

PROCEEDINGS

Open Access

Mathematical modeling of GATA-switching for regulating the differentiation of hematopoietic stem cell

Tianhai Tian*, Kate Smith-Miles

From The Twelfth Asia Pacific Bioinformatics Conference (APBC 2014)
Shanghai, China. 17-19 January 2014

Abstract

Background: Hematopoiesis is a highly orchestrated developmental process that comprises various developmental stages of the hematopoietic stem cells (HSCs). During development, the decision to leave the self-renewing state and selection of a differentiation pathway is regulated by a number of transcription factors. Among them, genes *GATA-1* and *PU.1* form a core negative feedback module to regulate the genetic switching between the cell fate choices of HSCs. Although extensive experimental studies have revealed the mechanisms to regulate the expression of these two genes, it is still unclear how this simple module regulates the genetic switching.

Methods: In this work we proposed a mathematical model to study the mechanisms of the GATA-PU.1 gene network in the determination of HSC differentiation pathways. We incorporated the mechanisms of GATA switch into the module, and developed a mathematical model that comprises three genes *GATA-1*, *GATA-2* and *PU.1*. In addition, a novel multiple-objective optimization method was designed to infer unknown parameters in the proposed model by realizing different experimental observations. A stochastic model was also designed to describe the critical function of noise, due to the small copy numbers of molecular species, in determining the differentiation pathways.

Results: The proposed deterministic model has successfully realized three stable steady states representing the priming and different progenitor cells as well as genetic switching between the genetic states under various experimental conditions. Using different values of GATA-1 synthesis rate for the GATA-1 protein availability in the chromatin sites during the time period of GATA switch, stochastic simulations for the first time have realized different proportions of cells leading to different developmental pathways under various experimental conditions.

Conclusions: Mathematical models provide testable predictions regarding the mechanisms and conditions for realizing different differentiation pathways of hematopoietic stem cells. This work represents the first attempt at using a discrete stochastic model to realize the decision of HSC differentiation pathways showing a multimodal distribution.

Background

Hematopoiesis is a highly orchestrated developmental process that comprises the proliferation, differentiation and maturation of a very small population of self-renewing, pluripotent hematopoietic stem cells (HSC) for producing different types of blood cells, including erythrocyte, megakaryocyte, granulocyte, and macrophage [1,2]. During

development, the decision to leave the self-renewing state and selection of a differentiation pathway are regulated by transcription factors (TFs) [3-6]. Intense experimental studies during the past two decades have suggested that tight regulation of HSC differentiation is controlled by the interaction of a number of genetic and epigenetic regulators of gene transcription, including the two TFs PU.1 and GATA-1. Although the precise mechanisms to initiate the transcriptional cascade leading to different differentiated cells are not clear currently, experimental studies have

* Correspondence: tianhai.tian@monash.edu
School of Mathematical Sciences, Monash University, Melbourne, VIC 3800, Australia

established that both *PU.1* and *GATA-1* 'autoregulate' themselves, i.e. they stimulate their own production, as well as they are mutually antagonistic, i.e. they repress the production of each other [7-9]. In the erythrocyte/megakaryocyte lineage high expression levels of gene *GATA-1* and low levels of *PU.1* were detected [6,10]; conversely, in the granulocyte/macrophage lineage higher expression levels of *PU.1* and low levels of *GATA-1* were measured [5,11]. However, the initial progenitor cells stay in the third state that has low-level activation of both genes *PU.1* and *GATA-1*. When a progenitor cell differentiates, it transitions from the initial indeterminate state into one of two differentiated states with high expression levels of either gene *GATA-1* or *PU.1*. Thus the full understanding of the *PU.1* and *GATA-1* interaction is important for studying of the differentiation process of HSCs, which may be very useful in the clinical application of "differentiation therapy" for re-establishing the correct expression of *PU.1* and/or *GATA-1* within immature leukemic cells [12].

To accurately describe the regulatory mechanisms controlling HSC differentiation, the important aspects that any mathematical model of the *GATA-1-PU.1* network must include are an indeterminate state for the progenitor cells and two stable attractors of the dynamical system for the differentiated lineages. Since the first modeling attempt to study the regulation in the *GATA-1-PU.1* network [13], a number of mathematical formalisms have been developed to realize the three stable steady states in HSCs. For example, Huang et al. used the Hill function with high co-operativity coefficients to qualitatively compare computer predictions with experimental evidence [14], giving support to the idea that lineage choice occurs as a two stage process, first priming and then differentiating. To remove the requirement of high co-operativity, Chickarmane et al. assumed that the autoregulation at both *PU.1* and *GATA-1* occurs through the binding of monomers [15]. It was predicted that an additional mechanism should be involved in the repressive interaction to create a bistable switch, and therefore a third unspecified gene was introduced to create bistability. Alternatively, a mathematical model was proposed to include the dynamics of the inactive heterodimer *GATA-1-PU.1* and the Michaelis-Menten function was used to represent the low co-operativity [16]. We have designed a model by separating the strength of co-operativity for autoregulation and repression, and successfully realized a rich variety of systems behavior that have been found cross the existing models [17]. Recently Huang and collaborators proposed a general model that assumed no explicit interaction between the two genes; and realized a degenerated steady state via a new type of bifurcation [18]. Since the assumptions in these models are based on either unrealistic high co-operativity or unspecified gene, additional mechanisms

are needed to accurately describe the differentiation decision of HSCs.

GATA-2 is one of the six members of the hematopoietic *GATA* factor family and is most abundantly expressed in HSCs as well as in immature progenitors in hematopoietic lineages [10]. Previous studies demonstrated a critical role of *GATA-2* in the emergence and maintenance of HSCs [19]. In addition, strict regulation of genes *GATA-1* and *GATA-2* is critical for proper lineage commitment and development of erythroid cells. It was reported that *GATA-2* directly activated *GATA-1* expression in early erythroid progenitors, and then *GATA-1* accelerated its expression after its own expression has been initiated [20]. The gene expression process involving *GATA-1*-mediated displacement of *GATA-2* from chromatin is termed a *GATA* switch [21-23]. This "*GATA* factor switch" suggests a model whereby *GATA-2* and *GATA-1* sequentially bind the same *cis*-elements with activating and repressive effects, respectively. During *GATA* factor switch, the *GATA-1* expression will increase due to the reciprocal decrease of *GATA-2*, which leads progenitor cells to adopt an erythroid lineage. However, when *GATA-1* is absent or its expression is delayed, the reduction of *GATA-2* will increase the expression levels of *PU.1* and leads progenitor cells to adopt a myeloid lineage [24,25]. Therefore the dynamic expression patterns of *GATA-1* and *GATA-2* may influence the erythroid-myeloid cell-fate selection by regulating the expression of gene *PU.1* [22]. Although accumulating experimental evidence has suggested the important role of gene *GATA-2* in regulating the cell-fate selection, only a simple Boolean network model has been proposed recently to include the regulatory function of gene *GATA-2* [26]. The kinetic dynamics of *GATA-2* and in particular the function of *GATA* switch has not been systematically studied so far.

Recent experimental studies have demonstrated that gene expression is a stochastic process. Key species of molecules such as DNA and mRNA may have small copy numbers, and the change of their copy numbers may cause significant variations of the system dynamics [27-30]. In particular, it has been shown that a variety of lineage-restricted genes in HSCs were expressed at low levels [31]. A recent study directly demonstrated that stochastic oscillation expression of lineage-associated genes could drive cell-fate commitment [32]. Accumulating experimental evidence also suggested that stem cells are heterogeneous, with cells moving between two or more metastable states [33]. Recently a minimal model has been designed by combining cell-extrinsic and cell-intrinsic elements of regulation to understand how both instructive and stochastic events could inform cell commitment decision in hematopoiesis [34]. However, compared with the advances in developing various stochastic models to investigate the key functions of noise in genetic and cellular

processes [35-37], the critical role of noise in determining the stem cell differentiation has not been well established. This work is aimed at developing the first stochastic model to explore the critical function of GATA switch and noise in determining the differentiation pathways of HSCs.

