Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 250 (2008) 606–620
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Understanding stem cell differentiation through self-organization theory

K. Qu, P. Ortoleva*

Department of Chemistry, Center for Cell and Virus Theory, Indiana University, Bloomington, IN 47405, USA

Received 25 April 2007; received in revised form 11 October 2007; accepted 18 October 2007
Available online 23 October 2007

Abstract

The mechanism underling stem cells' key property, the ability to either divide into two replicate cells or a replicate and a differentiated daughter, still is not understood. We tested a hypothesis that stem cell asymmetric division/differentiation is spontaneously created by the coupling of processes within each daughter and the resulting biochemical feedbacks via the exchange of molecules between them during mitotic division. We developed a mathematical/biochemical model that accounts for dynamic processes accompanying division, including signaling initiation and transcriptional, translational and post-translational (TTP) reactions. Analysis of this model shows that it could explain how stem cells make the decision to divide symmetrically or asymmetrically under different microenvironmental conditions. The analysis also reveals that a stem cell can be induced externally to transition to an alternative state that does not have the potentiality to have the option to divide symmetrically or asymmetrically. With this model, we initiated a search of large databases of transcriptional regulatory network (TRN), protein–protein interaction, and cell signaling pathways. We found 12 subnetworks (motifs) that could support human stem cell asymmetric division. A prime example of the discoveries made possible by this tool, two groups of the genes in the genetic model are revealed to be strongly over-represented in a database of cancer-related genes.

Keywords: Stem cell differentiation; Self-organization; Symmetry-breaking; Bifurcation; Cancer

1. Introduction

Stem cells have the unique capacities to develop into specialized cell types and proliferate extensively (Kirschstein et al., 2001; Yu et al., 2006). Due to these unique properties, stem cells have been studied in many contexts in the pure and applied life sciences, e.g. tissue transplantation, repair of tissue damages, embryonic development and cancer (Kirschstein et al., 2001; Yu et al., 2006; Jordan et al., 2006; Tannishtha et al., 2001; Passegue et al., 2003; Tenen, 2003; Broxmeyer et al., 2006; Atala, 2006). A stem cell can divide into two identical stem cells (replication/symmetric division) or one replicate and a differentiated daughter cell (differentiation/asymmetric division). For years, researchers have sought ways to delineate mechanisms of stem cell differentiation and to develop techniques to direct it to produce specific cell lines with clinical relevant applications (Masson et al., 2004; Heng et al., 2004; Odorico et al., 2001; Atala, 2006). Meanwhile, theoretical models have been proposed and developed to describe cell differentiation mechanisms. In 2001, Furusawa and Kaneko proposed a theory that describes the robustness of irreversibility in stem cell differentiation based on an analysis of a random network. Theise and d'Inverno (2004) described cell differentiation as an adaptive phenomenon, and Cinquin and Demongeot (2005) investigated a model involving high-dimensional switches. These models conceptually explained the cell differentiation phenomenon, however, they were all based on hypothetical networks instead of an experimentally verified biological (e.g. human cell) system, and therefore their predictive capability is limited. Loeffler and Roeder (2002) summarized previous works and proposed a framework for developing predictive quantitative stem cell models. More recent models focused on small regulatory networks with genes and transcription factors (TFs) that had been demonstrated experimentally to be involved in stem cell differentiation. Roeder and Glauche (2006) built a model based on the gene/TF interaction of GATA-1 and

*Corresponding author. Tel.: +1 812 855 2717; fax: +1 812 855 8300.
E-mail address: ortoleva@indiana.edu (P. Ortoleva).

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.10.019

PU.1, which were considered key TFs mediating differentiation in the hematopoietic stem cell system. Chickarmane et al. (2006) developed a network of feedforward regulation of genes OCT4, SOX2 and NANOG that exhibits dynamical switching for embryonic stem cells. While these studies focused on gene/TF interactions, it is expected that cell differentiation is the consequence of a complex gene/TF interactions that can be genome-wide in scope due to extensive gene–gene cross-talk in the human transcriptional regulatory network (TRN), instead of being the consequence of a subnetwork of two or three "key" gene/TF interactions. Although all the above models seem to explain selected features of differentiation, none of them explained a stem cell's key capacity to divide into a replicate daughter and a differentiated one, instead of two differentiated daughters. More importantly, we suggest that a viable model of cell differentiation should not only be able to explain the biological phenomenon, but should also take advantages of the wealth of experimental data on cytokine-induction, database of signaling pathways, protein–protein interactions, and human TRN information; as well as gene expression microarray, proteomic and metabolic data. A recent step in that direction had been taken by Qu et al. (2007) that use a genome-wide approach to predict dramatic transitions of a human cell behavior based on TRN information, microarray data and a nonlinear dynamical system analysis.

Recent experimental results show that cytokines, growth factors and other constituents in the extra-cellular medium activate signaling pathways involved in the stem cell's decision to replicate or differentiate (symmetric versus asymmetric division) (Li and Neaves, 2006; Fuchs et al., 2004; Tumbar et al., 2004; Rizvi and Wong, 2005). Presently, at least four signaling pathways, Wnt (Wodarz and Nusse, 1998; Polakis, 2000), Notch (Gaiano and Fishell, 2002; Artavanis-Tsakonas et al., 1999), Hh (Villavicencio et al., 2000), and BMP (Haramis et al., 2004), have been characterized and shown to mediate cell proliferation and differentiation (Rizvi and Wong, 2005). The common characteristic of these pathways is that they are initiated with a cytokine that activates a membrane-bound receptor, through which a sequence of transcriptional, translational, and post-translational (TTP) processes are initiated within the cytoplasm and the nucleus, with resultant activation/deactivation of transcription of a specific set of genes, and subsequent stem cell replication/differentiation.

Although it is clear that the composition of the microenvironment (or "niche") is vital in creating suitable conditions for cytokine-induced cell replication/differentiation (Li and Neaves, 2006; Fuchs et al., 2004; Tumbar et al., 2004; Rizvi and Wong, 2005), many other aspects of stem cell replication/differentiation behaviors are still not understood, for instance:

- What are the quantitative relationships between the microenvironment conditions and stem cell replication/differentiation?
- What genes and proteins are involved in stem cell proliferation/differentiation? What are their functions and how do they interact? How do the characteristics of a stem cell change if the transcription of selected genes is changed?
- How can a stem cell differentiate into two distinct daughter cells when both are subjected to the identical microenvironment?

In this study, we provide a possible answer to the above questions based on the concepts of "symmetry-breaking instability" and "self-organization". A symmetry-breaking instability occurs when differences among interacting subsystems, all having the same internal dynamical processes, are spontaneously amplified as the result of a feedback that arises through the coupling of the subsystems, i.e. systems self-organize the pattern of asymmetry (Ortoleva and Ross, 1973a, b; Nicolis and Prigogine, 1977). Hallmarks of self-organizing systems are (1) the existence of control parameters which, in a given range of values, promote self-organization, while it is repressed in another range of these parameters; and (2) the pattern that develops is not imposed from the environment. In the present study, the subsystems are the two daughters of a dividing stem cell, while the control parameters are the concentrations of molecular constituents in the microenvironment. The intracellular dynamic processes we believe to be important include TTP processes and the activation of cell and nuclear surface receptor sites. Processes of molecular exchange between the daughter cells and with the microenvironment are also accounted for in our approach.

The present work is based on three fundamental hypotheses.

H1. Biochemical feedback can allow a cell to support multiple states of composition and other variables, all for the same microenvironment (Rashevsky, 1960; Hahn et al., 1973; Chickarmane et al., 2006; Qu et al., 2007).

H2. Two cells interacting through the exchange of molecules can spontaneously evolve into distinct states through biochemical feedback, even though they are identical in the set of processes occurring within them, their architecture, their initial state, and the microenvironment to which they are subjected (Ortoleva and Ross, 1973a, b).

H3. H1 and H2 for stem cell replication/differentiation can be tested due to the current availability of experimental data on eukaryotic cells, bioinformatics tools, and our cell modeling software (https://systemsbiology.indiana.edu; Sayyed-Ahmad et al., 2007; Tuncay et al., 2006; Sun et al., 2007; Qu et al., 2007).

It is hypothesized that the stem cell asymmetric division/differentiation is mediated via the interplay of the signaling pathways, TTP processes, and the exchange of proteins and other molecules between the two daughter cells accompanying division of a stem cell. We suggest differentiation

is one of a stem cell's capacities that originates in its internal TTP network which can be trigged through signaling pathways, as is also noted in Loeffler and Roeder (2002). We suggest that differences between daughter cells accompanying asymmetric division/differentiation are created by minor random events and are spontaneously amplified by biochemical reaction/transport feedbacks during division. We term this scenario the self-organization of stem cell differentiation and, in particular, propose a 5-gene/mRNA/protein model (TTP5) that integrates cytokine signal initiation, TTP processes and intercellular communication. The methodology proposed shows mathematically and biologically how a stem cell decides to replicate or differentiate and offers a procedure for discovering the distinct factors that control this decision.

The great promise of our theory and automated discovery approach is that, given multiple experimental data to calibrate the chemical kinetic and transport coefficients, one can identify ranges of conditions (e.g. concentrations of cytokines and growth factors) in the microenvironment that support asymmetric versus symmetric division. This methodology integrates and takes full advantage of the existing experimental results on cytokines, signaling pathways, human TRN and protein–protein interaction databases, and genomic and proteomic microarray data. This tool may also illustrate the range of influence for each experimentally or therapeutically controllable factor for which data exists. The availability of this quantitative and predictive model we are attempting to build will enable the computer-aided design of stem cell technologies and therapies.

