Cell Selection as Driving Force in Lung and Colon Carcinogenesis

Helmut Schöllnberger, Niko Beerenwinkel, Rudolf Hoogenveen, and Paolo Vineis

Abstract

Carcinogenesis is the result of mutations and subsequent clonal expansions of mutated, selectively advantageous cells. To investigate the relative contributions of mutation versus cell selection in tumorigenesis, we compared two mathematical models of carcinogenesis in two different cancer types: lung and colon. One approach is based on a population genetics model, the Wright-Fisher process, whereas the other approach is the two-stage clonal expansion model. We compared the dynamics of tumorigenesis predicted by the two models in terms of the time period until the first malignant cell appears, which will subsequently form a tumor. The mean waiting time to cancer has been calculated approximately for the evolutionary colon cancer model. Here, we derive new analytic approximations to the median waiting time for the two-stage lung cancer model and for a multistage approximation to the Wright-Fisher process. Both equations show that the waiting time to cancer is dominated by the selective advantage per mutation and the net clonal expansion rate, respectively, whereas the mutation rate has less effect. Our comparisons support the idea that the main driving force in lung and colon carcinogenesis is Darwinian cell selection.

Introduction

Studying the timing of carcinogenic events has provided important hints on the putative mechanisms of cancer onset, that is, of the processes that lead to the first malignant cell (M-cell). Research on mathematical models of carcinogenesis in the 1970s and 1980s was based on the assumptions that (a) carcinogenesis is the effect of initiation and promotion, and (b) carcinogens can be distinguished into those affecting early stages with an irreversible effect and those affecting late stages with a reversible effect. This distinction was largely based on experimental research conducted in the previous decades that had shown that initiation was likely to be due to mutations and that promotion was likely based on nongenotoxic mechanisms.

Different models of carcinogenesis have been developed, and most involve mutation induction and cell expansion (3–10). Here, we compare two modeling approaches: an evolutionary model based on the Wright-Fisher (WF) process (11) and the MVK model of Moolgavkar, Venzon, and Knudson (7, 8), also known as the two-stage clonal expansion model (12). Earlier, the WF process has been used to describe the progression of a benign adenoma to an invasive carcinoma, and different values of the mutation and selection parameters have been explored based on tumor growth data (6). No analytic solution of the stochastic WF model is known, but an approximate closed-form expression of the mean waiting time until the first M-cell appears has been presented (6). In the present article, we derive a similar equation for the waiting time to the first M-cell based on the simplified MVK (S-MVK) model, a simplified deterministic variant of the stochastic MVK model (13). We compare the dynamics predicted by the S-MVK model and the WF model in terms of the time it takes until the first M-cell occurs. The main result of our study is that for both models, we derive similar equations for the median waiting time to the first M-cell and that the expressions emphasize the dominating effect of cell selection (i.e., cell proliferation) on carcinogenesis. We also compare waiting times numerically using the same biological end point, that is, the same cancer site, for both models. Using parameter estimates obtained from fitting epidemiologic lung and colon cancer data, the MVK model predicts waiting times that are consistent across several previously reported data sets with those obtained from the WF model assuming basic genomic parameters such as mutation rate and selective advantage per mutation. Thus, while both mutation and selection are necessary for carcinogenesis (14–16), the model comparison...
suggests that the formation of the first M-cell is driven mainly by clonal expansion of selectively advantageous cells, whereas the effect of the mutation rate appears much smaller.

**Materials and Methods**

**WF model**

The WF process is a stochastic model of an evolving asexual population that has been applied to the cells of a tumor (11). In this model, cancer cells evolve in discrete, nonoverlapping generations, and each cell independently derives from a cell of the previous generation with a probability proportional to the fitness of the parent. Each cell is either identical to its parent, or hit by an additional mutation with probability \( u \) per gene location per cell division (Fig. 1). The expectation of the waiting time until the first M-cell with \( k \) mutations appears, denoted \( T_{WF} \), has been approximated (6) as

\[
E[T_{WF}] = \frac{k \ln^2[s/(ud)]}{s \ln(N_{init}N_{fin})}.
\]

Here, \( s \) is the selective advantage per mutation. An exponential population growth is assumed from \( N_{init} \) cells to \( N_{fin} \) cells when the first M-cell occurs. The number of putative genes involved in the process is denoted by \( d \). The WF model has not been fitted to epidemiologic data. Therefore, we set the parameters to values obtained from published experimental data under additional assumptions as follows.

