BIOREPS 2016 Project B
Models of Tumor Progression:
Two-Type Branching Processes

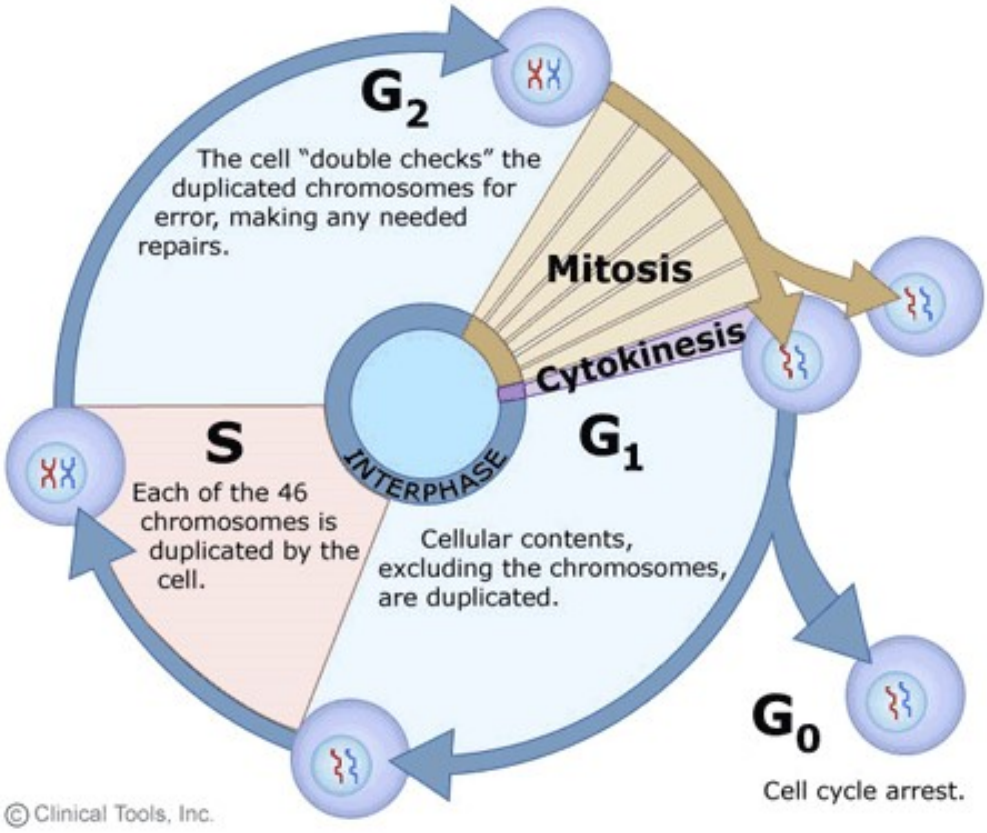

Figure 1: Schematic of the cell cycle. During Mitosis, somatic cells undergo division of cellular contents. The cellular membrane is then completely cleaved during cytokinesis resulting in the production of two 'identical' daughter cells. Most cells enter the G_0 phase to carry out biological functions while others continue through the remaining three phases in preparation for another mitotic division. Image Credit:http://www2.le.ac.uk/departments/genetics

1 Background

Cell division plays an important role in maintaining an equilibrium of cells within the body. The division process is controlled by the cell cycle (Figure 1) that has checkpoints in place that mediate progression between the different phases. Most cells in the body are found in the G_0 phase and are not actively dividing but are simply carrying out their biological functions. When needed, the cells can leave the G_0 phase and can continue through the cycle in preparation for another round of division. Once ready, the cell undergoes mitosis and the original parent cell divides into

two daughter cells. The daughter cells then have the option of returning to the G_0 phase where the parent cell originally was, or to continue on in the cycle and repeat the division process. This continuing cycle is under strict regulation to ensure proper cell production is achieved.

