UNIVERSITY OF CALIFORNIA,
IRVINE

Stochastic Modeling of Stem Cells

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematics

by

Jienian Yang

Dissertation Committee:
Professor Natalia L. Komarova, Chair
Associate Professor German A. Enciso Ruiz
Professor Michael C. Cranston

2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The text of this dissertation is a reprint of the material as it appears in Mathematical Biosciences and PLOS Computational Biology. The co-author, Natalia L. Komarova, listed in both publications directed and supervised research which forms the basis for the dissertation.

I would like to express my deepest appreciation to my committee chair, Professor Natalia L. Komarova, for her continuous advisory support of my Ph.D study and related research, for her patience, motivation and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. Without her guidance and persistent help, this dissertation would not have been possible.

I would like to thank the rest of my thesis committee members, Professor German A. Enciso Ruiz and Professor Michael C. Cranston, for their insightful comments, advising, and encouragement.

I thank Elsevier for permission to include Chapter 2 of my dissertation, which was originally published in Mathematical Biosciences. I also thank PLOS Computational Biology for permission to include Chapter 3 of the dissertation. Financial support was provided by NIH grant 1 U01 CA187956-01, grant no. 1R01CA12928601A1 and math department fellowships of University of California, Irvine.

Last but not the least, I would like to thank my parents for supporting me financially and spiritually throughout writing this thesis and my graduate life in general. I would also like to appreciate my wife for her appearance in my graduate life and sharing her wonderful life moments with me. Without their support and companion, I won't be at this stage.

CURRICULUM VITAE

Jienian Yang

EDUCATION

Doctor of Philosophy in Mathematics — **Mar. 2017**
University of California, Irvine — *Irvine, California*

Master of Science in Statistics — **June 2016**
University of California, Irvine — *Irvine, California*

Master of Science in Mathematics — **June 2013**
University of California, Irvine — *Irvine, California*

Bachelor of Arts in Applied Math — **Dec. 2010**
University of California, Berkeley — *Berkeley, California*

RESEARCH EXPERIENCE

Graduate Research Assistant — **Mar. 2012 – Mar. 2017**
University of California, Irvine — *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant — **Sept. 2011 – Mar. 2017**
University of California, Irvine — *Irvine, California*

INTERN EXPERIENCE

Biostatistician — **June 2016 – Sept. 2016**
Gilead Sciences — *Foster City, California*

REFEREED JOURNAL PUBLICATIONS

Analysis of Stochastic Stem Cell Models with Control **2015**
Mathematical Biosciences

The Role of Symmetric Stem Cell Divisions in Tissue Homeostasis **2015**
PLOS Computational Biology

ABSTRACT OF THE DISSERTATION

Stochastic Modeling of Stem Cells

By

Jienian Yang

Doctor of Philosophy in Mathematics

University of California, Irvine, 2017

Professor Natalia L. Komarova, Chair

Stem cells (SCs) are the body's raw materials - cells from which all other cells with specialized functions are generated. Understanding the dynamics of SC lineages is of central importance both for healthy and cancerous tissues. We study stochastic population dynamics of the two-compartment (stem and differentiated cells) system and the three-compartment (SCs, intermediate cell type and differentiated cells) model. Cell decisions such as proliferation vs. differentiation decisions are under regulation from surrounding cells. Successful maintenance of cellular lineages depends on the fate decision dynamics of SCs upon division. There are three possible strategies with respect to SC fate decision symmetry: (a) asymmetric mode, (b) symmetric mode and (c) mixed mode. Theoretically, either of these strategies can achieve lineage homeostasis. We start the whole project by only considering symmetric mode for the two-compartment system. We derive simple explicit expressions for the means and the variances of SC and differentiated cell number. The methodology is formulated without any specific assumptions on the functional form of the controls, and thus can be used for any biological system. We then extend the study of the two-compartment system to include asymmetric mode. In particular, we focus on minimal control mechanisms and networks of the two-compartment system. Through stochastic analysis and simulations we show that asymmetric divisions can either stabilize or destabilize the lineage system, depending on the underlying control network. Next, we propose an algorithm to identify a set

of candidate control networks that are compatible with (1) measured means and variances of cell populations, (2) qualitative information on cell population dynamics, and (3) statistical information on intra-crypt cell type correlations. We apply the algorithm on the data of human colon crypts, where lineages are an example of the three-compartment model. We start with 32 minimal control networks compatible with tissue stability, and zero in on only three networks that are most compatible with the measurements.

Chapter 1

Introduction

The stem cell lineage is a basic unit of hierarchical tissues and as such it has attracted the attention of many researchers. The question of stem cell control is at the center of our understanding of tissue functioning. Everyday, cells in hierarchical tissues perform their specific functions and die to be replaced by new cell divisions. This process is stochastic in nature and involves very large numbers of cellular events. The cells involved in the functioning and renewal of an organ differ from each other by their division and apoptosis capabilities, as well as the types of signals they send and the kinds of cell fate decisions they make. They organize into distinct phylogenetic lineages. At the end of each lineage are the non-dividing, terminally differentiated cells. Usually these cells, such as neurons, adipocytes or muscle fibers, are highly specialized and endow tissues with their respective functions. The origin of all differentiated cells can be traced back to their progenitors, the so-called stem cells (SCs) [2, 3].

Successful maintenance and repair of cellular lineages critically depends on the fate decision dynamics of SCs upon division. Long-term steady-state maintenance of lineages requires that only 50% of all SCs progenies remain as SCs, and even slight shift in fate outcomes over time

can lead to lineage exhaustion or uncontrolled expansion. For example, in the hair follicle, melanocyte SCs are more susceptible to exhaustion compared to epithelial SCs; therefore, commonly hair graying occurs faster than hair loss [4–6]. On the other hand, uncontrolled lineage expansion occurs upon myelodysplastic syndrome, a type of blood maligancy when mutated hematopoietic SCs increase their self-renewal rate to more than 50%. Over time, mutated SCs outcompete normal SCs, and accumulation of defective progeny cells leads to the loss of blood function and results in acute myeloid leukemia, a life-threatening complication of the myelodysplastic syndrome [7–9].

From these examples it is evident that tight control of SC fate decision dynamics is of paramount importance. In principle, steady-state maintenance of SCs can be achieved with three strategies [10], see figure 1.1:

a) Asymmetric mode, when each and every SC division produces one SC and one non-SC progeny [11–16];

b) Symmetric mode, when 50% of all divisions produce two SCs and another 50% – two non-SC progeny. In this case SC expansion is precisely balanced by SC exhaustion, and the long-term net balance of SCs and their lineages remains unchanged [10, 17, 18];

c) Mixed mode, when both the asymmetric and two types of symmetric SC divisions co-exist and are partitioned so that long-term net balance of the lineage output stays constant. As in (b), stability critically depends on the ratio of symmetric divisions: SC generating events should be precisely counterbalanced by SC exhaustion.

Assuming that individual cell division decisions are stochastic, at the tissue level, modes (b) and (c) result in neutral clone competition phenomenon, when some SC clones expand, some contract, while others stay constant [10, 19–22]. Theoretically, either one of the strategies (a-c) can achieve lineage homeostasis. However, it remains unclear which strategies are more

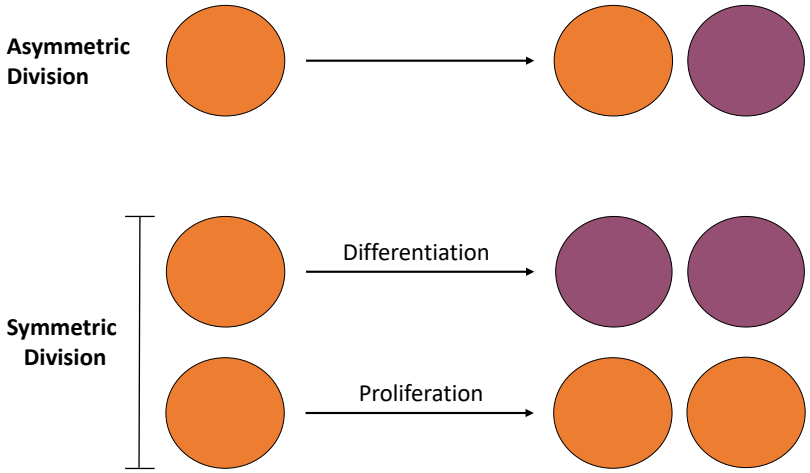


Figure 1.1: Symmetric and asymmetric SC divisions. In the asymmetric division model, a SC produces one differentiated cell and one SC. In the symmetric division model, a SC produces two differentiated cells or two SCs.

advantageous and under what specific circumstances, and what minimal control mechanisms are required to operate them.

1.1 Symmetric Mode

In the simplest lineage, there are SCs capable of self-renewing and regenerating the tissue, and differentiated cells which can perform the tissue's specific functions. We call this the two-compartment (stem and differentiated cells) system.

Differentiated cells are subject to relatively frequent cell death and need to be replenished by SC divisions. These divisions can be of several types. Specifically, a SC can divide symmetrically: a SC can differentiate by dividing into two differentiated cells, or it can proliferate, by dividing into two stem cells. Differentiation/proliferation decisions are thought to be under regulation coming from surrounding cells in the tissue. Various control loops help maintain a roughly constant overall tissue size, and keep variations in the numbers of stem and differentiated cells to a minimum.

We develop a general, analytical methodology for studying the behavior of two-compartment system with nonlinear control. We assume that division, death, and differentiation/proliferation decisions are given by some (unspecified) functions of the numbers of stem and differentiated cells, and provide tools to calculate the moments of the cell numbers, and importantly, the means and the variances of the numbers of cells. Our approach differs from previous theoretical literature on the subject because we do not make any prior assumptions on the type and direction of control loops, apart from the fact that a stable equilibrium exists, which biologically corresponds to the existence of homeostatic control.

The method developed here is algorithmically different, and simpler, than the linear noise approximation [23]. We studied the connection between the two methods and proved that they give the same result to all orders of accuracy. Therefore, our method could be considered a short-cut compared with the Van Kampen power series expansion. We developed a computer program which allows to apply our method to any two-compartment system with given control functions. In other words, if we assign the rates of divisions, differentiation/proliferation, and death to be some functions of the numbers of stem and differentiated cells, our tools allow to calculate analytically the means and the variances of the stem and differentiated cell numbers as functions of the system parameters, and to study stability and robustness of the system. The details will be covered in **Chapter 2**.

1.2 Asymmetric Mode

We then extend the study of the two-compartment system to include asymmetric mode: each and every SC division produces one SC and one differentiated cell. We focus on investigating how symmetric and asymmetric divisions contribute to lineage homeostasis/turnover. We provide analysis that allows to quantify the ability of these two types of divisions to maintain homeostasis. What SC division strategy is better at maintaining the nearly constant

population size? Quoting [10], "Asymmetric divisions are a key mechanism to ensure tissue homeostasis. In normal stem and progenitor cells, asymmetric cell division balances proliferation and self-renewal with cell-cycle exit and differentiation." At the intuitive level, it appears that asymmetric SC divisions should be associated with a more robust homeostatic maintenance. It can be argued that purely asymmetric SC divisions do not change the total number of SCs and therefore ensure the maintenance of a constant cell population, see e.g. [12]. It turns out however that tight homeostatic maintenance of the lineage (including differentiated cells) is not necessarily associated with purely asymmetric divisions. We show that asymmetric divisions can either stabilize or destabilize the lineage system, depending on the underlying control network. The details will be covered in **Chapter 3**.

We address the questions of SC division symmetry by means of mathematical modeling. Our approach is based on that developed in Chapter 2, and it contributes to the large theoretical literature on SC dynamics, see e.g. theoretical work of [24–27], and a review in [28]. Some of the important areas of mathematical modeling in the context of SCs include discrete and continuous models in the context of carcinogenesis [29–41]; modeling of SC in the hematopoietic system [42–46]; deterministic modeling of two-, three-, and multi-compartmental systems under various regulation functions [47–51]; stochastic modeling of SC systems and the analysis of fluctuations [52–57].

1.3 Application

Similar analysis can be extended to three compartments in a set cell lineage: SCs, transit amplifying cells (TACs) and differentiated cells (DCs). In the colon and intestinal crypts, as well as other structures, these are linearly ordered with SC at the bottom, DC cells at the top, and TAC in between. In order to maintain the number of each cell type, the rate of removal of the DCs from the top is balanced by division and differentiation of the SCs and

TACs below.

Theoretically, each cell population may influence (in a negative or positive way) each of the processes that happen in the system, which gives rise to a very large number of networks of cellular control. [1] examined such controls from the (linear) stability point of view and called the stable networks with the smallest possible number of loops the "minimal networks". In a three-compartment system, assume that the following five processes can be controlled: divisions of SCs and TACs, differentiations of SCs and TACs, and death of DCs. It turns out that in this case, the smallest number of control loops is three, and there are exactly 32 different three-loop control networks that are stable, see figures 4.2 and 4.3. These 32 stable networks have different topologies and different signs of control loops (positive or negative).

For our application, we are interested in determining the most likely control network(s) that govern the regulation of human colon crypt stem cell lineages. While the analysis of [1] restricts the total number of possibilities to 20 (plus additional 12 networks with non-constant DC death terms), the analysis did not indicate which of the 32 were most likely to describe the regulation in a real biological system.

Our approach determined the most likely regulatory network(s), among the 32 three-compartment stable networks, by using actual measurements of the number of SCs, TACs, and DCs in biopsies of human colon crypts, and additional mathematical methodology. [58] performed detailed measurements of the number of each of the three cell types (SCs, TACs, DCs) in 49 colon crypts in human biopsy specimens. We examined each of the 32 possible networks of [1] to determine whether it can produce the correct measured means and variances of cell population numbers. In addition to this static information we have also used data on the dynamics of injury recovery, as well as experimentally obtained intra-crypt correlations. Using these criteria, a selection algorithm was devised that identified three of the 32 possible control networks as most likely the ones corresponding to the regulation of homeostasis of human colon crypts. The details will be covered in **Chapter 4**.

This work demonstrates how theoretical analysis of control networks combined with only static biological data can shed light onto the inner workings of stem cell lineages, in the absence of direct experimental assessment of regulatory signaling mechanisms. The work contributes to the growing literature on the theory of stem cells, which ranges from ODE modeling [49, 51] to stochastic modeling [52–55, 57], and includes research of stem cells in the context of feedback mechanisms [59–62], carcinogenesis [35–37, 40, 63–67], modeling hematopoietic SC dynamics [41, 46, 68, 69], and cancer stem cells [40, 70–74].

Chapter 2

Symmetric Division of Stem Cells

2.1 General approach: modeling non-constant total cell populations

In a general setting, the total number of cells in the system is not a constant number. The number of stem cells, I, and the number of differentiated cells, J, vary independently, giving rise to a 2D Markov process. Let us suppose that in an infinitesimal time-interval, Δt, the following events can occur:

- With probability $L_{I,J}\Delta t$ a stem cell divides. Two types of division are possible.
  - With probability $L_{I,J}P_{I,J}\Delta t$ a stem cell differentiation takes place resulting in a creation of two differentiated cells, $(I, J) \to (I - 1, J + 2)$.
  - With probability $L_{I,J}(1 - P_{I,J})\Delta t$ a stem cell proliferation takes place resulting in a creation of a stem cell, $(I, J) \to (I + 1, J)$.
- With probability $D_{I,J}\Delta t$, a differentiated cell dies, $(I, J) \to (I, J - 1)$.

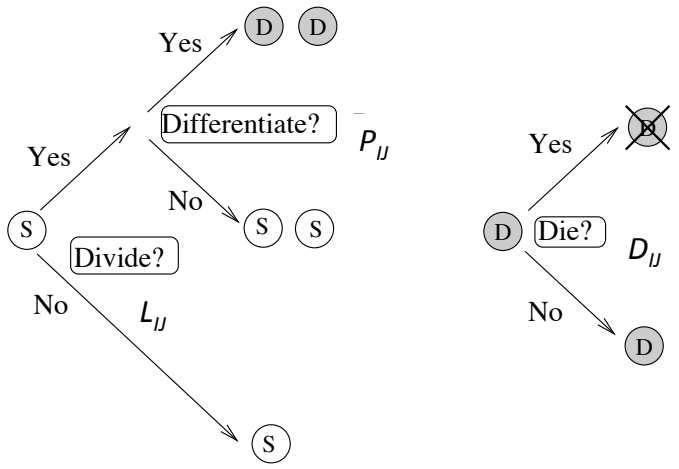


Figure 2.1: A non-constant total population process. A schematics showing the cellular processes and their probabilities. Circles represent stem cells ("S") and differentiated cells ("D"). A stem cell divides with probability $L_{I,J}$, where I and J are the current populations of stem and differentiated cells respectively. The division can be a differentiation event (with probability $P_{I,J}$) or a proliferation event (with probability $1 - P_{I,J}$). A differentiated cell dies with the rate $D_{I,J}$.

All other events are assumed to happen with zero probability. The processes described above are illustrated schematically in figure 2.1. Let us denote by $\varphi_{IJ}(t)$ the probability to have I stem cells and J differentiated cells at time t. The Kolmogorov forward equation corresponding to the above processes is given by:

$$\begin{aligned}\dot{\varphi}_{I,J} &= \varphi_{I,J+1}D_{I,J+1} + \varphi_{I-1,J}L_{I-1,J}(1 - P_{I-1,J}) + \varphi_{I+1,J-2}L_{I+1,J-2}P_{I+1,J-2} \\ &- \varphi_{I,J}(L_{I,J} + D_{I,J}).\end{aligned} \tag{2.1}$$

2.1.1 Previous results for a specific case

A specific form of this process was studied in [75], where we assumed

$$L_{I,J} = \frac{b}{1 + h(I + J)}, \quad P_{I,J} = \frac{r}{1 + gJ}, \quad D_{I,J} = 1 - L_{I,J}, \tag{2.2}$$

and parameters h and g are small. In this case, the probabilities of divisions and deaths are functions of the variable $N = I + J$, and the probability of differentiation is a function of the number of differentiated cells, J, only. The resulting means and variances are listed below:

$$
\begin{aligned}
E[N] &= \frac{1}{2} + \frac{2b-1}{h}, && (2.3)\\
Var[N] &= \frac{1}{4} + \frac{b}{h}, && (2.4)\\
E[J] &= \frac{2bg - 8h + 7gh + 20hr - \sqrt{4b^2g^2 + h^2(g-4r)^2 + 4bgh(4r-5g)}}{8gh},\\
Var[J] &= \frac{23}{32} - \frac{b^2}{8h^2} - \frac{b}{8h} - \frac{r}{4g} + \frac{br}{2gh} + \frac{3r^2}{2g^2}\\
&\quad + \frac{2gb + 7gh - 12hr}{32g^2h^2}\sqrt{4b^2g^2 + h^2(g-4r)^2 + 4bgh(4r-5g)}.
\end{aligned}
$$

It is easier to interpret the results for the differentiated cells if we consider the behavior in the limit of weak control, that is when $h \to 0$ or $g \to 0$. We have

$$
\text{If } h \to 0: \quad E[J] = \frac{2r-1}{g} + 1 + O(h), \quad Var[J] = \frac{b}{h} + \frac{r}{g} + \frac{1}{2} + O(h), \tag{2.5}
$$

$$
\text{If } g \to 0: \quad E[J] = \frac{2r-1}{g} + \frac{3}{2} + O(g), \quad Var[J] = \frac{3r}{g} - \frac{3}{4} + O(g). \tag{2.6}
$$

In the following we will develop a general method of calculating the means and the variances of non-constant population systems. It will be demonstrated how the above results can be obtained in a way much simpler than that of [75].

2.1.2 The general 2D model of control

Let us define the steady state of the system, (i_0, j_0), by the the following equations:

$$L_{i_0,j_0} = D_{i_0,j_0} \equiv L_0, \quad P_{i_0,j_0} = \frac{1}{2}. \tag{2.7}$$

Similar to the previous section, we will use lower-case letters to measure the difference between the current cell numbers and their equilibrium numbers: $i = I - i_0$, $j = J - j_0$.

Define $\tilde{\varphi}$ such that $\tilde{\varphi}_{i,j} = \varphi_{I,J}$, and $\tilde{Z}_{i,j} = Z_{I,J}$, where $Z_{I,J}$ denotes any of the functions $L_{I,J}, P_{I,J}, D_{I,J}$. Then equation (2.1) can be expressed as:

$$\begin{aligned} \dot{\tilde{\varphi}}_{i,j} &= \tilde{\varphi}_{i,j+1}\tilde{D}_{i,j+1} + \tilde{\varphi}_{i-1,j}\tilde{L}_{i-1,j}(1 - \tilde{P}_{i-1,j}) + \tilde{\varphi}_{i+1,j-2}\tilde{L}_{i+1,j-2}\tilde{P}_{i+1,j-2} \\ &- \tilde{\varphi}_{i,j}(\tilde{L}_{i,j} + \tilde{D}_{i,j}). \end{aligned} \tag{2.8}$$

Let us use the following short-hand notation for the moments:

$$x_{\alpha\beta} \equiv E[i^\alpha j^\beta] := \sum_{i,j} \tilde{\varphi}_{i,j} i^\alpha j^\beta. \tag{2.9}$$

Then, we obtain:

$$E[I] \equiv \sum_{I,J} I \cdot \varphi_{I,J} = \sum_{i,j} (i + i_0)\tilde{\varphi}_{i,j} = x_{10} + i_0; \tag{2.10}$$

$$E[I^2] \equiv \sum_{I,J} I^2 \cdot \varphi_{I,J} = \sum_{i,j} (i + i_0)^2 \tilde{\varphi}_{i,j} = x_{20} + 2i_0 x_{10} + i_0^2; \tag{2.11}$$

$$Var[I] = E[I^2] - E[I]^2 = x_{20} - x_{10}^2. \tag{2.12}$$

Similarily, we have:

$$E[J] = x_{01} + j_0; \quad Var[J] = x_{02} - x_{01}^2. \tag{2.13}$$

Our goal now is to find the quantities $x_{01}, x_{10}, x_{02}, x_{20}$, which are essential for calculating the expectation and variance for I and J. In order to derive equations for these quantities, we multiply equation (2.8) by $i^\alpha j^\beta$ with $\alpha + \beta \leq 2$, and perform the summation in i and j in the quasi-stationary state. Each of the resulting 5 moment equations involve higher moments, which means that the number of the unknowns is larger than the number of equations. More precisely, these 5 equations involve 20 unknown variables, $x_{\alpha\beta}$, with $1 \leq \alpha + \beta \leq 5$. In order to proceed, we need to implement a truncation methods to close the system. In other words, we need to derive the missing equations for the higher moments, $x_{\alpha\beta}$, $3 \leq \alpha + \beta \leq 5$.

As in the 1D case, there are at least 3 different truncation methods that can be readily implemented to solve this system:

- **Simple truncation method** assumes that all the higher moments are 0, that is, $x_{\alpha\beta} = 0$, for $\alpha + \beta \geq 3$. This method requires the least amount of calculations, because by assuming that all the higher moments are 0, we get a 5×5 linear system of equations.
- **Central moment truncation method** assumes that the central moments are equal to 0 for higher orders, that is, $E[(i - E[i])^\alpha (j - E[j])^\beta] = 0$, where $\alpha + \beta \geq 3$.
- **Cumulant truncation method** assumes that the higher order multivariate cumulants are equal to 0, that is, $\kappa_{\alpha,\beta} = 0$, where $\alpha + \beta \geq 3$.

The three methods are compared and contrasted in Appendix C.

In order to solve the resulting system of algebraic equations, we use the approximation of weak dependencies of the control functions on the cell numbers. Let us suppose that we

can represent the functions $L_{I,J}$, $D_{I,J}$, and $P_{I,J}$ near the equilibrium as $L_{I,J} = L(\epsilon I, \epsilon J)$, $P_{I,J} = P(\epsilon I, \epsilon J)$, and $D_{I,J} = D(\epsilon I, \epsilon J)$, where the parameter $\epsilon \ll 1$ defines the weakness of the dependence. It is convenient to denote $x = \epsilon I, y = \epsilon J$. Then we can expand the functions $L_{I,J}$, $P_{I,J}$ and $D_{I,J}$ around the steady state in Taylor series:

$$\begin{aligned}
\tilde{L}_{i,j} &\equiv L_{I,J} = L(\epsilon i_0 + \epsilon(I - i_0), \epsilon j_0 + \epsilon(J - j_0)) = L_0 + l_x i + l_y j + \\
&\quad \frac{1}{2}(l_{xx} i^2 + l_{yy} j^2 + 2 l_{xy} ij) + \cdots , \qquad (2.14) \\
\tilde{D}_{i,j} &\equiv D_{I,J} = L_0 + d_x i + d_y j + \frac{1}{2}(d_{xx} i^2 + d_{yy} j^2 + 2 d_{xy} ij) + \cdots , \qquad (2.15) \\
\tilde{P}_{i,j} &\equiv P_{I,J} = \frac{1}{2} + p_x i + p_y j + \frac{1}{2}(p_{xx} i^2 + p_{yy} j^2 + 2 p_{xy} ij) + \cdots , \qquad (2.16)
\end{aligned}$$

where the subscripts denote the partial derivatives of the functions with respect to its argument, evaluated at the equilibrium, $(I, J) = (i_0, j_0)$, and $I = i_0 + i, J = j_0 + j$. We further adopt the following convention: $l_x = L_x \epsilon, l_{xx} = L_{xx} \epsilon^2$, etc. In this description, the upper case constants $L_x = O(1)$, $L_{xx} = O(1)$, etc are all of order one, and all the derivatives expressed by lower-case letters contain a power of ϵ. In particular, the first derivatives l_x, l_y, p_x, etc contain a factor ϵ, and all the second derivatives l_{xx}, l_{xy}, l_{yy}, p_{xx}, etc contain a factor ϵ^2. In Appendix C we demonstrate that all three truncation methods give the same result in the highest order of expansion in terms of ϵ.

2.1.3 Results for the cell number means and variances

Here we present the results for the means and the variances of the cell numbers. Let use define the pair (i_0, j_0) by equation (2.7), and derive the equations for the moments (the summation equations) by expanding the probability functions around this point, see Appendix A. By the simple truncation method we set $x_{\alpha\beta} = 0$ for all $\alpha + \beta > 2$ and obtain the following five

equations for the first and second moments:

$$(1,0) \qquad L_0P_x x_{10} + L_0P_y x_{01} + \epsilon L_xP_x x_{20} + \epsilon(L_yP_x + L_xP_y)x_{11} + \epsilon L_yP_y x_{02} = 0, \tag{2.17}$$

$$(0,1) \qquad (-D_x + L_x + 2L_0P_x)x_{10} + (-D_y + L_y + 2L_0P_y)x_{01} + 2\epsilon L_xP_x x_{20} + 2\epsilon(L_yP_x + L_xP_y)x_{11} + 2\epsilon L_yP_y x_{02} = 0, \tag{2.18}$$

$$(2,0) \qquad L_0 + \epsilon L_x x_{10} + \epsilon L_y x_{01} - 4\epsilon L_0P_x x_{20} - 4\epsilon L_0P_y x_{11} = 0, \tag{2.19}$$

$$(1,1) \qquad L_0 + \epsilon(L_x + 2L_0P_x)x_{10} + \epsilon(L_y + 2L_0P_y)x_{01} + \epsilon(D_x - L_x - 2L_0P_x + 2\epsilon L_xP_x)x_{20} + \epsilon(D_y - L_y + 2L_0(P_x - P_y) + 2\epsilon(L_yP_x + L_xP_y))x_{11} + 2\epsilon(L_0P_y + \epsilon L_yP_y)x_{02} = 0, \tag{2.20}$$

$$(0,2) \qquad 3L_0 + \epsilon(D_x + 2L_x + 4L_0P_x)x_{10} + \epsilon(D_y + 2L_y + 4L_0P_y)x_{01} + 4\epsilon^2 L_xP_x x_{20} + 2\epsilon(-D_x + L_x + 2L_0P_x + 2\epsilon(L_yP_x + L_xP_y))x_{11} + 2\epsilon(-D_y + L_y + 2L_0P_y + 2\epsilon L_yP_y)x_{02} = 0. \tag{2.21}$$

Solving these to the highest order terms in ϵ and using (2.10), (2.12), and (2.13), we obtain the following result:

$$\begin{aligned} E[I] &= i_0 + O(\epsilon^0), && (2.22)\\ E[J] &= j_0 + O(\epsilon^0), && (2.23)\\ Var[I] &= \frac{q_y^2 + 8L_0^2 p_y^2 + 2L_0\Delta}{4B\Delta} + O(\epsilon^0), && (2.24)\\ Var[J] &= \frac{q_x^2 + 8L_0^2 p_x^2 + 6L_0\Delta}{4B\Delta} + O(\epsilon^0), && (2.25) \end{aligned}$$

where we used the following notations:

$$\begin{aligned} &q_x = l_x - d_x, \quad q_y = l_y - d_y,\\ &\Delta = p_y q_x - p_x q_y, \quad B = 2L_0(p_x - p_y) - q_y. \end{aligned} \tag{2.26}$$

Appendix B demonstrates the application of our methods to system (2.2). As discussed, all three methods yield the same result in the highest order of expansion, and the results coincide with the ones previously obtained. The correction terms are different in different methods.

Quantities Δ and B, equation (2.26), are key for determining the stability properties of the stem cell lineage. It was shown in [1] that conditions $\Delta > 0$, $B > 0$ are necessary and sufficient for stability. Furthermore, equations (2.24) and (2.25) relate these quantities with the size of variance experienced by the cells in the stem and differentiated compartment. Expressions (2.26) define a subset in the four-dimensional parameter space, (q_x, q_y, p_x, p_y), that is compatible with stability. Minimizing the variance in expressions (2.24, 2.25) restricts this subset further to identify the most general parameter region that is consistent with stable homeostasis. Note that only local properties (the derivatives at the steady state) of

the control functions are needed to characterize homeostasis.

2.2 Numerical Simulations

In this section, we will demonstrate that the formulas given by equations (2.22-2.25) agree with the results from numerical simulations via two examples on two different types of control.

Throughout this section, let us denote $x = \epsilon I,\ y = \epsilon J,\ \hat{q}_x = L_x - D_x,\ \hat{q}_y = L_y - D_y$. Thus, $\hat{q}_x$ and $\hat{q}_y$ are the partial derivatives of the net growth rate, $L - D$, with respect to x and y. To clarify the biological meaning of these parameters, consider the quantity L_y. If it is nonzero, it means that the probability of stem cell division is controlled by the differentiated cell population. Moreover, if $L_y < 0$, this means that the control is negative (the more differentiated cells in the system, the less likely the stem cells are to divide); $L_y > 0$ means the existence of a positive control loop. The other three quantities can be interpreted in a similar manner. Below are two examples.