Methods

Mathematical model of GATA-PU.1 gene network

In this work we propose a mathematical model for the GATA-PU.1 regulatory network including genes *GATA-1*, *GATA-2* and *PU.1* (Figure 1A). It was assumed that each of these three genes activates their own expression through positive autoregulation. In addition, TF GATA-2 activates the expression of gene *GATA-1* but inhibits the expression of gene *PU.1*, but the function of GATA-2 to regulate the expressions of genes *GATA-1* and *PU.1* is moderate [24]. Thus the expression levels of three genes *GATA-1*, *GATA-2* and *PU.1* may maintain at an intermediate state, which is compatible to the steady state of lineage priming. In addition, GATA-1 inhibits the expression of *GATA-2* and *PU.1*, whereas PU.1 inhibits the expression of *GATA-1* and *GATA-2*. Detailed assumptions

of regulatory mechanisms and the list of chemical reactions are given in Additional file 1.

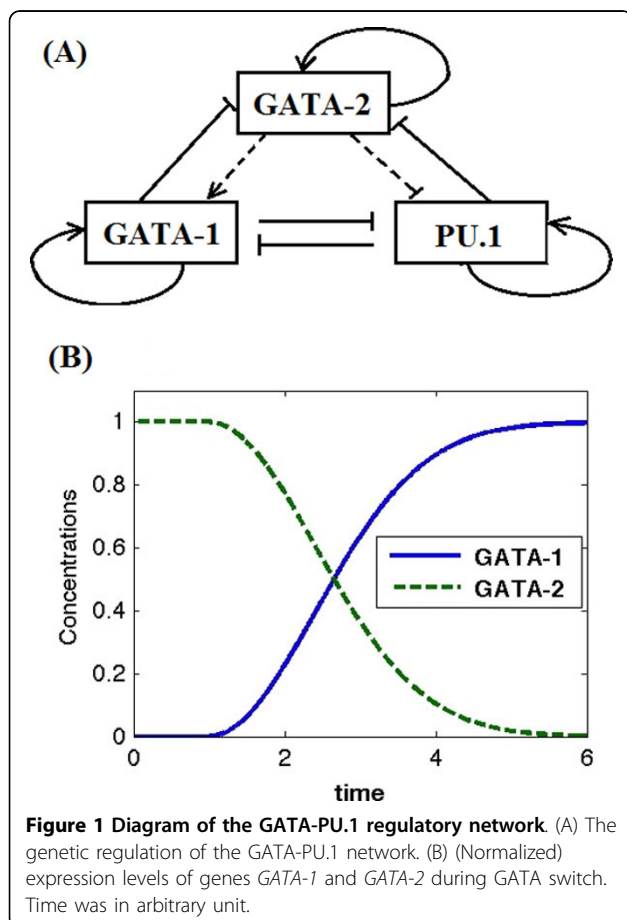
Based on the GATA switch model (Figure 1B), GATA-2 localizes at the chromatin sites in early stage erythroblasts. When the expression levels of GATA-1 increase as erythropoiesis progresses, GATA-1 displaces GATA-2 from chromatin sites and often (but not always) instigates a distant transcriptional output [21]. Although remaining in the cell, TF GATA-2 in fact is unable to access the chromatin sites. To model this spatial regulatory mechanism, it was assumed that GATA-2 proteins degrade during the process of GATA switch, which was realized by a large degradation rate constant k_2^* of GATA-2 in Eq. (1). Simultaneously, the concentration of GATA-1 increases by using an additional synthesis rate of GATA-1, which is proportional to the removal of GATA-2.

Experiments also showed that, when GATA-1 was not present, the shRNA knockdown of GATA-2 increases the expression of gene *PU.1* and reprograms the cells to become macrophages [24]. To realize genetic switching, the mechanism of the GATA switch and that of PU.1 knockdown were included in a single framework. Thus the knockdown of GATA-2 was realized by a large degradation rate constant k_2^* of GATA-2 during a particular time period. In addition, a very small synthesis rate of GATA-1 during that time period means the absence of GATA-1 protein in the DNA promoter region. Furthermore, for simplicity, the basal expression rates of these three genes are assumed to be zero. We use the Shea-Ackers formalism, which is a widely used thermodynamic approach, to represent the gene expression based on the structure of transcription machinery [38]. All these assumptions led to the following model to realize the genetic switching of the GATA-PU.1 regulatory network, given by

$$\begin{aligned} \frac{dx}{dt} &= \frac{a_1x + a_2y}{a_3 + a_4x + a_5y + a_6z + a_7xz} - k_1x + \mu k_2^*y \\ \frac{dy}{dt} &= \frac{b_1y}{b_2 + b_3x + b_4y + b_5z + b_6yz} - k_2y - k_2^*y \\ \frac{dz}{dt} &= \frac{c_1z}{c_2 + c_3x + c_4y + c_5z + c_6xz + c_7yz} - k_3z \end{aligned} \quad (1)$$

where x , y and z are the concentrations of TFs GATA-1, GATA-2 and PU.1, respectively, a_1 , b_1 and c_1 represent the expression rates of genes *GATA-1*, *GATA-2* and *PU.1* auto-regulated by itself, respectively, a_2 is the expression rate of gene *GATA-1* regulated by TF GATA-2, k_1 , k_2 and k_3 are the degradation rates of TFs GATA-1, GATA-2 and PU.1, respectively, k_2^* is the degradation rate constant to represent the displacement of GATA-2 proteins,

$$k_2^* = \begin{cases} k_{20} & t \in [t_1, t_2] \\ 0 & \text{else} \end{cases} \quad (2)$$



during the time period $[t_1, t_2]$, and μ is a parameter to adjust the availability of GATA-1 proteins in the chromatin sites. The GATA switch model is realized by using $k_2^* > 0, \mu > 0$ with a moderate value of parameter μ (e.g. $\mu = 1$). In contrast, the knockdown of GATA-2 is realized by using $k_2^* > 0, \mu = 0$. Another realization of the GATA-2 knockdown is to use rate constants $k_2^* = 0, b_1 = 0$ for the expression of gene *GATA-2*, namely to block the synthesis process of GATA-2 but maintain the degradation process of GATA-2 unchanged. Numerical results did not show any substantial difference between the computer simulations of these two types of realization (results not shown). Finally we note that the proposed model (Eq. 1) includes a recently designed model for the GATA-1-PU.1 module [18] as a special case if gene *GATA-2* is removed from the system.

Model parameters estimated from experimental data

There are 23 rate constants in the proposed mathematical model (Eq. 1). Part of the parameters was determined by the following experimental data.

- (1) The half-life of GATA-1 protein is one hour [39]. In addition, the half-life of GATA-2 is 30 min [40] that was confirmed by another observation in [33] therein. The half-life of PU.1 in Mel cell is ~ 2.4 h [41]. Thus the protein degradation rates constants are $\lambda_1 = 0.6931/h, \lambda_2 = 1.3863/h, \lambda_3 = 0.2888/h$.
- (2) The disassociate rate of GATA-1 binding to its DNA promoter is $K_d = 2.8$ nM, which is more stable than the binding of GATA-2 to its promoter site $K_d = 4.4$ nM [42]. In addition, the disassociate rate of PU.1 binding to its DNA promoter is $K_d = 170$ nM [43]. These rates were used to determine coefficients a_4, b_4 , and c_5 .
- (3) In addition, the heterodimer GATA-1-PU.1 has a 3-fold increase of the binding rate constant over GATA-1 to DNA [7]. It was assumed that $a_7 = 3a_4$.
- (4) The disassociation rates of GATA-1 and GATA-2 binding to the DNA promoter sites are very close to each other [44]. It is assumed that the binding rates of GATA-1 and GATA-2 to the same DNA binding sites are the same, namely $b_6 = a_7$, and $c_7 = c_6$.

