Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Theoretical Biology 231 (2004) 515 524
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

An ordinary differential equation model for the multistep transformation to cancer

Sabrina L. Spencer[a,b,*], Matthew J. Berryman[b], José A. García[c], Derek Abbott[b]

[a]Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109 0618, USA
[b]Centre for Biomedical Engineering, School of Electrical and Electronic Engineering, The University of Adelaide, SA 5005, Australia
[c]Laboratory of Theoretical Biology, Universidad La Salle, 06140 México, D.F., México

Received 29 February 2004; received in revised form 23 June 2004; accepted 12 July 2004
Available online 25 September 2004

Abstract

Cancer is viewed as a multistep process whereby a normal cell is transformed into a cancer cell through the acquisition of mutations. We reduce the complexities of cancer progression to a simple set of underlying rules that govern the transformation of normal cells to malignant cells. In doing so, we derive an ordinary differential equation model that explores how the balance of angiogenesis, cell death rates, genetic instability, and replication rates give rise to different kinetics in the development of cancer. The key predictions of the model are that cancer develops fastest through a particular ordering of mutations and that mutations in genes that maintain genomic integrity would be the most deleterious type of mutations to inherit. In addition, we perform a sensitivity analysis on the parameters included in the model to determine the probable contribution of each. This paper presents a novel approach to viewing the genetic basis of cancer from a systems biology perspective and provides the groundwork for other models that can be directly tied to clinical and molecular data.

Keywords: Cancer; Oncogenesis; Tumor progression; Genetic instability; Hallmarks of cancer

1. Introduction

The standard perspective on cancer progression is that it is a form of somatic evolution where certain mutations give one cell a selective growth advantage (Cahill et al., 1999). Oncogenesis is thought to require several independent, rare mutation events to occur in the lineage of one cell (Nowell, 1976). Kinetic analyses have shown that four to six rate-limiting stochastic mutational events are required for the formation of a tumor (Armitage and Doll, 1954; Renan, 1993). Hanahan and Weinberg (2000) proposed the following six hallmark capabilities that a normal cell must acquire to become a cancer cell: (i) self-sufficiency in growth signals, (ii) insensitivity to anti-growth signals, (iii) evasion of apoptosis, (iv) limitless replicative potential, (v) sustained angiogenesis, and (vi) tissue invasion and metastasis. They define genetic instability as an "enabling characteristic" that facilitates the acquisition of other mutations due to defects in DNA repair processes. We reduce these characteristics to the following four: angiogenesis (A), immortality, including evasion of cell death (D), genetic instability, a function of mutation rates (G), and increased replication rate (R). We consider invasion and metastasis (M) as a final step that allows the spread of a localized tumor. In line with the views of Hanahan and Weinberg (2000), we foresee cancer research developing into a logical science, where the molecular and clinical complexities of the disease will be understood in terms of a few underlying principles. We therefore explore the

*Corresponding author. Currently affiliated with Computational and Systems Biology at the Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
E mail address: spencers@mit.edu (S.L. Spencer).

0022 5193/$ see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.07.006

multistep progression to cancer with an ordinary differential equation (ODE) model which, despite the apparent complexity of the equations, is based on basic principles and a minimal set of parameters.

2. Structure and parameters of the model

Although the model applies to the process of oncogenesis in general, the parameters are loosely based on breast cancer data. We consider the following cell populations: a population of 10^8 normal cells (N), (Nowak et al., 2002), cells which have acquired the ability to induce angiogenesis (A), cells with mutations which allow them to avoid death (D), cells with mutations that lead to genetic instability (G), cells with mutations which increase their replication rate (R), and cells with two or more of these mutations. Cell populations that have acquired two or three mutations are denoted by listing the mutations together in alphabetical order (state DRA would be listed as state ADR, for example). We label a cell which has acquired all four mutations a primary tumor cell (T). Although our model only addresses development of a primary tumor in genetic detail, we allow a primary tumor cell that has acquired the capability to invade and metastasize to become a metastatic cell (M).

The spontaneous mutation rate in human cells has been estimated to be in the range of 10^{-7}–10^{-6} mutations/gene/cell division (Jackson and Loeb, 1998). We assume a spontaneous mutation rate of $k_1 = 10^{-7}$ mutations/gene/cell division. The loss of DNA repair genes can increase the mutation rate by a factor ranging from 10^1 to 10^4 (Tomlinson et al., 1996). We assume that the mutation rate after a genetic instability mutation increases 1000-fold to $k_2 = 10^{-4}$ mutations/gene/cell division. Successful invasion and metastasis depend upon acquisition of the other hallmark capabilities, as well as several new capabilities (Hanahan and Weinberg, 2000). To simplify the model, we do not address the multistep progression of a tumor cell to a metastatic cell, and instead consider this complex process as one step. It has been estimated that the rate of successful metastasis is in the range of 10^{-9}–10^{-7} per cell division (Luebeck and Moolgavkar, 2002), and so we use a conservative estimate of $k_3 = 10^{-9}$ for the transition from a primary tumor cell to a metastatic cell.

A tumor cannot grow past about 10^6 cells without angiogenesis supplying blood to the tumor (Folkman, 1990). We thus cap the size of the tumor at 10^6 cells until greater than 10% of the population of non-normal, non-metastatic cells have acquired a mutation in an A gene. This accounts for the fact that only a fraction of the cells in a tumor need to send angiogenesis signals in order to develop an adequate blood supply for the tumor. In addition, populations of non-normal, non-metastatic cells are always capped by a lethal tumor burden limit of 10^{13} cells (Friberg and Mattson, 1997), irrespective of the angiogenesis cap.