Negative control of differentiation and division. Consider the following functional forms of negatively controlled rates of division and differentiation:

$$\begin{aligned} L_{I,J} &= L(\epsilon I, \epsilon J) = e^{-\epsilon I}, \quad P_{I,J} = P(\epsilon I, \epsilon J) = e^{-\epsilon J}, \\ D_{I,J} &= D(\epsilon I, \epsilon J) = 1 - L_{I,J}. \end{aligned} \tag{2.27}$$

We therefore have $P_x = 0,\ P_y = -e^{-\epsilon J} < 0,\ \hat{q}_x = -2e^{-\epsilon I} < 0,\ \hat{q}_y = 0$. The steady state of the system can be obtained by solving $P(x, y) = \frac{1}{2}$, and $L(x, y) = D(x, y)$:

$$i_0 = j_0 = \frac{\log 2}{\epsilon}.$$

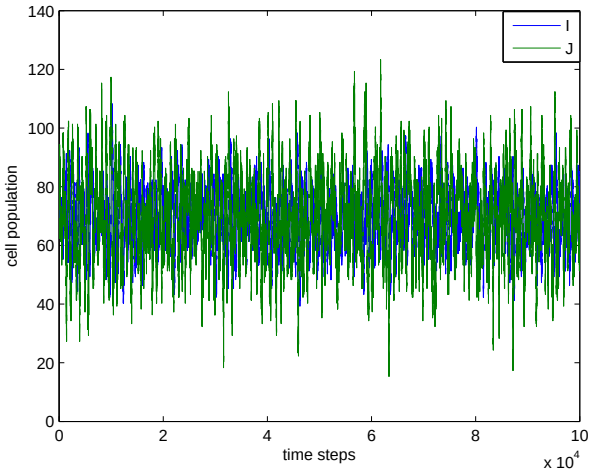


Figure 2.2: A numerical simulation of the system in (2.27) with $\epsilon = 0.01$ ran for 10^5 time steps. ('I') stands for the stem cell population, and ('J') stands for the differentiated cell population.

By equations (2.22-2.25), we can obtain the means and the variances of the system:

$$E[I] = i_0, \tag{2.28}$$

$$E[J] = j_0, \tag{2.29}$$

$$Var[I] = \frac{8L_0^2P_y^2 + 2L_0P_y\hat{q}_x}{-8L_0P_y^2\hat{q}_x} \cdot \frac{1}{\epsilon}, \tag{2.30}$$

$$Var[J] = \frac{\hat{q}_x^2 + 6L_0P_y\hat{q}_x}{-8L_0P_y^2\hat{q}_x} \cdot \frac{1}{\epsilon}, \tag{2.31}$$

where all the partial derivatives are evaluated at (i_0, j_0), and $L_0 = L(\epsilon i_0, \epsilon j_0) = 1/2$.

For each value of ϵ, we ran numerical simulations starting at the expected values of the cell population given above, and finishing either when the number of time-steps reached $2 \cdot 10^5$, or if any of the cell types went extinct. We then computed the means and the variances of the cell population over the time-course of each simulation. A typical run for a particular value of ϵ is presented in figure 2.2.

From figure 2.3, we observe that the theoretical results for the means and the variances show

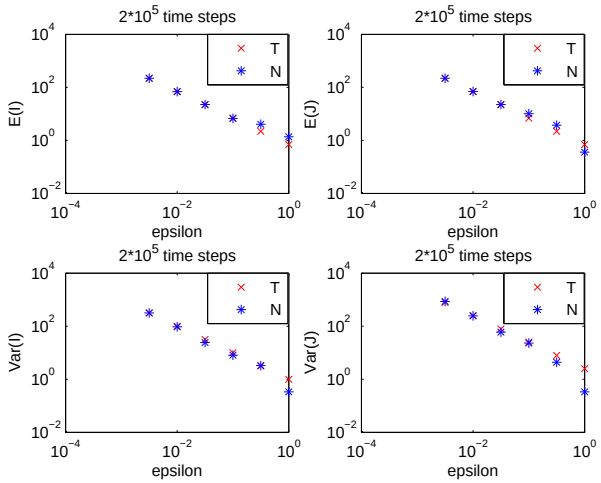


Figure 2.3: The behavior of the means and the variances of the cell population described by equation (2.27). The analytical results given by equations (2.28-2.31) (X's) are compared with the values obtained by numerical simulations (stars), for different values of ϵ. ('T') stands for the theoretical results, and ('N') stands for the numerical results.

a good agreement with the numerical results for smaller values of ϵ, which is what we expect. Also, the means and the variances of the cell population decrease as the value of ϵ increases, which is already predicted by the formulas given by equations (2.28-2.31).

Positive control of differentiation and negative regulation of division. The second example is given by equations:

$$\begin{aligned} L_{I,J} &= L(\epsilon I, \epsilon J) = \frac{1}{1+\epsilon J}, \quad P_{I,J} = P(\epsilon I, \epsilon J) = 0.7 \cdot \tanh(\epsilon I), \\ D_{I,J} &= D(\epsilon I, \epsilon J) = 1 - L(\epsilon I, \epsilon J). \end{aligned} \tag{2.32}$$

A typical stochastic simulation of system (2.32) for a particular value of ϵ is presented in figure 2.4.

To calculate the variances, we calculate $P_x = 0.7 \cdot \mathrm{sech}^2(\epsilon I) > 0$, $P_y = 0$, $\hat{q}_x = 0$, $\hat{q}_y = -2(1+\epsilon J)^{-2} < 0$. The steady state of the system can be obtained by solving $P(x,y) = \frac{1}{2}$,

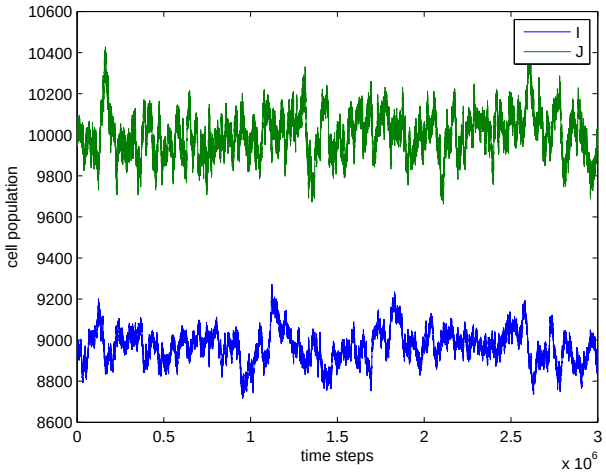


Figure 2.4: A typical numerical simulation of example (2.32) with $\epsilon = 10^{-4}$ and $3 \cdot 10^6$ time steps. ('I') stands for the stem cell population, and ('J') stands for the differentiated cell population.

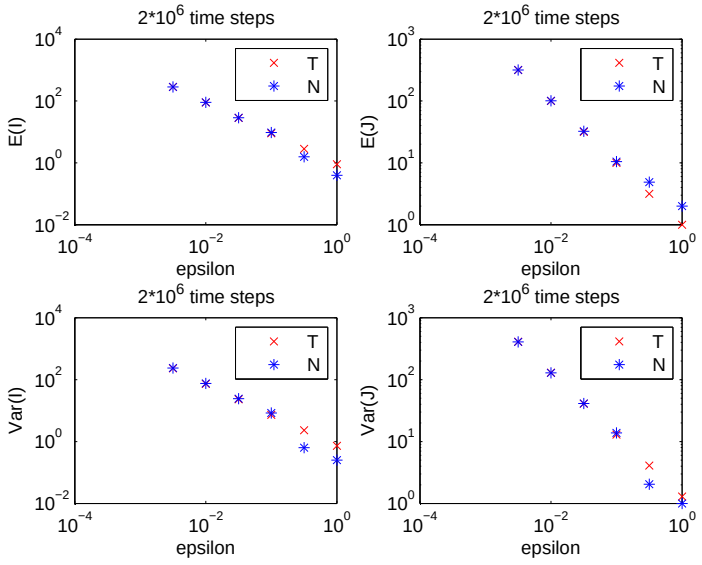


Figure 2.5: Same as in figure 2.3, except the means and variances are calculated of system (2.32).

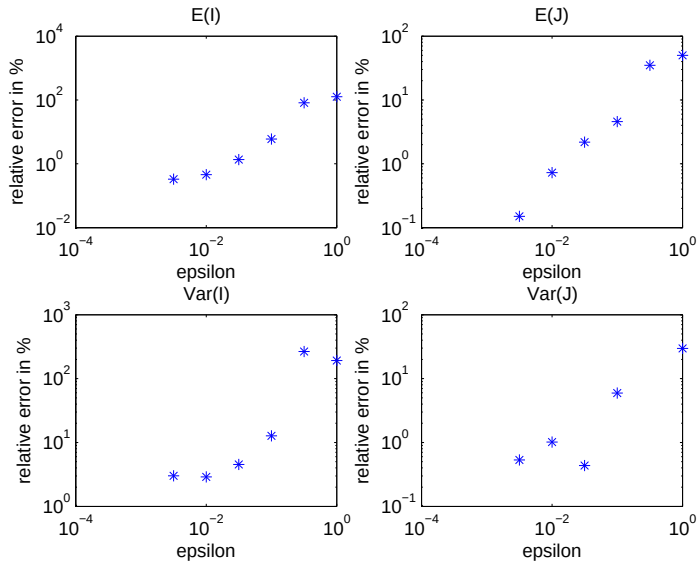


Figure 2.6: The behavior of the relative error of the means and the variances for different values of ϵ. We used the relative error $= \frac{|theoretical\ result - numerical\ result|}{numerical\ result} \times 100\%$.

and $L(x, y) = D(x, y)$:

$$i_0 = \frac{\log 6}{2\epsilon}, \quad j_0 = \frac{1}{\epsilon}.$$

By equations (2.22-2.25), we can obtain the means and the variances of the system:

$$E[I] = i_0, \tag{2.33}$$

$$E[J] = j_0, \tag{2.34}$$

$$Var[I] = \frac{\hat{q}_y^2 + 2L_0\hat{\Delta}}{4\hat{B}\hat{\Delta}} \cdot \frac{1}{\epsilon}, \tag{2.35}$$

$$Var[J] = \frac{8L_0^2P_x^2 + 6L_0\hat{\Delta}}{4\hat{B}\hat{\Delta}} \cdot \frac{1}{\epsilon}, \tag{2.36}$$

where all the partial derivatives are evaluated at (i_0, j_0), and $L_0 = 1/2$, $\hat{\Delta} = P_y\hat{q}_x - P_x\hat{q}_y = -P_x\hat{q}_y$, $\hat{B} = 2L_0(P_x - P_y) - \hat{q}_y = 2L_0P_x - \hat{q}_y$.

We used the same numerical scheme as in the previous example with $2 \cdot 10^6$ time steps. As observed in figure 2.5, the theoretical results are in good agreement with the numerical results

for smaller values of ϵ, which is consistent with the previous example. The means and the variances of the cell population decrease as the value of ϵ increases, which is foretold by equations (2.33-2.36). From figure 2.6, we can see the overall pattern of the relative error: the smaller the value of ϵ, the smaller the relative error, which is what we expect.

2.3 Connection with the power series expansion method of Van Kampen

In this section, we will show that the simple truncation method described here and the well-known power series expansion method of Van Kampen [23] give exactly the same results, up to any order of expansion. First, we demonsrate how the Van Kampen method can be used for our system of stem and differentiated cells, and then argue that the two methods give the same results.

2.3.1 The method of Van Kampen: review and notations

Let us introduce the operators $\mathbf{E}_I^k$ and $\mathbf{E}_J^k$, such that

$$\mathbf{E}_I^k[f_{I,J}] = f_{I+k,J}, \quad \mathbf{E}_J^k[f_{I,J}] = f_{I,J+k}.$$

Then equation (2.1) can be rewritten more conveniently,

$$\dot{\varphi}_{I,J} = (\mathbf{E}_J^{+1}-1)[\varphi_{I,J}D_{I,J}]+(\mathbf{E}_I^{-1}-1)[\varphi_{I,J}L_{I,J}(1-P_{I,J})]+(\mathbf{E}_I^{+1}\mathbf{E}_J^{-2}-1)[\varphi_{I,J}L_{I,J}P_{I,J}]. \quad (2.37)$$

Equation (2.37) is nonlinear, and a general solution cannot be found. Therefore, we will use approximate methods to solve it. Let us assume that the functions $L_{I,J}$, $P_{I,J}$, and $D_{I,J}$

depend weakly on their arguments:

$$L_{I,J} = \mathcal{L}(\epsilon I, \epsilon J), \quad P_{I,J} = \mathcal{P}(\epsilon I, \epsilon J), \quad D_{I,J} = \mathcal{D}(\epsilon I, \epsilon J),$$

where $\epsilon \ll 1$. We will use this parameter to perform the Van Kampen master equation expansion, in order to formulate the linear noise approximation [76]. We expect that in the long run, the probability distribution, $\varphi_{I,J}$, will have a peak somewhere around the (large) values

$$i_0 = \frac{\phi_I}{\epsilon}, \quad j_0 = \frac{\phi_J}{\epsilon},$$

with $\phi_I \sim \epsilon^0$, $\phi_J \sim \epsilon^0$. Let us suppose that the width of those peaks scales with $1/\epsilon^{1/2}$. This is expressed in the following change of variables,

$$I = \frac{\phi_I(t)}{\epsilon} + \frac{\xi(t)}{\epsilon^{1/2}}, \quad J = \frac{\phi_J(t)}{\epsilon} + \frac{\eta(t)}{\epsilon^{1/2}}. \tag{2.38}$$

This change of variables will be used in the master equation (2.37). First of all, the probability function $\varphi_{I,J}(t)$ is now a function of ξ and η:

$$\varphi_{I,J}(t) = \Pi(\xi, \eta; t).$$

Its time-derivative can be written as follows,

$$\frac{d\varphi_{I,J}(t)}{dt} = \frac{\partial \Pi}{\partial t} + \frac{\partial \Pi}{\partial \xi}\dot{\xi} + \frac{\partial \Pi}{\partial \eta}\dot{\eta}.$$

Because the left hand sides of expressions (2.38) are time-independent, we have $\dot{\xi} = -\dot{\phi}_I/\epsilon^{1/2}$, $\dot{\eta} = -\dot{\phi}_J/\epsilon^{1/2}$. Also, we will introduce a slow time-scale,

$$\tau = \epsilon t$$

(the necessity for this rescaling will become apparent once all the terms at different orders of ϵ are collected in the master equation). Therefore, we have for the time-derivative of $\varphi_{I,J}$:

$$\dot{\varphi}_{I,J} = \epsilon \frac{\partial \Pi}{\partial \tau} - \epsilon^{1/2} \left(\frac{\partial \Pi}{\partial \xi} \frac{\partial \phi_I}{\partial \tau} + \frac{\partial \Pi}{\partial \eta} \frac{\partial \phi_J}{\partial \tau} \right). \tag{2.39}$$

Next, we evaluate the shift operators. A jump of size k in the value of I is reflected by the jump of size $k\epsilon^{1/2}$ in the value of ξ:

$$I + k = \frac{\phi_I(t)}{\epsilon} + \frac{\xi(t)}{\epsilon^{1/2}} + k = \frac{1}{\epsilon} \left(\phi_I + \epsilon^{1/2} (\xi + k\epsilon^{1/2}) \right).$$

Similar arguments hold for the values of J. This allows us to express the shift operators $\mathbf{E}_I^k$ and $\mathbf{E}_J^k$ in terms of a (Taylor) series of differential operators,

$$\mathbf{E}_I^k = 1 + k\epsilon^{1/2} \frac{\partial}{\partial \xi} + \frac{k^2 \epsilon}{2} \frac{\partial^2}{\partial \xi^2} + \dots, \tag{2.40}$$

and similarly for the shift in the J-direction.

Finally, we use ansatz (2.38) to expand the functions $L_{I,J}$, $P_{I,J}$, and $D_{I,J}$. We have

$$\mathcal{L}(\epsilon I, \epsilon J) = \mathcal{L}(\phi_I + \epsilon^{1/2}\xi, \phi_J + \epsilon^{1/2}\eta).$$

It is convenient to denote $x = \epsilon I$, $y = \epsilon J$, such that $L_{I,J} = \mathcal{L}(x, y)$, and denote by the subscripts the derivatives of this function with respect to its argument, evaluated at (ϕ_I, ϕ_J): $L_x = \partial \mathcal{L}/\partial x$, $L_y = \partial \mathcal{L}/\partial y$, etc. We have

$$\mathcal{L}(\phi_I + \epsilon^{1/2}\xi, \phi_J + \epsilon^{1/2}\eta) = \mathcal{L}(\phi_I, \phi_J) + \epsilon^{1/2}\xi L_x + \epsilon^{1/2}\eta L_y + \frac{\epsilon}{2}\xi^2 L_{xx} + \frac{\epsilon}{2}\eta^2 L_{yy} + \epsilon \xi \eta L_{xy} + \dots.$$

Similarly, we expand the functions $\mathcal{P}$ and $\mathcal{D}$. These expressions, together with the operator expansions (2.40) and the time-derivative (2.39), are substituted into the master equation

(2.37). Then the terms in different orders of ϵ are equated. At order $\epsilon^{1/2}$ we have

$$\frac{\partial\Pi}{\partial\xi}\frac{d\phi_I}{d\tau}+\frac{\partial\Pi}{\partial\eta}\frac{d\phi_J}{d\tau}=\frac{\partial\Pi}{\partial\xi}\mathcal{L}(\phi_I,\phi_J)(1-2\mathcal{P}(\phi_I,\phi_J))+\frac{\partial\Pi}{\partial\eta}(2\mathcal{L}(\phi_I,\phi_J)\mathcal{P}(\phi_I,\phi_J)-\mathcal{D}(\phi_I,\phi_J)).$$

This equation gives rise to two "macroscopic laws",

$$\frac{d\phi_I}{d\tau}=\mathcal{L}(\phi_I,\phi_J)(1-2\mathcal{P}(\phi_I,\phi_J)),\quad \frac{d\phi_J}{d\tau}=2\mathcal{L}(\phi_I,\phi_J)\mathcal{P}(\phi_I,\phi_J)-\mathcal{D}(\phi_I,\phi_J). \tag{2.41}$$

or in steady state simply

$$\mathcal{P}(\phi_I,\phi_J)=1/2,\quad \mathcal{D}(\phi_I,\phi_J)=\mathcal{L}(\phi_I,\phi_J)\equiv L_0. \tag{2.42}$$

Let us introduce the notations

$$\tilde{q}_x=\frac{L_x-D_x}{L_0}=\frac{q_x}{\epsilon L_0},\quad \tilde{q}_y=\frac{L_y-D_y}{L_0}=\frac{q_y}{\epsilon L_0};$$

where q_x, q_y are defined in section 2.1.3. At order ϵ of the master equation expansion, after rescaling time once more by

$$T=L_0\tau=L_0\epsilon t,$$

we obtain the following linear Fokker-Planck equation:

$$\begin{aligned}\frac{\partial\Pi}{\partial T} &= -(\tilde{q}_y+2P_y)(\eta\Pi)_\eta-(\tilde{q}_x+2P_x)(\xi\Pi)_\eta \\ &+ 2P_x(\xi\Pi)_\xi+2P_y(\eta\Pi)_\xi+\frac{1}{2}(3\Pi_{\eta\eta}-2\Pi_{\xi\eta}+\Pi_{\xi\xi}).\end{aligned} \tag{2.43}$$

This is the linear noise approximation of Van Kampen [76]. The validity of this approximation has been studied extensively, see e.g. [77, 78]. Here we mention that the relative size of typical fluctuations scales with $\epsilon^{1/2}$, and thus for sufficiently small values of ϵ, the system

will remain near the equilibrium and stochastic extinction is an unlikely event, at least for a time-duration which grows with $1/\epsilon$. For a rigorous study of extinction times of birth-death processes see e.g. [79, 80].

From equation (2.43) we can obtain the equations for the first and second moments in a standard way:

$$\begin{aligned}
\frac{d\langle\xi\rangle}{dT} &= -2(P_x\langle\xi\rangle + P_y\langle\eta\rangle), && (2.44)\\
\frac{d\langle\eta\rangle}{dT} &= (\tilde{q}_y + 2P_y)\langle\eta\rangle + (\tilde{q}_x + 2P_x)\langle\xi\rangle, && (2.45)\\
\frac{d\langle\xi^2\rangle}{dT} &= -4(P_x\langle\xi^2\rangle + P_y\langle\xi\eta\rangle) + 1, && (2.46)\\
\frac{d\langle\eta^2\rangle}{dT} &= 2(\tilde{q}_y + 2P_y)\langle\eta^2\rangle + 2(\tilde{q}_x + 2P_x)\langle\xi\eta\rangle + 3, && (2.47)\\
\frac{d\langle\xi\eta\rangle}{dT} &= (\tilde{q}_y + 2P_y)\langle\xi\eta\rangle + (\tilde{q}_x + 2P_x)\langle\xi^2\rangle - 2(P_x\langle\xi\eta\rangle + P_y\langle\eta^2\rangle) - 1. && (2.48)
\end{aligned}$$

As we will show in the next section, the above moment equations are exactly the same as the summation equations (2.17)-(2.21) if we only keep the leading order terms, and hence Van Kampen method and simple truncation method give the same results to the leading order, see (2.22)-(2.25).

2.3.2 Comparison of the simple truncation method and the Van Kampen method

Before we illustrate the equivalence of the two methods, we state for convenience some fundamental facts that we will use later:

- Taylor series expansion. We know the Taylor expansion of $f(j) = j^n$ center at $j = a$ is: $j^n = a^n + na^{n-1}(j-a) + ... + na(j-a)^{n-1} + (j-a)^n$. Notice that the second to the

last term is a product of the jump size a and the derivative of the last term $(j-a)^n$.

- Integration by parts. When we compute $\int \eta^n \ (\cdot)_\eta \ d\eta$, integration by parts will give $\eta^n \ (\cdot) - \int n \ \eta^{n-1} \ (\cdot) \ d\eta$.

The leading order. By using the ansatz (2.38) in the previous section, we have in steady states:

$$x_{\alpha\beta} = \sum_{I,J} \varphi_{I,J} (I - i_0)^\alpha (J - j_0)^\beta = \frac{\langle \xi^\alpha \eta^\beta \rangle}{\epsilon^{(\alpha+\beta)/2}}; \tag{2.49}$$

where $x_{\alpha\beta}$ is defined in (2.9).

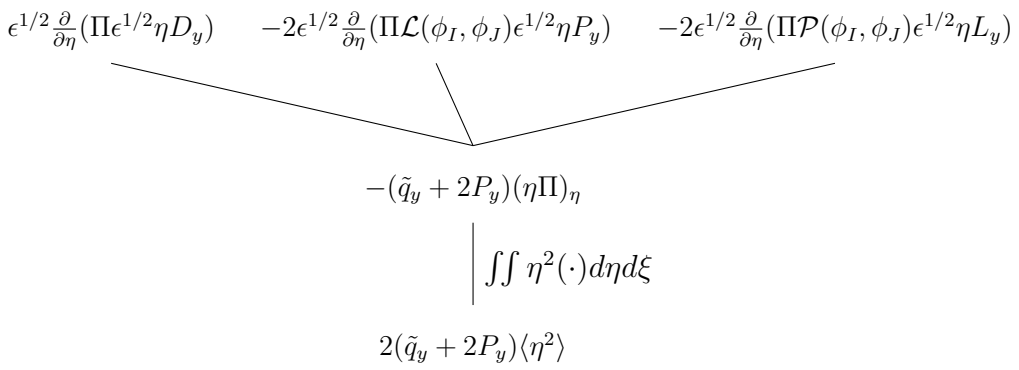
By using (2.49), we can rewrite the summation equations from section 2.1.3 in terms of $\langle \xi^\alpha \eta^\beta \rangle$. It turns out that they are the same as moment equations (2.44-2.48) if we only keep the leading order terms. Since the moment equations agree to the leading order, the two methods give the same results to the leading order.

The next order. Next, we investigate if the two methods provide the same result in the next order of accuracy. To this end, we will investigate the structure of a moment equation in depth from both methods by looking at a particular term. For illustration, we will analyze the first term in equation (2.47) and its counterpart in equation (2.21) from section 2.1.3:

$$\frac{d\langle \eta^2 \rangle}{dT} = 2(\tilde{q}_y + 2P_y)\langle \eta^2 \rangle + ... \tag{2.50}$$

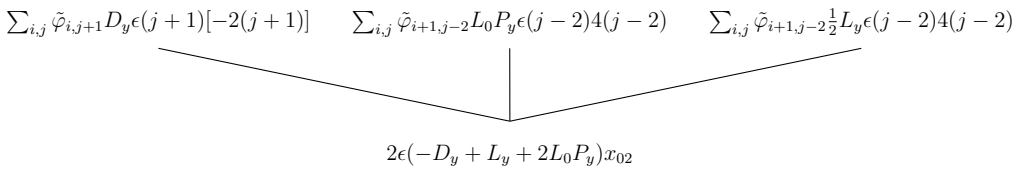
$$\frac{dx_{02}}{dt} = 2\epsilon(-D_y + L_y + 2L_0P_y)x_{02} + ... \tag{2.51}$$

If we trace back the terms which contribute to $2(\tilde{q}_y + 2P_y)\langle\eta^2\rangle$ in (2.50), we will have the following diagram:



Let us focus on the third term at the first level, we have the following observation: $-2\epsilon^{1/2}\frac{\partial}{\partial\eta}$ is the second term of Taylor series of E_J^{-2}; $\mathcal{P}(\phi_I, \phi_J)$ is the first term of $P_{I,J}$ in Taylor expansion; $\epsilon^{1/2}\eta L_y$ is the third term of Taylor series of $L_{I,J}$.

To draw comparison, we will also trace back the terms which contribute to $2\epsilon(-D_y + L_y + 2L_0P_y)x_{02}$ in (2.51). We obtain the following picture:



Let us look at the corresponding counterpart at the first level. Observe that $1/2$ is the first term of Taylor expansion of $P_{i+1,j-2}$; $L_y\epsilon(j-2)$ is the third term of $L_{i+1,j-2}$ in Taylor

expansion; $4(j-2)$ is the second to the last term of Taylor expansion of j^2 centered at $j=2$. It is not hard to see that the term $4(j-2)$ "captures" the jump size of E_J^{-2}; namely 2, and the derivative of η^2 (obtained from integration by parts in the first diagram); moreover, the differential operator $-\frac{\partial}{\partial \eta}$ will be offset after integration by parts. Hence, the two terms are exactly the same. In fact, it can be shown that any two corresponding terms at the first level (from both diagrams) are the same by similar analysis, which are due to the fundamental properties mentioned at the beginning of this section. This methodology is essentially carried over in any two corresponding terms in a homologous pair of moment equations.

To see whether the two methods agree to the next order corrections, we need to assume γ is sufficiently large, where γ is the constant for which $x_{\alpha\beta}=0$ for $\alpha+\beta\geq\gamma$ in the simple truncation method. It turns out that the moment equations that are used to compute the next order correction are the same for the two methods if we set $\gamma=5$, see Appendix E for details.

Generalization to higher orders of accuracy. By the methodology presented in the previous section, we can deduce that the two methods will produce the same moment equations for computing any order corrections if we set γ sufficiently large, hence the two methods give exactly the same results (up to any order) to the general two-step model as stated at the beginning. However, as the value of γ increases, the computation of moment equations will become more and more tedious. So, there is a trade off between efficiency and accuracy.

Under the same value of γ, cumulant truncation method is the most accurate among the three truncation methods presented in section 2.1.2. To see this, set $\gamma=3$ and look at the leading order of the terms $x_{\alpha\beta}$ for $\alpha+\beta=4$. From Appendix D, we see that the leading order is $O(\frac{1}{\epsilon})$ and $O(\frac{1}{\epsilon^2})$ from central moment truncation and cumulant closure method, respectively. On the other hand, we can obtain the leading order of $\langle \xi^\alpha \eta^\beta \rangle$ is $O(1)$ for $\alpha+\beta=4$ from Van Kampen method, as shown in E of appendix. By (2.49), the leading order of $x_{\alpha\beta}$ is $O(\frac{1}{\epsilon^2})$

for $\alpha + \beta = 4$. Therefore, cumulant truncation method is the most accurate though it's the most expensive in terms of computation, which verdicts the trade off between efficiency and accuracy.

In conclusion, simple truncation method produces the same results as Van Kampen method up to any order by setting γ sufficiently large. The advantage of simple truncation method is the straightforward calculations that it involves. To see this, we can compare the steps used to obtain the moment/summation equations in each method. For simple truncation method, we only use Taylor expansion on the probability functions of the master equation. On the other hand, Van Kampen method uses integration besides the master equation expansion in Taylor series. Our method could be regarded as a short-cut compared to the Van Kampen derivation. To see this, recall that we multiplied the equation (2.43) by $\xi^\alpha \eta^\beta$, and then integrated to obtain the moment equations given by (2.44-2.48), for $\alpha + \beta \leq 2$. This extra step requires more computational work. Clearly, there will be more terms to integrate in order to compute the moment equations for the next order corrections, since we extend the equation (2.43) to order $O(\epsilon^{1/2})$, see appendix E.1.

Chapter 3

Asymmetric Division of Stem Cells

In this chapter, we address the questions of SC division symmetry by means of mathematical modeling. Our approach is based on that developed in Chapter 2. The focus is to investigate how different division types contribute to lineage homeostasis/turnover. We provide analysis that allows to quantify the ability of two types of divisions (symmetric and asymmetric) to maintain homeostasis. Intuitively, Asymmetric SC division appears to be more associated with a robust homeostatic maintenance. It can be argued that purely asymmetric SC divisions do not change the total number of SCs and therefore ensure the maintenance of a constant cell population, see e.g. [12]. However, in this chapter we show that it is not necessarily the case: asymmetric divisions can either stabilize or destabilize the lineage system, depending on the underlying control network. In the remaining chapter, the results of our findings are stated at first, followed by the methodology.

3.1 Results

We study a stochastic model with various control loops that distinct cell populations impose on the prevalence of different processes. For example, consider the simplest lineage, which only consists of SCs and one type of daughter cells. We postulate that such a system consists of two compartments, that of SCs and the differentiated cells. We further assume that the rate of SC divisions, and also the probability of differentiation/proliferation (see figure 1.1), are controlled by chemical factors (such as morphogenetic growth factors) secreted by cells of different compartments, as well as exogenous factors coming from outside of the lineage (such as distinct niche cells). In figure 3.1, the endogenous controls are illustrated by using a simple example of symmetric divisions of SCs. An individual decision tree of a SCs is depicted schematically. It consists of the decision to undergo a division, followed by the decision about the nature of this division (that is, whether daughter cells will maintain SC fate or undergo differentiation). In the example in figure 3.1, the probability to divide is limited by the population of daughter cells. If there are too many of them, this will reduce the chances of further divisions. In the same system, the probability of differentiation is influenced by the number of SCs. The more SCs there are, the more likely they will be to differentiate, thus reducing the total SC number. We refer to this system (which in the case of figure 3.1 consists of only two controls) as a control network. In the above scenario, one control is positive, and the other is negative.