Based on genomic data (6, 17), we assume a total of \( d = 100 \) susceptible cancer-associated genes. We consider a cell initiated if \( k = 5 \) of those genes are mutated and malignant if \( k = 20 \). Both the number of driver genes (i.e., those that confer a selective advantage) available in the genome, \( d \), and the number of required mutations per initiated or M-cell, \( k \), are not known with certainty today and will eventually be determined in large cancer whole-genome sequencing projects. The values used here are motivated by the cancer genome study Sjöblom and colleagues (17), in which among 13,000 genes analyzed in 22 tumors, \( \sim 80 \) driver mutations have been identified, but only up to 18 and 23 per patient, for late-stage colon and breast cancer, respectively. These numbers were consistent with follow-up studies (18–20). Because only a few of those driver mutations occurred with high frequency (termed “mountains”) whereas most occurred at low frequencies (called “hills”), we assume that adenoma formation requires \( k = 5 \) mutated genes, which are likely to be more specific changes and therefore appear as mountains in the histogram provided by Sjöblom and colleagues (17). Assuming a normal mutation rate (no genetic instability), the average per-gene mutation rate has been estimated as \( u = 1 \times 10^{-7} \) per generation (6). In general, no experimental data are available for the fitness advantage \( s \) conferred by a mutation. Plausible values of \( s \) have been determined as those that cause expected waiting times (Eq. 1) consistent with observed clinical progression times between 5 to 20 years. This approach suggests a selective advantage per mutation on the order of \( 1 \times 10^{-3} \) to \( 1 \times 10^{-4} \) (6).

Under the WF model, genetic progression proceeds in successive approximately equidistant mutational sweeps (6). Based on this regular arrival of new mutant waves and the near clonality of the population at any point in time, the stochastic WF process can be approximated by a linear multistage process with transition rate \([2s \ln N_{clone}]/[\ln^2(s/(ud))]\) in which stages correspond to clonal expansions (21, 22). Here, \( N_{clone} = \sqrt{N_{init}N_{fin}} \) is the (geometric) mean number of adenoma cells. This deterministic approximation to the WF process can be regarded as an Armitage-Doll (AD) model with a series of transitions starting from a pool of \( N \) normal cells (N-cells; see Supplementary Material). The resulting approximated hazard function for this model is

\[
h(t) = N \left\{ \frac{2s \ln N_{clone}}{\ln^2[s/(ud)]} \right\}^k \frac{t^{k-1}}{(k-1)!}.
\]

The survival function is \( S(t) = \exp[-H(t)] \), in which \( H(t) \) is the integrated hazard. The median waiting time to the first cancer cell, denoted \( \tau \), is the solution of the equation \( S(\tau) = 1/2 \). It is the time when half of the population has acquired the first M-cell. The following expression for \( \tau \) is derived in the Supplementary Material:

\[
\tau_{WF} = \left( \frac{k \ln 2}{N} \right)^{1/k} \frac{\ln^2[s/(ud)]}{2s \ln N_{clone}}.
\]

We denote this waiting time \( \tau_{WF}^{cell} \), if the starting point of the carcinogenic process is a set of \( N \) normal cells and

![Figure 1. Schematic representation of the WF process. Illustrated are five generations of a single realization of the WF process with a constant population size of \( N = 6 \) cells. In each generation, cells are drawn randomly from the previous generation. Initially, in generation 1, all cells are wild-type (white). In generation 2, the first cell with one mutation appears (gray). Cells with additional mutations have a selective advantage; they are more likely to generate offspring and will, on average, outcompete cells with fewer mutations. In this realization, the first cell with two mutations occurs in generation 5 (black), and the waiting time for \( k = 2 \) mutations is \( T_{WF} = 4 \) generations; see Eq. 1.](image)
$N_{last} = N$. We also apply Eq. 3 to the process that starts with a clone of $m$ adenoma cells and denote the resulting median waiting time $\tau_{WF}^{[m]}$.

