Theory of genes network in reprogramming of iPS cells

Kazuhisa Nishi
School of Engineering,
University of Hyogo

1. Introduction

Reprogramming of somatic cells into iPS cells can be achieved by expression of several transcriptional factors (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Okita et al., 2007; Wernig et al., 2007; Maherali et al., 2007; Yu et al., 2007). There are some distinct properties in the reprogramming of iPS cells compared with nuclear transfer or cell fusion in somatic cells (Gurdon et al., 1958; Wilmut et al., 1997; Tada et al., 2001; Cowan et al., 2005). The first is that the specified transgenes (*Oct4, Sox2 etc.*) are needed in reprogramming, the second is that the endogenous loci of the transgenes should be expressed, and the third is that iPS cell generation requires several cell division. These facts imply that if the genes transcriptional network of somatic cells is activated by the specific factors, they could be finally reprogrammed by changing the total expressing pattern of genome. However, its mechanism still remains unsolved problem. In this paper I suggest a reprogramming mechanism of iPS cells, emphasizing how the dynamics of genes transcriptional network can be related with deterministic and stochastic processes.

2. Equations of gene network

The discovery of the iPS cells leads to the astonishing facts that tranceducing trigger genes into somatic cells can change the expressing patterns of genome drastically. In order to find the reprogramming mechanism of iPS cels, we will have to consider not only the genes network of the specified factors such as *Sox2*, *Oct3 etc* but also the total dynamics of genome with interacting between these pluripotency relevant factors. First let us consider basic equations used in the genes networks. Figure 1 shows the model of transcription factors network in iPS cells which is the positive feedback co-operated with *Oct4, Sox2*, and *Nanog*. Externally tranceduced *Oct4* and *Sox2* activate Oct4 and Sox2 enhancer, which expresses endogenous *Oct4, Sox2*, *Nanog* and so on. Further, they begin to activate Oct4-Sox2 enhancer, and a basic network of iPS cells is finally stabilized in the expression of the endogenous genes. Mathematical models of the gene-network have been almost applied for the system in which

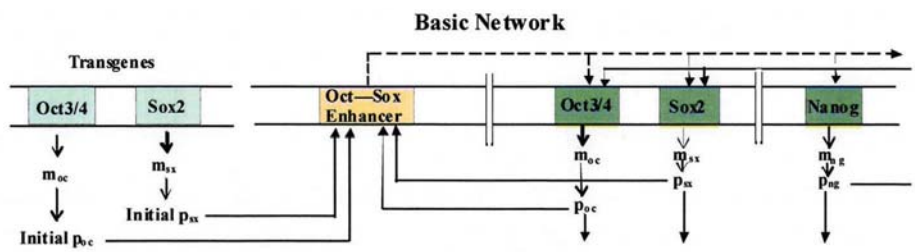

Fig. 1. Genes transcription network constructed with pluripotency related genes and proteins.

the genes are directly interacting with transcription factors (Elowitz and Leibler, 2000; Gadner et al., 2000). However, since in the eukaryote cells, genes are constituted with the chromatin structures, the control system of the gene transcription is not the same as the case of the prokaryote cells. Nevertheless, it is for simplicity assumed here that the main frame of the model equations adopted in the prokaryote cells can be also applied for the case of the eukaryote cells. To carry out this treatment efficiently, the system parameters or coefficients in equations will be re-defined so as to reflect the interaction between the transcriptional factors and the chromatin structures. The network starts with a mathematical model of transcriptional regulation, which is represented by Michaelis-Menten equation that describes the chemical reaction of the feedforward control constituted with *Oct4, Sox2* and *Nanog*. It is considered that Oct4-Sox2 enhancer is co-operatively activated by the product of the Oct4 and Sox2 expressing quantity.

Basic equations of the system are given by

$$\frac{dm_{oc(sx)}}{dt} = -m_{oc(sx)} + \frac{\alpha p_{oc} p_{sx}}{1 + p_{oc} p_{sx}}, \quad \frac{dp_{oc(sx)}}{dt} = -\beta(p_{oc(sx)} - m_{oc(sx))}).$$
$$\frac{dm_{ng}}{dt} = -m_{ng} + \frac{\alpha p_{ext}}{1 + p_{ext}}, \quad \frac{dp_{ng}}{dt} = -\beta(p_{ng} - m_{ng}). \tag{1}$$

