The Evolution of Two Mutations During Clonal Expansion

Hiroshi Haeno,* Yoh Iwasa* and Franziska Michor[†,1]

*Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka 812-8581, Japan and †Society of Fellows, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT

Knudson's two-hit hypothesis proposes that two genetic changes in the RB1 gene are the rate-limiting steps of retinoblastoma. In the inherited form of this childhood eye cancer, only one mutation emerges during somatic cell divisions while in sporadic cases, both alleles of RB1 are inactivated in the growing retina. Sporadic retinoblastoma serves as an example of a situation in which two mutations are accumulated during clonal expansion of a cell population. Other examples include evolution of resistance against anticancer combination therapy and inactivation of both alleles of a metastasis-suppressor gene during tumor growth. In this article, we consider an exponentially growing population of cells that must evolve two mutations to (i) evade treatment, (ii) make a step toward (invasive) cancer, or (iii) display a disease phenotype. We calculate the probability that the population has evolved both mutations before it reaches a certain size. This probability depends on the rates at which the two mutations arise; the growth and death rates of cells carrying none, one, or both mutations; and the size the cell population reaches. Further, we develop a formula for the expected number of cells carrying both mutations when the final population size is reached. Our theory establishes an understanding of the dynamics of two mutations during clonal expansion.

THE concept of a tumor-suppressor gene originated from a statistical analysis of retinoblastoma incidence (KNUDSON 1971). This and later work (MOOLGAVKAR and KNUDSON 1981; FRIEND *et al.* 1986; VOGELSTEIN and KINZLER 2002) led to Knudson's two-hit hypothesis suggesting that retinoblastoma develops due to the inactivation of both alleles of the RB1 gene. The inherited form of the disease results from a germ-line mutation in one allele followed by inactivation of the second allele during somatic cell divisions. In sporadic retinoblastoma, both mutations arise during retina development. An understanding of the tumorigenesis of sporadic retinoblastoma requires the study of the dynamics of RB1 mutations during the growth of the retina. What is the chance that both alleles have been inactivated before the retina reaches its final size? How does this probability scale with mutation rates and cell turnover? And how large is a retinoblastoma tumor expected to be?

Tumor metastasis is a significant contributor to death in cancer patients. Metastases arise when cancer cells leave the primary tumor site and form new colonies elsewhere (CHAMBERS *et al.* 2002). Metastasis formation can be driven by genetic alteration of many genes, including activation of oncogenes like RAS and MYC (POZZATTI *et al.* 1986; WYLLIE *et al.* 1987) and inactivation of metastasis-suppressor genes such as NM23 (STEEG *et al.* 1988; STEEG 2004). Metastasis suppressors maintain the normal, noninvasive state of cells, and their inactivation promotes metastasis formation. It is clinically relevant to know whether a growing tumor has already inactivated a metastasis suppressor when it is diagnosed. What is the probability that a clonally expanding tumor cell population has accumulated two mutations in a metastasis suppressor gene before detection? How many metastasis-enabled cells does such a tumor contain?

Acquired drug resistance is a threat for the successful treatment of cancer (GOTTESMAN 2002). Depending on therapy, the type of cancer, and its stage, one or several (epi)genetic alterations are necessary to confer resistance to treatment. Some mechanisms of resistance require two genetic alterations—either because of haplosufficiency of a gene such that one recessive mutation cannot confer resistance or because of the use of combination therapy that targets two different positions in the cancer genome. Examples of the former are inactivation of p53, ATM, and RB1 (LOWE *et al.* 1994; VOLM and STAMMLER 1996; WESTPHAL *et al.* 1997). An example of the latter is emerging resistance of chronic myeloid leukemia cells against combination therapy with imatinib (Gleevec, STI571) and dasatinib (BMS-35482) (SHAH *et al.* 2004, 2007). Although both agents target the BCR–ABL kinase domain, the spectra of mutations conferring resistance to these drugs do not overlap—apart from one point mutation, which causes resistance to both (TOKARSKI *et al.* 2006). It is important to know whether patients already have resistant cells at diagnosis because this determines treatment strategies.

These examples lead to the following two questions: (i) What is the probability that an exponentially

expanding cell population evolves two mutations before reaching a certain size?, and (ii) What is the expected number of such cells at that time? In this article, we study the dynamics of two mutations emerging in a growing population of cells. Earlier, we analyzed the dynamics of one mutation arising during clonal expansion (IWASA *et al.* 2006) as well as the dynamics of two mutations emerging in a population of constant size (MICHOR and IWASA 2006). Our studies are akin to Luria and Delbrück's investigation of the mutations conferring bacterial resistance to phages (LURIA and DELBRÜCK 1943). The distribution of mutants in an exponentially growing population is known as the Luria–Delbrück distribution and has been studied extensively (TLSTY *et al.* 1989; ZHENG 1999; FRANK 2003). These studies are based on pure birth processes and neglect the possibility of cell death. In most situations in cancer, disease, and development, however, cell death does occur. Therefore, we introduce a birth-and-death model and calculate the probability that both mutations have arisen once the population reaches its final size, as well as the expected number of such cells at that time. This work is part of a growing effort to study cancer with mathematical techniques (NORDLING 1953; ARMITAGE and DOLL 1954, 1957; FISHER 1959; GOLDIE and COLDMAN 1979; MOOLGAVKAR and KNUDSON 1981; LUEBECK and MOOLGAVKAR 2002; MICHOR *et al.* 2004, 2005; WODARZ and KOMAROVA 2005; IWASA *et al.* 2006; MICHOR and IWASA 2006).

THE MODEL

Consider an exponentially growing population of cells starting from a single cell that is wild type at both genomic positions of interest. Cells wild type at both positions are called type-0 cells, and they give rise to cells harboring a mutation at one position with probability u_1 per cell division. Cells carrying one mutation are called type-1 cells, and they produce cells harboring mutations at both positions with probability u_2 per cell division. Such cells are called type-2 cells (Figure 1). The term "mutation" is collectively used to describe point mutations, insertions, deletions, inversions, translocations, loss of heterozygosity, and other genetic aberrations that can occur during one cell division. We focus on only two genomic positions and consider the mutational background as constant across cell types.

The cell population follows a continuous-time branching process. Denote the growth rates of type-0, type-1, and type-2 cells by r, a_1, and a_2 and their death rates by d, b_1, and b_2. If $a_1 > r$, then the first mutation is advantageous and increases the fitness of the cell; if $a_1 = r$, then the first mutation is neutral and does not change the fitness; and if $a_1 < r$, then the mutation is disadvantageous and decreases the fitness of the cell. Similar comparisons apply to the fitness of type-2 cells. Detection
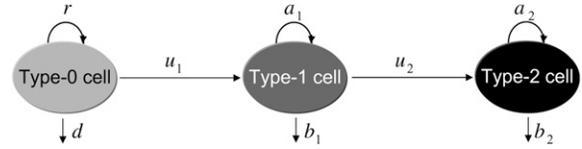

FIGURE 1.—Evolution of two mutations. A cell wild type at both positions is called a type-0 cell; it divides at rate r and dies at rate d. A cell carrying one mutation is called a type-1 cell. It arises by mutation of a type-0 cell at rate u_1 per cell division; it divides at rate a_1 and dies at rate b_1. A cell carrying two mutations is called a type-2 cell. It arises by mutation of a type-1 cell at rate u_2 per cell division; it divides at rate a_2 and dies at rate b_2.

occurs once the total population size—the sum of the number of type-0, type-1, and type-2 cells—reaches size M.

Computer simulations: We perform exact computer simulations of the stochastic process. There are three types of cells: type-0, type-1, and type-2 cells. Their respective numbers are denoted by x, y, and z. A change in x, y, and z occurs by cell division (possibly with mutation) or by cell death. Initially, there is one type-0 cell, $x = 1$, and no mutant cell, $y = z = 0$.

