PROGRESS

Elite and stochastic models for induced pluripotent stem cell generation

Shinya Yamanaka[1,2]

Induced pluripotent stem cells offer unprecedented potential for disease research, drug screening, toxicology and regenerative medicine. However, the process of reprogramming is inefficient and often incomplete. Here I consider reasons for bottlenecks in induced pluripotent stem cell generation, and propose a model in which most or all cells have the potential to become pluripotent.

In 2000, my laboratory began testing the idea that factors that maintain pluripotency in embryonic stem (ES) cells might induce pluripotency in somatic cells. Twenty-four factors were selected as initial candidates, on the basis of their important roles or specific expression in mouse ES cells.

To evaluate these factors, we introduced combinations of the genes into mouse embryonic fibroblasts (MEFs) using a retroviral vector. The cells were engineered to carry an antibiotic-resistance gene that would only be expressed when *Fbxo15*, one of the mouse ES cell pluripotency genes, was turned on. We predicted that this gene could also be turned on if pluripotency was induced by a combination of the 24 genes[1].

When each candidate gene was individually introduced into the fibroblasts, no colonies grew. However, when we mixed retroviruses expressing all 24 of the candidate genes, several colonies emerged. Much to our surprise, we found that only four of these factors were needed to generate a small number of stem-cell-like colonies. All of the four genes were transcription factors: namely Oct3/4 (also known as Pou5f1), Sox2, Klf4 and c-Myc, and we found they could reprogram fibroblasts from both embryonic and adult mice[2].

The reprogrammed cells, which we termed 'induced pluripotent stem (iPS) cells', were similar to ES cells in their morphology, expression of important ES cell marker genes, and their ability to form teratomas (tumours comprised of diverse tissue types) when injected into mouse testes. However, the initial iPS cells possessed different global gene expression patterns, and when injected into early mouse embryos, they failed to populate the embryo and form adult chimaeric mice. These characteristics indicated that the iPS cells were not fully reprogrammed. However, eventually, by modification of the induction protocols, our group and others generated mouse iPS cells competent for adult chimaeric mice and germline transmission[3–5]. In 2007, human fibroblasts were reprogrammed using the same or slightly modified combination of genes introduced by retroviruses or lentiviruses[6–9].

Although it is possible to reproducibly generate iPS cells by viral transduction of these defined factors, only a small portion of the transduced cells become pluripotent. In the original report of germline-competent iPS cells, the efficiency was only ~0.05%—that is, on average only one out of 2,000 plated fibroblasts formed pluripotent cells[4]. Furthermore, several groups reported that the cells that seemed to be pluripotent were in fact often only partially reprogrammed, being dependent on continuous transgene expression of the reprogramming factors for their self-renewal[2,10,11].

Low efficiency and partial reprogramming are barriers for applying human iPS cells to basic research, drug screening, toxicology and regenerative medicine. In this Progress article, I will consider two models, the elite and stochastic models, to explain the low efficiency and partial nature of iPS cell generation (Fig. 1). The elite model predicts that direct reprogramming only take place in a subset of

