Progress in understanding reprogramming to the induced pluripotent state

Kathrin Plath[*‡§||] *and William E. Lowry*[‡§||¶]

Abstract | Induction of pluripotency by transcription factors has become a commonplace method to produce pluripotent stem cells. Great strides have been made in our understanding of the mechanism by which this occurs — particularly in terms of transcriptional and chromatin-based events — yet only a small part of the complete picture has been revealed. Understanding the mechanism of reprogramming to pluripotency will have important implications for improving the efficiency and quality of reprogramming and advancing therapeutic application of induced pluripotent stem cells. It will also help to reveal the machinery that stabilizes cell identity and to instruct the design of directed differentiation or lineage switching strategies. To inform the next phase in understanding reprogramming, we review the latest findings, highlight ongoing debates and outline future challenges.

Induced pluripotent stem cells
Pluripotent cells that can be generated from many different types of somatic cells by expression of only a few pluripotency-related transcription factors, and that have properties of embryonic stem cells. They serve as an ideal platform to produce patient-specific pluripotent cells.

*David Geffen School of Medicine, Department of Biological Chemistry, ‡Jonsson Comprehensive Cancer Center, §Molecular Biology Institute, ||Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, ¶Department of Molecular Cell and Developmental Biology, University of California Los Angeles, California, USA.
e-mails: kplath@mednet.ucla.edu; blowry@ucla.edu
doi:10.1038/nrg2955

In 2006, Takahashi and Yamanaka published their milestone strategy to reprogramme somatic mammalian cells to induced pluripotent stem cells (iPSCs) by overexpression of only four transcription factors: OCT4 (also known as POU5F1), SOX2, Krüppel-like factor 4 (KLF4), and MYC (also known as c-MYC)[1]. The huge therapeutic potential of iPSCs makes understanding the mechanisms underlying the reprogramming process of paramount importance. Although reprogramming with transcription factors has become routine, we are only beginning to define how these factors induce pluripotency, but during the last few years several points have become clear. Many studies have demonstrated that mouse and human iPSCs are highly similar to their respective embryo-derived embryonic stem cell (ESC) counterparts morphologically, functionally and molecularly at the level of transcription and genome-wide distribution of chromatin modifications[2–16]. Therefore, the key mechanistic question of transcription factor-induced reprogramming to the iPSC state is how the somatic programme is erased and the ESC-like transcriptional network established to confer pluripotent capabilities.

Despite the development of numerous methods to introduce the reprogramming factors into somatic cells, only a small percentage of cells expressing the factors make the complete trip to the pluripotent state. It is now believed that the inefficiency of reprogramming is attributable to epigenetic hurdles that are only overcome infrequently[17,18]. Steps are being defined that precede the activation of the endogenous pluripotency network, and each step appears to be conquered by fewer and fewer cells[18–24]. Recent data also demonstrate that repressive chromatin states comprise a major mechanistic barrier to the induction of pluripotency[6,23,25–29]. Various extrinsic signals can modulate reprogramming and even affect the activity of the reprogramming factors, demonstrating the close relationship of extrinsic and intrinsic pathways in regulating reprogramming and cell identity[21,22,24,30–35]. In addition, testing is needed to ascertain whether a molecule that appears to accelerate reprogramming acts by changing the cell cycle or by lowering reprogramming barriers[17,36]. Also, the notion that iPSCs carry an epigenetic memory of the starting cell may shed light on processes that are difficult to reset during reprogramming[37,38].

In this Review, we highlight recent important work on understanding transcription factor-induced reprogramming to iPSCs. Although iPSCs can now be derived by various combinations of transcription factors and small molecules (for a review, see REF. 39), we concentrate mostly on lessons learned from experiments performed with the original reprogramming factor cocktail (OCT4, SOX2, KLF4 and MYC). We discuss steps leading to faithful reprogramming, the function of the reprogramming

Pluripotency
The ability of a cell to give rise to all cells of the embryo.

Embryonic stem cells
Pluripotent cells derived from epiblast cells of the blastocyst upon explantation in culture.

Reprogramming factors
Four transcription factors (OCT4, SOX2, KLF4 and MYC), first described by Shinya Yamanaka, that when forcibly expressed in somatic cells are capable of driving these cells into the induced pluripotent stem cell state.

Epigenetic memory
The idea that at least a portion of somatic post-translational modifications on histones and DNA is retained despite reprogramming to a more immature state. This memory is thought to make cells adopt facets of physiology that are representative of a previous cellular state.

factors in relation to the transcriptional network of pluripotent cells, and what is known about chromatin regulation in this process. We also consider whether analyses of molecular and functional similarities and differences between ESCs and iPSCs can illuminate mechanisms of reprogramming. Finally, we speculate on the best strategies to generate a complete account of the reprogramming process.