Futreal et al. (2004) state that 291 genes have been reported to be implicated in the causation of human cancer and note that many more cancer genes remain to be identified. The number listed on the Wellcome Trust Sanger Institute web site at this time is 298 genes. We therefore estimate that there are approximately 400 genes involved in the development of a primary tumor. Dividing by four categories, we assume that there are approximately 100 genes involved in each of categories A, D, G, and R. To account for the fact that some genes function in more than one category, we allow double and triple state transitions but reduce the number of genes involved to the order of 10 and 1, respectively, to reflect the likelihood that a single mutation would affect more than one category. For example, we assume that there are 100 genes involved transitions where only one mutation is acquired (e.g. $N \to A$, $D \to DR$, $AG \to AGR$, or $ADG \to ADGR$), 10 genes involved in transitions where two mutations are acquired in one step (e.g. $N \to AD$, $G \to ADG$, or $AR \to ADGR$), and 1 gene involved in transitions where three mutations are acquired in one step (e.g. $N \to ADG$ or $G \to ADGR$). This feature accounts for a mutational hit in p53, for example, which could take a cell directly from N to DGR, as p53 is involved in apoptosis, DNA repair, and cell cycle progression (Vogelstein et al., 2000).

We estimate that the relative contribution to increased net proliferation for mutations in the D and R categories is 7 and 3, respectively, an inference made from work by Tomlinson and Bodmer (1995). Using this D:R ratio of 7:3, a tumor volume doubling time for breast cancer of 500 days (Friberg and Mattson, 1997), and a cell division rate for breast cancer of $\frac{1}{10}$ days^{-1} (Rew and Wilson, 2000), we calculate (see Section 4) the following: cells without a mutation in an R gene divide every $b = 10$ days, cells with a mutation in an R gene divide every $b_R = 9.92$ days, the lifetime of cells without a mutation in a D gene is $d = 10$ days, and the lifetime of cells with a mutation in a D gene is $d_D = 10.11$ days. The birth and death rates are equal for normal cells and for all cells without a mutation in D or R.

The above information is depicted in a unified fashion in Fig. 1 and the parameters appearing in the ODE model are given in Table 1.

In the next section, we present an ODE model that will be used to explore the following areas:

(1) The kinetics of various paths to cancer.
(2) The effect of inherited mutations on cancer development.
(3) A sensitivity analysis of variations in the parameters.

3. Construction of the ODE model

Although the development of cancer has inherent stochasticity and several previous cancer models include stochastic components (Speer et al., 1984; Koscielny et al., 1985), ordinary differential equations can accurately model the rate changes associated with cancer progression. Adding a stochastic component would not change the mean time to cancer onset that our model produces using deterministic differential equations. Our goal is only to create one "generic" model to better understand the kinetics of cancer progression, not to create a model that captures the variability of many different types of cancer all at once. We here outline the ODE model that we use.

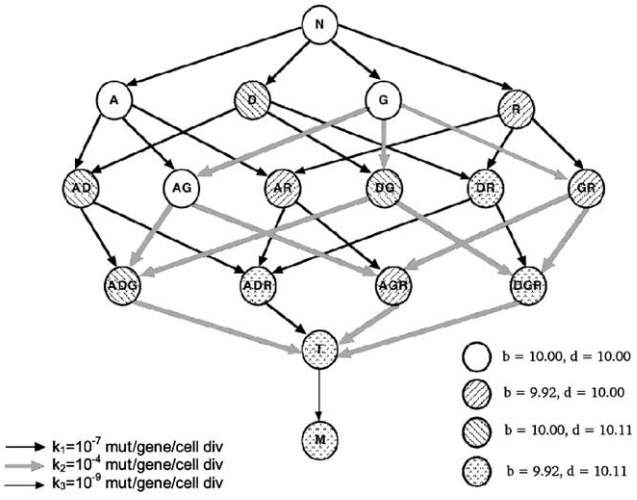

Fig. 1. State diagram of the model. Normal cells (N) can acquire mutations which give the cell the capability to induce angiogenesis (A), mutations which give the cell the capability to avoid death (D), mutations which lead to genetic instability (G), or mutations which increase the proliferation rate (R). These mutations are acquired at rate k_1. After a mutation in G, the mutation rate increases to k_2. Cells with one mutation go on to acquire two, three, and four mutations, denoted by listing the mutations together in alphabetical order for the cases of two and three mutations. When a cell has acquired all four mutations, it becomes a primary tumor cell (T). Finally, tumor cells become metastatic cells (M) at rate k_3. Double and triple state transitions are also allowed, as detailed in the text, but are not shown in this diagram for simplification. Cell birth rates ($1/b$) and cell death rates ($1/d$) have units days^{-1}.

Based on the basic rules outlined in the state diagram in Fig. 1, we construct 17 ODEs to model a heterogeneous population of cells undergoing the multistep process of tumorigenesis. Each equation represents one of the 17 populations of cells depicted in the state diagram and has the following format: the population of cells in a state is increased by cells gaining mutations and entering that state from previous states, is increased by cells replicating and remaining in that state, and is decreased by cells gaining new mutations and leaving that state for a new state. The populations are capped by two logistic terms, as detailed below.

We condense the ODEs into vector format as follows:

$$\frac{d\mathbf{y}}{dt} = \left(\mathrm{diag}\left(\mathrm{diag}\left(\mathbf{y}^{\mathrm{T}}\mathbf{k}\right)^{\mathrm{T}}\mathbf{b}\right)\mathbf{M} + \mathrm{diag}\left((\mathbf{b}-\mathbf{d})^{\mathrm{T}}\mathbf{y}\right)\right)\mathbf{S}\left(1 - a(\mathbf{y})\frac{P_{\overline{NM}}}{10^6}\right) \times \left(1 - \frac{P_{\overline{NM}}}{10^{13}}\right) + \mathbf{m_m}, \quad (1)$$

where $\mathbf{y}$ is the row vector of cell populations; y_1 is the population of normal cells, $y_2, y_3, \ldots, y_{15}$ are the populations of cells with single, double, and triple mutations, y_{16} is the number of primary tumor cells (cells with all four mutations), and y_{17} is the number of metastatic cells. Here, diag $(\cdot)$ is the operator that forms the row vector of the main diagonal of the matrix. The corresponding rate (row) vector is $\mathbf{k}$, with mutation rates

