Near-criticality underlies the behavior of early tumor growth

2016 Phys. Biol. 13 026005

(http://iopscience.iop.org/1478-3975/13/2/026005)

IOP Publishing Phys. Biol. 13 (2016) 026005 doi:10.1088/1478-3975/13/2/026005

Physical Biology

RECEIVED
21 August 2015

REVISED
9 February 2016

ACCEPTED FOR PUBLICATION
16 February 2016

PUBLISHED
4 April 2016

PAPER

Near-criticality underlies the behavior of early tumor growth

Guillaume Remy[1,2] **and Philippe Cluzel**[1,3]

[1] FAS Center for Systems Biology, Department of Molecular and Cellular Biology, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, United States of America
[2] ICFP Physics Master Program, Ecole Normale Superieure de Paris, 45 Rue d'Ulm, Paris 75005, France
[3] Author to whom any correspondence should be addressed.

E-mail: cluzel@mcb.harvard.edu

Keywords: branching process, phase transition, stochastic behavior, tumor growth, stem cells, Ising, systems biology

Supplementary material for this article is available online

Abstract

The controlling factors that underlie the growth of tumors have often been hard to identify because of the presence in this system of a large number of intracellular biochemical parameters. Here, we propose a simplifying framework to identify the key physical parameters that govern the early growth of tumors. We model growth by means of branching processes where cells of different types can divide and differentiate. First, using this process that has only one controlling parameter, we study a one cell type model and compute the probability for tumor survival and the time of tumor extinction. Second, we show that when cell death and cell division are perfectly balanced, stochastic effects dominate the growth dynamics and the system exhibits a near-critical behavior that resembles a second-order phase transition. We show, in this near-critical regime, that the time interval before tumor extinction is power-law distributed. Finally, we apply this branching formalism to infer, from experimental growth data, the number of different cell types present in the observed tumor.

1. Introduction

Over the last several decades there have been series of models that characterize the growth of cancerous tumors. However, these models use a large number of parameters, which makes the identification of the principles that govern the growth intractable. The difficulty of modeling tumor growth is also compounded with the fact that a wide array of different situations have to be taken into account: the type of cancer, the stage of the illness, the type or shape of tumors, and the different cell types that constitute the tumor. For all these reasons, methods used to model tumor growth vary widely and depend on these specific situations. Similarly, mathematical tools differ greatly from one model to another, and some approaches use partial differential equations to model the tumor growth as a continuous mass [1, 2], while others use a more agent based approach where each individual cell is a discrete object [3–6] (some even do a mix of both [7]). Some models use continuous time while others use discrete intervals built from a fixed cell division time [4, 8]. Overall, the real challenge of theoretical modeling lies in reproducing quantitatively the behavior of a tumor growth without using too many parameters to make the computational approach tractable. Consequently, a balance needs to be found between using a large number of parameters with the risk of over-fitting the experimental data, or too few, a condition under which key features of the system can be missed. Here, we provide a unified view that aims at identifying the minimum number of parameters that underlies early tumor growth.

For sake of simplicity, our study will focus on one class of models of tumor growth: discrete time branching processes. In these models, at given time steps, different cell types can divide, differentiate or die. In this paper, we will not consider any physical or biochemical interactions between cells and the fate of each cell at each step will be chosen regardless of its past, which is the signature of a memoryless system. Such simple models have been used to describe recent experimental data with single cell resolution [9, 10]. As we will show, however, it is possible to achieve the same level of prediction with much fewer parameters. We will first demonstrate that the branching model used in [9] can be reduced to one cell type to reproduce the key properties observed in the experimental system.