The stochastic simulation is performed by first determining which of the possible events (production or death of a type-0, a type-1, or a type-2 cell) is likely to occur first; the chance of each event to be first is proportional to its rate normalized by the sum of the rates of all possible events. Let us denote this sum by R. Then the timing of the first event is given by a negative exponential distribution with mean $1/R$. The process is continued either until all cells go extinct, $x = y = z = 0$, or until the total cell number reaches the final size, $x + y + z = M$. The transition probabilities between states are determined as follows. The number of type-0 cells increases if a type-0 cell divides without mutating. Hence the probability that the number of type-0 cells increases by one is given by

$$\Pr[(x, y, z) \to (x + 1, y, z)] = rx(1 - u_1)/\Gamma, \qquad (1a)$$

where $\Gamma = (r + d)x + (a_1 + b_1)y + (a_2 + b_2)z$. The number of type-1 cells increases by mutation of a type-0 cell or by division of a type-1 cell without mutation. The probability that the number of type-1 cells increases by one is given by

$$\Pr[(x, y, z) \to (x, y + 1, z)] = (rxu_1 + a_1y(1 - u_2))/\Gamma. \qquad (1b)$$

The number of type-2 cells increases by mutation of a type-1 cell or by division of a type-2 cell. We neglect the possibility of two mutations emerging during one cell division because this probability is sufficiently small. Hence the probability that the number of type-2 cells increases by one is given by

$$\Pr[(x, y, z) \to (x, y, z + 1)] = (a_1yu_2 + a_2z)/\Gamma. \qquad (1c)$$

The probabilities that the numbers of type-0, type-1, and type-2 cells decrease by one are given by

$$\Pr[(x, y, z) \rightarrow (x - 1, y, z)] = dx/\Gamma$$
$$\Pr[(x, y, z) \rightarrow (x, y - 1, z)] = b_1 y/\Gamma$$
$$\Pr[(x, y, z) \rightarrow (x, y, z - 1)] = b_2 z/\Gamma. \quad (1d)$$

For each parameter set, we perform many independent runs of the stochastic process to account for random fluctuations and count the fraction of runs that reach the final size, M, and have produced at least one type-2 cell. We also record the number of type-2 cells in those runs.

The probability of two mutations: Let us now derive an analytic expression for the probability that an exponentially growing population, starting from one type-0 cell, has produced at least one type-2 cell until the total population size reaches M. For simplicity, we assume that the death rates are constant across cell types, $d = b_1 = b_2$.

Branching-process formula: Let us first use a multistate branching process to calculate the probability of two mutations. This calculation is based on the assumption that the number of type-1 and type-2 cells is much smaller than the number of type-0 cells; hence we adopt the approximation that the final population size is reached once the number of type-0 cells becomes M. This approach represents an extension of earlier work (IWASA *et al.* 2006). The calculation is subdivided into two parts: (i) considering the number of type-1 cells produced from the exponentially expanding population of type-0 cells and (ii) studying the behavior of a cell lineage originating from a single type-1 cell. The generating function of the total number of type-2 cells is given by

$$G(s_1, s_2) = \exp\left[-\sum_{x=1}^{M-1} R_x(1 - g_x^1(s_1, s_2))\right]. \quad (2)$$

Here $g_x^1(s_1, s_2)$ is the generating function of a lineage starting from a single type-1 cell, and R_x is the expected number of newly created type-1 cells when there are x type-0 cells; it is given by $R_x \approx u_1/(1 - d/r)$ (IWASA *et al.* 2006).

The generating function, Equation 2, can be used to calculate important quantities. The probability that there are no type-2 cells—irrespective of the number of type-1 cells—once the total population size reaches M is given by

$$G(1, 0) = \exp\left[-\frac{u_1}{1 - d/r}\sum_{x=1}^{M-1}(1 - g_x^1(1, 0))\right]. \quad (3)$$

Hence the probability that there is at least one type-2 cell when the population reaches its final size is given by

$$P = 1 - G(1, 0) = 1 - \exp\left[-\frac{u_1}{1 - d/r}\sum_{x=1}^{M-1}(1 - g_x^1(1, 0))\right]. \quad (4)$$

APPENDIX A contains the derivation of $g_x^1(1, 0)$ and the derivation of the probability of two mutations in the neutral case, $r = a_1 = a_2$, which is given by

$$P = \frac{M u_1 u_2 \ln(1/u_2)}{(1 - d/r)^2}. \quad (5)$$

Let us now compare Equation 5 with the direct computer simulation. Figure 2 shows that the prediction of Equation 5 (solid curve) is a slight overestimation of the probability of two mutations represented by the results of the direct computer simulation (open circles). The formula for deleterious mutations is more accurate, while the formula for advantageous mutations gives a significant overestimation (data not shown). The deviation stems from the assumption of the analytic solution that the final population size is reached once there are M type-0 cells, rather than once the sum of type-0, type-1, and type-2 cells reaches M. This assumption leads to accurate predictions of the probability of having type-1 cells (IWASA *et al.* 2006), but to an overestimation of the probability of having type-2 cells—particularly if type-1 cells have a fitness advantage and can reach significant fractions of the final population size. Therefore, we consider a different analytic approach in the following.

Alternative formula: Let us consider the two steps required for the emergence of type-2 cells: (i) production of a type-1 cell and the survival of its lineage and (ii) production of a type-2 cell and its persistence. Denote by P_x the probability that the first step occurs when there are x type-0 cells. With $\beta = (1 - b_1/a_1)u_1/(1 - d/r)$, we have

$$P_x = e^{-\beta(x-1)}(1 - e^{-\beta}). \quad (6)$$

See APPENDIX B for the derivation of P_x. The population of cells now contains x type-0 cells and one (eventually successful) type-1 cell. It is in this lineage that a surviving type-2 cell emerges. A type-2 cell may arise by mutation during each division of a type-1 cell. The expected number of mutational events is proportional to the total number of cell divisions the type-1 cell lineage experiences until the final population size is reached. The total number of type-1 cell divisions is on average $y/(1 - b_1/a_1)$, where y is the abundance of type-1 cells at detection. Therefore the probability, Q_x, of producing a type-2 cell in a type-1 cell lineage is given by

$$Q_x = 1 - e^{-u_2 y/(1 - b_1/a_1)}. \quad (7)$$

Here $y = \exp[(a_1 - b_1)\tau_x]$, and τ_x is the time between the emergence of a type-1 cell from x type-0 cells and when the total number of type-0 and type-1 cells reaches M. See APPENDIX B for the formula of τ_x and the derivation of Q_x.

With these results, the probability that there is at least one type-2 cell when the total population size reaches M is given by
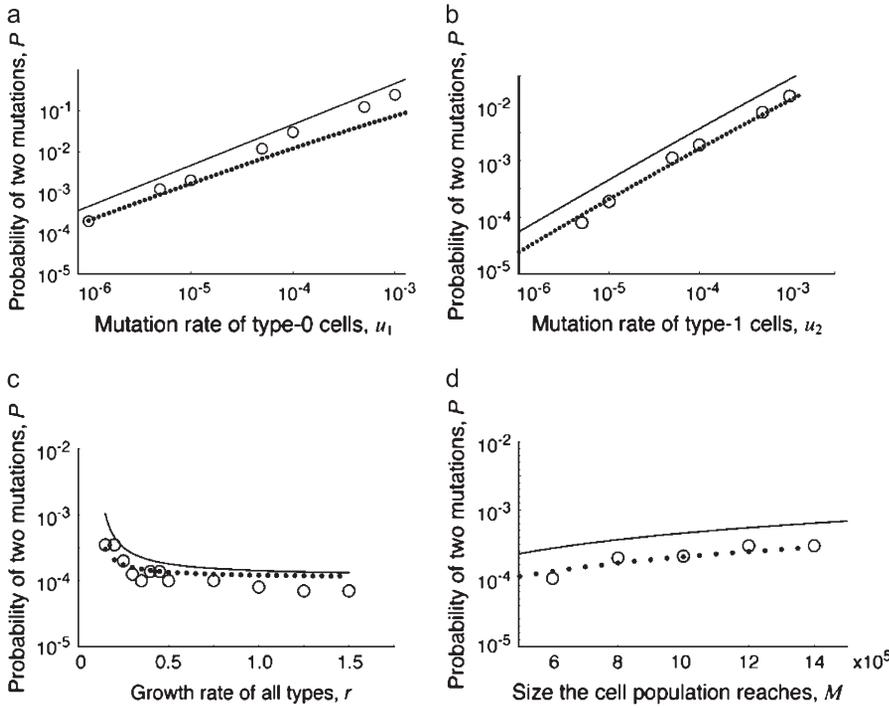
Figure 2.—The probability of two neutral mutations ($r = a_1 = a_2$). The probability of having produced at least one type-2 cell once the population reaches size M in dependence of the mutation rate of type-0 cells, u_1 (a), the mutation rate of type-1 cells, u_2 (b), the growth rate of all cell types, $r = a_1 = a_2$ (c), and the final population size, M (d), is shown. The solid curve shows the prediction of Equation 5, the open circles show the results of the exact computer simulation, and the dots show the prediction of the alternative formula, Equation 9. Parameter values are $d = b_1 = b_2 = 0.1$ and (a) $r = a_1 = a_2 = 0.2$, $u_2 = 10^{-5}$, and $M = 10^6$; (b) $r = a_1 = a_2 = 0.2$, $u_1 = 10^{-6}$, and $M = 10^6$; (c) $u_1 = 10^{-6}$, $u_2 = 10^{-5}$, and $M = 10^6$; and (d) $r = a_1 = a_2 = 0.2$, $u_1 = 10^{-6}$, and $u_2 = 10^{-5}$.