Table 1
Parameters appearing in the ODE model

Characteristic	Parameter	Range in literature	Default value	Reference
Mutation rate without a G mutation	k_1	10^{-7} 10^{-6} mut./gene/cell div.	10^{-7}	Jackson and Loeb (1998)
Mutation rate with a G mutation	k_2	10^{-6} 10^{-2} mut./gene/cell div.	10^{-4}	Tomlinson et al. (1996)
Metastasis rate	k_3	10^{-9} 10^{-7} /cell division	10^{-9}	Luebeck and Moolgavkar (2002)
Number of genes involved in cancer		291+ genes	400	Futreal et al. (2004)
Genes per single, double, triple transitions		Unknown	100, 10, 1	N/A
Tumor volume doubling time		88 523 (sometimes >5000) days	500	Friberg and Mattson (1997)
Relative contribution of D:R		7:3 8:2 (inferred)	7:3	Tomlinson and Bodmer (1995)
Cell division rate without an R mutation	$1/b$	Once every 1.8 47.5 days	$\frac{1}{10}$ days^{-1}	Rew and Wilson (2000)
Cell division rate with an R mutation	$1/b_R$		$\frac{1}{9.92}$ days^{-1}	See Section 4
Cell death rate without a D mutation	$1/d$	Once every 1.8 47.5 days	$\frac{1}{10}$ days^{-1}	Rew and Wilson (2000)
Cell death rate with a D mutation	$1/d_D$		$\frac{1}{10.11}$ days^{-1}	See Section 4
% of cells needed to signal for A		Unknown	10%	N/A
Angiogenesis cap		10^6 cells	10^6	Folkman (1990)
Lethal tumor burden cap		10^{13} cells	10^{13}	Friberg and Mattson (1997)

The default value is that used in our ODEs, unless otherwise specified.

k_i (mutations/gene/cell division) corresponding to the mutation rate for element y_i in $\mathbf{y}$. The same applies to the birth rates $\mathbf{b}$ (day ^{1}) and death rates $\mathbf{d}$ (day ^{1}). The metastasis rate vector is

$$\mathbf{m_m} = (0, 0, \ldots, 0, 10^{\ 9} \times y_{16}) + (1/b_R - 1/d_D)y_{17},$$

corresponding to cells leaving y_{16} for y_{17} at rate $10^{\ 9}$, and a doubling of metastatic cells at rate $(1/b_R - 1/d_D)$ for $1/b_R$ and $1/d_D$ as given in Table 1. The 17×17 upper triangular matrix $\mathbf{M}$ consists of elements $M_{i,j}$ $(j \neq i)$ for the number of genes associated with going from state i to state j, and

$$M_{i,i} = -\sum_{j \neq i} M_{i,j} \tag{2}$$

is the main diagonal containing the number of genes for leaving each of the states. $\mathbf{S}$ is the 17×17 matrix

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \ldots & & 1 & 0 \\ 0 & \ldots & & 0 & 0 \end{bmatrix} \tag{3}$$

used to apply the cell population caps to the non-normal, non-metastatic cells. Non-normal, non-metastatic cells are denoted by $P_{\overline{NM}}$, where

$$P_{\overline{NM}} = \left(\sum_{i=2}^{16} y_i \right). \tag{4}$$

The system is capped at 10^6 cells using a logistic term if 10% or fewer of the non-normal, non-metastatic cells are in states with angiogenesis mutations, otherwise this term is removed. This is expressed in the term $a(\mathbf{y})$, defined as

$$a(\mathbf{y}) = \begin{cases} 0, & \frac{P_A}{P_{\overline{NM}}} > 10\%, \\ 1 & \text{otherwise}, \end{cases} \tag{5}$$

where P_A is the number of non-metastatic cells with mutations in A category genes. The populations of non-normal, non-metastatic cells are also capped by a lethal tumor burden limit of 10^{13} cells (Friberg and Mattson, 1997), irrespective of the angiogenesis cap. The ODEs are solved using the Runge–Kutta method of order 5, with a variable step size between 1 and $10^{\ 5}$, to guarantee that the errors in calculating the populations remain within $10^{\ 4}$ at each step.

Below, we have reproduced the normal cell and four single state ODEs from the compact vector form for ease of comprehension. Note that our equations assume a constant, renewing population of normal cells, since we assume that cells leaving state N for other states are few enough in number so as not to affect the population of N cells.

$$\frac{dP_N}{dt} = 0, \tag{6a}$$

$$\frac{dP_A}{dt} = \left(\frac{100 P_N k_1}{b} - \frac{(3 \times 100 + 3 \times 10 + 1) P_A k_1}{b} \right) \times \left(1 - \frac{P_{\overline{NM}}}{10^6} \right) \left(1 - \frac{P_{\overline{NM}}}{10^{13}} \right), \tag{6b}$$

$$\frac{dP_D}{dt} = \left(\frac{100 P_N k_1}{b} + P_D \left(\frac{1}{b} - \frac{1}{d_D} \right) - \frac{(3 \times 100 + 3 \times 10 + 1) P_D k_1}{b} \right) \times \left(1 - \frac{P_{\overline{NM}}}{10^6} \right) \left(1 - \frac{P_{\overline{NM}}}{10^{13}} \right), \tag{6c}$$

$$\frac{dP_G}{dt} = \left(\frac{100 P_N k_1}{b} - \frac{(3 \times 100 + 3 \times 10 + 1) P_G k_2}{b} \right) \times \left(1 - \frac{P_{\overline{NM}}}{10^6} \right) \left(1 - \frac{P_{\overline{NM}}}{10^{13}} \right), \tag{6d}$$

$$\frac{dP_R}{dt} = \left(\frac{100 P_N k_1}{b} + P_R \left(\frac{1}{b_R} - \frac{1}{d} \right) - \frac{(3 \times 100 + 3 \times 10 + 1) P_R k_1}{b_R} \right) \times \left(1 - \frac{P_{\overline{NM}}}{10^6} \right) \left(1 - \frac{P_{\overline{NM}}}{10^{13}} \right), \tag{6e}$$