Roadblocks to reprogramming

During reprogramming to pluripotency using OCT4, SOX2, KLF4 and MYC, it takes at least 1 to 2 weeks for the first reprogrammed cells to emerge in the culture dish. Importantly, only a few of the somatic cells that initially express the reprogramming factors eventually convert to the pluripotent state within this timeframe[17]. In fact, an experiment that plated single pre-B cells into individual culture wells and quantified reprogramming in hundreds of these clonal cell populations demonstrated successful induced pluripotency in only 3–5% of the wells within 2 weeks. Even in 'successful' wells, only a small subset of daughter cells had undergone reprogramming[17].

Even when most of the cells are induced to express all the reprogramming factors — for example, by applying polycistronic cassettes that encode all four factors in a single construct, or by using secondary reprogramming systems to control for transgene expression — the number of faithfully reprogrammed colonies remains low relative to the number of dividing cells in the culture dish. This is contrary to the idea that the low efficiency of the process is attributable to heterogeneous transgene expression across the starting cell population[40–50]. The hypothesis that only non-lineage-committed cells or adult stem cells are amenable for reprogramming has also been discarded as an explanation for the low efficiency, based on the ability of terminally differentiated cells, such as pancreatic islets or terminal blood lineages, to give rise to iPSCs[51–57]. Additional evidence against this model is provided by the aforementioned clonal reprogramming experiment which demonstrated that, given time, virtually all cells in a donor pre-B-cell population have the potential to give rise to a reprogramming event: after 18 weeks in culture, more than 90% of the wells contained at least a few cells that are positive for a pluripotency marker[17]. However, debate continues as to whether the degree of differentiation of cells in a lineage influences the efficiency and kinetics of the process[17,52]. Initially, it was also suspected that insertional mutagenesis upon viral insertion of the reprogramming factor transgene was required for reprogramming, but non-integrative reprogramming studies[58–61] (for a review, see REF. 62), mapping of viral insertion sites[57,63,64] and the development of the 'reprogrammable' mouse model with a defined integration site for a single inducible, polycistronic reprogramming factor cassette[43,50] argue against this idea.

Together, these findings have led to a model which proposes that expression of the reprogramming factors *per se* is not sufficient to permit the transition to pluripotency, and that additional events are required to overcome major epigenetic barriers that prevent reprogramming[17,18].

From the differentiated to the pluripotent state

Owing to the low efficiency of reprogramming and the timescale involved, determining the events that occur between the initial expression of the reprogramming factors in somatic cells and the establishment of the pluripotent programme has been challenging. To probe for mechanistic insights, mouse embryonic fibroblasts are commonly used as starting cells for reprogramming experiments, and partially reprogrammed cells (pre-iPSCs) have been particularly valuable for analysing some stages of the process (BOX 1). In addition, the development of improved technologies, particularly of various tetracycline-inducible expression systems for the reprogramming factors (BOX 2) and, most recently, of live imaging analysis, have had a huge impact on mechanistic studies[6,17–20,22,37,43–50]. This demonstrates how technology development and mechanistic insight are intimately connected in this field.

Steps in reprogramming. Considering only those reprogramming events that occur within the first couple of weeks, many reports now suggest that successful reprogramming of fibroblasts requires stepwise transition through key intermediate steps, and at each step fewer and fewer cells advance owing to secondary events that are still being discovered[6,18–24]. The steps of fibroblast reprogramming are described in more detail in the subsections below, and the changes observed at each stage are shown in FIG. 1. Briefly, induction of proliferation and downregulation of fibroblast-specific transcription are

Box 1 | Tools for studying reprogramming mechanisms

Mouse embryonic fibroblasts are most commonly used to probe for mechanistic insights because they can be easily generated from reporter mice that carry a knock-in of the GFP coding sequence in the endogenous *Nanog* or *Oct4* locus or a transgenic GFP reporter driven by *Oct4* regulatory regions. Such reporter systems enable pluripotent cells to be identified easily. Fibroblasts also have a relatively high viral infection capacity for the delivery of the four reprogramming factors and a reasonable reprogramming efficiency.

Understanding the function of the reprogramming factors is particularly difficult at the intermediate and late phases of reprogramming as fewer and fewer cells are routed towards the pluripotent state. Thus, the isolation of intermediate stages has been informative. Partially reprogrammed induced pluripotent stem cells (pre-iPSCs) have been established as a tool to study the late step of reprogramming — that is, the upregulation of the pluripotency network[6,23,24,31]. These pre-iPSCs can be clonally expanded and form embryonic stem cell-like colonies. They have largely acquired the proliferative capacity and biosynthetic properties of pluripotent cells and silenced many somatic genes, but they fail to express many endogenous pluripotency genes, such as *Oct4* and *Nanog*[6,23,24,31]. Pre-iPSC clones obtained from different starting cell types share similar transcription profiles, which suggests that reprogramming from various starting cell types converges to the pluripotent state and stalls at a similar barrier[6]. Although it is not absolutely clear that pre-iPSCs represent an intermediate that occurs transiently during the reprogramming process, they are not simply an aborted reprogramming artefact, as they can be converted into iPSCs with small-molecule treatments that also improve the efficiency and kinetics of the reprogramming process[6,23,24,31,34]. Currently, pre-iPSCs are a useful platform for the identification of molecular mechanisms guiding the final steps of reprogramming and, because of their defined nature, they allow population-based genomics approaches.