The two daughter cells produced during each mitotic division are supposed to be identical to one another so that homogeny is conserved. However, variation can be introduced in a variety of ways. An array of mutations can be introduced during the S phase of DNA replication, they can be as minor as a nonsense or as serious as a missense mutation. Other sources of variation can occur during mitotic recombination or in the event of nondisjunction. Variation can be beneficial to organisms as it is at the heart of evolution. However, the benefits of variation also comes with a price. A mutation introduced to any part of the cell cycle can result in the formation of cancerous cells. The mutation does not have to be directly a result of reproduction error, but can also take form through exposure to carcinogens or other disease causing antigens. The formation of cancer cells are usually not due to the result of a single mutation but rather encompasses an array of mutations that later effect the kinetics of the cell cycle.

Cancer cells are distinguished from normal cells in that they continually and uncontrollably divide. The amount of time they spend in the G_0 phase is characteristically low when compared to that of a normal cell. In addition, the checkpoints and regulations that normal cells follow when proceeding through the cell cycle are dysfunctional in cancer cells. In order to proceed in the cell cycle, cells must meet requirments are three different checkpoints throughout the cycle. The G_1 checkpoint ensures the accuracy in the newly synthesized DNA strand, later on, the G_2 checkpoint ensures proper chromosome duplication has occured, and finally, the M checkpoint ensures the proper attachment of kintechore fibers before the cell undergoes mitosis. Unlike healthy cells, Cancerous cells can compromise the chekpoints and can pass through with reduced regulation. This results in the continuous and uncontrollable division process previously described. The continual division results in masses of cells called tumors. These clumps of cells fall into one of two classifications. Benign tumors are composed of cells that are non-invasive and continue to grow within a centralized location. While these tumors can be troublesome, they are often less lethal than their counterpart. Malignant tumors on the other hand, are invasive and have the ability to spread throughout the body. They can enter blood vessels or even lymphatic nodes through a process called intravasation. Once they have penetrated these systems, the malignant tumor cells can be transported and then introduced to new parts of the body where they can be deposited through a process called extravasation. The malignant tumor cells can then continue their rapid cellular division and grow into tumors by utilizing the entire body as a resource of energy to fuel their wild cellular division (Figure 2). The focus of this problem set will be to create a physical model that can be used to describe the range of tumor progression as well as the kinetics of cell death while focusing on simple cases to develop this framework.

References

[1] Antal, T. and Krapivsky, P. L. Exact solution of a two-type branching process: models of tumor progression. *J. Stat. Mech.* **2011**, 8018–8040 (2011).

2 Questions

2.1 Simple model for cellular division and mutation

We can create a general model for these processes of cellular proliferation, mutation, and death. Using A to represent progenitor cells, or cells that might mutate irreversibly into another cell type, and B to represent that mutated type, we can summarize this birth/death process using the rates

$$\begin{aligned} A &\longrightarrow AA \qquad \text{rate } \alpha_1 \\ A &\longrightarrow \emptyset \qquad \text{rate } \beta_1 \\ A &\longrightarrow B \qquad \text{rate } \nu \\ B &\longrightarrow BB \qquad \text{rate } \alpha_2 \\ B &\longrightarrow \emptyset \qquad \text{rate } \beta_2. \end{aligned} \tag{1}$$

Given a single initial A cell, we can model the probability distribution $P_{m,n}(t)$ representing the likelihood of finding m copies of A and n copies of B at some time t. We can set the division rate of A, α_1, to be 1, simply rescaling time increments to achieve this simplification.

a) Using this shorthand notation for the rate differences

$$\begin{aligned} \lambda_1 &= 1 - \beta_1 - \nu \\ \lambda_2 &= \alpha_2 - \beta_2 \end{aligned} \tag{2}$$

