

Dynamic System Biology Final Report: The Genetic Landscape from the aspect of Toggle Switch Model Simulation and Reversal Gene Pairs in Expression Data Analysis

Ko, Kuei Yueh

July 2, 2015

1 Introduction

Since the first proposed by Waddington [Waddington, 1957], the concept of genetic landscape have long been fascinated by biologists, due to its simplicity in sketching the process of cell differentiation. In the concept of epigenetic landscape, the cell differentiation could be reduced by a series of multiple choices. The stem cells, like many marbles in the game of pinball, roll down and fall into each valley of the landscape. The valley represents the declining developmental potential of each cell population (fig. 1) [Hochedlinger and Plath, 2009].

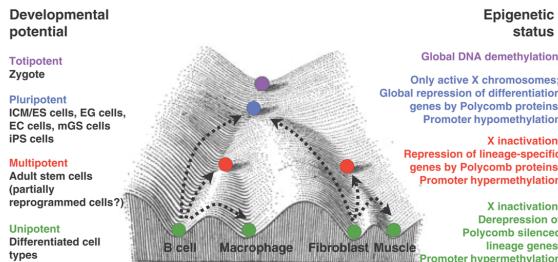


Figure 1: Illustrate the programming and reprogramming using the genetic landscape. The stem cell differentiation decrease its developmental potential, which is the number of cell types it could possibly becomes. The reprogramming of a cell could be classified as vertically (green ball to red ball or green ball to blue ball), increasing the developmental potential of the cells, and horizontally, converting to another cell types.(green ball to green ball)

The landscape is also used to explain the reprogramming of a cell. The cell reprogramming results in the conversion of a cell from one cell type to another, including vertical conversion on the landscape, such as the induced pluripotent stem cells, and horizontal transformation on the landscape, such as the process from fibroblast into a muscle cell [Davis et al., 1987]. However, the landscape is still quite an abstract concept, due to the complexity in a cell regulatory system, heterogeneity response to the same signal in a population of cells, and the noise of gene expression pattern throughout the population. It is hard to construct an actual landscape to depict the details behind the cell differentiation.

There are several points of view on the genetic landscape. The conventional molecular biology view emphasized on the epigenetic code such as histone modification and DNA methylation. It turns out that the acetylation and methylation on different histone residues produce different outcomes on the downstream gene expression. Likewise, the DNA methylation on gene body and promoter results in the different regulation on gene expression [Yang et al., 2014].

Another view of the genetic landscape is more systematic than the conventional view point. Since the popularization of high-throughput technologies, such as microarray and NGS, scientists are now enable to adopt a more systematic view on the cellular regulatory system. Imagine each cell type as an attractor during cell differentiation. The genetic landscape is actually the barriers and relative altitudes among each attractor.

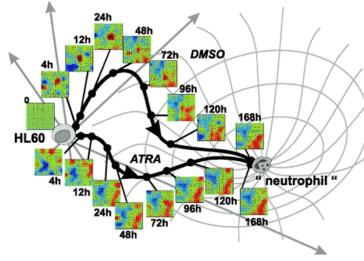


Figure 2: The experiments conducted by Sui Huang et al. The differentiation of HL60 was performed under two different cultural environment. As the time went by, the distance of gene expression patterns between the two populations increased at first and decreased in the end.

However, the question still remains how scientists may sketch the landscape and attractors using gene expression profile of each type of cells.

The relation between a cell type in cell differentiation and an attractor in dynamical system was first emphasized and discussed in detail by the theoretical biologist Stuart Kauffman [Kauffman, 2004]. In his discussion, theoretical models provided the tools to understand how thousands of genes orchestrate to produce hundreds of cell types. Each cell types are robust and differ more one another. Imaging an expression set of N genes as a sample point, or more abstractly, a state in an N -dimensional space, Kauffman proposed the hypothesis that cell types are in fact the attractors in the dynamic system of gene regulations. A year after the Kauffmans hypothesis, Sui Huang et al provided the experimental results that demonstrate the concept of attractors do exist when a cell type differentiates to another one [Huang et al., 2005]. The authors traced the differentiation of the human promyelocytic HL60 cells to neutrophils under two different conditions. It turned out that although the expression patterns of these two populations differed due to their cultural environment, the patterns were similar at the start and in the end of the process (fig. 2). These results and further publications [Chang et al., 2008] show that the state of cell types could be viewed as the stable steady states in the dynamic gene regulatory network.