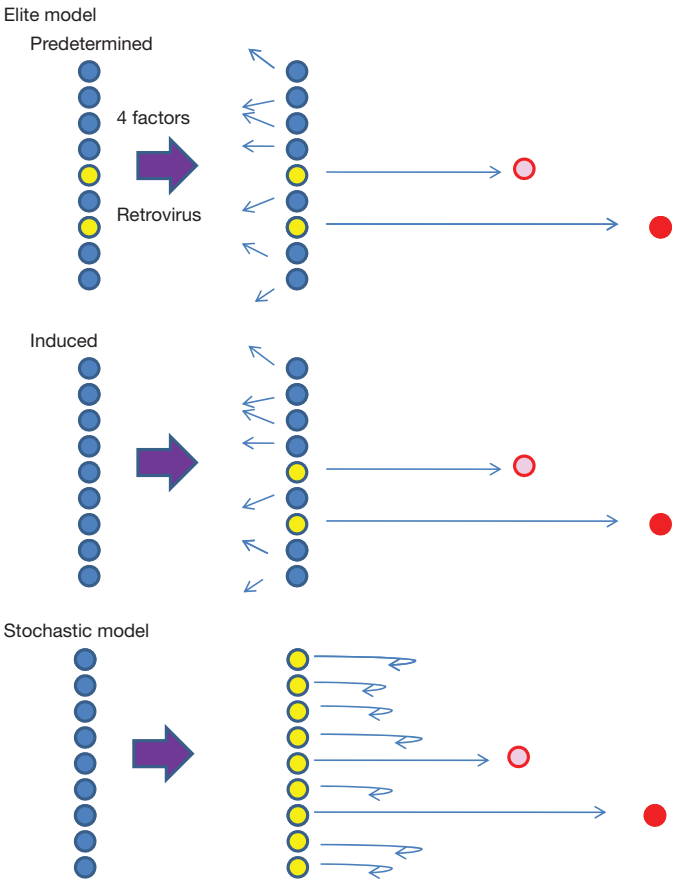

Figure 1 | Two models explaining the low efficiency of iPS cell generation. In the elite model, only a small number of cells, determined either before or after retroviral transduction, can be reprogrammed either partially or completely. In the stochastic model, most cells initiate the reprogramming process, but only a few can achieve complete reprogramming. Yellow, cell competent for reprogramming; pink, partially reprogrammed cells; red, iPS cells.

[1]Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto 606-8507, Japan. [2]Gladstone Institute of Cardiovascular Disease, San Francisco, California 94158, USA.

transduced cells, whereas the stochastic model predicts that most or all cells are competent for reprogramming. My review of the current literature supports the stochastic model.

Elite model

This model presupposes that only a few cells are competent for reprogramming. This model can be further divided into two models: a 'predetermined elite' and an 'induced elite' model (Fig. 1).

Predetermined elite model. In the predetermined model, small numbers of cells are competent for reprogramming even before retroviral transduction of the four factors. Tissue stem cells and other undifferentiated cells existing in regenerative tissues are good candidates for 'elite' cells, predisposed to reprogramming. In reprogramming by nuclear transfer (somatic cell nuclear transfer, or SCNT), higher efficiencies were achieved using nuclei from less-differentiated, progenitor-like neural stem cells, as well as from ES cells, than with nuclei from terminally differentiated neuronal donor cells such as lymphocytes[12]. Similarly, in inducing pluripotency using defined factors, more primitive stem cells may be preferentially reprogrammed. Multipotent stem cells are known to exist in adult organs and tissues such as the skin. In skin, stem cells comprise ~0.067% of the population[13], a number markedly similar to the efficiency of iPS cell generation in our initial report.

However, four lines of evidence contradict the predetermined elite model. First, the efficiency of iPS cell generation is now much higher than originally reported. By simply delaying the timing of drug selection for Nanog expression, we increased the efficiency 10-fold, showing that ~0.5% of MEFs can become iPS cells[14]. Furthermore, retroviral infection of MEFs routinely results in ~2% reprogramming efficiency in my laboratory. This can be further increased by the use of specific chemicals. It has been reported that up to 10% of MEFs are reprogrammed by the four factors when treated with a small molecule, valproic acid[15]. It is unlikely that tissue stem cells or otherwise undifferentiated cells comprise 2–10% of primary fibroblast cultures. Although reprogramming factors or chemical treatments might preferentially enhance the proliferation of tissue stem or progenitor cells, it is more likely that a wider spectrum of somatic cells is reprogrammed in the formation of iPS cells.

A separate line of evidence for the stochastic model arose from genetic lineage tracing analyses. In addition to fibroblasts, iPS cells have been generated from various tissues, including the liver[16] and pancreas[17]. Genetic lineage tracing analyses, using the Cre-*loxP* system, showed that most iPS cells obtained from liver originate from cells that have expressed albumin. Similarly, many iPS cells from pancreas originate from cells that have expressed insulin. These data do not prove that terminally differentiated cells can become iPS cells, because albumin and insulin are expressed in precursor cells as well as mature hepatocytes and pancreatic beta-cells. Nevertheless, these findings clearly demonstrate that the four factors can reprogram lineage-committed cells that had at least differentiated into stages in which the expression of either albumin or insulin is turned on.

