Pre-Selective and Selective Phase in Tumor Development

Marek Kimmel[a,b] and Tomasz Wojdyla[b]

[a]*Department of Statistics, Rice University, 6100 Main St., Houston, TX 77005, USA*
[b]*Systems Engineering Group, Silesian Univ. of Technology, Akademicka 16, 44-100 Gliwice, Poland*

Abstract. This contribution presents a brief study of conditions for fixation of a cancer driver mutation if selective pressures change during individual's lifetime. A branching process model is used to represent the pre-selection stage, which leads to creation of highly heterogeneous subclones of transformed cells. A Moran model with selection is used to represent the second phase, which leads to development of a primary tumor. Computations and simulations allow determining feasibility of the mechanisms proposed.

Keywords: Cancer dynamics, driver mutations, selection, branching process, Moran Model
PACS: 02.50.Fz, 87.10.Mn, 87.17.Ee, 87.19.xj

INTRODUCTION: DRIVER AND PASSENGER MUTATIONS

Cancer is a disease in which cells derived from a specific lineage stop responding to hormonal cues, proliferate uncontrollably and not only de-differentiate but also frequently return to molecular programs which have been suppressed in the evolution from single-cell to multicellular eukaryotic organisms. This includes transition to anaerobic metabolism and to motility and invasion of new environments. The process initiating these transitions, called carcinogenesis, may be in part due to inherited mutations and in part to environmental or behavioral pressures such as radiation or exposure to chemicals, for example to the tobacco smoke. However, progression of cancer from early to advanced stage proceeds through a succession of genomic and metabolic changes that, as far as we know, are purely somatic.

The somatic mutations driving progression of cancer have been broadly characterized as the driver and passenger mutations [1]. Although the terminology is not quite firm, the driver mutations are those that confer the selective advantage to cells within the tumor environment but also against the healthy cells of the organism. Passenger mutations are neutral or slightly deleterious. Whether passenger mutations play active role in tumor evolution or serve exclusively as molecular clocks, has not been determined.

This paper is concerned with evolution of driver mutations in cancer, or more precisely with the mechanism by which a single driver may become fixed in in the tumor population. We are interested in the mechanism that leads from a mutation that occurs in a neutral environment, such as during the expansion of stem cell populations over the embryonic and fetal life, to a mutant clone that starts taking over when conditions create selective advantage for the mutant. An example may be exposure to toxic chemicals such as the polycyclic aromatic hydrocarbons (PAH) in tobacco smoke (a risk factor for lung cancer), or exposure to therapeutic doses of granulocyte colony stimulating factor (GCSF) used to treat immune disorders (a risk factor for acute myeloid leukemia (AML) [2]).

We hypothesize that random fluctuations at the tissue expansion level create enough mutants that although not detectable even by high depth genome sequencing, lead to detectable tumors at realistic ages of onset.

STOCHASTIC MODELING

We will be using two established paradigms of stochastic modeling. One is appropriate when the population of cells is expanding without environmental limit; these are the stochastic processes, and more precisely the Two-type Age-Dependent Markov Branching Process. The other one is appropriate when the mutant population evolves in a strictly confined environment; this is the Moran Process, adapted from population genetics. We will use a special version of the Moran process that we introduced; the Moran process with co-localization, which accounts in a simplified manner for selection in a non-ideally mixing environment. In what follows we will track the fates of a mutation that may become advantageous in and for the tumor and thus a driver mutation.

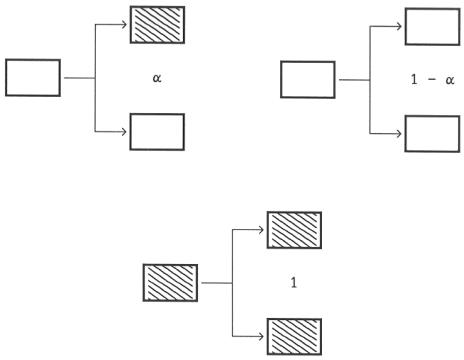
Neutral Expansion Phase: Branching Process

We use the branching process which has been developed to model drug resistance, and which has become influential in the cancer research community. It was originally published by Coldman and Goldie [3]. As demonstrated in the book by Kimmel and Axelrod [4] it is a Markov time-continuous branching process.