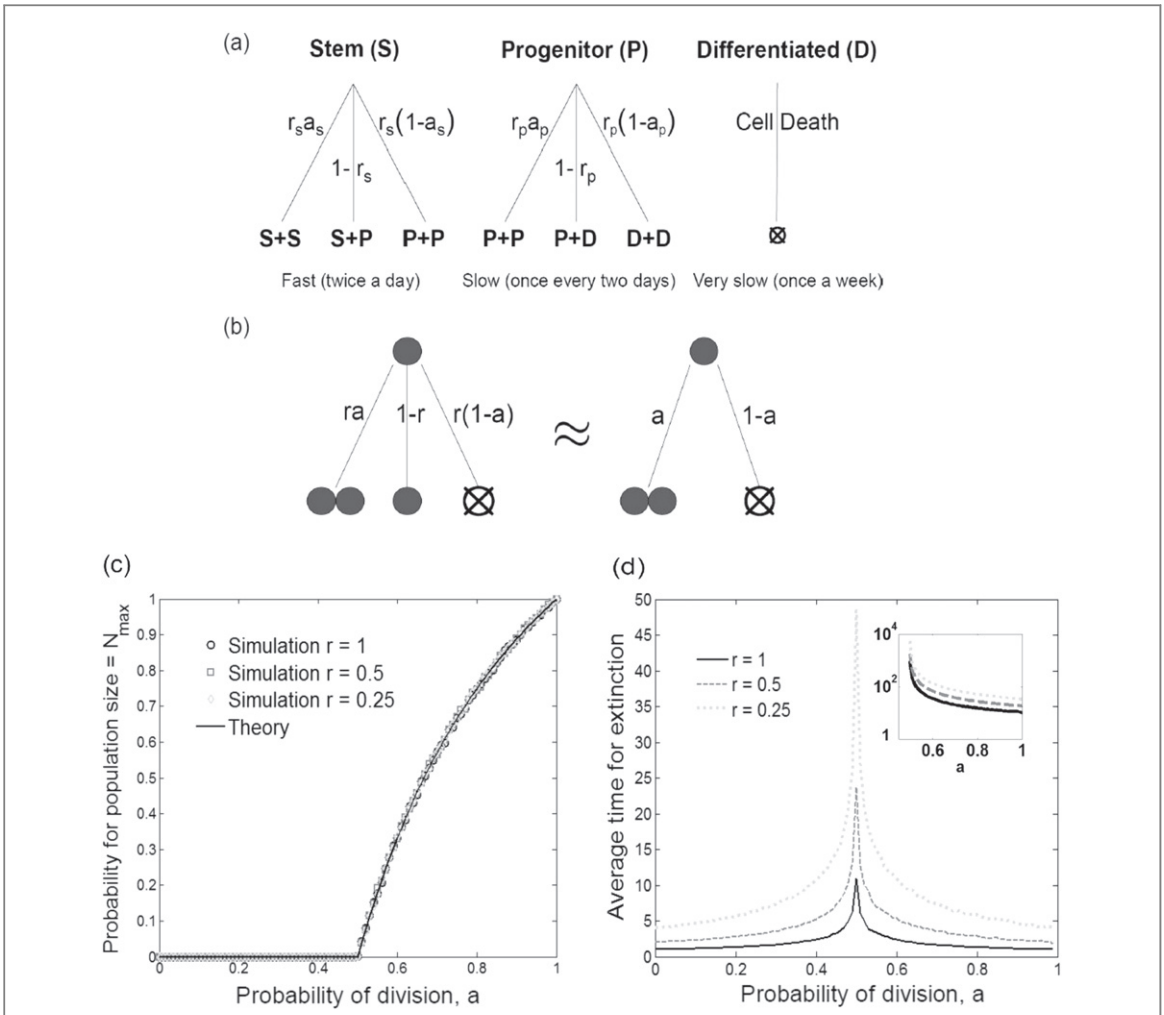

Figure 1. (a) This diagram describes an intricate branching process model used in [9]. This model describes the clonal size distributions of certain tumor cells observed *in vivo*. The parameters are set equal to: $a_s = a_p = 0.5$ and $r_s = r_p = 0.2$. (b) This cartoon shows the first procedure we will follow to simplify models such as (a). Given a 3-branch model where the cells have three possible fates at each step (division, death and doing nothing), we want to show we obtain the same general behavior when we reduce it to a 2-branch model where the cell can only die or divide at each step. Plots (c) and (d) illustrate this statement: r is just a delay parameter, a is the parameter that determines the behavior of the system. (c) Probability for the population size to cross $N_{\max}$ for different values of r (the 2-branch model corresponds to the case $r = 1$). (d) The average time for extinction as well as the average time to reach $N_{\max}$ (plotted in inset) have similar graphs for the different values of r. For smaller r the delay and these times get longer. (Both (c) and (d) have been plotted with $N_0 = 1$, $N_{\max} = 1000$ and using 10000 simulation runs.)

Using this simpler model allows us to identify a control parameter. We find for a specific value of the key control parameter a behavior similar to that of a second order-phase transition, at which the average time for tumor extinction diverges. Finally, we will extend our model to several cell types [9] and we will argue that our formalism allows us to predict from the growth rate the number of distinct cellular types present in growing tumors.

1.1. Using branching models

We start by characterizing a branching model (figure 1(a)) that was used to explain clonal size distributions in skin tumors [9]. The model considers three cell types: stem cells, progenitor cells and differentiated cells: (i) differentiated cells cannot differentiate further and eventually die off with a constant rate (once a week in our model). (ii) Stem cells can divide into stem cells or differentiate into progenitors. (iii) Progenitors divide into progenitors or differentiated cells. We model these different cell types by implementing two categories of cell divisions: symmetric divisions (produces two cells of the same kind) and asymmetric divisions (produces two cells of different types). Furthermore, according to [9], the different types of cells do not divide with the same rate: stem cells divide faster, twice a day, whereas progenitors divide once every two days. Let us note that in this model, there is no fundamental biological difference between stem and progenitor cells except that stem cells are at the top of the hierarchy (more details are given in the discussion). This model uses four parameters to control the probabilities of the different events (a_s, r_s, a_p, r_p) and two more to control the time scales at which the different cells operate. Because of the intricacy of the model, the dynamical behavior of

this system is hardly tractable, and it is therefore difficult to isolate the role of a single parameter from the dynamics of the growth. In order to identify the essential parameters that govern the growth dynamics, we have reduced their number to a bare minimum without losing the richness of the behavior observed experimentally.

We first consider a model that has only one type of cell (left of figure 1(b)). This model is a discrete time model: at each time step, there are three possible fates for every cell of the population: division, no-division or death. This 3-branch model has only two control parameters: a and r. The parameter a governs the probabilities for a given cell to divide or to die. If $a = 1$ (respectively $a = 0$), the probability of death (respectively to divide) is equal to 0. The second parameter r acts as a delay parameter: it controls how likely the cells are to remain unchanged instead of dividing or dying. If $r = 0$, the cells will never divide or die and the total number of cells will never change. On the other hand, if $r = 1$, then the middle branch of this 3-branch model disappears and we are left with a simple one-parameter 2-branch model (right panel, figure 1(b)). Our goal is to study the similarities and the differences between these 3-branch and 2-branch models. We will show that the simple 2-branch model captures all of the interesting dynamics present in the 3-branch model demonstrating only one parameter, a, controls the qualitative behavior of the growth dynamics.