2. Methods

2.1. Biological context

The cell cycle consists of DNA replication (Synthetic, S phase), division into two cells (Mitotic, M phase), and two intervening gap phases (G1 and G2). We hypothesize that asymmetry is self-organized in the M phase (Fig. 1). By our definition, a stem cell symmetry-breaking motif is a set of genes and proteins that enables the following sequence of events. Early in M phase, the two daughter cells are strongly coupled as the membrane separating them has not yet completely formed. At this stage, molecular exchange between the two daughters is so rapid that differences in their composition cannot develop. However, as the membrane forms, the symmetric state can be destabilized and differences in levels of proteins and other components are amplified through biochemical reaction-transport feedback. Once the membrane separates the daughter cells, they must be left in distinct states which, as for a hematopoietic stem cell, may result in complete isolation of the two daughters by physical separation. Thus, it seems evident that the single cell system must itself support multiple distinct stable states of the TTP system (including a stem cell state and a distinct transformed state). In stem

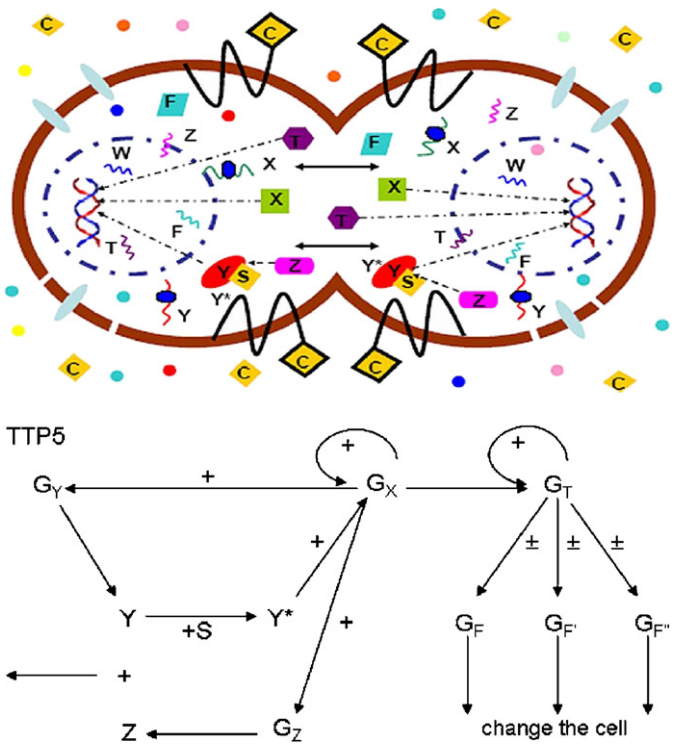

Fig. 1. Schematic depiction of a stem cell dividing into two daughters and TTP5 model. Top: stem cell dividing. The brown curve shows the cell membrane, the blue dotted line indicates the nuclei. The DNA has been replicated. mRNAs (colored strands) are being transcribed and translated into proteins (colored shapes). Cytokines (yellow diamond C) are sending signals into the cell by activating surface receptors. Resulting signals (yellow diamond S) within the cell complex with proteins Y and transform them into active transcription factors Y^*, which regulate the transcription of selected genes (e.g. G_X). Here, for simplicity, we assume the signal concentration equals to the cytokine concentration. In the stage shown above, the two daughter cells are still connected, and proteins and other molecules move between them. The colored dots represent small molecules (e.g. ATP, ions and hormones) that exist in both the microenvironment and the cell. Bottom: TTP5 model. Gene/protein interactions are shown. Auto-regulated gene G_X up-regulates (+) G_Y which encodes protein Y. S represents signaling factor activate protein Y to TF Y^*, which up-regulates G_X. G_X up-regulates G_Z, which encodes protein Z. Protein Z complexes with protein Y. G_T is also an auto-regulated gene, and is up-regulated by G_X. G_T up/down (±) regulates the expression of a lot functional genes G_F, $G_{F'}$, and $G_{F''}$, etc., which determine the cell behavior.

cell asymmetric division/differentiation, the symmetry-breaking mechanism guides one daughter into a stem cell state and the other into a differentiated state. In contrast, stem cell replication does not have this "deliverance" property, and both daughter cells are guided to the stem cell state. Therefore, a stem cell model must support both the symmetric and asymmetric cases under distinct extracellular conditions, as implied by the abilities of a stem cell to divide symmetrically or asymmetrically depending on the microenvironment.

A number of simple models were used to illustrate symmetry-breaking in reaction-transport systems (Turing, 1952; Ortoleva and Ross, 1973a, b), frog development (Borisuk and Tyson, 1998; Zhao et al., 1999; Valles, 2002), and electrical unicellular systems (Larter and Ortoleva,

1982). The model we developed is mainly inspired by the Brusselator (Nicolis and Prigogine, 1977), although it would be interesting to carry out a similar test of statistical significant for the other mechanisms as a best left to a future study. The Brusselator is one of the most well-studied models that support self-organized symmetry-breaking phenomena and its chemical reaction network is

$$A \rightarrow X \tag{1}$$

$$B + X \rightarrow Y + D \tag{2}$$

$$2X + Y \rightarrow 3X \tag{3}$$

$$X \rightarrow E \tag{4}$$

If A, B, D and E are kept constant, and species X and Y are allowed to exchange between adjacent cells, than this system can display symmetry-breaking instability under appropriate ranges of conditions (i.e. the concentrations of A and B). The special feedback/feedforward topology of the two species X and Y (e.g. X up-regulates itself through Y) in the Brusselator enables the system to support symmetry-breaking.

The following 5-gene/mRNA/protein network (TTP5) (see Fig. 1) is constructed to have the similar feedback/feedforward topology of that of the Brusselator and is shown here to support symmetry-breaking stem cell replication/differentiation. However, this model is built from the commonly accepted elements of the signaling pathways induced stem cell differentiation scenario. The latter include cytokine activation, TTP processes, with resultant activation/deactivation of transcription of specific target genes, and subsequent stem cell replication/differentiation. Genes G_X, G_Y, G_Z, G_T and G_F are transcribed to mRNAs R_X, R_Y, R_Z, R_T and R_F, and translated into proteins X, Y, Z, T and F, respectively. The auto-regulated gene G_X up-regulates G_Y, which encodes protein Y. Signaling factor S reacts with protein Y and activates it to TF Y^*, which up-regulates G_X. Meanwhile, G_X up-regulates G_Z, which encodes protein Z. Protein Z complexes with protein Y and competes with protein Y activation process. G_X, G_Y and G_Z form both feedforward and feedback loops, which are similar to the Brusselator topology. Another auto-regulated gene G_T, which is also up-regulated by G_X, up/down regulates the expression of a lot functional genes G_F, G_F, and $G_{F''}$, etc. In this model, the auto-regulation of gene G_T enables a cell to support multiple states (H1). G_X, G_Y and G_Z form a Brusselator-like structure that enables a symmetry-breaking behavior of the expression of G_X, which then delivers the expression of G_T from a stem cell state to a differentiated state in the differentiated daughter and keeps that of the replicate daughter unchanged. Since G_T controls a lot functional genes (G_F, $G_{F'}$, and $G_{F''}$, etc.), the different expression of G_T causes the different expression of those functional genes and thus change the cell behavior. This 5 gene/RNA/protein model constitutes of the minimum requirements for symmetry-breaking delivering differentiation. Auto-regulated genes G_X and G_T are the two key genes in the differentiation scenario. Post-translational reactions accounted-for include protein degradation and complexing of signaling molecules with protein Y.

2.2. Reaction/transport equations and rate parameters

$$\frac{dR_{X_1}}{dt} = K_{X_1}\left(\frac{Q_{X_1}X_1^2}{1+Q_{X_1}X_1^2}\right)\left(\frac{Q_{X_1}Y_1^*}{1+Q_{X_1}Y_1^*}\right) + \xi_{X_1} - \lambda_{X_1}R_{X_1} \tag{5}$$

$$\frac{dX_1}{dt} = \alpha_{X_1}R_{X_1} - \beta_{X_1}X_1 + h_X(X_2 - X_1) \tag{6}$$

$$\frac{dR_{Y_1}}{dt} = K_{Y_1}\frac{Q_{Y_1}X_1}{1+Q_{Y_1}X_1} + \xi_{Y_1} - \lambda_{Y_1}R_{Y_1} \tag{7}$$

$$\frac{dY_1}{dt} = \alpha_{Y_1}R_{Y_1} - \beta_{Y_1}Y_1 - \varepsilon CY_1 + h_Y(Y_2 - Y_1) \tag{8}$$

$$\frac{dY_1^*}{dt} = \varepsilon CY_1 - \gamma ZY_1^* + h_{Y^*}(Y_2^* - Y_1^*) \tag{9}$$

$$\frac{dR_{Z_1}}{dt} = K_{Z_1}\frac{Q_{Z_1}X_1^2}{1+Q_ZX_1^2} + \xi_{Z_1} - \lambda_{Z_1}R_{Z_1} \tag{10}$$

$$\frac{dZ_1}{dt} = \alpha_{Z_1}R_{Z_1} - \beta_{Z_1}Z + h_Z(Z_2 - Z_1) \tag{11}$$

$$\frac{dR_{T_1}}{dt} = K_{T_1}\left(\frac{Q_{T_1}T_1^2}{1+Q_{T_1}T_1^2}\right)\left(\frac{X_1}{1+X_1}\right) + \xi_{T_1} - \lambda_{T_1}R_{T_1} \tag{12}$$

$$\frac{dT_1}{dt} = \alpha_{T_1}R_{T_1} - \beta_{T_1}T_1 + h_T(T_2 - T_1) \tag{13}$$

$$\frac{dR_{F_1}}{dt} = K_{F_1}\frac{Q_{F_1}T_1^2}{1+Q_{F_1}T_1^2} + \xi_{F_1} - \lambda_{F_1}R_{F_1} \tag{14}$$

$$\frac{dF_1}{dt} = \alpha_{F_1}R_{F_1} - \beta_{F_1}F_1 + h_F(F_2 - F_1) \tag{15}$$

$$h_i = \begin{cases} h_{i0} & (t<0) \\ h_{i0} - \frac{t}{\tau}(h_{i0} - h_f) & (0 \leqslant t < \tau) \\ h_f & (t > \tau) \end{cases} \quad (i = X, Y, Y^*, Z, T, F) \tag{16}$$

Ordinary differentiation equations provide a quantitative description of the above processes. Sub note 1 represent RNAs/proteins/parameters/constants of daughter cell 1, and 2 means those of daughter cell 2. Only the equations of daughter cell 1 are shown. A more general genome-wide model had been developed in Qu et al. (2007) (Appendix A), where chemical kinetic model for transcriptional, translational and post-translational processes were deduce from statistical assumption and bioinformatics.