It is possible to construct many other control networks that consist of different numbers of positive and/or negative controls. [1] shows that the control network in figure 3.1 (along with many other networks) is compatible with stable maintenance of a constant cell population size. The resulting system of cells is characterized by a stochastic behavior, where the numbers of stem and daughter cells fluctuate around certain mean values. The size of these fluctuations is an important characteristic of a biological system. If these fluctuations are too large (compared to the means) then the population is running a danger of going extinct, which will

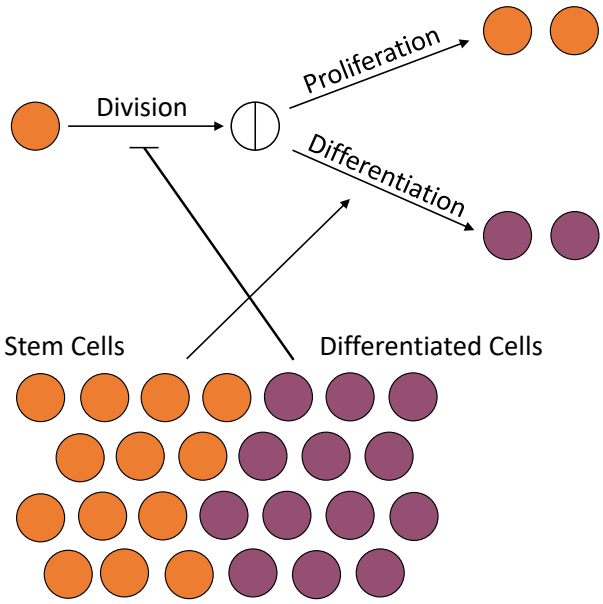


Figure 3.1: An example of endogenous control loops regulating SC decisions with the all symmetric division mode. Division events are negatively regulated by daughter cells and differentiation decisions are positively regulated by SCs.

be a catastrophic outcome for a biological system. The smaller the fluctuations, the more robust is the system and the tighter is homeostatic maintenance. We are interested in the general question of design: what features of control improve the robustness of the system in the sense described above.

In figure 3.1, only symmetric divisions are considered. At the next level of complexity we also consider the possibility of asymmetric divisions. Thus, we can assume that SCs can divide both symmetrically and asymmetrically, with a given relative probability. Here we study how the balance between symmetric and asymmetric SC divisions can change the robustness properties of the lineage. What percentage of divisions should be symmetric to minimize fluctuations for tighter homeostatic control?

Qualitative intuitive reasoning suggests that asymmetric divisions must be associated with

the highest level of stability of SC lineages. In the case of asymmetric divisions, the number of SCs does not change, because every time a SC divides, it replaces itself with exactly one SC, and also produces a differentiated cell. Therefore, it might seem that under fully asymmetric divisions, as long as the production of differentiated cells is balanced on average by their deaths, the system will be stable. It turns out however (see the Methods section) that depending on the exact control loops acting in the system of SCs and non-SC daughter cells, asymmetric divisions might either increase or decrease lineage size fluctuations. This is what we demonstrate next.

3.1.1 The role of division symmetry in stable homeostasis: the case of minimal control systems

Let us suppose that the lineage consists of two types of cells (two compartments), SCs and daughter (differentiated) cells. Let us denote by I and J the current number of stem and daughter cells, respectively. The processes of division (including differentiation/proliferation decisions) and death are dictated by probabilities and rates defined in table 3.1(a). Next, we need to quantify the control loops that exist in a given system.

We assume that $L_{I,J} = L(\epsilon I, \epsilon J)$, $D_{I,J} = D(\epsilon I, \epsilon J)$, etc, where ϵ measures the strength of dependence of the probabilities and rates on the cell population numbers. It is convenient to introduce the continuous variables $x = \epsilon I$, $y = \epsilon J$. To define the control network, we consider the partial derivatives of the rates and probabilities with respect to x and y, evaluated at the equilibrium. We will use the subscripts x and y to denote such partial derivatives, see table 3.1(b). A two-compartment system is characterized by the following four derivatives: p_x, p_y, q_x, and q_y, which we call *controls*. To clarify the biological meaning of these parameters, consider the quantity p_y. If it is nonzero, it means that the probability of SC differentiation is controlled by the differentiated cell population. Moreover, if $p_y < 0$, this means that the

Table 3.1: Notations used in the models

(a) Processes	
$L_{I,J}$	Division rate of SCs
$S_{I,J}$	Probability that the division is symmetric
$P_{I,J}$	Probability that a symmetric division is a differentiation event
$D_{I,J}$	Death rate of differentiated cells
(b) Controls	
$q_x\,(q_y)$	Partial derivative of $L_{I,J} - D_{I,J}$ with respect to the argument I (J), evaluated at the equilibrium
$p_x\,(p_y)$	Partial derivative of $P_{I,J}$ with respect to the argument I (J), evaluated at the equilibrium

(a) Definitions of rates and probabilities. Subscripts denote functional dependence on the cell populations I and J. (b) The four partial derivatives evaluated at the equilibrium comprise the four controls in a two-compartment system.

control is negative (the more differentiated cells in the system, the less likely the SCs are to differentiate); $p_y > 0$ means the existence of a positive control loop. The other three quantities can be interpreted in a similar manner.

It was shown in [1] that at least two of the four controls must be nonzero in order for the system to have a stable homeostatic equilibrium. Minimal control systems are defined as models with a restricted number of nonzero controls, and are presented in figure 3.2. In the schematic, round cells and star-like cells represent stem and differentiated cells respectively. The first horizontal arrow in each diagram indicates the division decision, and the second horizontal arrow the differentiation decision. Arch-like positive and negative arrows depict the dependence of the two decisions on each population. For example, if a negative arrow originates at SCs and points at the divisions decision, this means that the divisions are negatively controlled by the SC numbers, $q_x < 0$ (see diagram #1 in figure 3.2). It was shown in [1] that with two compartments, there are two distinct minimal control systems with two controls, and three systems with three controls (see also Appendix F).

The first two models (#1 and #2) in figure 3.2 are the only two systems that can be stable in

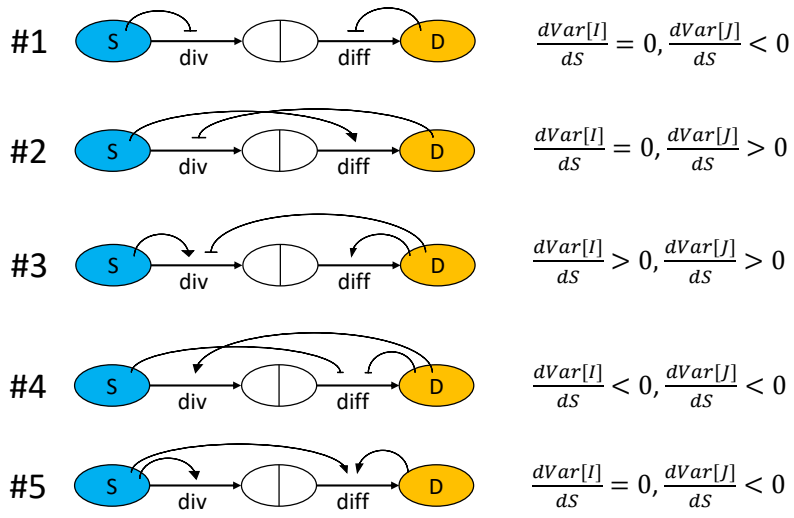


Figure 3.2: Classification of minimal control systems in two-compartment models. Symbol "div" refers to the rate of symmetric stem cell divisions (both proliferations and differentiations). Symbol "diff" refers to the probability of differentiation; the probability of proliferation is 1-Prob(diff). Models #1–2 are the two-control systems. Models #3–5 are three-control systems. Division and differentiation decisions can be positively or negatively controlled by the population sizes of SCs or differentiated cells, as indicated by arch-like arrows that originate at the relevant cell population and point toward the process that this population controls. The rightmost column indicates how cell number variances depend on the symmetry of divisions, as obtained from the analysis of section 3.2.3.

the presence of no more than two controls. The other three models (#3–5 in figure 3.2) are the only three irreducible three-control systems, that is, they cannot be reduced to models #1 or #2 by setting one of the controls to zero. While from the point of view of stability, all five of the networks are possible, further biological considerations are required to identify which control network is relevant for a particular tissue. Some of those considerations may include the matching of various moments of compartment sizes with the observations, robust recovery dynamics, etc.

Next we demonstrate how by varying the proportion of symmetric vs asymmetric SC divisions, one can change homeostatic properties of the system in the context of models #1–5. We will focus on the analysis of variance of the cell populations. A relatively small variance indicates stable, robust homeostasis. A large variance increases the probability of extreme events, such as extinction or growing out of control. By using stochastic analysis (see section 3.2.2) we can calculate the variance of the number of SCs, $Var[I]$, and the variance in the number of

differentiated cells, $Var[J]$, as functions of the parameters. In particular, it is possible to determine how these quantities depend on the four controls (Table 3.1(b)) and the frequency of symmetric SC divisions, S. It turns out that in two out of five control systems in figure 3.2, the variance increases with S. Namely, in systems #2 and #3, $Var[J]$ increases with S, and in addition, in #3 $Var[I]$ also increases with S (in #2, the variance of SC numbers is independent of the symmetry), see equations (3.31) and (3.32). Therefore, in these two control systems, purely asymmetric divisions are optimal from the viewpoint of minimizing fluctuations in cell numbers at homeostasis.

The opposite result is observed for systems #1, #4, and #5. There, purely symmetric divisions turn out to be the optimal choice. In those three systems, the variance of differentiated cell numbers is a decreasing function of S, and in addition, in #4, the variance of SC numbers is also a decreasing function of S, see equations (3.30), (3.33), and (3.34). In these three qualitatively different control networks, symmetric divisions are associated with the most stable homeostatic state. Next, we demonstrate this theoretical finding by numerical simulations.

3.1.2 Application to two control systems

The results reported in the previous section hold for any functional forms of controls. Here we illustrate these findings by considering two specific examples. Some technical details about the simulation setup are provided in Appendix I. Recall that ϵ measures the strength of control of the various processes by the cell population, and $x = \epsilon I$, $y = \epsilon J$; we further denote $\Delta = q_x p_y - q_y p_x$, and $B = 2L_* S_*(p_x - p_y) - q_y$, where the partial derivatives with respect to x and y are defined in table 3.1(b) and the star indicates that the quantity is evaluated at the equilibrium. The quantities Δ and B appear in the expressions for the variances (see section 3.2.2). Throughout this section, we will assume $S_{I,J}$ takes some constant value c,

where $0 < c \leq 1$. Although $S_{I,J}$ is not necessarily constant, its derivatives do not enter the stability conditions or expressions for population variances (as explained in section 3.2.2), and therefore we make the simplest assumption on this function. Below are two examples, where in order to illustrate the theory numerically, we chose some specific functional forms for the controls.

Model #3. Consider three-control model #3 from figure 3.2, which is characterized by negative regulation of division (by differentiated cells) and positive regulation of division (by SCs) and differentiation (by differentiated cells). As an example of this kind of a model, we assign the following functional forms of the controls:

$$\begin{aligned} L_{I,J} &= L(\epsilon I, \epsilon J) = \frac{1 - e^{-\epsilon I}}{1 - e^{-\epsilon I} + \epsilon J}, \quad P_{I,J} = P(\epsilon I, \epsilon J) = 1 - e^{-3\epsilon J}, \\ D_{I,J} &= D(\epsilon I, \epsilon J) = 1 - L_{I,J}, \quad S_{I,J} = c. \end{aligned} \tag{3.1}$$

We therefore have $p_x = 0$, $p_y = \epsilon 3 e^{-3y} > 0$, $q_x = \epsilon 2 y e^{-x} \cdot (1 - e^{-x} + y)^{-2} > 0$, $q_y = \epsilon \frac{2e^{-x} - 2}{(1 - e^{-x} + y)^2} < 0$. The steady state of the system can be obtained by solving $P(x, y) = 1/2$, $L(x, y) = D(x, y)$ (system (3.14) in Methods):

$$i_0 = -\frac{\log(1 - \log 2^{1/3})}{\epsilon}, \quad j_0 = \frac{\log 2}{3\epsilon}.$$

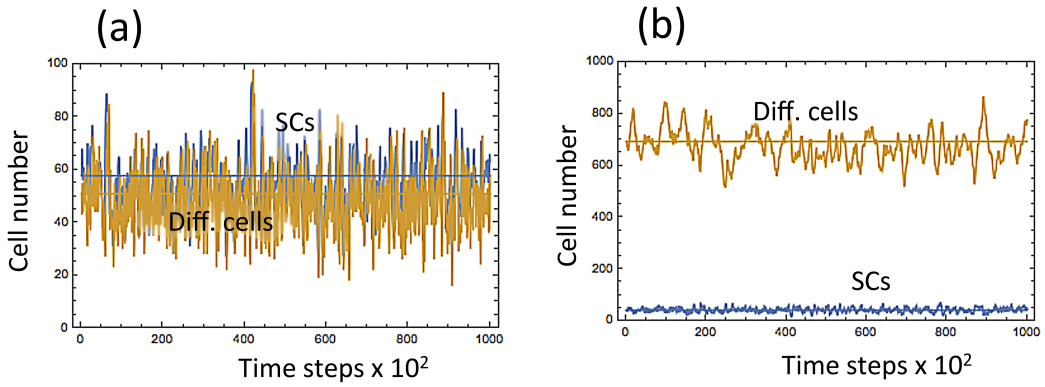


Figure 3.3: Typical numerical simulations of cell dynamics. (a) System (3.1) with $\epsilon = 0.005$ and $S_* = S = 0.5$; (b) system (3.6) with $\epsilon = 0.005$ and $S_* = S = 0.8$. Simulations are run for $2 \cdot 10^5$ time steps.

By equations (3.22), we can obtain the means and the variances of the system:

$$E[I] = i_0, \tag{3.2}$$

$$E[J] = j_0, \tag{3.3}$$

$$Var[I] = \frac{2L_*S_*\Delta + q_y^2 + 8L_*^2S_*p_y^2}{4B\Delta}, \tag{3.4}$$

$$Var[J] = \frac{2L_*(2+S_*)\Delta + q_x^2}{4B\Delta}, \tag{3.5}$$

where all the partial derivatives are evaluated at (i_0, j_0), and $L_* = 1/2$, $\Delta = q_x p_y$, $B = -2L_*S_*p_y - q_y$.

For each fixed pair (ϵ, S_*), we ran numerical simulations starting at the (rounded up) expected values of the cell population given above, and finishing either when the number of time-steps reached $2 \cdot 10^5$, or if any of the cell types went extinct. We then computed the means and the variances of the cell population over the time-course of each simulation. A typical run for a particular parameter set is presented in figure 3.3(a). In other simulations, both ϵ and S_* varied between 10^{-3} and 10^0.

From figure 3.4, we observe that the theoretical results for the means and the variances show

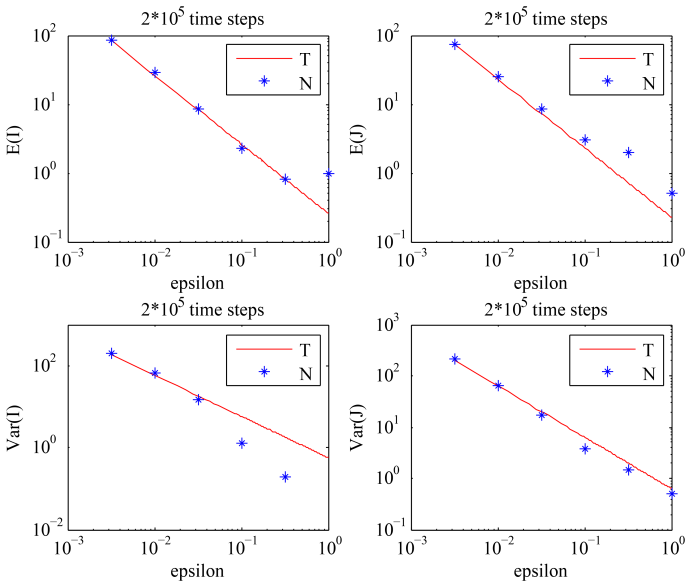


Figure 3.4: The behavior of the means and the variances of the cell population described by equations (3.1). The analytical results given by equation (3.2-3.5) (solid line) are compared with the values obtained by numerical simulations (stars), for different values of ϵ with the fixed value of S: $S = 0.5$. The choice of S should satisfy: $S = S_* < S_c$, where S_c is given by (3.27). ('T') stands for the theoretical results, and ('N') stands for the numerical results.

a good agreement with the numerical results for smaller values of ϵ, which is what we expect. We further observe that the means and the variances of the cell population decrease as the value of ϵ increases, exactly as predicted by equations (3.2-3.5). From figure 3.5, we can see that the variances of the cell population increase as the value of S increases, which is consistent with the analytical results given by (3.32).

Model #5. As the second example, we consider three control model #5 in figure 3.2, which is characterized by positive control of differentiation and division. We will assign the following equations to the probability and rate functions:

$$\begin{aligned} L_{I,J} &= L(\epsilon I, \epsilon J) = \frac{2\tanh(\epsilon I)}{2\tanh(\epsilon I) + 0.4}, \; P_{I,J} = P(\epsilon I, \epsilon J) = \tanh(\epsilon I + 0.1\epsilon J), \\ D_{I,J} &= D(\epsilon I, \epsilon J) = 1 - L_{I,J}, \quad S_{I,J} = c. \end{aligned} \tag{3.6}$$

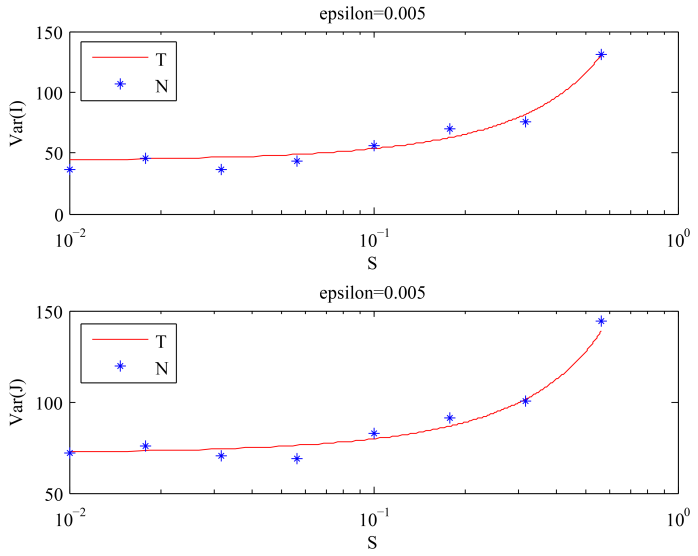


Figure 3.5: The behavior of the variances of the cell population described by equations (3.1) with $\epsilon = 0.005$, for different values of S. The analytical results given by equations (3.2-3.5) (solid line) are compared with the values obtained by numerical simulations (stars). ('T') stands for the theoretical results, and ('N') stands for the numerical results.

A typical stochastic simulation of system (3.6) for a particular parameter set is presented in figure 3.3(b). To calculate the variances, we find $p_x = \epsilon\,\mathrm{sech}^2(x + 0.1y) > 0$, $p_y = \epsilon 0.1\,\mathrm{sech}^2(x + 0.1y) > 0$, $q_x = \epsilon 1.6\,\mathrm{sech}^2(x) \cdot (2\tanh(x) + 0.4)^{-2} > 0$, $q_y = 0$, and hence $p_x > p_y > 0$. The steady state of the system is

$$i_0 = \frac{\log 1.5}{2\epsilon}, \quad j_0 = \frac{\log 3 - \log 1.5}{0.2\epsilon}.$$

By equations (3.22), we can obtain the means and the variances of the system:

$$E[I] = i_0, \tag{3.7}$$

$$E[J] = j_0, \tag{3.8}$$

$$Var[I] = \frac{2L_*S_*\Delta + 8L_*^2S_*p_y^2}{4B\Delta}, \tag{3.9}$$

$$Var[J] = \frac{2L_*(2+S_*)\Delta + q_x^2 + 8L_*^2S_*p_x^2}{4B\Delta}, \tag{3.10}$$

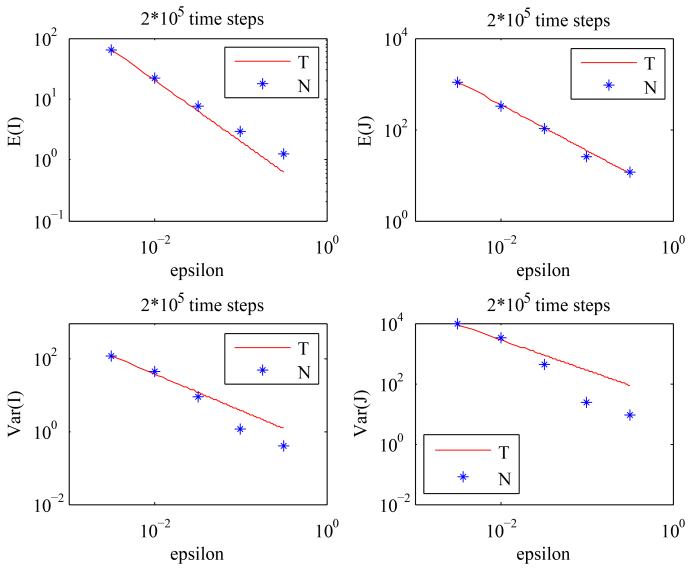


Figure 3.6: The behavior of the means and the variances of the cell population described by equations (3.6). The analytical results given by equation (3.7-3.10) (solid line) are compared with the values obtained by numerical simulations (stars), for different values of ϵ with the fixed value of S: $S = 0.5$. The choice of S should satisfy: $S > S_c = 0$ in this case, where S_c is given by (3.27). ('T') stands for the theoretical results, and ('N') stands for the numerical results.

where all the partial derivatives are evaluated at (i_0, j_0), and $L_* = 1/2$, $\Delta = q_x p_y$, $B = 2L_* S_* (p_x - p_y)$. Note that S_* in equation (3.9) cancels out.

We used the same numerical scheme as in the previous example. As observed in figure 3.6, the theoretical results are again in good agreement with the numerical results for smaller values of ϵ. The means and the variances of the cell population decrease as the value of ϵ increases, which is foretold by equations (3.7-3.10). From figure 3.7, we observe that $Var[I]$ stays approximately constant, whereas $Var[J]$ decreases as S increases. The first two points in figure 3.7 that appear to be inconsistent with the theory are explained by the analysis of the purely asymmetric divisions solution for smaller values of S, see Appendix H. The results of both numerical experiments are summarized in figure 3.8, which shows that:

- Increasing the fraction of symmetric division destabilizes the system given by equations (3.1);

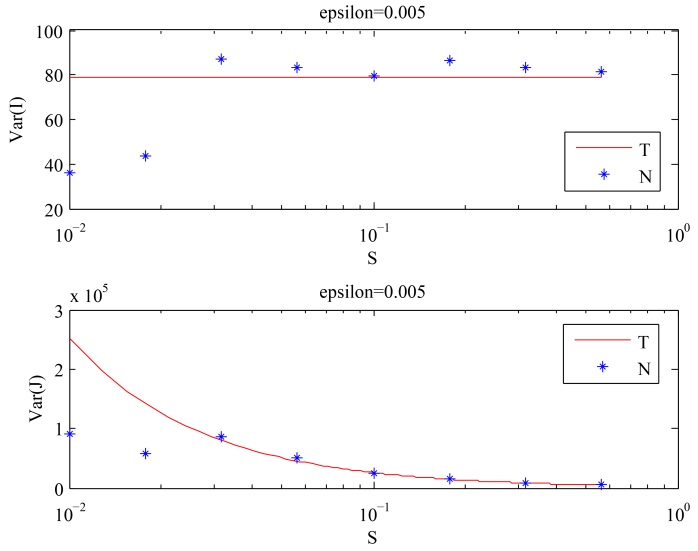


Figure 3.7: The behavior of the variances of the cell population described by equations (3.6) with $\epsilon = 0.005$ and time steps $= 2 \cdot 10^6$, for different values of S. The analytical results given by equations (3.7-3.10) (solid line) are compared with the values obtained by numerical simulations (stars). ('T') stands for the theoretical results, and ('N') stands for the numerical results.

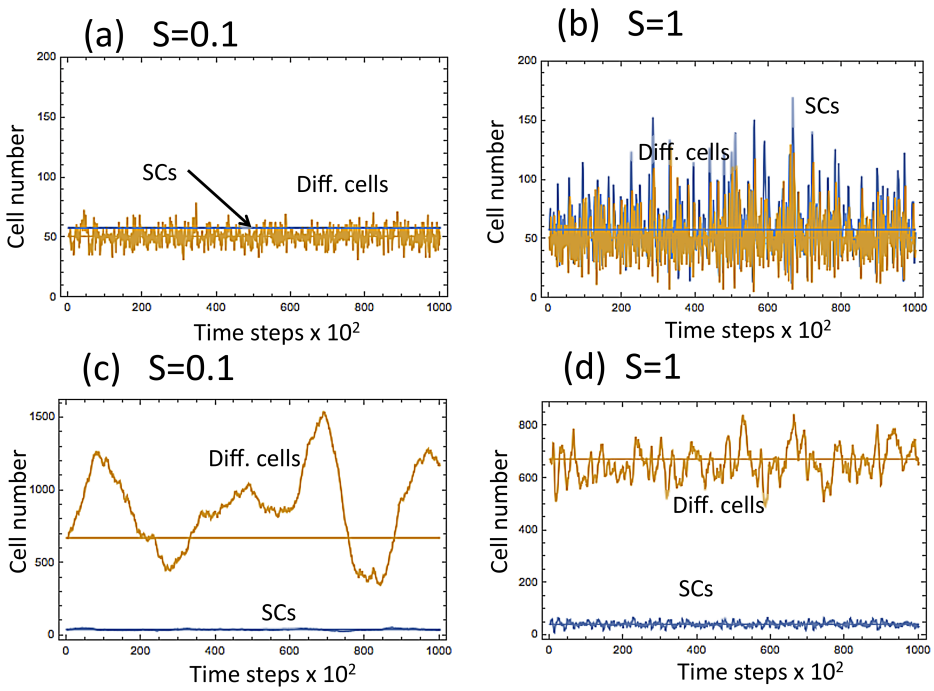


Figure 3.8: The behavior of the two systems described by equations (3.1) and (3.6) with $S = 0.1$ and $S = 1$. The top two diagrams, (a) and (b) correspond to the first example, and the bottom two diagrams, (c) and (d), correspond to the second example. In Panels (a) and (c), $S = 0.1$ (mostly asymmetric divisions). In (b) and (d), $S = 1$ (symmetric divisions).

- Increasing the fraction of asymmetric division destabilizes the system given by equations (3.6).

3.2 Methods

3.2.1 Stochastic model formulation

A stochastic model of cell population renewal is considered (see Chapter 2). The cells are subject to the following changes in a Poisson process with an infinitesimally small time-increment, Δt:

- With probability $L_{I,J}\Delta t$ a SC divides. Divisions can be symmetric (with probability $S_{I,J}$) or asymmetric (with probability $1 - S_{I,J}$).
  - With probability $L_{I,J}S_{I,J}P_{I,J}\Delta t$ a SC differentiation takes place resulting in a creation of two differentiated cells, $(I, J) \to (I - 1, J + 2)$.
  - With probability $L_{I,J}S_{I,J}(1 - P_{I,J})\Delta t$ a SC proliferation takes place resulting in a creation of a SC, $(I, J) \to (I + 1, J)$.
  - With probability $L_{I,J}(1 - S_{I,J})\Delta t$ a SC undergoes an asymmetric division resulting in a creation of a differentiated cell, $(I, J) \to (I, J + 1)$.
- With probability $D_{I,J}\Delta t$, a differentiated cell dies, $(I, J) \to (I, J - 1)$.

A deterministic model that captures these events can be expressed as the following system of

ordinary differential equations:

$$\dot{x} = LS(1-P) - LSP = LS(1-2P), \tag{3.11}$$

$$\dot{y} = 2LSP + L(1-S) - D, \tag{3.12}$$

where x and y refer to the numbers of stem and differentiated cells, and L, P, and S are all functions of x and y.

The stochastic description in terms of the Kolmogorov forward equation is given by the following equation for the variable $\varphi_{I,J}(t)$, the probability to find the system in state (I, J) at time t:

$$\begin{aligned}\dot{\varphi}_{I,J} &= \varphi_{I+1,J-2}L_{I+1,J-2}S_{I+1,J-2}P_{I+1,J-2} + \varphi_{I-1,J}L_{I-1,J}S_{I-1,J}(1-P_{I-1,J}) \\ &+ \varphi_{I,J-1}L_{I,J-1}(1-S_{I,J-1}) + \varphi_{I,J+1}D_{I,J+1} - \varphi_{I,J}(L_{I,J}+D_{I,J}),\end{aligned} \tag{3.13}$$

where the processes of the right hand side are presented in the same order as they appear in the list above. Note that system (3.11-3.12) is the "macroscopic law" obtained at the zeroth order of the "linear noise approximation", see equation (2.41).

We are interested in deriving equations for the mean values of the cell populations and their variances. To do this, we first define the steady states of the system, (i_0, j_0), by the following

equations (which are obtained by solving (3.11) and (3.12)) :

$$L_{i_0,j_0} = D_{i_0,j_0} = L_*, \quad P_{i_0,j_0} = \frac{1}{2}, \quad S_* = S_{i_0,j_0}. \tag{3.14}$$

(*mixed divisions steady state*)

$$L_{i_0,j_0} = D_{i_0,j_0} = L_*, \quad S_{i_0,j_0} = 0, \quad P_* = P_{i_0,j_0}. \tag{3.15}$$

(*purely asymmetric divisions steady state*)

Both equilibria are characterized by a balance between divisions and deaths (the first equation in (3.14) and (3.15)). In the first (mixed divisions) equilibrium, the probability of differentiation events is equal to the probability of proliferation events, thus ensuring that the expected change in the number of SCs is zero. The first two equations in (3.14) define the equilibrium population sizes i_0 and j_0. The fraction of symmetric divisions, $S_{I,J}$, does not influence the solution for i_0 and j_0, but, as shown below, can affect its stability properties and the size of fluctuations in the system.

The second (purely asymmetric) equilibrium is attained if the fraction of symmetric divisions can be made zero. The population sizes are determined by the first two equations in (3.15), and the probability of differentiations, formally defined by the last equation, becomes irrelevant at equilibrium. Below we focus on the mixed divisions steady state. Calculations pertaining to steady state (3.15) can be found in Appendix H.

3.2.2 Stability analysis and variance calculations

The methodology presented here is based on the assumption of weak dependencies of the functions $L_{I,J}$, $D_{I,J}$, etc on their variables. It is developed in [1] and justified rigorously in Chapter 2. Let us use the symbol $Z_{I,J}$ to denote any of the functions $L_{I,J}$, $P_{I,J}$, $D_{I,J}$, and $S_{I,J}$.