$$\vdots$$

The equations for the other populations follow the same format and can be derived from the state diagram and from the vector form of the ODEs. In words, Eq. (6c), for example, says that the population of cells with a mutation in D is increased by normal cells gaining a mutation in one of 100 genes in D at a rate of k_1 every b days. The population is also increased by cells in state D replicating (but not mutating) every b days and dying every d_D days. The population is decreased by cells leaving state D and gaining a single mutation in one of 3 other categories (AD, DG, DR) each with 100 genes, by gaining a double mutation in one of 3 ways (DGR, ADG, ADR), with 10 genes being involved in each transition, or by gaining a triple mutation to go to state $ADGR$ with 1 gene being involved in the transition. The logistic term caps the total population of non-normal, non-metastatic cells at 10^6 cells. What is not visible in this standard form of the ODEs but is present in the vector form is the fact that the logistic angiogenesis cap is only imposed when 10% or fewer of the non-normal, non-metastatic cells have a mutation in the A category. Finally, populations of non-normal, non-metastatic cells are always capped by a lethal tumor burden limit of 10^{13} cells.

4. Calculation of cell division and death rates

In order to calculate the change in cell division and death rates for mutations in R and D, we begin with the assumption that birth and death rates are equal for normal cells, that is, the cell division rate $= 1/b =$ cell death rate $= 1/d = \frac{1}{10}$ days ^{1}.

We use an approximation to the ODEs where we treat the rate of cells entering and leaving the state as negligible compared with the tumor volume doubling time, since they are several orders of magnitude different. Thus, for cells with mutations in D but not R, for example, we consider

$$\frac{dD}{dt} = D\left(\frac{1}{b} - \frac{1}{d_D}\right). \tag{7}$$

Solving this gives $D = D_0 \exp((1/b - 1/d_D)t)$. A doubling corresponds to $2 = \exp((1/b - 1/d_D)T_D)$. Similarly, for cells with mutations in R and not D, we arrive at $R = R_0 \exp((1/b_R - 1/d)t)$. Taking the natural logarithm of both sides leads to Eqs. (8a) and (8b),

$$\frac{\ln 2}{1/b_R - 1/d} = T_R, \tag{8a}$$

$$\frac{\ln 2}{1/b - 1/d_D} = T_D, \tag{8b}$$

where T_R is tumor volume doubling time for cells with a mutation in R but not D and equals $T + 50 \times 3$, where T_D is the tumor volume doubling time for cells with a mutation in D but not R and equals $T + 50 \times 7$, and where the $D : R$ importance ratio is $7 : 3$. The base tumor volume doubling time is T when the growing tumor has mutations in both D and R (500 days). The value 50 is chosen to give realistic doubling times for cells with mutations in D (but not R) and R (but not D), on the upper bound of observed tumor volume doubling times, where the cells typically have both mutations.

5. Kinetics of various paths to cancer

Given that multiple mutations are necessary to form a tumor, we are interested in whether the specific order of mutations is important. It is currently believed that the temporal sequence of mutations determines the propensity of tumor development (Arends, 2000). The extent to which genetic instability (G) determines the timing of tumorigenesis has been a controversial issue in cancer biology. Some have argued that an increased premalignant mutation rate (that is, acquiring a mutation in G early) is necessary for tumor development (Loeb, 1991; Rajagopalan et al., 2003). Others have argued that an increased cell division rate, offering more opportunities to accumulate mutations, is sufficient for tumorigenesis (Tomlinson and Bodmer, 1995, 1999; Sieber et al., 2003). Although the extent to which angiogenesis (A), decreased apoptosis (D), genetic instability (G), and increased replication rate (R) contribute to the development of cancer depends on the type of cancer involved, a better general understanding of the kinetics of various paths to cancer can be informative about their relative importance.

We explore the kinetics of various pathways to cancer by analysing the dynamics of the different cell populations. By plotting different sets of cell populations, we are able to identify the individual contribution of each mutation to the development of cancer. The growing populations of cells plateau at various points in the graphs due to the imposed 10^{13} cell population cap. In this model, the fastest pathway for tumor progression starts with a mutation in D (Fig. 2(a)) which increases the population of potential tumor cells. Next, a mutation in R is acquired, further increasing the population of cells by clonal expansion (Fig. 2(b)). After acquiring these two mutations, the tumor is sufficiently large to be inhibited by the angiogenesis cap imposed by the model. For this reason, a mutation in the angiogenesis category occurs next in the fastest path (Fig. 2(c)). Finally, a mutation in G follows. Fig. 2(d) shows the populations of tumor cells, T, and metastatic cells, M. Although the rate $k_3 = 10^{9}$ is very low, the large increase in population of T cells guarantees that eventually some cells successfully metastasize.

Our model predicts that genetic instability is more likely to be a feature of later-stage sporadic tumors, in accordance with the view of Tomlinson and Bodmer (1999). This is because a mutation in G has no direct selective advantage, only an indirect advantage through increasing the mutation rates in other genes. Although genetic instability can aid tumorigenesis, selection and clonal expansion are the main driving force for tumor progression in this model, a conclusion which has been proposed previously by Siber et al. (2003).

6. Effect of inherited mutations on cancer development

Here, we examine the effect of different inherited mutations on cancer development by varying our initial conditions. Since most inherited cancers are the result of mutations in tumor suppressors (Knudson, 2002), we model this situation by increasing the rate of transition from a normal cell to the appropriate mutated cell to 10^{5} mutations/gene/cell division. This models a case where a person inherits an inactivating mutation in one copy of the gene. These cells are still functionally "normal" (thus they begin in state N), but the chance of acquiring the second "hit" and losing functionality of the protein (moving into the mutated state) is much increased.