In 2013, Sui Huang, together with Stuart Kauffman, proposed a method of finding the steady state from the gene expression of different cell types using the concept of "solving the inverse problem" [Kauffman, 2004]. Among the model for cell-lineage control, the toggle switch model and its derived versions have been proposed as the underlying mechanism during the binary decision in cell differentiation. According to the simple version of the model, two lineage-specifying transcription factors (TFs) repress each other and thus form a toggle switch. The toggle-switch circuit contributes two steady states of the system where either one of the TF dominates the other in expression level. Based on the model, a common progenitor falls into these two states. This model is extensively studied in the hematopoietic system. During hematopoietic differentiation, common myeloid precursor cell (CMP) differentiate into two lineages of cells, the erythroid/megakaryocyte lineage and myelomonocytic lineage (fig. 3a). The components of the switch consist of gene GATA1 and gene SPI1 (PU.1). The excess of GATA1 guides the CMP into the erythroid/megakaryocyte fate, while the increase of SPI1 promotes the CMP into myelomonocytic fate [Huang et al., 2007, Wang et al., 2010]. M Heinniemi et al compared the expression profile using the pattern produced in the toggle switch model and clustered the cell types to illustrate the the differentiation process 3b.

In this article, I would first introduce two toggle switch model and illustrate the pattern of the circuits. Then, I would re-run the code provided by M Heinniemi et al [Heinäniemi et al., 2013]. to demonstrate how they found the toggle switch patterns from the expression data. Most of the concepts were introduced in the publication of Sui Huang's Group [Huang et al., 2007, Wang et al., 2010, Heinäniemi et al., 2013].

2 Toggle Switch Model

The toggle switch model (1):

$$x_1 = b_1 \frac{k_{21}^n}{k_{21}^n + x_2^n} - k_1 x_1$$

$$x_2 = b_2 \frac{k_{12}^n}{k_{12}^n + x_1^n} - k_2 x_2$$

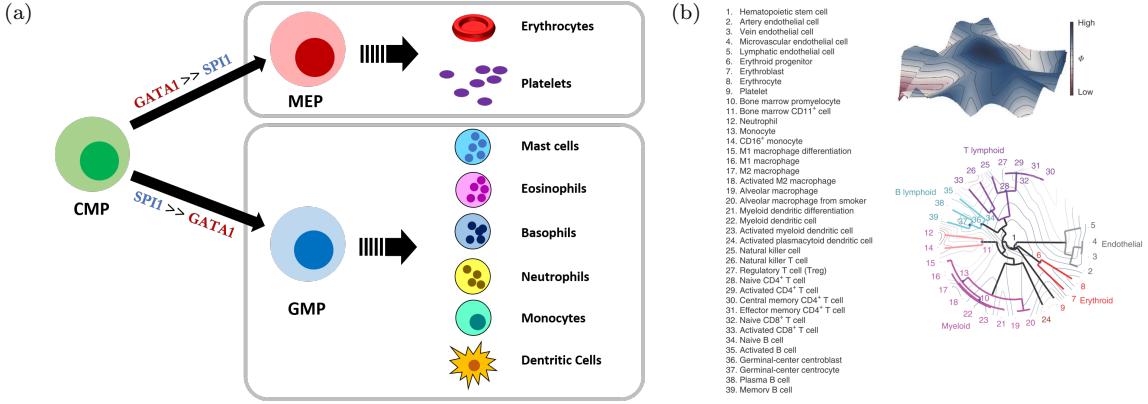


Figure 3: The Hematopoietic Differentiation (A) The binary decisions of common myeloid progenitor cell (CMP) and the cell types that are generated from these two different cell lineages. (B) The method proposed by M Heinniemi et al. The authors searched for the gene pairs that have reversal behaviors among the expression profiles of 166 cell types. The gene pairs were used to construct the large dendrogram to sketch the genetic landscape of hematopoietic differentiation.