$$P = \sum_{x=1}^{M-1} P_x \cdot Q_x = \sum_{x=1}^{M-1} e^{-\beta(x-1)}(1 - e^{-\beta})\left(1 - \exp\left[-\frac{u_2 y}{1 - b_1/a_1}\right]\right). \quad (8)$$

This expression holds if the death rates of type-2 cells can be neglected as compared to the growth rates. If the mortality of type-2 cells cannot be neglected, there is an additional factor in the equation (see APPENDIX B). In the neutral case ($r = a_1 = a_2$), we can simplify Equation 8 as

$$P = \sum_{x=1}^{M-1} e^{-u_1(x-1)}(1 - e^{-u_1})\left(1 - \exp\left[-\frac{Mu_2}{(1 - d/r)(x+1)}\right]\right). \quad (9)$$

Let us now compare Equation 8 with the stochastic computer simulation. Figure 3 shows the agreement between the predictions of the formula (dots) and the results of the simulation (open circles). The model includes seven parameters: the population size at detection (M), the probability of mutating either genomic position per cell division (u_1 and u_2), the growth rate of type-0, type-1, and type-2 cells (r, a_1, and a_2), and the death rate (d). If the four rates r, a_1, a_2, and d are multiplied by the same factor, then the whole process proceeds faster but the probability P remains constant. Hence P is determined by the three ratios a_1/r, a_2/r, and d/r rather than by the four rates independently. In the following analysis of the parameter dependence, we examine the case with $r = 1$ (without loss of generality) and consider the formula in which a_1, a_2, and d are replaced by a_1/r, a_2/r, and d/r, respectively:

i. Probability of mutating the first position, u_1: As shown in Figure 3a, the probability of two muta-

tions, P, increases with u_1. High values of u_1 increase the risk of producing a type-1 cell lineage that can eventually give rise to type-2 cells.

ii. Probability of mutating the second position, u_2: As shown in Figure 3b, the probability of two mutations, P, increases with u_2. High values of u_2 increase the chance that type-2 cells emerge.

iii. Relative growth rate of type-1 cells, a_1/r: As shown in Figure 3, a and b, the probability of two mutations, P, increases with a_1/r. A larger growth rate of type-1 cells relative to that of type-0 cells enhances the chance of two mutations because type-1 cells can reach higher frequencies.

iv. Relative growth rate of type-2 cells, a_2/r: The probability of two mutations, P, is almost independent of the relative growth rate of type-2 cells, a_2/r (data not shown). This effect emerges because here we focus on the existence of type-2 cells rather than on their abundance. The growth rate of type-2 cells does not significantly affect the probability of successfully establishing a lineage from a single cell; as long as the growth rate is clearly greater than the death rate, the newly produced mutant will almost certainly survive. Once a lineage of type-2 cells has emerged, the growth rate of type-2 cells has no effect on the chance of their existence once the population reaches its final size—it does, however, have an important effect on their abundance, as discussed later.

v. Relative death rate, d/r: As shown in Figure 3c, the probability of two mutations, P, increases with d/r. A large death rate prolongs the time it takes until the total population reaches size M. It increases the
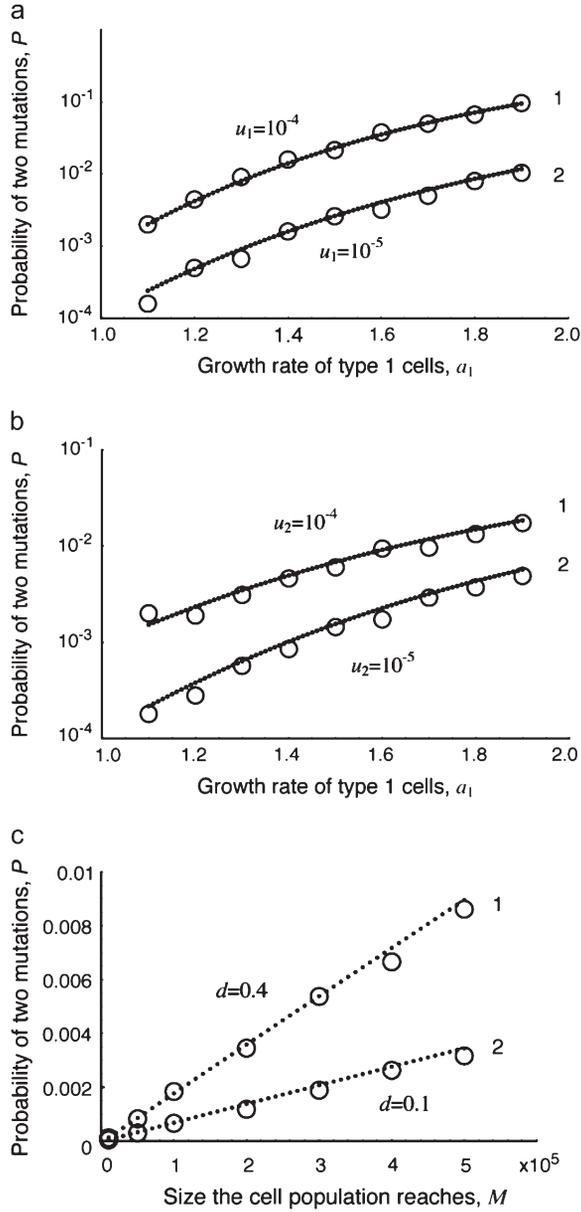
FIGURE 3.—The probability of two mutations. The probability of having produced at least one cell mutated in both positions when the cell population reaches size M in dependence of the growth rate of cells mutated in one position, a_1 (a and b), and the final population size, M (c), is shown. Equation 8 is shown as dots and the direct computer simulation (system 1) as circles. Parameter values are (a) $r = 1$, $M = 10^6$, $a_2 = 1.1a_1$, $d = b_1 = b_2 = 0.1$, $u_2 = 10^{-6}$, $u_1 = 10^{-4}$ (line 1), and $u_1 = 10^{-5}$ (line 2); (b) $u_1 = 10^{-6}$, $u_2 = 10^{-4}$ (line 1), and $u_2 = 10^{-5}$ (line 2); and (c) $u_1 = 10^{-6}$, $u_2 = 10^{-4}$, $a_1 = 1.5$, $a_2 = 1.65$, $d = b_1 = b_2 = 0.4$ (line 1), and $d = b_1 = b_2 = 0.1$ (line 2).

number of cell divisions and therefore enhances the chance of accumulating mutations.

vi. Final population size, M: As shown in Figure 3c, the probability of two mutations, P, increases almost linearly with M. This effect emphasizes the importance of detecting a growing population of cancer cells as early as possible to reduce the risk of having evolved two mutations.