Similar formulas have been used to describe TTP processes in other sources (Cinquin and Demongeot, 2005; Hasty et al., 2000, 2001; Vu and Vohradsky, 2007; Zak et al., 2003). Eqs. (5), (7), (10), (12), and (14) model the transcription and degradation of mRNA R_X, R_Y, R_Z, R_T and R_F, respectively; Eqs. (6), (11), (13), and (15) account for the translation and degradation of proteins X, Z, T, and F. Eqs. (8) and (9) generate the dynamics of protein Y and TF Y^*, where C represents the cytokine concentration. h_i is the permeability of protein i ($i = X$, Y, Y^*, Z, T and F) for exchange between the two daughter cells times the shared area divide by the single cell volume, and follows a hill function as Eq. (16). t is time and τ is the typical time for mammalian cell mitosis. Here $t = 0$ means the stem cell begins to divide, thus two daughter cells have extensive communication and the permeability is large (h_{i0}). $t = \tau$ means the division has been completed, and two daughter cells have restricted communication h_f. We assume the signal molecule S concentration in the cell is proportional to the cytokine concentration in the micro-environment, thus we use cytokine concentration C to represent the effect that is made by signal molecules. K_i is the overall transcription rate forefactor of gene i, in considering RNA polymerase binding, reading and elongation. λ_i, α_i and β_i are the rate constants for mRNA degradation, translation and protein degradation respectively of their corresponding gene/RNA/protein i. ε and γ are rate constants for cytokine–protein and protein–protein binding. Q_i is the binding constant for a TF to a gene, which, for simplicity we take to be the same for all genes and TFs. ξ_i accounts for a small residual transcription rate for gene i. The values of these parameters are provided in Table 1.

Table 1
Parameter values and references

Parameters	Value	Data source
K transcription rate forefactor		
K_X (nM/h)	~320	Slutsky and Mirny (2004), Proudfoot (2000)
K_Y (nM/h)	~16	
K_Z (nM/h)	~16	
K_T (nM/h)	~16	
K_F (nM/h)	~16	
Q transcription factor binding constant	$0.05\,\mathrm{nM}^{-1}$ or $0.05\,\mathrm{nM}^{-2}$	None
λ mRNA degradation rate (h^{-1})	~1.0	⟨http://www.wisc.edu/molpharm/Courses/pharm620/Lecture_16.web.ppt⟩
α Translation rate (h^{-1})	~0.1	Ujvaro et al. (2001)
β Protein degradation rate constant (h^{-1})	~0.075	Averaged from the literature
h_{i0} permeability before division		
h_{X_0}	0.015	None
h_{Y_0}	2.625	
$h_{Y_0^*}$	2.625	
h_{Z_0}	0.20	
h_{T_0}	0.02	
h_{F_0}	0.02	
h_f permeability after division	10−4	None
ε Rate constant of binding for cytokine to proteins $(\mathrm{nM\,h})^{-1}$	~0.1	None
γ Rate constant of binding for two proteins $(\mathrm{nM\,h})^{-1}$	~0.08	Uetz et al. (2000)
ξ Residual transcription rate factor	0.05	None

2.3. Bifurcation analysis and AUTO

Bifurcation analysis is accomplished using an augmented version of AUTO software (Doedel et al., 1991) which delineates the types and range of existence of distinct states for a prescribed interval of physical parameters of a model. Our augmentation of AUTO involve the use of more efficient numerical procedures and created a preprocessor that automatically generates an AUTO Fortran input file for a general form of models and allows for an arbitrary number of molecular species and processes. This system is available at ⟨https://systemsbiology.indiana.edu⟩.

2.4. TRN construction and protein–protein interaction database

Gene regulation can be described via a TRN, which includes (1) a list of genes, each of which contains a set of proteins or protein complexes (the TFs) that up or down regulate their expressions and (2) for each of the TFs, the gene(s) that encode its components. We have developed a preliminary version of an automated systems biology approach to discover TRNs and implemented it as a web-based service (https://systemsbiology.indiana.edu and Qu et al., 2007), and we also have a video demo on the website tutoring of how to use our TRND system. Our construction workflow now contains several bioinformatics modules including: a TF-based method using cDNA microarray data to correct, extend and calibrate a TRN (FTF/KAGAN), gene ontology (GO) promoter analyses module, as well as a correlation method and Phylogenic similarity analysis. The results of the individual modules are integrated via a Bayesian approach to discover gene/TF regulatory interactions with the highest confidence (Sayyed-ahmad et al., 2007; Tuncay et al., 2006; Sun et al., 2007). The TRN we constructed and used in this study contains 3058 genes, 1187 TFs and 6166 experimentally verified and predicted gene/TF interactions for human cells.

The Human Protein Reference Database (HPRD: www.hprd.org) is developed by Peri et al. (2003). It records protein–protein interactions using Entrez Gene ID. We modify the ID-index-based interaction file to their corresponding encoding gene-name-based file through UniGene ID and UniGene name and get 37,385 pairs of human protein–protein interaction.

3. Results

3.1. Bifurcation and full time-dependent analysis

Behaviors supported by the TTP5 model described in Section 2 were explored via numerical simulation. Bifurcation analysis was used to delineate families of cell "states" (i.e. mRNAs, protein, TF intercellular contents) change with "control" variables (i.e. microenvironmental composition, contact with neighboring cells, or intracellular architecture, etc.). A dynamical approach was used to discover how the symmetry breaking mechanism delivers the daughter cells to the appropriate replicate or differentiated states.

We focus on events starting from late Gap 2 (G2) phase, through Mitosis (M) and ending in early Gap 1 (G1) phase. The integration of the bifurcation (Fig. 2a–e) and dynamic (Fig. 3a–e) approaches reveals how a stem cell undergoes

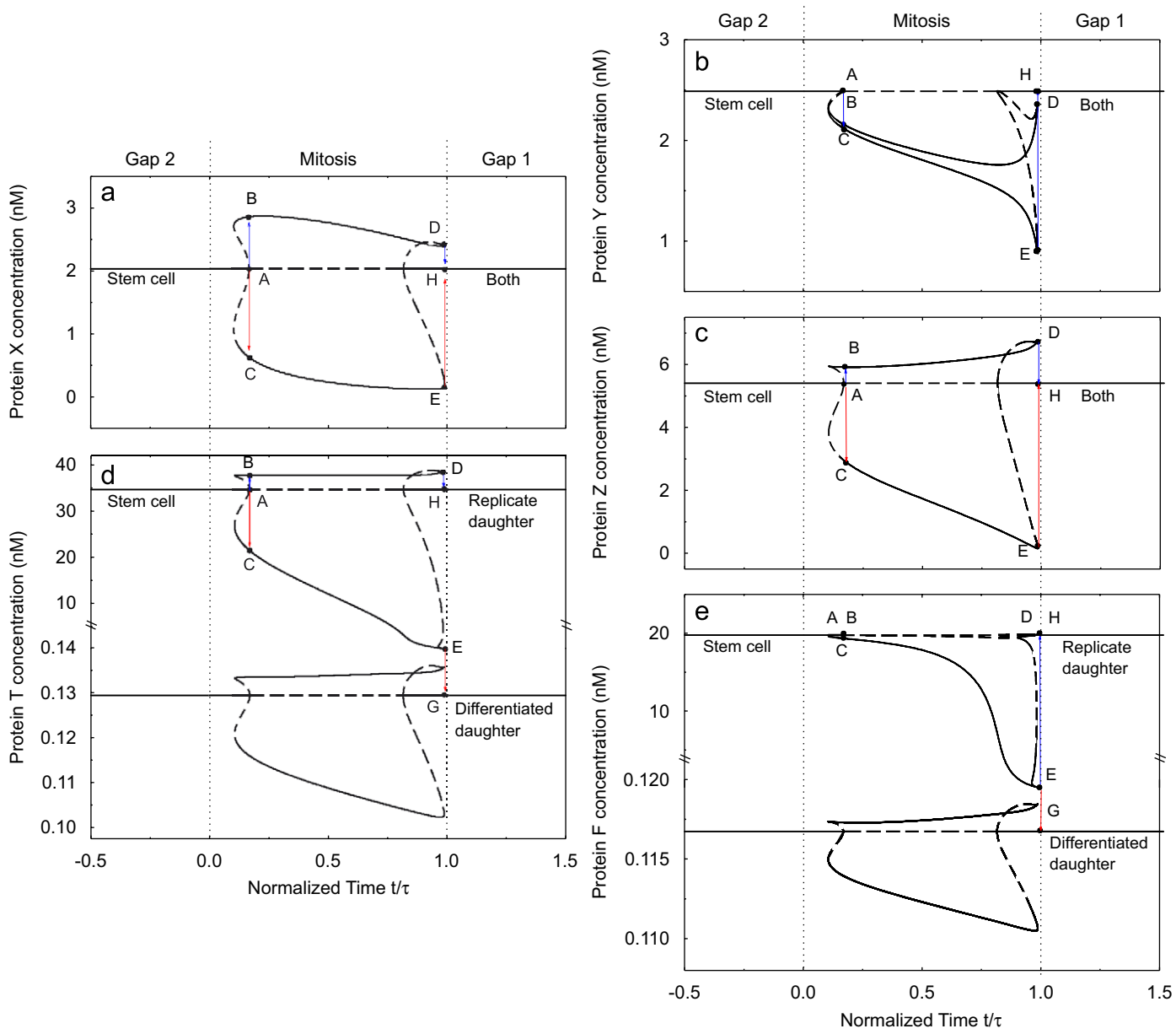

Fig. 2. Bifurcation diagram showing protein concentrations in two interacting daughter cells starting from late Gap 2 phase, proceeding through Mitosis, and ending in early Gap1 phase. Parts (a), (b), (c), (d), (e) are for proteins X, Y, Z, T and F concentrations versus normalized time t/τ, respectively, which corresponds to the change in permeability, i.e. the ease of communication between the two daughter cells. Solid lines are the stable steady states; dashed lines are the unstable steady states. The two-cell system stays on a stable symmetric steady state until it reaches the first bifurcation point A at early M phase, when one of the daughters transitions to state B, while the other goes to state C (i.e. symmetry-breaking occurs). The two daughters remain in distinct states until they reach states D/E, when the replicate daughter transitions back to the original stem cell state (H); the differentiated daughter simultaneously transitions to a state in which some of its proteins (e.g. X, Y and Z) are at the same concentration as that of the original stem cell, while others are different (e.g. T and F). Transitions of protein levels in the replicate/differentiated daughters are indicated by blue/red arrows. Here, we set cytokine concentration $C = 1.0$ nM.

asymmetric division. The stem cell system stays on a stable steady state until it enters M phase, and reaches the first bifurcation point A, when one of the daughter cells transitions to state B, while the other transitions to state C (i.e. symmetry-breaking occurs). The two daughters remain in distinct states from that of the original stem cell until they reach states D/E, when the replicate daughter transitions to a state that is the same as the original stem cell (state H), and the differentiated daughter transitions to a state with some of its proteins (e.g. X, Y and Z) at the same concentration levels as those of the original stem cell, while other proteins (e.g. T and F) are delivered to different levels and thereby distinguishing that cell to a different phenotype. As mentioned in Section 2, the symmetry-breaking of the expression of protein X helps to deliver the expression of protein T and F from their levels in stem cell to that characteristics of the differentiated daughter while those in the replicate daughter end up to be unchanged.

