

Regulation of Pluripotency and Reprogramming by Transcription Factors*

Published, JBC Papers in Press, September 26, 2008, DOI 10.1074/jbc.R800063200
Duanqing Pei¹

From the Stem Cell and Cancer Biology Group, Laboratory of Regenerative Biology, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institute of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510663, China

Living organisms, from virus to human, rely on the transcription machinery to express specific parts of their genomes to execute critical biological functions during their life cycle by responding to environmental or developmental signals. Thus, transcription constitutes a critical step in regulating biological processes, and transcription factors have been considered as master switches for cell fate determination. Stem cell biology has benefited from rapid advances in recent years, largely because of the characterization of several transcription factors as master regulators of stem cell pluripotency. The same factors, *viz.* Oct4, Sox2, Nanog, Klf4, and Myc, have been shown to possess the magic power to reprogram somatic cells into pluripotent ones, a remarkable achievement with both practical and theoretical implications. This minireview summarizes recent advances in pluripotency and reprogramming by focusing on key transcription factors and the likely mechanisms.

Transcription factors often act in concert with cofactors and modifiers and turn on or off the expression of downstream genes in response to developmental cues or environmental signals (1). As such, a significant number of transcription factors have been shown to specify cell fate during development, presumably by controlling the expression of cell type-specific genes (2). ES² cells are good model systems for the studies of cell differentiation and fate determination and logically for the biochemical analysis of relevant transcription factors.

The first ES cells were isolated from the inner cell mass of mouse blastocysts in 1981 by Evans and Kaufman (3) and Martin (4), who devised methods to grow them indefinitely. These cells are pluripotent because they can form chimera when reintroduced into mouse blastocysts and contribute to the formation of all tissues, including the germ line (5). This breakthrough led to gene targeting by homologous recombination in

ES cells and the generation of knock-out animals (6). In 1998, human ES cells were isolated successfully by Thomson *et al.* (7) and heralded in a new era of hope that stem cell technology may eventually benefit human disease therapy.

Stem cell research over the past decade or so has begun to infiltrate into many disciplines in biology and medicine. The trend may continue to crown stem cells as the central paradigm of biomedical research. First, stem cells, both embryonic and adult, hold the key for regenerative medicine, which may be considered the third therapeutic modality after drug therapy and surgery. Bone marrow transplantations have been successful in treating multiple diseases through replacement of diseased or deficient hematopoietic stem cells. Second, stem cells, especially ES cells, are ideal models for basic research in fields such as signal transduction, development, and epigenetics. Last, stem cells could be useful tools for drug screening and safety assessments. Despite the excitement associated with stem cell research, we are still in the early phase of our exploration toward a molecular understanding of stem cells in normal development, diseases, and regenerations.

Recent advances in understanding the molecular mechanisms governing ES cell pluripotency have provided insights into the role of transcription factors such as Oct4 and Nanog in maintaining ES cells in the undifferentiated state (8–11). Remarkably, these pluripotency factors, Oct4, Sox2, and Nanog, have also been shown to participate in the reprogramming of differentiated cells back to pluripotent states (12, 13). In this minireview, I will attempt to explain pluripotency and reprogramming in the context of transcription factors and will discuss the related challenges and opportunities.

Basic Concepts

Pluripotency is the central property of all ES cells. It refers to the ability to generate any type of cells in the body. Developmentally, zygotes are totipotent, capable of giving rise to a whole animal, including all cell types. After several cell divisions, zygotes differentiate into blastocysts, from which the inner cell mass can be isolated and cultured into ES cells. Thus, ES cells are developmentally arrested at the pluripotent stage and can be propagated indefinitely *in vitro* (14). Although less potent than totipotent zygotes, ES cells have been experimentally proven to be able to contribute to all cell types except the trophectoderm in mouse (15). Using a dominant-negative form of Sox2, we have demonstrated that ES cells can be differentiated into trophoblast-like cells (16). Therefore, ES cells appear to have the potential to generate all cell types of the body. However, ES cells are generally described as being pluripotent, not totipotent, reflecting the fact that no animal has been generated by ES cells alone (16).