We performed a regression analysis of Equation 8 to analyze the sensitivity of the probability P with respect to the parameters. Using many different parameter sets, we obtained the following regression formula:

$$\log P = 0.80 \log u_1 + 0.76 \log u_2 + 6.4 \log\left(\frac{a_1}{r}\right)$$

$$+ 1.2 \log\left(\frac{d}{r}\right) + 0.94 \log M + 0.11. \qquad (10)$$

The ratio of the results from the regression formula to that from Equation 8 ranges from -0.3 to 0.3 in logarithmic units. The formula indicates that every parameter except a_2/r (the relative growth rate of type-2 cells) affects the probability P. Note that the coefficient for $\log(a_1/r)$ is the largest among all terms, implying that an increase in the growth rate of type-1 cells most effectively enhances the risk of two mutations. Note also that the coefficient for $\log(d/r)$ is positive, signifying that a higher death rate increases the risk. A larger death rate necessitates a larger number of cell divisions to reach a certain population size, and hence the chance of mutations increases. Finally, Equation 10 indicates that P decreases with the growth rate of type-0 cells, r. Hence a large growth rate of type-0 cells reduces the risk of two mutations—fast-growing tumors are more likely to be sensitive to treatment than slowly growing ones.

The expected number of cells with two mutations: Let us now calculate the expected mean number of type-2 cells once the population reaches its final size.

Branching-process formula: The mean number of type-2 cells in a lineage starting from a single type-1 cell is calculated from Equations 2 and A1 by taking the derivative with respect to s_2 and setting $s_1 = s_2 = 1$. According to the calculation in APPENDIX C, we have

$$\frac{\partial g^1}{\partial s_2}(1, 1, t) = \frac{a_1 u_2}{\delta - \gamma}(e^{\delta t} - e^{\gamma t}), \qquad (11)$$

with $\gamma = a_2 - b_2$ and $\delta = a_1(1 - u_2) - b_1$. In the neutral case, $a_1 = a_2 = r$ and $b_1 = b_2 = d$, the difference between γ and δ becomes very small, $\gamma - \delta = ru_2 \ll 1$, and we can approximate Equation 11 by

$$\frac{\partial g^1}{\partial s_2}(1, 1, t) = ru_2 t e^{(r-d)t}. \qquad (12)$$

The duration between the time when there are x type-0 cells and the time when the final population size is reached is given by $t \approx \ln(M/x)/(r - d)$ (IWASA *et al.* 2006). Then we have

$$\frac{\partial g^1}{\partial s_2}\left(1, 1, \frac{1}{r-d}\ln(M/x)\right) = \frac{ru_2}{r-d}\frac{M}{x}\ln\left(\frac{M}{x}\right). \qquad (13)$$

The expected number of type-2 cells once the population reaches its final size is given by the derivative of

TABLE 1

Comparison of Equation 15 with the computer simulation in the neutral case

	Standard[a]	u_1	u_2	M	d/r
		Parameter enhanced by 50%			
Computer simulation	15.5	17.6	18.0	19.0	31.4
Equation 15	15.4	16.6	17.3	17.5	25.7

[a] The standard parameter set is given by $u_1 = 10^{-4}$, $u_2 = 10^{-4}$, $M = 10^6$, and $d/r = 0.5$. Each specified parameter is enhanced by 50% while all the other parameters remain constant. Equation 15 gives good predictions of the expected number of type-2 cells in the neutral case.

Equation 2 with respect to s_2 evaluated at $s_1 = s_2 = 1$ and $t \approx \ln(M/x)/(r - d)$, which is

$$\begin{bmatrix} \text{mean no. of} \\ \text{type-2 cells} \end{bmatrix}$$
$$= \frac{u_1}{1 - d/r} \sum_{x=1}^{M-1} \frac{\partial g^1}{\partial s_2} = \frac{u_1 u_2}{(1 - d/r)^2} \sum_{x=1}^{M-1} \frac{M}{x} \ln\left(\frac{M}{x}\right). \tag{14}$$

The mean number of type-2 cells among those with at least one type-2 cell is given by Equation 14 divided by the probability to have any type-2 cells, P:

$$\bar{Z} = \frac{u_1 u_2}{P(1 - d/r)^2} \sum_{x=1}^{M-1} \frac{M}{x} \ln\left(\frac{M}{x}\right). \tag{15}$$

It can be seen from Equation 14 that the expected number of type-2 cells decreases with the ratio r/d. When r is close to d (and greater than one), the number of type-2 cells is very large. When the ratio r/d becomes very large, the cell number converges to an asymptote independent of r and d. A formula for the expected

number of cells in the nonneutral case can be derived similarly (see APPENDIX C).

A comparison between the formula and the direct computer simulation shows that the formula is accurate for neutral mutations (Table 1), but tends to overestimate the number of type-2 cells when the mutations are advantageous (data not shown) for the same reason as above. Therefore, we again consider an alternative approach.

Alternative formula: Note that we have already derived the probability of type-2 cells (P) and the number of type-1 cells (y) once the final population size is reached, as well as the length of time between the emergence of type-1 cells and when the sum of type-0 and type-1 cells reaches M (τ_x). Denote by Z the number of type-2 cells once the final size is reached. If we adopt the assumption that type-2 cells experience deterministic growth, then the number Z is determined by the time of emergence of the first type-2 cell. Hence the probability that the number of type-2 cells, Z, is between z_1 and z_2 equals the probability that the first type-2 cells emerges between times t_1 and t_2 (for a schematic explanation, see Figure 4). Then we obtain a formula for $\Pr[z_1 \leq Z < z_2]$ and can further derive the mean number of type-2 cells, $\bar{Z}$ (see APPENDIX D).

The probability distribution of the number of type-2 cells is shown in Figure 5. The prediction of Equation D4 (red line) is in good agreement with the results of the computer simulation (black dots, see APPENDIX D for details). Table 2 shows the parameter dependence of the expected number of type-2 cells by comparing the standard parameter set with sets in which one of the six parameters u_1, u_2, M, a_1/r, a_2/r, and d/r is enhanced by 50%. For each parameter set, we performed >100,000 runs and calculated the conditional mean number of type-2 cells. This number increases with the enhancement of each parameter except the growth rate of type-1
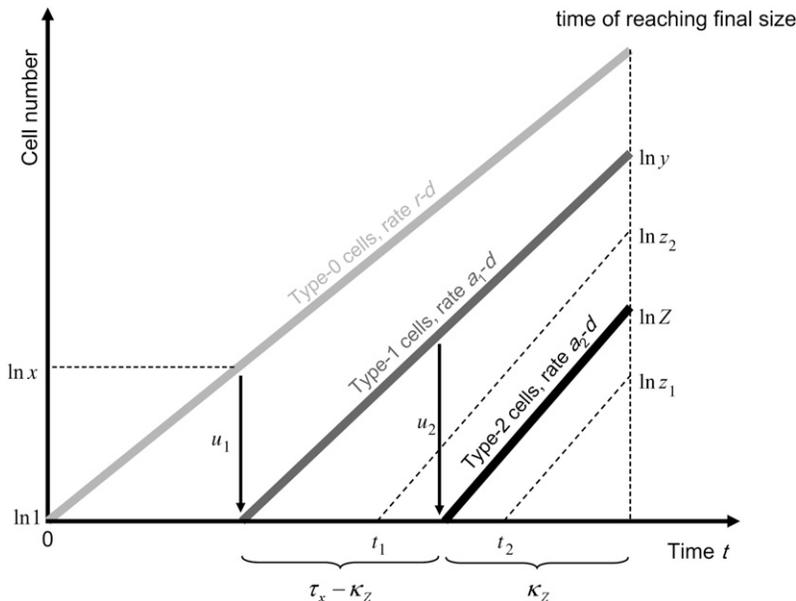