Suppose that we can represent the functions $Z_{I,J}$ near the equilibrium as $Z_{I,J} = Z(\epsilon I, \epsilon J)$, where the parameter $\epsilon \ll 1$ defines the weakness of the dependence. It is convenient to denote $x = \epsilon I,\ y = \epsilon J,\ i = I - i_0,\ j = J - j_0$, then we can expand the functions $Z_{I,J}$ around the steady state in Taylor series:

$$Z_{I,J} = Z_{i_0,j_0} + z_x i + z_y j + \frac{1}{2}(z_{xx} i^2 + z_{yy} j^2 + 2 z_{xy} ij) + \cdots , \tag{3.16}$$

where the subscripts x and y denote partial derivative of the function with respect to its argument, evaluated at (i_0, j_0), and $z_x = Z_x \epsilon, z_{xx} = Z_{xx}\epsilon^2$, etc. In this description, while constants $Z_x = O(1),\ Z_{xx} = O(1)$, etc are all of order one, all the first derivatives $z_x,\ z_y$ contain a factor ϵ, and all the second derivatives $z_{xx},\ z_{xy},\ z_{yy}$ contain a factor ϵ^2.

Define $\tilde{\varphi}_{i,j} = \varphi_{i+i_0,j+j_0} = \varphi_{I,J}$, and $\tilde{Z}_{i,j} = Z_{i+i_0,j+j_0} = Z_{I,J}$, then (3.13) can be reformulated as:

$$\begin{aligned} \dot{\tilde{\varphi}}_{i,j} &= \tilde{\varphi}_{i+1,j-2}\tilde{L}_{i+1,j-2}\tilde{S}_{i+1,j-2}\tilde{P}_{i+1,j-2} + \tilde{\varphi}_{i-1,j}\tilde{L}_{i-1,j}\tilde{S}_{i-1,j}(1 - \tilde{P}_{i-1,j}) \\ &+ \tilde{\varphi}_{i,j-1}\tilde{L}_{i,j-1}(1 - \tilde{S}_{i,j-1}) + \tilde{\varphi}_{i,j+1}\tilde{D}_{i,j+1} - \tilde{\varphi}_{i,j}(\tilde{L}_{i,j} + \tilde{D}_{i,j}). \end{aligned} \tag{3.17}$$

Using expansion (3.16) in equation (3.17), we can derive the moment equations for this system. In what follows, we use the following notations for the moments:

$$X_{\alpha\beta} = \sum_{i,j} i^\alpha j^\beta \tilde{\varphi}_{i,j}(t). \tag{3.18}$$

Multiplying equation (3.17) by i and by j, performing a summation in the two indices, and keeping only the highest order terms in ϵ, we obtain equations for the first moments in

steady-state:

$$
\begin{aligned}
0 &= -2L_*S_*(p_yX_{01} + p_xX_{10}), && (3.19)\\
0 &= (2L_*S_*p_x + q_x)X_{10} + (2L_*S_*p_y + q_y)X_{01}. && (3.20)
\end{aligned}
$$

For the second moments we have:

$$
\begin{aligned}
0 &= (S_*l_x + L_*s_x)X_{10} + (S_*l_y + L_*s_y)X_{01} - 4L_*S_*(p_yX_{11} + p_xX_{20}) + L_*S_*,\\
0 &= -(S_*l_x + L_*s_x + 2L_*S_*p_x)X_{10} - (S_*l_y + L_*s_y + 2L_*S_*p_y)X_{01}\\
&+ 2L_*S_*[p_xX_{20} - p_yX_{02} + (p_y - p_x)X_{11} - 1/2] + q_yX_{11} + q_xX_{20},\\
0 &= [S_*(4L_*p_x + l_x) + l_x + d_x + L_*s_x]X_{10} + [S_*(4L_*p_y + l_y) + l_y + d_y + L_*s_y]X_{01}\\
&+ (4L_*S_*p_x + 2q_x)X_{11} + (4L_*S_*p_y + 2q_y)X_{02} + L_*(2 + S_*). && (3.21)
\end{aligned}
$$

Solving this system, we can obtain the expressions for the means and variances: $E[I] = X_{10} + i_0 = i_0, E[J] = X_{01} + j_0 = j_0; Var[I] = X_{20} - X_{10}^2$, $Var[J] = X_{02} - X_{01}^2$. The highest order terms for the variances are given by

$$
Var[I] = \frac{K_x}{4B\Delta}, \quad Var[J] = \frac{K_y}{4B\Delta}, \tag{3.22}
$$

where we defined the quantities:

$$\Delta = q_x p_y - q_y p_x, \tag{3.23}$$

$$B = 2L_* S_* (p_x - p_y) - q_y, \tag{3.24}$$

$$K_x = 2L_* S_* \Delta + q_y^2 + 8L_*^2 S_* p_y^2, \tag{3.25}$$

$$K_y = 2L_*(2 + S_*)\Delta + q_x^2 + 8L_*^2 S_* p_x^2. \tag{3.26}$$

Details of stability analysis are given in Appendix H.2. It follows that mixed division steady state is stable as long as $\Delta > 0$ and $B > 0$; constants K_x and K_y are always positive quantities. Increasing Δ and B makes the system more robust by decreasing the variation of population sizes.

Notes. There are two important conclusions from the above analysis.

- The numbers of stem and differentiated cells at the equilibrium does not depend on the quantity S (the fraction of symmetric divisions). This is because regardless of the proportion of symmetric divisions, there are only two requirements for the constancy of the population: (1) Probability of differentiation under symmetric divisions is 1/2 (this keeps the number of stem cells constant), and (2) The rate of divisions equals to the rate of death (this keeps the number of differentiated cells constant). Both conditions are independent of S.

- Related to this, the fraction of symmetric divisions, S, only enters the expressions for the cell number variances. In the first order analysis above, only the equilibrium value, S_*, and not the derivatives, appear in the expressions for the second moments.

Below we explore how the probability of symmetric divisions affects the homeostatic control.

3.2.3 The role of asymmetric divisions in cell number regulation

The equilibrium values for the numbers of stem and differentiated cells are unaffected by the presence of asymmetric divisions, as illustrated by equations (3.14). On the other hand, the probability of symmetric divisions, S_*, can influence two important properties of the SC system: (a) stability of the equilibrium and (b) the size of fluctuations (the amount of variance), which is related to the robustness of homeostatic control.

Stability. The only way in which the fraction of symmetric divisions can influence stability of the system is by changing the sign of the quantity B, equation (3.24). From equation (3.24):

$$B > 0 \iff S_* > S_c \equiv \frac{q_y}{2L_*(p_x - p_y)} \tag{3.27}$$

If the value S_c is between 0 and 1, then we have the following trends:

- Increasing the fraction of asymmetric divisions can destabilize the system if $q_y < 2L_*(p_x - p_y)$ and $p_x > p_y$;
- Increasing the fraction of symmetric divisions can destabilize the system if $q_y < 0$ and $p_x < p_y$.

The size of fluctuations (robustness). In order to study the influence of asymmetric divisions on the behavior of cell populations, we consider the derivatives of the variances of I

and J with respect to parameter S_*:

$$\frac{dVar[I]}{dS_*} = \frac{L_*}{2B^2\Delta} p_y q_y (q_y - q_x - 4L_* p_y), \tag{3.28}$$

$$\frac{dVar[J]}{dS_*} = \frac{L_*}{2B^2\Delta} (p_x q_x - \Delta)(q_y - q_x - 4L_* p_y). \tag{3.29}$$

The signs of these derivatives can be different, depending on parameters. For a fixed set of parameters, the dependence on S_* is monotonic, that is, each of the variances either grows or decays with S_*.

Application to the five minimal controls.

- Two-control model #1, figure 3.2. In this case, $S_c = 0$ from (3.27), therefore the steady state is stable for any $S_* > 0$. Further, we have

$$\frac{dVar[I]}{dS_*} = 0, \quad \frac{dVar[J]}{dS_*} = \frac{L_*}{2B^2}(q_x + 4L_* p_y) < 0. \tag{3.30}$$

  In other words, increasing the share of symmetric divisions reduces the fluctuation size in the system. Thus, symmetric divisions (i.e. $S_* = 1$) will be optimal for this system.

- Two-control model #2, figure 3.2. In this case, $S_c < 0$ from (3.27), therefore the steady state is stable for any value of S_*. Further, we have

$$\frac{dVar[I]}{dS_*} = 0, \quad \frac{dVar[J]}{dS_*} = \frac{L_*}{2B^2}(-q_y) > 0. \tag{3.31}$$

  That is, increasing the share of asymmetric divisions makes the fluctuations smaller. Thus, purely asymmetric divisions (i.e. $S_* \to 0$) will be optimal.

- Three-control model #3, figure 3.2. In this case, we have

$$\frac{dVar[I]}{dS_*} > 0, \quad \frac{dVar[J]}{dS_*} > 0. \tag{3.32}$$

For the first three-control minimal system, $p_x < p_y$ and $q_y < 0$. From the viewpoint of stability, it is disadvantageous to increase the value of S_*. From the perspective of robustness, small values of S_* are best, since $Var[I]$ and $Var[J]$ grow with it. Thus, asymmetric divisions will be optimal.

- Three-control model #4, figure 3.2. In this case, we have $p_x > p_y$ and $q_y < 2L_*(p_x - p_y)$. Therefore, decreasing the value of S_* may destabilize the system. Further, we have

$$\frac{dVar[I]}{dS_*} < 0, \quad \frac{dVar[J]}{dS_*} < 0, \tag{3.33}$$

that is, fluctuations decay with S_*. Thus, symmetric divisions will be optimal.

- Three-control model #5, figure 3.2. Again, $p_x > p_y$ and $q_y < 2L_*(p_x - p_y)$, and

$$\frac{dVar[I]}{dS_*} = 0, \quad \frac{dVar[J]}{dS_*} < 0. \tag{3.34}$$

Therefore, as in the previous case, symmetric divisions will be optimal.

Chapter 4

Application in Colonic Crypts

In human colon crypts, the stem cell lineage is ordered with SCs at the bottom, DCs at the top, and TAC in bettween. In order to maintain the number of each cell type, the rate of removal of the DCs from the top is balanced by division and differentiation of the SCs and TACs below. As mentioned in chapter 1, we are interested in determining the most likely control network(s) given in figures 4.2 and 4.3 that govern the regulation of human colon crypt stem cell lineages. In this chapter, we determined the most likely regulatory network(s), using actual measurements of the number of SCs, TACs, and DCs in 49 colon crypts in human biopsy specimens. We examined each of the 32 possible networks in figures 4.2 and 4.3 to determine whether it can produce the measured means and variances of the cell population. In addition, we also used data on the dynamics of injury recovery, as well as experimentally obtained intra-crypt correlations. Using these criteria, a selection algorithm was devised that identified three of the 32 possible control networks as most likely the ones corresponding to the regulation of homeostasis of human colon crypts.

4.1 Materials and Methods

4.1.1 Data Description

In [58] Bravo and Axelrod measured the number and location of dividing cells (Ki-67 positively stained cells) and non-dividing cells (Ki-67 non-stained cells) in 49 colon crypts in human biopsy specimens. The non-dividing cells at the bottom of the crypt are considered quiescent stem cells, the non-dividing cells in the top two-thirds of the crypt are considered differentiated cells. The dividing cells near the bottom third of the crypt are considered to consist of transient amplifying cells and active stem cells [81]. The experimental details of the source of the specimens, measurement of each cell type, and determination of reliability of measurements, have previously been described [58].

For our model we need an approximate distribution of active cells into active stem cells and transit amplifying cells. This can be done by using experimental observations, as described below. [81] have reviewed evidence for the existence of quiescent stem cells at the bottom of the crypt and, in addition, of active stem cells among the dividing cells above the bottom. The existence of a quiescent stem cell population is consistent with the observation that mTert-expressing slowly cycling cells are resistant to intestinal injury and function in intestinal regeneration [82]. And the existence of an active stem cells population is consistent with the observation that rapidly cycling Lgr5+ cells are highly sensitive to intestinal damage [83]. We will denote the fraction of dividing cells that are active stem cells as W, and true fraction of active cells that are transient amplifying cells as $(1 - W)$. The value of W can be estimated using the following considerations.

Cells staining positive for various stem cell markers (Musashi-1, Bmi1, Lgr5, Lrig1) have been observed in the region of dividing cells in the small intestine of the mouse [83–86].

W, the proportion of active stem cells among all of the dividing cells in human colon crypts,

can be determined from data available about human colon crypts stained with the stem cell marker Musashi-1, and stained separately with the proliferating cell marker Ki-67.

The percentages of Musashi-1 positively staining cells at different positions in the human colon crypt were reported in [87] figure 4. According to this study, 69% of all positively staining cells are in positions 1-7 at the bottom of the crypt, and 31% of all positively staining cells are in positions 8 and above. [58] have reported, in Table 1, the number of Ki-67 positively and negatively staining cells at different positions of the human colon crypt. The average number of negatively staining cells in positions 1-7, is 35.7 ± 36.3 s.d. [1], and the average number of positively stained cells in positions 8 and above is 623.9 ± 234.1 s.d. The negatively stained cells at the bottom of the crypt are considered quiescent stem cells. The positively stained cells are considered to comprise all of the dividing cells, including both active stem cells and transient amplifying cells.

Table 4.1: Notations used to calculate W, the fraction of active SCs.

$\%SCq$	The percentage of all of the stem cells that are at the bottom of the crypt, e.g. quiescent stem cells
$\%SCa$	The percentage of all of the stem cells that are above the bottom of the crypt, e.g. active stem cells in the region of dividing cells
$\#SCq$	The number of quiescent stem cells at the bottom of the crypt
$\#SCa$	The number of active stem cells above the bottom of the crypt in the region of dividing cells
$\#Ki67+$	The total number of dividing cells

The probability distribution for the values of W, as well as the lower and upper bounds of W, can be determined. Using the notations defined in table 4.1, the number of active stem cells is given by

$$\#SCa = \#SCq \, \frac{\%SCa}{\%SCq},$$

[1] The considerable variation in the number of quiescent stem cells has previously been noted (Table 1, [58]). The method of measuring cell numbers and the measurement reliability has been described in detail in (Additional File 5, [58]). The experimental error in the measurement of the number of stem cells was determined by 49 repeated measurements of one crypt, C.V. = 7.7%. Since the experimental error in repeated measurements of one crypt was much less than the measured variation between 49 adjacent crypts, C.V.= 102%, it is likely that the measured variation between adjacent crypts is really indicative of a large variation between crypts, and not just due to experimental error.

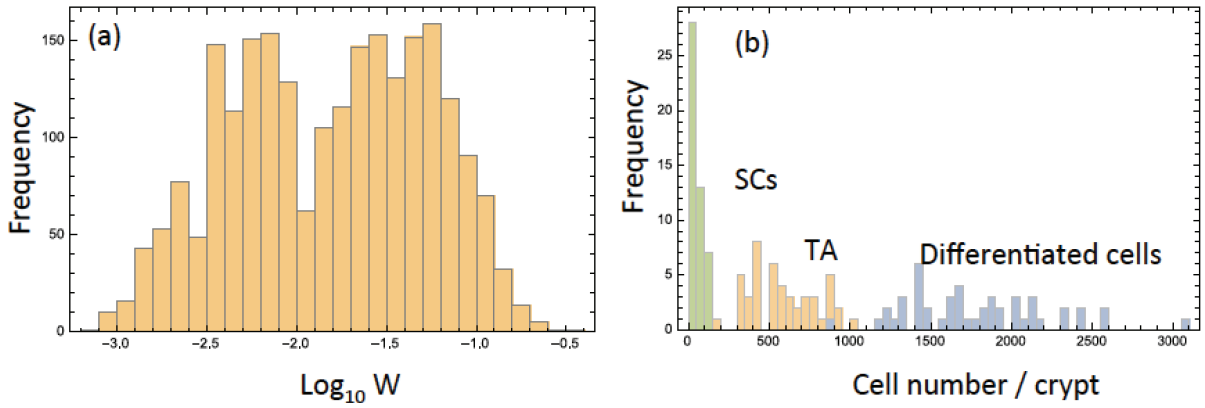


Figure 4.1: Estimating cell numbers. (a) Histogram of the experimentally calculated values of W, equation (4.1). (b) Histograms showing the distribution of cell numbers per crypt, by cell type, where $W = 0.03$ was assumed.

and the parameter W can be calculated as follows:

$$W = \frac{\#SCa}{\#Ki67+} = \frac{\#SCq}{\#Ki67+}\frac{\%SCa}{\%SCq}. \tag{4.1}$$

The first quotient in the right hand side of this expression can be calculated from the data of [58], and the second quotient from [87]. The probability distribution of the estimated values of W are shown in figure 4.1(a). It was approximated numerically using all the realizations of the cell numbers measured experimentally in [58]. The mean of this distribution corresponds to $W = 0.03$, the value used in the calculations presented here, unless otherwise noted. In figure 4.1(b) we show the frequency histograms of the three cell types obtained from the data by [58] using $W = 0.03$; again, this was created using all the experimentally obtained values of the cell numbers. We also investigated the effect of $W = 0$, as discussed in the last paragraph of section 4.2.

4.1.2 Stochastic model formulation

Consider a three-compartment model consisting of stem cells (SCs), transient amplifying cells (TACs), and differentiated cells (DCs). We will refer to the number of stem cells as I_1, the number of transient amplifying cells as I_2, and the number of differentiated cells as I_3. We assume only symmetric divisions of stem cells, see section 4.3, and employ a Poisson process to describe the dynamics (Poisson processes, and a related birth-death process, are conventionally used to describe cellular processes, see e.g. [88, 89]). The cells are subject to the following changes in a Poisson process with an infinitesimally small time-increment, Δt:

- With probability $L_1(I_1, I_2, I_3)\Delta t$ a stem cell (SC) divides.
  - With probability $L_1(I_1, I_2, I_3)P_1(I_1, I_2, I_3)\Delta t$ a SC differentiation takes place resulting in a creation of two transient amplifying cells (TACs), $(I_1, I_2, I_3) \rightarrow (I_1 - 1, I_2 + 2, I_3)$.
  - With probability $L_1(I_1, I_2, I_3)(1 - P_1(I_1, I_2, I_3))\Delta t$ a SC proliferation takes place in a creation of SC, $(I_1, I_2, I_3) \rightarrow (I_1 + 1, I_2, I_3)$.
- With probability $L_2(I_1, I_2, I_3)\Delta t$ a transient amplifying cell (TA) divides.
  - With probability $L_2(I_1, I_2, I_3)P_2(I_1, I_2, I_3)\Delta t$ a TA differentiation takes place resulting in a creation of two differentiated cells (DCs), $(I_1, I_2, I_3) \rightarrow (I_1, I_2 - 1, I_3 + 2)$.
  - With probability $L_2(I_1, I_2, I_3)(1 - P_2(I_1, I_2, I_3))\Delta t$ a TA proliferation takes place resulting in a creation of a TA, $(I_1, I_2, I_3) \rightarrow (I_1, I_2 + 1, I_3)$.
- With probability $D(I_1, I_2, I_3)\Delta t$, a differentiated cell dies, $(I_1, I_2, I_3) \rightarrow (I_1, I_2, I_3 - 1)$.

A deterministic model that captures these events can be expressed as the following system of

ordinary differential equations:

$$\dot{I}_1 = -L_1P_1 + L_1(1 - P_1), \tag{4.2}$$

$$\dot{I}_2 = 2L_1P_1 - L_2P_2 + L_2(1 - P_2), \tag{4.3}$$

$$\dot{I}_3 = 2L_2P_2 - D. \tag{4.4}$$

The equilibrium of this system, $(\bar{I}_1, \bar{I}_2, \bar{I}_3)$, can be obtained by solving equations (4.2) - (4.4) in steady state:

$$L_1(\bar{I}_1, \bar{I}_2, \bar{I}_3) = L_0, \quad D(\bar{I}_1, \bar{I}_2, \bar{I}_3) = D_0, \quad L_2(\bar{I}_1, \bar{I}_2, \bar{I}_3) = D_0 - L_0, \tag{4.5}$$

$$P_1(\bar{I}_1, \bar{I}_2, \bar{I}_3) = \frac{1}{2}, \quad P_2(\bar{I}_1, \bar{I}_2, \bar{I}_3) = \frac{D_0}{2(D_0 - L_0)}. \tag{4.6}$$

4.1.3 Stochastic analysis

There are five distinct processes that can take place in this system: differentiation divisions of SCs (Q_1), proliferation divisions of SCs (Q_2), differentiation divisions of TACs (Q_3), proliferation divisions of TACs (Q_4), and death (Q_5). The rates of these processes are given by:

$$Q_1 = L_1P_1; Q_2 = L_1(1 - P_1); Q_3 = L_2P_2; Q_4 = L_2(1 - P_2); Q_5 = D. \tag{4.7}$$

The stochastic description in terms of the Kolmogorov forward equation is given by the following equation for the variable φ_{I_1,I_2,I_3}, the probability to find the system in state (I_1, I_2, I_3)

at time t:

$$\begin{aligned}
\dot{\varphi}_{I_1,I_2,I_3} &= \varphi_{I_1+1,I_2-2,I_3} * Q_1(I_1+1, I_2-2, I_3) + \varphi_{I_1-1,I_2,I_3} * Q_2(I_1-1, I_2, I_3) \\
&+ \varphi_{I_1,I_2+1,I_3-2} * Q_3(I_1, I_2+1, I_3-2) + \varphi_{I_1,I_2-1,I_3} * Q_4(I_1, I_2-1, I_3) \\
&+ \varphi_{I_1,I_2,I_3+1} * Q_5(I_1, I_2, I_3+1) - \varphi_{I_1,I_2,I_3} * \sum_{n=1}^{5} Q_n(I_1, I_2, I_3), \qquad (4.8)
\end{aligned}$$

where the processes of the right hand side are presented in the same order as they appear in section 4.1.2.

The methodology presented here was developed in Chapter 2 and [1], and is related to the linear noise approximation of [76]. A detailed derivation and justification can be found in [90]. Let us use the symbol H_{I_1,I_2,I_3} to denote any of the functions $L_1(I_1, I_2, I_3), L_2(I_1, I_2, I_3)$, $P_1(I_1, I_2, I_3), P_2(I_1, I_2, I_3)$, and $D(I_1, I_2, I_3)$. Suppose that we can represent the functions H_{I_1,I_2,I_3} near the equilibrium as $H_{I_1,I_2,I_3} = H(\epsilon I_1, \epsilon I_2, \epsilon I_3)$, where the parameter $\epsilon \ll 1$ defines the weakness of the dependence of these rates on the populations that control them. Note that in this methodology, the peak of the probability distribution of the number of cells is assumed to be located near population sizes of the order $1/\epsilon$ and has a width of the order of $1/\epsilon^{1/2}$, see the derivation in [1]. While the validity of this approach has been studied extensively, see e.g. [77, 78], in our context it is important to note that typical fluctuations (of size $1/\epsilon^{1/2}$) must remain sufficiently small compared with the typical population size ($\sim 1/\epsilon$), such that the system will remain near the equilibrium and stochastic extinction is an unlikely event (for a time-duration which grows with $1/\epsilon$). These are conditions of homeostasis in a biological system; our approximation will technically break down outside the homeostatic regime.

It is convenient to denote the continuous variables

$$x_1 = \epsilon I_1, \quad x_2 = \epsilon I_2, \quad x_3 = \epsilon I_3,$$

and further shift the cell counts to be equal to zero at the equilibrium:

$$i_1 = I_1 - \bar{I}_1, i_2 = I_2 - \bar{I}_2, i_3 = I_3 - \bar{I}_3. \tag{4.9}$$

We can expand the functions H_{I_1,I_2,I_3} around the equilibrium $(\bar{I}_1, \bar{I}_2, \bar{I}_3)$ in Taylor series:

$$H_{I_1,I_2,I_3} = H_{\bar{I}_1,\bar{I}_2,\bar{I}_3} + H_{x_1}\epsilon i_1 + H_{x_2}\epsilon i_2 + H_{x_3}\epsilon i_3 + \ldots, \tag{4.10}$$

where the subscripts x_1, x_2, and x_3 denote the partial derivative of the function with respect to its argument, evaluated at $(\bar{I}_1, \bar{I}_2, \bar{I}_3)$.

To obtain the equations for the means and variances, we will follow the stochastic calculus of stem cells methodology developed in [90]. The stochastic processes defined in section 4.1.2 can be characterized by the following cell number changes:

Q_1: Differentiation of SCs, $\Delta_1 I_1 = -1, \Delta_1 I_2 = 2, \Delta_1 I_3 = 0,$

Q_2: Proliferation of SCs, $\Delta_2 I_1 = 1, \Delta_2 I_2 = 0, \Delta_2 I_3 = 0,$

Q_3: Differentiation of TACs , $\Delta_3 I_1 = 0, \Delta_3 I_2 = -1, \Delta_3 I_3 = 2,$

Q_4: Proliferation of TACs, $\Delta_4 I_1 = 0, \Delta_4 I_2 = 1, \Delta_4 I_3 = 0,$

Q_5: Death of DCs, $\Delta_5 I_1 = 0, \Delta_5 I_2 = 0, \Delta_5 I_3 = -1.$

Then equation (4.8) can be expressed as:

$$\begin{aligned}\dot{\varphi}_{I_1,I_2,I_3} &= \sum_{k=1}^{5} \varphi_{I_1-\Delta_k I_1, I_2-\Delta_k I_2, I_3-\Delta_k I_3} Q_k(I_1-\Delta_k I_1, I_2-\Delta_k I_2, I_3-\Delta_k I_3) \\ &- \varphi_{I_1,I_2,I_3} \sum_{k=1}^{5} Q_k(I_1,I_2,I_3). \end{aligned} \tag{4.11}$$

Define $\tilde{\varphi}_{i_m,i_n} = \varphi_{I_m,I_n}$, and $\tilde{Q}_k(i_m,i_n) = Q_k(I_m,I_n)$, then equation (4.11) can be rewritten as:

$$\begin{aligned}\dot{\tilde{\varphi}}_{i_1,i_2,i_3} &= \sum_{k=1}^{5} \tilde{\varphi}_{i_1-\Delta_k I_1, i_2-\Delta_k I_2, i_3-\Delta_k I_3} \tilde{Q}_k(i_1-\Delta_k I_1, i_2-\Delta_k I_2, i_3-\Delta_k I_3) \\ &- \tilde{\varphi}_{i_1,i_2,i_3} \sum_{k=1}^{5} \tilde{Q}_k(i_1,i_2,i_3). \end{aligned} \tag{4.12}$$

Here we use a standard technique to derive equations for the moments. Let us adopt the following notations for the first moments and the second moments of cell numbers:

$$y_m = \sum_{i_1,i_2,i_3} \tilde{\varphi}_{i_1,i_2,i_3} i_m, \quad y_{qp} = \sum_{i_1,i_2,i_3} \tilde{\varphi}_{i_1,i_2,i_3} i_q i_p.$$

We multiply both sides of Kolmogorov forward equation (4.12) by i_m and by $i_p i_q$, and sum over the indices i_1, i_2, i_3, to obtain:

$$\sum_{k=1}^{5} \sum_{i_1,i_2,i_3} \tilde{\varphi}_{i_1,i_2,i_3} \tilde{Q}_k(i_1,i_2,i_3)(i_m+\Delta_k I_m) - \sum_{k=1}^{5} \sum_{i_1,i_2,i_3} \tilde{\varphi}_{i_1,i_2,i_3} \tilde{Q}_k(i_1,i_2,i_3) i_m = 0, \tag{4.13}$$

$$\sum_{k=1}^{5} \sum_{i_1,i_2,i_3} \tilde{\varphi}_{i_1,i_2,i_3} \tilde{Q}_k(i_1,i_2,i_3)(i_p+\Delta_k I_p)(i_q+\Delta_k I_q) - \sum_{k=1}^{5} \sum_{i_1,i_2,i_3} \tilde{\varphi}_{i_1,i_2,i_3} \tilde{Q}_k(i_1,i_2,i_3) i_p i_q = 0, \tag{4.14}$$

where $m = 1,2,3$ and $p,q = 1,2,3$. The right hand side of the equations is zero because we

consider the equilibrium state and the time-derivatives in the Kolmogorov forward equation are zero.

Next, we use expansions (4.10) in equations (4.13-4.14), and truncate the expressions by keeping terms of order ϵ and ϵ^2 in equations for the first and second moments respectively. This results in the following moment equations of the cell numbers:

$$a_{m1}y_1 + a_{m2}y_2 + a_{m3}y_3 = 0, \quad m = 1, 2, 3, \tag{4.15}$$

$$\sum_{j=1}^{3} a_{pj}y_{jq} + \sum_{j=1}^{3} a_{qj}y_{pj} = -s_{pq}, \quad p, q = 1, 2, 3, \tag{4.16}$$

where $a_{mj} = \sum_{k=1}^{5} \frac{\partial Q_k}{\partial x_j}\Delta_k I_m$, $s_{pq} = \sum_{k=1}^{5} Q_{k*}\Delta_k I_p \Delta_k I_q$, and Q_{k*} is the equilibrium of Q_k.

Because of definition (4.9), the means y_m are all zero, and solving equations (4.16), we can obtain the expressions for the second moments, which are equal to the cell number variances,

$$Var[I_m] = y_{mm}, \quad m = 1, 2, 3.$$

4.2 Selection Algorithm

[1] has identified 20 different 3-compartment minimal control networks that are compatible with stable homeostatic control, see figure 4.2. These networks are characterized by constant death terms. In addition, there are 12 minimal control networks with non-constant death terms, see figure 4.3. All of these networks have exactly three controls (it was shown that this is the minimal number of controls compatible with stability), and include the 5 processes described in Section (4.1.2). We will use a selection algorithm to determine the most likely

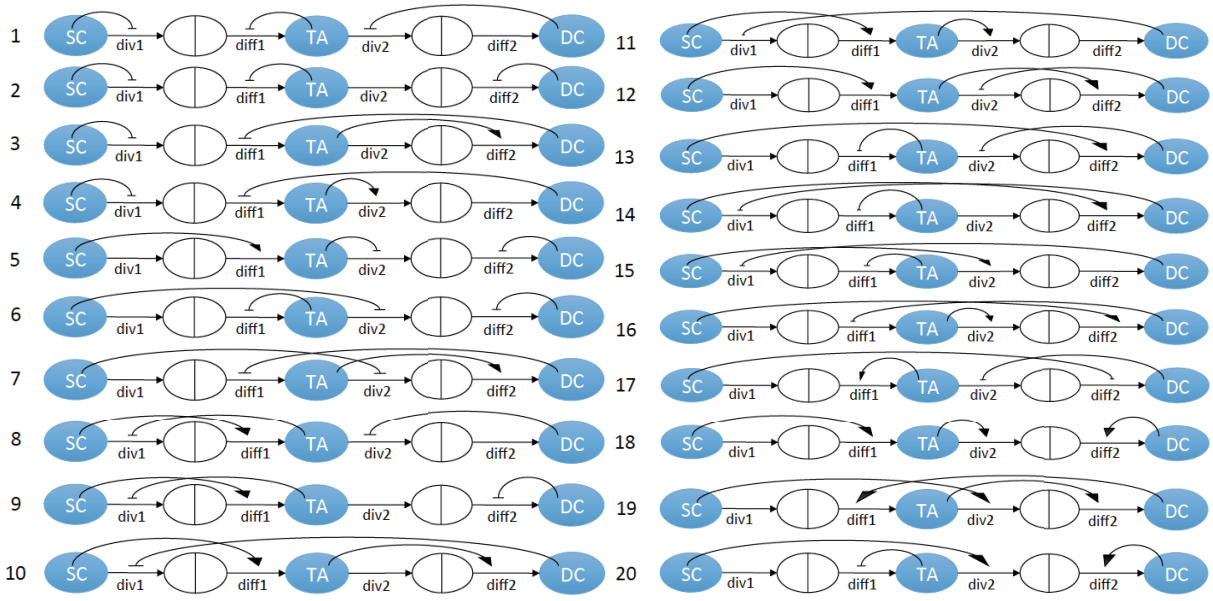


Figure 4.2: Twenty 3-compartment minimal control networks identified in [1], which are characterized with constant death rates. The three types of cells are marked by SC (stem cells), TA (transient amplifying cells), DC (differentiated cells). Horizontal arrows indicate the cell fate decisions: div1 and div2 the division process of SCs and TACs; diff1 and diff2 are the probability of the division to be a differentiation, as apposed to proliferation, for SCs and TACs respectively. The curved positive and negative arrows indicate control. The point of the arrow corresponds to the process that is being controlled, and the base of the arrow corresponds to the cell type controlling the process.