Pluripotency can be experimentally verified by the ability of stem cells to contribute to embryonic development and to generate chimeric animals after being injected into blastocysts (15). This can be achieved routinely for mouse ES cells. The pluripotency of human or primate ES cells is assessed by teratoma formation upon transplantation into severe combined immunodeficient or nude mice (15). Less stringent tests for pluripo-

* This work was supported in part by National Natural Science Foundation of China Grants 30725012 and 30630039; Chinese Academy of Sciences Grant KSCX2-YW-R-48; Guangzhou Science and Technology Grant 2006A50104002; and Ministry of Science and Technology 973 Grants 2006CB701504, 2006CB943600, 2007CB948002, and 2007CB947804. This minireview will be reprinted in the 2009 Minireview Compendium, which will be available in January, 2010.

¹ To whom correspondence should be addressed. E-mail: pei_duanqing@gibh.ac.cn.

² The abbreviations used are: ES, embryonic stem; LIF, leukemia inhibitory factor; iPS, induced pluripotent stem (cell).

MINIREVIEW: Transcription Control of Pluripotency/Reprogramming

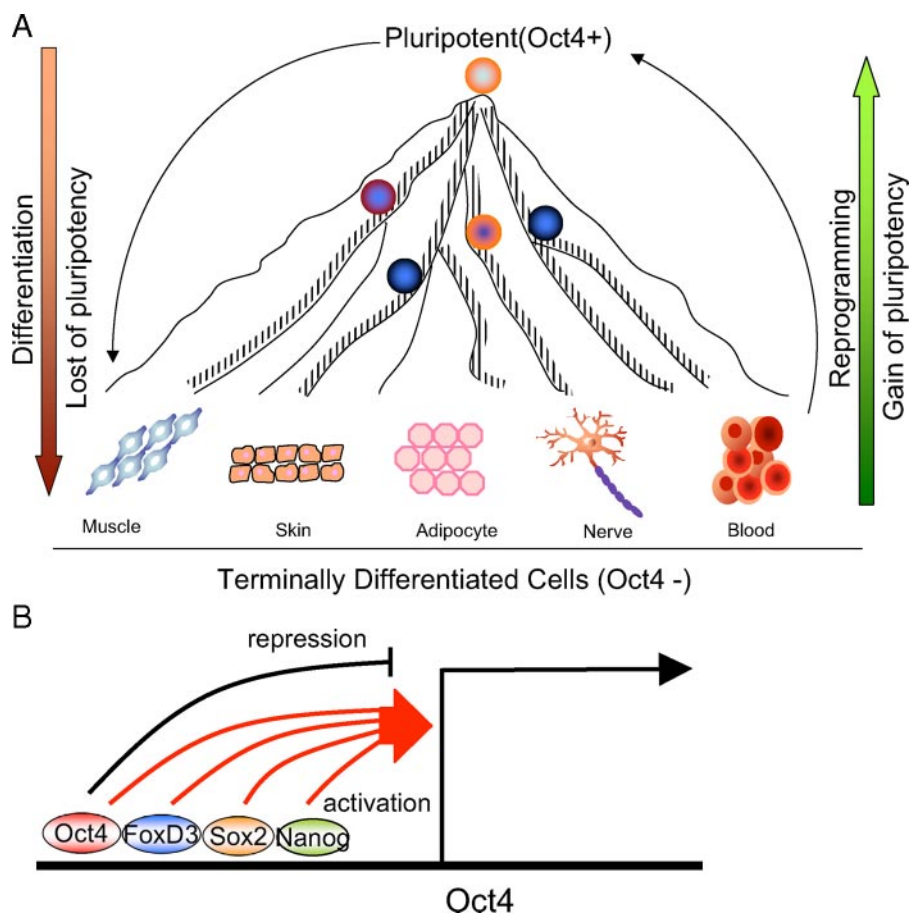


FIGURE 1. Stem cell pluripotency. *A*, the three emphases of stem cell biology, self-renewal, differentiation, and reprogramming, are depicted here to show their relationship to pluripotency. The pluripotent stem cells may be viewed as being positioned at the top, ready to differentiate spontaneously into various cell types of tissues and organs (*bottom*). The pluripotent state is maintained by self-renewal at the top by maintaining the expression of *oct4* at the optimal level. Pluripotent stem cells differentiate by losing pluripotency into a specific lineage accompanied by the gradual loss of expression for pluripotent genes and the activation of differentiation genes. Differentiated somatic cells regain pluripotency by reprogramming back to the pluripotent state (on the *right*). *B*, multiple factors are involved in the maintenance of *oct4* expression in ES cells.