The formulas represents a simple toggle switch models. Each of the formula consists of two parts: one is the Hill function that represents the inhibition of the transcription rate from the other gene, and another one is the first-order degradation with coefficients k_1 and k_2 . Figure 4a represents the vector fields under different Hill coefficient n . As n becomes larger than one, the bifurcation occurs and the steady state splits from one to three. The vector field shows that two of them are stable fixed points and the one in the middle is the unstable fixed point.

To analyze the relationship between the parameters and the steady state, the vector fields are drawn using different combinations of parameter values. Figure 5a shows how different k_1 and k_2 affect the location of fixed points. The parameters k_1 and k_2 represents the different degradation rate of x_1 and x_2 . It turns out that the stable steady state is closer to the gene with lower degradation coefficient. On the other hand, k_{12}^n and k_{21}^n represents the dissociation constants of the binding between TFs and promoters. Under the same concentration of TF, the higher the dissociation constant is, the lower proportion of TF-DNA binding complex exists. The results of different combinations of value k_{12} and k_{21} is plotted in figure 5b. The trajectories and nullcline of the system in figure 5b are shown in figure 5c.

To illustrate how the population of cells distribute on the landscape constructed using the expression of x_1 and x_2 , the numerical simulation of the switch model (1) was performed using the Gillespie algorithm (fig. 6a). The simulation was plotted on the 2-dimensional Cartesian coordinates, and each axis represents the expression level of each TF. Using the numerical simulation, we could model a cell population and trace the expression level of TFs through the iterations. In figure 6b, the color represents the iteration time, flowing from yellow to red. It turns out that all cells approach to the saddle node fixed point and then separate to the two stable fixed points where either one of the TF's expression level exceeds the other one.

Another version of toggle switch model is shown as follow. The toggle switch model (2) includes the self-stimulation of x_1 and x_2 :

$$x_1 = a_1 \frac{x_1^n}{k_{11}^n + x_1^n} + b_1 \frac{k_{21}^n}{k_{21}^n + x_2^n} - k_1 x_1$$

$$x_2 = a_2 \frac{x_2^n}{k_{22}^n + x_2^n} + b_2 \frac{k_{12}^n}{k_{12}^n + x_1^n} - k_2 x_2$$

Unlike the first toggle switch model, when the bifurcation occurs, this model has six fixed points, where three of them are stable (fig. 7a; the orange arrows) and the others are unstable (fig. 7a; the blue arrows). Based on the model, Sui Huang proposed how the change of genetic landscape guides the direction of the differentiation [Huang, 2009](fig. 7b).

To simulate how cell population on the vector fields, the results of Gillespie simulation of the second toggle switch model is shown in figure 8a and figure 8b. According to the figure 8b, as the time goes by, the cells (red dots) fluctuate around the three stable steady states, which is consistent to the model proposed in [Huang, 2009].