Multiple-objective optimization approach

To infer the unknown parameters in model (Eq. 1), we designed a novel multiple-objective optimization approach to estimate unknown model parameters (Figure 2). The criteria in this approach include the conditions of tristability, genetic switching and robustness property of the model. This approach includes the following major steps:

Step 1. Generate a set of model parameters in the GA. We first used the uniform distributed random

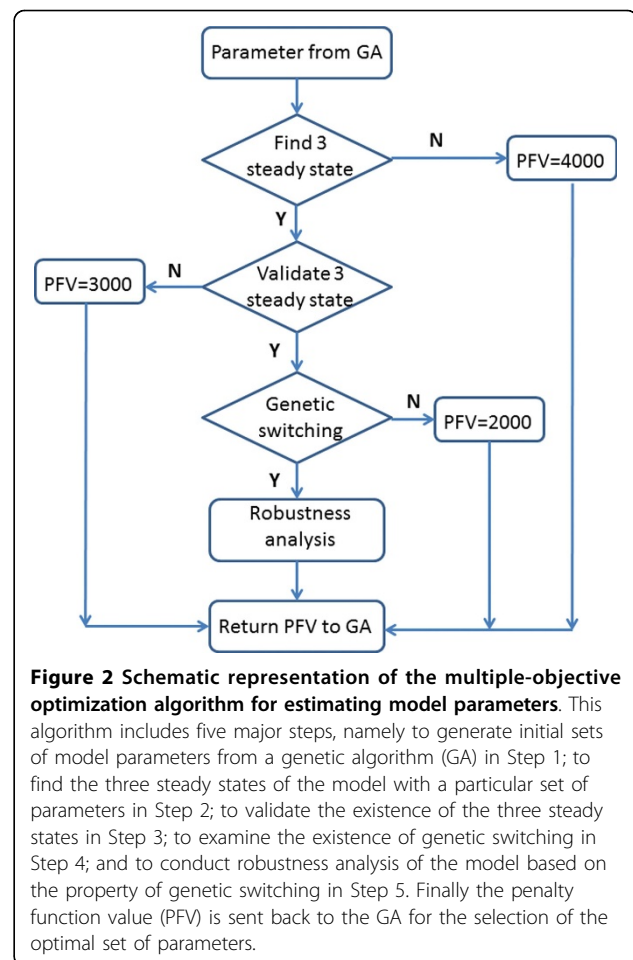


Figure 2 Schematic representation of the multiple-objective optimization algorithm for estimating model parameters. This algorithm includes five major steps, namely to generate initial sets of model parameters from a genetic algorithm (GA) in Step 1; to find the three steady states of the model with a particular set of parameters in Step 2; to validate the existence of the three steady states in Step 3; to examine the existence of genetic switching in Step 4; and to conduct robustness analysis of the model based on the property of genetic switching in Step 5. Finally the penalty function value (PFV) is sent back to the GA for the selection of the optimal set of parameters.

variable $U(0,1)$ generated in the genetic algorithm (GA) to determine the initial model rate constants. Note that the mutation manipulation in the GA was carried out on the random samples rather than model parameters. Since there are 7 restricted conditions for the existence of stable steady states (Eq. 6~8), other parameters were directly determined by the random samples $a_i = r_i$, where $r_i \sim U(0,1)$. Seven parameters were determined by the random samples together with the restricted conditions. For example, according to condition (Eq. 7), parameter b_3 is determined by

$$b_3 = \frac{1}{r_{10}x_1^*} \left[\frac{b_1}{k_2} - b_2 \right]$$

where $r_{10} \sim U(0,1)$ and x_1^* is the steady state (Eq. 4). In addition, based on condition (Eq. 6), the four synthesis rates were also determined by a factor k ,

$$a_1 = \frac{k_1 a_3}{r_1} \times k, \quad a_2 = r_2 k, \quad b_1 = \frac{k_2 b_2}{r_9} \times k, \quad c_1 = \frac{k_3 c_2}{r_{16}} \times k,$$

where $r_i \sim U(0,1)$. We tested different values of k in order to realize genetic switching. When k is not large

($1 < k < 100$), we failed to find a set of parameters that could realize genetic switching. Thus in this work we used $k = 1000$ to search the unknown synthesis rate.

Step 2. Find the third steady state. The generated parameter set in Step 1 ensures the existence of the two stable steady states in which either GATA-1 or PU.1 has high expression levels. To find the third stable steady state representing the primed progenitor state, we used MATLAB function *fsolve.m* to solving the nonlinear system for the steady state of the model (Eq. 1). If we can find the third stable steady state, we go to Step 3. Otherwise, we set the penalty function value to 4 and go to Step 6.

Step 3. Validate the existence of three steady states. Since the third steady state found in Step 2 was obtained by a numerical method, it may not exist due to the computational error. To examine the existence of the three steady states, we perturbed each steady state using samples of the uniformly distributed random variable and used the perturbed steady state as the initial condition to simulate system (Eq. 1). If the simulation converges to the steady state, it means the steady state exists, and then we go to Step 4. Otherwise, we set the penalty function value to 3 and go to Step 6.

Step 4. Validate the existence of genetic switching. We next examined whether the system can realize genetic switching using the mechanisms of GATA switch and GATA-2 knockdown. The large degradation rate constant of GATA-2 in model (Eq. 1) for these two mechanisms is

$$k_2^* = \begin{cases} 20, & 500 \leq t \leq 1500 \\ 0, & \text{else.} \end{cases}$$

Simultaneously, the additional synthesis rate constant of GATA-1 in model (Eq. 1) is

$$\mu = \begin{cases} 1, & \text{GATA switch} \\ 0, & \text{GATA - 2 knockdown.} \end{cases}$$

If the system realizes genetic switching through these two mechanisms, go to Step 5. Otherwise, set the penalty function value to 2 and go to Step 6.

Step 5. Robustness analysis. The inferred parameter set now satisfies the requirement for realizing three stable steady states and genetic switching. Next we used robustness property of the system as an additional criterion to choose the optimal model parameters. We tested the robustness property of mathematical model using the perturbed model parameters [45,46]. Each parameter in the model was perturbed by a uniformly distributed random number

$$a_{jk} = a_j(1 + \sigma^*(U_{jk} - 0.5)), \quad j = 1, \dots, 23, k = 1, \dots, 1000,$$

where $U_{jk} \sim U(0,1)$ and $\sigma = 0.5$ is the perturbation strength. For each set of model parameters, we obtained 1000 sets of perturbed parameters and then examined whether the mathematical model with the perturbed model parameters still maintained the three steady states. The model with a particular set of model parameters is more stable if the model maintains the three steady states with more sets of perturbed model parameters. To make an unbiased comparison, we used the same random samples U_{jk} for different sets of parameters. The penalty function value at this step is the percentage of the parameter sets (from the 1000 sets of perturbed parameters) with which the model does not maintain the three steady states. Thus a smaller value of penalty function means better robustness property of the model.

Step 6. Return the value of penalty function to the genetic algorithm.

Note that there are four penalty functions in this proposed inference method, namely the existence of three steady states in Step 2 (denoted as event O_1), validation of these three steady states in Step 3 (denoted as O_2), validation of the existence of genetic switching in Step 4 (denoted as O_3), and robustness property of the model over 1000 perturbed sets of model parameters in Step 5 (denoted as O_4). Thus, if a numerical test in Steps 2~4 is not satisfied, the penalty score of that step should be larger than the maximal score of the next step. Since the maximal penalty score of Step 5 is set to unit one if none of the perturbed parameter set maintains genetic switching, the scores of the first three penalty functions are set to 4, 3 and 2, respectively. Using the notations of objective functions [47], the multi-objective function is represented by

$$\min \{4P(O_1) + 3P(O_2|P(O_1) = 0) + 2P(O_3|P(O_2) = 0) + P(O_4|P(O_3) = 0)\}.$$

Here the probability of objective O_1 is defined by, for example,

$$P(O_1) = \begin{cases} 0 & \text{existing three steady states} \\ 1 & \text{otherwise} \end{cases}.$$

Similar definitions are applied to the probabilities of objectives O_2 and O_3 .