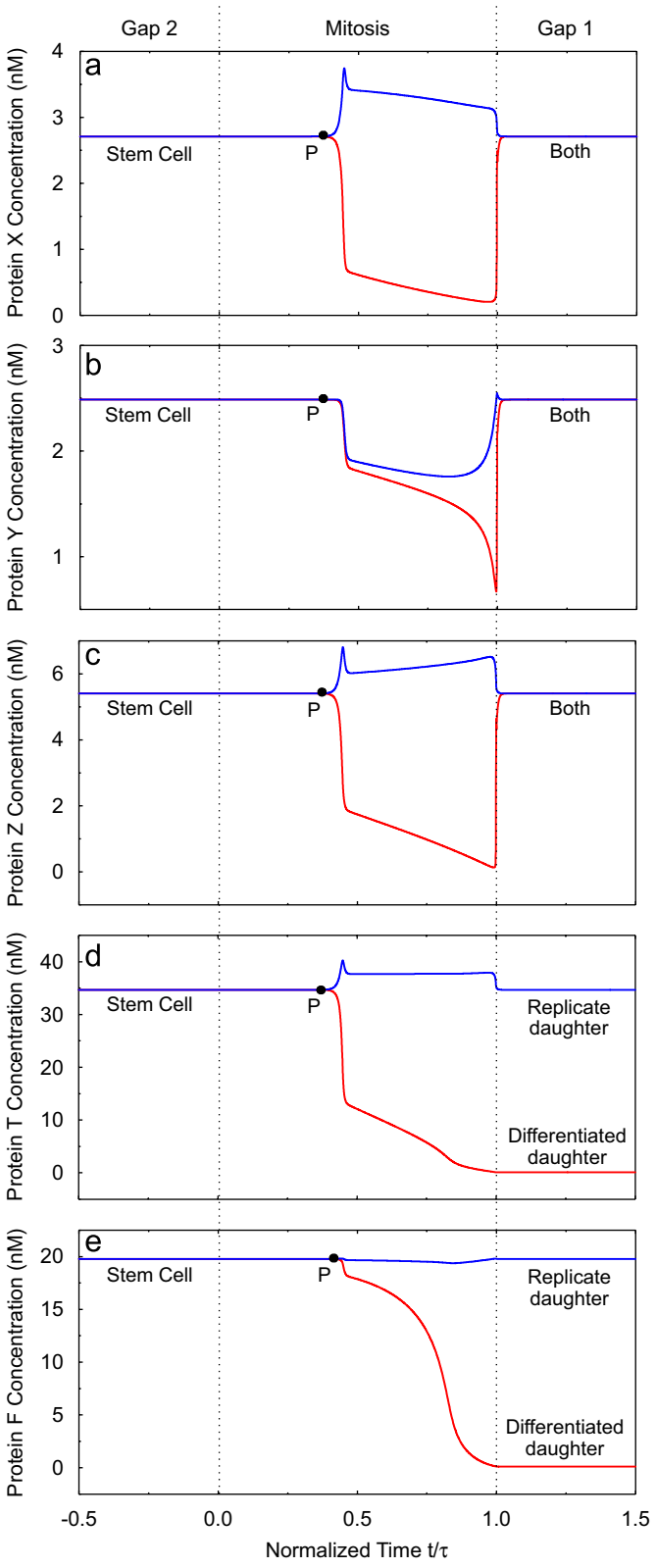

Fig. 3. Dynamic simulations of protein concentrations in two interacting daughter cells as in Fig. 2. These time series show how differences in the daughters develop dynamically. Proteins X, Y, Z, T and F profiles versus normalized time t/τ are shown in graph (a), (b), (c), (d), and (e), respectively. The blue line shows the time series for the replicate pathway; red lines show the differentiation pathway. Point P is the time when symmetry-breaking starts. It happens slightly later than the bifurcation point A, because the system must accumulate enough fluctuation in order to get out of the unstable symmetric steady state. After division is complete, the levels of proteins X, Y, and Z in both daughters return to their initial values in the original stem cell. In contrast, proteins T and F levels differ in the replicate and differentiated daughters. Here, the noise intensity $D = 10^{-6}$. Here, we set cytokine concentration $C = 1.0\,\text{nM}$. Simulation starts from the higher steady state.

In arriving at the bifurcation diagram (Fig. 2), TTP processes are assumed to be relatively fast compared to other processes that dictate the rate decrease in the cell-to-cell communication. Dynamical simulations of the protein concentrations in the two daughter cells during division are shown in Fig. 3. Here, Gaussian white noise has been added in each differential equation to trigger an initial difference. $noise = D \times \delta$, D is intensity of noise in the net reaction rate, and δ is a random number that follows Gaussian distribution.

Small (as small as $D = 10^{-10}$) random noise can trigger the symmetry-breaking of the expression of protein X, and therefore initiates stem cell differentiation process. In reality, asymmetry does not develop instantaneously when the bifurcation point (point A) is traversed; rather it occurs gradually, being completed at point P in Fig. 3. Asymmetry onset time (normalized time value of point P) for different noise intensities were recorded and shown in Fig. 4a. It is clear that the larger the noise intensity, the earlier the asymmetry occurs. This implies that the system uses fluctuations to initiate asymmetric division. To ensure the deliverance of state E to state G shown in Fig. 2d, state G must have a large enough basin of attraction. We probed the size of basin of attraction by running 1000 different random noise scenarios, and counted the percentage of successful deliverances (and therefore differentiation). When noise intensity $D \leqslant 10^{-2}$, 100% of the stem cell differentiated. This means the basin of attraction of state G is larger. When noise intensity is greater than 10^{-2} the

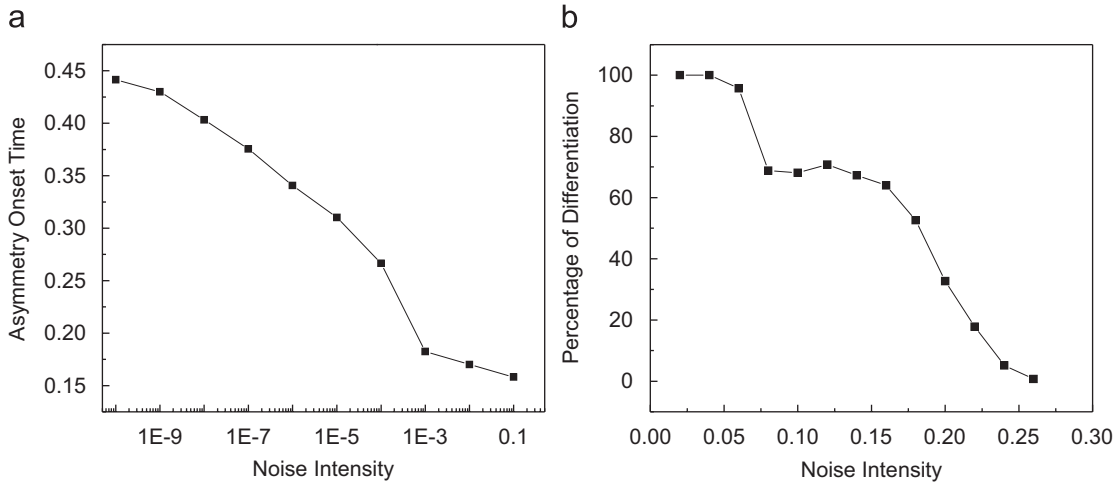

Fig. 4. Noise intensity influence (a) asymmetry onset time (the normalized time at Point P in Fig. 3) with noise intensity. The larger the noise intensity, the earlier is the asymmetry onset. (b) Percentage of successful deliverance (therefore differentiation) from state E to state G with noise intensity. The larger the noise intensity, the less the chance the stem cell differentiates.

percentage of deliverance decreases as noise intensity increases. The result is shown in Fig. 4b.

These results clearly show that (1) small random fluctuation can trigger symmetry-breaking of the system and (2) the large basin of attraction of state G ensures the successful deliverance from state E to state G. As the ubiquity of small random fluctuation in a real biological system (notably concentration fluctuations of low population species such as RNA), these dynamical simulations illustrate the robustness of the symmetry-breaking self-organization mechanism of asymmetric stem cell division/differentiation scenario.

3.2. Microenvironment cytokine regulation

To enable a stem cell differentiation scenario, two key processes are required: (1) the symmetry-breaking of the expression of protein T, which is controlled by the symmetry-breaking of the expression of protein X, as we stressed in Section 2; and (2) the deliverance of protein T expression from one stable steady state E to another previously unconnected stable steady state G (Fig. 2d), which is controlled by the basin of attraction of state G, as we discussed above. In order to illustrate the effect of extracellular conditions on regulating stem cell replication/differentiation, we study the relationship between the microenvironmental cytokine concentration and these two key processes. We consider how the protein expressions, especially that of protein T, changes as the microenvironmental cytokine concentration changes. Bifurcation and dynamical simulations of protein T concentration under different cytokine concentration values are shown in Figs. 5 and 6. We divide each bifurcation graph into different zones (defined in Table 2). Symmetry-breaking instability happens when the stem cell starts in a stable state (zone I), as division progresses, the two daughter cells evolve into two distinct stable states (zone III); and deliverance happens when the stem cell passes through a zone that contains both symmetric and asymmetric stable states (zone II), and ultimately, one of them transitions to the original stem cell state (zone I) while the other is committed to a previously unconnected stable state (also zone I). That is to say, in order to have differentiation, a stem cell has to pass through zones I, III, II and return to I during division. According to this criteria and graphs shown in Fig. 5, a stem cell replicates when cytokine concentration $C = 0.25$, 0.90, or 1.60 nM, but differentiates when $C = 1.00$ nM (reason shown in Table 3). These predictions have been verified by the dynamical simulations shown in Fig. 6.