More directly, Hanna *et al.* generated iPS cells from B lymphocytes[18]. They confirmed the origin to be terminally differentiated B cells by showing genetic recombination of the globulin locus. Although the ectopic expression of the myeloid transcription factor CCAAT/enhancer binding protein alpha (CEBPα) or specific knockdown of the B cell transcription factor Pax5 was required, as well as the introduction of the four reprogramming factors, their data nevertheless demonstrated that lineage-committed cells can be reprogrammed by these defined factors.

The lines of evidence described earlier suggest that many, if not all, lineage-committed somatic cells can become iPS cells. However, some cell types might be more easily reprogrammed by defined factors than others. Indeed, mouse neural stem cells can be directly reprogrammed by only one factor, Oct3/4 (ref. 19).

Induced elite model. In the induced model, genes of factor(s) other than the four factors must be activated or inactivated by viral integration into the host genome (Fig. 1). Therefore, only cells with specific viral integration sites are competent for reprogramming. Many lines of evidence, however, do not support this model.

iPS cells derived from epithelial cells and tissues, such as the liver, gastric mucosa and skin[20], have less retroviral integration than iPS cells derived from fibroblasts. Taking advantage of this phenomenon, we performed inverse PCR to determine the retroviral integration sites in two iPS cell clones from hepatocytes and two from the stomach[16]. No common integration sites were identified. More recently, it was reported that six iPS cell clones derived from mouse fibroblasts showed no common retroviral vector insertions[21]. These data suggest that retroviral integration into specific sites is not required for iPS cell generation.

Further direct evidence came from several groups who generated iPS cells without retroviruses. iPS cells were generated from mouse adult hepatocytes by introducing the four reprogramming factors with adenoviruses[22]. Another group generated iPS cells from MEFs with two plasmids—one with *Oct3/4*, *Sox2* and *Klf4* complementary DNAs connected by the 2A self-cleavage peptides, and the other with *c-Myc* (also known as *Myc*) cDNA[23]. Both groups showed that these iPS cells did not contain the integration of adenoviruses or plasmids into the genomes of the established iPS cells. More recently, mouse iPS cells were generated by protein transduction of the four factors[24]. Furthermore, human iPS cells have been generated with episomal expression vectors[25]. In some of their iPS cell clones, the episomal DNA spontaneously disappeared during the course of culture.

However, iPS cell generation without retroviruses is far less efficient. For the episomal induction, seven factors, including the potent oncogene SV40, were required. This may suggest that insertional mutagenesis is not required for iPS cell generation but does promote the process. In this case, integration sites do not have to be common among different clones. The generation of iPS cells could be promoted by activating or inactivating endogenous genes by retroviruses or lentiviruses, so as to enhance proliferation, decrease apoptosis, or enhance reprogramming. Furthermore, the amount, balance, continuity and silencing of transgene expression can be greatly influenced