Results

Steady states of the mathematical model

When $k_2^* = 0$, the proposed mathematical model (Eq. 1) has up to four steady states. When $k_1a_4 \neq 0$ and $k_3c_5 \neq 0$, three steady states can be obtained analytically, given by

$$(x_0, \gamma_0, z_0) = (0, 0, 0) \tag{3}$$

$$(x_1, y_1, z_1) = \left(\frac{a_1 - k_1 a_3}{k_1 a_4}, 0, 0 \right) \quad (4)$$

$$(x_2, y_2, z_2) = \left(0, 0, \frac{c_1 - k_3 c_2}{k_3 c_5} \right). \quad (5)$$

The existence conditions of these three steady states are given in Theorem 1.

Theorem 1 (1) The trivial steady state (Eq. 3) is unstable if any one of the following conditions is satisfied

$$a_1 > a_3 k_1, \quad b_1 > b_2 k_2, \quad \text{and} \quad c_1 > c_2 k_3. \quad (6)$$

(2) The steady state with high expression levels of gene *GATA-1* (Eq. 4) is stable if the following conditions are satisfied

$$b_1 < k_2(b_2 + b_3 x_1), \quad \text{and} \quad c_1 < k_3(c_2 + c_3 x_1). \quad (7)$$

(3) The steady state with high expression levels of gene *PU.1* (Eq. 5) is stable if the following conditions are satisfied

$$a_1 < k_1(a_3 + a_6 z_2), \quad \text{and} \quad b_1 < k_2(b_2 + b_5 z_2). \quad (8)$$

The proof of this theorem is provided in the Additional file 1.

However, it is difficult to derive an analytical expression of the fourth potential steady state. For a given set of model parameters, we have to examine the existence of the fourth steady state numerically on a case-by-case basis. The existence conditions of the stable steady states will be used as criteria to search the unknown model parameters.

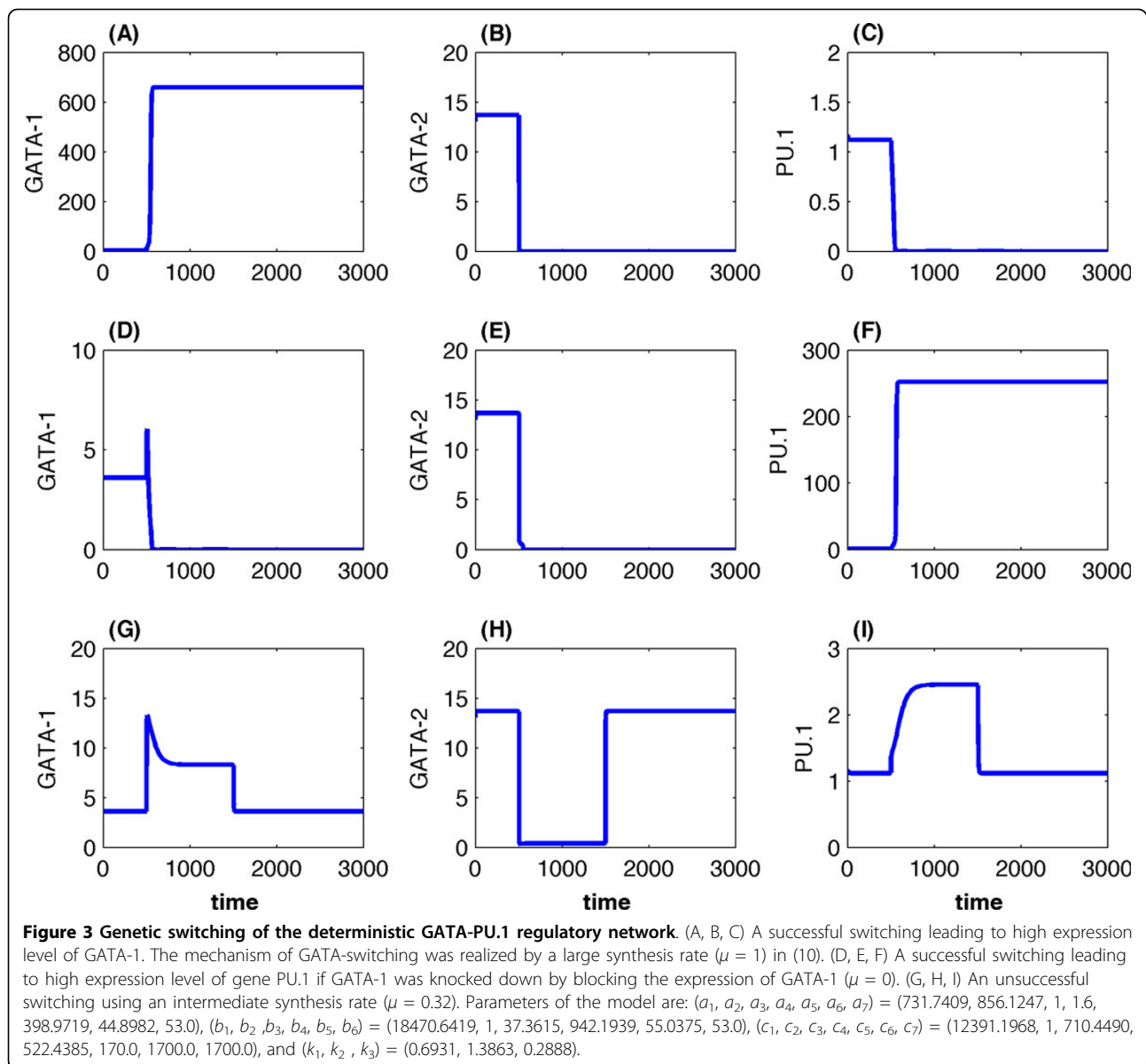
Inference of model parameters

The proposed model (Eq. 1) has 20 unknown parameters by setting $a_3 = b_2 = c_2 = 1$. We first estimated the values of 9 model parameters, namely $(a_4, a_7, b_4, b_6, c_5, k_1, k_2, k_3)$ and $(c_6 = c_7)$ from the published experimental data discussed in the model section. Since there is not any published data for the temporal dynamics of gene expression, we used the designed multiple-objective optimization approach to estimate the remaining 11 unknown parameters (Figure 2). We used the genetic algorithm as an effective tool to search the optimal model parameters to realize genetic switching. A MATLAB toolbox developed by Chipperfield et al. [48] was used to infer the unknown model parameters. This toolbox used MATLAB functions to build a set of versatile routines for implementing a wide range of genetic algorithms. The initial estimate of rate constants can be changed by using different random seeds in the MATLAB toolbox, leading to different final estimates of the rate constants [49]. The genetic algorithm was run over 100 generations for each estimate, and we

used a population of 1000 individuals in each generation. Our tests showed that a smaller population of individuals (i.e. 100 or 200) failed to produce an estimate of model parameters with which the model has tristability property. We implemented the genetic algorithm with different initial kinetic rates and obtained a number of estimated parameter sets. The parameter set having the best robustness property was selected as the final estimate.

Figure 3 gives simulations of the deterministic model (Eq. 1) using the degradation rate constant of GATA-2 (Eq. 2) and different values of the GATA-1 synthesis rate μ . The concentrations of GATA-2 decreased substantially during $t \in [500, 1500]$ when a large degradation rate $k_2^* = 6$ was applied during that time period. Figures 3A, 3B and 3C give a simulation using the GATA switch module that was realized by $\mu = 1$. When the concentration of GATA-2 decreases during $t \in [500, 1500]$ in Figure 3B, GATA-1 reaches the high expression level quickly in Figure 3A. However, when GATA-1 is absent ($\mu = 0$) and GATA-2 is knockdown, Figures 3D, 3E and 3F show a simulation leading to the high expression level of PU.1. The third scenario is that the expression of GATA-1 maintains at a low level during the period of GATA-2 decreases when $\mu = 0.32$. In this case the expression levels of these three genes in Figures 3G, 3H and 3I maintain at an intermediate state that is dependent on the value of μ . Neither GATA-1 nor PU.2 reached the steady state of high expression levels. When the expression level of GATA-2 returned to the normal state at $t = 1500$, the system also returned to the primed steady state, which is an unsuccessful genetic switching.

