TTP: Tool for Tumor Progression

Johannes G. Reiter[1], Ivana Bozic[2,3], Krishnendu Chatterjee[1],
and Martin A. Nowak[2,3,4]

[1] IST Austria (Institute of Science and Technology Austria), Klosterneuburg, Austria
[2] Program for Evolutionary Dynamics, Harvard University, Cambridge, USA
[3] Department of Mathematics, Harvard University, Cambridge, USA
[4] Department of Organismic and Evolutionary Biology,
Harvard University, Cambridge, USA

Abstract. In this work we present a flexible tool for tumor progression, which simulates the evolutionary dynamics of cancer. Tumor progression implements a multi-type branching process where the key parameters are the fitness landscape, the mutation rate, and the average time of cell division. The fitness of a cancer cell depends on the mutations it has accumulated. The input to our tool could be any fitness landscape, mutation rate, and cell division time, and the tool produces the growth dynamics and all relevant statistics.

1   Introduction

Cancer is a genetic disease which is driven by the somatic evolution of cells [1], where driver mutations for cancer increase the reproductive rate of cells through different mechanisms, e.g. evading growth suppressors, sustaining proliferative signaling, or resisting cell death [2]. Tumors are initiated by some genetic event which increases the reproductive rate of previously normal cells. The evolution of cancer (malignant tumor) is a multi-step process where cells need to receive several mutations subsequently [3]. This phase of tumor progression is characterized by the uncontrolled growth of cells [2,4]. The requirement to accumulate multiple mutations over time explains the increased risk of cancer with age.

There are several mathematical models to explain tumor progression and the age incidence of cancer [5,6,7]. The models have also provided quantitative insights in the evolution of resistance to cancer therapy [8]. The models for tumor progression are multi-type branching processes which represent an exponentially growing heterogeneous population of cells, where the key parameters for the process are: (i) the fitness landscape of the cells (which determine the reproductive rate), (ii) the mutation rate (which determines the accumulation of driver mutations), and (iii) the average cell division time (or the generation time for new cells). The fitness landscapes allow the analysis of the effects of interdependent (driver) mutations on the evolution of cancer [9].

In this work, we present a very flexible tool (namely, TTP, tool for tumor progression) to study the dynamics of tumor progression. The input to our tool could be any fitness landscape, mutation rate, and cell division time, and the tool generates the growth dynamics and all relevant statistics (such as the expected

tumor detection time, or the expected appearance time of surviving mutants, etc). Our stochastic computer simulation is an efficient simulation of a multi-type branching process under all possible fitness landscapes, driver mutation rates, and cell division times.

Our tool provides a quantitative framework to study the dynamics of tumor progression in different stages of tumor growth. Currently, the data to understand the effects of complex fitness landscapes can only be obtained from patients or animals suffering the disease. With our tool, playing with the parameters, once the real-world data is reproduced, the computer simulations can provide many simulation examples that would aid to understand these complex effects. Moreover, once the correct mathematical models for specific types of cancer are identified where the simulations match the real-world data, verification tools for probabilistic systems can be used to further analyze and understand the tumor progression process (such an approach has been followed in [10] for the verification of biological models). In this direction, results of specific fitness landscapes of our tool have already been used in a biological application paper [11]. While we present our tool for the discrete-time process (which provides a good approximation of the continuous-time process), results of our tool for the special case of a uniform fitness landscape in the continuous-time process have also been shown to have excellent agreement with the real-world data for the time to treatment failure for colorectal cancer [8]. Full version available at [12].

2   Model

Tumor progression is modeled as a discrete-time branching process (Galton-Watson process [13]). At each time step, a cell can either divide or die. The phenotype i of a cancerous cell determines its division probability b_i and is encoded as a bit string of length four (i.e. $\{0,1\}^4$). The death probability d_i follows from b_i as $d_i = 1 - b_i$. If a cell divides, one of the two daughter cells can receive an additional mutation (i.e., a bit flips from wildtype 0 to the mutated type 1) with probability u in one of the wildtype positions (e.g., cells of phenotype 1010 can receive an additional mutation only at positions two and four; cells of phenotype 1111 can not receive any additional mutations). The branching process is initiated by a single cell of phenotype $i = 0000$ (resident cell). The resident cells are wildtype at all four positions and have a strictly positive growth rate (i.e., $b_{0000} - d_{0000} > 0$).