The first quantity we compute in the 3-branch model (figure 1(b)) is the average size of a population of growing cells. If we start from a population of N_0 cells and call the average number of cells N_t after t discrete time steps, then we have:

$$N_t = N_0 (2ra + 1 - r)^t$$

From this formula we can see that, if $a < 1/2$, the population will die off on average exponentially fast with time, whereas if $a > 1/2$ the population will grow exponentially fast. When $a = 1/2$, then $2ra + 1 - r = 1$ and so on average, the number of cells remains constant as a function of time (see appendix B for more details). As we could have expected, the tipping point $a = 1/2$ is key to understanding the behavior of the system: it separates two phases of exponential behavior and most importantly it is the only point where the population growth is not exponential as observed in experiments of [9] on benign tumors. We will study this specific regime, $a = 1/2$, in more details and discuss its significance from a biological point of view in the section 2.2. For a mathematical background on braching processes, see [11, 12], or [13].

1.2. Introducing a maximum number of cells

Before studying the properties of these models in details, we introduce a maximum number of cells that we will call N_{max}. We start off each simulation with an initial number of cells N_0 and we let the population evolve until one of these two events happen: all the cells die or the total number of cell reaches N_{max}. While N_{max} may seem at first like an arbitrary parameter, we will see that the general behavior of the system does not depend on its precise value as long as we choose it much bigger than N_0. There is also a biological justification to the existence of a maximum number of cells. When the number of cells is small, stochasticity of cellular growth has a big impact on the dynamic of the system and we will see that the total number of cells can then exhibit very large temporal fluctuations. On the other hand, once the tumor reaches a certain size, it will become more self-sustainable and will no longer die off just as a byproduct of stochastic effects. For this reason, we will make the simplifying assumption that once a tumor reaches our threshold N_{max} it will necessarily continue to grow and reach macroscopic size even if our model stops being valid for a large number of cells. Furthermore, as size increases, additional effects have to be taken into account, such as the geometry of the tumor (see [14]), pressure from surronding tissues (see [15]), the total amount of nutrients available, etc. Therefore, a simple branching process is relevant only to explain the early growth of tumors with a small number of cells (at most 10000 cells in [9]).

2. Numerical study of a one cell type population

2.1. Probability of survival and average times for growth and extinction

Using the dynamical rules controlled by the two parameters a and r of the branching process described earlier, we let the system evolve from an initial number of cells N_0. We call 'a run' of the simulation one time series of the population growth that spans from N_0 up to N_{max} or to extinction. A large number of different realizations of these runs are used to compute three quantities: the probability for the population to cross the threshold, the time it takes to cross it if this event happens before extinction of the tumor, and finally the average extinction time (the average time it takes for all the cells to die if this happens). To compute the probability to cross the threshold, we use 10000 runs and we normalize the number of times the threshold was crossed by the total number of runs. We use a similar approach to compute the time to extinction or the duration before crossing the threshold, the only difference being that the two average times are conditional quantities: the specific runs for which the threshold was actually crossed are used to compute the time to reach the threshold and the other runs are used to compute the extinction time.

Figure 1(c) shows the probability to cross the threshold for different values of r (0.25, 0.5, and 1). Let us recall that $r = 1$ corresponds to no delay (2-branch

model) while $r = 0.5$ and $r = 0.25$ correspond to the delayed cases (3-branch model). For the three values of r, the three corresponding curves collapse on top of the curve directly calculated from the theory of branching processes (see appendix B). Additionally, figure 1(c) shows two distinct behaviors: for $a < 1/2$, the probability to cross $N_{\max}$ is equal to 0, whereas if a becomes larger than $1/2$, the probability is non-zero. These two distinct dynamics reflect two different biological behaviors: if a is smaller than $1/2$, the tumor will never be able to grow to a macroscopic size. By contrast, if a is larger than $1/2$, then the tumor has a good chance to survive, and under this condition it will grow exponentially fast. Importantly, $a = 1/2$ is a very peculiar point that separates the two latter regimes for which the system behaves in completely different manners. We will give more details on what happens when the system is exactly on this near-critical point $a = 1/2$ in the next section.

Figure 1(d) shows the average extinction time as well as the average time to cross the threshold (plotted in inset in semi-log scale). Both plots show that the smaller r, the longer it takes for the population of cells to die off or to cross the threshold. This result supports the idea that r acts as a delay parameter: the smaller r, the longer the delay and therefore the longer the two times get. But the shapes of both plots stay relatively unchanged for the different values of r. Let us note that a similar result holds if we plot the variance of the time for extinction as a function of a for different values of r: a is the important parameter, r only shifts the curves. We can intuitively understand the shape of the extinction time: for a smaller than $1/2$, the cells have a small probability to divide so the extinction time is very small (the population size decreases exponentially fast in this case). For a larger than $1/2$, the population grows exponentially fast (the chance of division being larger than the chance of death) so if extinction happens, it must happen very early during growth, right at the beginning before the exponential growth kicks off. Therefore, for a larger than $1/2$, the extinction time is also very small. The interesting behavior happens when a is very close to $1/2$: in this case the system is fluctuating between growing and dividing, therefore the associated extinction time tends to be very long. In figure 1(d) the extinction time has a finite value at $a = 1/2$. However, if we relax the constraint of maximum number of cells, the extinction time becomes mathematically infinite at $a = 1/2$ (for more details see appendix B).