We also made a 2-D bifurcation diagram depicting the effect of conditions in the extracellular cytokine concentration (C)/cell–cell communication (h) (Fig. 7), by combining a number of one dimensional bifurcation diagrams as in Fig. 5 along various trajectories in the C/h plane. Behavioral zones (defined in Table 2) are also indicated. Based on the criteria we mentioned above, only when the extracellular cytokine concentration is between ~0.95 and ~1.1 nM (as is ΔC shown in Fig. 7), which is neither too large nor too small, a stem cell differentiates, otherwise the stem cell will replicate. We believe similar 2-D bifurcation diagrams will be very useful in predicting stem cell behavior under various microenvironmental conditions.

All dynamical simulations (Figs. 3 and 6) start from the higher unconnected steady state (if there are two or more disconnected steady state). Interestingly, (1) the number of unconnected steady state is influenced by microenvironmental condition, see Fig. 5, when cytokine concentration is too low, there is only one unconnected steady state Fig. 5a and b; (2) when there are two attractors supported by our model for the single cell or symmetric state with appropriate cytokine concentration, see Fig. 5c, it seems that only the higher one of them could act as a stem cell

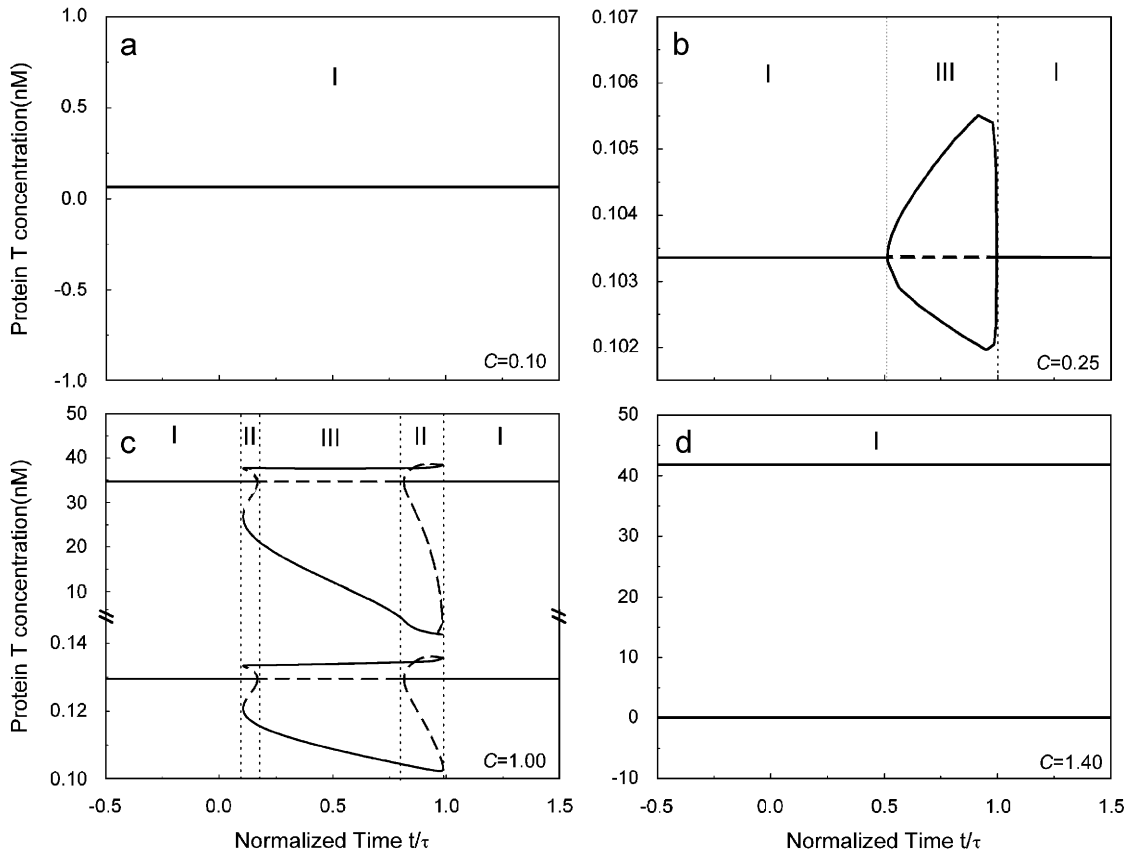

Fig. 5. Bifurcation diagrams of protein T expression under different cytokine concentrations: (a–d) show the bifurcation of protein T concentrations versus the permeability for microenvironmental cytokine concentrations $C = 0.10$, 0.25, 1.00 and 1.40 nM, respectively. Solid lines are stable steady states, and dashed lines are unstable steady states. Each graph is divided into different zones by dotted lines. Zones are defined in Table 2.

with the option of dividing symmetrically or asymmetrically. The lower one, on the contrary, does not appear to enable differentiation. This may suggest that (1) the stem cell potentiality is inherited in at the systems level and not simply by genetics; and (2) it is difficult for a differentiated daughter to return to its progenitor stem cell.

3.3. TTP5 motif in human cells and their relation to cancer

GenDat (https://systemsbiology.indiana.edu) is a database containing 3058 genes, 1187 TFs and 6166 experimentally verified and predicted gene/TF interactions for human cells. The HPRD (www.hprd.org) contains 37,385 pairs of protein–protein interactions for human cells. We search these two databases for stem cell TTP5 motifs and 12 such networks were found (see Appendix A). We believe that these genes are most likely to be involved in self-organized differentiation, for instance, gene RELB and NFKB1 which encode the protein components of the key TF (NF-kB) that is known to control T/B cell differentiation through Notch signaling pathway in hematopoietic stem cell system (Bray, 2006; Osborne and Minter, 2007). This set of networks provides guidance for targets of future RNA and protein expression experiments that delineate human stem cell profiles. Given knowledge of the TFs that regulate the involved genes (see GenDat), our result provides ways to direct stem cell behavior. Gene functions are from GO website (http://www.geneontology.org) and NCBI website (http://www.ncbi.nlm.nih.gov) and were integrated in GenDat.

To validate our TTP5 network, we made two types of random networks, in random network type 1, we fixed the total number of genes, TFs and the number of up/down interactions for each gene to be the same as the actual network we used, and randomly chose TFs regulate each gene. In 15,000 constructed random networks, the probability to find 10 or more TTP5 motifs is less than 5.8% (Fig. 8a). Type 2 network is more randomized than type 1, we fixed the total number of genes, TFs and up/down interactions, and randomly assign up/down interactions for all genes/TFs, thus the number of genes with given number of TFs regulating them is not constrained. The probability to find six or more TTP5 motifs is only 6.67×10^{-5} (Fig. 8b). From Fig. 8 we see, for a network chose at random, the probability that it support N TTP5 motifs decrease with N, and for a network with the least human TRN character is least likely to support TTP5 motifs. These results show that the TTP5 networks in the human TRN is over-represented relative to random networks.

As experimental evidence suggests the existence of cancer stem cells (Jordan et al., 2006; Tannishtha et al., 2001), common properties of stem and cancer cells suggest that