Box 2 | **Inducible expression systems and their use in reprogramming**

Listed below are tetracycline (tet)-inducible systems for reprogramming factor expression and a summary of their advantages and applications.

Individual, tet-inducible reprogramming factors, encoded on separate lenti- or retroviruses[19,20,23,46,47,49].

- Used to test the temporal requirement of each reprogramming factor.
- Demonstrated that the reprogrammed pluripotent state is independent of transgene expression but intermediate stages are not.
- Demonstrated that reprogramming factors must be repressed in induced pluripotent stem cells (iPSCs) to allow efficient differentiation.
- Allow titration of reprogramming factor levels.
- Withdrawal of doxycycline is a selection strategy against incompletely reprogrammed intermediates.

tet-inducible polycistronic cassettes encoding all four reprogramming factors in one transcript[44,45,48,112].

- Reprogramming factor stochiometries are fixed, thereby overcoming cellular heterogeneity of randomly infected fibroblasts regarding which reprogramming factors are expressed.
- Allow for more efficient deletion of reprogramming cassette using the *loxP*/Cre or transposon systems.

'Secondary' tet-inducible reprogramming system and tet-inducible 'reprogrammable mouse'. In the 'secondary' system, in a first reprogramming round, iPSCs are obtained from somatic cells using tet-inducible lentiviral vectors or transposons and then, in the absence of doxycycline, differentiated *in vitro* (human cells) or *in vivo* by chimaera formation (mouse cells). Differentiated cells are then induced with doxycycline to trigger a second reprogramming round yielding 'secondary' iPSCs[18,22,37,44,46,47,49,52,56,66]. Alternatively, mouse models have been generated that carry a single inducible polycistronic transgene in a defined genomic position[43,45,77].

- Homogeneous cell populations can be derived, harbouring the same integration event(s) in each cell, thereby overcoming genetic heterogeneity of a randomly infected primary reprogramming culture.
- No need for direct viral transduction.
- Reprogramming efficiency is higher compared to the direct infection systems.
- Permits reprogramming of cells that are typically difficult to infect with retro- or lentiviruses.
- Better scalability, allowing chemical and genetic high-throughput screening approaches.
- Largely homogeneous transgene expression levels across the cell population.
- Reprogrammable mouse permits the comparison of genetically matched embryonic stem cells and iPSCs from different tissues.
- Reprogrammable mouse allows easy transfer of the reprogramming factor cassette onto different genetic backgrounds.
- Highly reproducible kinetics and efficiencies facilitate mechanistic dissection of the reprogramming process.
- Efficient reprogramming systems can be generated that carry only a subset of reprogramming factors to identify small molecules that replace the missing factor.

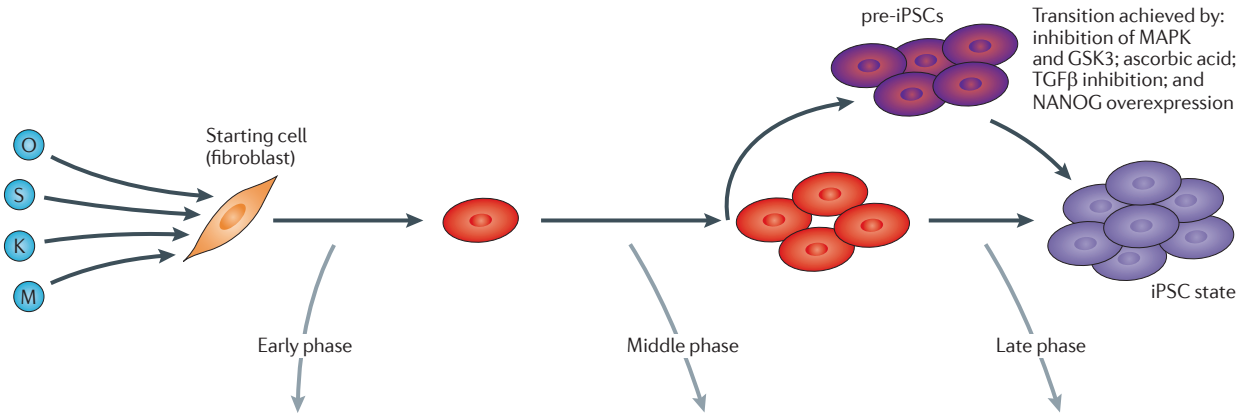
Faithful reprogramming
Complete reprogramming to induced pluripotent stem cells, defined by endogenous expression of pluripotency-related genes (such as *Nanog* and *Oct4*). Presence of markers such as alkaline phosphatase or the surface antigen stage-specific embryonic antigen 1 (SSEA1), often used to assess reprogramming, mark partially as well as faithfully reprogrammed cells.

followed by acquisition of epithelial characteristics and activation of some ESC markers. Later, pluripotency-related genes are activated. Markers used to detect these steps include alkaline phosphatase and stage-specific embryonic antigen 1 (SSEA1, also known as FUT4) at an intermediate stage[19,20] and, for the final stages, *Nanog* in the mouse system[6,7,9] or the surface marker TRA-1-60 in human cells[65]. The steps en route to the iPSC state require continuous expression of the reprogramming factors, but maintenance of iPSCs is independent of their overexpression; this independence indicates a stable conversion of cell fate[19,20].