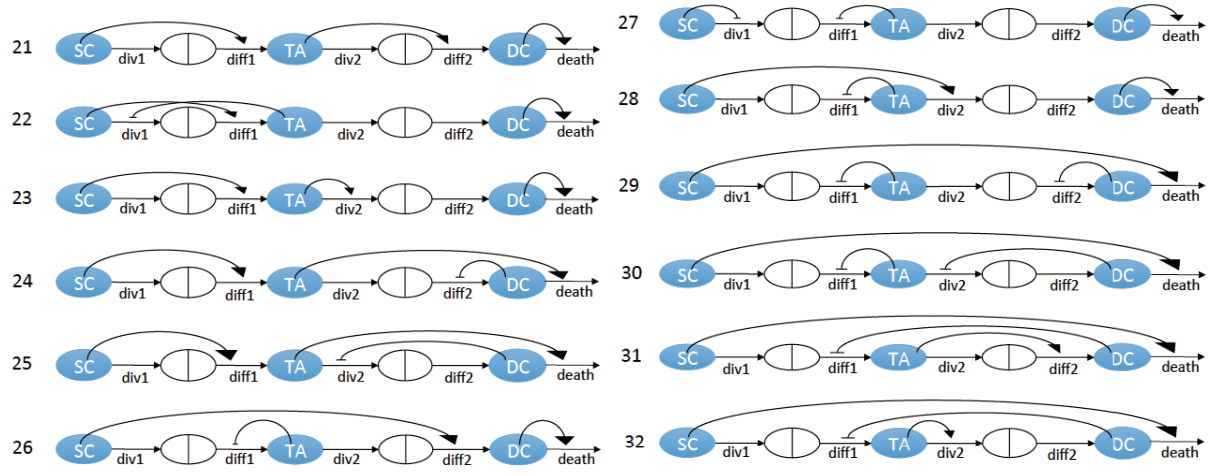


Figure 4.3: Twelve additional 3-compartment minimal control networks with non-constant death rates. Notations are as in figure 4.2.

minimal control network that matches the distribution of the measured data, see figure 4.4. This algorithm allows us to use biological criteria to exclude many of the possible control networks depicted in figures 4.2, 4.3. It uses the data generated for the distribution of the cell numbers as well as other considerations from the literature. The algorithm is demonstrated here using the 20 constant death rate networks of figure 4.2. The 12 networks of figure 4.3 are considered in Appendix L. The following is the step-by-step procedure used.

Choose networks with local controls. Cell-cell communication may occur by direct mechanical contact or by dispersal of molecules from a source cell. Mechanisms for the dispersal of molecules include transport through cell membranes, extracellular Brownian motion, or transport on the outer cell surface [91]. In each situation the effect of one cell is greatest on an adjacent cell and decreases on cells further away. Therefore, we assume that SCs can only control SCs and TACs; that TACs can only control TACs, SCs, and DCs; and that DCs can only control DCs and TACs. It follows that 11 out of the 20 control networks have local controls, see figure 4.4. We used figure 4.2 to select local controls, and the results are independent of the values of W.

Choose networks with stable solution and measured means and variances. Let us assume that all the control functions are linear (or consider the linearization of nonlinear controls, see Appendix J), and the death rate of DCs is constant in our analysis (this assumption is relaxed in Appendix L). Using equations (4.5-4.6) as constraints, we define

linearized control functions as

$$
\begin{aligned}
L_1 &= L_0\left(1 + a_{L_1} i_1 + b_{L_1} i_2 + c_{L_1} i_3\right), && (4.17)\\
P_1 &= \frac{1}{2}\left(1 + a_{P_1} i_1 + b_{P_1} i_2 + c_{P_1} i_3\right), && (4.18)\\
L_2 &= L_0\left(\frac{D_0}{L_0} - 1\right)\left(1 + a_{L_2} i_1 + b_{L_2} i_2 + c_{L_2} i_3\right), && (4.19)\\
P_2 &= \frac{1}{2(1 - L_0/D_0)}\left(1 + a_{P_2} i_1 + b_{P_2} i_2 + c_{P_2} i_3\right), && (4.20)\\
D &= D_0, && (4.21)
\end{aligned}
$$

where coefficients a, b, and c with the appropriate subscripts are constants. Functions (4.17-4.20) are the most general linear functions compatible with identities (4.5-4.6). We however are interested in "minimal controls", which is a restricted subset of such functions. It was shown in [1] that for stability of a three-compartment system, it is necessary to have at least three control loops, and all three populations must be involved in the control. There are exactly 20 systems with minimal control (that is, only 3 control loops) with constant death terms, see figure 4.2 (the non-constant death terms are included in networks of figure 4.3 and analyzed in Appendix L). Each of these control networks is characterized by exactly one nonzero coefficient a, one nonzero coefficient b, and one nonzero coefficient c in system (4.17-4.20). For example, the topmost network in the left column of figure 4.2 contains control of SC divisions by SCs, control of SC differentiation probabilities by TACs, and control of TAC divisions by DCs. This means that the corresponding linear system of controls, equations (4.17-4.20), contains nonzero coefficients

$$a_{L_1}, \quad b_{P_1}, \quad c_{L_2},$$

with the rest of coefficients being zero. For each minimal control network, there are 5 unknown constants: the equilibrium values L_0, D_0, and the nonzero controls (a, b, c). Let us denote

$$q = L_0/D_0,$$

where q is the ratio between the division rate of the SCs and the death rate of DCs at equilibrium. Equations (4.5) and (4.6) imply that $q \in (0, 1/2)$ (since the probability $P_2 \leq 1$). Further, by rescaling the time unit, we can set $D_0 = 1$. Therefore, only four unknown coefficients remain:

$$a, b, c, q.$$

Let us use definitions (4.17-4.20) and solve the linear algebraic system of equations given by (4.16). In particular, we can obtain the expressions for the three variances y_{11}, y_{22}, and y_{33}. For each control network, these expressions depend on the unknowns q, a, b, and c.

Using the data on the numbers of dividing and non-dividing cells for a given value of W, we can compute the numerical distributions of SCs, TACs, and DCs, and measure their means and variances. In this study we will focus on the case: $W = 0.03$ (most of the dividing cells are TACs).

Let us pick a control network, and also fix a q value; in our simulations we took $q = 0.1, 0.2, 0.4, 0.5$. For each q value, we have a system of equations

$$y_{11} = Var(I_1)_{exp}, \quad y_{22} = Var(I_2)_{exp}, \quad y_{33} = Var(I_3)_{exp},$$

where the left hand sides are functions of coefficients (a, b, c) (under fixed control network and the q value), and the right hand sides are numerically measured values of the cell number variances (under the fixed value of W, the fraction of active SCs among all dividing cells). This linear system of three equations with three unknowns can be solved to find the unknown

controls a, b, c, and only the networks that have real stable solutions for at least one q value will be considered.

In conclusion, networks #6, 17, 20 do not give stable solutions, and therefore are eliminated. The other 8 listed in figure 4.4 give realistic stable solutions.

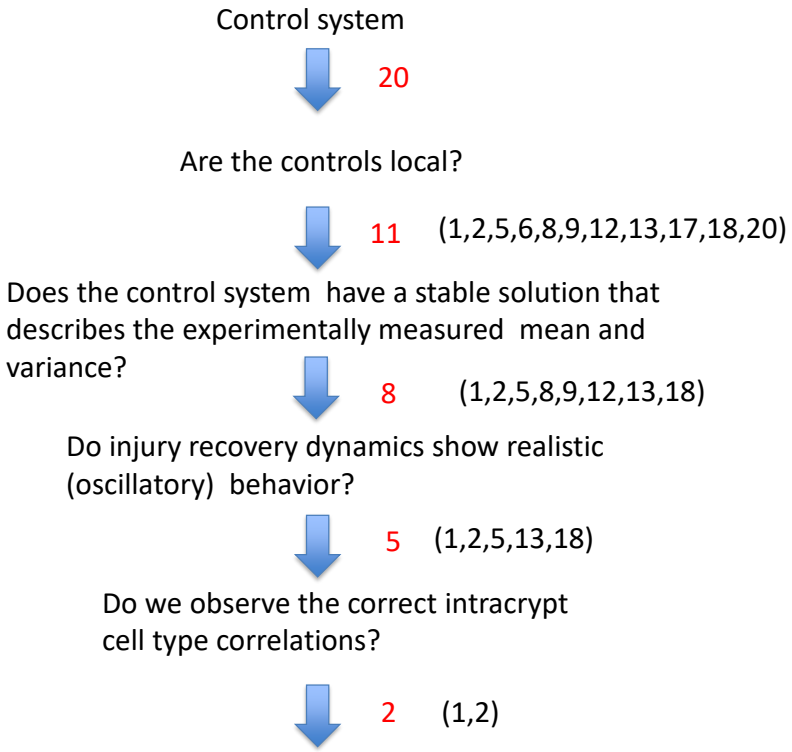


Figure 4.4: The outcome of the selection algorithm for $W = 0.03$.

To confirm the theoretical results, for each network, we then run numerical simulations with the coefficients obtained as described above. Note that the analysis presented here is local, in the sense that only the derivatives of the control at the equilibrium can be determined. We do not have any information on the global shapes of the control functions $L_1(I_1, I_2, I_3), L_2(I_1, I_2, I_3)$, etc. A numerical simulation requires further assumptions on the actual functional form for all the four control functions. The simplest way is to use linear functions, equations (4.17-4.20), for controls for all values of the arguments (cell numbers).

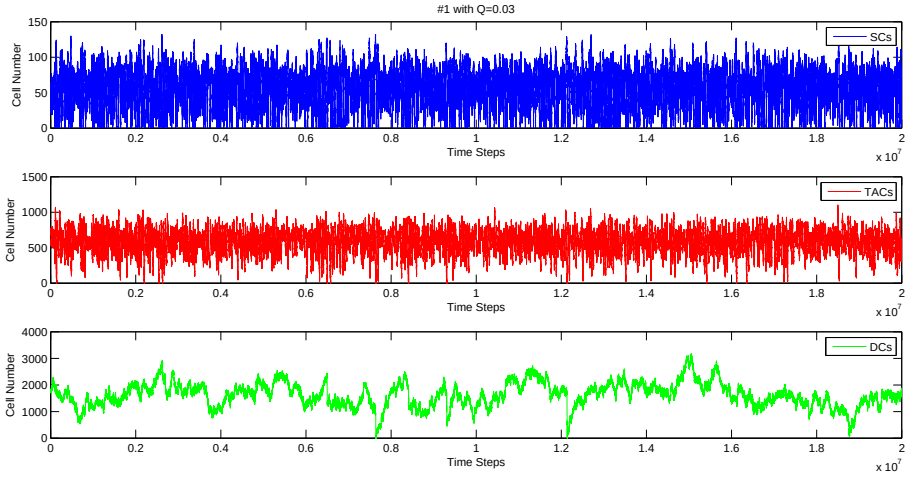


Figure 4.5: A typical simulation of network #1. Simulation starts at the experimentally measured means and finishes when the number of time steps reaches $2 \cdot 10^7$. Here we used $q = 0.1$ and $W = 0.03$; a set of control coefficients which produces means and variances similar to those measured in human crypts was determined by solving system (4.16). For the nonlinear control function see Appendix J.

This assumption works for many control networks, but sometimes it was observed that stochastic deviations of cell numbers from the mean forced the linear control function to take values outside the realistic range (e.g. a division rate may become negative). In such cases we used nonlinear functions which have the correct values of the derivatives at the equilibrium, but are defined in the biologically relevant range, see Appendix J. A typical simulation of the control dynamics for network #1 is presented in figure 4.5. The variance of the number of each cell type, is similar to the measured values reported in [58].

Choose networks with appropriate dynamics of recovery from perturbation.

Next, we perform the eigenvalue analysis of the networks to study the injury recovery dynamics. The recovery dynamics of the measured data are oscillatory, see e.g. [92]. Theoretically, we will study the eigenvalues of the linearized system around the equilibrium for each network. Using the deterministic equations (4.2)-(4.4), we can compute the Jacobian of each network (evaluating at the steady state) and its eigenvalues, thus obtain the condition of stability and osillatory behavior, see Appendix K for details. Complex eigenvalues indicate robust

oscillations. The results are listed as follows (see equations (4.17)-(4.20) for the notations):

- system 1: The system is always stable, and it is oscillatory if $a_{L_1} < \frac{b_{P_1}}{4}$.
- system 2: The system is always stable, and it is oscillatory if $a_{L_1} < \frac{b_{P_1}}{4}$.
- system 8: The system is always stable but not oscillatory.
- system 9: The system is always stable but not oscillatory.
- system 12: The system is always stable but not oscillatory.
- system 13: The system is always stable, and it is oscillatory if $a_{P_2} > -\frac{L_0 b_{P_1}}{4D_0}$.
- system 18: The system is stable if $b_{L_2} > \frac{D_0 c_{P_2}}{L_0}$, and it is oscillatory if $4b_{L_2}c_{P_2}D_0^2 > (L_0 b_{L_2} + D_0 c_{P_2})^2$.

From this analysis we conclude that networks #8, 9 and 12 do not have appropriate oscillatory dynamics and are therefore eliminated.

Choose networks with observed intra-crypt correlations of cell types. From the measurements, the sum of the number of SCs (total stem cells) and TACs is not correlated with the number of DCs, see figure 4.6(a). For each candidate network, we evaluate the absence of this correlation. We do not need to investigate correlations between TACs and SCs, because these numbers are strongly correlated due to the assumption that a fraction W of dividing cells is SCs and $(1 - W)$ is TACs. We use $W = 0.03$, from section 4.1.1.

From the previous step, the remaining networks are: $1, 2, 5, 13$, and 18. For each of these networks, we perform simulations of the dynamics in homeostatic conditions, figure 4.6(b-f). Each simulation starts at the experimentally measured mean and finishes when time steps reach $2 \cdot 10^7$. Each data point in figure 4.6 is collected every $4 \cdot 10^5$ time steps: the x-value is

the sum of SC and TAC population numbers, and y-value is the DC population number (that is, for each panel in figure 4.6, we plotted 50 data points). Using the statistical package R, we can check the correlation for each network: we fit a linear regression model of the simulated DCs against the sum of simulated SCs and TACs. We then perform hypothesis testing on the linear relation: suppose the model is expressed as $y = \alpha x + \beta$, then the null hypothesis is $H_0 : \alpha = 0$, and the alternative hypothesis is $H_a : \alpha \neq 0$. P-values of α can be obtained from R. A p-value less than 0.05 indicates that a linear correlation was unlikely due to chance, and a p-value greater than 0.05 indicates that the correlation could have been due to chance. The result is presented in figure 4.6, and p-values are given in the caption of the figure.

From this part of the analysis we conclude that networks #5, 13, and 18 have significant correlations between the number of DCs and the sum of SCs and TACs, unlike the observed data, and therefore are eliminated. However, for networks #1 and #2, the intra-crypt correlations are not significant, as are the observed data not significant, and they are retained.

Revisit injury recovery dynamics. Besides the eigenvalue analysis, we also look at the actual trajectories of the cell numbers following in injury. Injury recovery measurements are available in the literature, see e.g. [92], where the cell numbers in the mouse small intestine were measured following a perturbation of homeostasis of cell dynamics by a dose of radiation. Further, [93] shows very similar results, see Fig.6(b) in their paper that depicts oscillations of DNA synthesizing cells in mouse intestinal crypts after irradiation.

We observe that oscillation trajectories in recovery dynamics feature a certain overshoot followed by diminishing oscillations around the mean number of the cells. In particular, Fig.1(c) in [92] shows recovery oscillation of (clonogenic) stem cells, and Fig.1(d) in the same paper shows the recovery of total cells per crypt. Even though direct measurements of oscillatory crypt recovery dynamics are only available for murine crypts, we expect that similar behavior will be observed in human crypts, see also simulations of the human colon crypt

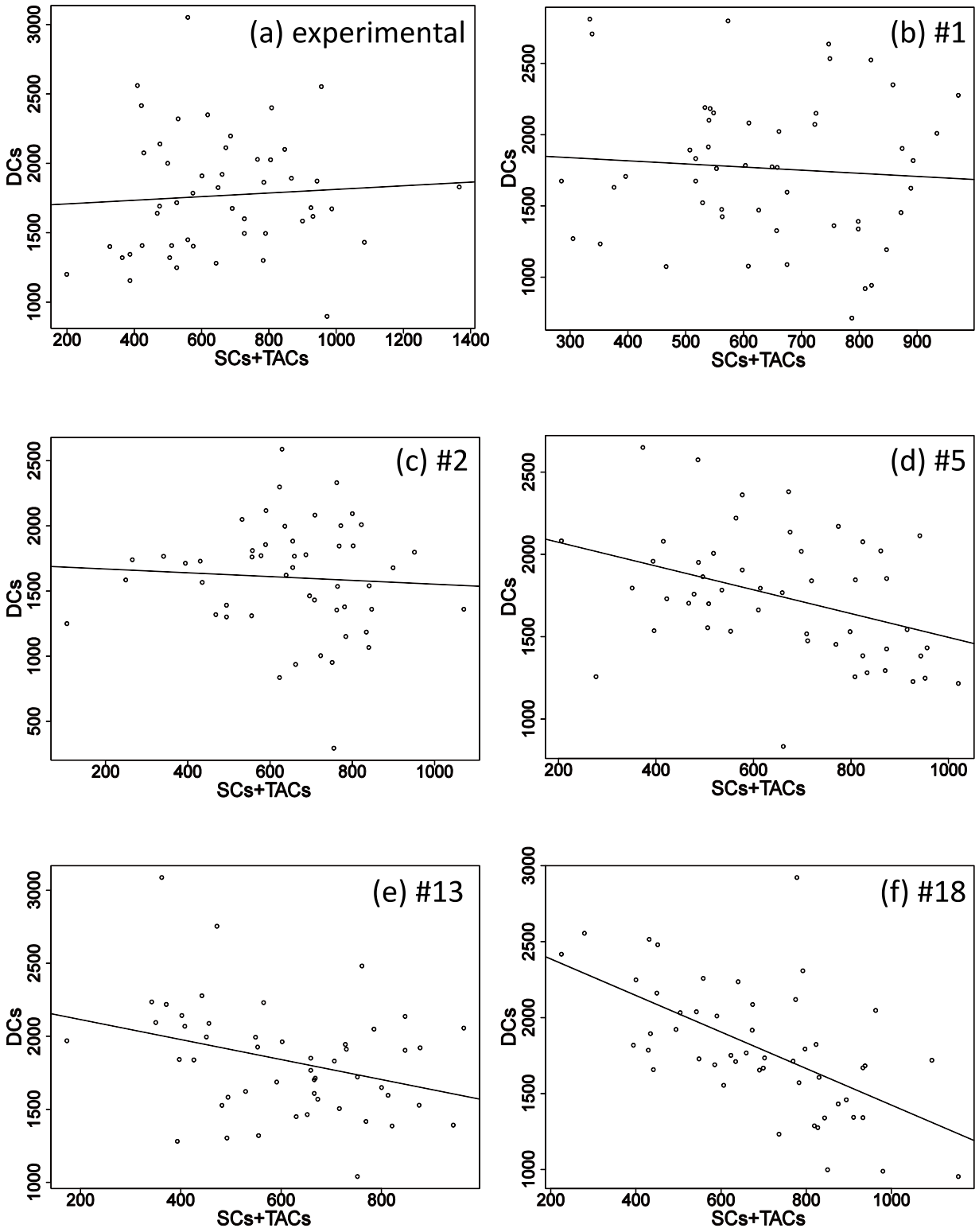


Figure 4.6: The plots of DCs vs. the sum of SCs + TACs. (a) is for the experimental data, and p-value = 0.64 (linear correlation coefficient); (b) is for network 1 when $q = 0.1$, and p-value = 0.598; (c) is for network 2 when $q = 0.1$, and p-value = 0.661; (d) is for network 5 when $q = 0.1$, and p-value = 0.006; (e) is for network 13 when $q = 0.2$, and p-value = 0.021; (f) is for network 18 when $q = 0.3$, and p-value = 4.89×10^{-6}. The straight lines indicate the fitted regression line for each. Please note that different values of q are used in the subfigures (a-f) because different networks describe the measured mean and variances for different subsets of the possible q values.

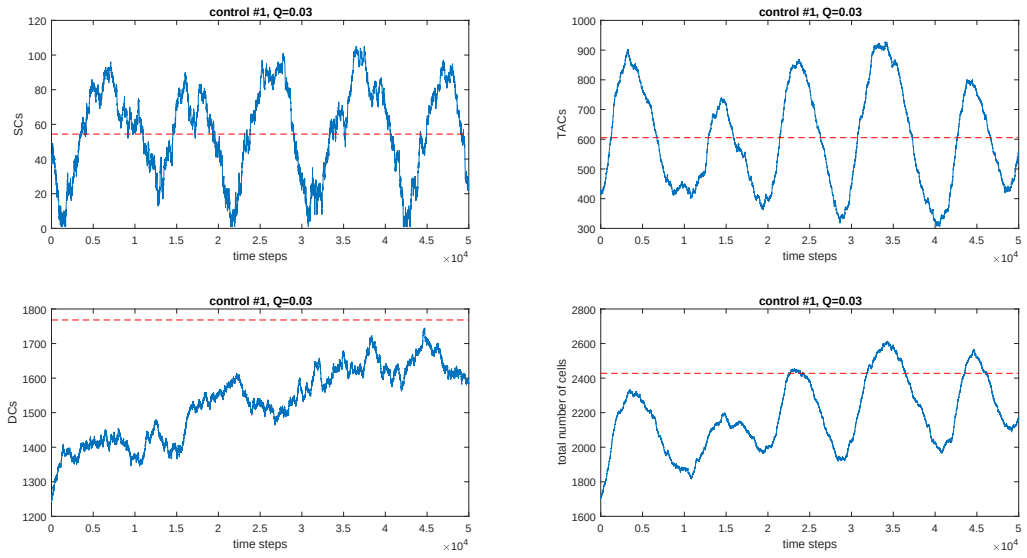


Figure 4.7: The simulated recovery trajectories of network #1 when $W = 0.03$ and $q = 0.1$. The red lines indicate the measured equilibrium values of cell numbers. The behavior of network 2 is qualitatively similar.

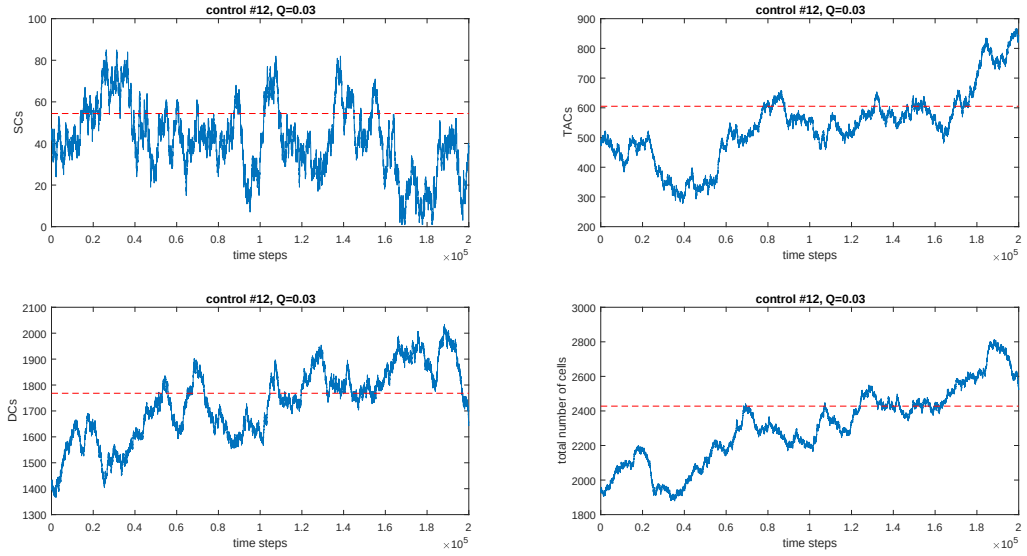


Figure 4.8: The simulated recovery trajectories of network #12 when $W = 0.03$ and $q = 0.1$. The red lines indicate the measured equilibrium values of cell numbers.

in [58], where Fig.5 shows oscillatory behavior similar to that of mouse crypts. Therefore, we argue that a reasonable control network should exhibit oscillatory behavior both in the numbers of stem cells and in the total number of cells.

In the previous stage of the algorithm, we have used the eigenvalue analysis to study the oscillatory behavior of each candidate network, and now we check if each remaining network produces a recovery trajectory that is qualitatively similar to the measurements. For each network, we start a numerical simulation at the state of equilibrium. We reduce the total cell number to 0.8 of the unperturbed value, run the simulation when the time steps reach 2×10^6, and then plot the number of cells over the time course. The numerical results for network #1 are presented in figure 4.7 (for network #2, the dynamics look very similar and are not shown). We observe that networks #1 and #2, each produce recovery dynamics of total cell numbers similar to the experimentally observed recovery dynamics reported in [92, 93]. In particular, both the number of SCs and the total number of cells are characterized by oscillatory recovery trajectories consistent with the experiments. In contrast with that, other (non-oscillatory) networks exhibit qualitatively different behavior, which is illustrated by the example of network #12, see figure 4.8.

A note on the variance of SCs. The calculated mean number of SCs in the crypts was only slightly larger than the standard deviation. As long as parameter W (the fraction of dividing cells that are active SC) was not too much lower, the system under the minimal controls studied here was able to maintain robust homeostasis. If we used $W = 0$, however, we observed that under the parameter values that produced the experimentally measured variance, the SCs were subject to relatively frequent extinction events. This observation allows several interpretations. (i) If we were to interpret the inter-crypt variation as an indicator of temporal intra-crypt variation, and assumed that $W = 0$, the minimal control networks studied here are not enough to explain the system behavior, and additional processes

such as TAC de-differentiation activated by SCs falling below a certain level would have to be included. (ii) Alternatively, it is possible that the relatively high inter-crypt variance of the SC numbers is a consequence of inter-crypt parameter variation, and the actual temporal homeostatic variability of crypts is lower than this. (iii) The value of W does not fall much below the measured mean of $W = 0.03$, in which case no further model modifications are necessary.

Summary of findings. Of the 20 possible constant-death control networks (figure 4.2), only networks #1 and #2 have stable dynamics that reproduce the measured mean and variance of each cell type, exhibit the correct intra-crypt correlation patterns, and show realistic oscillatory recovery dynamics. Further, as demonstrated in Appendix L, out of the 12 additional, non-constant death control networks (figure 4.3), only network #27 satisfies the same criteria.

4.3 Results and Discussion

Investigating possible regulation of stem cell dynamics in colon and intestinal crypts has been a popular subject for computational and mathematical modeling [94–97]. This is because crypts have a few distinguishable cell types organized in a hierarchy of fewer than 2500 cells that maintain homeostasis, and can recover after perturbation. It is the kind of dynamical system that lends itself to formulating possible regulatory models, and testing the model behaviors by comparing simulation results to experimental observations of real biological crypts.

In this study we investigated the 20 theoretically possible minimal control networks for a three-compartment system consisting of SCs, TACs, and DCs. We used the data obtained on 49 human colonic crypts, where the numbers of dividing and non-dividing cells were measured.

From this information assuming that the fraction W of all dividing cells were active SCs, and fraction $1 - W$ were TACs, we obtained the distributions of the three cell types. Using this information as well as observations of crypt recovery from injury and intra-crypt correlations, we devised an algorithm which allowed us to test all 20 networks. All but two were excluded based on their inconsistency with the measured data. In particular we found that control networks #1 and #2 are the best systems that describe the measured data, see figure 4.4.

A conceptually surprising outcome is that an argument about the interactions among the compartments of a stem cell lineage can be made based on only a very limited set of measurements, which does not contain any direct assessment of signaling mechanisms. The biological input consists of quantitative static measurements (the sample means and the variances of the cell numbers together with their correlations) and qualitative dynamic measurements (the existence of oscillations in tissue recovery process). Based on these pieces of evidence, and on our analysis of mathematically possible networks, we were able to restrict the number of possible regulatory networks to only two.

The resulting candidate networks #1 and #2 (figure 4.2) are among only three networks (among the 32 networks) that consist entirely of negative loops (network #6 is the third such network, and it was eliminated at the first step of the analysis because it failed to produce a stable root with the means and variances matching the observations, see figure 4.4). Network #27 is one of only two networks (among those where death of DCs is controlled, figure 4.3) that contain two negative control loop (non-local network #29 is the second one). In general, negative feedback controls are common in biological systems at many levels, from repressor protein effects on transcription of the lac operon in DNA to the effect of insulin on glucose in the blood. Negative controls have also been invoked to model other cell lineages [27, 48]. [98] emphasizes the important role that negative signaling loops play in oscillatory behavior in a wide variety of biological systems.

