J. Cancer 11, 161–169.

Bishop, J.M., 1983. Cellular oncogenes and retroviruses. Annu. Rev. Biochem. 52, 301–354.

Chang, S., Khoo, C., DePinho, R.A., 2001. Modeling chromosomal instability and epithelial carcinogenesis in the telomerase-deficient mouse. Semin. Cancer. Biol. 11, 227–239.

Chang, S., Khoo, C., Naylor, M.L., Maser, R.S., DePinho, R.A., 2003. Telomere-based crisis: functional differences between telomerase activation and ALT in tumor progression. Genes Dev 17, 88–100.

Fisher, J.C., 1958. Multiple-mutation theory of carcinogenesis. Nature 181, 651–652.

Frank, S.A., 2004. Age-specific acceleration of cancer. Curr. Biol. 14, 242–246.

Friend, S.H., Bernards, R., Rogeli, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M., Dryja, T.P., 1986. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature 323, 643–646.

Gatenby, R.A., Vincent, T.L., 2003. An evolutionary model of carcinogenesis. Cancer Res 63, 6212–6220.

Goldie, J.H., Coldman, A.J., 1979. A mathematic model for relating the drug sensitivity of tumors to their spontaneous mutation rate. Cancer Treat. Rep. 63, 1727–1733.

Goldie, J.H., Coldman, A.J., 1983. Quantitative model for multiple levels of drug resistance in clinical tumors. Cancer Treat. Rep. 67, 923–931.

Grist, S.A., McCarron, M., Kutlaca, A., Turner, D.R., Morley, A.A., 1992. In vivo human somatic mutation: frequency and spectrum with age. Mutat. Res. 266, 189–196.

Iwasa, Y., Michor, F., Nowak, M.A., 2004. Stochastic tunnels in evolutionary dynamics. Genetics 166, 1571–1579.

Kinzler, K.W., Nilbert, M.C., Vogelstein, B., Bryan, T.M., Levy, D.B., Smith, K.J., Preisinger, A.C., Hamilton, S.R., Hedge, P., Markham, A., Carlson, M., Joslyn, G., Groden, J., White, R., Miki, Y., Miyoshi, Y., Nishisho, I., Nakamura, Y., 1991. Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. Science 251, 1366–1370.

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

Knudson, A.G., 1993. Antioncogenes and human cancer. Proc. Natl Acad. Sci. USA 90, 10914–10921.

Komarova, N.L., Wodarz, D., 2003. Evolutionary dynamics of mutator phenotypes in cancer: implications for chemotherapy. Cancer Res 63, 6635–6642.

Komarova, N.L., Wodarz, D., 2004. The optimal rate of chromosome loss for the inactivation of tumor suppressor genes in cancer. Proc. Natl. Acad. Sci. USA 101, 7017–7021.

Komarova, N.L., Sengupta, A., Nowak, M.A., 2003. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. J. Theor. Biol. 223, 433–450.

Lengauer, C., Kinzler, K.W., Vogelstein, B., 1998. Genetic instabilities in human cancers. Nature 396, 623–649.

Levine, A.J., 1993. The tumor suppressor genes. Annu. Rev. Biochem. 62, 623–651.

Luebeck, E.G., Moolgavkar, S.H., 2002. Multistage carcinogenesis and the incidence of colorectal cancer. Proc. Natl Acad. Sci. USA 99, 15095–15100.

Maser, R.S., DePinho, R.A., 2002. Connecting chromosomes, crisis, and cancer. Science 297, 565–569.

Michor, F., Iwasa, Y., Komarova, N.L., Nowak, M.A., 2003. Local regulation of homeostasis favors chromosomal instability. Curr. Biol. 13, 581–584.

Michor, F., Iwasa, Y., Rajagopalan, H., Lengauer, C., Nowak, M.A., 2004a. Linear model of colon cancer initiation. Cell Cycle 3, 358–362.

Michor, F., Iwasa, Y., Nowak, M.A., 2004b. Dynamics of cancer progression. Nature Rev. Cancer 4, 197–206.

Moolgavkar, S.H., Knudson, A.G., 1981. Mutation and cancer: a model for human carcinogenesis. J. Natl Cancer Inst. 66, 1037–1052.

Moran, P., 1962. The statistical processes of evolutionary theory (Clarendon Press).

Nordling, C.O., 1953. A new theory on cancer-inducing mechanism. Br. J. Cancer 7, 68–72.

Nowak, M.A., Komarova, N.L., Sengupta, A., Jallepalli, P.V., Shih, I.-M., Vogelstein, B., Lengauer, C., 2002. The role of chromosomal instability in tumor initiation. Proc. Natl Acad. Sci. USA 99, 16226–16231.

Nowak, M.A., Michor, F., Komarova, N.L., Iwasa, Y., 2004. Evolutionary dynamics of tumor suppressor gene inactivation. Proc. Natl Acad. Sci. USA 101, 10635–10638.

Nunney, L., 1999. Lineage selection and the evolution of multistage carcinogenesis. Proc. R. Soc. London B 266, 493–498.

Otsuka, K., Suzuki, T., Shibata, H., Kato, S., Sakayori, M., Shimodaira, H., Kanamaru, R., Ishioka, C., 2003. Analysis of the human APC mutation spectrum in a *Saccharomyces cerevisiae* strain with a mismatch repair defect. Int. J. Cancer 103, 624–630.

Owen, M.R., Sherratt, J.A., 1999. Mathematical modelling of macrophage dynamics in tumours. Math. Models Methods Appl. Sci. 9, 513–539.

Rajagopalan, H., Nowak, M.A., Vogelstein, B., Lengauer, C., 2003. The significance of unstable chromosomes in colorectal cancer. Nature Rev. Cancer 3, 695–701.

Strauss, B.S., 1998. Hypermutability in carcinogenesis. Genetics 148, 1619–1626.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H., Godelle, B., 1997. Role of mutator alleles in adaptive evolution. Nature 387, 700–702.

Tomlinson, I., Sasieni, P., Bodmer, W., 2002. How many mutations in a cancer? Am. J. Pathol. 160, 755–758.

Vogelstein, B., Kinzler, K.W., 2001. The genetic basis of human cancer. McGraw-Hill, Toronto.

Weinberg, R.A., 1991. Tumor suppressor genes. Science 254, 1138–1146.

Wodarz, D., Krakauer, D.C., 2001. Genetic instability and the evolution of angiogenic tumor cell lines. Oncol. Rep. 8, 1195–1201.