where the variable p_{oc}, p_{sx}, p_{ng} are Nanog, Oct4, Sox2 protein number per cell in units of K_m (the number of proteins necessary to half-maximally activate a promoter), respectively, and m_{oc}, m_{sx}, m_{ng} are the corresponding mRNA number per cell rescaled by the average number of proteins produced per mRNA molecule. The coefficient $\alpha_1,\alpha_2,\alpha_3$ are maximum numbers of proteins per cell of Nanog, Oct4, Sox2, respectively. The coefficient β is the ratio of the protein decay rate to the mRNA decay rate. The action of the network depends on several factors, including the dependence of transcription rate on activator or repressor concentration, the translation rate, and the decay values of the protein and the messenger RNA. Figure 2 shows that Oct4 or Sox2 proteins number per cell in units of K_m (the number of proteins necessary to half-maximally activate a promoter) is represented as a function of time rescaled in units of the mRNA decay lifetime. Here, $\alpha_1 + \alpha_2$, α_3 are maximum numbers of proteins per cell of Oct4 or Sox2 and Nanog, respectively, and β is the ratio of the protein decay rate to the mRNA decay rate. The values of p_1, p_2, p_3 indicate the initial numbers of these proteins, respectively. It shows whether the expression of Oct4 or Sox2 can be stabilized in a cell or not is determined by the initial quantity of these proteins.

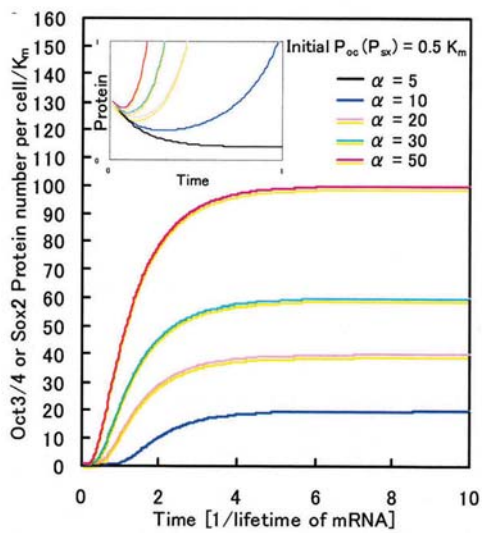

Fig. 2. Oct3/4 or Sox2 protein number.

Effective deterministic equations are defined as the formula of input and output constructed with only basic equations.

$$\frac{dm_{out}}{dt} = -m_{out} + \frac{\alpha p_{in}}{1 + p_{in}}, \quad \frac{dp_{out}}{dt} = -\beta (p_{out} - m_{out}). \tag{2}$$

Further, stochastic connection of deterministic equations are also defined as

3. Reprogramming process

Now let us consider the reprogramming process of somatic cells. In order to understand the reprogramming mechanism, the total dynamics of genome has to been revealed. First three groups of cells which are somatic, partially reprogrammed, and iPS cells will be considered and be factorized into G1~G5 groups, shown in the left side of Fig. 3. The right side of Fig. 3 shows the process model indicating how the expression patterns of genome can change from somatic state to reprogrammed state as the result of the interaction between reprogrammed related genes, master genes of G1~G5 groups and other genes as environment.

It is well known that requiring several cell divisions in reprogramming can be one reason why iPS cell generation is so slow and inefficient. These experimental facts lead us to the conjecture that the wild types of somatic cells can repair the epigenetic modifications after cell division, but the somewhat artificial cells modified by the basic network cannot sufficiently repair them. The epigenetic modifications of eukaryote cells can be also

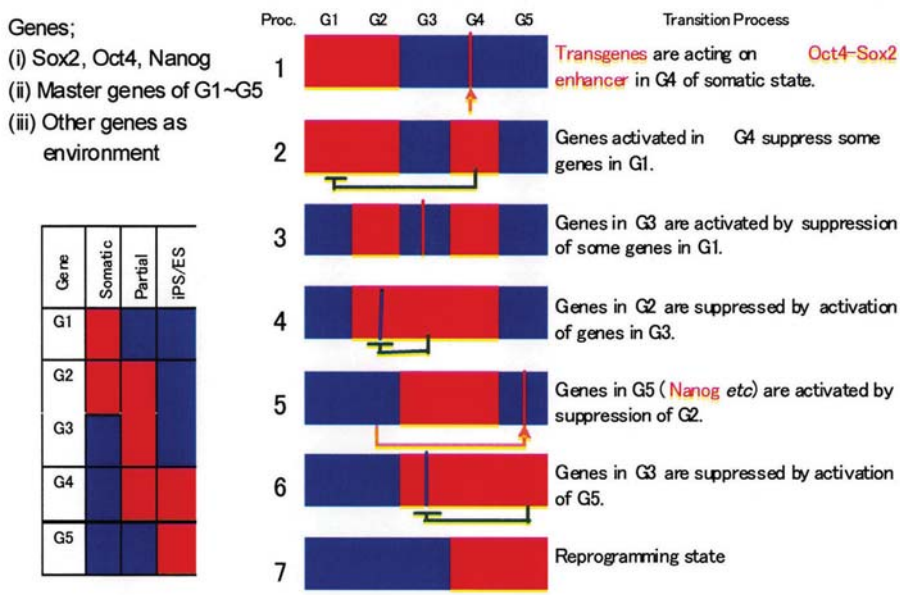

Fig. 3. left: Genes group G1~G5, right: Process model of reprogramming mechanism.