The assumptions of the model are as follows (Fig. 1(a)).

1. Cancer cell population is initiated by a single non-mutant cell. The cell population proliferates without losses.
2. Interdivision time of cells is a random variable with exponential distribution with a fixed parameter λ.
3. At each division, with given probability α, a single progeny cell mutates and becomes resistant to the cytotoxic agent.
4. Mutations are irreversible; mutants proliferate at the same rate as wild-type cells.

(a)



(b)

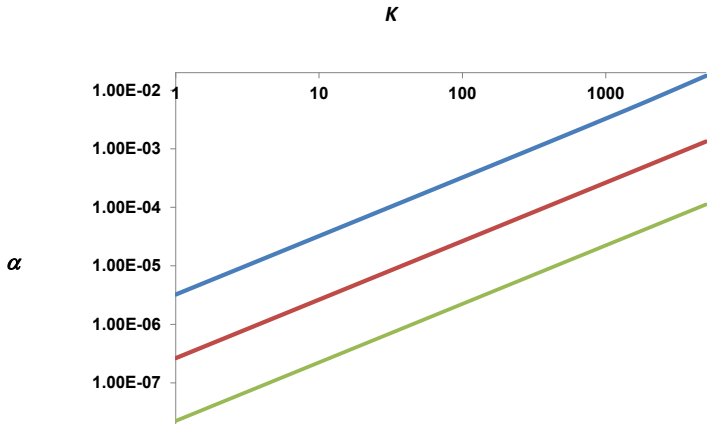


FIGURE 1. (a) Transition rules in the branching process with mutation (Goldie and Coldman model). (b) Relationship between K (final mutant count) and α (mutation rate) variables, for a process that reaches the finit size of N cells (*blue*, N = 30,000; *red*, N = 300,000; *green*, N = 3,000,000) which demonstrates that mutation rate α required for K = 100 varies around 10^{-5} per division.

This model implies, among other, that in a population started by a single non-mutant cell, at any time t, the expected count of mutant cells (K) and the total count of cells (N) are tied by a simple time-independent algebraic relationship

$$K(t) = N(t) - N(t)^{1-\alpha}$$

This relationship (Fig. 1(b)) allows realizing that in population of stem cells, corresponding to a specialized tissue such as large intestine or lung epithelium, or bone marrow, even low frequency mutations produce enough mutant cells that, although not detectable by deep sequencing, may come to dominate in selective conditions.

Selection Phase: Moran Process with Co-Localization

For a simplified model of adult-marrow competition between mutant and wild type (WT) cells, we will use the Moran process with directional selection [5]. In this process (Fig. 2a), the total population of cells is considered constant, with variable in time proportion of mutants and time in discrete units (such as days). We consider a population of N biological cells, which at time 0 contains K mutant cells. The mutant has selective advantage expressed by the relative fitness r, equal to the ratio of average progeny count of the mutant to that of the WT. For an advantageous mutant, $r > 1$. Under Moran Model, the probability of fixation of the mutant is equal to $\Pr[T_N < T_0] = [1-(1-s)^i]/[1-(1-s)^N]$, where T_0 and T_N are times to extinction or fixation of the mutant and $r = (1-s)^{-1} \cong 1+s$. For large N, the expected time to fixation, given fixation occurs, is asymptotically equivalent to $E_1[T_N \mid T_N < T_0] \sim (2\ln N - \ln K)/s$ [5].