Silencing of the somatic programme and change in the cell division rate. A high-resolution time-lapse imaging approach that enabled retroactive tracking of faithful reprogramming events demonstrated that an increase in proliferation rate and a concomitant decrease in cell size are the first noticeable changes in the reprogramming of mouse fibroblasts, and occur as early as 24 hours after induction of the reprogramming factors[18]. These morphological and proliferative changes are accompanied molecularly by the induction of proliferation genes and downregulation of the somatic expression programme[6,22,66]. Interestingly, a cell sorting experiment for THY1 (a marker present on the surface of fibroblasts) suggests that expression changes in this early phase of reprogramming occur in the majority of cells[19]. Thus, whereas the transcriptional response to the reprogramming factors may be population-wide, only a few cells undergo the rapid shift in proliferation that coincides with the reduction of cell size in this early phase of reprogramming, which can be tracked as the first morphological event in all successful reprogramming events. To what extent the expression changes seen at the population level reflect changes in these fast-dividing cells remains unclear at this point.

Most cells expressing the reprogramming factors fail to successfully induce the first morphological change of proper reprogramming events, remain fibroblast-like and often undergo apoptosis, senescence and cell-cycle arrest. Notably, apoptosis, senescence and cell-cycle arrest are thought to be barriers to reprogramming, as methods that suppress these responses are associated with higher reprogramming efficiency[67–72]. Specifically, the silencing of central regulators of these responses, such as p53 and p21 or p16/INK4A (also known as CDKN2A), is observed upon reprogramming and their experimental depletion enhances the efficiency and kinetics of iPSC generation[17,67–72]. It is important to note that the role of p53 in reprogramming has been debated and that one study linked the positive effect of p53 depletion in pre-B-cell reprogramming solely to an increased proliferation rate[17,73]. This would be consistent with data suggesting that, even in wild-type cells, promoting the cell cycle improves reprogramming efficiency[17,36]. It is likely that active promotion of cell proliferation — that is, more transitions through S phase — enhances efficient resetting of the transcriptional and chromatin landscapes. Intriguingly, monitoring the effect of p53 knockdown at the single-cell level suggested that although proliferation is induced in more cells than in the control sample, most of the p53-depleted cells derail from the reprogramming path later, yielding a lower overall reprogramming efficiency when normalized to the number of cells that initially responded[18]. This study highlighted how single-cell analysis can provide novel insights that cannot be obtained from typical population studies.

It is possible that the extinction of the somatic programme is a lower barrier to reprogramming than the acquisition of the ESC programme, and it is conceivable that the induction of the pluripotent state may only be possible after drivers of the somatic state are efficiently silenced. Recent studies have supported this idea by



		Early phase	Middle phase	Late phase
Phenotypic changes		• Rapid induction of proliferation • Acquisition of an ESC-like cell cycle • Smaller, round cell appearance	• Occurrence of tightly packed clusters of rounded cells concomitant with the mesenchymal-to-epithelial transition • Epithelial cell character • ESC-like biosynthetic properties	• ESC-like colonies are obvious
Key transcriptional changes	Down	• Efficient downregulation of the somatic expression programme, including mesenchymal genes such as *Snai1* and *Snai2* and the THY1 surface marker		
	Up	• Upregulation of proliferation genes such as *Ccnd1*, *Ccnd2* and DNA replication genes	• Many embryonic genes upregulated reflecting the proliferating, biosynthetic, epithelial properties of ESCs (including E-cadherin and the surface marker SSEA1)	• Upregulation of key transcriptional and developmental regulators of the pluripotent state, including endogenous *Oct4* and genes such as *Zfp42* and *Esrrb*, possibly following an ordered sequence
Chromatin events		• Expression changes are restricted to pre-existing H3K4me3-positive (euchromatic) regions • Gain of H3K4me2 at promoter and enhancer regions of many pluripotency genes that are hypomethylated and CpG-rich, without transcriptional activation or loss of the repressive mark H3K27me3 in surrounding regions		• Xi reactivation • Loss of repressive chromatin character at many pluripotency genes
Reprogramming factor dependence		Dependent		Independent
Defined requirements			• TGFβ inhibition and BMP signalling to promote mesenchymal-to-epithelial transition	• NANOG expression