interpreted to be evolutionarily conserved, and this situation may not be applicable for the artificial cells such as iPS/ES. With these considerations on mind, several conditions will be considered in the reprogramming process.

There are epigenetic states such as DNA methylation and histon modificstion. Since the reprogramming transition from these epigenetic states has two properties of active and passive processes. Then, let us consider the two stochastic processes which are in the absence and presence of cell divisions. In the absence of cell divisions including both active and passive reprogramming processes, stochastic relation between deterministic genes network equations is given by

$$P = p_1 \times \cdots \times p_i \tag{3}$$

where i shows the number of effective deterministic equations. Next let us consider in the presence of cell divisions (passive reprogramming process). Given L the number of genes epigenetically modified in somatic cells. Here r is defined as the probability attaining the reprogrammed state after one cell division per one gene. The expectation value Q_1 of the genes in a single DNA, which are in the reprogrammed state after one cell division, on the average, is given by

$$Q_1 = \sum_{k=1}^{L} {}_L\mathrm{C}_k k r^k (1-r)^{L-k} \equiv Lr. \tag{4}$$

After one cell division, the number of cells which have the partial reprogrammed region is given by

$$N_1 = p_1 N. \tag{5}$$

where N is the number of initial somatic cells and p_1 is the probability defined by the relation (3). This means the existence of N_1 cells which have reprogrammed region Q_1. The value Q_n after n times cell divisions and the number N_n of partially reprogrammed cells are represented as

$$Q_n = r(L - Q_{n-1}) + Q_{n-1} \equiv L[1-(1-r)^n], \qquad N_n = p_1 \times \cdots p_i N. \tag{6}$$

The reprogrammed ratio in a cell is represented as a function of cell divisions and is shown in Fig. 4a.

$$R_n = \frac{\partial}{\partial n}\left(\frac{Q_n}{L}\right) = (1-r)^n \ln\frac{1}{1-r}. \tag{7}$$

The minimum cell divisions M for the complete iPS cells is obtained by the following relations

$$NPL(1-r)^{M-1} \geq 1, \quad NPL(1-r)^M < 1. \tag{8}$$

Then, M is given by (shown in Fig. 4b)

$$M = 1 - \frac{\log NPL}{\log(1-r)}. \tag{9}$$

The efficiency E_f producing iPS cells is also given by (shown in Fig. 4c)

$$E_f = PL(1-r)^{M-1}. \tag{10}$$

where $P = p_1 \times \cdots \times p_i$. The unknown parameters of P, L and r determining the reprogramming

processes can be calculated from equations (7), (9), (10). Choosing the values of R_1 = 0.08, M = 50 and E_f = 1% which is realistic in the experimental data, the values of parameters are evaluated as P = 0.001, L = 300 and r = 0.08. Let us interpret briefly some experimental facts related to the success of reprogramming and its producing efficiency. Why *Klf4* and *c-Myc* enhance the iPS producing can be interpreted by two possible mechanisms. One of these is to decrease reprogramming time by accelerating the cell division rate, and the other is to prevent the epigenetic modifications by the expressing factors of these genes. It is well known that the inhibition of p53/p2 pathway does not change necessary cell divisions to produce