When inferring the model parameters, we only tested the robustness property of the model with perturbation strength $\sigma = 0.5$ (see Step 5 in the Optimization Approach section). The next question is whether different values of the perturbation strength may lead to different selections of the estimated model parameters. To answer this question, we selected 10 sets of the estimated model parameters that have better robustness properties than the others, and examined the robustness property of the model using different perturbation strengths ranging from 0.1 to 1. Figure 4 shows that, the robustness property of the model with a smaller strength ($\sigma = 0.3$) is slightly different from that using the perturbation strength ($\sigma = 0.5$) because only a small numbers of perturbed parameters sets did not maintain the tristability property. However, when $\sigma \geq 0.4$, the robustness property based on different values of perturbation strength σ is consistent with that using ($\sigma = 0.5$). Figure 4 also presents the averaged robustness properties based on 10 values of the perturbation strength ($\sigma = 0.1, 0.2, \dots, 1$), which is well consistent with that using $\sigma = 0.5$. Thus it is suggested that the inferred model parameter is independent of the perturbation strength.



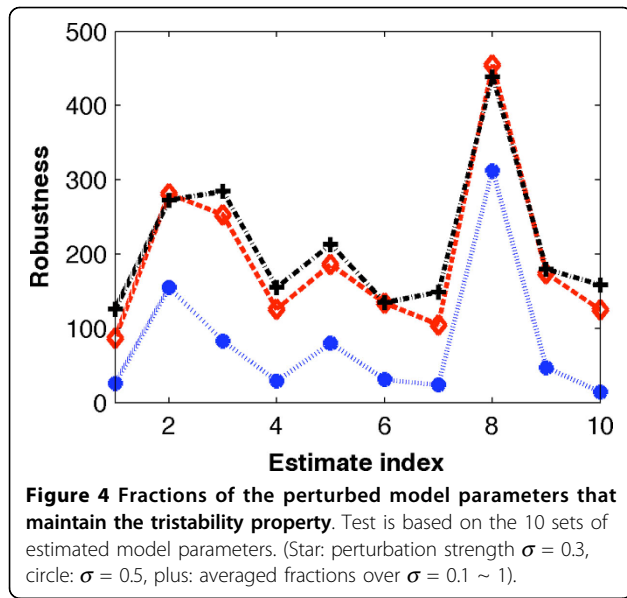
Bifurcation analysis

An important question in the estimation of model parameters is to what extent the model with the varied parameter still maintains the three stable steady states. To answer this question, we examined the tristability of the proposed model (Eq. 23) under the variations of one model parameter. For each rate constant, the perturbed values range from 0 to the 2-fold of the estimated value. We first tested the influence of the four synthesis rates a_1 , a_2 , b_1 and c_1 since these parameters are important in regulating the expression of these three genes. Figure 5 shows that the tristability property is not sensitive to the change of synthesis rate a_2 , which suggested that the expression of GATA-1 is not sensitive to the regulation of GATA-2. The result is consistent with the experimental observation

showing that this regulation is relatively weaker than other genetic regulations in the system. Interestingly, the system maintains tristability if the synthesis rate of GATA-1 (a_1) or PU.1 (c_1) increased, or the synthesis rate of GATA-2 (b_1) decreased. However, the tristability property of the system disappears when these three rate constants vary along the opposite direction, in particular for the synthesis rates of GATA-2 and PU.1. The discontinuity of the curves in Figure 5C suggested that the tristability property of the system disappears when the value of b_1 is below a threshold value.

Stochastic dynamics

We have successfully realized the three steady states of the HSCs and genetic switching using the genetic



regulations between GATA genes and gene *PU.1*. However, the proposed deterministic model failed to describe the heterogeneity in the differentiation decision of the HSCs. To tackle the challenge, we designed a stochastic model to realize the function of intrinsic noise arising from key species with small copy numbers, such as DNA and mRNA. Using our proposed stochastic modeling approach [50], we designed the following stochastic model based on the developed deterministic model (Eq. 1), given by

$$\begin{aligned} X(t + \tau) &= X(t) + P \left[\frac{a_1 X + a_2 Y}{a_3 + a_4 X + a_5 Y + a_6 Z + a_7 XZ} \tau \right] - P[k_1 X \tau] + P[\mu k_2^* Y \tau] \\ Y(t + \tau) &= Y(t) + P \left[\frac{b_1 Y}{b_2 + b_3 X + b_4 Y + b_5 Z + b_6 YZ} \tau \right] - P[k_2 Y \tau] - P[k_2^* Y \tau] \\ Z(t + \tau) &= Z(t) + P \left[\frac{c_1 Z}{c_2 + c_3 X + c_4 Y + c_5 Z + c_6 XZ + c_7 YZ} \tau \right] - P[k_3 Z \tau] \end{aligned} \quad (9)$$

where X , Y , and Z are the copy numbers of protein GATA-1, GATA-2 and PU.1 respectively, τ is the step-size, $P(\lambda)$ is a Poisson random variable with mean λ . Protein concentrations in model Eq. (1) were transferred into the molecular copy numbers in this stochastic model. To reduce the computing time of stochastic simulations, the mechanisms of genetic switching were realized in the time period

$$k_2^* = \begin{cases} 25 & 50 \leq t \leq 200 \\ 0 & \text{else} \end{cases} \quad (10)$$

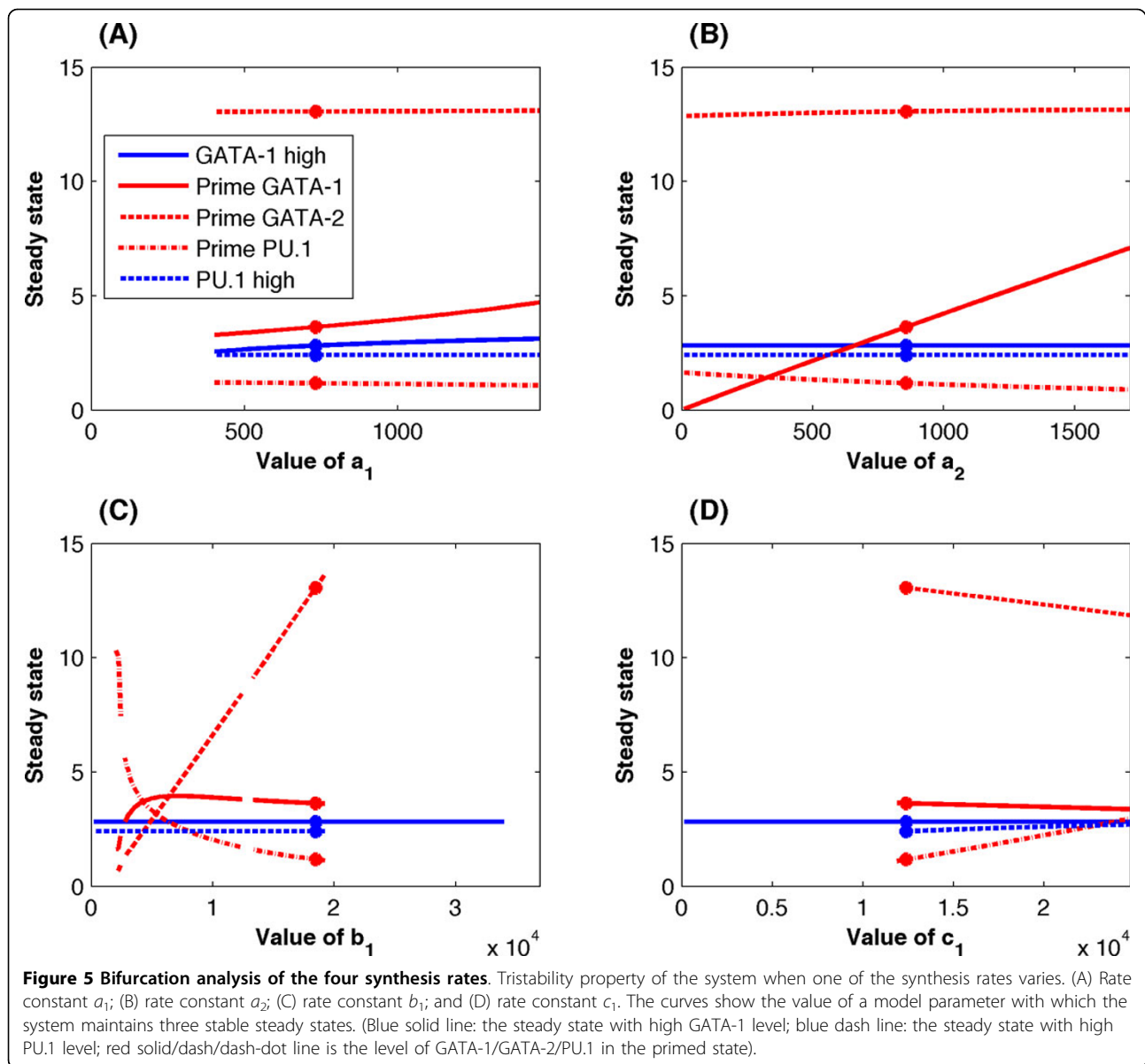
Figure 6 gives three stochastic simulations using the same rate constants in Figure 4 and $\mu = 0.28$. At time point $t = 50$, the expression level of GATA-2 decreased quickly to the basal level due to the large degradation rate k_2^* . Although the expression levels of genes *GATA-1* and *PU.1* increased quickly from time $t = 50$, the high