Figure 1 | **The generation of induced pluripotent stem cells is a multistep process.** Known events occurring in early, middle and late phases of reprogramming mouse embryonic fibroblasts to induced pluripotent stem cells (iPSCs) are shown in the table beneath the schematic of reprogramming. Reprogramming is driven in this case by OCT4 (also known as POU5F1), SOX2, Krüppel-like factor 4 (KLF4) and MYC (also known as c-MYC), which are represented as O, S, K and M, respectively. Events in the early and middle phases may be less clearly separated from each other compared to those that occur later. Even though the initial response to reprogramming factor expression — for example, downregulation of somatic expression programmes and changes in histone H3 lysine 4 dimethylation (H3K4me2) — may occur population-wide, none of the early and middle steps alone is sufficient for the induction of pluripotency and only a subset of cells makes it from one step to the next, accounting for the low overall efficiency of the process. Although events that eventually lead to the reprogrammed state are initiated early, successful reprogramming requires the expression of the four transcription factors until the iPSC state is established, otherwise cells revert back to a differentiated state. The partially reprogrammed state (pre-iPSC) is often identified in reprogramming cultures as embryonic stem cell (ESC)-like colonies that do not express *Nanog* and other pluripotency-related genes that are induced only during the last step of reprogramming. Pre-iPSCs can be converted to iPSCs with treatments that enhance the late phase of reprogramming. The close cooperation between reprogramming factors and the extracellular milieu during reprogramming is indicated by the fact that certain signalling pathways affect the reprogramming process at the indicated places. Further details are provided in the main text; see also REFS 6,18–24,31,34,35,66. BMP, bone morphogenetic protein; *Ccnd*, cyclin D; E-cadherin, epithelial cadherin (also known as cadherin 1); *Esrrb*, oestrogen-related receptor-β; GSK3, glycogen synthase kinase 3; MAPK, mitogen-activated protein kinase; *Snai*, Snail homologue; SSEA1, stage-specific embryonic antigen 1 (also known as FUT4); TGFβ, transforming growth factor-β; Xi, inactive X chromosome; *Zfp42*, zinc finger protein 42 (also known as *Rex1*).

demonstrating that the expression of lineage-specific transcription factors blocks reprogramming of the somatic genome in a dominant fashion[6,74].

Gaining epithelial cell character. ESCs and iPSCs have characteristics of epithelial cells, with tight intercellular contacts and surface expression of the key gene epithelial cadherin (E-cadherin; also known as cadherin 1). Thus, mesenchymal cells such as fibroblasts need to gain an epithelial character during reprogramming. After suppression of the somatic transcriptional programme, small, fast-cycling cells cluster tightly and undergo coordinated changes in cell–cell and cell–matrix interactions, which correspond with a loss of mesenchymal features and acquisition of epithelial cell characteristics[18,21]. This supports the idea that fibroblasts undergo a mesenchymal-to-epithelial transition (MET) during reprogramming, thus reversing the epithelial-to-mesenchymal transition (EMT) that occurred during the differentiation of fibroblasts *in vivo*[21,22]. Signalling pathways that are known to promote or suppress MET affect the efficiency of the reprogramming process. For example, inhibition of signalling by transforming growth factor-β (TGFβ) improves reprogramming, because TGFβ activity prevents MET by inhibiting the upregulation of epithelial markers and the downregulation of the mesenchymal transcriptional repressor zinc finger protein SNAI1 (REF. 21). By contrast, bone morphogenetic protein (BMP) signalling enhances reprogramming through the upregulation of pro-MET microRNAs[22]. Furthermore, E-cadherin is crucial for ESC pluripotency[75] and its knockdown interferes with reprogramming[21]. Of course, not all cell types would have to go through MET during reprogramming. For instance, keratinocytes and hepatocytes are epithelial; they can be reprogrammed with higher efficiencies than fibroblasts[57,76], so it can be speculated that this is because they do not have to overcome the MET barrier.

Establishing pluripotency. After epithelial cell character has been established and as larger colonies are formed, other ESC markers such as SSEA1 are induced, probably only in a subset of E-cadherin-positive cells[19–21]. At this point, many embryonic genes seem to be upregulated already, particularly those that are involved in housekeeping functions[6,22,23]. Experiments in which SSEA1-positive and -negative cell populations were isolated from reprogramming cultures demonstrated that only SSEA1-positive cells can give rise to faithfully reprogrammed cells and activate the expression of the pluripotency network — that is, the expression of transcriptional or developmental regulators that are highly expressed in ESCs, including endogenously encoded *Oct4*, *Sox2* and *Nanog*, and many other pluripotency-related genes[19,20]. The upregulation of this core pluripotency network is considered to be the final step of reprogramming and, as with the other steps, only a few SSEA1-positive cells make this final transition[19].