We further notice that the candidate control networks identified by our algorithm are all very

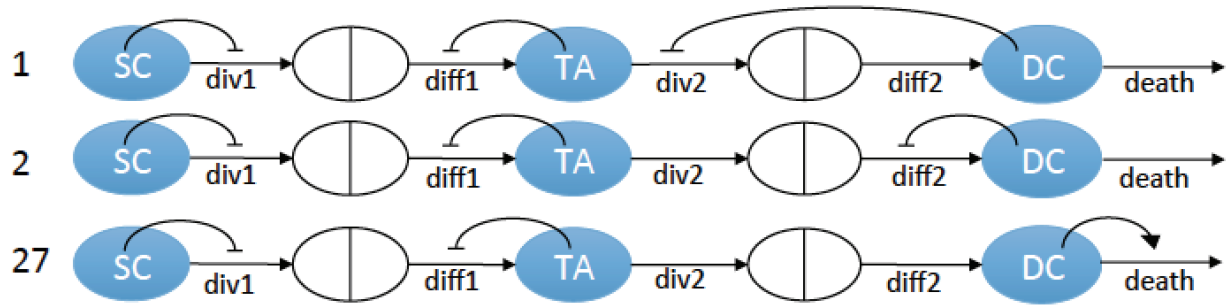


Figure 4.9: The three minimal control networks selected by our algorithm as candidate networks for colonuc crypt lineages.

similar: out of three control loops, two occur in all the three networks. These are: (1) the negative regulation of SC divisions rate by the cells in the SC compartment, and (2) the negative regulation of SC differentiation probability by the TAC compartment. These control loops have a simple explanation through the mechanism of crowding: having too many SCs prevents them from further divisions, and too many TACs restricts differentiation divisions in favor of proliferation divisions. The DCs in the three candidate networks perform different tasks: they exhibit negative control of divisions (#1) and proliferations (#2) of TACs and they regulate their own death (#27). Our current algorithm cannot distinguish between these three possibilities; further biological information may further narrow the set of possibilities.

Understanding the design principles of control networks is a fascinating research direction, to which this work can contribute. Given the results of our analysis, we hypothesize that perhaps the pattern of two negative control loops that occurs in all the three networks selected by our algorithm may be important for effective control of hierarchically organized tissues such as crypts.

This work should be considered more of a demonstration of principle than a final result. For example, we cannot conclude that one of the networks #1, #2 and #27 is definitely the one that acts in human colonic crypts. We have only tested the minimal (3-control) regulatory networks. It is still possible that a more complicated network with more than three essential control mechanisms is in place. All we can say is that we found the three simplest (in the

sense of having the minimal number of loops) control networks that are compatible with (both dynamic and static) observations in colonic crypts. The same holds for additional cellular processes that were not included in the present model, such as cell de-differentiation. The methods proposed here however can be extended to such systems, see [90]. The process of de-differentiation can be especially relevant for the system in question, as the measured inter-crypt standard deviation of the SC numbers is not much smaller than their mean, and in certain regimes it may be necessary to include TAC de-differentiation to compensate for stochastic loss of SCs.

In principle, SCs are capable of three types of division: (1) asymmetric divisions, where one of the daughter cells retains the stemness property while the other is more differentiated; (2) symmetric proliferation, where both offspring are stem cells, and (3) symmetric differentiation, where both offspring have higher degree of differentiation compared to the dividing cell. A large number of studies has been devoted to understanding the symmetry of SC divisions, and it appears that in some organisms SCs divide mostly asymmetrically, and in others both division types happen, depending on the specific context. The prevalence of symmetric divisions depends on the tissue. For example, in the mouse epidermis, it has been reported that about 20% of SC divisions are symmetric in the ear and tail epidermis, while 40% of SC divisions are symmetric in the paw epidermis. In the epidermis it has been argued in the recent years that SCs divide predominantly asymmetrically [19–22]. In crypts, however, SC symmetrical divisions play an important role [99–102]. Therefore, in the present work we have used the model with symmetric divisions (types (2) and (3) above). Again, asymmetric divisions can be added by using the present methodology, see Chapter 3.

In the present work the measurements of Ki-67, where the positively stained cells were identified as dividing, and the non-stained cells as non-dividing, were used to study cellular control networks. Different cell types (such as SCs, TACs, DCs) control various cell fate decisions. In order to convert the measurements into information on the numbers of SCs,

TACs, and DCs, we used the assumption that some stem cells are quiescent and others are active. [81] has reviewed evidence for this understanding of stem cells. They describe evidence for the co-existence of quiescent and active stem cell populations in the hair follicle, small intestine, and bone marrow. The model includes interconversion between the two types of stem cells, with quiescent stem cells replenishing damaged active stem cell population, and the possibility of active cells converting to quiescent stem cells. The number of quiescent cells are regulated by negative feedback from the active stem cells and/or their progeny. This model extends the previous models that have described each cell in the stem cell population as having a probability of not dividing, or dividing symmetrically to produce more stem cells or asymmetrically to produce a stem cell and a transient amplifying cell [103].

Our model only includes spatial considerations in the most rudimentary sense. We exclude the networks with nonlocal control, bearing in mind the geometry of the crypts. A more detailed, spatial model can be designed to test if our conclusions still hold. A limitation of such numerical studies is that complex spatial models do not allow for analytical solutions, and a comprehensive parameter study of the kind presented in our algorithm cannot be implemented. The advantage of our approach is that we obtained analytical expressions for the means and the variances of numbers of each cell type, and this allowed us to recover the values for the controls, given the experimentally obtained cell population measurements.

In conclusion, we have compared the simulated behavior of several possible regulatory networks with the numbers of cell types measured in human biopsy specimens, and have determined that the most likely form of regulation is by local control of stem cells on their own division, transient amplifying cells on the differentiation of stem cells, differentiated cells on the division and differentiation of transient amplifying cells and on their own death.

Bibliography

[1] Natalia L Komarova. Principles of regulation of self-renewing cell lineages. *PloS one*, 8(9):e72847, 2013.

[2] Cédric Blanpain and Elaine Fuchs. Plasticity of epithelial stem cells in tissue regeneration. *Science*, 344(6189):1242281, 2014.

[3] Nick Barker, Sina Bartfeld, and Hans Clevers. Tissue-resident adult stem cell populations of rapidly self-renewing organs. *Cell stem cell*, 7(6):656–670, 2010.

[4] Wei Chin Chou, Makoto Takeo, Piul Rabbani, Hai Hu, Wendy Lee, Young Rock Chung, John Carucci, Paul Overbeek, and Mayumi Ito. Direct migration of follicular melanocyte stem cells to the epidermis after wounding or uvb irradiation is dependent on mc1r signaling. *Nature medicine*, 19(7):924–929, 2013.

[5] Ken Inomata, Takahiro Aoto, Nguyen Thanh Binh, Natsuko Okamoto, Shintaro Tanimura, Tomohiko Wakayama, Shoichi Iseki, Eiji Hara, Takuji Masunaga, Hiroshi Shimizu, et al. Genotoxic stress abrogates renewal of melanocyte stem cells by triggering their differentiation. *Cell*, 137(6):1088–1099, 2009.

[6] Emi K Nishimura, Scott R Granter, and David E Fisher. Mechanisms of hair graying: incomplete melanocyte stem cell maintenance in the niche. *Science*, 307(5710):720–724, 2005.

[7] Matthew J Walter, Dong Shen, Li Ding, Jin Shao, Daniel C Koboldt, Ken Chen, David E Larson, Michael D McLellan, David Dooling, Rachel Abbott, et al. Clonal architecture of secondary acute myeloid leukemia. *New England Journal of Medicine*, 366(12):1090–1098, 2012.

[8] Li Ding, Timothy J Ley, David E Larson, Christopher A Miller, Daniel C Koboldt, John S Welch, Julie K Ritchey, Margaret A Young, Tamara Lamprecht, Michael D McLellan, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, 481(7382):506–510, 2012.

[9] Thomas Walenda, Thomas Stiehl, Hanna Braun, Julia Fröbel, Anthony D Ho, Thomas Schroeder, Tamme W Goecke, Björn Rath, Ulrich Germing, Anna Marciniak-Czochra, et al. Feedback signals in myelodysplastic syndromes: increased self-renewal of the malignant clone suppresses normal hematopoiesis. *PLoS Comput Biol*, 10(4):e1003599, 2014.

[10] Sean J Morrison and Judith Kimble. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature*, 441(7097):1068–1074, 2006.

[11] Shukry J Habib, Bi-Chang Chen, Feng-Chiao Tsai, Konstantinos Anastassiadis, Tobias Meyer, Eric Betzig, and Roel Nusse. A localized wnt signal orients asymmetric stem cell division in vitro. *Science*, 339(6126):1445–1448, 2013.

[12] Mayu Inaba and Yukiko M Yamashita. Asymmetric stem cell division: precision for robustness. *Cell stem cell*, 11(4):461–469, 2012.

[13] Yoh-ichi Kawabe, Yu Xin Wang, Iain W McKinnell, Mark T Bedford, and Michael A Rudnicki. Carm1 regulates pax7 transcriptional activity through mll1/2 recruitment during asymmetric satellite stem cell divisions. *Cell stem cell*, 11(3):333–345, 2012.

[14] Juergen A Knoblich. Mechanisms of asymmetric stem cell division. *Cell*, 132(4):583–597, 2008.

[15] Shihuan Kuang, Kazuki Kuroda, Fabien Le Grand, and Michael A Rudnicki. Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell*, 129(5):999–1010, 2007.

[16] Juergen A Knoblich. Asymmetric cell division during animal development. *Nature reviews Molecular cell biology*, 2(1):11–20, 2001.

[17] Fabien Le Grand, Andrew E Jones, Vanessa Seale, Anthony Scimè, and Michael A Rudnicki. Wnt7a activates the planar cell polarity pathway to drive the symmetric expansion of satellite stem cells. *Cell stem cell*, 4(6):535–547, 2009.

[18] Jessica Yingling, Yong Ha Youn, Dawn Darling, Kazuhito Toyo-Oka, Tiziano Pramparo, Shinji Hirotsune, and Anthony Wynshaw-Boris. Neuroepithelial stem cell proliferation requires lis1 for precise spindle orientation and symmetric division. *Cell*, 132(3):474–486, 2008.

[19] Xinhong Lim, Si Hui Tan, Winston Lian Chye Koh, Rosanna Man Wah Chau, Kelley S Yan, Calvin J Kuo, Renée van Amerongen, Allon Moshe Klein, and Roel Nusse. Interfollicular epidermal stem cells self-renew via autocrine wnt signaling. *Science*, 342(6163):1226–1230, 2013.

[20] Guilhem Mascré, Sophie Dekoninck, Benjamin Drogat, Khalil Kass Youssef, Sylvain Brohée, Panagiota A Sotiropoulou, Benjamin D Simons, and Cédric Blanpain. Distinct contribution of stem and progenitor cells to epidermal maintenance. *Nature*, 489(7415):257–262, 2012.

[21] David P Doupé, Allon M Klein, Benjamin D Simons, and Philip H Jones. The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. *Developmental cell*, 18(2):317–323, 2010.

[22] Elizabeth Clayton, David P Doupé, Allon M Klein, Douglas J Winton, Benjamin D Simons, and Philip H Jones. A single type of progenitor cell maintains normal epidermis. *Nature*, 446(7132):185–189, 2007.

[23] NG van Kampen. A power series expansion of the master equation. *Canadian Journal of Physics*, 39(4):551–567, 1961.

[24] Markus Loeffler and Ingo Roeder. Tissue stem cells: definition, plasticity, heterogeneity, self-organization and models–a conceptual approach. *Cells Tissues Organs*, 171(1):8–26, 2002.

[25] Mark d'Inverno, Neil Theise, Jane Prophet, et al. Mathematical modeling of stem cells: A complexity primer for the stem-cell biologist. *Tissue Stem Cells: Biology and Applications, 2nd edn. Marcell Dekker Inc: New York, NY*, pages 1–16, 2006.

[26] Ingo Roeder, Joerg Galle, and Markus Loeffler. Theoretical concepts of tissue stem cell organization. *Christopher S Potten*, 2006.

[27] Marc Mangel and Michael B Bonsall. Phenotypic evolutionary models in stem cell biology: replacement, quiescence, and variability. *PLoS one*, 3(2):e1591, 2008.

[28] Monika Joanna Piotrowska, Heiko Enderling, U van der Heiden, and Michael C Mackey. Mathematical modeling of stem cells related to cancer. *Complex Systems in Biomedicine*, 2008.

[29] IP Tomlinson and WF Bodmer. Failure of programmed cell death and differentiation as causes of tumors: some simple mathematical models. *Proceedings of the National Academy of Sciences*, 92(24):11130–11134, 1995.

[30] Alberto dOnofrio and Ian PM Tomlinson. A nonlinear mathematical model of cell turnover, differentiation and tumorigenesis in the intestinal crypt. *Journal of theoretical biology*, 244(3):367–374, 2007.

[31] Matthew D Johnston, Carina M Edwards, Walter F Bodmer, Philip K Maini, and S Jonathan Chapman. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proceedings of the National Academy of Sciences*, 104(10):4008–4013, 2007.

[32] Bruce M Boman, Jeremy Z Fields, Kenneth L Cavanaugh, Arthur Guetter, and Olaf A Runquist. How dysregulated colonic crypt dynamics cause stem cell overpopulation and initiate colon cancer. *Cancer Research*, 68(9):3304–3313, 2008.

[33] K Hardy and J Stark. Mathematical models of the balance between apoptosis and proliferation. *Apoptosis*, 7(4):373–381, 2002.

[34] Yasushi Yatabe, Simon Tavaré, and Darryl Shibata. Investigating stem cells in human colon by using methylation patterns. *Proceedings of the National Academy of Sciences*, 98(19):10839–10844, 2001.

[35] R Ganguly and IK Puri. Mathematical model for the cancer stem cell hypothesis. *Cell proliferation*, 39(1):3–14, 2006.

[36] R Ganguly and IK Puri. Mathematical model for chemotherapeutic drug efficacy in arresting tumour growth based on the cancer stem cell hypothesis. *Cell proliferation*, 40(3):338–354, 2007.

[37] Rina Ashkenazi, Sara N Gentry, and Trachette L Jackson. Pathways to tumorigenesismodeling mutation acquisition in stem cells and their progeny. *Neoplasia*, 10(11):1170–IN6, 2008.

[38] Franziska Michor. Mathematical models of cancer stem cells. *Journal of Clinical Oncology*, 26(17):2854–2861, 2008.

[39] Cristian Tomasetti and Doron Levy. Role of symmetric and asymmetric division of stem cells in developing drug resistance. *Proceedings of the National Academy of Sciences*, 107(39):16766–16771, 2010.

[40] Heiko Enderling and Philip Hahnfeldt. Cancer stem cells in solid tumors: Is evading apoptosisa hallmark of cancer? *Progress in biophysics and molecular biology*, 106(2):391–399, 2011.

[41] Thomas Stiehl and Anna Marciniak-Czochra. Mathematical modeling of leukemogenesis and cancer stem cell dynamics. *Mathematical Modelling of Natural Phenomena*, 7(1):166–202, 2012.

[42] Ingmar Glauche, Michael Cross, Markus Loeffler, and Ingo Roeder. Lineage specification of hematopoietic stem cells: mathematical modeling and biological implications. *Stem cells*, 25(7):1791–1799, 2007.

[43] Franziska Michor, Timothy P Hughes, Yoh Iwasa, Susan Branford, Neil P Shah, Charles L Sawyers, and Martin A Nowak. Dynamics of chronic myeloid leukaemia. *Nature*, 435(7046):1267–1270, 2005.

[44] Caroline Colijn and Michael C Mackey. A mathematical model of hematopoiesisi. periodic chronic myelogenous leukemia. *Journal of Theoretical Biology*, 237(2):117–132, 2005.

[45] Mostafa Adimya, Fabien Craustea, and Shigui Ruanb. Modelling hematopoiesis mediated by growth factors with applications to periodic hematological diseases. *Bulletin of mathematical biology*, 68(8):2321–2351, 2006.

[46] Anna Marciniak-Czochra, Thomas Stiehl, Anthony D Ho, Willi Jäger, and Wolfgang Wagner. Modeling of asymmetric cell division in hematopoietic stem cells-regulation of self-renewal is essential for efficient repopulation. *Stem cells and development*, 18(3):377–386, 2009.

[47] Arthur D Lander, Kimberly K Gokoffski, Frederic YM Wan, Qing Nie, and Anne L Calof. Cell lineages and the logic of proliferative control. *PLoS Biol*, 7(1):e1000015, 2009.

[48] Wing-Cheong Lo, Ching-Shan Chou, Kimberly K Gokoffski, Frederic Y-M Wan, Arthur D Lander, Anne L Calof, and Qing Nie. Feedback regulation in multistage cell lineages. *Mathematical biosciences and engineering: MBE*, 6(1):59, 2009.

[49] Yukihiko Nakata, Philipp Getto, Anna Marciniak-Czochra, and Tomás Alarcón. Stability analysis of multi-compartment models for cell production systems. *Journal of biological dynamics*, 6(sup1):2–18, 2012.

[50] Ovide Arino and Marek Kimmel. Stability analysis of models of cell production systems. *Mathematical Modelling*, 7(9):1269–1300, 1986.

[51] Thomas Stiehl and Anna Marciniak-Czochra. Characterization of stem cells using mathematical models of multistage cell lineages. *Mathematical and Computer Modelling*, 53(7):1505–1517, 2011.

[52] Heiko Enderling, Mark AJ Chaplain, Alexander RA Anderson, and Jayant S Vaidya. A mathematical model of breast cancer development, local treatment and recurrence. *Journal of theoretical biology*, 246(2):245–259, 2007.

[53] Heiko Enderling, Alexander RA Anderson, Mark AJ Chaplain, Afshin Beheshti, Lynn Hlatky, and Philip Hahnfeldt. Paradoxical dependencies of tumor dormancy and progression on basic cell kinetics. *Cancer research*, 69(22):8814–8821, 2009.

[54] Heiko Enderling, Derek Park, Lynn Hlatky, and Philip Hahnfeldt. The importance of spatial distribution of stemness and proliferation state in determining tumor radioresponse. *Mathematical Modelling of Natural Phenomena*, 4(3):117–133, 2009.

[55] H Enderling, L Hlatky, and P Hahnfeldt. Migration rules: tumours are conglomerates of self-metastases. *British journal of cancer*, 100(12):1917–1925, 2009.

[56] Zvia Agur, Yoaz Daniel, and Yuval Ginosar. The universal properties of stem cells as pinpointed by a simple discrete model. *Journal of mathematical biology*, 44(1):79–86, 2002.

[57] David Dingli, Arne Traulsen, and Jorge M Pacheco. Stochastic dynamics of hematopoietic tumor stem cells. *Cell cycle*, 6(4):461–466, 2007.

[58] Rafael Bravo and David E Axelrod. A calibrated agent-based computer model of stochastic cell dynamics in normal human colon crypts useful for in silico experiments. *Theoretical Biology and Medical Modelling*, 10(1):1, 2013.

[59] Arthur D Lander, Kimberly K Gokoffski, Frederic YM Wan, Qing Nie, and Anne L Calof. Cell lineages and the logic of proliferative control. *PLoS Biol*, 7(1):e1000015, 2009.

[60] H Youssefpour, X Li, AD Lander, and JS Lowengrub. Multispecies model of cell lineages and feedback control in solid tumors. *Journal of theoretical biology*, 304:39–59, 2012.

[61] Anna Konstorum, Thomas Hillen, and John Lowengrub. Feedback regulation in a cancer stem cell model can cause an allee effect. *Bulletin of mathematical biology*, 78(4):754–785, 2016.

[62] Sameeran Kunche, Huaming Yan, Anne L Calof, John S Lowengrub, and Arthur D Lander. Feedback, lineages and self-organizing morphogenesis. *PLoS Comput Biol*, 12(3):e1004814, 2016.

[63] Yasushi Yatabe, Simon Tavaré, and Darryl Shibata. Investigating stem cells in human colon by using methylation patterns. *Proceedings of the National Academy of Sciences*, 98(19):10839–10844, 2001.

[64] K Hardy and J Stark. Mathematical models of the balance between apoptosis and proliferation. *Apoptosis*, 7(4):373–381, 2002.

[65] Matthew D Johnston, Carina M Edwards, Walter F Bodmer, Philip K Maini, and S Jonathan Chapman. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proceedings of the National Academy of Sciences*, 104(10):4008–4013, 2007.

[66] Bruce M Boman, Jeremy Z Fields, Kenneth L Cavanaugh, Arthur Guetter, and Olaf A Runquist. How dysregulated colonic crypt dynamics cause stem cell overpopulation and initiate colon cancer. *Cancer Research*, 68(9):3304–3313, 2008.

[67] Rina Ashkenazi, Trachette L Jackson, Gabriela Dontu, and Max S Wicha. Breast cancer stem cells-research opportunities utilizing mathematical modeling. *Stem cell reviews*, 3(2):176–182, 2007.

[68] Ingmar Glauche, Michael Cross, Markus Loeffler, and Ingo Roeder. Lineage specification of hematopoietic stem cells: mathematical modeling and biological implications. *Stem cells*, 25(7):1791–1799, 2007.

[69] Jasmine Foo, Mark W Drummond, Bayard Clarkson, Tessa Holyoake, and Franziska Michor. Eradication of chronic myeloid leukemia stem cells: a novel mathematical model predicts no therapeutic benefit of adding g-csf to imatinib. *PLoS Comput Biol*, 5(9):e1000503, 2009.

[70] David Dingli and Franziska Michor. Successful therapy must eradicate cancer stem cells. *Stem cells*, 24(12):2603–2610, 2006.

[71] Matthew D Johnston, Philip K Maini, S Jonathan Chapman, Carina M Edwards, and Walter F Bodmer. On the proportion of cancer stem cells in a tumour. *Journal of theoretical biology*, 266(4):708–711, 2010.

[72] Thomas Hillen, Heiko Enderling, and Philip Hahnfeldt. The tumor growth paradox and immune system-mediated selection for cancer stem cells. *Bulletin of mathematical biology*, 75(1):161–184, 2013.

[73] Jacob G Scott, Anita B Hjelmeland, Prakash Chinnaiyan, Alexander RA Anderson, and David Basanta. Microenvironmental variables must influence intrinsic phenotypic parameters of cancer stem cells to affect tumourigenicity. *PLoS Comput Biol*, 10(1):e1003433, 2014.

[74] Heiko Enderling. Cancer stem cells: small subpopulation or evolving fraction? *Integrative Biology*, 7(1):14–23, 2015.

[75] Zheng Sun and Natalia L Komarova. Stochastic control of proliferation and differentiation in stem cell dynamics. *Journal of mathematical biology*, 71(4):883–901, 2015.

[76] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[77] Crispin W Gardiner et al. *Handbook of stochastic methods*, volume 3. Springer Berlin, 1985.

[78] EWJ Wallace, DT Gillespie, KR Sanft, and LR Petzold. Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET systems biology*, 6(4):102–115, 2012.

[79] Charles R Doering, Khachik V Sargsyan, and Leonard M Sander. Extinction times for birth-death processes: Exact results, continuum asymptotics, and the failure of the fokker–planck approximation. *Multiscale Modeling & Simulation*, 3(2):283–299, 2005.

[80] Michael Assaf and Baruch Meerson. Spectral theory of metastability and extinction in birth-death systems. *Physical review letters*, 97(20):200602, 2006.

[81] Linheng Li and Hans Clevers. Coexistence of quiescent and active adult stem cells in mammals. *Science*, 327(5965):542–545, 2010.

[82] Robert K Montgomery, Diana L Carlone, Camilla A Richmond, Loredana Farilla, Mariette EG Kranendonk, Daniel E Henderson, Nana Yaa Baffour-Awuah, Dana M Ambruzs, Laura K Fogli, Selma Algra, et al. Mouse telomerase reverse transcriptase (mtert) expression marks slowly cycling intestinal stem cells. *Proceedings of the National Academy of Sciences*, 108(1):179–184, 2011.

[83] Nick Barker, Johan H van Es, Jeroen Kuipers, Pekka Kujala, Maaike van den Born, Miranda Cozijnsen, Andrea Haegebarth, Jeroen Korving, Harry Begthel, Peter J Peters, et al. Identification of stem cells in small intestine and colon by marker gene lgr5. *Nature*, 449(7165):1003–1007, 2007.

[84] Christopher S Potten, Catherine Booth, Gregory L Tudor, Dawn Booth, Gerard Brady, Patricia Hurley, Gary Ashton, Robert Clarke, Shin-ichi Sakakibara, and Hideyuki Okano. Identification of a putative intestinal stem cell and early lineage marker; musashi-1. *Differentiation*, 71(1):28–41, 2003.

[85] Javier Muñoz, Daniel E Stange, Arnout G Schepers, Marc van de Wetering, Bon-Kyoung Koo, Shalev Itzkovitz, Richard Volckmann, Kevin S Kung, Jan Koster, Sorina Radulescu, et al. The lgr5 intestinal stem cell signature: robust expression of proposed quiescent + 4cell markers. *The EMBO journal*, 31(14):3079–3091, 2012.

[86] Anne E Powell, Yang Wang, Yina Li, Emily J Poulin, Anna L Means, Mary K Washington, James N Higginbotham, Alwin Juchheim, Nripesh Prasad, Shawn E Levy, et al. The pan-erbb negative regulator lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. *Cell*, 149(1):146–158, 2012.

[87] Satoshi Nishimura, Naoki Wakabayashi, Kazuyuki Toyoda, Kei Kashima, and Shoji Mitsufuji. Expression of musashi-1 in human normal colon crypt cells: a possible stem cell marker of human colon epithelium. *Digestive diseases and sciences*, 48(8):1523–1529, 2003.

[88] Martin A Nowak. *Evolutionary dynamics*. Harvard University Press, 2006.

[89] Dominik Wodarz and Natalia L Komarova. Mathematical foundations of oncology. 2014.

[90] Zheng Sun, Maksim V Plikus, and Natalia L Komarova. Near equilibrium calculus of stem cells in application to the airway epithelium lineage. *PLoS Comput Biol*, 12(7):e1004990, 2016.

[91] Alexander F Schier and Daniel Needleman. Developmental biology: Rise of the source–sink model. *Nature*, 461(7263):480–481, 2009.

[92] U Paulus, Christopher S Potten, and M Loeffler. A model of the control of cellular regeneration in the intestinal crypt after perturbation based solely on local stem cell regulation. *Cell proliferation*, 25(6):559–578, 1992.

[93] Guoqiang Hua, Tin Htwe Thin, Regina Feldman, Adriana Haimovitz-Friedman, Hans Clevers, Zvi Fuks, and Richard Kolesnick. Crypt base columnar stem cells in small intestines of mice are radioresistant. *Gastroenterology*, 143(5):1266–1276, 2012.

[94] IMM Van Leeuwen, HM Byrne, OE Jensen, and JR King. Crypt dynamics and colorectal cancer: advances in mathematical modelling. *Cell proliferation*, 39(3):157–181, 2006.

[95] Sophie K Kershaw, Helen M Byrne, David J Gavaghan, and James M Osborne. Colorectal cancer through simulation and experiment. *IET systems biology*, 7(3):57–73, 2013.

[96] Giovanni De Matteis, Alex Graudenzi, and Marco Antoniotti. A review of spatial computational models for multi-cellular systems, with regard to intestinal crypts and colorectal cancer development. *Journal of mathematical biology*, 66(7):1409–1462, 2013.

[97] Alexis J Carulli, Linda C Samuelson, and Santiago Schnell. Unraveling intestinal stem cell behavior with models of crypt dynamics. *Integrative Biology*, 6(3):243–257, 2014.

[98] Béla Novák and John J Tyson. Design principles of biochemical oscillators. *Nature reviews Molecular cell biology*, 9(12):981–991, 2008.

[99] Benjamin D Simons and Hans Clevers. Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell*, 145(6):851–862, 2011.

[100] Carlos Lopez-Garcia, Allon M Klein, Benjamin D Simons, and Douglas J Winton. Intestinal stem cell replacement follows a pattern of neutral drift. *Science*, 330(6005):822–825, 2010.

[101] Hugo J Snippert, Laurens G Van Der Flier, Toshiro Sato, Johan H Van Es, Maaike Van Den Born, Carla Kroon-Veenboer, Nick Barker, Allon M Klein, Jacco Van Rheenen, Benjamin D Simons, et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing lgr5 stem cells. *Cell*, 143(1):134–144, 2010.

[102] Nick Barker. Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nature reviews Molecular cell biology*, 15(1):19–33, 2014.

[103] Adam Humphries and Nicholas A Wright. Colonic crypt organization and tumorigenesis. *Nature Reviews Cancer*, 8(6):415–424, 2008.

Appendices

A The summation equations

Here we consider the patterns for the 5 summation equations which are derived from equation (2.8) by multiplying by $i^\alpha j^\beta$ and summing over i and j, for $\alpha + \beta \leq 2$. Let us denote the summation equation derived from multiplying by $i^\alpha j^\beta$ by the pair of numbers (α,β).

We present the example of equation $(1,0)$. Multiplying equation (2.8) by i and summing over i and j, we obtain

$$\begin{aligned} 0 \;=\; & \sum_i \sum_j \tilde{\varphi}_{ij}(-\frac{1}{2}\epsilon(4iL_0P_x + 4jL_0P_y) - \frac{1}{2}\epsilon^2[2i^2(2L_xP_x + L_0P_{xx}) \\ & +4ij(L_yP_x + L_0P_{xy} + L_xP_y) + 2j^2(2L_yP_y + L_0P_{yy})] - \frac{1}{2}\epsilon^3[i^3(2L_{xx}P_x + L_xP_{xx}) \\ & +i^2j(4L_{xy}P_x + 2L_yP_{xx} + 4L_xP_{xy} + 2L_{xx}P_y) + ij^2(4L_{xy}P_y + 2L_xP_{yy} + 4L_yP_{xy} + 2L_{yy}P_x) \\ & +j^3(2L_{yy}P_y + 2L_yP_{yy})] - \frac{1}{2}\epsilon^4[i^4L_{xx}P_{xx} + 2i^3j(L_{xy}P_{xx} + L_{xx}P_{xy}) \\ & +i^2j^2(L_{yy}P_{xx} + 4L_{xy}P_{xy} + L_{xx}P_{yy}) + j^4L_{yy}P_{yy}]) \end{aligned} \tag{A.1}$$

From the above equation, we notice that the power of ϵ is consistent with the power of $i^\alpha j^\beta$. Indeed, every term containing ϵ^k term multiplies a term of the form $i^\alpha j^\beta$, where $\alpha + \beta = k$.

The reason for this can be seen by examining the structure of equation (2.8). In this equation, every term containing the power k of ϵ (as follows from the number of derivatives of the probability functions), is multiplied by $i^\alpha j^\beta$ with $\alpha + \beta = k$.

The same property holds for the other four summation equations, which we do not present here. To derive all the summation equations, we need to perform the summations in i and j and use definition (2.9) for the moments. Each equation will be coupled to other equations containing higher order moments. The easiest way to close the system is to use the simple truncation method. The resulting system of 5 equations is given by (2.17-2.21). The other two truncation methods are worked out below.

B A case study

In this section, we will consider a special case of the general model equation (2.1). We will see that the results for the general case coincide with the results found previously by a different methodology in [75] and reported in Section 2.1.1. In particular, we will demonstrate that all three truncation methods have the same result for $a^1_{\alpha\beta}$, but different results for $a^2_{\alpha\beta}$, where $x_{\alpha\beta} = \frac{a^1_{\alpha\beta}}{\epsilon^n} + \frac{a^2_{\alpha\beta}}{\epsilon^{n-1}}$, $\alpha + \beta \leq 2$.