expression level of one gene inhibited the expression of the other gene. Figures 6A, 6B and 6C show that the high expression levels of *GATA-1* inhibited the expression of gene *PU.1* and *GATA-2*, leading the cell into the erythrocyte lineage; whereas the expression levels of *GATA-1* failed to increase further in Figures 6D and its expression was inhibited by the growing expression levels of *PU.1* in Figure 6F, leading the cell into the granulocyte lineage. In addition, it was possible that the expression levels of neither gene *GATA-1* nor *PU.1* were high enough to inhibit the expression of the other gene. Figures 6G, 6H and 6I gave a simulation of unsuccessful switching in which the expression levels of these three genes maintained at an intermediate state that was determined by the value of μ . When the degradation rate of GATA-2 returned to the normal value at $t = 300$, the expression levels of these three genes returned to the original priming state.

To obtain the statistical property of genetic switching, we generated 1000 stochastic simulations for each value of μ over the range of $[0, 1.5]$. When the value is small ($\mu < 0.11$), which mimicked the experimental condition of gene *GATA-2* knock-down, Figure 7 shows that all simulations have high expression levels of gene *PU.1*. On the contrary, when the value of μ is large ($\mu \geq 1.4$), all simulations were switched to the state with high expression levels of gene *GATA-1*, which represents the mechanisms of GATA switch that lead HSCs to the erythrocyte lineage. In addition, there is a range of values ($0.12 < \mu \leq 0.22$) with which the system failed to realize genetic switching and all stochastic simulations stay at the priming state. When the value of μ locates between these windows, stochastic simulations can realize two or three different steady states. For example, when the value of μ is $0.11 \leq \mu \leq 0.12$, the system state is in either the priming state or granulocyte lineage. An important prediction in Figure 7 is that, when the value of μ is ($0.22 < \mu \leq 0.37$), stochastic simulation of the system may lead to anyone of the three steady states, namely the erythrocyte, priming and granulocyte lineages, such as the case demonstrated in Figure 6.

Discussion

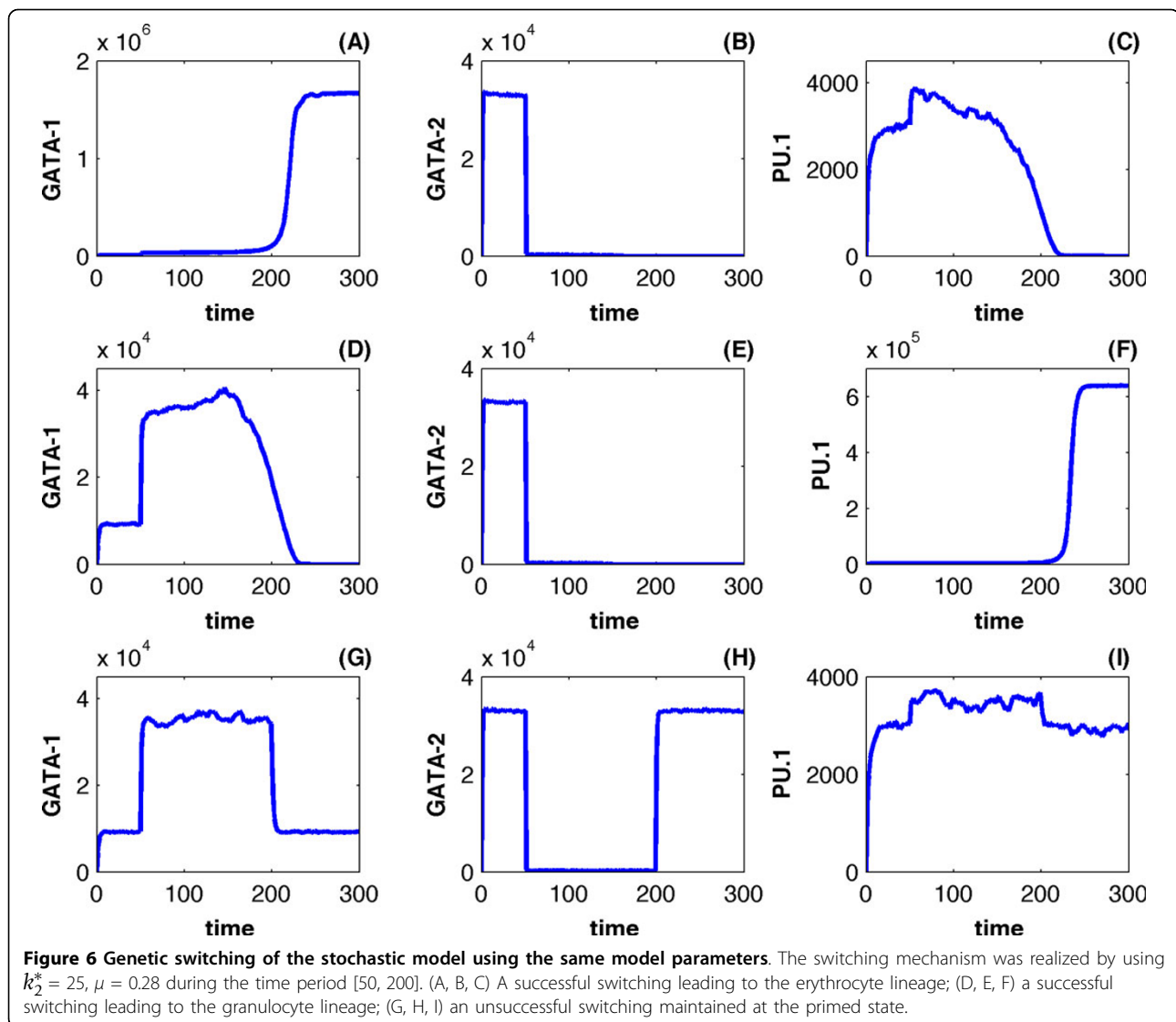
In this work we proposed a mathematical model to study the mechanisms of the GATA-PU.1 gene network in the determination of HSC differentiation pathways. The novelty of this model is the inclusion of gene *GATA-2* and the GATA switch model based on the experimentally determined regulatory mechanisms. Our simulation results suggested that, based on the experimental determined regulatory mechanisms, the addition of the third gene, namely gene *GATA-2*, is necessary and adequate to realize the three stable steady states of the HSCs. This result is consistent with the prediction that a connector



gene *X* is required to realize the primed state [15]. Although the third gene negatively regulates the expression of *PU.1* in these two models, the regulatory mechanisms of the third gene in these two models are not completely the same. For example, it was assumed that gene *GATA-1* positively regulate the expression of the unknown gene *X* [15], rather than the negative regulation of gene *GATA-2* by gene *GATA-1* in this work. Compared with the model in [15] in which additional and unknown external signals are required to maintain the tristability property, regulations between these three genes in our proposed model are adequate to maintain the tristability property of the system, which represents a successful approach in utilizing experimentally confirmed regulatory

mechanisms to realize tristability property of the HSCs in regulating the erythroid-myeloid lineage decision.