FIGURE 4.—Schematic representation of how to calculate the expected number of type-2 cells. The horizontal axis shows the time since initiating growth with one type-0 cell. The vertical axis shows the logarithmic number of the three cell types. Type-0, type-1, and type-2 cells undergo exponential growth at rates $r - d$, $a_1 - d$, and $a_2 - d$, respectively, which appear as three lines with different slopes. A mutation is created at a Poisson rate, as indicated by the two arrows. It takes τ_x time units for a lineage of type-1 cells to reach y and κ_Z time units for a lineage of type-2 cells to reach Z.
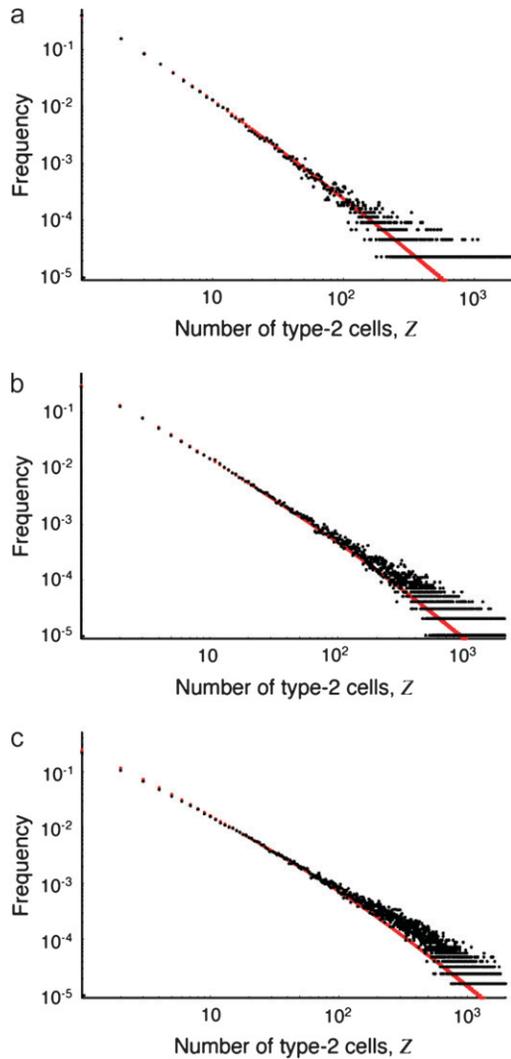
FIGURE 5.—The probability distribution of the number of type-2 cells. We show the probability distribution of type-2 cells for different values of the growth rate of type-1 cells, a_1. Red lines indicate the predictions of Equation D4 and black dots show the results of the stochastic computer simulation. The horizontal axis shows the number of type-2 cells, Z, and the vertical axis is $\Pr[Z = z]$, the probability that Z is equal to z. Parameter values are $r = 1.0$, $u_1 = 10^{-5}$, $u_2 = 10^{-4}$, $M = 10^6$, $d = 0.1$, and $a_2 = 1.1a_1$ and (a) $a_1 = 1.1$, (b) $a_1 = 1.5$, and (c) $a_1 = 1.9$.

cells, a_1/r; interestingly, the latter enhancement reduces the cell number while it increases the probability of having at least one type-2 cell. This effect emerges because the total population reaches size M before type-2 cells gain a significant frequency if the growth rate of type-1 cells is large. Therefore, the consequence of a large growth rate of type-1 cells is a high probability of producing type-2 cells, but a small number of such cells that may be difficult to observe. In contrast, a higher relative death rate (d/r) increases the number of type-2 cells because in that case, it takes longer for the total population to reach M and type-2 cells are likely to increase during that time. The most effective parameter

to influence the expected number of type-2 cells is their growth rate; this parameter, however, does not affect the probability of having type-2 cells.

DISCUSSION

In this article, we study the dynamics of two mutations arising in an exponentially growing population of cells. We calculate the probability of having produced at least one cell with two mutations once the population reaches a certain size, as well as the distribution of the number of such cells at that time. These results are important in many medical scenarios. For instance, chronic myeloid leukemia (CML) patients treated with imatinib and dasatinib readily evolve several resistance mutations (SHAH *et al.* 2007). A clinically relevant question is whether such compound resistance mutations are preexisting to diagnosis of CML, because this influences treatment choices. Our mathematical framework allows us to calculate the probability that a patient harbors an imatinib- and a dasatinib-resistant mutation, as well as the number of such cells, at the time of diagnosis (Table 3). We find that the chance is nonnegligible that both mutations emerge before diagnosis.

We have presented two approaches to calculating the probability of having two mutations once the total population reaches a certain size. One approach is based on a multistate branching process. This methodology can be used to derive an analytic expression for the probability of two mutations as well as the probability distribution of cells. It is very accurate for mutations that confer a fitness disadvantage to the cell or that are neutral as compared to wild-type cells. The second approach requires some numerical calculation, but is very accurate when the mutations confer a fitness advantage to the cell and when they are neutral. To provide mathematical solutions for all scenarios, we present both approaches in this article.

An investigation of the probability of two mutations informs us about the importance of particular parameter values in the process. The relative growth rate of cells harboring one mutation (type-1 cells) is decisive for the probability of eventually producing two mutations. This feature suggests experimentally determining the growth rates of type-0 cells (which carry no mutation) and type-1 cells to predict the presence of two mutations. If the ratio of the growth rate of type-1 cells to the growth rate of type-0 cells is high, then the probability of two mutations is large. Therefore, the abundance of type-1 cells is a good proxy for the presence of type-2 cells (which harbor both mutations). Furthermore, large death rates increase the chance of accumulating mutations because the risk increases with the number of cell divisions generating the cell population; a population of a particular size will contain many more mutants if the cell turnover is large. Hence tumors with high apoptosis rates are at particular risk of containing resistant cells. Last, the mutation rates themselves increase the probability of mutations, and hence

TABLE 2

The mean number of type-2 cells

	Standard[a]	Parameter enhanced by 50%					
		u_1	u_2	M	a_1/r	a_2/r	d/r
Mean no. of type-2 cells	82.0	90.7	112.5	131.1	41.9	1675.0	101.5

[a] The standard parameter set is given by $u_1 = 10^{-5}$, $u_2 = 10^{-4}$, $M = 10^6$, $a_1/r = 1.5$, $a_2/r = 1.65$, and $d/r = 0.1$. Each specified parameter is enhanced by 50% while all the other parameters remain constant. Enhancement of the relative growth rate of type-1 cells reduces the mean number of type-2 cells, but enhancement of each of the other parameters increases the mean number once the final population size is reached.

therapies that induce genetic instabilities or work by damaging DNA enhance the chance of treatment failure.

We have also investigated the probability distribution of the number of type-2 cells. The parameter that most effectively influences the number of those cells is their relative growth rate, which in turn does not affect the probability of having any type-2 cells. Interestingly, high relative growth rates of type-1 cells increase the probability that type-2 cells exist, but reduce their number. On the basis of our theory, the worst case is represented by a large relative growth rate of type-1 cells (which increases the probability that a type-2 cell emerges) and an even larger relative growth rate of type-2 cells (which ensures that such cells can grow). A situation requiring particular attention occurs when the growth rate of type-1 cells is

TABLE 3

The probability of a compound-resistant mutation

Cell turnover	Mutation rates		
	$u_1 = 50 \times 10^{-8}$, $u_2 = 10 \times 10^{-8}$	$u_1 = 50 \times 10^{-7}$, $u_2 = 10 \times 10^{-7}$	$u_1 = 50 \times 10^{-6}$, $u_2 = 10 \times 10^{-6}$
$d = 0.003$ $r = 0.008$ $a_1 = 0.010$ $a_2 = 0.012$	$P = 1.3 \times 10^{-6}$ $\bar{Z} = 44.6$	$P = 0.0001$ $\bar{Z} = 46.5$	$P = 0.0091$ $\bar{Z} = 60.7$
$d = 1.003$ $r = 1.008$ $a_1 = 1.010$ $a_2 = 1.012$	$P = 1.6 \times 10^{-6}$ $\bar{Z} = 44.7$	$P = 0.0002$ $\bar{Z} = 46.7$	$P = 0.0112$ $\bar{Z} = 61.3$
$d = 1.003$ $r = 1.008$ $a_1 = 1.016$ $a_2 = 1.032$	$P = 6.9 \times 10^{-5}$ $\bar{Z} = 774.5$	$P = 0.0063$ $\bar{Z} = 816.4$	$P = 0.3008$ $\bar{Z} = 1312.6$
$d = 1.003$ $r = 1.008$ $a_1 = 1.080$ $a_2 = 1.800$	$P = 0.0093$ $\bar{Z} = 8068.3$	$P = 0.2183$ $\bar{Z} = 8326.7$	$P = 0.9395$ $\bar{Z} = 9487.8$

We show the probability P that a patient harbors an imatinib- and a dasatinib-resistant mutation when diagnosed with chronic myeloid leukemia (CML) and the expected number of such cells, $\bar{Z}$, at diagnosis. The net growth rate of leukemic stem cells before diagnosis is about $(r - d) = 0.005$/day (MICHOR *et al.* 2005), and we provide several examples of growth and death rates. At diagnosis, there are $\sim M = 250,000$ leukemic stem cells (HOLYOAKE *et al.* 2002).

larger than the growth rates of both type-0 and type-2 cells, because then a small number of type-2 cells exist with high probability, but those cells will be difficult to detect.