tency include the generation of embryonic bodies *in vitro* and the detection of molecular markers representing the three germ layers (15). In addition, special protocols can be utilized to differentiate ES cells into a particular cellular lineage such as muscle or neurons. Thus, the differentiation potential of ES cells can be demonstrated experimentally both *in vitro* and *in vivo*.

Pluripotency is maintained by a process called self-renewal. Self-renewal allows ES cells to duplicate themselves without losing the ability to differentiate, thus maintaining pluripotency. This can be achieved through both symmetric and asymmetric cell divisions (17). *In vitro*, ES cells undergo self-renewal through symmetric divisions. *In vivo*, tissue stem cells tend to self-renew through asymmetric divisions to generate one exact copy and another one for differentiation.

Experimentally, ES cells have to be grown under special conditions to be kept in a pluripotent state (3, 4). Mouse ES cells should be cultured on top of a feeder layer of cells, presumably supplying unknown factors to the ES cells (3, 4). In addition, LIF or other cytokines are routinely added to prevent ES cells from undergoing spontaneous differentiation, a phenomenon encountered on a daily basis during ES cell culturing. Human ES cells appear to have

different requirement for cytokines. Instead of LIF, human ES cells require both bone morphogenetic protein and fibroblast growth factors to prevent differentiation (7). The removal of feeders or cytokines leads to spontaneous differentiation and the loss of pluripotency.

Differentiation is the process during which pluripotency is expressed, e.g. the generation of all 220 or so cell types of our body from ES cells. During differentiation, stem cells commit to one cell lineage while losing the ability to commit to the rest of the cell lineages.

Reprogramming is the process that converts differentiated cells back to pluripotent ones, effectively the reversal of differentiation (15, 18). Experimentally, reprogramming has been achieved through somatic cell nuclear transfer or cloning. More recently, iPS cell technology has accomplished the same feat via the introduction of pluripotency factors, including Oct4 and Sox2, into somatic cells (see below). Taken together, self-renewal, differentiation, and reprogramming can be viewed as three different aspects, *i.e.* maintenance, expression, and acquisition, of pluripotency (Fig. 1A).

Control of Stem Cell Pluripotency by Transcription Factors

oct4 was the first gene to be identified as a master regulator of pluripotency (9). Nichols *et al.* (9) demonstrated that *oct4*-deficient embryos develop to the blastocyst stage but that the inner cell mass cells are not pluripotent. In fact, *oct4* was originally discovered by Scholer *et al.* (19) as a member of the murine octamer-binding protein family that interacts specifically with the octamer motif, a transcription regulatory element found in the promoter and enhancer regions of many genes. The expression profile of Oct4 suggests that it may regulate cell fate during early developmental control (19). Biochemically, Oct4 has been shown to be a DNA-binding protein with a bipartite POU/homeodomain encoded by a 324-amino acid open reading frame (19). Oct4 relies on two transactivation domains flanking the DNA-binding domain to exert its transcription activities (20). Oct4 protein is synthesized in the cytosol and transported into the nuclei via a typical nuclear localization signal (21). The nuclear localization signal of Oct4 is required for its transcription activity, and its ablation leads to the generation of a dominant-negative form of Oct4, which is capable of inducing ES cell differentiation by interfering with wild-type Oct4 activity (21).

Downloaded from www.jbc.org at Biblioteca biomedica Universita' di Torino, on October 25, 2009

In ES cells, *oct4* appears to regulate cell fate in a dosage-dependent fashion (10). Using a conditional expression and repression system, Niwa *et al.* (10) demonstrated that the level of Oct4 activity specifies three distinct fates of ES cells: 1) a <2-fold increase in expression turns ES cells into primitive endoderm and mesoderm; 2) repression of *oct4* induces the formation of trophectoderm; and 3) only an optimal amount of Oct3/4 can sustain stem cell self-renewal. These results suggest that ES cells must possess a network of regulators to keep *oct4* expression at the optimal level to ensure pluripotency (20).