Two-stage model with clonal expansion

In the MVK model (7, 8), normal stem cells can be transformed into cells of an intermediate form at an event rate $\mu_1$, the first mutation rate. These intermediate cells can divide symmetrically into two intermediate cells at rate $\alpha$, die, or differentiate at rate $\beta$, and divide asymmetrically into one initiated cell (I-cell) and one M-cell at rate $\mu_2$ (Fig. 2). Clonal expansion of I-cells is formulated as a stochastic process (7, 8, 12). M-cells are assumed to develop into a detectable tumor after a constant lag time, $t_{lag}$. The hazard function of the MVK model has been derived (23, 24) and can be used to describe cancer incidence and cancer mortality rates (25). It is

$$h(t) = \frac{\mu_1\mu_2N}{\alpha} \frac{e^{\delta_2 t} - 1}{1 - A + (B - 1)e^{\delta_2 t}}, \quad (4)$$

with $A, B = [\alpha + \beta + \mu_2 \pm \sqrt{\alpha + \beta + \mu_2}^2 - 4\alpha\beta]/(2\alpha)$, $\delta_2 = \alpha(B - A)$, and $\lim_{t \to \infty} h(t) = (\mu_1\mu_2N)/[\alpha(B - 1)]$. If the background estimates for $\mu_1, \mu_2, \alpha, \beta$ are used (spontaneous rates), then the hazard function describes the baseline cancer incidence. Recently, the MVK model has been fitted to lung cancer incidence data from the large European Prospective Investigation into Cancer and Nutrition cohort study (26).

The S-MVK model (13) is defined by the hazard

$$h(t) = \frac{N\mu_1\mu_2}{\alpha - \beta} \left( e^{(\alpha - \beta)t} - 1 \right), \quad (5)$$

in which $N$ is the number of normal cells. The S-MVK model is unbound at high ages, that is, $\lim_{t \to \infty} h(t) = \infty$, whereas the hazard of the MVK model reaches a plateau at high ages because of the stochasticity of the birth-death process of I-cells. The S-MVK model can be derived from the MVK model (Eq. 4) by assuming that $\mu_2$ is negligible (13).

In the MVK model, the expected waiting time to cancer can be calculated as the expected value of the time $T$ to the first M-cell, $E[T] = \int_0^\infty P(T > t) dt$. A closed-form expression for the survival function, $P(T > t)$, has been reported (7, 24), but the integral $E[T]$ cannot be solved analytically. By contrast, we obtained the following closed-form expression for the median waiting time from a pool of $N$-cells to the first M-cell for the S-MVK model (see Supplementary Material),

$$\tau_{S-MVK}^{[m]} = \frac{N}{\alpha - \beta} \ln \left[ \frac{(\alpha - \beta)^2 \ln 2}{\mu_1\mu_2N} \right]. \quad (6)$$

We are also interested in the time to the first M-cell starting with a clone of $m$ I-cells, that is, in the transition from adenoma to carcinoma, in the MVK model. Such an expression would be analogous to Eq. 1, which gives an approximation to the expected waiting time to the first M-cell for the WF process starting with an adenoma (a
c

c

Figure 2. Conceptual view of the two-stage model with clonal expansion. Normal cells (N) can develop into initiated (or intermediate) cells (I). I-cells have sustained the first rate-limiting event in the pathway to malignancy at rate $\mu_1$, the parameter defining the rate of critical genomic events involved in initiation. An I-cell can divide into two I-cells with rate $\alpha$; it dies or differentiates with rate $\beta$; it divides asymmetrically into one I-cell and one cell that has sustained the second event (malignant cells (M)) with rate $\mu_2$. The M-cells are assumed to develop into a tumor (T) after a deterministic lag time, $t_{lag}$.

Comparison of the two models

The WF and the MVK model are different stochastic models of carcinogenesis (Figs. 1 and 2). In the MVK model, carcinogenesis is viewed as the result of two critical, irreversible, rate-limiting, and hereditary (at the level of somatic cells) events (23). By contrast, the WF model describes carcinogenesis explicitly as a series of genetic changes in a reproducing cell population. The genetic progression stages that eventually lead to the first M-cell are defined by the number of mutated cancer-associated genes. These mutant types sweep through the population in several subsequent clonal waves (6). Tumor growth is also modeled differently. In the S-MVK model, the net clonal expansion rate $(\alpha - \beta)$ describes the growth of initiated cells, whereas in the WF model, a deterministic exponential growth from adenoma to carcinoma is assumed. The background rate $\mu_0$ in the MVK model corresponds to the composite parameter $u \times d$ in the WF model because in the WF process, no carcinogenic environmental factors such as smoking or radiation are included. Likewise, the growth rate, $(\alpha - \beta)$, plays a similar role in the MVK model as the selective advantage, $s$, in the WF model. The term “initiated cell” used in the MVK model corresponds to the notion of “adenoma cell” in the context of the WF model.
Results