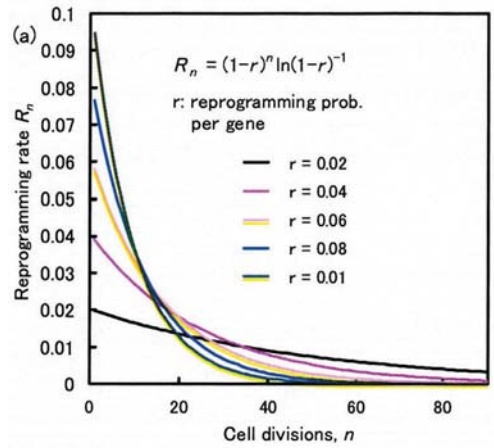

Fig. 4 (a) Reprogramming rate

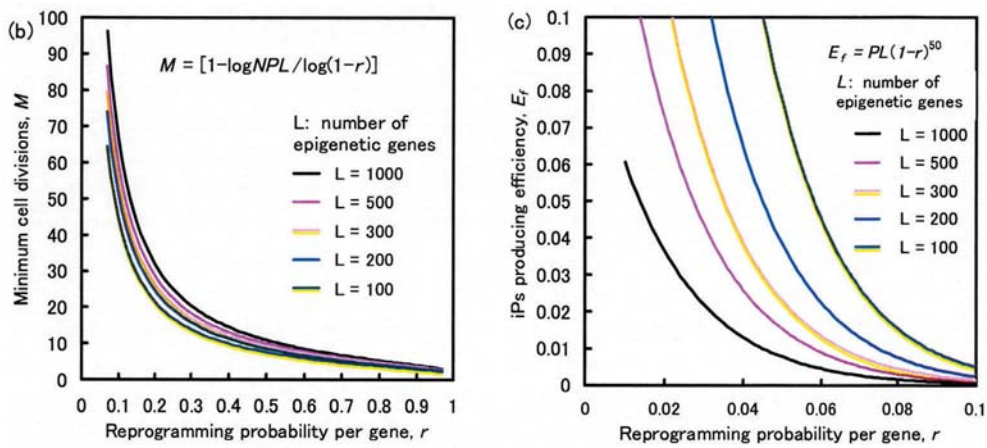

Fig. 4 (b) Minimum cell dvision for reprogramming; (c) iPS producing efficiency

iPS cells, but it decreases the necessary time to produce iPS cells (Hanna et al., 2009). This corresponds to the case that M is unchanged and τ is smaller in the model parameters. When Nanog, that is one of proteins expressing in the basic network, increases, it will increase the value of r because the repairing of epigenetic modification is prevented by operation of the basic network. This explains the experimental fact that Nanog overexpression decreases the necessary time to produce iPS cells. Only a small portion of the transduced cells become pluripotent and the others become almost partially reprogrammed state. This phenomenon may be corresponding to the case of $r << 1$ in the model. Also, the partially reprogrammed cells not only express the genes expressing in both somatic and iPS cells, but also express the genes non-expressing in these cells (Mikkelsen at al., 2006) In this model, the suppressed genes in somatic cells can activate by de-epigenetic process after cell divisions. Of these

genes, there exist the genes to be suppressive in iPS cells. However, there will be some cells which are not suppressed by basic network. This case is then considered to be the partially reprogrammed cells.

4. Conclusions

This model can compute the minimum cell divisions for attaining iPS cell and its generating efficiency, and can explain various experimental facts which have been by now reported in iPS cells. The reprogramming mechanism considered here assumes the existence of the undefined determinants or environmental conditions suppressing the repair enzyme for the epigenetic modifications. If suppressing determinants can be identified, it will make the efficiency of producing iPS cells more improved. More speculatively, since genes control system is not still closed, it will be modified by the interaction with the unknown environment or material which must be evolutionally determined. Thus, if the genes control system experiences the environment which can be never realized in wild types of cells, it will be possible to change the genes control system drastically. This viewpoint may open a new approach to control the gene network including reprogramming treatment.

References

1. K. Takahashi, S. Yamanaka, *Cell* **126,** 663 (2006).
2. K. Takahashi, *et al.*, *Cell* **131,** (2007).
3. K. Okita, T. Ichisaka, S. Yamanaka, *Nature* **448**, 313 (2007).
4. M. Wernig, *et al.*, *Nature* **448**, 318 (2007).
5. N. Maherali, *et al.*, *Cell Stem Cell* **1,** 55 (2007).
6. J. Yu, *et al.*, *Science* **318**, 1917 (2007).
7. J. B. Gurdon, T. R. Elsdale, M. Fischberg, *Nature* **182**, 64 (1958).
8. I. Wilmut, *et al.*, *Nature* **385**, 810 (1997).
9. M. Tada, *et al.*, *Curr. Biol.* **11**, 1553 (2001).
10. C. A. Cowan, *et al.*, *Science* **309**, 1369 (2005).
11. L. A. Boyer, *et al.*, *Cell* **122,** 947 (2005).
12. H. Niwa, *et al.*, *Development* **134**, 635 (2007).
13. J. Kim, *et al.*, *Cell* **132,** 1049 (2008).
14. M. B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
15. T. S. Gardner, *et al.*, *Nature* **403**, 339 (2000).
16. J. Hanna, *et al.*, *Nature* **462**, 595 (2009).
17. T. S. Mikkelsen, *et al.*, *Nature* **454**, 49 (2006).

School of Engineering
University of Hyogo
Himeji, Hyogo 671-2201
Japan
E-mail address: tyus51kzhs@ares.eonet.ne.jp