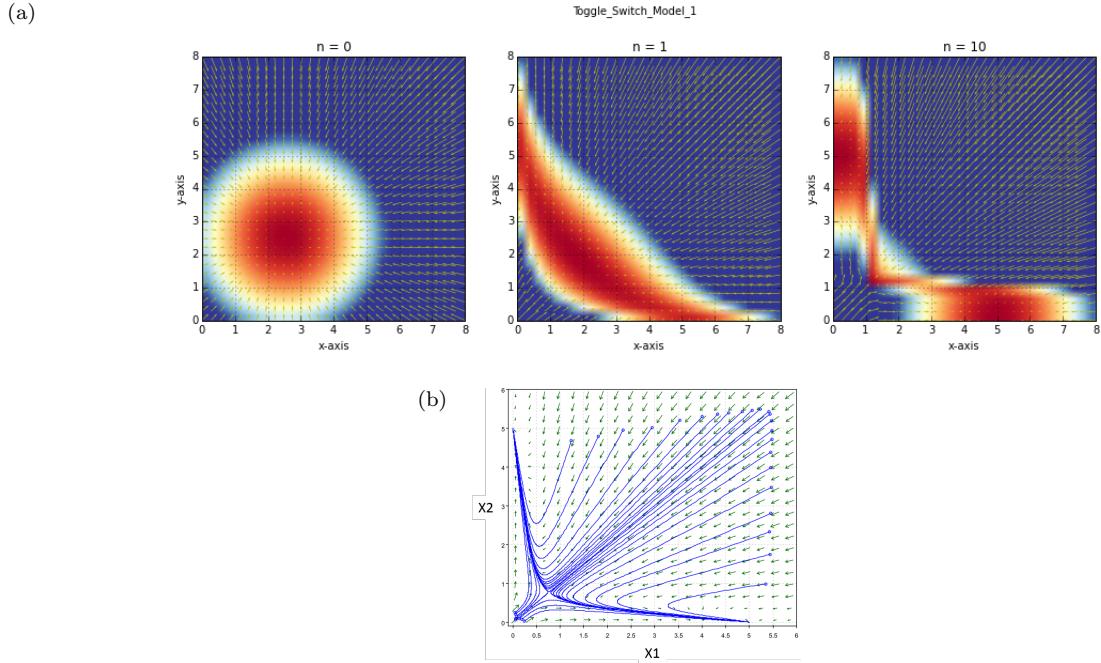


Figure 4: The toggle switch model (1) (A) As the Hill coefficient n is smaller than one, there is only one steady state, where the expression level of gene x_1 and gene x_2 are equal to each other. However, when n reaches one, the superior pitchfork bifurcation occurs and the steady state splits into three. (B) The trajectories with different initial state was plotted in the vector field of the system. There are three fixed points. Two are the stable fixed points where either one of the gene exceeds the other in expression level. The middle fixed point is a unstable saddle fixed point.

3 Gene-Pair Expression-Reversal Analysis

The above two toggle switch models illustrate the possible mechanism underlying the cell differentiation. The M Heinniemi et al advanced the idea of bifurcation and search for the toggle-switch patterns among the expression profile of 166 cell types (fig. 9b). The authors define the reversal score of each gene pair in each cell types (fig. 9a). The reversal score represents how the two genes are ranked differently between cell types. They then found the top reversal pairs for specific lineage split. In the end of their work, the authors define a similarity index that measures the similarity of each pair of cell types. The similarity index is calculated based on the number of reversal pairs identified using reversal score. The results are graphed in figure 10.

4 Conclusion

This article was aimed on reviewing the ideas of constructing the genetic landscape based on gene dynamic analysis and expression profiles. First, I introduced some aspects on how researchers tackle the problems of cell differentiation. Then, from the view of gene dynamic regulatory system, I simulated the two versions of toggle switch models, which provide the foundation of constructing the genetic landscape. Finally, I reproduced the work by M Heinniemi et al and illustrated how the reversal gene patterns are searched throughout the expression profile. Despite the noise of data, the method is still able to sketch the outlines of the genetic landscape.

References

- [Chang et al., 2008] Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453(7194):544–547.
- [Davis et al., 1987] Davis, R. L., Weintraub, H., and Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, 51(6):987–1000.