The proposed model in this work for a network of three genes is a general framework that includes a recently published model for a network of two genes as a special case [18]. Unlike the model of two genes, which realized tristability using the bifurcation of the system by increasing the ratio of two model parameters, our model maintains the tristability property over a wide range of model parameter values. A related interesting question is the minimal motif to realize the stability property of a regulatory network with different numbers of steady states. It has been demonstrated that the two-gene module with self-activation and mutual repression can realize



stability property with two steady states [50-53]. Although attempts have been made to realize tristability property of this two-gene module using various assumptions of regulatory mechanisms [13,14,17], our research demonstrated that a regulatory module with three genes could maintain stability property with three steady states without the assumption of the autoregulation via high order multimers. A similar result is that a three-component motif with four links can realize tristability [54]. However, the GATA-PU.1 module has six links together with more complex regulatory mechanisms. It would be interesting to analyze theoretically the stability property of a regulatory module with the maximal number of steady states under various regulatory mechanisms.

Stochastic simulations in this work predicted that the synthesis rate of GATA-1 during the decreasing process of GATA-2 determines the probability of erythroid-myeloid

lineage decision. Since this synthesis rate represents the availability of GATA-1 in the genomic regulatory regions during the GATA switch, there are two potential resources of noise to explain the variations of GATA-1 proteins in the genomic regions. First, recent experimental research investigating cellular processes at single cells has revealed convincing evidence showing large heterogeneity in protein abundance and dynamics among genetically identical cells [55]. The variations of protein concentration in different cells may lead to different rates for GATA-1 to enter the regulatory regions during the process of GATA switch. The heterogeneity in protein abundance may be one of the reasons to explain the differentiation preference of HSCs for the erythroid lineage or the myeloid lineage. In addition, the intrinsic noise due to the small copy numbers of molecular species in the GATA-PU.1 module may further contribute the heterogeneity to the HSC lineage

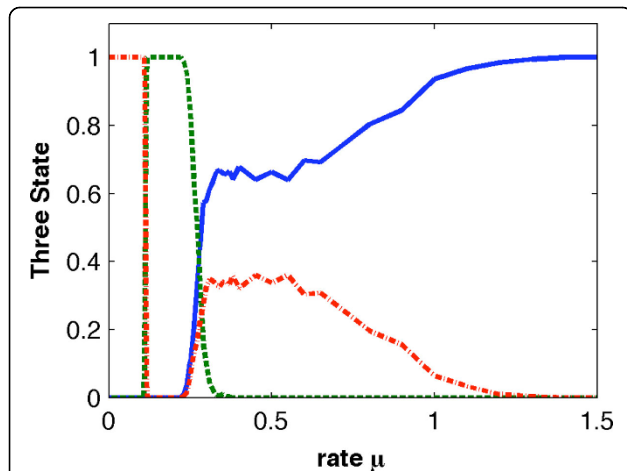


Figure 7 Fractions of cells showing different lineage states based on different values of the rate constant μ and 1000 simulations. (blue solid line: system showing the erythrocyte lineage; green dash line: system showing the priming state; red dash-dot line: system showing the granulocyte lineage).

differentiation decision even for cells having the same lineage decision preference. These various resources of noise in gene expression may be part of the transcriptome-wide noise proposed by Huang and colleagues [32]. Therefore more comprehensive stochastic models are needed to explain the functions of other types of noise in the decision of lineage selection.

Conclusions

In summary, we proposed a mathematical model to study the mechanisms of the GATA-PU.1 gene network in the determination of HSC differentiation pathways. In addition, a multiple-objective optimization approach was developed to infer model parameters in order to realize the three stable steady states representing the three different types of blood cells and genetic switching. A stochastic model was also designed to describe the function of noise in determining the differentiation pathways. Stochastic simulations successfully realized different proportions of cells leading to different developmental pathways under the same experimental conditions, and provided testable predictions regarding the conditions and mechanisms to realize different differentiation pathways. This work represents the first attempt at using a discrete stochastic model to realize the decision of HSC differentiation pathways with a multimodal distribution.

Additional material

Additional file 1: Supplementary Information. It provides the detailed assumptions of the mathematical model, the list of chemical reactions, and proof of Theorem 1 in the paper.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

T.T. conceived the research. T.T. and K.S.-M. conducted research, analyzed data and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work is supported by grants from the Australian Research Council (ARC) (DP1094181 and DP120104460). T.T. is also in receipt of an ARC Future Fellowship (FT100100748).

Declarations

The publication costs for this article were funded by the corresponding author.

This article has been published as part of *BMC Systems Biology* Volume 8 Supplement 1, 2014: Selected articles from the Twelfth Asia Pacific Bioinformatics Conference (APBC 2014): Systems Biology. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcsystbiol/supplements/8/S1>.

Published: 24 January 2014

References

- Wang LD, Wagers AJ: Dynamic niches in the origination and differentiation of haematopoietic stem cells. *Nat Rev Mol Cell Biol* 2011, **12**(10):643-655.
- Cedar H, Bergman Y: Epigenetics of haematopoietic cell development. *Nat Rev Immunol* 2011, **11**(7):478-488.
- Mercer EM, Lin YC, Murre C: Factors and networks that underpin early hematopoiesis. *Semin Immunol* 2011, **23**(5):317-325.
- Dore LC, Crispino JD: Transcription factor networks in erythroid cell and megakaryocyte development. *Blood* 2011, **118**(2):231-239.
- Friedman AD: Transcriptional control of granulocyte and monocyte development. *Oncogene* 2007, **26**(47):6816-6828.
- Goldfarb AN: Transcriptional control of megakaryocyte development. *Oncogene* 2007, **26**(47):6795-6802.
- Liew CW, Rand KD, Simpson RJ, Yung WW, Mansfield RE, Crossley M, Proetorius-Ibba M, Nerlov C, Poulsen FM, Mackay JP: Molecular analysis of the interaction between the hematopoietic master transcription factors GATA-1 and PU.1. *J Biol Chem* 2006, **281**(38):28296-28306.
- Nishimura S, Takahashi S, Kuroha T, Suwabe N, Nagasawa T, Trainor C, Yamamoto M: A GATA box in the GATA-1 gene hematopoietic enhancer is a critical element in the network of GATA factors and sites that regulate this gene. *Mol Cell Biol* 2000, **20**(2):713-723.
- Okuno Y, Huang G, Rosenbauer F, Evans EK, Radomska HS, Iwasaki H, Akashi K, Moreau-Gachelin F, Li Y, et al: Potential autoregulation of transcription factor PU.1 by an upstream regulatory element. *Mol Cell Biol* 2005, **25**(7):2832-2845.
- Akashi K, Traver D, Miyamoto T, Weissman IL: A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* 2000, **404**(6774):193-197.
- Gupta P, Gurudutta GU, Saluja D, Tripathi RP: PU.1 and partners: regulation of haematopoietic stem cell fate in normal and malignant haematopoiesis. *J Cell Mol Med* 2009, **13**(11-12):4349-4363.
- Burda P, Laslo P, Stopka T: The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia* 2010, **24**(7):1249-1257.
- Roeder I, Glauche I: 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* 2006, **241**(4):852-865.
- Huang S, Guo YP, May G, Enver T: Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev Biol* 2007, **305**(2):695-713.
- Chickarmane V, Enver T, Peterson C: Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. *PLoS Comput Biol* 2009, **5**(1):e1000268.
- Bokes P, King JR, Loose M: A bistable genetic switch which does not require high co-operativity at the promoter: a two-timescale model for the PU.1-GATA-1 interaction. *Math Med Biol* 2009, **26**(2):117-132.