In this example, we assume that

$$Q(N) = \frac{b}{1 + hN}, \quad P(J) = \frac{r}{1 + gJ},$$

where $h, g \ll 1$ and $b, r = O(1)$ are two constants. The corresponding Kolmogorov forward

equation is given by:

$$\begin{aligned}\dot{\varphi}_{J,N} &= \varphi_{J+1,N+1}(1-Q(N+1)) + \varphi_{J,N-1}Q(N-1)(1-P(J)) + \\ &\quad \varphi_{J-2,N-1}Q(N-1)P(J-2) - \varphi_{J,N}. \end{aligned} \tag{B.2}$$

Let n_0 and j_0 be the steady state, and $n = N - n_0$, $j = J - j_0$. n_0 and j_0 are defined to satisfy the deterministic equations:

1. $\dot{N} = Q(N) - (1 - Q(N))$,
2. $\dot{J} = 2Q(N)P(J) - (1 - Q(N))$.

We can easily see that $n_0 = \frac{2b-1}{h}$.

Define $\tilde{\varphi}_{j,n}$ such that $\tilde{\varphi}_{j,n} = \varphi_{J,N}$, and $\tilde{Q}(n) = Q(N)$, $\tilde{P}(j) = P(J)$, then (B.2) can be rewritten as:

$$\begin{aligned}\dot{\tilde{\varphi}}_{j,n} &= \tilde{\varphi}_{j+1,n+1}(1-\tilde{Q}(n+1)) + \tilde{\varphi}_{j,n-1}\tilde{Q}(n-1)(1-\tilde{P}(j)) + \\ &\quad \tilde{\varphi}_{j-2,n-1}\tilde{Q}(n-1)\tilde{P}(j-2) - \tilde{\varphi}_{j,n}. \end{aligned} \tag{B.3}$$

Expanding $Q(N)$ and $P(J)$ in Taylor series, we obtain

$$Q(N) = \frac{1}{2} + r_1 n + r_2 n^2 + \cdots,$$

$$P(J) = \frac{1}{2} + s_1 j + s_2 j^2 + \cdots,$$

Let us multiply both sides of equation (B.3) by $j^\alpha n^\beta$ and sum over j, n in the quasi-stationary

state. We obtain the following 5 summation equations, where $x_{\alpha\beta} = \sum_j \sum_n j^\alpha n^\beta \tilde{\varphi}_{j,n}$:

$$r_1 x_{01} + r_2 x_{02} = 0$$

$$2r_1 x_{01} + s_1 x_{10} + 2r_2 x_{02} + 2r_1 s_1 x_{11} + 2r_2 s_1 x_{12} + s_2 x_{20} + 2r_1 s_2 x_{21} + 2r_2 s_2 x_{22} = 0$$

$$1 + 4r_1 x_{02} + 4r_2 x_{03} = 0$$

$$1 + s_1 x_{10} + 2r_1 x_{02} + (2r_1 + s_1 + 2r_1 s_1) x_{11} + s_2 x_{20} + 2r_2 x_{03} + 2(r_2 + r_1 s_1 + r_2 s_1) x_{12}$$

$$+(s_2 + 2r_1 s_2) x_{21} + 2r_2 s_1 x_{13} + 2s_2 (r_1 + r_2) x_{22} + 2r_2 s_2 x_{23} = 0$$

$$\frac{3}{2} + r_1 x_{01} + 2s_1 x_{10} + r_2 x_{02} + 4r_1 (1 + s_1) x_{11} + 2(s_1 + s_2) x_{20} + 4r_2 (1 + s_1) x_{12} + 4r_1 (s_1 + s_2) x_{21}$$

$$+2s_2 x_{30} + 4r_2 (s_1 + s_2) x_{22} + 4r_1 s_2 x_{31} + 4r_2 s_2 x_{32} = 0$$

As in the general case, we expand every term in the Taylor series:

$$Q(N) = \frac{1}{2} - \frac{h_0}{4} n + \frac{h_0^2}{8} n^2,$$

$$P(J) = \frac{1}{2} - \frac{g_0}{4} j + \frac{g_0^2}{8} j^2,$$

where we introduce the following short-hand notations:

$$h_0 = \frac{h}{b}, \quad g_0 = \frac{g}{r}, \quad \eta = \frac{g_0}{h_0}.$$

We will use the truncation equations of Appendix D for central moment and cumulant closure method to solve the system for $x_{\alpha\beta} = \frac{a^1_{\alpha\beta}}{h_0^n} + \frac{a^2_{\alpha\beta}}{h_0^{n+1}}$, $\alpha + \beta \leq 2$. The solutions are presented below.

1. Simple truncation method

$$
\begin{aligned}
E[j] &= x_{10} = \frac{4+3\eta}{2(2+\eta)} + O(h_0^2) \\
E[n] &= x_{01} = \frac{1}{2} + O(h_0^2) \\
E[j^2] &= x_{20} = \frac{2+3\eta}{\eta(2+\eta)} * \frac{1}{h_0} + O(h_0) \\
E[jn] &= x_{11} = \frac{2}{2+\eta} * \frac{1}{h_0} + O(h_0) \\
E[n^2] &= x_{02} = \frac{1}{h_0} + O(h_0)
\end{aligned}
$$

Comparing with the old results in formulas (2.3)-(2.6), here we have

$$
E[N] = \frac{2b-1}{h} + \frac{1}{2}, Var[N] = \frac{b}{h} - \frac{1}{4}
$$

2. Central moment truncation method

$$
\begin{aligned}
x_{10} &= \frac{4+3\eta}{2(2+\eta)} + \frac{44\eta + 108\eta^2 + 93\eta^3 + 27\eta^4}{8(2+\eta)^3} h_0 + O(h_0^2) \\
x_{01} &= \frac{1}{2} + \frac{3}{8} h_0 + O(h_0^2) \\
x_{20} &= \frac{2+3\eta}{\eta(2+\eta)} * \frac{1}{h_0} + \frac{8 + 92\eta + 138\eta^2 + 93\eta^3 + 27\eta^4}{4\eta(2+\eta)^3} + O(h_0) \\
x_{11} &= \frac{2}{2+\eta} * \frac{1}{h_0} + \frac{16 + 24\eta + 22\eta^2 + 9\eta^3}{2(2+\eta)^3} + O(h_0) \\
x_{02} &= \frac{1}{h_0} + \frac{3}{4} + O(h_0)
\end{aligned}
\tag{B.4}
$$

$$
\begin{aligned}
x_{10} &= \frac{4+3\eta}{2(2+\eta)} + \frac{\eta(228+556\eta+483\eta^2+135\eta^3)}{32(2+\eta)^3}h_0 + O(h_0^2) \\
x_{01} &= \frac{1}{2} + \frac{3}{8}h_0 + O(h_0^2) \\
x_{20} &= \frac{2+3\eta}{\eta(2+\eta)} * \frac{1}{h_0} + \frac{328+580\eta+438\eta^2+135\eta^3}{16(2+\eta)^3} + O(h_0) \\
x_{11} &= \frac{2}{2+\eta} * \frac{1}{h_0} + \frac{128+140\eta+184\eta^2+105\eta^3}{16(2+\eta)^3} + O(h_0) \\
x_{02} &= \frac{1}{h_0} + \frac{3}{4} + O(h_0)
\end{aligned}
$$

From the above results, all three truncation methods have the same solution for $a^1_{\alpha\beta}$, but different solutions for $a^2_{\alpha\beta}$, where $x_{\alpha\beta} = \frac{a^1_{\alpha\beta}}{h_0^n} + \frac{a^2_{\alpha\beta}}{h_0^{n+1}}$, $\alpha+\beta \leq 2$.

As we can see, the simple truncation, central moment truncation and cumulant truncation methods yield the same result for the highest order terms in $x_{\alpha\beta}$, where $\alpha+\beta \leq 2$. Therefore, all three methods have the same result for the mean and variance of N and J, if we only keep the highest order term, as summarized below:

$$
\begin{aligned}
E[N] &= \frac{2b-1}{h} + \frac{1}{2} + O(h), && \text{(B.5)} \\
Var[N] &= \frac{b}{h} + O(1), && \text{(B.6)} \\
E[J] &= \frac{2r-1}{g} + \frac{4+3\eta}{2(2+\eta)} + O(h), && \text{(B.7)} \\
Var[J] &= \frac{2+3\eta}{\eta(2+\eta)} * \frac{1}{h_0} + O(1). && \text{(B.8)}
\end{aligned}
$$

To compare the above results with formulas (2.3)-(2.6), derived from direct calculations in

Section 2.1.1, we expand formulas (2.3)-(2.6) with respect to h_0, and then only keep the highest order terms. The results are identical to equations (B.5-B.8).

C Comparison of the three truncation methods

Each of the three truncation methods has its own advantages and disadvantages. The advantage of the simple truncation method is its simplicity. Compared to the simple truncation method, both the central moment and cumulant truncation method require more extensive calculations. However, these methods can give more accurate results, as shown below.

The order of magnitude for the lower moments. Let us expand the moments $x_{\alpha\beta}$ in a power series in terms of ϵ, $x_{\alpha\beta} = \sum_{n=-\infty}^{\infty} X_{\alpha\beta}^{(n)} \epsilon^n$. We will consider only the two highest order terms in this expansion, with the corresponding coefficients denoted as $a_{\alpha\beta}^1$ and $a_{\alpha\beta}^2$. That is, we write

$$x_{\alpha\beta} = \frac{a_{\alpha\beta}^1}{\epsilon^n} + \frac{a_{\alpha\beta}^2}{\epsilon^{n-1}} + O(1/\epsilon^{n-2}),$$

where $a_{\alpha\beta}^1, a_{\alpha\beta}^2 = O(1)$ are unknown constants that we need to find. Next, we prove that for $x_{\alpha\beta}$, where $\alpha + \beta \leq 2$, all three truncation methods yield the same result for $a_{\alpha\beta}^1$, but the results of the three methods differ for $a_{\alpha\beta}^2$.

To determine the largest contributions to the expansions for $x_{\alpha\beta}$, we consider the five moment equations. At order (m, k), we multiply equation (2.8) by $i^m j^k$ and perform a double-summations in i and j. We call the resulting equations the *summation equations*. Because of expansion (2.14-2.16), coefficients in front of different variables $x_{\alpha\beta}$ will have a different order in terms of ϵ. In general, such an equation will contain terms multiplying $x_{m+s,k+r}$ with $s = 0, 1, \ldots$ and $r = 0, 1, \ldots$. The coefficient in front of the term $x_{m+s,k+r}$ is of the order

ϵ^{s+r}. The summation equations may also contain a nonhomogeneous (constant) term of order $O(1)$.

Because $x_{01} = E[j]$, and $j = J - j_0$ is a small perturbation around the steady state, it is reasonable to assume that $x_{01} = a^1_{01} + a^2_{01}\epsilon$ and $x_{10} = a^1_{10} + a^2_{10}\epsilon$. Because $x_{02} = E[j^2]$, x_{02} should be at least of order $O(\frac{1}{\epsilon})$. But if x_{02} is of order $O(\frac{1}{\epsilon^2})$ or higher, then from the 5 summation equations, the coefficient of the terms with power $O(\frac{1}{\epsilon^k})$ with $k \geq 2$ should be 0, which is not the case. Therefore, $x_{02} = \frac{a^1_{02}}{\epsilon} + a^2_{02}$. Similarly, we have $x_{11} = \frac{a^1_{11}}{\epsilon} + a^2_{11}$ and $x_{20} = \frac{a^1_{20}}{\epsilon} + a^2_{20}$. Calculations presented in Appendix A demonstrate these arguments in detail. As it will be discussed in E.3 of Appendix, these assumptions are indeed valid.

The order of magnitude of the higher moments. For the 3rd order truncation equations, the central moment and the cumulant closure methods have the same truncation equations, because for $\alpha + \beta = 3$ we have $E[(i - E[i])^\alpha (j - E[j])^\beta] = \kappa_{\alpha,\beta}$. Setting these moments to zero yields 4 equations (equations (D.9)-(D.12) in Appendix D). Because $x_{\alpha\beta} = O(1)$ for $\alpha + \beta = 1$ and $x_{\alpha\beta} = O(\frac{1}{\epsilon})$ for $\alpha + \beta = 2$, in order to balance equations (D.9-D.12), $x_{\alpha\beta}$ with $\alpha + \beta = 3$ has to be of the order of $O(\frac{1}{\epsilon})$.

For $\alpha + \beta \geq 4$, the central moment and the cumulant closure methods have different truncation equations (see Appendix D.2). For the central moment truncation method, we have $x_{\alpha\beta} = O(\frac{1}{\epsilon})$ with $\alpha + \beta = 4,\ 5$. For the cumulant truncation method, we have $x_{\alpha\beta} = O(\frac{1}{\epsilon^2})$, where $\alpha + \beta = 4,\ 5$ (see Appendix D).

All three methods coincide for $a^1_{\alpha\beta}$, but differ for $a^2_{\alpha\beta}$. Next, we will show that all three truncation methods give the same result for $a^1_{\alpha\beta}$, but different results for $a^2_{\alpha\beta}$, where $x_{\alpha\beta} = \frac{a^1_{\alpha\beta}}{\epsilon^n} + \frac{a^2_{\alpha\beta}}{\epsilon^{n-1}} + O\left(\frac{1}{\epsilon^{n-2}}\right)$ and $\alpha + \beta \leq 2$.

First, we look for the highest order term of ϵ in the 5 summation equations. By the above and equation (A.1) in Appendix A, the highest order term in the 5 summation equations is

$$\epsilon f_1(x_{01}, x_{10}) + \epsilon^2 f_2(x_{02}, x_{11}, x_{20}),$$

where f_1 is a function of x_{01}, x_{10} and f_2 is a function of x_{02}, x_{11}, x_{20}. Presenting $x_{\alpha\beta}$ with $\alpha + \beta \leq 2$ as a series, and keeping only the highest order terms in the summation equations, we obtain a linear system for $a^1_{\alpha\beta}$. Because the linear system derived from the moment equations is independent of the truncation methods, we can see that all three truncation methods yield the same result for $a^1_{\alpha\beta}$, where $\alpha + \beta \leq 2$. This is also the reason why the highest order contributions to the expectation and variance for i and j are the same for all the three methods, and so are the leading order to the expectation and variance for I and J by equations (2.10), (2.12), and (2.13).

Next, we will show that $a^2_{\alpha\beta}$ with $\alpha + \beta \leq 2$ are different in the 3 truncation methods by looking for the second highest order terms of ϵ in the 5 summation equations.

In the simple truncation method, the second highest order terms are $a^2_{\alpha\beta}\epsilon^2$, $\alpha+\beta \leq 2$, because we simply assume that all the higher moments are 0. Thus, we get a linear system for $a^2_{\alpha\beta}$, $\alpha + \beta \leq 2$.

In the central moment truncation method, the second highest order terms are $a^2_{\alpha\beta}\epsilon^2$ for $\alpha + \beta \leq 2$, and $a^1_{\alpha\beta}\epsilon^2$ for $\alpha + \beta = 3$, because when $\alpha + \beta \geq 4$, the power is at least $O(\epsilon^3)$, by equation (A.1). Therefore, we obtain a linear system for $a^2_{\alpha\beta}$, $\alpha + \beta \leq 2$, which contains terms $a^1_{\alpha\beta}$ with $\alpha + \beta = 3$. We can solve the system for coefficients $a^1_{\alpha\beta}$ with $\alpha + \beta = 3$ from the third order central moment truncation equations.

Finally, for the cumulant truncation method, the second highest order terms are $a^2_{\alpha\beta}\epsilon^2$ for

$\alpha+\beta \leq 2$, and $a^1_{\alpha\beta}\epsilon^2$ for $\alpha+\beta = 3,\ 4$ (the argument is similar to the one presented above). The difference from the central moment closure method is that the linear system for $a^2_{\alpha\beta}$ with $\alpha+\beta \leq 2$ does not only contain terms $a^1_{\alpha\beta}$ with $\alpha+\beta = 3$, but also contains terms $a^1_{\alpha\beta}$ with $\alpha+\beta = 4$. Similarly, $a^1_{\alpha\beta}$ with $\alpha+\beta = 4$ can be obtained from the cumulant truncation equations.

From the above considerations, we can see that the equations for $a^2_{\alpha\beta}$ with $\alpha+\beta \leq 2$ for the simple truncation method contain no information about the higher order terms $x_{\alpha\beta}$, $\alpha+\beta \geq 3$. For the central moment truncation method, these equations contain some information about the higher order terms, which is $x_{\alpha\beta}$ with $\alpha+\beta = 3$. Finally, for the cumulant truncation method, these equations contain information about $x_{\alpha\beta}$ with $\alpha+\beta = 3,\ 4$.

In conclusion, the three truncation methods produce the same result to the leading order of the mean and variance for the cell population. While we expect the central moment and cumulant truncation methods to give more accurate results to the next order correction, they require more extensive computations. In section 2.3.2, we show that the cumulant truncation is the most accurate among the three methods. The advantage of the simple truncation method is the straightforward calculations that it involves.

D Truncation equations

For more sophisticated truncation techniques employed here, we need to use truncation equations that express the higher moments in terms of the lower moments. Here we present these truncation equations for the central moment truncation method and the cumulant truncation method.

D.1 Truncation equations for $x_{\alpha\beta}$, where $\alpha + \beta = 3$

These two methods have the same truncation equations for $x_{\alpha\beta}$, where $\alpha + \beta = 3$:

$$2x_{10}^3 - 3x_{10}x_{20} + x_{30} = 0 \tag{D.9}$$

$$2x_{01}x_{10}^2 - 2x_{10}x_{11} - x_{01}x_{20} + x_{21} = 0 \tag{D.10}$$

$$2x_{10}x_{01}^2 - 2x_{01}x_{11} - x_{10}x_{02} + x_{12} = 0 \tag{D.11}$$

$$2x_{01}^3 - 3x_{02}x_{01} + x_{03} = 0 \tag{D.12}$$

Because $x_{10}, x_{01} = O(1)$ and $x_{20}, x_{11}, x_{02} = O(\frac{1}{\epsilon})$, we can see that $x_{\alpha\beta} = O(\frac{1}{\epsilon})$ with $\alpha + \beta = 3$.

D.2 Truncation equations for $x_{\alpha\beta}$, where $\alpha + \beta = 4, 5$

When $\alpha + \beta = 4, 5$, central moment closure method has different truncation equations compared to cumulant closure method.

(a). Central moment closure method.

$$-3x_{01}x_{10}^3 + 3x_{10}^2x_{11} + 3x_{01}x_{10}x_{20} - 3x_{10}x_{21} - x_{01}x_{30} + x_{31} = 0$$

$$-3x_{01}^2x_{10}^2 + x_{02}x_{10}^2 + 4x_{01}x_{10}x_{11} - 2x_{10}x_{12} + x_{01}^2x_{20} - 2x_{01}x_{21} + x_{22} = 0$$

$$-3x_{10}x_{01}^3 + 3x_{01}^2x_{11} + 3x_{01}x_{10}x_{02} - 3x_{01}x_{12} - x_{10}x_{03} + x_{13} = 0$$

$$4x_{01}^2x_{10}^3 - x_{02}x_{10}^3 - 6x_{01}x_{10}^2x_{11} + 3x_{10}^2x_{12} - 3x_{01}^2x_{10}x_{20} + 6x_{01}x_{10}x_{21} - 3x_{10}x_{22}$$
$$+x_{01}^2x_{30} - 2x_{01}x_{31} + x_{32} = 0$$

$$4x_{10}^2x_{01}^3 - x_{20}x_{01}^3 - 6x_{10}x_{01}^2x_{11} + 3x_{01}^2x_{21} - 3x_{10}^2x_{01}x_{02} + 6x_{01}x_{10}x_{12} - 3x_{01}x_{22}$$
$$+x_{10}^2x_{03} - 2x_{10}x_{13} + x_{23} = 0$$

Because $x_{10}, x_{01} = O(1)$, $x_{20}, x_{11}, x_{02} = O(\frac{1}{\epsilon})$ and $x_{30}, x_{21}, x_{12}, x_{03} = O(\frac{1}{\epsilon})$, by the above truncation equations, $x_{\alpha\beta} = O(\frac{1}{\epsilon})$, for $\alpha + \beta = 4, 5$.

(b). Cumulant closure method. The truncation equations for cumulant closure method for $x_{\alpha\beta}$, where $\alpha + \beta = 4, 5$ are:

$$x_{31} - 3x_{10}x_{21} - x_{01}x_{30} - 3x_{20}x_{11} + 6x_{20}x_{01}x_{10} + 6x_{11}x_{10}^2 - 6x_{10}^3x_{01} = 0$$

$$x_{22} - 2x_{10}x_{12} - 2x_{01}x_{21} - x_{20}x_{02} - 2x_{11}^2 + 2x_{20}x_{01}^2 + 2x_{02}x_{10}^2 + 8x_{11}x_{10}x_{01} - 6x_{10}^2x_{01}^2 = 0$$

$$x_{13} - 3x_{01}x_{12} - x_{10}x_{03} - 3x_{02}x_{11} + 6x_{02}x_{01}x_{10} + 6x_{11}x_{01}^2 - 6x_{01}^3x_{10} = 0$$

$$24x_{01}^2x_{10}^3 - 6x_{02}x_{10}^3 - 36x_{01}x_{10}^2x_{11} + 12x_{10}x_{11}^2 + 6x_{10}^2x_{12} - 18x_{01}^2x_{10}x_{20}$$

$$+6x_{02}x_{10}x_{20} + 12x_{01}x_{11}x_{20} - 3x_{12}x_{20} + 12x_{01}x_{10}x_{21} - 6x_{11}x_{21} - 3x_{10}x_{22} + 2x_{01}^2x_{30}$$

$$-x_{02}x_{30} - 2x_{01}x_{31} + x_{32} = 0$$

$$24x_{10}^2x_{01}^3 - 6x_{20}x_{01}^3 - 36x_{10}x_{01}^2x_{11} + 12x_{01}x_{11}^2 + 6x_{01}^2x_{21} - 18x_{10}^2x_{01}x_{02}$$

$$+6x_{20}x_{01}x_{02} + 12x_{10}x_{11}x_{02} - 3x_{21}x_{02} + 12x_{01}x_{10}x_{12} - 6x_{11}x_{12} - 3x_{01}x_{22} + 2x_{10}^2x_{03}$$

$$-x_{20}x_{03} - 2x_{10}x_{13} + x_{23} = 0$$

By these truncation equations, we get $x_{\alpha\beta} = O(\frac{1}{\epsilon^2})$, for $\alpha + \beta = 4, 5$.

E Moment Equations

E.1 Linear Noise Approximation

To find the next order correction, we will collect terms up to $O(\epsilon^{3/2})$ in the master equation expansion. After rescaling time by $T = L_0\tau = L_0\epsilon t$, we will extend the Fokker-Planck equation (2.43) to $O(\epsilon^{1/2})$. Then, we can obtain the moment equations by integrations. Here we will only illustrate several of them:

$$\begin{aligned}
\frac{d\langle\xi\rangle}{dT} &= -2(P_x\langle\xi\rangle + P_y\langle\eta\rangle) + \epsilon^{1/2}[-(P_{xx} + \frac{2L_xP_x}{L_0})\langle\xi^2\rangle - (P_{yy} + \frac{2L_yP_y}{L_0}) \\
&\quad \langle\eta^2\rangle - 2(P_{xy} + \frac{L_xP_y + L_yP_x}{L_0})\langle\xi\eta\rangle], \qquad \text{(E.13)} \\
\frac{d\langle\xi^2\rangle}{dT} &= -4(P_x\langle\xi^2\rangle + P_y\langle\xi\eta\rangle) + 1 + \epsilon^{1/2}[-2(P_{xx} + \frac{2L_xP_x}{L_0})\langle\xi^3\rangle - 2(P_{yy} + \\
&\quad \frac{2L_yP_y}{L_0})\langle\xi\eta^2\rangle - 4(P_{xy} + \frac{L_xP_y + L_yP_x}{L_0})\langle\xi^2\eta\rangle + \frac{L_x\langle\xi\rangle}{L_0} + \\
&\quad \frac{L_y\langle\eta\rangle}{L_0}], \qquad \text{(E.14)} \\
\frac{d\langle\xi^3\rangle}{dT} &= -6P_x\langle\xi^3\rangle - 6P_y\langle\xi^2\eta\rangle + 3\langle\xi\rangle + \epsilon^{1/2}[-3\langle\xi^4\rangle(P_{xx} + \frac{2L_xP_x}{L_0}) - \\
&\quad 3\langle\xi^2\eta^2\rangle(P_{yy} + \frac{2L_yP_y}{L_0}) - 6\langle\xi^3\eta\rangle(P_{xy} + \frac{L_xP_y + L_yP_x}{L_0}) + 3\langle\xi^2\rangle\frac{L_x}{L_0} \\
&\quad +3\langle\xi\eta\rangle\frac{L_y}{L_0}], \qquad \text{(E.15)} \\
\frac{d\langle\xi^4\rangle}{dT} &= -8P_x\langle\xi^4\rangle - 8P_y\langle\xi^3\eta\rangle + 6\langle\xi^2\rangle. \qquad \text{(E.16)}
\end{aligned}$$

E.2 Simple Trunciation

Correspond with (E.13) - (E.16), we have:

$$(1,0) \quad -2\epsilon[L_0(P_y x_{01} + P_x x_{10})] + [-2L_y(P_y x_{02} + P_x x_{11}) - 2L_x(P_y x_{11} + P_x x_{20}) - L_0(P_{yy} x_{02} + 2P_{xy} x_{11} + P_{xx} x_{20})]\epsilon^2 = 0, \tag{E.17}$$

$$(2,0) \quad L_0 - 4L_0(P_y x_{11} + P_x x_{20})\epsilon + (L_y x_{01} + L_x x_{10})\epsilon + \frac{1}{2}\epsilon^2[-4L_0(P_{xx} + \frac{2L_x P_x}{L_0})x_{30} - 4L_0(P_{yy} + \frac{2L_y P_y}{L_0})x_{12} - 8L_0(P_{xy} + \frac{L_y P_x + L_x P_y}{L_0}) x_{21} + L_{yy} x_{02} + L_{xx} x_{20} + 2L_{xy} x_{11}] = 0, \tag{E.18}$$

$$(3,0) \quad 3L_0 x_{10} + \epsilon(3L_y x_{11} + 3L_x x_{20} - 2L_0 P_y x_{01} - 6L_0 P_y x_{21} - 2L_0 P_x x_{10} - 6L_0 P_x x_{30}) + \epsilon^2(-2L_x P_y x_{11} + \frac{3}{2}L_{yy} x_{12} - 2L_x P_x x_{20} + 3L_{xy} x_{21} + \frac{3}{2}L_{xx} x_{30} - 6L_x P_y x_{31} - 2L_y P_y x_{02} - 6L_y P_y x_{22} - 2L_y P_x x_{11} - 6L_y P_x x_{31} - 6L_x P_x x_{40} - L_0 P_{yy} x_{02} - 3L_0 P_{yy} x_{22} - 2L_0 P_{xy} x_{11} - 6L_0 P_{xy} x_{31} - L_0 P_{xx} x_{20} - 3L_0 P_{xx} x_{40}) = 0, \tag{E.19}$$

$$(4,0) \quad L_0 + 6L_0 x_{20} + \epsilon(L_y x_{01} + 6L_y x_{21} + L_x x_{10} + 6L_x x_{30} - 8L_0 P_y x_{11} - 8L_0 P_y x_{31} - 8L_0 P_x x_{20} - 8L_0 P_x x_{40}) = 0. \tag{E.20}$$

By using (2.49), equations (E.13)-(E.15) coincide with equations (E.17)-(E.19) to $O(\epsilon^{1/2})$, and equations (E.16) agrees with (E.20) to $O(1)$. In fact, all the first, second, and third order moment equations from the two methods coincide to $O(\epsilon^{1/2})$, and so are the fourth order moment equations to $O(1)$.

E.3 Methodology

To find $O(\epsilon^{1/2})$ terms of the first order moments $\langle\xi\rangle, \langle\eta\rangle$, (E.13) shows we only need $O(1)$ terms of $\langle\xi^2\rangle, \langle\eta^2\rangle, \langle\xi\eta\rangle$. Since the two methods give the same results to the leading order of the second order moments, $O(\epsilon^{1/2})$ terms of the first order moments (which turn out to be non-zero) are the same for the two methods. It follows that the next order correction to the first order moments are the same for the two methods.

To find $O(\epsilon^{1/2})$ terms of the second order moments $\langle\xi^2\rangle, \langle\xi\eta\rangle, \langle\eta^2\rangle$, (E.14) shows we only need $O(1)$ terms of $\langle\xi^{\alpha\beta}\rangle$ for $\alpha+\beta=3$, and $O(1)$ terms of $\langle\xi\rangle, \langle\eta\rangle$. Through (E.15) and the other third order moment equations, we have $O(1)$ terms of $\langle\xi^{\alpha\beta}\rangle$, for $\alpha+\beta=3$, are all zero since $O(1)$ terms of $\langle\xi\rangle, \langle\eta\rangle$ are zero. It follows that $O(\epsilon^{1/2})$ terms of the second order moments $\langle\xi^2\rangle, \langle\xi\eta\rangle, \langle\eta^2\rangle$ are all zero.

To find $O(\epsilon)$ terms of the second order moments $\langle\xi^2\rangle, \langle\xi\eta\rangle, \langle\eta^2\rangle$, (E.14) shows we need $O(\epsilon^{1/2})$ terms of $\langle\xi^{\alpha\beta}\rangle$ for $\alpha+\beta=3$, and $O(\epsilon^{1/2})$ terms of $\langle\xi\rangle, \langle\eta\rangle$. We know $O(\epsilon^{1/2})$ terms of $\langle\xi\rangle, \langle\eta\rangle$ are the same for the two methods. To find $O(\epsilon^{1/2})$ terms of $\langle\xi^{\alpha\beta}\rangle$ for $\alpha+\beta=3$, we need $O(1)$ terms of $\langle\xi^{\alpha}\eta^{\beta}\rangle$ for $\alpha+\beta=4$, and $O(1)$ terms of $\langle\xi^{\alpha}\eta^{\beta}\rangle$ for $\alpha+\beta=2$ by (E.15). Now, (E.16) and the other fourth order moment equations show that $O(1)$ terms of $\langle\xi^{\alpha}\eta^{\beta}\rangle$ coincide for the two methods, for $\alpha+\beta=4$, since both methods give the same results to the leading order of the second order moments. It follows that the next order corrections to the second order moments (which turn out to be non-zero) are the same for the two methods.