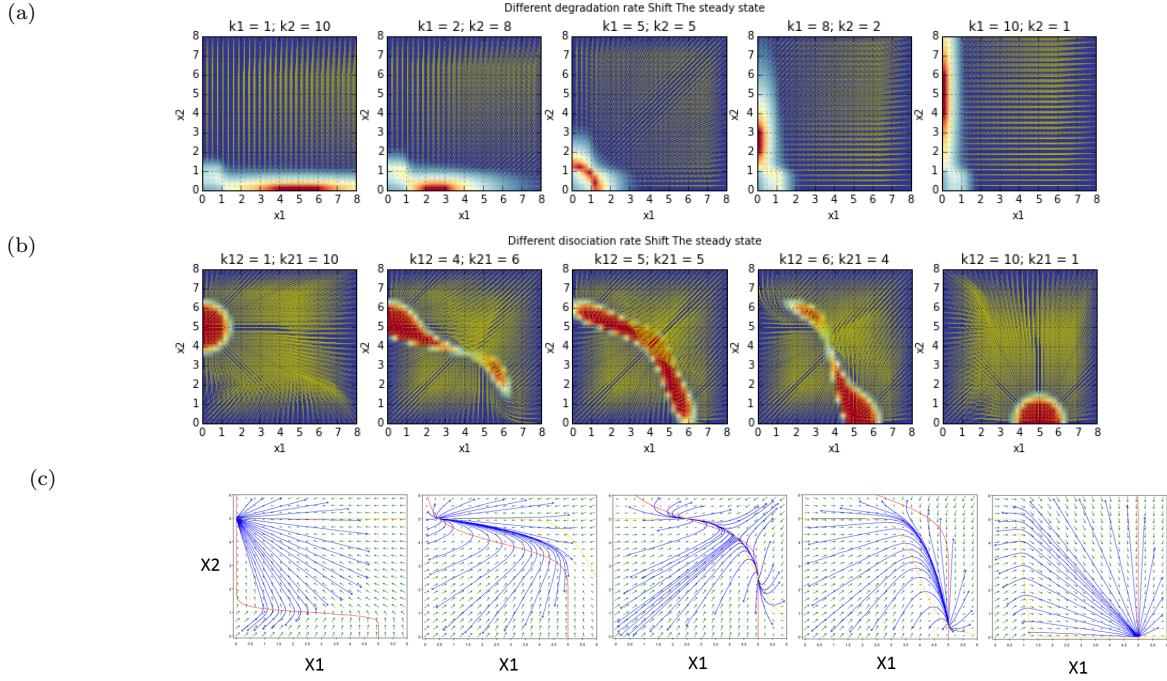


Figure 5: The toggle switch model (1) (A) The vector fields and the heatmap under different values of k_1 and k_2 . The stable fixed point is closer to the gene with less degradation coefficient. (B) The vector fields and the heatmap under different values of k_{21} and k_{12} , resulting in the dissociation rate between the TF and promoter. The stronger the binding of TF and promoter, the more efficiency of the inhibition. Therefore, if the x_1 binds to the promoter of x_2 with higher dissociation rate than the x_1 binds to the promoter of x_2 , the stable fixed point would tend to have higher expression level of x_2 . (C) The vector fields and trajectories of (B). The red and yellow line shows the nullcline of the model. Note that the figure runs the model using the parameter values $b = b_1 = b_2 = 5$. The k_{12} and k_{21} were assigned to 1 and k_1 and k_2 were assigned to 1 if those values are unspecified

[Heinäniemi et al., 2013] Heinäniemi, M., Nykter, M., Kramer, R., Wienecke-Baldacchino, A., Sinkkonen, L., Zhou, J. X., Kreisberg, R., Kauffman, S. a., Huang, S., and Shmulevich, I. (2013). Gene-pair expression signatures reveal lineage control. *Nature methods*, 10(6):577–83.

[Hochedlinger and Plath, 2009] Hochedlinger, K. and Plath, K. (2009). Epigenetic reprogramming and induced pluripotency. *Development (Cambridge, England)*, 136(4):509–523.

[Huang, 2009] Huang, S. (2009). Reprogramming cell fates: Reconciling rarity with robustness. *BioEssays*, 31(5):546–560.

[Huang et al., 2005] Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D. E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical Review Letters*, 94(12):1–4.

[Huang et al., 2007] Huang, S., Guo, Y. P., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology*, 305(2):695–713.