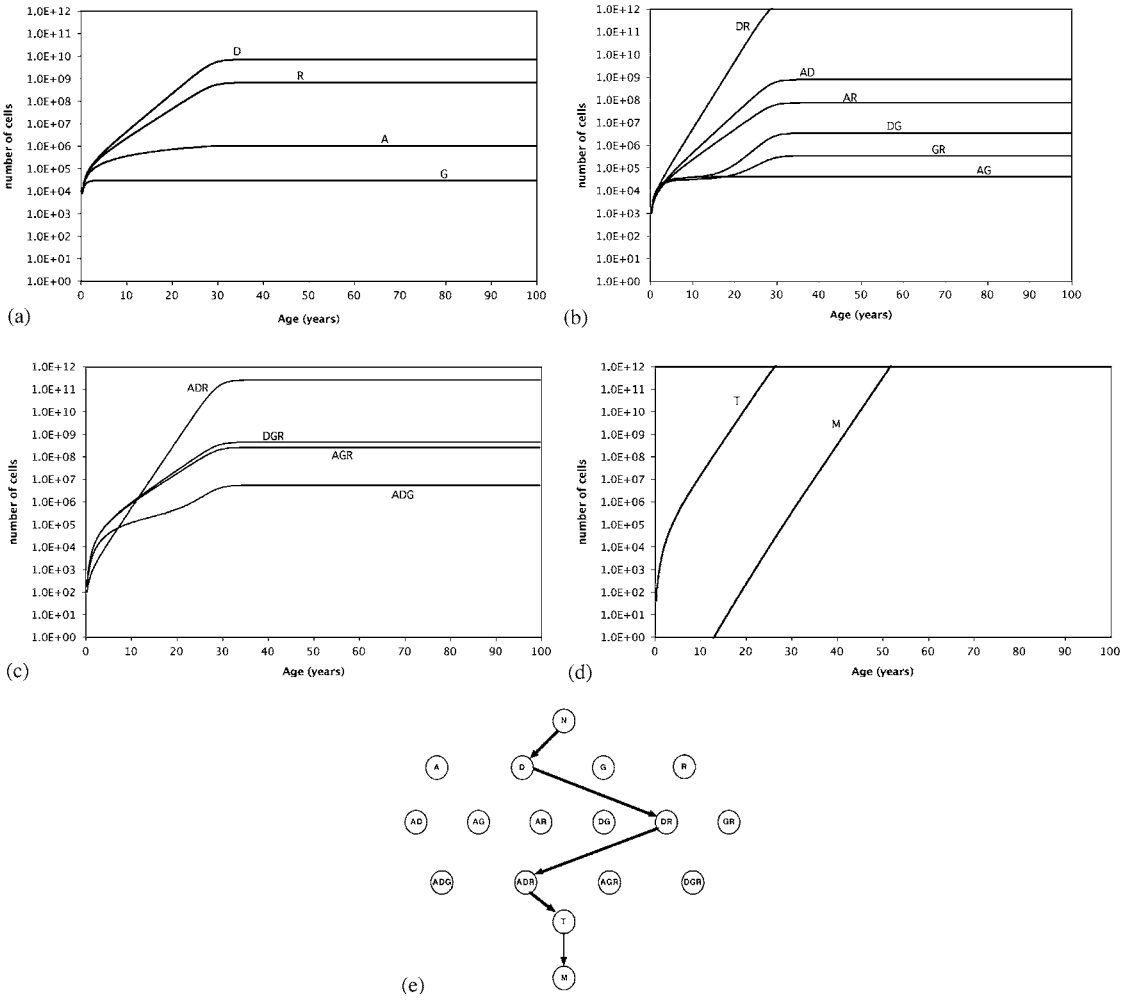

Fig. 2. Fastest path to cancer. (a) Dynamics of cell populations with one type of mutation. (b) Dynamics of cell populations with two types of mutations. (c) Dynamics of cell populations with three types of mutations. (d) Dynamics of cell populations with four types of mutations (T), and those that have metastasized (M). (e) The fastest path to cancer is by acquiring a mutation in D, then R, then A, then G.

As expected, inheriting a mutation in a cancer-critical gene decreases the time to cancer onset. The effects of inheriting a mutation in each category on time to reach 10^9 primary tumor cells, 10^{12} primary tumor cells, and 10^{12} metastatic cells are shown in Fig. 3. A tumor volume of 1 cm^3 weighs about 1 g and represents about 10^9 cells (Friberg and Mattson, 1997). This tumor size is regarded as relatively small in a clinical setting and it is at this size that a tumor may give rise to the first symptoms and may first become detectable by palpation (Friberg and Mattson, 1997). A tumor that weighs about 1 kg (10^{12} cells) is approaching the lethal tumor burden for a patient (Friberg and Mattson, 1997). The 10^{12} metastatic cells plotted in Fig. 3 are not necessarily localized to one site in the body; they could represent 10^{12} cells present in one location or 10^{11} cells present in each of 10 different locations, for example.

In contrast to the results obtained in Fig. 2, where the increased population of cells caused by mutations in D and R dominates the fastest path to sporadic cancer, *inheriting* a mutation in a G gene causes cancer onset at the earliest age. There is no observable difference between inheriting a mutation in one of the other categories (A, D, or R) and inheriting no mutations at all (N), partly due to the robustness of the model to changes in parameters, discussed in the Fig. 3 caption and in Section 7. In the fastest path plots (Fig. 2), there is equal probability of acquiring a mutation in A, D, G or R. D will dominate over G due to the fact that the transition from one state to another is a function not only of the mutation rates k_1 and k_2, but also the cell population size. Both D and G are equally likely to begin with, but since D increases the net cell population very quickly, it soon dominates over the rate k_2 associated with G. Therefore, the fastest path to sporadic cancer is through a mutation in D first.