Fitness Landscapes. Our tool provides two predefined fitness landscapes for driver mutations in tumor progression: (1) Multiplicative Fitness Landscape (MFL) and (2) Path Fitness Landscape (PFL). Additionally, the user can also define its own general fitness landscape (GFL). A fitness landscape defines the birth probability b_i for all possible phenotypes i. Following the convention of the standard modeling approaches, we let $b_{0000} = 1/2(1 + s_0)$ be the birth probability of the resident cells (i.e., cells of phenotype 0000) [9,11]. The growth coefficient s_j indicates the selective advantage provided by an additional mutation at position j in the phenotype.

Multiplicative Fitness Landscape. In the MFL a mutation at position j of the phenotype i of a cell results in a multiplication of its birth probability by $(1+s_j)$. Specifically, the birth probability b_i of a cell with phenotype i is given by

$$b_i = \frac{1}{2}(1 + s_0) \prod_{j=1}^{4} (1 + \widehat{s}_j);$$

where $\widehat{s}_j = 0$ if the j-th position of i is 0; otherwise $\widehat{s}_j = s_j$. Hence, each additional mutation can be weighted differently and provides a predefined effect $(1+s_1), (1+s_2), (1+s_3)$, or $(1+s_4)$ on the birth probability of a cell. Additional mutations can also be costly (or neutral) which can be modeled by a negative s_j (or $s_j = 0$). If $s_0 = s_1 = s_2 = s_3 = s_4$ the fitness landscape reduces to the model studied by Bozic et al. [9], which we call EMFL (Equal Multiplicative Fitness Landscape) and is also predefined in our tool.

Path Fitness Landscape. The PFL defines a certain path on which additional mutations need to occur to increase the birth probability of a cell. The predefined path can be $0000 \rightarrow 1000 \rightarrow 1100 \rightarrow 1110 \rightarrow 1111$, and again the growth coefficients s_j determine the multiplicative effect of the new mutation on the birth probability (see appendix of [12] for more details). Mutations not on this path are deleterious for the growth rate of a cell and its birth probability is set to $1/2(1 - v)$. The parameter v $(0 \leq v \leq 1)$ specifies the disadvantage for cells of all phenotypes which do not belong to the given path.

General Fitness Landscapes. Our tool allows to input any fitness landscape as follows: for b_i for $i \in \{0,1\}^4$, our tool can take as input the value of b_i. In this way, any fitness landscape can be a parameter to the tool.

Density Limitation. In some situations, a tumor needs to overcome current geometric or metabolic constraints (e.g. when the tumor needs to develop blood vessels to provide enough oxygen and nutrients for further growth [14,11]). Such growth limitations are modeled by a density limit (carrying capacity) for various phenotypes. Hence, the cells of a phenotype i grow first exponentially but eventually reach a steady state around a given carrying capacity K_i. Only cells with another phenotype (additional mutation) can overcome the density limit. Logistic growth is modeled with variable growth coefficients $\widetilde{s}_j = s_j(1 - X_i/K_i)$ where X_i is the current number of cells of phenotype i in the tumor. In this model, initially $\widetilde{s}_j \approx s_j$ $(X_i \ll K_i)$, however, if X_i is on the order of K_i, $\widetilde{s}_j$ becomes approximately zero (details are given in the appendix of [12]).

3   Tool Implementation and Experimental Results

Our tool provides an efficient implementation of a very general tumor progression model. Essentially, the tool implements the above defined branching processes to simulate the dynamics of tumor growth and to obtain statistics about the

expected tumor detection time and the appearance of additional driver mutations during different stages of disease progression. TTP can be downloaded from here: http://pub.ist.ac.at/ttp.

