A Markov model of induced pluripotency

Zhirui hu[2], Minping Qian[3], and Michael Q. Zhang[1,2]

***Short Abstract* —Somatic cells can be reprogrammed to embryonic stem cell-like state by introducing few reprogramming factors. Inspired by the phenomenological reprogramming model of Artyomov et al. (2010), we proposed a novel Markov model with more realistic gene regulation rules and explored various properties of the model with Monte Carlo simulation. Furthermore, we demonstrated the utility of our model by testing it with the real dynamic gene expression data spanning across different intermediate stages in the iPSC reprogramming process. The data can be predicted and explained by our model reasonably well; in turn, it lands further support on our general rules of gene regulation in iPSC reprogramming.**

***Keywords* —Markov model, iPSC reprogramming**

I. Background

Somatic cells can be reprogrammed to induced-pluripotent stem cells (iPSCs) by introducing few reprogramming factors (such as Oct3/4, Sox2, c-Myc and Klf4) [1]. However, the mechanism of induced pluripotency is still unknown.

Experiments showed that gene expression and epigenetic modifications change dramatically in the reprogramming process [2,3], which makes us depict the reprogramming process as pushing the cell going up epigenetic energy landscape by a set of reprogramming factors. Although all the cells have the potency to be reprogrammed, only the cells having overcome all the epigenetic barriers can be reprogrammed to the iPSC state, which partly explains the low efficiency of reprogramming.

In this framework, we proposed a novel Markov model, stepwise reprogramming Markov (SRM) model, depicting that the somatic cells dedifferentiate towards pluripotent state gradually and stochastically in the reverse order of pluripotent stem cell differentiation.

II. Methods

We modeled the reprogramming process by a Markov chain whose states are different differentiation stages of a cell, cell death or intermediate state in reprogramming. We assumed that cell differentiation forms a binary tree and group of cell type specific genes can be granulated as a module to represent each cell type. Thus states in the Markov chain can be defined by the "open" or "close" of expression and epigenetic states of the modules. The transition probability can be calculated by two phenomenological rules of gene and epigenetic interaction of within a module or between a pair of modules. As reprogramming factors activate or repress a module stochastically, the cells starting from somatic state transit between the states in the Markov chain and eventually reprogram to iPSC state or die.

A. Estimating reprogramming rate and time

Reprogramming rate is estimated by the probability of cells getting to iPSC state and reprogramming time is the expectation of number of transitions getting to iPSC state.

Furthermore, we modified our model and showed that reprogramming rate would increase in the condition of knockdown of somatic transcription factors or inhibition of DNA methylation globally. Our estimation of reprogramming rate and time is consistent with the real experiments.

B. Simulating gene expression changes in reprogramming

Expression of cell type specific genes in each module can be computed by the probability of cells reaching each state of the Markov chain from somatic cell state at different time points. The results agree with the real time-serial reprogramming expression profiles of gene clusters enriched in corresponding cell type specific genes.

III. Conclusion

Our model explored general rules of gene regulation in iPSC reprogramming, correctly predicted reprogramming rate under several typically experiment conditions and partly explained gene expression change during reprogramming, which may help uncover the basic mechanism of reprogramming and improve the efficiency of converting somatic cells to iPSCs.

Acknowledgements: This work was funded by NSFC grants 91019016, 31061160497 and NIH grant HG001696.

[1]Department of Molecular and Cell Biology, Center for Systems Biology, the University of Texas at Dallas, 800 West Campbell Road, RL11 Richardson, TX 75080-3021, USA. E-mail: michael.zhang@utdallas.edu

[2]MOE Key Laboratory of Bioinformatics and Bioinformatics Div, TNLIST /Department of Automation, Tsinghua University, Beijing 100084, China. E-mail: hzr07@mails.tsinghua.edu.cn

[3]School of Mathematics, Peking University, Beijing 100871, China. E-mail: qianmp@math.pku.edu.cn

References

[1] Kazutoshi T, Shinya Y (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676.

[2] Richard PK, et al. (2011) Reprogramming Factor Expression Initiates Widespresd Targeted Chromatin Remodeling. *Cell Stem Cell* **8**, 96-105.

[3] Tarjei SM, et al. (2008) Dissecting direct reprogramming through integrative genomic analysis. *Nature*, **454**, 49-55.

[4] Maxim NA, Alexander M, Arup KC (2010) A model for Genetic and Epigenetic Regulatory Networks Identifies Rare Pathways for Transcription Factor Induced Pluripotency. *PLoS Comput Biol* **6(5)**, e1000785.