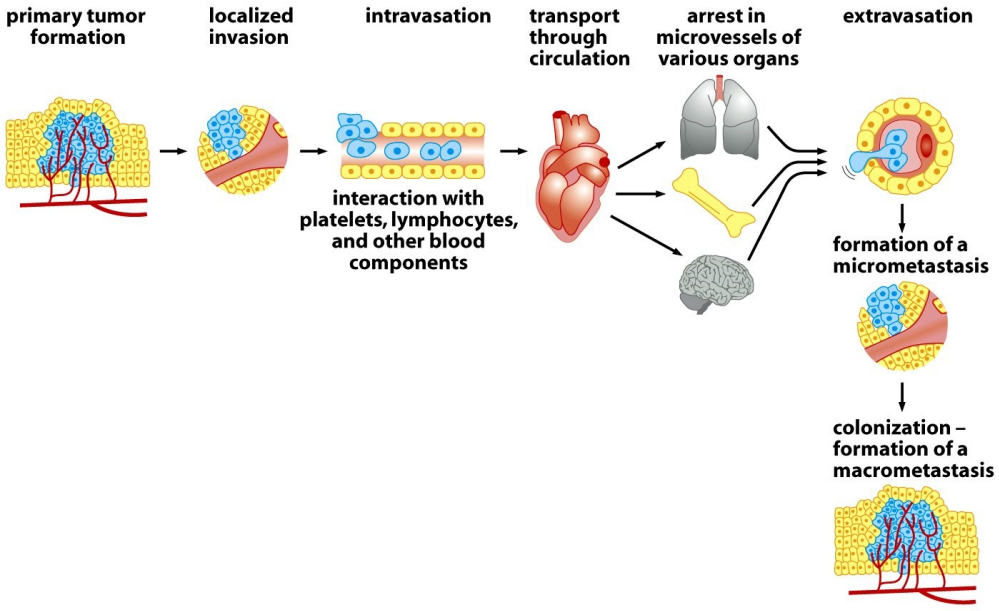

Figure 2: Schematic of Malignant Tumor Cells utilizing the circulatory system as a means of transportation throughout an organism. Intravasation is the process of the malignant cancer cells entering the circulatory system, they then circulate and then return to the body through a process called extravasation. This results in the formation of tumors away from the primary tumor site. Image credit: Figure 14.4 The Biology of Cancer (©Garland Science 2007) .

which can be thought of as the fitness values for each cell type. A can be thought of as representing a benign tumor cell; upon acquiring another mutation, it transforms into a malignant B cell. Given that the average number of A cells, m, and the average number of B cells, n, follow the rate equations

$$\begin{aligned} \frac{d\langle m\rangle}{dt} &= \lambda_1\langle m\rangle \\ \frac{d\langle n\rangle}{dt} &= \nu\langle m\rangle + \lambda_2\langle n\rangle \end{aligned} \tag{3}$$

Starting with one A cell ($\langle$m$\rangle$(0) = 1) and given $\langle n\rangle(0) = 0$, solve (3) to find equations for $\langle m\rangle$ and $\langle n\rangle$ defined in terms of ν, λ_1, and λ_2.

b) The dynamics of the two-type branching process is found using probabilities satisfying the forward and backward Kolmogrov equations. After taking into account all possible transitions for an initial A or B cell and taking the limit of tau going to zero, we have the following rate equations.

$$\begin{aligned} \frac{dP^A_{m,n}}{dt} &= P^{AA}_{m,n} + \beta_1\delta_{m,0}\delta_{n,0} + \nu P^B_{m,n} - (1+\beta_1+\nu)P^A_{m,n} \\ \frac{dP^B_{m,n}}{dt} &= \alpha_2 P^{BB}_{m,n} + \beta_2\delta_{m,0}\delta_{n,0} - (\alpha_2+\beta_2)P^B_{m,n} \end{aligned} \tag{4}$$

Using the generating function

$$\mathcal{P}(x,y,t) = \sum_{m,n\geq 0} x^m y^n P_{m,n}(t) \tag{5}$$

show that equation (refe5) can turn into the pair of coupled non-linear ordinary differential equations below.