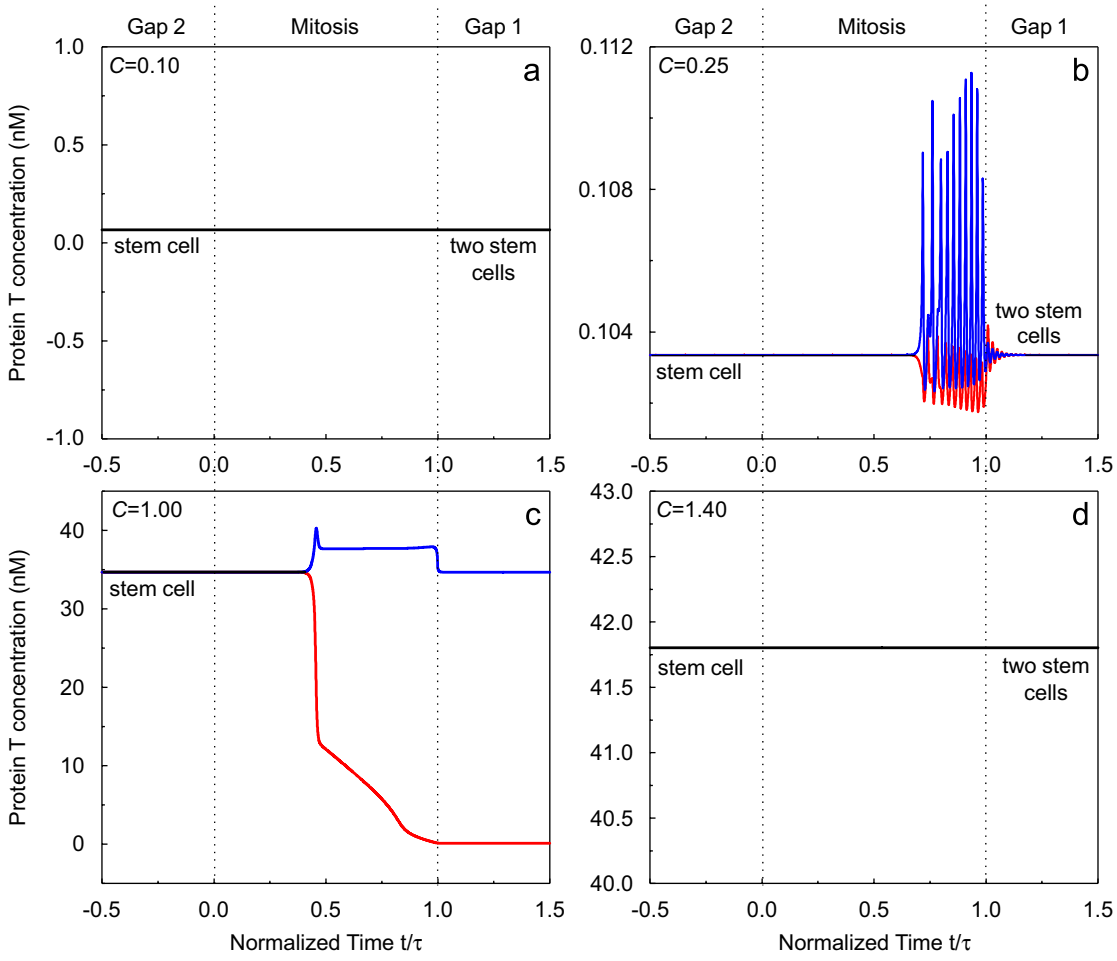

Fig. 6. Dynamical simulations of protein T expression under different cytokine concentrations: (a–d) dynamical simulations of protein T concentrations versus the permeability, for microenvironmental cytokine concentrations $C = 0.10$, 0.25, 1.00 and 1.40, respectively. Simulations start from the higher steady state if there is more than one steady state. Blue lines show the time series of protein T expression in the replicate cell and red lines showing that of the differentiated cell. Black lines indicate that the protein T expressions in both daughter cells are the same. For cytokine concentrations $C = 0.10$, 0.25 and 1.40 nM, the stem cell replicates. For cytokine concentration $C = 1.00$ nM, the stem cell differentiates. The dynamical simulations verified the predictions made on the bifurcation diagrams in Fig. 5. See Section 3 and Table 3 for further discussions.

Table 2
Definition of the types of steady states supported by domains in Figs. 5 and 7

Zone	Type of steady states supported in each unconnected branch
I	One stable symmetric steady state
II	One stable symmetric steady state and two stable asymmetric steady states and two unstable asymmetric steady states
III	Two stable asymmetric steady states and one unstable symmetric steady state

A stem cell can support multiple unconnected branches of steady states (H1). In this case, both of them share the same boundaries between these three defined zones. Note that in zone I, one of the stable states corresponds to the stem cell while the other is the yet-unrealized differentiated state.

Table 3
Stem cell fate under different cytokine concentrations

Cytokine concentration, C (nM)	Stem cell fate	Reason
0.10	Replication	Only across zone I during division
0.25	Replication	Not across zone II during division
1.00	Differentiation	Across all the required zones
1.40	Replication	Only across zone I during division

some mechanisms of stem cell replication/differentiation can be discovered via studies of cancer, particularly cancer-related genes and their interactions. To investigate this hypothesis, we initiated a search to determine whether oncogenes and tumor repressor genes (cancer genes) (http://embryology.med.unsw.edu.au/DNA/DNA10.htm) are significantly over-represented in our TTP5 networks. Cancer genes are observed more often in the G_X, G_T, and

G_Y gene categories. We see (Fig. 9) 50% (6 out of 12) G_X genes are cancer genes, and the probability to find cancer genes in G_T group is even greater, which is 57.9% (11 out of 19), whereas, only 2.65% (81 out of 3058) of the genes in the database are cancer genes. The result implies that many genes involve in stem cell differentiation might also be involved in cancer, and that cancer cells' proliferation/differentiation might have some of the same mechanism as that of stem cells.

One may suspect that the database we use contains much more information on cancer genes and therefore cancer genes are over-represented among highly connected genes, in other words, highly connected genes are more likely to be cancer genes. Since G_X, G_Y, G_Z, G_T genes (defined here as core genes) are highly connected, it is not surprising that cancer genes are over-represented among these core gene groups. Does that mean the high probability of cancer genes among the core genes we found through our TTP5

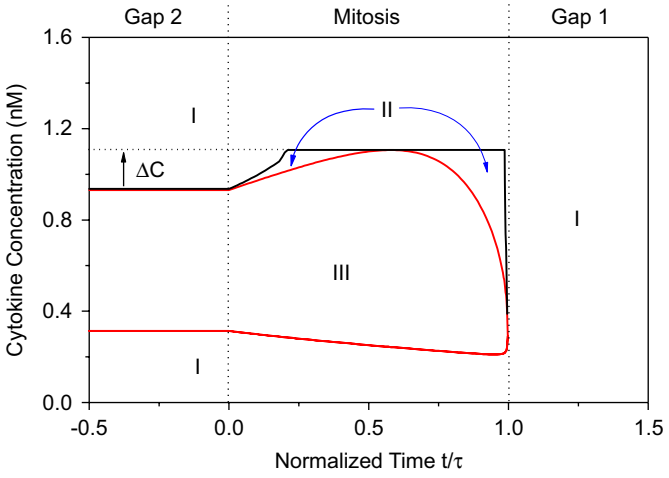

Fig. 7. Extra-cellular cytokine concentration/cell–cell communication 2-D bifurcation diagram. Influence of cytokine concentration (from 0 to 0.16 nM) to the stem cell differentiation (from late Gap 2 to early Gap 1) is shown. Solid lines indicate the boundaries of the different behavior zones, as defined in Table 2. The red line shows the location of the bifurcation points. A stem cell differentiates as it passes through zones I, III, II and returns to zone I during division. In this case, only when extra-cellular cytokine concentration lies between ~0.95 and ~1.1 nM, will the stem cell differentiate, otherwise, the stem cell will just replicate.

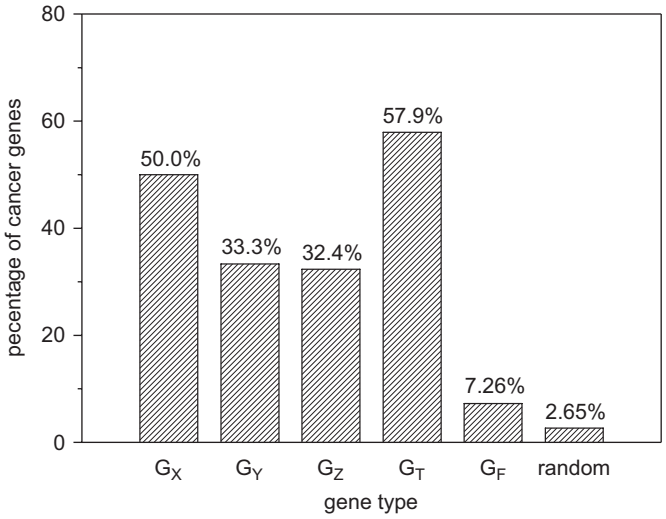

Fig. 9. Percentage of cancer genes in gene category: we searched our TRN database for TTP5 motifs and counted the number of genes that are oncogenes or tumor repressor genes (cancer genes). We found that the probability to find a cancer gene in gene category G_X is 50.0% (6 out of 12), G_Y is 33.3% (6 out of 18), G_Z is 32.4% (11 out of 34), G_T is 57.9% (11 out of 19) and G_F is 7.26% (40 out of 551). For a randomly selected gene, the probability that it is a cancer gene is only 2.65% (81 out of 3058). Cancer genes are clearly over-represented in category G_X and G_T.

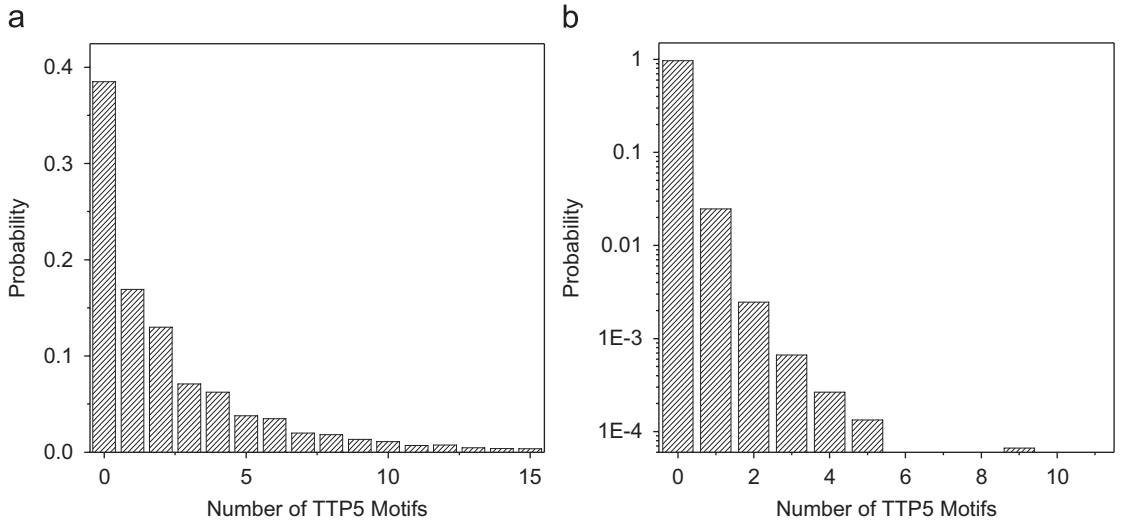

Fig. 8. Probability of finding a given number of TTP5 motifs in a random network: (a) random network generated by fixing the total number of genes, TFs and the number of gene/TF up/down interactions for each gene to be the same as the actual network we used, and randomly chose TFs regulate each gene. The probability to find 10 or more TTP5 motifs is less than 5.8%; (b) random network generated by fixing the total number of genes, TFs and gene/TF up/down interactions, and randomly assign up/down interactions for all the genes/TFs. The probability to find six or more TTP5 motifs is only 6.67×10^{-5}.