[Kauffman, 2004] Kauffman, S. (2004). A proposal for using the ensemble approach to understand genetic regulatory networks. *Journal of Theoretical Biology*, 230(4 SPEC. ISS.):581–590.

[Waddington, 1957] Waddington, C. (1957). *The strategy of the genes; a discussion of some aspects of theoretical biology*. New York, Macmillan [c1957].

[Wang et al., 2010] Wang, J., Xu, L., Wang, E., and Huang, S. (2010). The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophysical Journal*, 99(1):29–39.

[Yang et al., 2014] Yang, X., Han, H., DeCarvalho, D., Lay, F., Jones, P., and Liang, G. (2014). Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. *Cancer Cell*, 26(4):577–90.

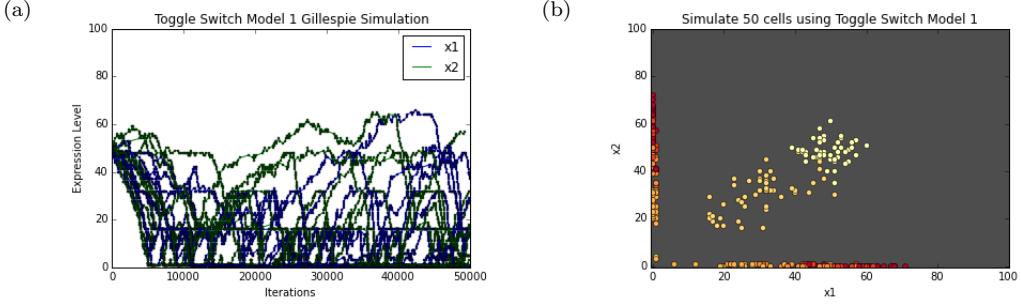


Figure 6: Gillespie simulation of toggle switch (1) (A) Perform ten rounds of simulation and the number of x_1 and x_2 is plotted during the simulation. (B) The simulation was drawn on the 2D plane where each axis represents number of x_1 and x_2 . The initial values were assigned based on the normal distribution with mean=50 and std=5. Each cell is plotted for every 5000 iterations and the color represents the flow of the time (from yellow to red). From the results, we could observe that in the end (the red dots), the red dots were around the two fixed points. Note that to increase the distance between the two stable steady states, parameter b_1 and b_2 is set to 100.