Despite their differences, the WF model and the MVK model are conceptually similar. Both explain the appearance of the first malignant cancer cell by accumulating mutations with a selective advantage in a cell population, and both distinguish the generation of a new cell type (mutation) from its growth (clonal expansion). The dynamics of the carcinogenic process can be directly compared between the two models in terms of the time to appearance of the first M-cell. Because the expectation of this waiting time cannot be calculated analytically for the MVK model, we have calculated the median waiting time to the first M-cell, \( \tau_{\text{WF}}^{\text{N-cells}} \) and \( \tau_{\text{S-MVK}}^{\text{N-cells}} \), for the WF model and the S-MVK model, respectively (see Materials and Methods and Supplementary Material),

\[
\tau_{\text{WF}}^{\text{N-cells}} = \left( \frac{k! \ln 2}{N} \right)^{1/3} \frac{\ln^2 \left[ s/(ud) \right]}{2s \ln N_{\text{clone}}},
\]

\[
\tau_{\text{S-MVK}}^{\text{N-cells}} = \frac{1}{(\alpha - \beta)} \ln \left[ \frac{(\alpha - \beta)^2 \ln 2}{\mu_1 \mu_2 N} \right].
\]

The two equations are strikingly similar. In both expressions the parameter governing the growth of clones harboring advantageous mutations, namely the selective advantage, \( s \), and the net clonal expansion rate, \( (\alpha - \beta) \), appears in the denominator, whereas all other parameters enter only logarithmically. Thus, in both models, the selection parameters have the strongest impact on the waiting time distributions, suggesting that cell selection is the major driving force of carcinogenesis.

The observed prominent role of cell selection is confirmed by numerical evaluation of the waiting time equations using best estimates from various model fits to individuals with full grown tumors. Sets of best estimated values, which were obtained from several studies by fitting the MVK or two-stage clonal expansion model to lung (26, 28–31) and colon cancer data (32, 33), were used to calculate \( \tau_{\text{MVK}}^{\text{clone}(m - 1)} \) (Supplementary Material, section 4; Table 1). In the present study, no model fits were performed. We have analyzed whether the predicted time from a clone of I-cells to the first M-cell is consistent between the two models. We performed calculations using \( m = 1 \), that is, a clone consisting of one I-cell and \( \beta = 0 \), assuming nonextinction (see Section 5 of the Supplementary Material for the effect of setting \( \beta = 0 \) on the parameter estimates).

For lung cancer, we obtained values for \( \tau_{\text{MVK}}^{\text{clone}(m - 1)} \) between 99 and 192 years, and for colon cancer between 62 and 69 years (Table 1). For the WF process, we evaluated the same median waiting time \( \tau_{\text{WF}}^{\text{clone}(m = 1)} \) with \( N = N_{\text{init}} = 1 \), \( N_{\text{init}} = 1 \times 10^6 \), and \( k = 15 \), because initiated cells are assumed to harbor 5 of the 20 required mutations already (see Methods). With \( s = 1 \times 10^{-3} \) and \( 1 \times 10^{-4} \), the resulting waiting times are 17.6 and 43.9 years, respectively. We also evaluated Eq. 1 for the same parameter values and obtained \( E[T_{\text{WF}}] = 42.0 \) and 105.1 years for \( s = 1 \times 10^{-3} \) and \( 1 \times 10^{-4} \), respectively. Given that the WF model has not been fitted to any cancer incidence or mortality data, the values for \( \tau_{\text{WF}}^{\text{clone}(m = 1)} \) and \( E[T_{\text{WF}}] \) based on the WF model compare well to those for \( \tau_{\text{clone}(m = 1)}^{\text{MVK}} \) based on the MVK model. Our values for \( \tau_{\text{MVK}}^{\text{clone}(m = 1)} \) also compare very well with those for the sojourn time, \( T_s \), from the study by Meza and colleagues (27). For colon cancer, those authors found values between 55 and 64 years. For \( m = 1 \), we obtain \( \tau_{\text{MVK}}^{\text{clone}(m = 1)} = 62 \) and 69 years for the two colon cancer data sets (Table 1).