In comparison, when a mutation in G is *inherited*, the cell has already surpassed the initial probability hurdle of acquiring a mutation in G. The rate of subsequent mutation is now 1000-fold higher and once a mutation

in D or R is obtained, the cell population will begin to increase. For this reason, an inherited mutation in G has the greatest effect. This is consistent with the fact that many inherited cancer syndromes are the result of a mutation in the G category. These include xeroderma pigmentosum, ataxia telangiectasia, Nijmegen breakage syndrome, hereditary non-polyposis colorectal cancer, and Bloom syndrome (Siber et al., 2003).

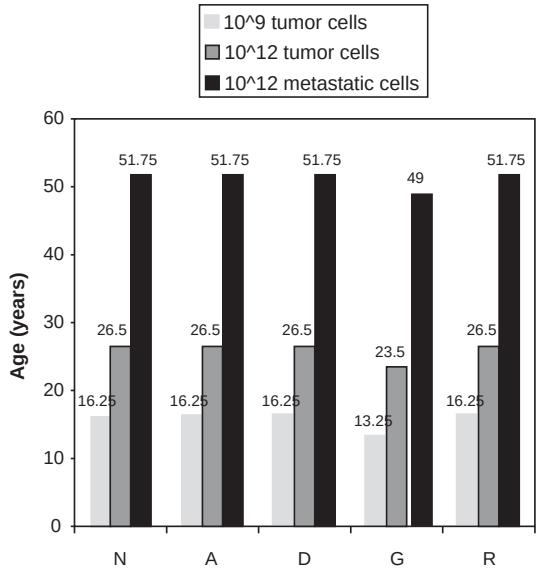

Fig. 3. Inherited mutations in cancer critical genes. Age at which a person may acquire 10^9 primary tumor cells, 10^{12} primary tumor cells, and 10^{12} metastatic cells with different inherited mutations. For reference, the case where no mutations are inherited is also shown (N). An inherited mutation in G is the only case which produces an earlier onset of cancer than the case where no mutations are inherited. Inheriting a mutation in G has a large effect because it allows the cell to surpass the initial probability hurdle of acquiring a mutation in G which then confers a 1000 fold increase in the rate of subsequent mutation. Increasing the initial probability of getting into state A, D, or R (that is, inheriting a mutation in A, D, or R) does not increase the time to cancer onset due to the number of cells that normally already build up in states D and R and due to the fact that mutations in A confer no benefit until 10^6 cells are obtained.

The time to develop a palpable primary tumor (10^9 cells) in this model is 16.25 years if no mutations are inherited (Fig. 3). Even taking into account the fact that detection of the tumor would not occur until several years later (Friberg and Mattson, 1997), this age of cancer onset is significantly earlier than the average age of cancer onset in the human population (DePinho, 2000). This is an indication of the need for more accurate information on cell division, cell death, and tumor doubling rates. Importantly, this may also be an indication that acquisition of mutations in more than four categories is necessary for development of a primary tumor. Adding two more steps to the multistep model would certainly delay the time to cancer, however to add those as detailed in Hanahan and Weinberg (2000) would be better left for an agent-based model (see Discussion). Consideration of the role of the immune system in curbing the growth of a tumor would also slow the time to cancer onset. Our model does not directly consider this factor, although category D does allow for apoptosis initiated by the immune system. Consideration of these three factors would allow the model to be more appropriately scaled to the timing of human cancer.

7. Sensitivity analysis of variations in the parameters

In order to determine the relative contributions of the parameters to the model, we vary each parameter listed in Table 2 while holding all others constant at the default value. The default values chosen are our best estimate from the literature. Except for *D:R importance ratio* where we use 3:7 to determine the effect on the fastest path, and *% A cells needed to remove cap* where we test the range from 0% to 100%, the other values are chosen to be near the upper and lower bounds of the

Table 2
Sensitivity of the model to changes in changes in parameters

Parameter	Default	Time	Path	Other	Time	Path	Other	Time	Path
Cell birth and death rates	$\frac{1}{10}$	51.75	DRAG	$\frac{1}{5}$	50.25	DRAG	$\frac{1}{30}$	54.75	DRAG
Tumor volume doubling time	500	51.75	DRAG	300	38.25	DRAG	700	64.75	DRAG
% A cells needed to remove cap	10%	51.75	DRAG	30%	51.75	DRAG	40%	57.50	DRAG
$D:R$ importance ratio	7:3	51.75	DRAG	8:2	50.50	DRAG	3:7	51.75	RDAG
Mut. rate with a G mutation	10^{-4}	51.75	DRAG	10^{-3}	50.50	DRAG	10^{-2}	50.50	DRAG
Mut. rate without a G mutation	10^{-7}	51.75	DRAG	10^{-6}	48.50	DRAG			
# of genes involved in transitions	100, 10, 1	51.75	DRAG	500, 100, 10	48.25	DRAG			

Cell birth and death rates have units days^{-1}. Tumor volume doubling times are measured in days. Mutation rates are measured as mutations/gene/cell division. Number of genes involved in transitions are listed as number involved in single, double, triple transitions. "Other" refers to different values tested in the sensitivity analysis. "Time" refers to age at acquisition of 10^{12} M cells for a variation in that parameter, measured in years. "Path" refers to the fastest path to cancer for a variation in that parameter.

range given in the literature. We assess the contribution of each by examining their effect on time to reach 10^{12} M cells and on the fastest pathway to cancer in Table 2.

The most salient result of the sensitivity analysis is the robustness of the model. Despite trials with very high values for k_2, the fastest path to somatic cancer is always via a mutation in D then R then A then G, except in the case where we flip the D:R importance ratio. As expected, a ratio of 3:7 flips the roles of D and R in the fastest path to give $RDAG$, but does not change the time to 10^{12} M cells from the default value of 51.75 years. Using a D:R ratio of 8:2 decreases the time to reach 10^{12} M cells due to the increased weight given to D.