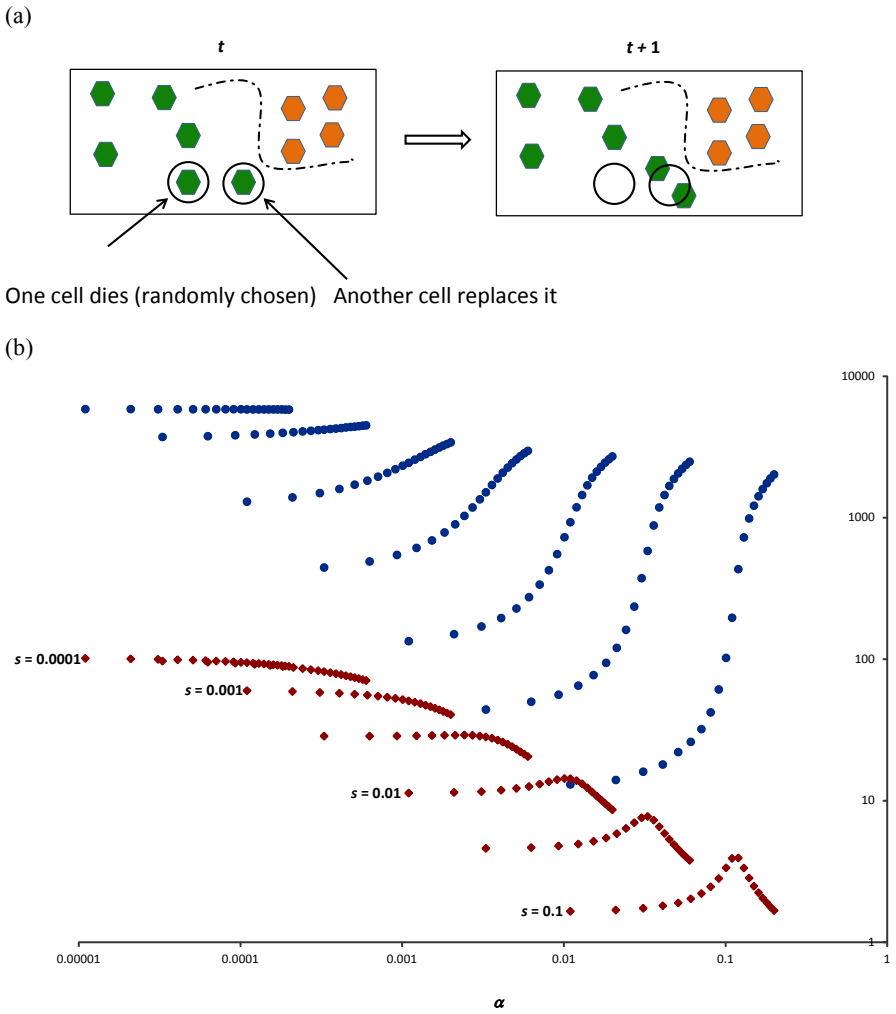


FIGURE 2. (a) Schematic depiction of the Moran model with co-localization (*green*, wild type cells; *orange*, mutant cells). (b). Numerical results of computing the expected age at fixation of mutant (red diamonds) and the initial count of mutant cells (*blue circles*) with probability 0.7, using the Moran model with co-localization, with a range of values of selection and co-localization coefficients.

In our version of the Moran process we also include the impact of what may be called "co-localization" (Fig. 2(a)). This is a property, which reflects, in a simplified and synthetic manner, the spatial aspects of selection. Intuitively, it is a correction to the Moran model, in which, in addition to selection, there exists a predisposition to replace a deceased individual by an individual of like type (mutant by mutant, wildtype by wildtype). In other words, it is not only that the event of replacing of the deceased cell by a mutant is more probable than the event of

replacing it by a wildtype, but these events have different conditional probabilities given the type of the deceased. This happens for example when the competing wild type and mutant cells are not perfectly mixed but occur in clumps. Then, if the space liberated by a deceased cell is preferentially occupied by its closest neighbor, this neighbor is more likely to be of like type. Co-localization is represented by a coefficient $\alpha \geq 0$, with the zero value corresponding to the standard Moran model.