In addition to the waiting times that start with a clone of I-cells, we also evaluated the median waiting time, \( \tau_{\text{S-MVK}}^{\text{N-cells}} \), which starts with a pool of N-cells. For lung cancer, \( \tau_{\text{S-MVK}}^{\text{N-cells}} \) ranges from 108 to 135 years. For colon cancer, the calculated waiting times are also consistent across data sets (88–106 years; Table 1).

### Table 1. Median times to failure

<table>
<thead>
<tr>
<th>References and exact information from which sets of best estimates values were used for the calculations of</th>
<th>( \tau_{\text{MVK}}^{\text{clone}(m - 1)} ) (y)</th>
<th>( \tau_{\text{S-MVK}}^{\text{N-cells}} ) (y)</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meza and colleagues (28); “NHS” from Table 2 in cited article</td>
<td>109.8</td>
<td>119.5</td>
<td>Lung</td>
</tr>
<tr>
<td>Schöllnberger and colleagues (26); fit-RIM1 from Table 4 in cited article</td>
<td>192.3</td>
<td>134.8</td>
<td>Lung</td>
</tr>
<tr>
<td>Hazelton and colleagues (29); “CPS-I males” from Table 2 in cited article</td>
<td>99.7</td>
<td>132.0</td>
<td>Lung</td>
</tr>
<tr>
<td>Moolgavkar and colleagues (30); “Model B” from Table 2 in cited article</td>
<td>99.3</td>
<td>128.6</td>
<td>Lung</td>
</tr>
<tr>
<td>Moolgavkar and colleagues (31); “Full Model” from Table 2 in cited article</td>
<td>126.1</td>
<td>107.6</td>
<td>Lung</td>
</tr>
<tr>
<td>Luebeck and Moolgavkar (32); “White males, 2 stage (k = 2)” from Table 3 in appendix to cited article</td>
<td>88.7</td>
<td>87.5</td>
<td>Colom</td>
</tr>
<tr>
<td>Moolgavkar and Luebeck (33); “Males, Two-mutation model” from Table 1 in cited article</td>
<td>61.7</td>
<td>106.4</td>
<td>Colon</td>
</tr>
</tbody>
</table>

NOTE: Values for median time to the first M-cell (Eqs. 7 and 9) calculated from different sets of best estimated values to lung and colon cancer. We calculated \( \tau_{\text{S-MVK}}^{\text{N-cells}} \) for \( m = 1 \) and \( \beta = 0 \), i.e., starting with one initiated cell and assuming that there is no extinction of clones. For the lung cancer calculations of \( \tau_{\text{clone}(m - 1)}^{\text{MVK}} \), \( N = 1 \times 10^6 \); for colon cancer, \( N = 1 \times 10^8 \). These values are taken from the referenced studies. Further details of the calculations are provided in section 4 of the Supplementary Material.
Discussion

We have applied two different mathematical models to study the relative contribution of mutation induction and cell selection to carcinogenesis. The WF model is based on an evolutionary theory of carcinogenesis (6, 34), and the MVK model is a stochastic two-stage model with clonal expansion (7, 8). Both models share essential biologically motivated characteristics, namely the generation of new cell types with increased replication capacities (Figs. 1 and 2). However, some specific model features are conceptually very different. For example, the MVK model applies two stages to malignancy with typical values for the background mutation rates of $1 \times 10^{-6}$ to $1 \times 10^{-8}$ per year and background net clonal expansion rates of 0.01 to 0.1 per year (26, 33). In general, the two mutation rates $\mu_1$ and $\mu_2$ in the MVK model do not refer to genomic mutation rates, but rather to transition rates between biophysical states of the carcinogenic process that might be the result of several genetic changes. This interpretation is in line with the original definition of $\mu_1$ and $\mu_2$ as event rates (7, 8) and consistent with Moogavkar’s interpretation of Vogelstein’s multistage genetic model for colorectal tumorigenesis within the two-stage model (35). By contrast, the WF process assumes $\sim 20$ genetically defined stages, and the mutation rate between these stages directly reflects the DNA mutation rate in humans. Typical values for the selective advantage $s$ are 0.01 and 0.001 per mutation, that is, 1% and 0.1% selective advantage, respectively. Because $s$ can also be regarded as a growth rate (refer to Eq. 9 in the Supplementary Material of ref. 6), these values correspond to 3.65 and 0.365 per year, respectively. The normal mutation rate of $\mu = 1 \times 10^{-7}$ per gene per cell division corresponds to a total of $u \times d = 3.65 \times 10^{-3}$ mutations per year. The mutation and selection parameters of the MVK and the WF model are not directly comparable because they refer to differently defined stages of carcinogenesis. However, obviously, the 20 stages of the WF model need to be taken much faster than the two stages of the MVK model to arrive at similar waiting times, implying higher values for the mutation rate or the selective advantage (or both) in the WF model, consistent with the values for $s$ and $\alpha \times d$ mentioned above.