By using (2.49), we have just shown that Van Kampen method and simple truncation give the same results to the next order correction of the mean and variance of the cell population. In particular, let I and i_0 denote the number of stem cells and its steady state; respectively,

then we have the following results:

$$\langle \xi \rangle = a_{10}^1 \, \epsilon^{1/2} + \cdots ; \tag{E.21}$$

$$\langle \xi^2 \rangle = a_{20}^1 + a_{20}^2 \, \epsilon + \cdots ; \tag{E.22}$$

$$E[I] = i_0 + a_{10}^1 + \cdots ; \tag{E.23}$$

$$Var[I] = \frac{a_{20}^1}{\epsilon} + a_{20}^2 - (a_{10}^1)^2 + \cdots . \tag{E.24}$$

F Minimal control systems

In section 3.1.1 we review the analysis of [1], see also figure 3.2 of the main text. For the two-compartment model, we note that at least two of the four quantities, (q_x, q_y, p_x, p_y), must be nonzero to satisfy the stability condition: $\Delta \equiv p_y q_x - p_x q_y > 0$. In fact, there are exactly two cases where only two of the four derivatives are nonzero and satisfy the other stability condition: $B \equiv 2L_* S_* (p_x - p_y) - q_y > 0$:

[1] $q_x < 0, \quad p_y < 0, \quad q_y = p_x = 0;$

[2] $q_y < 0, \quad p_x > 0, \quad q_x = p_y = 0.$

Extending the analysis to three nonzero controls, we find that there are exactly three cases that satisfy both stability conditions:

[3] $q_y < 0, \quad q_x > 0, \quad 0 < p_y < -\frac{q_y}{2L_* S_*}, \quad p_x = 0;$

[4] $q_y > 0, \quad q_x = 0, \quad p_x < 0, \quad p_y < p_x - \frac{q_y}{2L_* S_*} < 0;$

[5] $q_y = 0, \quad q_x > 0, \quad p_x > p_y > 0.$

Note that the case where $p_y = 0$ yields a system of controls that is reducible to the two-control model [2] by setting $q_x = 0$.

G Stability analysis

A deterministic description of the system is given by equations

$$\dot{x} = LS(1-P) - LSP = LS(1-2P), \tag{G.25}$$

$$\dot{y} = 2LSP + L(1-S) - D, \tag{G.26}$$

The equilibria are defined by

$$L_{i_0,j_0} = D_{i_0,j_0} = L_*, \quad P_{i_0,j_0} = \frac{1}{2}, \quad S_* = S_{i_0,j_0}. \tag{G.27}$$

$$L_{i_0,j_0} = D_{i_0,j_0} = L_*, \quad S_{i_0,j_0} = 0, \quad P_* = P_{i_0,j_0}. \tag{G.28}$$

We can compute the Jacobian of the system, evaluated at mixed divisions steady state (G.27):

$$J = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix},$$

where $f = LS(1-2P), g = 2LSP + L(1-S) - D$, and all derivatives are evaluated at the equilibrium (i_0, j_0).

Let det, τ be the determinant and the trace of J, respectively. Then,

$$\begin{aligned} det &= \frac{\partial f}{\partial x} \cdot \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \cdot \frac{\partial g}{\partial x} \\ &= \frac{2L_* S_* (q_x p_y - q_y p_x)}{\epsilon^2}; \\ \tau &= \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = -\frac{2L_* S_* (p_x - p_y) - q_y}{\epsilon}. \end{aligned}$$

The stability of the system requires $det > 0$ and $\tau < 0$. It follows that mixed divisions steady state is stable as long as $\Delta > 0$ and $B > 0$, where we defined

$$\Delta = q_x p_y - q_y p_x, \quad B = 2L_* S_* (p_x - p_y) - q_y.$$

Similarly, we evaluate the Jacobian at purely asymmetric divisions steady state (G.28), and we obtained in this case:

$$\begin{aligned} det &= \frac{L_* (2P_* - 1)(q_x s_y - q_y s_x)}{\epsilon^2}; \\ \tau &= \frac{L_* (s_y - s_x)(2P_* - 1) + q_y}{\epsilon}. \end{aligned}$$

It follows that purely asymmetric divisions steady state is stable as long as $\delta > 0$ and $b > 0$, where

$$b = -L_* (s_y - s_x)(2P_* - 1) - q_y, \quad \delta = (2P_* - 1)(q_x s_y - q_y s_x).$$

H Purely asymmetric divisions steady state

The analysis of the mixed divisions steady state is presented in the main text, and the result for the variances is given by

$$Var[I] = \frac{K_x}{4B\Delta}, \quad Var[J] = \frac{K_y}{4B\Delta}. \tag{H.29}$$

Here we present analysis of the dynamics in the vicinity of the purely asymmetric divisions equilibrium, equation (G.28).

H.1 Results for the cell number means and variances

It is important to note that in the limit of $S_* \to 0$ (asymmetric divisions only) formulas (H.29) break down. When $S_* = 0$, the numbers of stem cell cannot change in the model. The state space becomes one-dimensional. By using the same approach of "linear noise approximation" as in Chapter 2, we can obtain the means and the variances of the cell population to the highest order:

$$E[I] = i_0, \quad E[J] = j_0; \tag{H.30}$$

$$Var[I] = \frac{\{L_*(2P_* - 1)s_y\}^2}{b\delta}, \tag{H.31}$$

$$Var[J] = \frac{L_*}{b} + \frac{\{L_*(2P_* - 1)s_x\}^2}{b\delta}. \tag{H.32}$$

Evaluating this at the equilibrium, we obtain the correct answer for strictly asymmetric divisions (by taking the limit $S_* \to 0$)

$$Var[I] = 0, \quad Var[J] = -\frac{L_*}{q_y}.$$

H.2 A case study with a purely asymmetric divisions equilibrium

In this section, we will demonstrate that the analytic results of section H.1 agree with the numerical results. We will use the example of two-control model [2] in figure 3.2 of the main text, and equip it with a control of division symmetry parameter, S. This model is characterized by negative control on division and positive control on differentiation. We consider the following functional forms:

$$\begin{aligned} L(x,y) &= \frac{1}{1+y}, \quad P(x,y) = 0.7 \cdot \tanh(x), \\ D(x,y) &= 1 - L(x,y), \quad S(x,y) = \Big(1 - \frac{1}{cx}\Big)^2, \end{aligned} \tag{H.33}$$

where c is a positive constant.

The equilibria. We have $q_x = p_y = s_y = 0,\ q_y = -2\epsilon(1+y)^{-2} < 0,\ p_x = 0.7\epsilon \cdot \mathrm{sech}^2(x) > 0,$ and $s_x = 2\epsilon\Big(1 - \frac{1}{cx}\Big) \cdot \frac{1}{cx^2}$. The mixed divisions steady state can be obtained by solving $P(x,y) = 1/2$, and $L(x,y) = D(x,y)$:

$$x_m = \frac{\log(6)}{2}, \quad y_m = 1 \ (or\ i_m = \frac{\log(6)}{2\epsilon}, \quad j_m = \frac{1}{\epsilon}). \tag{H.34}$$

The purely asymmetric divisions steady state can be obtained by solving $S(x,y) = 0$, and $L(x,y) = D(x,y)$:

$$x_a = \frac{1}{c}, \quad y_a = 1 \ (or\ i_a = \frac{1}{c\epsilon}, \quad j_a = \frac{1}{\epsilon}). \tag{H.35}$$

By using the theory in the main text, we can obtain the means and the variances for the

mixed divisions solution:

$$E[I] = \frac{\log(6)}{2\epsilon}, \tag{H.36}$$

$$E[J] = \frac{1}{\epsilon}, \tag{H.37}$$

$$Var[I] = \frac{2L_*S_*\Delta + q_y^2}{4B\Delta}, \tag{H.38}$$

$$Var[J] = \frac{2L_*(2+S_*)\Delta + 8L_*^2S_*p_x^2}{4B\Delta}, \tag{H.39}$$

where all the partial derivates are evaluated at the mixed divisions steady state, and $L_* = 1/2$, $S_* = (1 - \frac{2}{c\log(6)})^2$, $\Delta = -q_y p_x$, and $B = 2L_*S_*p_x - q_y$.

Similarly, we obtain the means and the variances for the purely asymmetric solution by equations (H.30 -H.32):

$$E[I] = \frac{1}{c\epsilon}, \quad E[J] = \frac{1}{\epsilon}, \tag{H.40}$$

$$Var[I] = 0, \quad Var[J] = -\frac{L_*}{q_y}. \tag{H.41}$$

where $L_* = 1/2$, and q_y is evaluated at the purely asymmetric equilibrium.

Stability analysis. The mixed solution is stable in this case (since $det > 0$, and $\tau < 0$), but the stability of asymmetric solution is ambiguous (since $det = 0$, and $\tau < 0$), see Section G. Nevertheless, a nonlinear stability analysis can be performed in this case. Let $c_* = 1/x_m$, then if $c > c_*$, $x_m > x_a$, and if $c < c_*$, $x_m < x_a$. We have the following results by nonlinear stability analysis:

1. If $c > \frac{2}{\log(6)}$, and the system starts near (x_a, y_a) with $x > \frac{1}{c}$, then the solution will

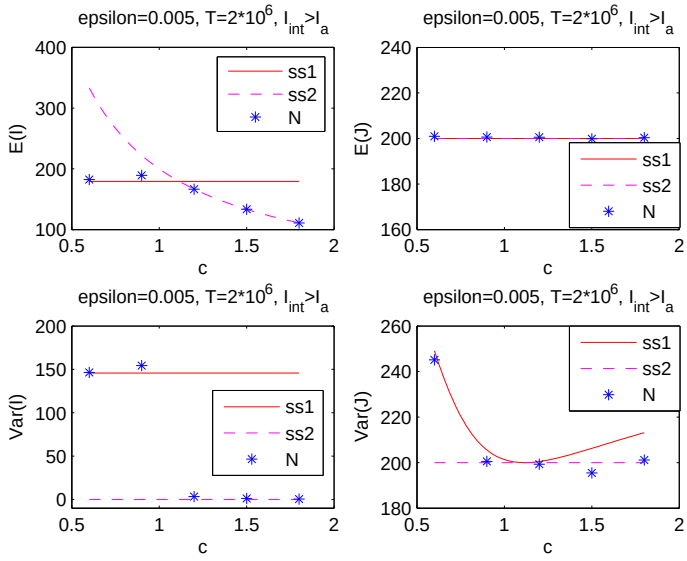


Figure H.1: The behavior of the means and variances of the system described by (H.33) with $\epsilon = 0.005$. The system starts at $(\frac{3i_m+i_a}{4}, j_a)$ for $c < c_* = 2/\log(6)$, and it starts at $(i_a - 20, j_a)$ when $c > c_*$. The analytical results given by (H.36-H.41) (solid line for mixed division and dashed line for purely asymmetric division) are compared with numerical results (stars), for different values of c. ('ss1') stands for the mixed division solution, ('ss2') stands for the purely asymmetric division solution, and ('N') stands for the numerical results.

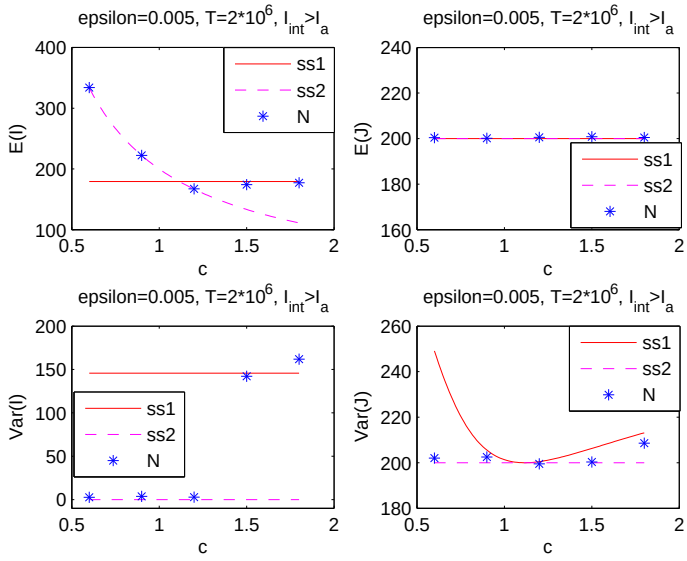


Figure H.2: Analogous graph as figure H.1, except the system starts at $(i_a + 10, j_a)$ for $c < c_*$, and it starts at $(\frac{3i_m+i_a}{4}, j_a)$ for $c > c_*$.

converge to (x_m, y_m) from below.

2. If $c > \frac{2}{\log(6)}$, and the system starts near (x_a, y_a) with $x < \frac{1}{c}$, then the solution will converge to (x_a, y_a) from below.

3. If $c < \frac{2}{\log(6)}$, and the system starts near (x_a, y_a) with $x > \frac{1}{c}$, then the solution will converge to (x_a, y_a) from above.

4. If $c < \frac{2}{\log(6)}$, and the system starts near (x_a, y_a) with $x < \frac{1}{c}$, then the solution will converge to (x_m, y_m) from above.

Numerical results. For each set of ϵ and c, we ran numerical simulations starting near (x_a, y_a), and finishing either when the number of time steps reached $2 \cdot 10^6$, or if any of the cell types went extinct. We then computed the means and the variances of the cell population over the time-course of each simulation. From figure H.1, we observe that when the system starts near (x_a, y_a) with $x < \frac{1}{c}$, the numerical results agree with the analytic results given by cases (2) and (4) in the above stability analysis. As observed in figure H.2, the numerical results are consistent with the analytic results given by cases (1) and (3) in the stability analysis when the system starts near (x_a, y_a) with $x > \frac{1}{c}$.

I Numerical simulations

The numerical simulations presented in this paper are set up in the following way. At each time-step, one of two events happen: either a SC divides with probability $\frac{L(i,j)}{L(i,j)+D(i,j)}$, or a differentiated cell dies with probability $\frac{D(i,j)}{L(i,j)+D(i,j)}$. In the case of a SC division, its nature is determined based on the probabilities $S(i,j)$ and $P(i,j)$. In this algorithm, each time-step corresponds to a cellular event (a division or a death). In other words, we have a non-uniform clock which only advances if a biological event takes place.

This is in contrast with a real-time, physical clock, which continues "ticking" even if no events take place. If we were to implement the latter, more realistic simulation, we could use the well-known Gillespie algorithm, where the probabilities of different events are determined as above, but the time-steps between the events, Δt_i, are assigned according to an exponential distribution. The latter type of a simulation would present a picture of population dynamics as it happens in physical time.

For our purposes, however keeping track of biological time is unnecessary. We are only concerned with calculating the means and the variances of the cell population, which are the same in our simulation and in the real-time simulation just described. To show this, let us suppose that we need to calculate the mean value of a stochastic variable f_i, which stands for the population size of interest at time-step i of our simulation. Then the mean value obtained from our simulation after n time-steps is simply given by

$$\langle f \rangle_1 = \frac{\sum_{i=1}^{n} f_i}{n}.$$

The value obtained from the simulation which keeps track of biological time is given by

$$\langle f \rangle_2 = \frac{\sum_{i=1}^{n} f_i \Delta t_i}{\sum_{i=1}^{n} \Delta t_i}.$$

The two expressions are equal because

$$\frac{\sum_{i=1}^{n} f_i}{n} \frac{\sum_{i=1}^{n} \Delta t_i}{n} = \frac{\sum_{i=1}^{n} f_i \Delta t_i}{n} \ ,$$

where the left hand side is $\langle f \rangle_1 \langle \Delta t \rangle_1$, and the right hand side is $\langle f \Delta t \rangle_1$. Because the two variables are independent, this equality holds.

The same argument holds for the moments of the variable f_i, where for kth moment we simply consider the mean value of the quantity $(f_i)^k$.

J Nonlinear control functions

For some parameter values it was possible to assume that the controls are described by functions (4.17-4.20), as long as the probability values (functions P_1 and P_2) are within the interval $[0, 1]$, and the rate functions (L_1 and L_2) are nonnegative. If in the course of stochastic dynamics the functions given by equations (4.17-4.20) fall outside these bounds, we implemented a rule where the functions were replaced by zeros or ones (that is, values inside the boundaries). For a set of parameters (specifically, for relatively small values of W), this rule resulted in a system failure, as depicted in figure J.3 for network #1, $W = 0.03$. As observed, SCs stop dividing after around some time; that is $L_1 \leq 0$. Using equation 4.17, one can see that the abnormal situation is due to the form of function L_1. We have for control system #1:

$$L_1 = L_0(1 + a_{L_1} i_1), \quad a_{L_1} < 0; \tag{J.42}$$

$$P_1 = \frac{1}{2}(1 + b_{P_1} i_2), \quad b_{P_1} < 0; \tag{J.43}$$

$$L_2 = (D_0 - L_0)(1 + c_{L_2} i_3), \quad c_{L_2} < 0; \tag{J.44}$$

where i_1, i_2 and i_3 are given by equation (4.9). We observe that L_1 drops below zero when the SC number is sufficiently large. In the case of this control system, the dynamics often drives the population of stem cells toward such values where L_1 drops below zero. In other words, linear function L_1 attains biologically unrealistic values within the range of normal fluctuations of x. Therefore, we need to replace the linear function L_1 with a nonlinear function, which remains biologically relevant within the range of x that is attained regularly by the stochastic system. Notice that such alternative, nonlinear function L_1^s should have the following properties:

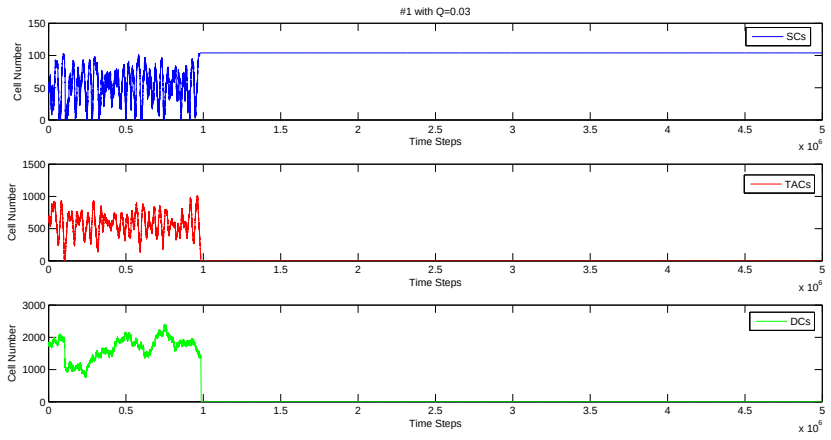


Figure J.3: A simulation of network #1 when $q = 0.1$ and $W = 0.03$. The figure is for linear function of L_1 described by equation (4.17). For the nonlinear control of L_1 described in Appendix J, see figure 4.5. In both cases, we hold other control functions as linear.

- L_1^s depends only on i_1, and it is decreasing;
- $L_1^s\Big|_{i_1=0} = L_0$ and $\frac{\partial L_1^s}{\partial i_1}\Big|_{i_1=0} = L_0 a_{L_1}$;
- $\lim_{i_1\to\infty} L_1^s = 0$.

One suitable choice of the substitute can be an exponential decay function: $L_1^s = L_0 e^{a_{L_1} i_1}$. As we observed from figure 4.5, network #1 behaves as expected with the nonlinear control L_1^s. For other networks, we perform a similar procedure whenever the assumption of linearity is violated.

K Eigenvalues Analysis

We illustrate the eigenvalues analysis on system 1 of figure 4.2. Using the figure and the notations in equations (4.17)-(4.20), system 1 can be characterized as:

$$\frac{\partial L_1}{\partial x_1} = \frac{L_0 a_{L_1}}{\epsilon} < 0, \quad \frac{\partial L_2}{\partial x_3} = \frac{(D_0 - L_0) c_{L_2}}{\epsilon} < 0, \quad \frac{\partial P_1}{\partial x_2} = \frac{b_{P_1}}{2\epsilon} < 0 \ .$$

The Jacobian of the system (evaluating at the steady state) can be obtained using the deterministic equations (4.2)-(4.4):

$$J = \begin{bmatrix} 0 & -\frac{L_0 b_{P_1}}{\epsilon} & 0 \\ \frac{L_0 a_{L_1}}{\epsilon} & \frac{L_0 b_{P_1}}{\epsilon} & -\frac{L_0 c_{L_2}}{\epsilon} \\ 0 & 0 & \frac{D_0 c_{L_2}}{\epsilon} \end{bmatrix}.$$

By solving the roots of the characteristic polynomial $|J - \lambda I|$, we obtain the eigenvalues:

$$\lambda_1 = \frac{D_0 c_{L_2}}{\epsilon}, \quad \lambda_2, \lambda_3 = \frac{L_0 b_{P_1} \pm \sqrt{L_0 b_{P_1}(L_0 b_{P_1} - 4 L_0 a_{L_1})}}{2\epsilon}.$$

It follows that system 1 is always stable (all eigenvalues have negative real part), and it is oscillatory if $L_0 b_{P_1} - 4 L_0 a_{L_1} > 0$ (complex eigenvalues indicate robust oscillations); that is, $a_{L_1} < \frac{b_{P_1}}{4}$.

We then performed the same analysis for system 2, 8, 9, 12, 13 and 18. The results are summarized below:

- system 2: $\frac{\partial L_1}{\partial x_1} = \frac{L_0 a_{L_1}}{\epsilon} < 0, \quad \frac{\partial P_1}{\partial x_2} = \frac{b_{P_1}}{2\epsilon} < 0, \quad \frac{\partial P_2}{\partial x_3} = \frac{D_0 c_{P_2}}{2\epsilon(D_0 - L_0)} < 0$
  - Eigenvalues are $\lambda_1 = \frac{D_0 c_{P_2}}{\epsilon}, \quad \lambda_2, \lambda_3 = \frac{L_0 b_{P_1} \pm \sqrt{L_0 b_{P_1}(L_0 b_{P_1} - 4 L_0 a_{L_1})}}{2\epsilon}$.
  - The system is always stable, and it is oscillatory if $a_{L_1} < \frac{b_{P_1}}{4}$.
- system 8: $\frac{\partial L_1}{\partial x_2} = \frac{L_0 b_{L_1}}{\epsilon} < 0, \quad \frac{\partial L_2}{\partial x_3} = \frac{(D_0 - L_0) c_{L_2}}{\epsilon} < 0, \quad \frac{\partial P_1}{\partial x_1} = \frac{a_{P_1}}{2\epsilon} > 0$
  - Eigenvalues are $\lambda_1 = \frac{2 D_0 c_{L_2}}{\epsilon}, \quad \lambda_2 = -\frac{L_0 a_{P_1}}{\epsilon}, \quad \lambda_3 = \frac{L_0 b_{L_1}}{\epsilon}$.
  - The system is always stable but not oscillatory.
- system 9: $\frac{\partial L_1}{\partial x_2} = \frac{L_0 b_{L_1}}{\epsilon} < 0, \quad \frac{\partial P_1}{\partial x_1} = \frac{a_{P_1}}{2\epsilon} > 0, \quad \frac{\partial P_2}{\partial x_3} = \frac{D_0 c_{P_2}}{2\epsilon(D_0 - L_0)} < 0$
  - Eigenvalues are $\lambda_1 = \frac{D_0 c_{P_2}}{\epsilon}, \quad \lambda_2 = -\frac{L_0 a_{P_1}}{\epsilon}, \quad \lambda_3 = \frac{L_0 b_{L_1}}{\epsilon}$.
  - The system is always stable but not oscillatory.

- system 12: $\frac{\partial L_2}{\partial x_3} = \frac{(D_0 - L_0)c_{L_2}}{\epsilon} < 0, \quad \frac{\partial P_1}{\partial x_1} = \frac{a_{P_1}}{2\epsilon} > 0, \quad \frac{\partial P_2}{\partial x_2} = \frac{D_0 b_{P_2}}{2\epsilon(D_0 - L_0)} > 0$
  - Eigenvalues are $\lambda_1 = -\frac{L_0 a_{P_1}}{\epsilon}, \quad \lambda_2, \lambda_3 = \frac{-D_0(b_{P_2} - c_{L_2}) \pm \sqrt{[D_0(b_{P_2} + c_{L_2})]^2 - 4D_0 L_0 b_{P_2} c_{L_2}}}{2\epsilon}$.
  - The system is always stable but not oscillatory.
- system 13: $\frac{\partial L_2}{\partial x_3} = \frac{(D_0 - L_0)c_{L_2}}{\epsilon} < 0, \quad \frac{\partial P_1}{\partial x_2} = \frac{b_{P_1}}{2\epsilon} < 0, \quad \frac{\partial P_2}{\partial x_1} = \frac{D_0 a_{P_2}}{2\epsilon(D_0 - L_0)} > 0$
  - Eigenvalues are $\lambda_1 = \frac{D_0 c_{L_2}}{\epsilon}, \quad \lambda_2, \lambda_3 = \frac{L_0 b_{P_1} \pm \sqrt{L_0 b_{P_1}(L_0 b_{P_1} + 4D_0 a_{P_2})}}{2\epsilon}$.
  - The system is always stable, and it is oscillatory if $a_{P_2} > -\frac{L_0 b_{P_1}}{4D_0}$.
- system 18: $\frac{\partial L_2}{\partial x_2} = \frac{(D_0 - L_0)b_{L_2}}{\epsilon} > 0, \quad \frac{\partial P_1}{\partial x_1} = \frac{a_{P_1}}{2\epsilon} > 0, \quad \frac{\partial P_2}{\partial x_3} = \frac{D_0 c_{P_2}}{2\epsilon(D_0 - L_0)} > 0$
  - Eigenvalues are $\lambda_1 = -\frac{L_0 a_{P_1}}{\epsilon}, \quad \lambda_2, \lambda_3 = \frac{(D_0 c_{P_2} - L_0 b_{L_2}) \pm \sqrt{(L_0 b_{L_2} + D_0 c_{P_2})^2 - 4D_0^2 b_{L_2} c_{P_2}}}{2\epsilon}$.
  - The system is stable if $b_{L_2} > \frac{D_0 c_{P_2}}{L_0}$, and it is oscillatory if $4 b_{L_2} c_{P_2} D_0^2 > (L_0 b_{L_2} + D_0 c_{P_2})^2$.

L Analysis of non-constant death rate minimal networks of figure 4.3

We followed the steps of the algorithm described in the main text to analyze the 12 minimal networks depicted in figure 4.3. The consecutive elimination steps are shown in figure L.4. First, we eliminated the non-local networks, which excluded networks 29-31. Out of the remaining networks 21-28, all eight allow for control coefficients compatible with the measured means and variances. The stochastic analysis is based on the linearization of the control functions given by equations (4.17-4.20) and the equation for the death rate,

$$D = D_0 \left(1 + a_D i_1 + b_D i_2 + c_D i_3\right).$$

Analysis of eigenvalues shows that networks #21, #22, and #23 are non-oscillatory. Networks #24 and #25 have complex eigenvalues, but the solution for the stem cell component is non-oscillatory. Similarly, network #26 has oscillatory eigenvalues but the total number of cells is non-oscillatory. Finally, networks #27 and #28 both have complex eigenvalues, and both the stem cells and total numbers of cells have oscillatory recovery dynamics. Of these two remaining networks, #28 has statistically significant intra-crypt correlations, while #27 does not, see figure L.5. Further, this network exhibits oscillatory recovery dynamics similar to that depicted in figure 4.7. Therefore, we conclude that the only remaining candidate network is #27.

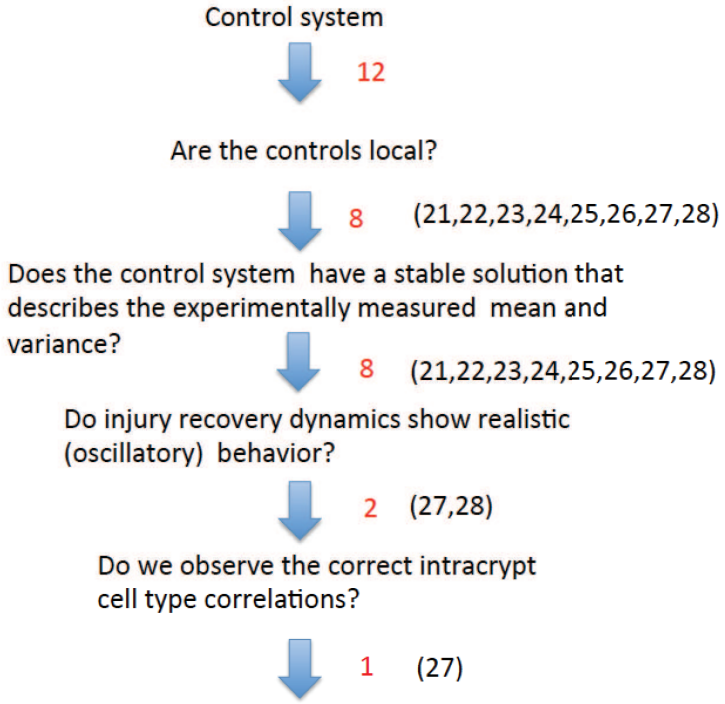


Figure L.4: The outcome of the selection algorithm for the non-constant death networks of figure 4.3 with $W = 0.03$.

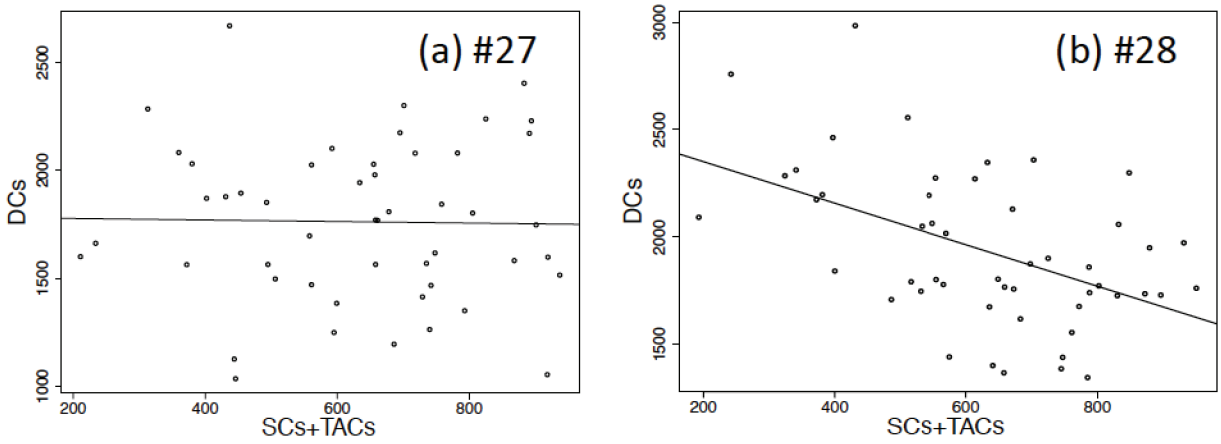


Figure L.5: The plots of DCs vs. the sum of SCs + TACs for two non-constant death networks. (a) Network 27 when $q = 0.2$, and p-value = 0.9. (b) Network 28 when $q = 0.2$, and the p-value is 0.000328 (other values of q also produced p< 0.05, not shown).