One could wonder how figures 1(c) and (d) would change if we had used a different value for $N_{\max}$ or N_0 but we show that the shapes of these curves do not change as long as $N_{\max}$ is large enough compared to N_0 (supplementary figure 1). Therefore, we have shown that out of the four parameters of our system, a, r, N_0 and N_{max}, a is the only parameter that governs the qualitative dynamics of tumor growth.

2.2. Detailed study of the near-critical behavior at $a = 1/2$

In the previous section, we have noted that under the condition $a = 1/2$, the system exhibits a unique behavior, (figures 1(c) and (d)). Interestingly the authors of [9] also chose this point to model their experimental data (division and death are perfectly symmetrical both for stem and progenitor cells). For these reasons, we will now present more numerical results to provide a detailed analysis of the system dynamics at this point. Instead of characterizing the system mean behavior, we now turn to the stochastic growth of individual tumors. In this section, N_0 is set equal to 1 and $N_{\max}$ to 10000. Figures 2(a) and (b) show some typical simulation runs (the plot of the total number of cells in the tumor as a function of the discrete time step) obtained for different values of a. When a is smaller than $1/2$, for each run we expect the tumor to go extinct very quickly (probability to survive is equal to 0, typical run figure 2(a), inset). When a is larger than $1/2$, many tumors will go extinct right at the beginning of the growth (just like the inset of figure 2(a)) and for the others they will grow exponentially (inset of figure 2(b)). Interestingly, for $a = 1/2$, most of the runs will look like the ones for a smaller than $1/2$, but in rare instances, the system will take a very long time either to go extinct (figure 2(a)) or to reach the threshold (figure 2(b)). The probability for these events to occur is very small: the probability for the system to reach $N_{\max}$ is approximately $\frac{1}{N_{\max}}$ and the probability for a run to take a time T to go extinct is approximately $\frac{1}{T}$ (see supplementary figure 2 and appendix B for more details of the theory). But this probability drops to nearly zero as soon as $a \neq 1/2$. Therefore, even though these long runs remain unlikely, they still occur enough to make both the time to extinction and the time to reach $N_{\max}$ diverge at $a = 1/2$ as illustrated in figure 1(d).

To further study the observed fluctuating behavior of the system at $a = 1/2$, we plotted (figure 2(c)) the average number of individual cell deaths and divisions before extinction as a function of a. This number of individual events (divisions and deaths) is not linked in any simple way to the actual time the system takes before extinction (number of discrete time steps). The advantage of introducing this new quantity, however, is that it can be computed analytically if we remove the threshold number of cells $N_{\max}$. The details of the computation are shown in appendix B. We find that this number of events diverges at the point $a = 1/2$ and hereby explains why the time to extinction is also infinite at $a = 1/2$.

Lastly, the plots of figures 2(d) and (e) show the distribution of the extinction times for different values of a (0.4, 0.5, 0.6). Our model uses discrete time steps so we can only consider the probability for the extinction to occur at a given time which we write P(Time). These plots were made with $N_0 = 1$, $N_{\max} = 10000$