How many transcription factors are involved in the regulation of *oct4* expression? This was the question asked by several groups in light of the observation that *oct4* must be maintained in a narrow range of expression levels to ensure stem cell pluripotency. The discovery of Nanog offered a clear candidate for *oct4* regulation. Named after Tir Nan Og (the Land of the Young), Nanog was discovered based on its ability to sustain stem cell self-renewal in the absence of LIF (8, 11). Although it was originally believed that Nanog prevents ES cell from differentiation in the absence of LIF by repressing the expression of differentiation genes, a simple reporter assay demonstrated that Nanog possesses two potent transactivators (22, 23), suggesting that Nanog could be an activator of *oct4* expression. Indeed, Nanog behaves as a strong activator of the *oct4* promoter, thus participating in the regulation of *oct4* expression in ES cells (Fig. 1B) (24).

Sox2 often partners with Oct4 to regulate gene expression (25, 26). Like Oct4, Sox2 has also been implicated in the regulation of *Fgf4* expression (27). Gene knock-out experiment demonstrated that Sox2 is required for epiblast and extraembryonic ectoderm formation, suggesting that Sox2 and Oct4 cooperatively specify the fate of pluripotent stem cells at implantation (28). Recent results demonstrated that Sox2 is necessary for regulating multiple transcription factors that affect *oct4* expression, thus stabilizing ES cells in a pluripotent state by maintaining the requisite level of *oct4* expression (16, 29).

FoxD3, a member of the Forkhead transcription factor family, also activates *oct4* expression (24). FoxD3 is also required for epiblast formation in blastocysts and stem cell pluripotency in mouse (30). Although FoxD3^{-/-} blastocysts express Oct4, FoxD3 appears to be capable of activating the promoter of *oct4* in a sequence-specific manner (24).

Oct4 represses itself when overexpressed. Because Nanog, Sox2, and FoxD3 are activators for expression of *oct4*, it would be difficult to maintain its expression level at the desired optimal level as reported (10). Apparently, one way to counterbalance the activating force of these positive factors is for Oct4 to repress its own promoter when overexpressed (24). Alternatively, additional repressors may be uncovered in the future to provide the counterbalance for maintaining *oct4* expression levels in ES cells.

Oct4, Sox2, and Nanog regulate overlapping targets. Employing genome-wide location analysis in human ES cells, Boyer *et al.* (31) identified potential targets of three core transcription factors, Oct4, Sox2, and Nanog. Interestingly, they found that Oct4, Sox2, and Nanog co-occupy a substantial portion of these targets, including many developmentally important homeodomain proteins (31). Based on these large-scale

data sets, Boyer *et al.* (31) proposed that Oct4, Sox2, and Nanog collaborate to form regulatory circuitry consisting of autoregulatory and feed-forward loops that contribute to pluripotency and self-renewal. Similar approaches were attempted by other groups to identify these networks by high throughput technologies, and more elaborate mechanisms were proposed (32–35). However, more detailed analysis would be required to validate the proposed mechanisms and to delineate the regulatory logic of these networks.

Reprogramming by Transcription Factors

One of the goals in dissecting the molecular networks that control pluripotency is to regain pluripotency lost during development and differentiation. This would entail reversing the well programmed process of development from a fertilized egg to a grown adult. In higher mammals, it was thought that the differentiation process is irreversible until the successful cloning of Dolly (36). The cloning experiment demonstrated that somatic cells can be reprogrammed back to the totipotent zygotic state by the cellular factors of unfertilized eggs. It could have taken a considerable amount of time and effort to identify those unknown factors responsible for reprogramming. Takahashi and Yamanaka (12) leapfrogged this hurdle through a candidate gene approach. By analyzing genes highly expressed in ES cells, a pool of 24 genes were delivered to reprogram fibroblasts via retroviral transduction. To enhance their chance of success, a selection marker driven by *fbx15*, a gene known to be specifically expressed in ES cells, was engineered into the recipient cells by gene targeting. Remarkably, ES-like colonies were recovered after ~2 weeks. Eventually, only four genes, *oct4*, *sox2*, *klf4*, and *myc*, were deemed sufficient to reprogram fibroblasts into ES-like cells (12). These ES-like cells were later coined as iPS cells for induced pluripotent stem cells to differentiate them from blastocyst-derived pluripotent ES cells (12). The recapture of pluripotency lost during differentiation by these four magic factors established a new paradigm for our understanding of pluripotency. The molecular mechanisms associated with reprogramming could now be dissected in greater detail, as the iPS process is amicable to many hypothesis-driven investigations. This elegant iPS approach opened a new era for stem cell and regenerative medicine (12, 13, 15, 37).