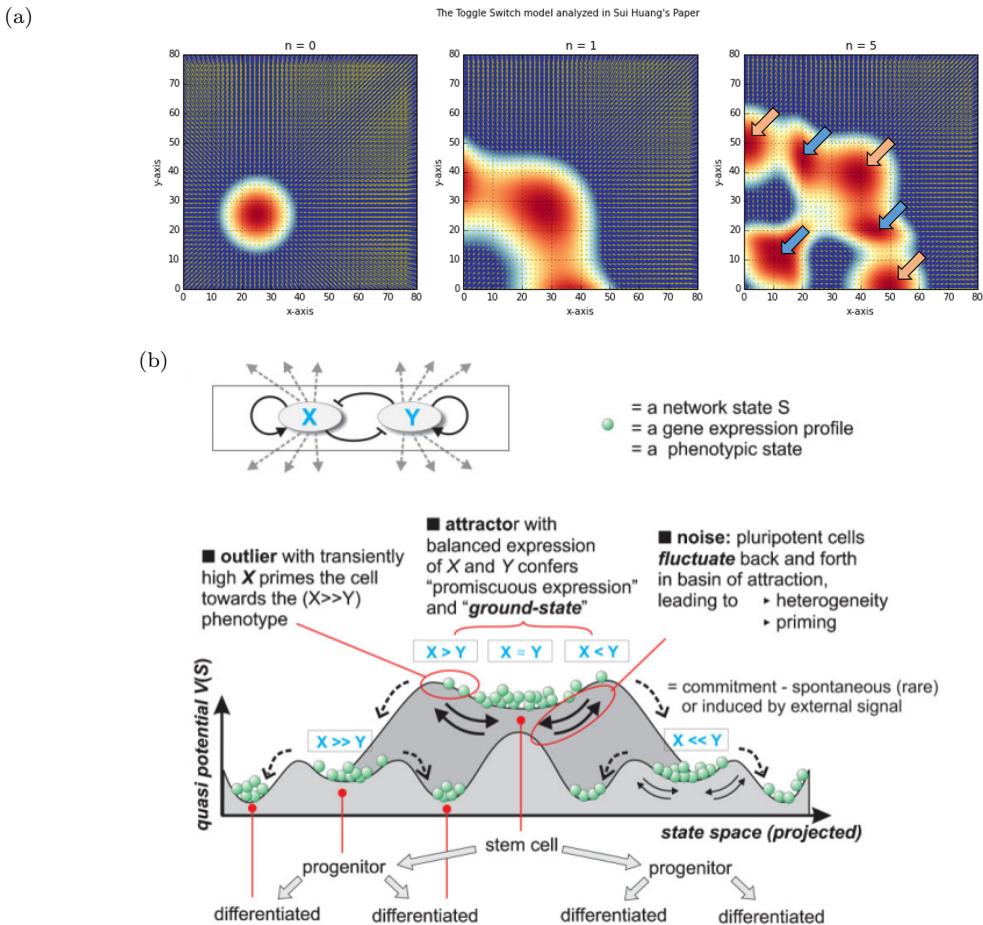


Figure 7: The toggle switch model (2) and the model introduced by [Sui Huang 2009 Review] (A) The vector field and steady state with different Hill coefficient n . (B) The model introduced by [sui Huang 2009 Review] is illustrated in the figure. In this model, the change of landscape creates three stable fixed point, and the each cell makes the binary decision due to noise. The process keeps going and more and more cell types form.

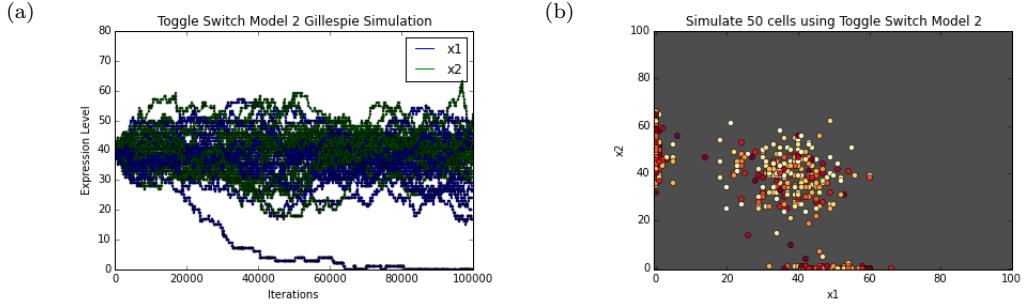


Figure 8: Gillespie simulation of the toggle switch model (2) (A) Perform ten rounds of simulation and the number of x_1 and x_2 is plotted during the simulation. (B) The initial values, similar to the simulating switch model (1), are independent random variables following normal distribution with mean=40 and std=10. The time flowed from yellow to red. At the end of simulation, the cells were near the three stable steady states. Note that the model is run with the parameter values listed as follow: $n = 4, a_1 = a_2 = 40, b_1 = b_2 = 10, k_{11} = k_{22} = k_{12} = k_{21} = 20, k_1 = k_2 = 1$