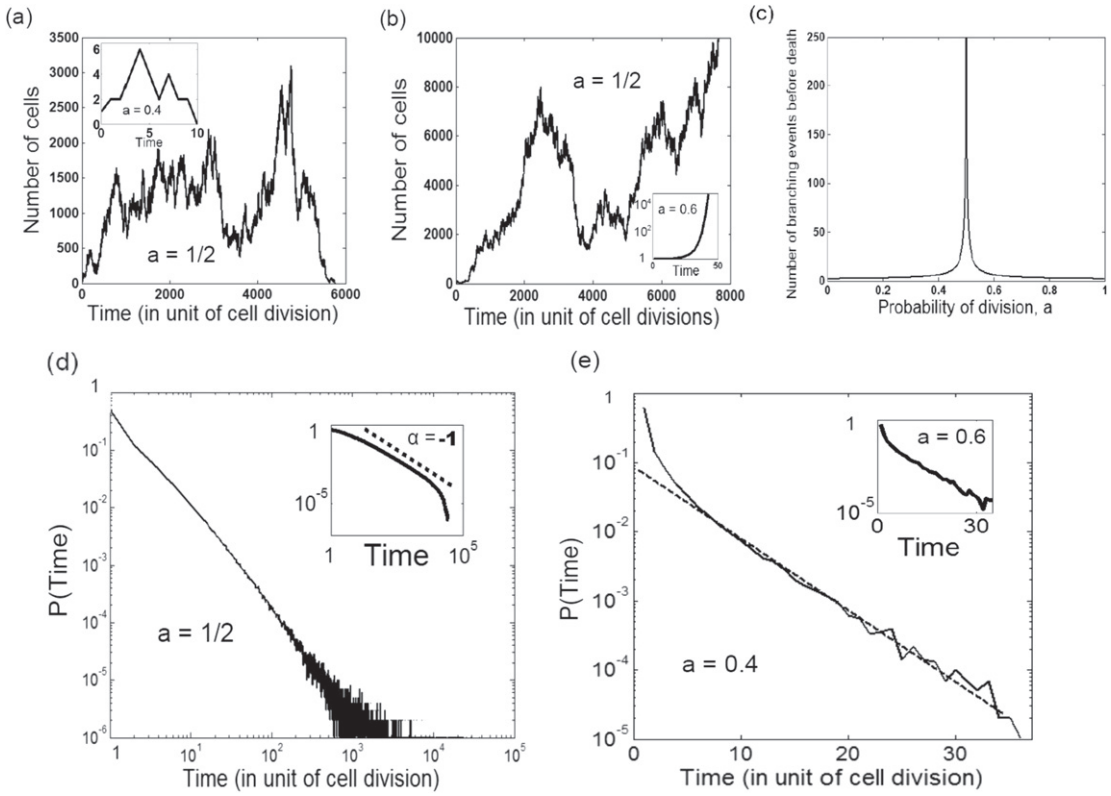

Figure 2. (a) We plot a simulation run where the population takes a very long time to die off at $a = 1/2$. Even if this event is not very likely at $a = 1/2$ (about one chance out of 6000), its probability of happening drops to zero as soon as $a \neq 1/2$. In that case, when the extinction happens, it happens very fast (see inset plot for $a = 0.4$). (b) Similar to (a) but for a run that reaches $N_{\max}$. This takes a very long time and is not very likely when $a = 1/2$ but it goes exponentially fast when $a > 1/2$ (see inset plot for $a = 0.6$). (c) Average number of events (number of individual deaths and divisions taken together). (d) To get a further understanding of the system, we plot the distribution of the time of extinction meaning that on the y-axis we have the probability p(Time) for the population to go extinct at that time. For $a = 1/2$, the time of extinction is power law distributed (straight line in log-log). We have plotted in inset the cumulative distribution which has a slope of $\alpha = -1$ meaning the distribution has a slope of -2 (in log-log). (e) The distribution is exponential for $a = 0.4$ (straight line in semi-log). $a = 0.6$ is shown in inset (we obtain the same plot as for $a = 0.4$). Both plots (d) and (e) were done with 10^6 simulation runs.

and 10^6 simulation runs. We find that for $a = 1/2$, the extinction time is power law distributed (straight line in log-log, see figure 2(d)). We have determined the slope by first plotting in inset a cumulative distribution that has a slope of $\alpha = -1$, which means the underlying distribution has a slope -2:

$$P(\text{Time}) \propto \frac{1}{\text{Time}^2}$$

In computing the average time for extinction (without an $N_{\max}$), we obtain an expression proportional to:

$$\sum_{T=1}^{+\infty} \frac{T}{T^2} = \sum_{T=1}^{+\infty} \frac{1}{T} = \infty$$

This result is consistent with the infinite extinction time that we previously discussed (for details, see the appendix B). In contrast, in figure 2(e), for $a = 0.4$ (main plot) and for $a = 0.6$ (inset), the distribution is exponentially distributed. This result is consistent with a finite extinction time when $a \neq 1/2$ (if we write the same sum in this case, it will converge to a finite value). To summarize, we found a drastically different behavior when $a = 1/2$, for which, there is no typical time scale for the extinction of the tumor while as soon as $a \neq 1/2$, the distribution becomes exponential and therefore it decays at fixed rate.

The observed long individual runs, diverging number of events, and power law distributed extinction time lead us to think that there is a link between the dynamical behavior of our system and second order phase transitions in physics. Taking the example of the ferromagnetic/paramagnetic phase transition, we know that the spontaneous magnetization becomes suddenly non-zero when the temperature goes below the critical temperature. This result is qualitatively similar to the probability for crossing $N_{\max}$ (figure 1(c)) that suddenly becomes non-zero when a crosses $1/2$. Furthermore, in the case of second order phase transitions, we know that there is a correlation length that diverges at the critical point. Similarly, in our case the number of time steps of the individual runs diverges (see plot 1d or 2a, 2b, 2c). This analogy between our simple branching model and phase transitions has actually already been worked out using a percolation model on the Bethe lattice (which is equivalent to a binary tree for a coordination number of $z = 3$), see for instance [16].