Cooperation of transcription factors

Roles during MET. Understanding the contribution of each reprogramming factor to the different steps of reprogramming is ultimately required to reveal the molecular mechanisms underlying the induction of pluripotency. It is now thought that each reprogramming factor has a distinct role[21,23,77–79]. This concept is exemplified by their respective contributions to the MET during fibroblast reprogramming[21]. It was shown that OCT4 and SOX2 suppress the pro-mesenchymal regulator *Snai1*, whereas KLF4 induces the epithelial programme by directly binding to and activating epithelial genes, including E-cadherin[21]. At the same time, MYC reduces TGFβ signalling by repressing *Tgfb1* and *Tgfbr1*. The fact that the reprogramming factors collaborate in the MET by suppressing different pro-EMT molecules and promoting various pro-MET mechanisms may explain why the four canonical Yamanaka factors constitute such an efficient reprogramming cocktail.

Reprogramming factors and the induction of the pluripotency network. To address how pluripotency-related genes are upregulated during the final phase of reprogramming, we have mapped the binding sites of OCT4, SOX2, KLF4 and MYC in mouse iPSCs and pre-iPSCs[23]. We found that in iPSCs the target genes of these four transcription factors are similar to those previously defined in ESCs: SOX2 and OCT4 co-occupy promoters of highly expressed genes, including their own promoters, and KLF4 shares roughly half of its targets with these two transcription factors[23,80–83]. Notably, based on limited target overlap, it was proposed that the function of MYC differs from that of OCT4, SOX2 and KLF4 in ESCs and iPSCs[23,80,81]. Consistent with this notion, MYC targets are predominantly involved in the regulation of cellular proliferation, metabolism and biosynthetic pathways, whereas OCT4, KLF4 and SOX2 targets in pluripotent cells are skewed towards the transcriptional and developmental regulators that form the pluripotency network[23,80,81,84] (FIG. 2).

These results imply that MYC, unlike OCT4, SOX2 and KLF4, is not involved in the upregulation of the pluripotency network during the final step of reprogramming. A recent report also suggests that MYC promotes the release of promoter-proximal pausing of RNA polymerase II (Pol II) and thereby enhances transcriptional elongation, rather than mediating the initial recruitment of Pol II to promoters[85]. Therefore, MYC might enhance but not be absolutely required for the transcription of its target genes. Together, these findings could explain why MYC is dispensable for reprogramming but still able to enhance the efficiency and kinetics of the process[78,79]. MYC overexpression might lay the foundation for the function of the other factors in activating the pluripotency network (FIG. 2).

Furthermore, MYC already binds many of its iPSC target genes in pre-iPSCs, indicating that the MYC transcriptional network is already largely engaged at an intermediate step of reprogramming[23]. By contrast, many pluripotency-related genes that are occupied by OCT4, SOX2 and KLF4 in iPSCs completely lack binding by these three reprogramming factors in pre-iPSCs[23]. Consequently, the expression of genes belonging to the MYC target network is comparable between pre-iPSCs and iPSCs, but genes of the OCT4, SOX2 and KLF4 network are not activated in pre-iPSCs[84]. Thus, genes belonging to the

Polycistronic cassette
DNA-containing sequence that codes for multiple genes expressed from a single promoter. These coding regions are sometimes separated by sequences that are cleaved during translation to produce individual protein products.

Secondary reprogramming system
A system in which induced pluripotent stem cells are first generated from somatic cells with virally encoded inducible reprogramming factors, then differentiated again to obtain somatic cell populations that can express these factors in all cells and be used for secondary reprogramming experiments upon re-induction of the programming factors.

Pre-iPSCs
Partially reprogrammed cells that arise in reprogramming cultures. They have efficiently silenced somatic genes but have not induced the endogenous pluripotency programme.

NANOG
A transcription factor that is highly expressed in pluripotent cells and is essential for the establishment of embryonic stem cells but not for their maintenance. Although not belonging to the original Yamanaka set of reprogramming factors, NANOG overexpression enhances mouse cell reprogramming at the late step and has been used with OCT4 and SOX2 to reprogramme human cells.

Mesenchymal-to-epithelial transition
Mesenchymal and epithelial cells are distinguished by, among other traits, their gene expression, morphology and cell adhesion properties. Transitions between these two states are thought to have key roles in development, cancer and, more recently, in reprogramming.