iPS-mediated reprogramming of somatic cells removes the ethic as well as technical hurdles associated with therapeutic cloning, the use of human eggs for the generation of patient-specific pluripotent cells. Indeed, patient-specific iPS cells have been reported at an accelerated pace in the literature (38). These iPS cells may become important models for us to understand the mechanisms associated with a particular disease. However, given the use of viral delivery and four potent oncogenes in the iPS process, these patient-specific iPS cells are not safe for therapeutic purposes. Efforts are under way in many laboratories to identify small molecules that can functionally substitute for these four reprogramming factors (39, 40). As Oct4, Sox2, Myc, and Klf4 regulate specific signaling pathways, it is conceivable that a mixture of chemical regulators may function to reprogram somatic cells. The chemical approach, or ciPS for chemical iPS, to reprogramming may eventually yield

MINIREVIEW: Transcription Control of Pluripotency/Reprogramming

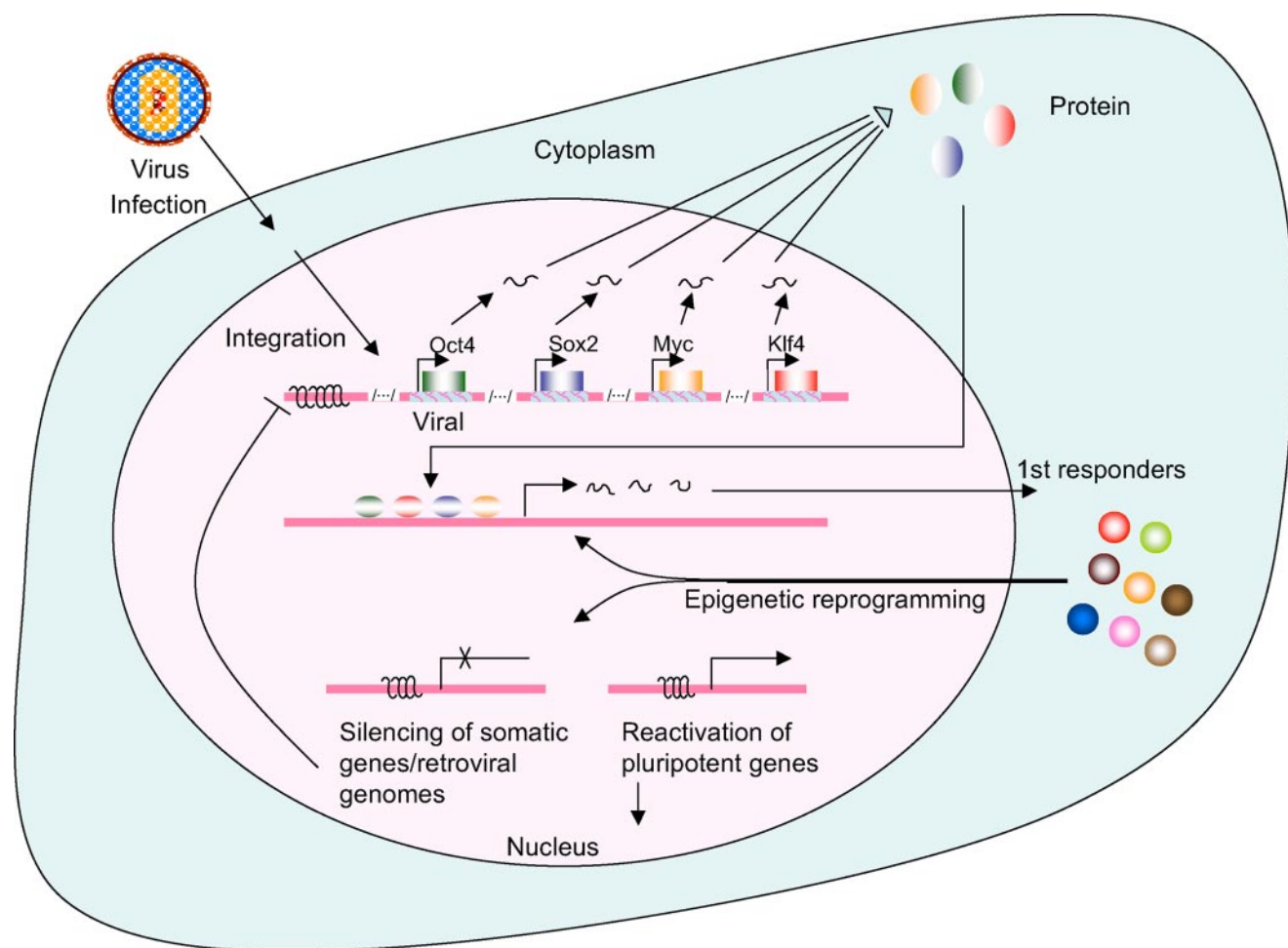


FIGURE 2. Mechanical steps involved in the reprogramming of somatic cells into pluripotent ones by Oct4/Sox2/Klf4/Myc (see description under “Challenges and Opportunities Ahead”).

clinical grade pluripotent stem cells for therapies and regenerative medicine.

Challenges and Opportunities Ahead

iPS is merely 2 years old, and much progress has been made in improving its efficiency and adapting it to various model organisms (15, 37, 38). Attempts have also been made to understand its mechanistic process, largely from high throughput technologies (41). However, we know very little how iPS works.