For an efficient processing of the discrete-time branching process, the stochastic simulation samples from a multinomial distribution for each phenotype at each time step [9,11]. The sample returns the number of cells which divided with and without mutation and the number of cells which died in the current generation (see the appendix of [12] for details). From the samples for each phenotype the program calculates the phenotype distribution in the next generation. Hence, the program needs to store only the number of cells of each phenotype during the simulation. This efficient implementation of the branching process allows the tool to simulate many patients within a second and to obtain very good statistical results in a reasonable time frame.
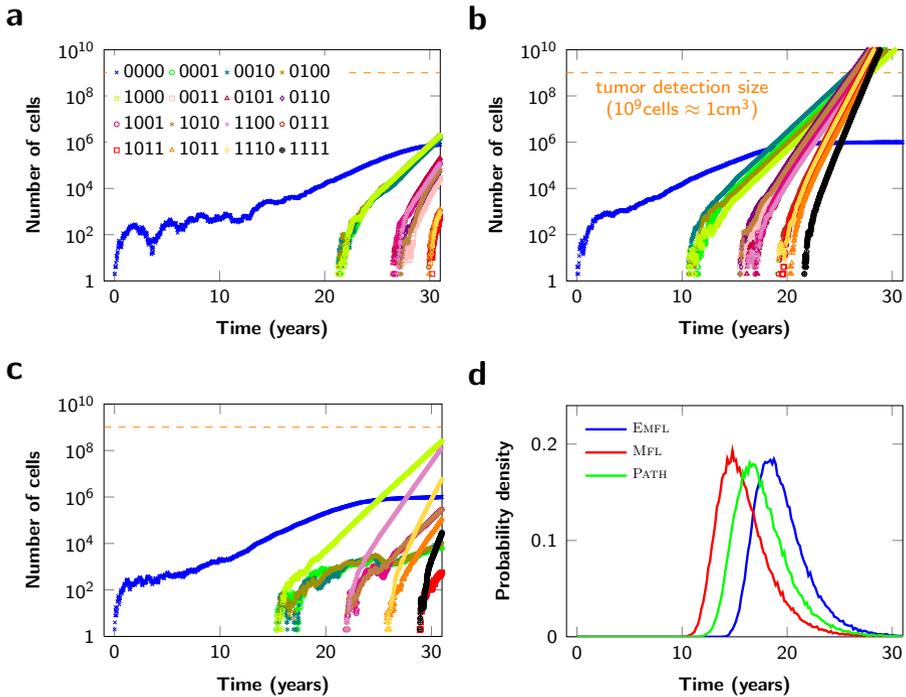
Fig. 1. Experimental results illustrating the variability of tumor progression. In panels a and b we show examples for two particular simulation runs where the cells grow according to the EMFL and resident cells (blue) are constrained by a carrying capacity of $K_{0000} = 10^6$. In panel c the cells grow according to the PFL. In panel d we show statistical results for the probability density of tumor detection when cells grow according to different fitness landscapes. Parameter values: growth coefficients $s_0 = 0.004$, $s_1 = 0.006$, $s_2 = 0.008$, $s_3 = 0.01$, $s_4 = 0.012$, and $v = 0.01$, mutation rate $u = 10^{-6}$, cell division time $T = 3$ days, tumor detection size 10^9 cells.

Modes. The tool can run in the following two modes: individual or statistics. In the *individual mode* the tool produces the growth dynamics of one tumor in a patient (see panels a, b, and c in Fig. 1). Furthermore, both the growth dynamics and the phenotype distribution of the tumor are depicted graphically. In the *statistics mode* the tool produces the probability distribution for the detection time of the tumor (see panel d in Fig. 1) both graphically and quantitatively. Additionally, the tool calculates for all phenotypes the appearance times of the first surviving lineage, the existence probability, and the average number of cells at detection time.

Features. TTP provides an intuitive graphical user interface to enter the parameters of the model and shows plots of the dynamics during tumor progression, the phenotype distribution, or the probability density of tumor detection. These plots can also be saved as files in various image formats. Furthermore, the tool can create data files (tab-separated values) of the tumor growth history and the probability distribution of tumor detection for any set of input parameters (details on the format are given in the appendix).

Input Parameters. In both modes, the tool takes the following input parameters: (i) growth coefficients s_0, s_1, s_2, s_3, and s_4 (and v in the case of PFL), (ii) mutation rate u, (iii) cell generation time T, (iv) fitness landscape (MFL, PFL, EMFL, or GFL with the birth probability for each phenotype), and optional (v) density limits for some phenotypes. In the individual mode, additionally, the user needs to provide the number of generations which have to be simulated. In the statistics mode, the additional parameters are: the tumor detection size and the number of patients (tumors which survive the initial stochastic fluctuations) which have to be simulated.