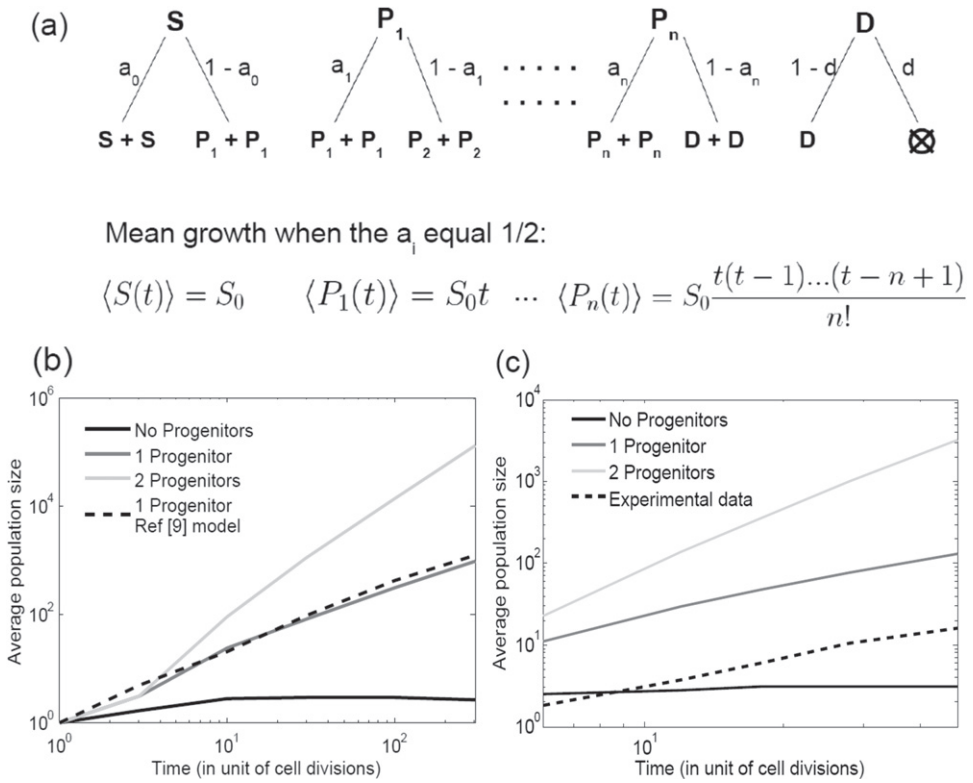

Figure 3. (a) A multiple cell type model that consists of stem cells, n levels of progenitor cells and differentiated cells. In the case of a non-exponential growth where all the a_i equal $1/2$, the mean growth of each type of cell is a polynomial given by the equations in (a). The mean growth of the whole population will therefore be a polynomial of the n-th degree, n being the number of progenitor levels. Comparison of branching models with different numbers of progenitor levels to the model of the paper [9] in (b) and to experimental data of the same paper in (c) (all numerical plots done with 10000 simulation runs and $N_0 = 1$). We see that models with different number of progenitors have different slopes in log-log. (b) shows that the model of the paper [9] that has 1 progenitor level but more parameters falls very close to our simple 1 progenitor model. (c) is the same plot as (b) except that we compare directly the data of [9] to our models. As expected, we see that the experimental data has the same slope as our 1 progenitor model.

3. Multiple cell type population

We now wish to extend our approach to a more elaborate branching model where many different cell types are present just as in [9]. Thanks to our previous results, we know that the middle branch controlled by r does not impact the general behavior of the system. Therefore, we will no longer include it in our model and consider that at every time step, each cell will either die or divide. The model of [9] also contains different time scales for the division of the different types of cells, but we will make a simplifying assumption and consider that all the cells divide according to the same effective time scale that encapsulates all the other time scales. Mathematically, growth will be a polynomial function of time and this additional simplification does not change the degree of the polynomial (see the last section of appendix B for more explanations). The most general model is shown in figure 3(a): it consists of stem cells, n levels of progenitor cells and differentiated cells. The model has in total $n + 2$ parameters: a_0 controls the probability for the stem cells to divide, a_1 through a_n control the division of the progenitors and one more parameter d for the death rate of the differentiated cells. Once again, the fact that we choose to call the first level of dividing cells stem cells and the n other levels progenitor cells is merely a convention and has no biological meaning in the model. However, the fact that stem cells are at the very top of the hierarchy means that they give rise to many more cells than the progenitors.

Let us note that in contrast with existing models such as [17], our model is stochastic, which means that all the cancer cells can give birth to an arbitrary number of descendants. In the model of [17], only the cancer stem cells have a stochastic fate, all other cancer cells in the model give birth to a deterministic number of descendants. We chose this stochastic approach because it is the one used in [9] to reproduce experimental data and also it allows us to infer the number of cell types in the growing tumor (the number of progenitor levels).

First, we compute the average growth of a population that starts off with stem cells and no progenitor cell or differentiated cell. By average growth we mean the average of the sum of all the different types of cells in the population. A full summary of the different regimes of growth is given in appendix B in the case of a one progenitor level model. In order for the

population not to grow exponentially fast, all the a_i have to be smaller or equal to 1/2. If some are larger than 1/2, the average growth will be an exponential with rate $2a$, a being the largest a_i. If an a_i is smaller than 1/2, then the corresponding type of cell will die off exponentially fast and the growth will be dominated by the other types of cells. Ignoring the cases of exponential growth or decay, we will therefore consider the specific condition for which all the a_i equal 1/2. In this case, the average growth of each type of cell is polynomial and is given by the formulas on figure 3(a). The growth of the total population is a polynomial of the n-th degree, n being the number of progenitor levels. This result is potentially powerful as it should, in principle, allow us to predict the number of progenitor levels just by characterizing the average growth of the system. Let us also note that the differentiated cells, because they cannot divide, never change the type of growth (exponential or polynomial) regardless of the value of d (if one models a system where the differentiated cells do divide, then a polynomial growth of degree n corresponds to n-1 progenitors levels plus the dividing differentiated cells). Finally, we have plotted in figure 3 of the supplementary information the three quantities discussed in section 2 (probability of crossing the threshold, time of extinction and time of crossing the threshold) for the model of figure 3(a) for $n = 1$. We obtain behaviors similar as in figures 1(c) and (d).