The parameter that has the largest effect on time to reach 10^{12} M cells is the tumor volume doubling time. A tumor volume doubling time of 300 days decreases the time to reach 10^{12} M cells by 13.50 years relative to the default of 500 days, and a tumor volume doubling time of 700 days increases the time to reach 10^{12} M cells by 13.00 years. This effect is seen in Fig. 4(a) as well as in Table 2. The large effect of this parameter on the model is due to its impact on the $(\frac{1}{b}-\frac{1}{d})$ term; when cells have mutations in R and/or D, the terms become $1/b_R$ and/or $1/d_D$, allowing the cell populations to increase at a rate that reflects the tumor volume doubling time chosen.

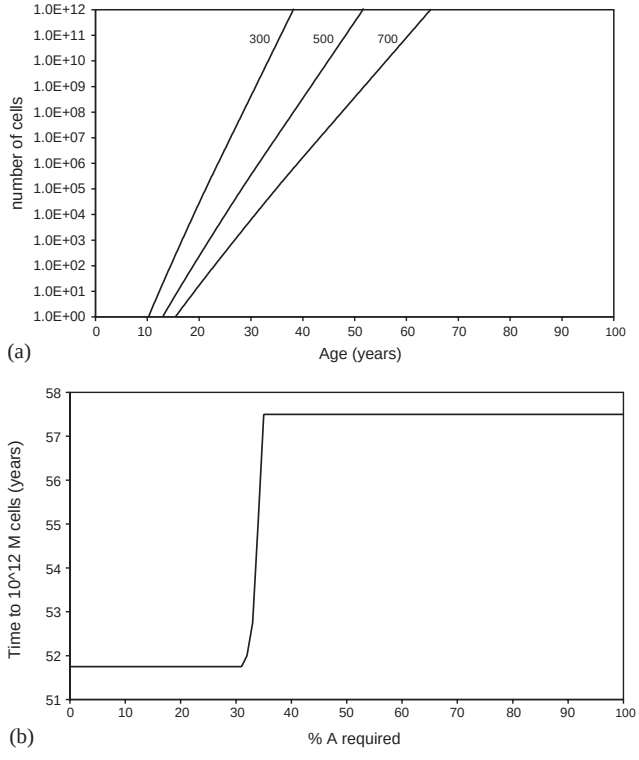

Fig. 4. Sensitivity of the model to changes in parameters. (a) Sensitivity to changes in tumor volume doubling time. Time to reach 10^{12} M cells is shown for tumor volume doubling times of 300, 500, and 700 days. (b) Effect of variations in percentage of A cells required to induce angiogenesis on time to reach 10^{12} M cells.

Variations in the birth and death rates to 1 in every 5 days and 1 in every 30 days also have an effect on time to reach 10^{12} M cells. This can be seen in row one of Table 2, but the effect is small when compared with the effect of tumor volume doubling time.

The time (51.75 years) to reach 10^{12} M cells does not change in varying the percentage of A cells needed to remove the angiogenesis cap from 0% to 31%. Between 31% and 35%, the time to reach 10^{12} M cells increases rapidly. The time (57.50 years) to reach 10^{12} M cells does not change in varying the percentage from 35% to 100%. This effect can be seen in Fig. 4(b). At 31%, the requirement for mutations in the A category begins to have an effect on the growing cell populations. At 35% and above, the percentage of A cells required is so large that the sum of the populations of non-normal, non-metastatic cells never goes above 10^6 because there are never at least 35% with A mutations. Thus the time to reach 10^{12} M cells depends only on a fixed number of T cells in each case, and remains constant at 57.50 years for percentages 35% and above.

A ten-fold change (from 10^{-7} to 10^{-6} mutations/gene/cell division) in mutation rate without a G mutation (k_1) has a larger effect on time to reach 10^{12} M cells than a ten-fold change (from 10^{-4} to 10^{-3} mutations/gene/cell division) in mutation rate with a G mutation (k_2). This is due to the fact that the effect of k_2 only becomes important later in tumorigenesis since G is last in the fastest path to cancer, whereas the effect of k_1 occurs at the beginning. There is no change in time to reach 10^{12} M cells when k_2 is increased from 10^{-3} to 10^{-2} mutations/gene/cell division because the effect of mutation rate has already saturated the system at a k_2 value of 10^{-3}.

Since the parameter *number of genes involved in transitions* is located in the numerator of the differential equations, increasing the number of genes involved in the transitions decreases time to reach 10^{12} M cells simply by making the numerator larger.

To determine whether we could obtain a more realistic age of cancer onset, we ran the model using all parameter values that would push back the time to cancer onset but are still in the biologically valid range. The parameters that we varied from the default settings are a tumor volume doubling time of 700 days, a cell birth and death rate of one every 30 days, and an angiogensis cap removal percentage of 40%. Running the model with these parameter adjustments results in a time to 10^9 tumor cells of 20.75 years, a time to 10^{12} tumor cells of 33.75 years, and a time to 10^{12} metastatic cells of 65.5 years. Although this time to cancer onset is more realistic, it does not represent our best estimates of parameter values from the current literature.

8. Discussion

This paper explores facets of the multistep model of oncogenesis. The key findings of this paper are:

(1) The fastest path to somatic cancer is predicted to be through gaining mutations in D (evasion of cell death), then R (increased replication rate), then A (angiogenesis), then G (increased mutation rate).
(2) Of the four categories of mutations, inheriting a mutation in G is predicted to produce cancer at the earliest age.
(3) The fastest path to somatic cancer is robust to realistic changes in parameters, with the model being most affected by variations in tumor volume doubling time.

The strength of our model lies not in its utility for predicting any one individual's time to cancer onset *per se*, but rather in the fact that it presents a novel approach to understanding the genetic basis of cancer from a systems biology perspective. Although a thorough testing of this model is not currently possible due to the "generic" nature of the model (i.e. specific types of cancer and specific genes are not mentioned), this model establishes the groundwork for future models that can be directly tied to clinical and molecular data pertaining to a specific type of cancer.