We find that the MVK and the WF model have similar dynamics in the development of the first M-cell mainly driven by cell selection, that is, clonal expansions of cells harboring selectively advantageous mutations. This conclusion is based on comparing Eqs. 8 and 9, the median times to the first M-cell in the WF model and in the S-MVK model, respectively. Both waiting times are inversely proportional to the selective advantage, $s$, and the net clonal expansion rate, $(\alpha - \beta)$, respectively, whereas all other parameters enter the equations only logarithmically. Thus, in both models, the waiting time is very sensitive to parameters associated with promotion, but less sensitive to all other parameters, including mutation rates. Formulas have also been developed for the waiting time to the first M-cell starting with a clone of I-cells (Eqs. 1 and 7). Again, these expressions both show the strongest dependence on cell selection confirming the main finding of the study. Other approximations to the average time between initiation and promotion and more generally to the speed of evolution in large asexual populations have been reported elsewhere (34, 36).

Our findings are in line with the results of several earlier approaches (14–16). The major role of selection, in the form of promotion, was indicated as the mechanism of caloric reduction in preventing experimental cancer 65 years ago (37). Rodin and Rodin (14) analyzed mutation spectra and concluded that the main driver of lung carcinogenesis is cell selection, not mutagenesis. Cell selection was also shown to be the driving force in spontaneous transformation of cells in culture (38, 39). Tomlinson and colleagues (16) mathematically analyzed the role of the mutation rate in the growth of sporadic colorectal cancer and concluded that selection without increased mutation rates is sufficient to explain the evolution of tumors (15), challenging the concept of a mutator phenotype proposed by Loeb and colleagues (40, 41).

The dominance of cell selection over mutation is also visible in the hazard functions (Eqs. 2 and 5; Supplementary Material). For the WF process (Eq. 2), the mutation rate enters only logarithmically, whereas the selective advantage enters in a much stronger way as $s^\alpha$. For the S-MVK model (Eq. 5), the mutation rates enter linearly, whereas the net clonal expansion rate, $(\alpha - \beta)$, enters exponentially. These qualitative differences reflect our finding based on analyzing the median times to the first M-cell. For the S-MVK model, the much stronger influence of clonal expansion of initiated cells over mutations is a consequence of the exponential growth rate implied by the model structure. The WF process, which is based on the basic evolutionary mechanisms of mutation and selection, comes to the same conclusion. Therefore, both the S-MVK and the WF model provide independent support for this conclusion.

The approximations made in the derivation of $\tau_{WF}$ (Eq. 3), namely the approximation of the stochastic WF process by a linear multistep process using approximate transition rates, together with the fact that it has to be evaluated with parameter values that were not determined by fitting incidence or mortality data, make the comparison with the MVK model difficult. It is therefore not surprising that somewhat different numerical values are obtained for $\tau_{WF}^{(m = 1)}$ compared with $\tau_{MVK}^{(m = 1)}$. However, all results are clearly in the same order of magnitude, and given the strong approximations in the derivation of Eq. 3, they appear in reasonable agreement.

The two-stage clonal expansion model has recently been advanced to include preinitiation stages (27, 32). It has been shown that this new multistage clonal expansion model is consistent with the linear phase of cancer incidence in the Surveillance Epidemiology and End Results (SEER) data (27). The authors also presented a formula for the mean duration, $T_m$, from the birth of a premalignant clone (i.e., $m = 1$) to its eventual development into a malignant tumor, conditional on nonextinction (equivalent to assuming that $\beta = 0$): $T_m = \ln(\alpha/\mu_2)/\alpha$ (here, we applied $\beta = 0$ in Eq. 3 of ref. 27). It can be shown that for $\beta = 0$ and $m = 1$, Eq. 7 yields the same expression: $\tau_{MVK}^{(m = 1)} = \ln(\alpha/\mu_2)/\alpha$ (refer to Supplementary Material, Section 6 for the mathematical details). It is plausible that because of the approximations necessary to obtain this closed-form expression, the median time to the first M-cell is identical to the mean time $T_m$. Values for $T_m$ between...
55 and 64 years obtained with the multistage clonal expansion model for colon cancer have been reported (27). These values are based on fitting three- and four-stage models to the SEER data. Their values for \( T_e \) need to be compared with evaluations of Eq. 7 for \( m = 1 \) and \( \beta = 0 \). For colon cancer, we obtained values between \( \tau_{\text{MVK}} \sim 62 \) and 69 years (Table 1). We suggest that the changes in the \( T_e \) values in ref. (27) stem from the fact that these authors used different models and data compared with (32, 33). Given these differences, it seems that their values are in excellent agreement with ours.