Mechanically, iPS reprogramming involves the following key steps (Fig. 2). First, the recombinant viruses carrying Oct4, Sox2, Klf4, and Myc enter the somatic cells and integrate into the host genomes. Following transcription driven by the viral promoters, all four proteins are produced in the cytosol and then imported back to the nuclei to activate the first wave of genes whose promoters are accessible to them. These first responders must then engage the epigenetic machinery to remodel the chromatin through the histone modification system and the DNA methylation system. Through this process, genes critical for pluripotency must be switched on by transcription factors and kept on through chromatin remodeling. Conversely, genes responsible for differentiation must be turned off by the transcription machinery and kept silent

through epigenetic mechanisms. One remarkable feature of the iPS process is the silencing of the integrated viral genomes carrying the reprogramming initiators Oct4, Sox2, Klf4, and Myc (15, 42). As such, iPS cells function indistinguishably from ES cells derived from blastocysts.

Each of these mechanical steps should be investigated to yield critical information for the reprogramming process. Biochemists may choose to focus on two critical events during iPS, *viz.* transcription activation/repression and epigenetic remodeling. As Oct4, Sox2, Myc, and Klf4 are known transcription factors, their mode of action now requires further investigation. For example, how do they turn their target genes on and off in concert with co-activators and the polymerase II complex? There is enough information on how each of these proteins contacts DNA and engages the transcription machinery. In light of the requirement for these four genes to work in concert, it is important to understand how they coordinate with each other on the chromatin and decide which ones to switch on and off. Unlike studies on individual promoter or single transcription factors, iPS involves multiple transcription factors and the whole genome. Therefore, new tools may be required to investigate both the transcription and epigenetic processes associated with iPS-mediated reprogramming.