So far, we have assumed that the mutation rates are constant throughout tumor growth. However, genomic instabilities lead to increasing rates of genetic changes and are a frequent property of tumors (LENGAUER *et al.* 1998). Figure 6 demonstrates how increasing mutation rates influence the evolutionary dynamics of two mutations. As compared to constant mutation rates, continuously increasing rates lead to elevated probabilities of harboring two mutations. However, the mean number of cells with two mutations decreases in such scenarios (Table 4). This effect emerges because large mutation rates enhance the production of cells carrying two mutations, particularly when the population size is large, and therefore lead to more patients harboring fewer such cells. Our finding emphasizes the danger of genetically unstable lesions for treatment outcomes, as well as the necessity to use anticancer drugs that do not increase genomic mutation rates.
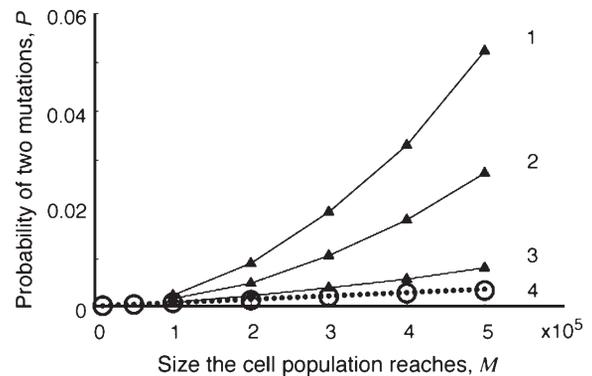

FIGURE 6.—The effect of increasing mutation rates on the probability of two mutations. The probability of having produced at least one type-2 cell once the population reaches size M in dependence of the final population size (M) is shown. We examine situations in which the mutation rates, u_1 and u_2, each increase by ε each 1000 times type-0 cells and type-1 cells divide, respectively. We show that the probability of two mutations increases with the value of ε. Line 4 represents the same results as line 2 in Figure 3c. The parameter values are $r = 1$, $u_1 = 10^{-6}$, $u_2 = 10^{-4}$, $a_1 = 1.5$, $a_2 = 1.65$, and $d = b_1 = b_2 = 0.1$ and $\varepsilon = 10^{-6}$ (line 1), $\varepsilon = 5 \times 10^{-7}$ (line 2), $\varepsilon = 10^{-7}$ (line 3), and $\varepsilon = 0$ (line 4).

TABLE 4

The effect of increasing mutation rates on the mean number of type-2 cells

	$\varepsilon = 0$	$\varepsilon = 10^{-7}$	$\varepsilon = 5 \times 10^{-7}$	$\varepsilon = 10^{-6}$
Mean no. of type-2 cells	70.9	28.5	11.3	8.3

The mean number of cells carrying two mutations for different enhancement factors of the mutation rates, ε, is shown. The mean number decreases as ε increases. Parameter values are $r = 1$, $u_1 = 10^{-6}$, $u_2 = 10^{-4}$, $a_1 = 1.5$, $a_2 = 1.65$, $d = b_1 = b_2 = 0.1$, and $M = 500,000$. We performed $>10,000$ runs in which at least one type-2 cell is present when M is reached.

This article increases our knowledge of the evolutionary dynamics of an exponentially expanding population. An important goal of the field is to generalize our algorithm to arbitrary mutation–selection networks such that the probability of having n mutations (emerging in a particular order) in a growing population can be calculated and the expected number of cells harboring these mutations can be predicted. Further, the calculation may be adapted to describe situations in which more than one offspring arises per division event; such scenarios arise in infectious diseases such as human immunodeficiency virus (HIV). Haeno and Iwasa (2007) consider the risk of drug resistance of an exponentially growing virus, assuming that infected cells give rise to a random number of virus particles per time interval. They derive a formula for the probability of one mutation at detection and show its implications for HIV primary infection. Our formulation of the probability of two mutations with regard to (cancer) cell division inspires an approach to calculate the probability of two mutations with regard to virus proliferation.

LITERATURE CITED

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

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

Chambers, A., A. Groom and I. MacDonald, 2002 Dissemination and growth of cancer cells in metastatic sites. Nat. Rev. Cancer **2:** 563–572.

Fisher, J. C., 1959 Multiple-mutation theory of carcinogenesis. Nature **181:** 651–652.

Frank, S. A., 2003 Somatic mosaicism and cancer: inference based on a conditional Luria-Delbrueck distribution. J. Theor. Biol. **223:** 405–412.

Friend, S. H., R. Bernards, S. Rogeli, R. A. Weinberg, J. M. Rapaport et al., 1986 A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature **323:** 643–646.

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

Gottesman, M. M., 2002 Mechanisms of cancer drug resistance. Annu. Rev. Med. **53:** 615–627.

Haeno, H., and Y. Iwasa, 2007 Probability of resistance evolution for exponentially growing virus in the host. J. Theor. Biol. **246:** 323–331.

Holyoake, T. L., X. Jiang, M. W. Drummond, A. C. Eaves and C. J. Eaves, 2002 Elucidating critical mechanisms of deregulated stem cell turnover in the chronic phase of chronic myeloid leukemia. Leukemia **16:** 549–558.

Iwasa, Y., M. A. Nowak and F. Michor, 2006 Evolution of resistance during clonal expansion. Genetics **172:** 2557–2566.

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

Lengauer, C., K. W. Kinzler and B. Vogelstein, 1998 Genetic instabilities of human cancers. Nature **396:** 623–649.

Lowe, S. W., S. Bodis, A. McClatchey, L. Remington, H. E. Ruley et al., 1994 p53 status and the efficacy of cancer therapy in vivo. Science **266:** 807–810.

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

Luria, S. E., and M. Delbrück, 1943 Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28:** 491–511.

Michor, F., and Y. Iwasa, 2006 Dynamics of metastasis suppressor gene inactivation. J. Theor. Biol. **241:** 676–689.

Michor, F., Y. Iwasa and M. A. Nowak, 2004 Dynamics of cancer progression. Nat. Rev. Cancer **4:** 197–205.

Michor, F., T. P. Hughes, Y. Iwasa, S. Branford, N. P. Shah et al., 2005 Dynamics of chronic myeloid leukemia. Nature **435:** 1267–1270.

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

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

Pozzatti, R., R. Muschel, J. Williams, R. Padmanabhan, B. Howard et al., 1986 Primary rat embryo cells transformed by one or two oncogenes show different metastatic potentials. Science **232:** 223–227.

Shah, N. P., C. Tran, F. Y. Lee, P. Chen, D. Norris et al., 2004 Overriding imatinib resistance with a novel ABL kinase inhibitor. Science **305:** 399–401.

Shah, N. P., B. J. Skaggs, S. Branford, T. P. Hughes, J. M. Nicoll et al., 2007 Sequential ABL kinase inhibitor therapy selects for compound drug-resistant BCR-ABL mutations with altered oncogenic potency. J. Clin. Invest. **117:** 2562–2569.

Steeg, P. S., 2004 Metastasis suppressor genes. J. Natl. Cancer Inst. **96:** E4.

Steeg, P. S., G. Bevilacqua, L. Kopper, U. P. Thorgeirsson, J. E. Talmadge et al., 1988 Evidence for a novel gene associated with low tumor metastatic potential. J. Natl. Cancer Inst. **80:** 200–204.

Tokarski, J. S., J. A. Newitt, C. Y. Chang, J. D. Cheng, M. Wittekind et al., 2006 The structure of dasatiuib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants. Cancer Res. **66:** 5790–5797.