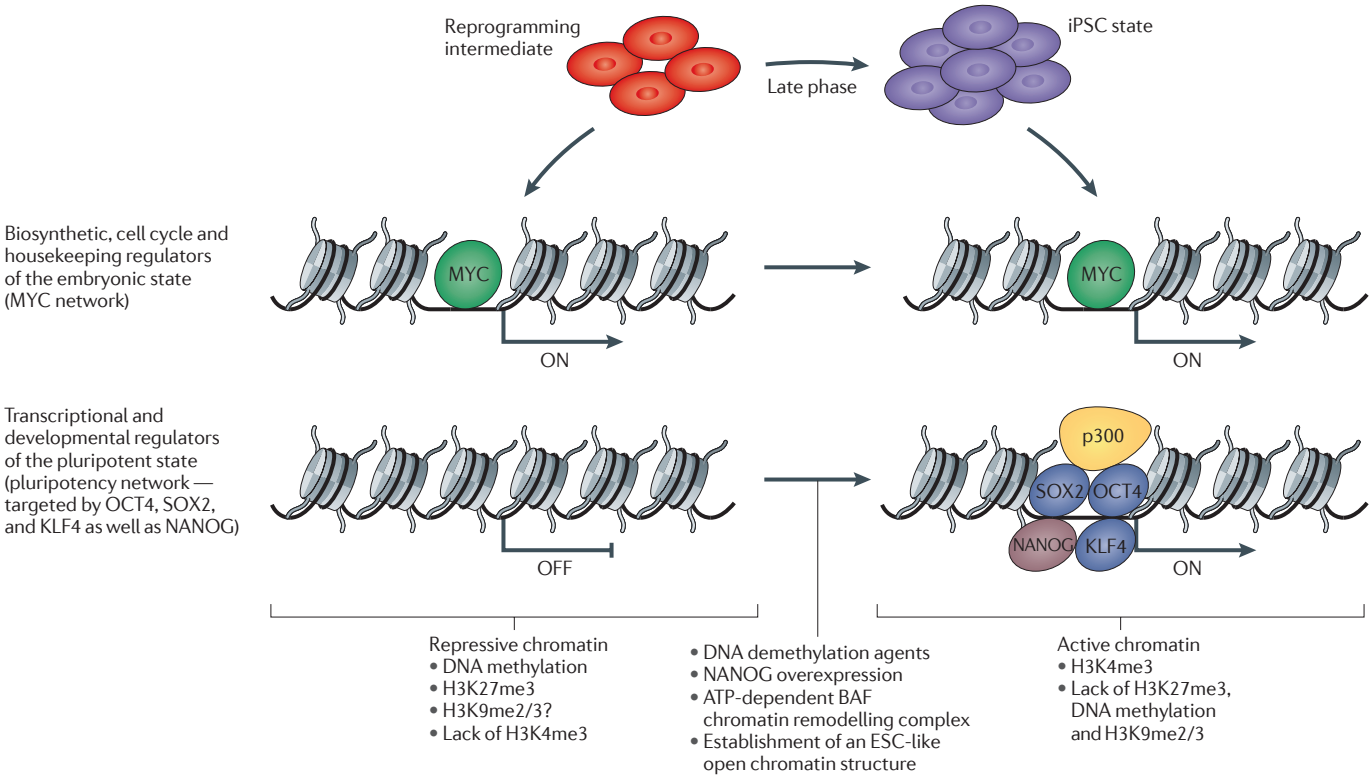


Figure 2 | **Roles of the reprogramming factors and their interaction with chromatin during the final step of reprogramming.** Scheme illustrating the different functions of the reprogramming factors in the late phase of reprogramming. OCT4 (also known as POU5F1), SOX2 and Krüppel-like factor 4 (KLF4) are implicated in mediating the upregulation of the pluripotency network (lower panels) and only bind many of these genes during the final step of reprogramming associated with their transcriptional activation[23,84] (compare lower left to lower right). By contrast, many MYC (also known as c-MYC) targets are bound and activated at an intermediate step before activation of the pluripotency network[23,84] (upper panels, left and right). It seems that the inability of OCT4, SOX2 and KLF4 to bind and activate pluripotency-related genes is, at least in some cases, associated with repressive chromatin signatures (such as histone H3 lysine 27 trimethylation (H3K27me3), other repressive histone marks or DNA methylation) found in the promoters of these genes in the partially reprogrammed state, which is reset to active chromatin in induced pluripotent stem cells (iPSCs)[23,84] (compare lower left to lower right). The loss of these repressive marks seems to be required for efficient reprogramming (BOX 3). Overexpression of the pluripotency transcription factor NANOG synergistically acts with DNA demethylating agents to enhance the final transition to the iPSC state[35,88]. NANOG co-binds many of the regulatory regions with OCT4, SOX2 and KLF4 and may promote their binding, as well as the recruitment of co-activators such as the histone acetyltransferase p300 (REFS 80–82,87). In addition, embryonic stem cell (ESC)-specific chromatin remodelling complexes have been implicated in the activation of pluripotency-related genes. For example, overexpression of components of the BAF chromatin-remodelling complex enhances reprogramming and may facilitate binding of OCT4 to pluripotency target genes[91]. Similarly, CHD1, another chromatin remodelling enzyme that is thought to be important for the maintenance of an open chromatin state in mouse ESCs, is required for reprogramming and may act during the late phase of the process[92].

pluripotency network are not accessible to the reprogramming factors at this intermediate stage of reprogramming, and we have proposed that the engagement of OCT4, SOX2 and KLF4 at pluripotency genes and their subsequent transcriptional upregulation represents a major hurdle to the completion of reprogramming[23] (FIG. 2).