In the branching model of [9], the authors use a one progenitor level model with $a_s = a_p = 1/2$, $r_s = r_p = 0.2$, see figure 1(a) (notice that $a_s = a_0$ and $a_p = a_1$, we have chosen to rename these parameters just to be able to consider a model with more levels of progenitors). The d that we introduced on figure 3(a) is the death rate of the differentiated cells: it is equal to one cell a week in the model of [9]. We already know from our previous results that we can ignore the value of r_s, r_p and d. Interestingly both a_s and a_p are set equal to 1/2 which means the system must be tuned to the polynomial regime (regardless of the values of all the other parameters). Our goal is now to test on this seemingly complicated branching model our reductive method to predict the number of progenitor levels used by the model. We will compare this model to three other models each with a different number of progenitor levels (0, 1, 2) and with very simple parameters: all a_i equal to 1/2 and $d = 1/2$ (here the value of d is chosen arbitrarily as we know that it does not impact the growth of the system). We numerically compute the average number of cells after 3, 10, 30, 100, 300 time steps for each of the 4 models. Here each run of the simulation starts from one stem cell and stops after 3, 10, 30, 100 or 300 time steps. We use 10000 simulation runs to compute these averages.

The results are shown in figure 3(b). In order to clearly show the different degrees of the polynomial growth, we have chosen a log-log scale. After about 10 time steps, we see that the plot discriminates well between three types of polynomial growth: a constant growth (straight line), a linear growth (slope 1), and a polynomial of degree 2 (slope 2). The first time steps are not very meaningful as the number of cells is too small to accurately observe the polynomial growth. Therefore we find that the model of [9] that has one progenitor level and large number of parameters falls very close to our simple one progenitor level model.

Finally, we now apply our method directly to the experimental data of [9]. The data available from this paper is the clonal size distributions at different times, which is a proxy for the number of clones (a population of cells that all came from one mother cell) of different size at distinct times (see figure 2(e) of [9]). We average those numbers for each time point to get an average tumor size after 6, 12, 18, 28, and 48 time steps. We then run numerical simulations (starting with N_0= one stem cell) to generate the same data for our different models and we plot the results along side with the experimental data. The data set clearly shows the same slope in log-log scale as the simple one progenitor level model. We observe a shift between experimental and theoretical curves and this is because in the experiment some tumors start from one stem cell and some from one progenitor cell while in our numerical simulation all the runs start from one stem cell. In order to get a perfect fit, the authors of [9] explain that one must start from a stem cell in roughly 20% of the cases and from a progenitor the rest of the time. The strength of our model is that even without the precise knowledge of the initial conditions it is still possible to show that there are no more than two types of dividing cells, stem cells and progenitors. This same conclusion is supported by the findings of [9].

4. Conclusion

In summary, we have shown that in branching processes modeling tumor growth the parameter that controls the balance between the probabilities of division and death of the cells is the only relevant parameter to describe the qualitative behavior of the system. We have also shown that in estimating fundamental quantities that includes probability of crossing $N_{\max}$, time to cross $N_{\max}$, time to extinction as a function of a, the plots obtained always have the same shapes regardless of the values of the detailed parameters of the models (number of cell types, delay r, time scales, etc). We have also seen that the system behaves like a phase transition system when the probabilities of cell division and death are perfectly balance ($a = 1/2$). At this critical point, the fluctuations of the number of cells become very large and there is no typical time scale for the time of tumor extinction. In the case of a multiple cell type model, we studied how the properties of the critical point could be used to predict the number of different cell types. We would like to stress the fact that the goal of our

paper was not to study the properties of the most complicated branching models (see for instance [11, 12], or [13]); rather we hoped to find the simplest model possible that was still rich enough to explain the biological data of [9].

Let us now briefly discuss the biological implications of all our results. In all the branching processes we considered, non-exponential growth only happens for a very specific value of the parameters: when division and death are perfectly balanced ($a = 1/2$). However, such a fine-tuned value is an unrealistic hypothesis because of the presence of inherent noise in biology and therefore a parameter can never be exactly equal to fixed value. But if we assume that a follows a probability distribution centered around $a = 1/2$, the parameter would spend as much time below as above the fixed value $1/2$, and then the system will have the same behavior as when $a = 1/2$. The only requirement is that the added noise must be independent of the time step and with no correlations between cells (see the last section of appendix B). A more radical solution would consist in introducing a feedback loop in the system in order to maintain the parameters at the desired fixed value. This approach has been studied in [18] and [19] for simple models. The limitation of this latter approach is that there is no guarantee that the results will hold for any type of feedback loops. Nevertheless, it seems that such branching models can be used at the critical value to predict the (non-exponential) behavior of a population of cells like in [9] or [10].

Finally, our model that allows us to predict the number of different types of dividing cells from experimental data could prove useful in the cancer stem cell debate [20]. The existence of cancer stem cells is still being debated and even the exact definition is unclear. This problem is closely linked to the epithelial-mesenchymal transition that has been extensively studied (see for example [21, 22], and [23]). At first, cancer stem cells were just defined from an experimental point of view as cells that could produce a whole tumor by themselves when introduced in mice. Some computational papers consider cancer stem cells as regular dividing cells that divide faster than the average ([9] or [10]) while others consider them as a slow cycling population that stays roughly constant [4]. The strength of our result is that we can estimate how many levels of dividing cells are present in a growing tumor without a precise knowledge of the actual parameters used (which are subject to much debate). From the data of [9], we confirm the hypothesis that not all the dividing cells of the tumor are equivalent: there are two types of dividing cells, stem and progenitor cells and the stem cells are those that really drive the growth of tumors.