We hope that the creation of this model for the multistep progression to cancer will encourage biologists to gather quantitative data and will suggest which experiments should be performed with highest priority. The only parameter values which are reasonably agreed upon in the literature are the spontaneous mutation rate and the size to which a tumor can grow before angiogenesis is required. All other parameter values could use experimental refinement. In particular, experimental data that would help include direct measurement of tumor volume doubling time and measurement of the number of mutations in various gene categories in heterogeneous cancer cell populations. The experimental data on mutations should consider the ordering of acquisition of key genes in each of the categories and include a study of inherited mutations in these key genes. A number of genes that could be categorized and tracked include those in the Wellcome Trust Sanger Institute database (Futreal et al., 2004). Tracking mutations in A category genes along with a study of the timing of angiogenesis would allow an estimate of the percentage of A cells required to signal for successful angiogenesis and tumor growth past 10^6 cells. Estimates of cell division and death rates in cells with these mutations would also be useful. Our model is kept as general as possible; to make use of much of this experimental data the model would need to focus on a particular type of cancer, as many of these parameters are highly dependent on the originating tissue type.

To push back the time to cancer onset, we did try building a six step ODE model more like the one proposed by Hanahan and Weinberg with the same 4 steps (A, D, G, R) and two additional ones: L (limitless replicative potential, e.g. turning on telomerase) and M (invasion and metastasis, e.g. loss of E-cadherin). The main problem was that we could not accurately model L or M using ODEs. To properly model the effect of a beneficial mutation in L, we would need to store how many times each individual cell divides. A population-based ODE model ignores individual cells and their cell divisions, and so we only monitor cell populations. To properly model invasion, we need to consider the cells existing in 3D space. The ODEs can only model change in population over time, not through space, and we were unable to create equations that could reflect invasion by just using a rate change. An agent-based model would allow stochastic modeling of the six-step transformation to cancer. In this type of model, it would be possible to track the number of times an individual cell divides (category L) and where the cells are located in 3D space (category M).

Better estimates of parameter values, inclusion of two additional categories to give a total of six steps in the multistep model, and consideration of the role of the immune system in curbing the growth of a tumor will allow future models to be more appropriately scaled to human cancers. Modeling the multistep accumulation of genetic mutations in cancer will give insight into topical questions about the progression of a normal cell to a cancerous cell, enabling cancer treatments to be better targeted to various stages of cancer progression, and suggesting the most important directions for future experimental research.

Acknowledgements

We thank Trachette Jackson, Gilbert Omenn, and John Holland (University of Michigan), Andrew Allison, Samuel Mickan, David Findlay, and Brendon Coventry (The University of Adelaide), Setayesh Behin-Ain (The Queen Elizabeth Hospital, Australia), and Roger Reddel (Children's Medical Research Institute, Australia) for useful discussions, Sara Dempster (Massachusetts Institute of Technology) for her early contributions to this project, and the Santa Fe Institute for the opportunity to attend the Complex Systems Summer School where we began our exploration of this topic. This work was supported in part by scholarship funding from The Queen Elizabeth Hospital, Adelaide, Australia, arranged through Setayesh Behin-Ain.

References

Arends, J.W., 2000. Molecular interactions in the Vogelstein model of colorectal carcinoma. J. Pathol. 190, 412 416.

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

Cahill, D.P., Kinzler, K.W., Vogelstein, B., Lengauer, C., 1999. Genetic instability and darwinian selection in tumours. Trends Cell Biol. 9, M57 M60.

DePinho, R.A., 2000. The age of cancer. Nature 408, 248 254.

Folkman, J., 1990. What is the evidence that tumors are angiogenesis dependent? J. Natl Cancer Inst. 82, 4 6.

Friberg, S., Mattson, S., 1997. On the growth rates of human malignant tumors: implications for medical decision making. J. Surg. Oncol. 65, 284 297.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., Stratton, M.R., 2004. A census of human cancer genes. Nat. Rev. Cancer 4, 177 183.

Hanahan, D., Weinberg, R., 2000. The hallmarks of cancer. Cell 100, 57 70.

Jackson, A.L., Loeb, L.A., 1998. The mutation rate and cancer. Genetics 148, 1483 1490.

Knudson, A.G., 2002. Cancer genetics. Am. J. Med. Genet. 111, 96 102.

Koscielny, S., Tubiana, M., Valleron, A. J., 1985. A simulation model of the natural history of human breast cancer. Br. J. Cancer 52, 515 524.

Loeb, L.A., 1991. Mutator phenotype may be required for multistage carcinogenesis. Cancer Res. 51, 3075 3079.

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

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. 99 (25), 16226 16231.

Nowell, P.C., 1976. The clonal evolution of tumor cell populations. Science 194, 23 28.

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

Renan, M.J., 1993. How many mutations are required for tumorigen esis? implications from human cancer data. Mol. Carcinogen. 7, 139 146.

Rew, D.A., Wilson, G.D., 2000. Cell production rates in human tissues and tumours and their significance. Part II: clinical data. Eur. J. Surg. Oncol. 26, 405 417.

Sieber, O.M., Heinimann, K., Tomlinson, I.P.M., 2003. Genomic instability the engine of tumorigenesis? Nat. Rev. Cancer 3, 701 708.

Speer, J.F., Petrosky, V.E., Retsky, M.W., Wardewll, R.H., 1984. A stochastic numerical model of breast cancer growth that simulates clinical data. Cancer Res. 44, 4124 4130.

Tomlinson, I., Bodmer, W., 1999. Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. Nat. Med. 5, 11 12.

Tomlinson, I.P.M., Bodmer, W.F., 1995. Failure of programmed cell death and differentiation as causes of tumors: some simple mathematical models. Proc. Natl Acad. Sci. 92, 11130 11134.

Tomlinson, I.P.M., Novelli, M.R., Bodmer, W.F., 1996. The mutation rate and cancer. Proc. Natl Acad. Sci. 93, 14800 14803.

Vogelstein, B., Lane, D., Levine, A.J., 2000. Surfing the p53 network. Nature 408, 307 310.