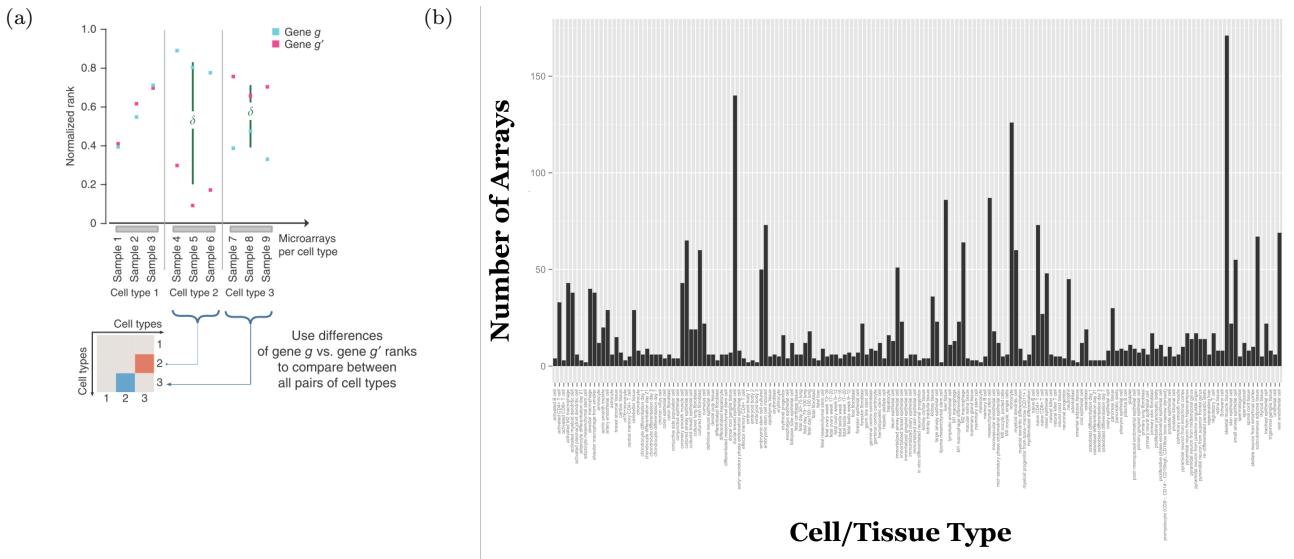


Figure 9: The concept of finding toggle pattern and the cell types used in M Heinniemi et al (A) The concept of searching the gene pairs with reversed pattern is shown in the figure. According to the figure, a gene pair that has reversed rank difference would be selected as reversal gene pair. (B) There are 166 cell types used in this publication, the number of array data for each cell types was counted and graph in the plot.

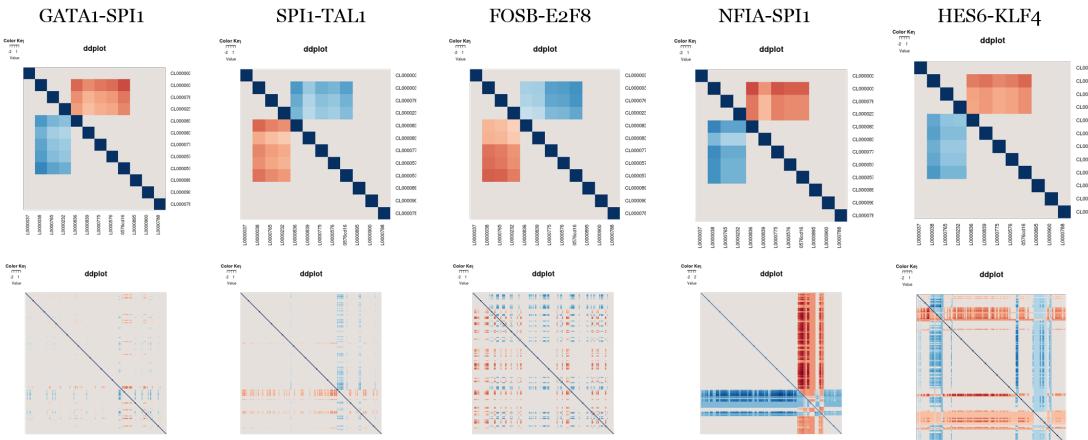


Figure 10: The pattern of reversal gene pairs are plotted using the cell types from the blood lineage (top row) and using all the cell types (bottom row)