Experimental Results. In panels a, b, and c of Fig. 1 we show examples of the growth dynamics during tumor progression. Although we used exactly the same parameters in panels a and b, we observe that the time from tumor initiation until detection can be very different. In panel d we show the probability density of tumor detection under various fitness landscapes. Further experimental results are given in the appendix.

Case Studies. Several results of these models have shown excellent agreement with different aspects of real-world data. In [9], results for the expected tumor size at detection time using a EMFL fit the reported polyp sizes of the patients very well. Similarly, using a continuous-time branching process and a uniform fitness landscape, results for the expected time to the relapse of a tumor after start of treatment agree thoroughly with the observed times in patients [8].

Future Work. In some ongoing work, we also investigate mathematical models for tumor dynamics occurring during cancer treatment modeled by a continuous-time branching process. Thus an interesting extension of our tool would be to model treatment as well. Another interesting direction is to model the seeding of metastasis during tumor progression and hence simulate a "full" patient rather than the primary tumor alone. Once faithful models of the evolution of cancer have been identified, verification tools such as PRISM [15] and theoretical results such as [16] might contribute to the understanding of these processes.

References

1. Vogelstein, B., Kinzler, K.W.: Cancer genes and the pathways they control. Nature Medicine 10(8), 789–799 (2004)
2. Hanahan, D., Weinberg, R.A.: Hallmarks of cancer: the next generation. Cell 144(5), 646–674 (2011)
3. Jones, S., Chen, W.D., Parmigiani, G., Diehl, F., Beerenwinkel, N., Antal, T., Traulsen, A., Nowak, M.A., Siegel, C., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Willis, J., Markowitz, S.D.: Comparative lesion sequencing provides insights into tumor evolution. PNAS 105(11), 4283–4288 (2008)
4. Nowak, M.A.: Evolutionary Dynamics: Exploring the equations of life. The Belknap Press of Harvard University Press, Cambridge (2006)
5. Komarova, N.L., Sengupta, A., Nowak, M.A.: Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. Journal of Theoretical Biology 223(4), 433–450 (2003)
6. Iwasa, Y., Michor, F., Nowak, M.A.: Stochastic tunnels in evolutionary dynamics. Genetics 166(3), 1571–1579 (2004)
7. Nowak, M.A., Michor, F., Komarova, N.L., Iwasa, Y.: Evolutionary dynamics of tumor suppressor gene inactivation. PNAS 101(29), 10635–10638 (2004)
8. Diaz, L.A., Williams, R.T., Wu, J., Kinde, I., Hecht, J.R., Berlin, J., Allen, B., Bozic, I., Reiter, J.G., Nowak, M.A., Kinzler, K.W., Oliner, K.S., Vogelstein, B.: The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature 486(7404), 537–540 (2012)
9. Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., Nowak, M.A.: Accumulation of driver and passenger mutations during tumor progression. PNAS 107(43), 18545–18550 (2010)
10. Sadot, A., Fisher, J., Barak, D., Admanit, Y., Stern, M., Hubbard, E., Harel, D.: Toward verified biological models. IEEE/ACM Transactions on Computational Biology and Bioinformatics 5(2), 223–234 (2008)
11. Reiter, J.G., Bozic, I., Allen, B., Chatterjee, K., Nowak, M.A.: The effect of one additional driver mutation on tumor progression. Evolutionary Applications 6(1), 34–45 (2013)
12. Reiter, J.G., Bozic, I., Chatterjee, K., Nowak, M.A.: TTP: Tool for Tumor Progression. CoRR abs/1303.5251 (2013), http://arxiv.org/abs/1303.5251
13. Haccou, P., Jagers, P., Vatutin, V.A.: Branching Processes: Variation, Growth, and Extinction of Populations. Cambridge University Press (2005)
14. Kerbel, R.S.: Tumor angiogenesis: past, present and the near future. Carcinogenesis 21(3), 505–515 (2000)
15. Hinton, A., Kwiatkowska, M.Z., Norman, G., Parker, D.: Prism: A tool for automatic verification of probabilistic systems. In: Hermanns, H., Palsberg, J. (eds.) TACAS 2006. LNCS, vol. 3920, pp. 441–444. Springer, Heidelberg (2006)
16. Etessami, K., Stewart, A., Yannakakis, M.: Polynomial time algorithms for multitype branching processesand stochastic context-free grammars. In: STOC, pp. 579–588 (2012)