Table 4
Connectivity distributions of all genes, cancer genes, core genes and core cancer genes for a given number of connections

Connections	NG1	F1	NG2	F2	NG3	F3	NG4	F4
1–5	2658	0.8692	42	0.51852	13	0.28889	1	0.0625
6–10	222	0.0726	14	0.17284	6	0.13333	1	0.0625
11–15	71	0.02322	2	0.02469	2	0.04444	0	0
16–20	34	0.01112	5	0.06173	3	0.06667	2	0.125
21–25	18	0.00589	3	0.03704	3	0.06667	1	0.0625
26–30	11	0.0036	1	0.01235	2	0.04444	0	0
31–35	6	0.00196	1	0.01235	1	0.02222	1	0.0625
36–40	4	0.00131	1	0.01235	1	0.02222	0	0
41–45	7	0.00229	1	0.01235	1	0.02222	1	0.0625
46–50	4	0.00131	1	0.01235	0	0	0	0
51–55	3	9.81033E−4	1	0.01235	1	0.02222	0	0
56–60	1	3.27011E−4	0	0	0	0	0	0
61–up	19	0.00621	9	0.11111	12	0.26667	9	0.5625

Here, we define G_X, G_Y, G_Z, G_T genes that follow our TTP5 motif are core genes. This table contains nine columns: the first column is the range of connections and the other columns are: NG1, the number of genes in our database that have connections in each range; F1, the fraction of genes in our database that have connections in each range; NG2, the number of cancer genes in our database that have connections in each range; F2, the fraction of cancer genes in our database that have connections in each range; NG3, the number of core genes that have connections in each range; F3, the fraction of core genes that have connections in each range; NG4, the number of core cancer genes with connections in each range; and F4, the fraction of core cancer genes that have connections in each range.

motif is coincidently caused by the high connectivity of those genes instead of any cancerous or other biological mechanisms? In order to clarify these issues, we made Table 4 showing the connectivity distribution for all genes, cancer genes, core genes, and core cancer genes.

The TRN we use contains 3058 genes, 81 of which are cancer genes (there are 159 known cancer genes in total, http://embryology.med.unsw.edu.au/DNA/DNA10.htm); we found 45 core genes that follow our TTP5 motif, 16 of which are cancer genes (therefore core cancer genes). From column NG1 and F1 in Table 4, we see there are only a few genes in the TRN we use are higher connected, and the majority of genes have connections of five or less (at a fraction of 0.8692). Although there are some genes that are highly connected (19 genes have connections of 61 or more), the fraction is low (0.00621). From column NG2 and F2, we see cancer genes are not necessarily highly connected; actually, the number of connections of cancer genes is quite a wide spectrum. There are over 50% cancer genes having connections of five or less, although there are also nine cancer genes that are highly connected (with a fraction of 0.1111). Comparing NG2/F2 to NG1/F1, we see (1) the average connectivity of cancer genes is slightly larger than that of all the genes and (2) the percentage of cancer gene is higher in highly connected genes than in less connected genes (42 cancer genes in 2658 genes with connections of five or less, but nine cancer genes in 19 genes with connections of 61 or more). Comparing NG3/F3 to NG2/F2, we see number of connections of core genes have a similar spectrum to that of the cancer genes with a slightly higher average number of connections (13 genes having connections of five or less, with a fraction ~0.29; 12 genes having connections of 61 or more, with a fraction ~0.27).

The above results prove that (1) core genes we located through our TTP5 mechanism are not all highly connected, instead the number of TF mediated connections of these genes vary from 1 to 61 or more; (2) although the highly connected genes are more probable to be cancer genes, cancer genes are not necessarily to be highly connected genes (see NG2 in Table 4); (3) most importantly, one would be able to locate cancer genes by searching for extremely highly connected genes in a network, but he will lose the majority of cancer genes that are much less connected, however, using our TTP5 mechanism, we could not only locate most of the highly connected cancer genes (here 9 out of 9), but also some of the much less connected ones (see NG4 in Table 4), which would have been missed via a search that assumed a strong coloration between genes with high connectivity and those which are cancer-related. Therefore, we believe that the over-representation of cancer genes in our TTP5 genes is biologically meaningful and our methodology to locate these genes is valid.

4. Discussion

4.1. Microenvironmental control of self-organized asymmetric division

In the present theory, differentiation accompanying division creates daughter cells in distinct stable states through "symmetry-breaking instability". Symmetry-breaking is a general phenomenon that can take place in systems involving two or more subsystems (i.e. daughters of a dividing stem cell). Conditions necessary for this self-organized asymmetry include nonlinearity in the reaction-transport processes, maintenance of far-from-equilibrium conditions, exchange of mass and energy between the

subsystems and feedback mechanisms among the operating processes (see the monographs Ortoleva, 1992; Ortoleva et al., 1994). A commonly quoted scenario is that the feedback in a multi-compartmented system amplifies omnipresent fluctuation into the symmetry-broken (i.e. patterned) state. However, the familiar reaction-transport laws are meant to apply to the average behavior of a system; as the level of noise increases, one might expect that the lows of average behavior change. In the case of the stem cell as modeled here, we found that this leads to a more complex role of noise. As noise increases from zero, the onset time of asymmetry decreases as one might expect. However, as noise increases further, the fraction of daughter cells that are delivered to different state actually decreases. This implies that the decision for a stem cell to divide symmetrically or asymmetrically depends on the noise in the environment. This could yield the evolutionary advantage for stem cell to have a higher percentage of replication relative to asymmetric division when uncertainties in the surroundings are high—i.e. the multi-cellular system should not put too much reliance on only a few stem cells in an uncertain environment.

A human cell has roughly 25,000 genes and an extremely complex network of nonlinear TTP processes. This network can support a tremendous number of distinct stable states (Qu et al., 2007). We believe it may be the special structure of a stem cell's TRN that enables symmetry-breaking during its division, and allows a stem cell's great potential to differentiate into a myriad of cell types. Therefore, constructing a human stem cell TRN using the methods introduced here and compare its implications with observed stem cell behavior seems to hold great promise for developing a predictive model for stem cell differentiation. It is appropriate to assume that while a stem cell has a number of special characteristics, the overwhelming majority of its TRN is common to many human cell lines. Thus we have constructed our stem cell TRN from data on a typical human cell but have sought special subnetwork motifs within it that have self-organizational potential. Proceeding in this way, we identified 12 motifs that could support taking stem cell differentiation from a genome-wide perspective.

A stem cell's microenvironment is critical in initiating its decision to replicate or differentiate (Li and Neaves, 2006; Fuchs et al., 2004; Tumbar et al., 2004; Rizvi and Wong, 2005). Multiple contributing influences are relayed via signaling pathways to the nucleus where transcriptional responses are elicited. The many factors in the microenvironment, the complexity of signaling pathways, and the vast scale of regulatory networks make delineating the conditions that favor a given scenario of stem cell behavior a grand challenge. For example, a given human gene is activated or deactivated by from 1 to several dozen different TFs. These TFs are directly or indirectly affected by many factors in the intracellular medium (e.g. proteins, enzymes, nutrients) and microenvironment (e.g. cytokine, estrogen, vitamin D). Thus, there are likely hundreds factors that can direct a stem cell to divide symmetrically or asymmetrically. Hence, delineating the favorable conditions for a given differentiation scenario is to discover zones in a 100-dimensional space of control parameters, a task as difficult as finding the proverbial needle in a 100-dimensional "hay stack".

In this study, we developed a theory that seems to explain how a stem cell can be directed to divide symmetrically or asymmetrically by the microenvironment, and provided a practical computer-aided methodology to discover ways to control stem cell differentiation. In our theory, there are subnetworks of genes whose TTP interactions enable symmetry-breaking and self-organization of the dividing stem cell into a replicate daughter and a differentiated daughter. The stem cell's decision is regulated by its microenvironment. We also developed and demonstrated an automated procedure for discovering self-organizing motifs and analyzing their consequences. This methodology holds great promise to solve complex problems in pure and applied stem cell research.

4.2. Gene functional category and the TTP5 motif

Molecules involved in stem cell differentiation have been classified into functional categories, including cell signaling, cell-cycle inhibitors, basal lamina and extra-cellular matrix molecules (Rizvi and Wong, 2005). In the TTP5 motif (see Appendix A), G_Y is involved in signaling pathways, and combines with G_Z and helps G_X to break symmetry, which then causes G_T and G_F to be differently expressed in the two daughter cells and eventually differentiates them. In our model G_X, G_Y, G_Z and G_T are the core genes that control/deliver cell differentiation, and the functions of these genes are usually transcription/translation regulation, signaling transduction and protein binding. G_F genes (available at GenDat) are controlled by the core genes and determine cell state, notably the phenotype of the differentiated daughter. Within each TTP5 network, there are hundreds of G_F genes that cover various functions from RNA binding, protein synthesis, intracellular signaling and cell surface reception to metabolism, cell growth, cell cycle/differentiation/proliferation regulation, etc. Thus, the different expressions of G_F genes distinguish the functionality of the two daughters.

A direct consequence of our theory is that not all genes are differently expressed in the differentiated daughters. Although this prediction is consistent with several experimental studies (Ramalho-Santos et al., 2002; Ivanova et al., 2002; Luo et al., 2002), comparing microarray data on a human stem cell and its differentiated daughter in detail (gene by gene) will be critical to prove/disprove our theory. However, the experimentally verified regulatory interaction for genes known to be involved in stem cell processes does not seem to be available in the literature. As a minimal amount of such data is needed for an analysis, such validation was not possible yet. This highlights a critical

research need that could be met by ChIP-on-chip or RT-PCR studies.

4.3. Strategy to discover controls on stem cell differentiation

Difficulties in control of stem cell behavior for clinical applications and in attaining a detailed knowledge of the operating processes involved arise from the complexity of the regulatory networks, signaling pathways and the many influential factors in their microenvironment. Our strategy for addressing these difficulties was to integrate as much of the intracellular and microenvironmental information as possible into a predictive TTP kinetic model. Predictions on transcriptional interactions were made using a training set of regulatory information, a microarray data-based information theory/chemical kinetic approach (Sayyed-ahmad et al., 2007), and a GO similarity bioinformatics method (Tuncay et al., 2006; Sun et al., 2007), all integrated via a Bayesian methodology (Sun et al., 2007). GO similarity and HPRD protein–protein interaction database (Peri et al., 2003; www.hprd.org) were used to suggest post-translational reactions and gene functions, while microarray data and cell biological observations were used to calibrate the parameters in the TTP model. Having reconstructed the TTP network and calibrated the rate parameters, computational simulation enabled the computer-aided discovery of genes and TFs involved in, and conditions favoring, a given stem cell's behavioral scenario. Our preliminary study demonstrates that such a computational approach holds great promise to delineate the mechanism of, and to discover controls on, stem cell replication/differentiation.