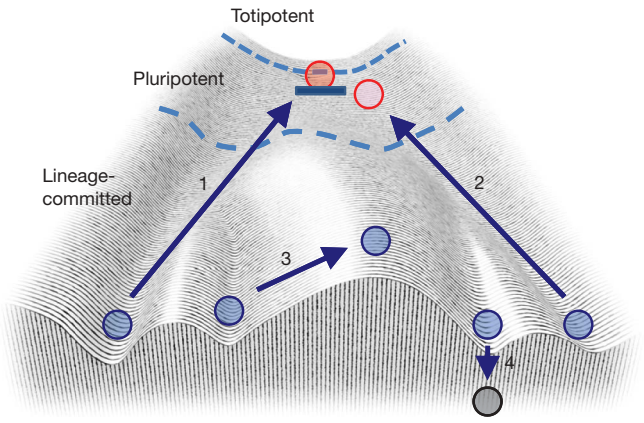

Figure 2 | Stochastic model. Totipotent fertilized eggs differentiate into various lineages through the pluripotent status. Here I depict this process in the context of the epigenetic landscape proposed by Conrad Waddington[26]. The iPS cell is like a ball rolling down the slope of a valley. The reprogramming factors cooperatively push cells up the slope to the pluripotent zone. Some cells are blocked by an epigenetic bump (closed rectangle) on the slope and thus become able to self-renew (1). Other cells are only partially reprogrammed and are not blocked by the bump; therefore, without the exogenous reprogramming factors, they would roll down again (2). When the expression of the reprogramming factors is not appropriate, cells may transform to other types of cell (3), or even undergo apoptosis or senescence (4). For cells that are located on the middle of the valley (that is, somatic stem cells) it might be easier to go back to the pluripotent state. Figure modified from Waddington, 1957 (ref. 26).

by the positions of retroviral integrations. This may explain why only a small portion of transduced cells complete the reprogramming process, as discussed later.

Stochastic model

In the stochastic model, most, if not all, differentiated cells have the potential to become iPS cells after the introduction of the four factors. Cell differentiation is often described as a ball rolling down an epigenetic landscape, first depicted by Conrad Waddington in 1957, starting from the totipotent state, going through the pluripotent state, and rolling down to a lineage-committed state[26] (Fig. 2). In normal development, pluripotent cells appear transiently. They cannot stop on the slope and are pulled by gravity to rapidly differentiate into various lineages. In contrast, ES cells can self-renew and maintain pluripotency for a long time. Thus, it is as if ES cells are blocked by a bump or a roadblock formed by their particular epigenetic status. In this metaphor, the four reprogramming factors cooperatively push the cells up into the pluripotent state.

There are at least two requirements for complete reprogramming. First, the four factors must be expressed in a pattern that provides a sufficient push in the right direction. Because available technologies cannot precisely control the expression levels of the four transgenes, this first requirement can only be achieved stochastically. Second, cells must be blocked by the epigenetic bump so that they can remain in the pluripotent zone even when the transgene expression disappears. Because the four reprogramming factors alone cannot constitute such an epigenetic roadblock, stochastic events again are required during iPS cell generation. As has been demonstrated in reprogramming by nuclear transfer, DNA methylation and histone modifications probably have important roles in iPS cell generation.

Expression patterns of reprogramming factor. Direct reprogramming probably depends on the amount, balance, continuity and silencing of the transgene expression of the four factors. The high copy numbers of proviruses in each iPS cell clone suggest that strong transgene expression is initially required. Klf4 is expressed in fibroblasts, but its endogenous expression is not sufficient for iPS cell generation. The endogenous *c-Myc* gene is also expressed in fibroblasts, but its ectopic expression significantly enhances the efficiency of iPS cell generation in both mice and humans[14,27].

iPS cell induction may also depend on the specific stochiometric balance of the four factors. For example, excess Oct-3/4 (ref. 28) and Sox2 (ref. 29) are detrimental to the maintenance of pluripotency. In fact, in neural stem cells that express the endogenous *Sox2* gene, the efficiency of iPS cell generation is higher with ectopic expression of three factors devoid of *Sox2* than with the combination of all four factors[30]. The balance between c-Myc and Klf4 may also be crucial for preventing apoptosis and senescence caused by the overexpression of these tumour-related genes[31]. An inappropriate balance of the four factors would result in improper reprogramming, senescence or apoptosis.

The continuity and silencing of the transgene expression are also important. Transgene expression must be maintained during the first 10 to 14 days[32,33]. In this respect, retroviral vectors clearly have an advantage over plasmids and adenoviruses. However, to achieve complete reprogramming, transgene expression should be silenced and then taken over by the endogenous genes after this initial stage. The failure to achieve such transgene silencing could result in so-called partially reprogrammed cells, which possess ES cell morphology and express some ES cell marker genes, but have a limited ability to differentiate[2,10,11].