17. Duff C, Smith-Miles K, Lopes L, Tian T: **Mathematical modelling of stem cell differentiation: the PU.1-GATA-1 interaction.** *J Math Biol* 2012, **64**(3):449-468.
18. Andrecut M, Halley JD, Winkler DA, Huang S: **A general model for binary cell fate decision gene circuits with degeneracy: indeterminacy and switch behavior in the absence of cooperativity.** *PLoS One* 2011, **6**(5):e19358.
19. Ling KW, Ottersbach K, van Hamburg JP, Oziemlak A, Tsai FY, Orkin SH, Ploemacher R, Hendriks RW, Dzierzak E: **GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells.** *J Exp Med* 2004, **200**(7):871-882.
20. Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH: **GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling.** *Proc Natl Acad Sci USA* 2003, **100**(15):8811-8816.
21. Bresnick EH, Lee HY, Fujiwara T, Johnson KD, Keles S: **GATA switches as developmental drivers.** *J Biol Chem* 2010, **285**(41):31087-31093.
22. Kaneko H, Shimizu R, Yamamoto M: **GATA factor switching during erythroid differentiation.** *Curr Opin Hematol* 2010, **17**(3):163-168.
23. Snow JW, Trowbridge JJ, Johnson KD, Fujiwara T, Emambokus NE, Grass JA, Orkin SH, Bresnick EH: **Context-dependent function of "GATA switch" sites in vivo.** *Blood* 2011, **117**(18):4769-4772.
24. Chou ST, Khandros E, Bailey LC, Nichols KE, Vakoc CR, Yao Y, Huang Z, Crispino JD, Hardison RC, Blobel GA, et al: **Graded repression of PU.1/Sfp1 gene transcription by GATA factors regulates hematopoietic cell fate.** *Blood* 2009, **114**(5):983-994.
25. Huang Z, Dore LC, Li Z, Orkin SH, Feng G, Lin S, Crispino JD: **GATA-2 reinforces megakaryocyte development in the absence of GATA-1.** *Mol Cell Biol* 2009, **29**(18):5168-5180.
26. Krumsiek J, Marr C, Schroeder T, Theis FJ: **Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network.** *PLoS One* 2011, **6**(8):e22649.
27. Raser JM, O'Shea EK: **Noise in gene expression: origins, consequences, and control.** *Science* 2005, **309**(5743):2010-2013.
28. Raj A, van Oudenaarden A: **Nature, nurture, or chance: stochastic gene expression and its consequences.** *Cell* 2008, **135**(2):216-226.
29. Balazsi G, van Oudenaarden A, Collins JJ: **Cellular decision making and biological noise: from microbes to mammals.** *Cell* 2011, **144**(6):910-925.
30. Hume DA: **Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression.** *Blood* 2000, **96**(7):2323-2328.
31. Miyamoto T, Iwasaki H, Reizis B, Ye M, Graf T, Weissman IL, Akashi K: **Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment.** *Dev Cell* 2002, **3**(1):137-147.
32. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S: **Transcriptome-wide noise controls lineage choice in mammalian progenitor cells.** *Nature* 2008, **453**(7194):544-547.
33. Graf T, Stadtfeld M: **Heterogeneity of embryonic and adult stem cells.** *Cell Stem Cell* 2008, **3**(5):480-483.
34. Palani S, Sarkar CA: **Integrating extrinsic and intrinsic cues into a minimal model of lineage commitment for hematopoietic progenitors.** *PLoS Comput Biol* 2009, **5**(9):e1000518.
35. Wilkinson DJ: **Stochastic modelling for quantitative description of heterogeneous biological systems.** *Nat Rev Genet* 2009, **10**(2):122-133.
36. Kaern M, Elston TC, Blake WJ, Collins JJ: **Stochasticity in gene expression: From theories to phenotypes.** *Nature Reviews Genetics* 2005, **6**(6):451-464.
37. Raj A, van Oudenaarden A: **Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences.** *Cell* 2008, **135**(2):216-226.
38. Shea MA, Ackers GK: **The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation.** *J Mol Biol* 1985, **181**(2):211-230.
39. Hernandez-Hernandez A, Ray P, Litos G, Ciro M, Ottolenghi S, Beug H, Boyes J: **Acetylation and MAPK phosphorylation cooperate to regulate the degradation of active GATA-1.** *Embo J* 2006, **25**(14):3264-3274.
40. Ferreira R, Wai A, Shimizu R, Gillemans N, Rottier R, von Lindern M, Ohneda K, Grosveld F, Yamamoto M, Philippsen S: **Dynamic regulation of Gata factor levels is more important than their identity.** *Blood* 2007, **109**(12):5481-5490.
41. Papetti M, Skoultchi AI: **Reprogramming leukemia cells to terminal differentiation and growth arrest by RNA interference of PU.1.** *Molecular cancer research: MCR* 2007, **5**(10):1053-1062.
42. Ghirlando R, Trainor CD: **Determinants of GATA-1 binding to DNA: the role of non-finger residues.** *J Biol Chem* 2003, **278**(46):45620-45628.
43. Pio F, Assa-Munt N, Yguerabide J, Maki RA: **Mutants of ETS domain PU.1 and GGAA/T recognition: free energies and kinetics.** *Protein science: a publication of the Protein Society* 1999, **8**(10):2098-2109.
44. Ko LJ, Engel JD: **DNA-binding specificities of the GATA transcription factor family.** *Mol Cell Biol* 1993, **13**(7):4011-4022.
45. Kitano H: **Biological robustness.** *Nat Rev Genet* 2004, **5**(11):826-837.
46. Kitano H: **Towards a theory of biological robustness.** *Mol Syst Biol* 2007, **3**:137.
47. Wang Y, Wu QF, Chen C, Wu LY, Yan XZ, Yu SG, Zhang XS, Liang FR: **Revealing metabolite biomarkers for acupuncture treatment by linear programming based feature selection.** *BMC Syst Biol* 2012, **6**(Suppl 1):S15.
48. Chipperfield AJ, Fleming PJ, Fonseca CM: **Genetic Algorithm Tools for Control Systems Engineering.** *Proceedings of Adaptive Computing in Engineering Design and Control* Plymouth Engineering Design Centre; 1994, 128-133.
49. Tian T, Xu S, Gao J, Burrage K: **Simulated maximum likelihood method for estimating kinetic rates in gene expression.** *Bioinformatics* 2007, **23**(1):84-91.
50. Tian T, Burrage K: **Stochastic models for regulatory networks of the genetic toggle switch.** *Proc Natl Acad Sci USA* 2006, **103**(22):8372-8377.
51. Tian T, Burrage K: **Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage lambda.** *J Theor Biol* 2004, **227**(2):229-237.
52. Gardner TS, Cantor CR, Collins JJ: **Construction of a genetic toggle switch in *Escherichia coli*.** *Nature* 2000, **403**(6767):339-342.
53. Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, Cantor CR, Collins JJ: **Programmable cells: interfacing natural and engineered gene networks.** *Proc Natl Acad Sci USA* 2004, **101**(22):8414-8419.
54. Tyson JJ, Novak B: **Functional motifs in biochemical reaction networks.** *Annu Rev Phys Chem* 2010, **61**:219-240.
55. Yuan TL, Wulf G, Burga L, Cantley LC: **Cell-to-cell variability in PI3K protein level regulates PI3K-AKT pathway activity in cell populations.** *Curr Biol* 2011, **21**(3):173-183.

doi:10.1186/1752-0509-8-S1-S8

Cite this article as: Tian and Smith-Miles: Mathematical modeling of GATA-switching for regulating the differentiation of hematopoietic stem cell. *BMC Systems Biology* 2014 **8**(Suppl 1):S8.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