$$\begin{aligned} \delta_t\mathcal{A} &= \mathcal{A}^2 + \beta_1 + \nu\mathcal{B} - (1+\beta_1+\nu)\mathcal{A} \\ \delta_t\mathcal{B} &= \alpha_2\mathcal{B}^2 + \beta_2 - (\alpha_2+\beta_2)\mathcal{B} \end{aligned} \tag{6}$$

with initial conditions

$$\begin{aligned} \mathcal{A}(x,y,t=0) &= x \\ \mathcal{B}(x,y,t=0) &= y \end{aligned}$$

c) From the master equation we can obtain the following differential equations:

$$\frac{\partial\mathcal{A}}{\partial t} = \mathcal{A}^2 + \beta_1 + \nu\mathcal{B} - (1+\beta_1+\nu)\mathcal{A} \tag{7}$$

$$\frac{\partial\mathcal{B}}{\partial t} = \alpha_2\mathcal{B}^2 + \beta_2 - (\alpha_2+\beta_2)\mathcal{B} \tag{8}$$

with initial conditions

$$\mathcal{A}(x, y, t = 0) = x$$
$$\mathcal{B}(x, y, t = 0) = y$$

The equation for $\mathcal{B}$ is easy to solve. Its solution is given by:

$$\mathcal{B} = 1 - \frac{\lambda_2}{\alpha_2(1-z)}, z = [1 - \frac{\lambda_2}{\alpha_2(1-y)}] \exp^{-\lambda_2 t} \tag{9}$$

However $\mathcal{A}$ is more complicated. Plug the solution for $\mathcal{B}$ into the differential equation for $\mathcal{A}$ and define $X = 1 - \mathcal{A}$. This new differential equation is called a Riccati Equation. Additionally it is useful to know that $\lambda_1 = 1 - \beta_1 - \nu$.

Let us now define $X \equiv \frac{d}{dt} \log Z = \frac{1}{Z}\frac{dZ}{dt}$. Use this definition for X in our Riccati Equation to get a new differential equation in terms of Z. Additionally, you should multiply by Z and disregard higher order terms. This differential equation is called the Sturm-Liouville Equation.

The end result of all this mathematical manipulation is the following:

$$\frac{d^2 Z}{dt^2} = \lambda_1 \frac{dZ}{dt} + \frac{\nu\lambda_2}{\alpha_2(1-z)} Z \tag{10}$$

Let us assume a solution of the form:

$$Z(t) \equiv z^{\omega/\lambda_2} \Phi(z)$$

Where ω will be defined later. Using this ansatz in our differential equation we derive a solution in terms of the hypergeometric functions. It has two linearly independent solutions. Thus,

$$\Phi(z) = F(a, b; c; z) + C z^{1-c} F(-b, -a; 2 - c; z)$$

Where the functions F are the hypergeometric functions. And:

$$a = \frac{\omega}{\lambda_2}, b = \frac{\omega + \lambda_1}{\lambda_2}, c = 1 + \frac{2\omega + \lambda_1}{\lambda_2}, \omega = -\frac{\lambda_1}{2} + \sqrt{\frac{(\lambda_1}{2})^2 + \frac{\nu\lambda_2}{\alpha_2}}$$

Using our previous definitions, write the function $\mathcal{A}$ in terms of Φ. If we wish to write $\mathcal{A}$ in terms of the hypergeometric functions we must first take the derivative of a hypergeometric function. The following defines the derivative of the hypergeometric function:

$$\frac{d}{dz} F(a, b; c; z) = \frac{ab}{c} F(1 + a, 1 + b; 1 + c; z)$$

Using this result write $\mathcal{A}$ in terms of the hypergeometric functions and the unknown constant C. Thus, we have derived a solution for $\mathcal{A}$ which gives us the generating function for an initial A cell. Therefore, we have solved our differential equations for $\mathcal{A}$ and $\mathcal{B}$.

d) Calculation of probability densities from generating functions

Now that we have derived generating functions for both $A(x, y, t)$, the general case of A and B cells given a single initial A cell, and $B(y, t)$, the case of just B cells given a single initial B cell, we will show how to use these generating functions to calculate probability densities. To develop our understanding, we will focus on the bi-critical $B(y, t)$ generating function

$$B(y,t) = \frac{(y + \alpha_2 t(1-y))}{(1 + \alpha_2 t(1-y))} \tag{11}$$

and show how to both analytically and numerically compute probabilities, $P_n(t)$. The concepts here apply generally to the conversion between generating functions and probabilities regardless of how complicated the generating function is or the number of dimensions.