REFERENCES

1. Kadonaga, J. T. (2004) *Cell* **116**, 247–257
2. Mannervik, M., Nibu, Y., Zhang, H., and Levine, M. (1999) *Science* **284**, 606–609
3. Evans, M. J., and Kaufman, M. H. (1981) *Nature* **292**, 154–156
4. Martin, G. R. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7634–7638
5. Kuehn, M. R., Bradley, A., Robertson, E. J., and Evans, M. J. (1987) *Nature* **326**, 295–298
6. Thomas, K. R., and Capecchi, M. R. (1990) *Nature* **346**, 847–850
7. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. (1998) *Science* **282**, 1145–1147
8. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003) *Cell* **113**, 643–655
9. Nichols, J., Zevnik, B., Anastasiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998) *Cell* **95**, 379–391
10. Niwa, H., Miyazaki, J., and Smith, A. G. (2000) *Nat. Genet.* **24**, 372–376
11. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003) *Cell* **113**, 631–642
12. Takahashi, K., and Yamanaka, S. (2006) *Cell* **126**, 663–676
13. Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2007) *Science* **318**, 1917–1920
14. Surani, M. A., Hayashi, K., and Hajkova, P. (2007) *Cell* **128**, 747–762
15. Jaenisch, R., and Young, R. (2008) *Cell* **132**, 567–582
16. Li, J., Pan, G., Cui, K., Liu, Y., Xu, S., and Pei, D. (2007) *J. Biol. Chem.* **282**, 19481–19492
17. Xi, R., and Xie, T. (2005) *Science* **310**, 1487–1489
18. Egli, D., Birkhoff, G., and Eggan, K. (2008) *Nat. Rev. Mol. Cell Biol.* **9**, 505–516
19. Scholer, H. R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990) *Nature* **344**, 435–439
20. Pan, G. J., Chang, Z. Y., Scholer, H. R., and Pei, D. (2002) *Cell Res.* **12**, 321–329
21. Pan, G., Qin, B., Liu, N., Scholer, H. R., and Pei, D. (2004) *J. Biol. Chem.* **279**, 37013–37020
22. Pan, G. J., and Pei, D. Q. (2003) *Cell Res.* **13**, 499–502
23. Pan, G., and Pei, D. (2005) *J. Biol. Chem.* **280**, 1401–1407
24. Pan, G., Li, J., Zhou, Y., Zheng, H., and Pei, D. (2006) *FASEB J.* **20**, 1730–1732
25. Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B., Ng, H. H., and Robson, P. (2005) *J. Biol. Chem.* **280**, 24731–24737
26. Ambrosetti, D. C., Basilico, C., and Dailey, L. (1997) *Mol. Cell. Biol.* **17**, 6321–6329
27. Scaffidi, P., and Bianchi, M. E. (2001) *J. Biol. Chem.* **276**, 47296–47302
28. Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003) *Genes Dev.* **17**, 126–140
29. Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., Ko, M. S., and Niwa, H. (2007) *Nat. Cell Biol.* **9**, 625–635
30. Hanna, L. A., Foreman, R. K., Tarasenko, I. A., Kessler, D. S., and Labosky, P. A. (2002) *Genes Dev.* **16**, 2650–2661
31. Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. (2005) *Cell* **122**, 947–956
32. Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W., and Orkin, S. H. (2006) *Nature* **444**, 364–368
33. Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K. Y., Sung, K. W., Lee, C. W., Zhao, X. D., Chiu, K. P., Lipovich, L., Kuznetsov, V. A., Robson, P., Stanton, L. W., Wei, C. L., Ruan, Y., Lim, B., and Ng, H. H. (2006) *Nat. Genet.* **38**, 431–440
34. Pan, G., and Thomson, J. A. (2007) *Cell Res.* **17**, 42–49
35. Zhou, Q., Chipperfield, H., Melton, D. A., and Wong, W. H. (2007) *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16438–16443
36. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J., and Campbell, K. H. (1997) *Nature* **385**, 810–813
37. Qin, D., Li, W., Zhang, J., and Pei, D. (2007) *Cell Res.* **17**, 959–962
38. Dimos, J. T., Rodolfa, K. T., Niakan, K. K., Weisenthal, L. M., Mitsumoto, H., Chung, W., Croft, G. F., Saphier, G., Leibel, R., Golland, R., Wichterle, H., Henderson, C. E., and Eggan, K. (2008) *Science* **321**, 1218–1221
39. Pei, D. (2008) *Cell Res.* **18**, 221–223
40. Shi, Y., Do, J. T., Desponts, C., Hahm, H. S., Scholer, H. R., and Ding, S. (2008) *Cell Stem Cell* **2**, 525–528
41. Wang, J., and Orkin, S. H. (2008) *Cells Tissues Organs* **188**, 23–30
42. Okita, K., Ichisaka, T., and Yamanaka, S. (2007) *Nature* **448**, 313–317