Tlsty, T. D., B. H. Margolin and K. Lum, 1989 Difference in the rates of gene amplification in nontumorigenic and tumorigenic cell-lines as measure by Luria-Delbrück fluctuation analysis. Proc. Natl. Acad. Sci. USA **86:** 9441–9445.

Volm, M., and G. Stammler, 1996 Retinoblastoma (rb) protein expression and resistance in squamous cell lung carcinomas. Anticancer Res. **16:** 891–894.

Westphal, C. H., S. Rowan, C. Schmaltz, A. Elson, D. E. Fisher et al., 1997 ATM and p53 cooperate in apoptosis and suppression of tumorigenesis, but not in resistance to acute radiation toxicity. Nat. Genet. **16:** 397–401.

Wodarz, D., and N. L. Komarova, 2005 *Computational Biology of Cancer: Lecture Notes and Mathematical Modeling.* World Scientific Publishing, Hackensack, NJ.

Wyllie, A. H., K. A. Rose, C. M. Steel, R. G. M. Foster and D. A. Spandidos, 1987 Rodent fibroblast tumors expressing human myc and ras genes: growth, metastasis and endogenous oncogene expression. Br. J. Cancer **56:** 251–259.

Vogelstein, B., and K. W. Kinzler, 2002 *The Genetic Basis of Human Cancer.* McGraw-Hill, New York.

Zheng, Q., 1999 Progress of a half century in the study of the Luria-Delbrück distribution. Math. Biosci. **162:** 1–32.

APPENDIX A: DERIVATION OF THE BRANCHING-PROCESS FORMULA FOR THE PROBABILITY OF TWO MUTATIONS

Denote by $g^1(s_1, s_2, t)$ the generating function for a lineage starting from a single type-1 cell and by $g^2(s_1, s_2, t)$ the generating function for a lineage starting from a single type-2 cell. Their equations satisfy the recursive formulas

$$\frac{\partial}{\partial t} g^1 = b_1(1 - g^1) + a_1((1 - u_2)(g^1)^2 + u_2 g^2 g^1 - g^1), \tag{A1a}$$

$$\frac{\partial}{\partial t} g^2 = b_2(1 - g^2) + a_2((g^2)^2 - g^2), \tag{A1b}$$

with the initial conditions $g^1(s_1, s_2, 0) = s_1$ and $g^2(s_1, s_2, 0) = s_2$.

Let $g(t) = g^1(1, 0; t)$ be the probability that there is one type-1 cell and no type-2 cell at time t. Similarly, let $h(t) = g^2(1, 0; t)$ be the probability that there is no type-1 cell and one type-2 cell at time t. Then system (A1) becomes

$$\frac{d}{dt} g = b_1(1 - g) + a_1((1 - u_2)g^2 + u_2 hg - g), \tag{A2a}$$

$$\frac{d}{dt} h = b_2(1 - h) + a_2(h^2 - h), \tag{A2b}$$

with the initial conditions $g(0) = 1$ and $h(0) = 0$. As Equation A2a does not have an explicit solution, we adopt the following approximation. Suppose we have a solution of $g(t) = g^1(1, 0; t)$ for $t > 0$. We convert this solution to a function $g_x^1(1, 0)$ of x by assuming deterministic exponential growth of type-0 cells, $t = \ln(M/x)/(r - d)$. Hence we have

$$g_x^1(1, 0) = g\left(\frac{1}{r - d} \ln\left(\frac{M}{x}\right)\right). \tag{A3}$$

The solution of Equation A2b is given by

$$h(t) = 1 - \frac{1 - b_2/a_2}{1 - (b_2/a_2)e^{-(a_2 - b_2)t}}. \tag{A4}$$

However, as shown in the following, $h(t)$ does not contribute to the result of P. Equation A2a is rewritten as

$$\frac{d}{dt} g = (1 - g)(b_1 - a_1 g) - a_1 u_2 g(g - h), \tag{A5}$$

with $g(0) = 1$. If we neglect the second term, which is of the order of u_2, this equation is similar to a logistic equation and has two equilibria: $g = 1$ and $g = b_1/a_1$. Starting from $g(0) = 1$, the solution stays in the unstable equilibrium $g = 1$. Now note that the second term of Equation A5 is negative because $g(0) = 1$ and $h(0) = 0$. For small t, this term is approximately $-a_1 u_2$ and pushes the system away from the unstable equilibrium $g = 1$. Once the system is perturbed, it converges to the other equilibrium at $g = b_1/a_1$ following the dynamics determined by the first term. The second term is therefore important only for small t, and we can approximate Equation A5 as

$$\frac{d}{dt} g = (1 - g)(b_1 - a_1 g) - a_1 u_2 \frac{a_1 g - b_1}{a_1 - b_1}. \tag{A6}$$

When $t = 0$, Equation A6 equals Equation A5. With $g(0) = 1$ and $\varepsilon = a_1 u_2/(a_1 - b_1) > 0$, we have

$$1 - g(t) = \frac{1 - b_1/a_1 + \varepsilon}{1 + ((1 - b_1/a_1 + \varepsilon)(1/\varepsilon) - 1)e^{-(a_1 - b_1 + a_1\varepsilon)t}} - \varepsilon. \tag{A7}$$

By noting that $\varepsilon > 0$ is a small quantity and $1/\varepsilon$ is a very large quantity, Equation A7 becomes

$$\begin{aligned}
1 - g(t) &\approx \frac{1 - b_1/a_1}{1 + (1 - b_1/a_1)(1/\varepsilon)e^{-(a_1 - b_1)t}} \\
&= \frac{1 - b_1/a_1}{1 + (1 - b_1/a_1)((a_1 - b_1)/a_1 u_2)e^{-(a_1 - b_1)t}} \\
&= \frac{1 - b_1/a_1}{1 + (1 - b_1/a_1)^2(1/u_2)e^{-(a_1 - b_1)t}}.
\end{aligned} \tag{A8}$$

Using $t = \ln(M/x)/(r - d)$, we have

$$P = 1 - \exp\left[-\frac{u_1}{1 - d/r}\sum_{x=1}^{M-1}\left(1 - g\left(\frac{1}{r - d}\ln\frac{M}{x}\right)\right)\right]$$

$$= 1 - \exp\left[-\frac{u_1}{1 - d/r}\sum_{x=1}^{M-1}\frac{1 - b_1/a_1}{1 + (1 - b_1/a_1)^2(1/u_2)(x/M)^{((a_1-b_1)/(r-d))}}\right]. \tag{A9}$$

In the neutral case, $r = a_1 = a_2$, we can further simplify the formula as follows:

$$P = 1 - \exp\left[-\frac{u_1}{1 - d/r}\sum_{x=1}^{M-1}\frac{1 - d/r}{1 + (1 - d/r)^2(1/u_2)(x/M)}\right]$$

$$= 1 - \exp\left[-Mu_1\int_0^1\frac{1}{1 + (1 - d/r)^2(1/u_2)y}dy\right]$$

$$= 1 - \exp\left[-\frac{Mu_1}{(1 - d/r)^2(1/u_2)}\ln(1 + (1 - d/r)^2(1/u_2))\right]. \tag{A10}$$

We can further simplify the formula to obtain Equation 5 in the text.

APPENDIX B: DERIVATION OF THE ALTERNATIVE FORMULA FOR THE PROBABILITY OF TWO MUTATIONS

Let P_x be the probability that the first successful lineage of type-1 cells arises when there are x type-0 cells. This probability can be written as the product of the probability that no successful type-1 cell lineage arises when there are $1, 2, 3, \ldots, x - 1$ type-0 cells and the probability that a successful type-1 cell lineage is created when there are exactly x type-0 cells. We assume that the number of mutants created when there are x type-0 cells follows a Poisson distribution with mean R_x (IWASA *et al.* 2006). A fraction of these new mutations survives the initial stochastic fluctuations and results in exponentially growing lineages. This fraction is given by $\sigma = 1 - b_1/a_1$, the nonextinction probability. Hence we have $P_x = \prod_{x'=1}^{x-1} e^{-\sigma R_{x'}}(1 - e^{-\sigma R_x})$. With $R_x = u_1/(1 - d/r)$ (IWASA *et al.* 2006) and $\beta = \sigma R_x$, we obtain Equation 6 in the text.