The amount, balance, continuity and silencing of the transgene expression can be greatly influenced by gene-delivery methods. With retroviruses and lentiviruses, positions of proviral integrations should have strong effects. This may explain why only a small portion of transduced cells complete the reprogramming process, because each transduced cell has a unique integration pattern. If this is the case, then one would expect a marked increase in efficiency when iPS cells are generated from fibroblasts containing the reprogramming-competent integration of transgenes.

This experiment was first performed by Wernig *et al.*, who isolated MEFs from chimaeric embryos after the injection of doxycycline-induced iPS cells into blastocysts, and then generated 'secondary' iPS cells by treating these MEFs with doxycycline[34]. A similar approach was also used in human cells[35]. Primary human iPS cells were generated with doxycycline-regulated lentiviruses, which induced differentiation into fibroblasts, and then secondary iPS cells were generated with doxycycline. In these experiments, up to 4% of fibroblasts became iPS cells. More recently, a comparable method using the *piggyBac* transposon system demonstrated ~20% reprogramming efficiency[36]. These findings indicate that cells containing the proper patterns of transgene integration could efficiently become iPS cells.

Epigenetic requirements for reprogramming. Even when the four factors are suitably expressed and able to move cells up the valleys, cells would roll down again without the transgene expression. Cells have to be blocked by a bump formed by a specific epigenetic status (Fig. 2). The endogenous loci of the four factors should be fully activated. An important requirement is proper DNA methylation. The promoter regions of many pluripotency associated genes are heavily methylated in fibroblasts and somatic cells, but are hypomethylated in ES cells and iPS cells[37]. Therefore, during direct reprogramming, DNA demethylation of these regions should be accomplished. Because the four factors do not have intrinsic DNA demethylation activity, the process is probably a secondary effect, requiring several cell divisions. This may be one reason why iPS cell generation is so slow and inefficient. The fact that demethylation-promoting agents, such as 5-azacytidine, promote iPS cell generation supports this model[10].

Generation of iPS cells should also require the proper reprogramming of histone modifications. In both ES cells and iPS cells, histone H3 and H4 are hyperacetylated in the promoter regions of pluripotency associated genes. In contrast, differentiated cells have hypoacetylated H4. Therefore, H4 of these regions should be acetylated during iPS cell generation. Because the three transcription factors themselves do not have histone modification activities, other factors are required. One function of c-Myc may be to recruit histone acetyltransferase to the target genes[38]. The fact that a histone deacetylase inhibitor, such as valproic acid, enhances the efficiency of iPS cell generation supports the notion that histone acetylation is important in direct reprogramming[15,39].

Histone methylation also is likely to be important in suppressing iPS cell generation. In both ES cells and iPS cells, H3 is methylated at lysine 4 and demethylated at lysine 9 in the promoter regions of pluripotency associated genes. Fibroblasts have the opposite patterns of histone modifications. Furthermore, both ES cells and iPS cells have a so-called bivalent chromatin structure of developmental genes, consisting of methylation of both of H3 lysines 27 and 4 (ref. 40). These histone methylation states have to be established for iPS cell generation.

The iPS technology is still at its infancy. Nevertheless, its potential is enormous. Patient- or disease-specific iPS cells should provide unprecedented cell sources for better understanding the pathogenesis of diseases and for developing safer and more effective drugs. Furthermore, iPS cell technology should some day make it possible to perform cell transplantation therapies for a wide variety of diseases and injuries, while circumventing ethical issues and immune rejection. To realize the clinical applications, we have to achieve complete and uniform reprogramming in iPS cells. Failure to do this would result in resistance to differentiation and increase the risk of teratoma formation. The stochastic model predicts that iPS cells can be generated from a variety of somatic cells with a variety of methods. We have to evaluate different original cells and induction methods to determine the best combination to allow us to generate the safest iPS cells for clinical application.

Received 20 February; accepted 29 May 2009.