$$p_{i,i-1} = \frac{i}{N}\frac{N-i}{(1+\alpha)ri+(N-i)}, \quad p_{i,i+1} = \frac{N-i}{N}\frac{ri}{ri+(1+\alpha)(N-i)}, \quad p_i = 1 - p_{i,i-1} - p_{i,i+1}.$$

To obtain the exact expression for the expected time to fixation we can use the general expression for the probability that in a Markov chain, the first instance of hitting state *j* at step *n* starting from state *i* equals to f_n^i, the *i*-th entry of the column vector f_n

$$f_n = (P^{(j)})^{n-1} f_1$$

with entries $(f_n^0,\ldots,f_n^N)$ and $P^{(j)}$ is the transition probability matrix with the *j*-th column replaced by a column of 0's. Vector f_1 is set as the *j*-th column of the transition probability matrix *P*. For technical purposes, instead of matrix *P*, we use matrix $\widetilde{P}$, which has $\widetilde{P}_{00} = \widetilde{P}_{NN} = 0$. This does not change the outcome of the iteration, but allows inversion of the matrices in expressions that follow. Finally, iterations yield

$$\sum_{n\geq 1} f_n = (I - \widetilde{P}^{(j)})^{-1} f_1, \quad \sum_{n\geq 1} n f_n = (I - \widetilde{P}^{(j)})^{-2} f_1.\cdot$$

Since the first entrance into *N* is equivalent to absorption, then the *N*-th entry of the two vectors above yield $P[T_0 > T_N] = \sum_n f_n^N$ and $E[T_N; T_0 > T_N] = \sum_n n f_n^N$. Application of these expressions requires inversion of very large matrices, but this is accomplished efficiently because of the matrices' band structure.

Figure 2(b) depicts the numerical results of computing the expected age at fixation, with probability 0.7, of mutant (*red diamonds*) and the initial count of mutant cells (*blue circles*) using the Moran model with co-localization, with a range of values of selection and co-localization coefficients. Let us note that first, given the fixation probability, colocalization decreases the age at which fixation occurs, but increases the number of initial mutants required. Second, if we take for example selection coefficient around 0.001, with a similar co-localization coefficient, this will result in around 100 mutant cells needed to obtain mutant fixation at the age of around 30 years. With different parameters, we obtain somewhat different results, but still within a realistic range.

DISCUSSION

We tracked the fate of a mutation which starts as a neutral and rare at the expansion phase of the stem cell population and which becomes fixed because of selective pressure in the selection phase. As indicated, employing simplest stochastic models, we find a realistic range of ages of onset of the disease (following fixation of the driver mutation). Our model is somewhat similar to that of Tomassetti et al. [6], and the topics related to passenger and driver mutations are discusses in many paper such as for example [7]. A specialized variant of this model fitting myelodysplastic syndrome data has been developed in reference [8].

ACKNOWLEDGMENTS

The authors were supported by the National Science Center (Poland) grant nr DEC-2012/05/N/NZ2/01647.

REFERENCES

1. C. Greenman, et al., *Nature* **446**, 153-158 (2007).
2. T. Glaubach and S. J. Corey, *Blood* **119,** 5063-5064 (2012).
3. J. H. Goldie A. J. Coldman, *Cancer Treatment Reports* **63,** 1727-1733 (1978).
4. D. E. Axelrod and M. Kimmel, *Branching Processes in Biology,* New York: Springer, 2015.
5. R. Durrett, *Probability models for DNA sequence evolution*, New York: Springer Science & Business Media, 2008.
6. C. Tomasetti et al., *Proceedings of the National Academy of Sciences* **110**, 1999-2004 (2013).
7. M. Kimmel and S. J. Corey, *Frontiers in Oncology*, **3** (2013) 89. doi:10.3389/fonc.2013.00089..
8. T. P. Wojdyla et al., *PLoS Computational Biology, submitted* (2015).