We remark that a hazard function of the AD-type has been shown to provide a poorer fit to age-specific incidence data of colorectal cancer and pancreatic cancer in the SEER registry than more complex models (27). We nevertheless used the AD model here because it has been shown (21) that the stochastic WF process can be approximated by a linear multistage process (i.e., an AD-type model) in which stages correspond to clonal expansions. In addition, the formula for \( \tau_{\text{WF}} \) (Eq. 3), which was derived using an AD model, was evaluated using values from ref. (6) and not with values obtained with the multistage clonal expansion model of Meza and colleagues (27).

The values for the median time to the first M-cell for the S-MVK model, \( \tau_{\text{S-MVK}} \), are consistent among various sets of best estimates for lung cancer (Table 1). Values between 110 and 140 years \( \{\text{S-MVK} + \text{lag} \text{ with } \text{lag} = 5, \text{ respectively,} \text{ } 3.5 \text{ y (refer to Supplementary Material, Section 4)}\} \) for half of a population of nonsmokers to get lung cancer are compatible with epidemiologic studies (26, 42). We also calculated \( \tau_{\text{S-MVK}} \) with parameter estimates relating to a male smoker who smokes 20 cigarettes per day starting at the age of 20 years for life \( \alpha = 0.133 \text{ y}^{-1}, \beta = 0, \mu_1 = \mu_2 = 2.03 \times 10^{-7} \text{ y}^{-1}, \mu_3 = 5.06 \times 10^{-7} \text{ y}^{-1}; \) these parameter estimates were taken from ref. (26). At an age of \( \tau_{\text{S-MVK}} + \text{lag} \) = 75.6 years, half of these life-long smokers would die from lung cancer. This model prediction is similar to the 88 years estimated in ref. (26) for the same event to happen. Two suitable sets of best estimates from MVK model fits to colon cancer were also identified. The calculated values for \( \tau_{\text{S-MVK}} \), are not much different from each other (Table 1) and are consistent with extrapolations from epidemiologic studies. For example, Fig. 3 in ref. (33) provides the lifetime probability of colon cancer in the male population of Birmingham, Alabama represented in the SEER data. When this curve is extrapolated to higher ages (after fitting a suitable function to selected data points on that curve), it is found that after \( \sim 120 \) years half of the population has died from cancer. This is consistent with \( \tau_{\text{S-MVK}} = 106 \) years from Table 1.

Summarizing, it can be said that the analytic and numerical analyses of the waiting times derived from two different mathematical models of carcinogenesis support the claim that cancer onset is dominated by the clonal expansion of mutated cells, that is, by cell selection, and that mutation induction has a smaller influence. We expect this conclusion to hold also for other solid tumors.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Dr. Annette Kopp-Schneider, German Cancer Research Center, for the fruitful discussions and Dr. Petra Wark, Imperial College London, for reading an earlier version of the manuscript.

Grant Support

Grant from the European Union to the ECNIS Network of Excellence (FOOD-CT-2005-513943, P. Vineis) and grants from the Austrian Science Fund FWF (P18055-N02 and P21630-N22) and financial support from Stiftungs- und Förderungsgesellschaft von Salzburg University (H. Schöllnberger). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 12/02/2009; revised 06/29/2010; accepted 06/30/2010; published OnlineFirst 07/23/2010.

References

Cell Selection as Driving Force in Lung and Colon Carcinogenesis

Helmut Schöllnberger, Niko Beerenwinkel, Rudolf Hoogenveen, et al.

Cancer Res 2010;70:6797-6803. Published OnlineFirst July 23, 2010.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-09-4392

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2010/07/23/0008-5472.CAN-09-4392.DC1

Cited articles
This article cites 39 articles, 17 of which you can access for free at:
http://cancerres.aacrjournals.org/content/70/17/6797.full#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/70/17/6797.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.