1. Tokuzawa, Y. *et al.* Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. *Mol. Cell. Biol.* **23,** 2699–2708 (2003).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
This paper is the first demonstration of direct reprogramming by defined factors.
3. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES cell-like state. *Nature* **448,** 318–324 (2007).
4. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germ-line competent induced pluripotent stem cells. *Nature* **448,** 313–317 (2007).
5. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodelling and widespread tissue contribution. *Cell Stem Cell* **1,** 55–70 (2007).
6. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
7. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–1920 (2007).
8. Lowry, W. E. *et al.* Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl Acad. Sci. USA* **105,** 2883–2888 (2008).
9. Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451,** 141–146 (2008).
10. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
11. Silva, J. *et al.* Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6,** e253 (2008).
12. Blelloch, R. *et al.* Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. *Stem Cells* **24,** 2007–2013 (2006).
13. Toma, J. G. *et al.* Isolation of multipotent adult stem cells from the dermis of mammalian skin. *Nature Cell Biol.* **3,** 778–784 (2001).
14. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnol.* **26,** 101–106 (2008).
15. Huangfu, D. *et al.* Induction of pluripotent stem cells from primary human fibroblasts with only *Oct4* and *Sox2*. *Nature Biotechnol.* **26,** 1269–1275 (2008).
16. Aoi, T. *et al.* Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321,** 699–702 (2008).
17. Stadtfeld, M., Brennand, K. & Hochedlinger, K. Reprogramming of pancreatic Beta cells into induced pluripotent stem cells. *Curr. Biol.* **18,** 890–894 (2008).
18. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133,** 250–264 (2008).
19. Kim, J. B. *et al.* Oct4-induced pluripotency in adult neural stem cells. *Cell* **136,** 411–419 (2009).
20. Aasen, T. *et al.* Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. *Nature Biotechnol.* **26,** 1276–1284 (2008).
21. Varas, F. *et al.* Fibroblast-derived induced pluripotent stem cells show no common retroviral vector insertions. *Stem Cells* **27,** 300–306 (2008).
22. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced pluripotent stem cells generated without viral integration. *Science* **322,** 945–949 (2008).
This paper is the first demonstration of integration-free iPS cell generation.
23. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322,** 949–953 (2008).
24. Zhou, H. *et al.* Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* **4,** 381–384 (2009).
This paper is the first demonstration of iPS cell generation by protein transduction.
25. Yu, J. *et al.* Human induced pluripotent stem cells free of vector and transgene sequences. *Science* **324,** 797–801 (2009).
26. Waddington, C. H. *The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology* (Alen & Unwin, 1957).
27. Wernig, M., Meissner, A., Cassady, J. P. & Jaenisch, R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* **2,** 10–12 (2008).
28. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24,** 372–376 (2000).
29. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* **26,** 903–911 (2008).
30. Eminli, S., Utikal, J. S., Arnold, K., Jaenisch, R. & Hochedlinger, K. Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. *Stem Cells* **26,** 2467–2474 (2008).
31. Rowland, B. D. & Peeper, D. S. KLF4, p21 and context-dependent opposing forces in cancer. *Nature Rev. Cancer* **6,** 11–23 (2006).
32. Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2,** 151–159 (2008).
33. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2,** 230–240 (2008).
34. Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotechnol.* **26,** 916–924 (2008).
35. Maherali, N. *et al.* A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* **3,** 340–345 (2008).
36. Woltjen, K. *et al. piggyBac* transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458,** 766–770 (2009).
37. Imamura, M. *et al.* Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.* **6,** 34 (2006).
38. Knoepfler, P. S. *et al.* Myc influences global chromatin structure. *EMBO J.* **25,** 2723–2734 (2006).
39. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnol.* **26,** 795–797 (2008).
This paper presents considerable improvement of iPS cell generation by a defined chemical.
40. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).

Acknowledgements I thank B. Conklin and G. Howard for critical reading of the manuscript, and members of my laboratory for scientific and administrative support. I apologize to researchers whose work could not be cited owing to space constraints.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to S.Y. (yamanaka@cira.kyoto-u.ac.jp).