4.4. Relation to cancer

To probe the TTP5$\rightleftharpoons$cancer relationship, we constructed a probability measure, i.e. the percentage of five gene groups we discovered from GenDat with a list of known oncogenes and tumor repressor (henceforth cancer) genes. The percentages of G_X, G_Y, G_Z, G_T and G_F genes and randomly selected genes that are cancer genes, are given in Fig. 9. We see that the probability of G_X and G_T genes that are cancer genes greatly exceeds that for genes chosen at random. It is important to emphasize that the TTP5 genes were identified via our TRN subnetwork structure, and were not chosen due to their GO description or connectivity. As in our earlier study (Qu et al., 2007), it appears that the concept of a "cancer gene" should be replaced with that of an "onconetwork", i.e. a subset of genes whose regulatory interactions creates a feedback loop that enables proliferation, thwarts apoptosis or can drive the cell into another distinct, abnormal state of the TTP system. Since the over-representation of TTP5 as a likely onconetwork motif is clear, we suggest that those genes in TTP5 networks not yet classified as oncogenes or tumor repressor genes should also be identified as cancer related.

Acknowledgments

We thank Major L. Viveros (US Air Force) and Dr. N. Kelley-Loughnane (US Air Force) for helpful discussions. L. Ensman and M. Trelinski for assistance in writing the TTP5 discovery software, and D. Ha and K. Riley for help in the signaling pathway research. We thank an anonymous referee for helpful suggestions in improving the manuscript. This project is supported in part by the United States Department of Energy (Genomics: Genome to Life Program), Oak Ridge Institute for Science and Education (Student Research Participation Program), and the College of Arts and Sciences (Indiana University) as general support of the Center for Cell and Virus Theory.

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jtbi.2007.10.019.

References

Artavanis-Tsakonas, S., Rand, M.D., Lake, R.J., 1999. Notch signaling: cell fate control and signal integration in development. Science 284, 770–776.

Atala, A., 2006. Recent developments in tissue engineering and regenerative medicine. Curr. Opin. Pediatr. 18 (2), 167–171.

Borisuk, M.T., Tyson, J., 1998. Bifurcation analysis of a model of mitotic control in frog eggs. J. Theor. Biol. 195, 69–85.

Bray, S.J., 2006. Notch signaling: a simple pathway becomes complex. Nat. Rev. Mol. Cell Biol. 7, 678–689.

Broxmeyer, H.E., et al., 2006. Cord blood stem and progenitor cells. Methods Enzymol. 419, 439–473.

Chickarmane, et al., 2006. Transcriptional dynamics of the embryonic stem cell switch. PLoS Comput. Biol. 2 (9), e123.

Cinquin, O., Demongeot, J., 2005. High-dimensional switches and the modeling of cellular differentiation. J. Theor. Biol. 233, 391–411.

Doedel, E.J., Keller, H.B., Kernevez, J.P., 1991. Numerical analysis and control of bifurcation problems, Part I & Part II. Int. J. Bifurc. Chaos 1, 493–520, 745–772.

Fuchs, E., Tumbar, T., Guasch, G., 2004. Socializing with the neighbors: stem cells and their niche. Cell 116, 769–778.

Furusawa, C., Kaneko, K. Theory of robustness of irreversible differentiation in a stem cell system: chaos hypothesis. J. Theor. Biol. 209, 395–416.

Gaiano, N., Fishell, G., 2002. The role of Notch in promoting glial and neural stem cell fates. Annu. Rev. Neurosci. 25, 471–490.

Hahn, H.-S., Ortoleva, P., Ross, J., 1973. Chemical oscillations and multiple steady states due to variable boundary permeability. J. Theor. Biol. 41, 503–521.

Haramis, A.P., et al., 2004. De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. Science 303, 1684–1686.

Hasty, J., Collins, J.J., et al., 2000. Noise-based switches and amplifiers for gene expression. Proc. Natl. Acad. Sci. 97, 2075–2082.

Hasty, J., Collins, J.J., et al., 2001. Designer gene networks: towards fundamental cellular control. Chaos 11, 207–220.

Heng, B.C., Cao, T., Lee, E.H., 2004. Directing stem cell differentiation into the chondrogenic lineage in vitro. Stem Cells 22, 1152–1167.

Ivanova, N.B., et al., 2002. A stem cell molecular signature. Science 298, 601–604.

Jordan, C.T., Guzman, M.L., Nobel, M., 2006. Cancer stem cells. N. Engl. J. Med. 355, 1253–1261.

Kirschstein, R., et al., 2001. Stem cells: scientific progress and future research directions. NIH Report, 〈http://stemcells.nih.gov/info/scireport〉.

Larter, R., Ortoleva, P., 1982. A study of instability to electrical symmetry-breaking in unicellular systems. J. Theor. Biol. 96, 175–200.

Li, L., Neaves, W.B., 2006. Normal stem cells and cancer stem cells: the niche matters. Cancer Res. 66, 4553–4557.

Loeffler, M., Roeder, I., 2002. Tissue stem cells: definition, plasticity, heterogeneity, self-organization and models—a conceptual approach. Cells Tissues Organs 171, 8–26.

Luo, Y., et al., 2002. Microarray analysis of selected genes in neural stem and progenitor cells. J. Neurochem. 83, 1481–1497.

Masson, S., et al., 2004. Potential of hematopoietic stem cell therapy in hepatology: a critical review. Stem Cells 22, 897–907.

Nicolis, G., Prigogine, I., 1977. Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuations. Wiley-Interscience, New York.

Odorico, J.S., Kaufman, D.S., Thomson, J.A., 2001. Multilineage differentiation from human embryonic stem cell lines. Stem Cells 19, 193–204.

Ortoleva, P., 1992. Nonlinear Chemical Waves. Wiley, New York.

Ortoleva, P., Ross, J., 1973a. A theory of asymmetric cell division (differentiation). Dev. Biol. 34, F19–F23.

Ortoleva, P., Ross, J., 1973b. A chemical instability mechanism for asymmetric cell differentiation. Biophys. Chem. 1, 87–96.

Ortoleva, P., Chen, Y., Chen, W., 1994. Agates, geodes, concretions and orbicules: self-organized zoning and morphology. In: Kruhl, J.H. (Ed.), Fractals and Dynamic Systems in Geoscience. Springer, New York, pp. 283–305.

Osborne, B.A., Minter, L.M., 2007. Notch signaling during peripheral T-cell activation and differentiation. Nat. Rev. Immunol. 7, 64–75.

Passegue, E., et al., 2003. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics. Proc. Natl. Acad. Sci. 100, 11842–11849.

Peri, S., et al., 2003. Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Research 13, 2363–2371.

Polakis, P., 2000. Wnt signaling and cancer. Genes Dev. 14, 1837–1851.

Proudfoot, N., 2000. Connecting transcription to messenger RNA processing [Review]. Trends Biochem. Sci. 25 (6), 290–293.

Qu, K., et al., 2007. Cancer onset and progression: a genome-wide, nonlinear dynamical systems perspective on onconetworks. J. Theor. Biol. 246, 234–244.

Ramalho-Santos, M., et al., 2002. ''Stemness'': transcriptional profiling of embryonic and adult stem cells. Science 298, 597–600.

Rashevsky, N., 1960. , third ed. Mathematical Biophysics: Physico-Mathematical Foundations of Biology, vol. 1. Dover Publications, New York.

Rizvi, A.Z., Wong, M.H., 2005. Epithelial stem cells and their niche: there is no place like home. Stem Cells 23, 150–165.

Roeder, I., Glauche, I., 2006. Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors GATA-1 and PU.1. J. Theor. Biol. 241, 852–865.

Sayyed-ahmad, A., Tuncay, K., Ortoleva, P., 2007. Transcriptional regulatory network refinement and quantification through kinetic modeling, gene expression micorarray data and information theory. BMC Bioinformaitcs 8, 20.

Slutsky, M., Mirny, L.A., 2004. Kinetics of protein–DNA interaction: facilitated target location in sequence-dependent potential. Biophys. J. 87, 4021–4035.

Sun, J., et al., 2007. Transcriptional regulatory network discovery via multiple method integration: application to *E. coli* K12. Algorithms Mol. Biol. 2, 2.

Tannishtha, R., et al., 2001. Stem cell, cancer and cancer stem cell. Nature 414, 105–111.

Tenen, D.G., 2003. Disruption of differentiation in human cancer: AML shows the way. Nat. Rev. 3, 89–101.

Theise, N.D., d'Inverno, M., 2004. Understanding cell lineages as complex adaptive systems. Blood Cells Mol. Dis. 32, 17–20.

Tumbar, T., et al., 2004. Defining the epithelial stem cell niche in skin. Science 303, 359–363.

Tuncay, K., et al., 2006. Transcriptional regulatory networks via gene ontology and expression data. In Silico Biol. 7, 0003.

Turing, A., 1952. The chemical basis of morphogenesis. Phil. Trans. R. Soc. London B 327, 37{72}.

Uetz, P., et al., 2000. A comprehensive analysis of protein±protein interactions in *Saccharomyces cerevisiae*. Nature 403, 623–627.

Ujvaro, A., et al., 2001. Translation rate of human tyrosinase determines its N-linked glycosylation level. J. Biol. Chem. 276 (8), 5924–5931.

Valles, J.M., 2002. Model of magnetic field-induced mitotic apparatus reorientation in frog eggs. Biophys. J. 82, 1260–1265.

Villavicencio, E.H., Walterhouse, D.O., Iannaccone, P.M., 2000. The sonic hedgehog-patched-gli pathway in human development and disease. Am. J. Hum. Genet. 67, 1047–1054.

Vu, T.T., Vohradsky, J., 2007. Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of *Saccharomyces cerevisiae*. Nucleic Acids Res. 35, 279–287.

Wodarz, A., Nusse, R., 1998. Mechanisms of Wnt signaling in development. Annu. Rev. Cell. Dev. Biol. 14, 59–88.

Yu, J., Domen, J., Panchision, D., Zwaka, T.P., Rohrbaugh, M.L., et al., 2006. Regenerative Medicine. NIH Report, 〈http://stemcells.nih.gov/info/scireport〉.

Zak, D.E., et al., 2003. Continuous-time identification of gene expression models. OMICS: J. Integrative Biol. 7, 373–386.

Zhao, M., Forrester, J.V., McCaig, C.D., 1999. A small, physiological electric field orients cell division. Proc. Natl. Acad. Sci. USA 96, 4942–4946.