Next we consider the probability that a type-2 cell is created within a lineage that starts from one type-1 cell. Let y be the number of type-1 cells present when the total cell population reaches size M, and let N be the mean number of type-1 cell division events that occur while type-1 cells increase from 1 cell to y cells. The per capita rate of increase in their number is given by $a_1 - b_1$, and the per capita rate of cell division is given by a_1. Hence we have $(a_1 - b_1)/a_1 = (y - 1)/N$. With $y \gg 1$, this expression becomes

$$N = \frac{y}{1 - b_1/a_1}. \tag{B1}$$

During each cell division, a type-2 cell emerges with probability u_2. Hence the probability that at least one type-2 cell arises in such a lineage is given by $Q_x = 1 - \exp[-u_2y/(1 - b_1/a_1)]$ with $y = \exp[(a_1 - b_1)\tau_x]$. Note that τ_x is the length of time between the emergence of a successful type-1 cell and when the total number of type-0 and type-1 cells reaches M. If $y \gg 1$, we have

$$xe^{(r-d)\tau_x} + e^{(a_1-b_1)\tau_x} = M. \tag{B2}$$

Here $x\exp[(r - d)\tau_x]$ is the number of type-0 cells, $\exp[(a_1 - b_1)\tau_x]$ is the number of type-1 cells, and their sum equals M. Equation B2 is a transcendental equation and has to be solved numerically to obtain τ_x.

With these results, the probability that there is at least one type-2 cell when the total population size reaches M is given by Equation 8 in the text. This expression holds if the death rates of type-2 cells can be neglected as compared to their growth rates. If the mortality of type-2 cells cannot be neglected, we replace Q_x by the formula in IWASA *et al.* (2006) and obtain

$$P = \sum_{x=1}^{M-1}e^{-\beta(x-1)}(1 - e^{-\beta})\left(1 - \exp\left[-\frac{u_2yF}{1 - b_1/a_1}\right]\right). \tag{B3}$$

Here $F = \int_0^1(1 - b_2/a_2)/(1 - (b_2/a_2)z^\alpha)dz$ and $\alpha = (a_2 - b_2)/(a_1 - b_1)$.

APPENDIX C: BRANCHING-PROCESS FORMULA FOR THE EXPECTED NUMBER OF TYPE-2 CELLS—DELETERIOUS AND ADVANTAGEOUS CASES

The expected number of type-2 cells is obtained by taking the derivative of the generating function, Equation 2, with respect to s_2 and evaluating it at $s_1 = s_2 = 1$. We first take the derivative of the differential equations for the generating functions, Equation A1, with respect to s_2. Then we obtain a pair of linear equations, which leads to the solution given by Equation 11 in the text.

After having discussed the neutral case in the main text, we focus here on nonneutral cases. From Equation 11 and with $t \approx \ln(M/x)/(r - d)$, we have

$$\frac{\partial g^1}{\partial s_2}\left(1, 1, \frac{1}{r - d}\ln(M/x)\right) = \frac{a_1 u_2}{\delta - \gamma}\left(\left(\frac{M}{x}\right)^{\delta/(r-d)} - \left(\frac{M}{x}\right)^{\gamma/(r-d)}\right), \tag{C1}$$

with $\gamma = a_2 - b_2$ and $\delta = a_1(1 - u_2) - b_1$. Hence the expected number of type-2 cells once the total population reaches size M is

$$\begin{bmatrix} \text{mean number of} \\ \text{type-2 cells} \\ \text{at detection} \end{bmatrix} = \bar{Z}P = \frac{u_1}{1 - d/r}\sum_{x=1}^{M-1}\frac{a_1 u_2}{\delta - \gamma}\left(\left(\frac{M}{x}\right)^{\delta/(r-d)} - \left(\frac{M}{x}\right)^{\gamma/(r-d)}\right)$$

$$= \frac{a_1 u_1 u_2}{(1 - d/r)(\delta - \gamma)}\sum_{x=1}^{M-1}\left(\left(\frac{M}{x}\right)^{\delta/(r-d)} - \left(\frac{M}{x}\right)^{\gamma/(r-d)}\right). \tag{C2}$$

The mean number of type-2 cells among those with at least one type-2 cell is

$$\bar{Z} = \frac{a_1 u_1 u_2}{P(1 - d/r)(\delta - \gamma)}\sum_{x=1}^{M-1}\left(\left(\frac{M}{x}\right)^{\delta/(r-d)} - \left(\frac{M}{x}\right)^{\gamma/(r-d)}\right). \tag{C3}$$

APPENDIX D: ALTERNATIVE FORMULA FOR THE EXPECTED NUMBER OF TYPE-2 CELLS

Let us consider the number of type-2 cells once the population reaches its final size by approximating their growth with deterministic dynamics (see Figure 4 for a schematic explanation). Note that we have already derived the probability that at least one type-2 cell exists, P, the number of type-1 cells, y, and the length of time between the emergence of the first type-1 cell and when the total number of type-0 and type-1 cells reaches M, τ_x. Now we consider the probability that the number of type-2 cells is less than a given value once the total population size reaches M. Let Z be the number of type-2 cells, and let κ_Z be the time it takes from the appearance of the first type-2 cell until the final size is reached (Figure 4). From the assumption that type-2 cells grow exponentially, we have $\kappa_Z = \ln Z/(a_2 - d)$. Then the probability that the number of type-2 cells, Z, is less than z' with $1 < z' \leq M$ is given by

$$\Pr[Z < z'] = \Pr[\kappa_Z < \kappa_{z'}]. \tag{D1}$$

Note that we have already calculated τ_x, and this expression approximately represents the length of time between the emergence of a type-1 cell from x type-0 cells and when the final population size is reached. Then the probability that the number of type-2 cells, Z, is less than z' is expressed as the probability that no type-2 cell lineage emerges during $\tau_x - \kappa_{z'}$ and that a type-2 cell lineage appears thereafter. Let S be the probability that a type-2 cell lineage appears during $\tau_x - \kappa_{z'}$. The probability, Q_x, that a type-2 cell lineage appears during τ_x has already been derived. Therefore, the probability that no type-2 cell lineage emerges during $\tau_x - \kappa_{z'}$ and that a type-2 cell lineage appears thereafter is expressed as $Q_x - S$. Considering that the population starts from one type-0 cell and excluding the cases in which no type-2 cells emerge, we have

$$L_{z'} = \Pr[Z < z'] = \frac{1}{P}\sum_{x=1}^{M-1} P_x(Q_x - S)$$

$$= \frac{1}{P}\sum_{x=1}^{M-1} e^{-\beta(x-1)}(1 - e^{-\beta})(1 - \exp[-(u_2 e^{(a_1-b_1)\tau_x}/(1 - b_1/a_1))] - S). \tag{D2}$$

Since we sum all cases from $x = 1$ to $x = M - 1$, the cases where $\kappa_{z'} \geq \tau_x$ must be considered. When $\kappa_{z'} \geq \tau_x$, the number of the type-2 cells is always less than z'. Hence S is given by

$$S = \begin{cases} 0 & \text{if } \tau_x - \kappa_{z'} \leq 0 \\ 1 - \exp\left[-\dfrac{u_2 \, e^{(a_1 - b_1)(\tau_x - \kappa_{z'})}}{1 - b_1/a_1} \right] & \text{if } \tau_x - \kappa_{z'} > 0. \end{cases} \tag{D3}$$

Note that the probability is conditional to the existence of several type-2 cells.

Finally, let us discuss the probability distribution of the number of type-2 cells. According to Equation D2, the probability that the number of type-2 cells is between z_1 and z_2 with $1 \leq z_1 < z_2 < M$ is given by

$$\Pr[z_1 \leq Z < z_2] = L_{z_2} - L_{z_1}. \tag{D4}$$

From using Equation D4, we can derive the expected number of type-2 cells numerically as

$$\bar{Z} = \sum_{i=1}^{M-2} z_i \left(L_{z_{i+1}} - L_{z_i} \right). \tag{D5}$$

In Figure 5, the dots represent the results of the computer simulation, system 1. We record the number of type-2 cells once the total population size (including type-2 cells) reaches M and calculate the distribution frequency. We perform >40,000 runs of this process to generate each figure.