First, let's return to a general definition of our generating function:

$$B(y,t) \equiv \sum_{n\geq 0}^{\infty} y^n P_n(t). \tag{12}$$

This probably looks quite intimidating to the uninitiated, but if one is familiar with discrete Fourier transforms one can quite easily convert by substituting $y = e^{-i\theta}$,

$$B(e^{-i\theta}, t) = \sum_{n\geq 0}^{\infty} e^{-in\theta} P_n(t). \tag{13}$$

We note that $B(e^{-in\theta}, t)$ is periodic with θ and that we can fully describe the function over the interval $\theta = [0, 2\pi]$. Thus, when discretizing θ for our numerical implementation we choose $\theta_k = 2\pi k/N$ where N is the number of samples and $k = 0, 1, 2, , N-1$. This discretized $B(e^{-i\theta_k}, t)$ corresponds to a truncated $P_n(t)$ with N samples which is implicitly assumed to repeat infinitely outside of the $k = [0, N-1]$ window (this is the typical discrete-time Fourier transform assumption). This gives us

$$B(e^{-i\theta_k}, t) = \sum_{n=0}^{N-1} e^{-ink2\pi/N} P_n(t). \tag{14}$$

Thus, $B(e^{-i\theta_k}, t)$ and $P_n(t)$ are, in the large N limit, Fourier transform pairs. This also means that there exists an inverse discrete Fourier transform of B which can transform back into $P_n(t)$,

$$P_n(t) = \frac{1}{N} \sum_{n=0}^{N-1} e^{ink2\pi/N} B(e^{-i\theta_k}, t). \tag{15}$$

This relationship holds for conversion of any generating function to the corresponding probability density. In this way, we can use the very powerful general function, $A(x, y, t)$, the probability of m cells of type A and n cells of type B, to determine $P_{m,n}(t)$ by two inverse discrete Fourier transforms over x and y. Theoretically, we could compute the probability density for any number of cells of any number of cell types, provided the generating function is well defined. However, here we will focus on the simpler generating function $B(y, t)$ for the bi-critical case and demonstrate that the Fourier transform approach yields the expected results for short time

$(t = 0)$ and long time $(t \to \infty)$.

1. Show that $P_0(t = 0) = 0$, $P_1(t = 0) = 1$, $P_0(t \to \infty) = 1$, and $P_1(t \to \infty) = 0$ using the inverse discrete Fourier transform. First, find the limit of $B(e^{-i\theta_k}, t)$ at long and short times and then compute the probabilities analytically with $N = 2$. Keep in mind the Euler relation $e^{-i\pi} = e^{i\pi} = -1$.

2. Compute $P_n(t)$ at $t = 0$,$t = 1$,$t = 10$, and $t = 100$ using the numerical implementation of the inverse Fourier transform with $N = 16$, $\alpha_2 = 0.1$. Essentially, compute a vector $B(e^{-i\theta_k}, t)$ and perform the inverse Fourier transform (ifft in Matlab). Plot your results and describe what happens to the distribution $P_n(t)$ at the short time, long time, and intermediate times.

e) Calculation of critical B cell behavior from Kolmogorov equation
The above bi-critical equation (11) demonstrating the behavior of critical $\mathcal{B}$ cells can be calculated from equation (9) by determining when $\lambda_2 = 0$. Verify the solution of equation (11) by calculating the limit $\lambda_2 \to 0$ of equation (9). Hint: you will get an indeterminate form, so recall L'Hopital's Rule and differentiate before taking the limit.