Lastly, we point out that our approach is not solely restricted to cancer cells: our results hold for any other biological systems whose growth can be described by branching models, such as those found in stem cell research or development (see [12, 24] and [25]).

Acknowledgments

We thank Mark Kim, Kevin Wood and Lisa Marshall for their useful advice and support. We are also grateful to Cedric Blanpain and Gregory Driessens for providing us with their experimental data. Guillaume Remy was supported by the Physics of Living Systems Program (NSF Award 1219334).

References

[1] Ribba B, Saut O, Colin T, Bresch D, Grenier E and Boissel J P 2006 A multiscale mathematical model of avascular tumor growth to investigate the therapeutic benefit of anti-invasive agents *Journal of Theoretical Biology* **243** 532–41
[2] Sachs R K, Hlatky L R and Hahnfeldt P 2001 Simple ode models of tumor growth and anti- angiogenic or radiation treatment *Mathematical and Computer Modeling* **33** 1297–305
[3] Ciampi A, Kates L, Buick R, Kriukov Y and Till J E 1986 Multitype galton-watson process as a model for proliferating human tumor cell populations derived from stem cells estimation of stem cell self-renewal probabilities in human ovarian carcinomas *Cell and Tissue Kinetics* **19** 129–40
[4] Dingli D, Traulsen A and Michor F 2007 (A)Symmetric stem cell replication and cancer *PLOS Computational Biology* **3** 482–7
[5] Gupta P B, Fillmore C M, Jiang G, Shapira S D, Tao K, Kuperwasser C and Lander E S 2011 Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells *Cell* **146** 633–44
[6] Durrett R and Moseley S 2010 Evolution of resistance and progression to disease during clonal expansion of cancer *Theoretical Population Biology* **77** 42–8
[7] Kim Y and Stolarska M A 2007 A hybrid model for tumor spheroid growth *in vitro*: theoretical development and early results *Mathematical Models and Methods in Applied Sciences* **17** 1773–98
[8] Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, Karchin R, Kinzler K W, Vogelstein B and Nowak M A 2010 Accumulation of driver and passenger mutations during tumor progression *PNAS* **107** 18545–50
[9] Driessens G, Beck B, Caauwe A, Simons B D and Blanpain C 2012 Defining the mode of tumour growth by clonal analysis *Nature* **488** 527–31
[10] Doupé D P, Alcolea M P, Roshan A, Zhang G, Klein A M, Simons B D and Jones P H 2012 A single progenitor population switches behavior to maintain and repair esophageal epithelium *Science* **337** 1091–3
[11] Harris T E 1964 *The Theory of Branching Processes* (Berlin: Springer-Verlag)
[12] Kimmel M and Axelrod D E 2002 *Branching Processes in Biology* (New York: Springer)
[13] Haccou P, Jagers P and Vatutin V A 2005 *Branching Processes: Variation, Growth, and Extinction of Populations* (Cambridge: Cambridge University Press)
[14] Norton L 2007-2008 Cancer stem cells, self-seeding, and decremented exponential growth: Theoretical and clinical implications *Breast Disease, IOS Press* **29** 27–36
[15] Basan M, Risler T, Joanny J-F, Sastre-Garau X and Prost J 2009 Homeostatic competition drives tumor growth and metastasis nucleation *HFSP* **3** 265–72
[16] Stauffer D and Aharony A 1994 *Introduction to Percolation Theory* (Oxford: Taylor & Francis)
[17] La Porta C A M, Zapperi S and Sethna J P 2012 Senescent cells in growing tumors: Population dynamics and cancer stem cells *PLOS Computational Biology* **8**

[18] Johnston M D, Edwards C M, Bodmer W F, Maini P K and Chapman S J 2007 Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer *PNAS* **104** 4008–13
[19] Johnston M D, Maini P K, Chapman S J, Edwards C M and Bodmer W F 2010 On the proportion of cancer stem cells in a tumour *Journal of Theoretical Biology* **266** 708–11
[20] Hanahan D and Weinberg R A 2011 Hallmarks of cancer: the next generation *Cell* **144** 646–74
[21] Kalluri R and Weinberg R A 2009 The basics of epithelial-mesenchymal transition *The Journal of Clinical Investigation* **119** 1420–8
[22] Thiery J P, Acloque H, Huang R Y J and Nieto M A 2009 Epithelial-mesenchymal transitions in development and disease *Cell* **139** 871–90
[23] Mani S A *et al* 2008 The epithelial-mesenchymal transition generates cells with properties of stem cells *Cell* **133** 704–15
[24] Werner B, Dingli D, Lenaerts T, Pacheco J M and Traulsen A 2011 Dynamics of mutant cells in hierarchical organized tissues *PLOS Computational Biology* **7**
[25] Klein A M, Doupé D P, Jones P H and Simons B D 2007 Kinetics of cell division in epidermal maintenance *PRE* **76**