At least two models can be envisioned to explain the lack of binding to and upregulation of pluripotency genes in pre-iPSCs. The first suggests that additional transcription factors are required to cooperatively bind with OCT4, KLF4 and SOX2 and recruit co-activators, and that these additional factors are not yet available at this intermediate stage[23]. This model is supported by studies of the NANOG transcription factor (FIG. 2). NANOG has extensive protein–protein interactions with SOX2, OCT4 and other pluripotency transcription factors[86] and co-binds many of their targets in ESCs[80–82,87]. It is essential for the generation of iPSCs but is only upregulated during the final step of reprogramming[6,23,35,88]. However, low levels of *Nanog* transcripts can be detected early in the transition of pre-iPSCs to iPSCs, which might be sufficient to promote OCT4, SOX2 and KLF4 function[88]. Accordingly, NANOG overexpression in pre-iPSCs and during reprogramming enhances the induction of pluripotency by lowering cell-intrinsic barriers[17,35,88]. A second model suggests that repressive chromatin at pluripotency gene promoters and enhancers (which forms to silence these genes during differentiation[89]) interferes with the binding of reprogramming factors (FIG. 2).

Enhancers
DNA regions that positively control gene expression and that can be located upstream, downstream or even within the genes that they regulate. They are often bound by cell-type-specific transcription factors (such as OCT4 and NANOG in embryonic stem cells) and have a specific chromatin signature.

Box 3 | **Repressive chromatin inhibits reprogramming**

Small molecules have been useful in showing that repressive chromatin states contribute to the stability of differentiated cell identity. However, how they affect reprogramming remains largely unclear as they are likely to lead to global changes in chromatin structure as well as changes at specific genes, and they might influence several steps of reprogramming.

Histone deacetylases (HDACs) catalyse the removal of acetyl groups from lysine residues of histones, which is classically associated with chromatin condensation and transcriptional repression[113]. Four HDAC inhibitors — suberoylanilide hydroxamic acid (SAHA), trichostatin A (TSA), butyrate and valporic acid (VPA) — greatly improve the reprogramming efficiency of mouse and/or human fibroblasts[25,27–29]. Additionally, VPA enabled the efficient induction of mouse induced pluripotent stem cells (iPSCs) in the absence of ectopic MYC (also known as c-MYC) and the reprogramming of human fibroblasts without ectopic expression of Krüppel-like factor 4 (KLF4) and MYC, and made possible the generation of iPSCs using cell-penetrable recombinant proteins[114]. By contrast, butyrate requires ectopically expressed MYC to exert its positive effect and functions in the early phase of mouse reprogramming[29] but late in the human reprogramming, in which it can efficiently substitute for ectopically expressed KLF4 or MYC[25]. Although these studies come to different conclusions regarding the temporal requirement of HDAC inhibition and reprogramming factor replacement, they all agree that treatment of reprogramming cultures with VPA or butyrate induces a transcriptional change towards the embryonic stem cell state. This is consistent with the idea that inhibiting HDACs shifts the balance towards histone acetylation and activation of transcription[25,27,29].

Interestingly, the reprogramming factors have been shown to interact with various histone acetyltransferases, which could partially explain why their expression can be replaced by HDAC inhibition. However, in addition to regulating the acetylation state of histones, HDACs can deacetylate and regulate the activity of a number of other proteins, including the transcription factor p53 (REF. 115), which has been implicated as a barrier to reprogramming. Given that all the HDAC inhibitors listed above block the activity of several HDAC family members, the particular HDAC (or HDACs) implicated in reprogramming and its substrate (or substrates) remain to be determined.

Similar to HDAC inhibition, an inhibitor of DNA methylation (5-azacytidine), or knockdown of DNA methyltransferase 1 (*Dnmt1*; which encodes an enzyme that is responsible for maintaining DNA methylation through DNA replication), enhances reprogramming and promotes the conversion of partially reprogrammed iPSCs (pre-iPSCs) to the iPSC state[6,27]. In addition, a small molecule (BIX-01294) that can inactivate the repressive histone H3 lysine 9 methyltransferases G9a and GLP[26,116,117], and parnate, which is an inhibitor of lysine-specific demethylase 1 (a histone H3 lysine 4 demethylase)[90], have similar effects and can compensate for the loss of various reprogramming factors.

If repressive chromatin marks interfere with reprogramming, how are they removed during successful reprogramming events? A passive mechanism could require DNA replication and lead to the dilution of repressive marks by simply preventing the re-establishment of the parental chromatin pattern on newly incorporated histones and DNA. In support of this notion, reprogramming is accelerated by an increased cell division rate and inhibited by cell cycle arrest[17,36]. Alternatively, DNA replication could facilitate the resetting of chromatin states, potentially by allowing the reprogramming factors to engage their target sites more effectively. However, active mechanisms might be more likely given that demethylating enzymes have been identified for almost every 'repressive' methylated lysine within histones and are now also being uncovered for methylated DNA.

Chromatin states and reprogramming

In agreement with the second model, it is currently believed that repressive chromatin comprises a major mechanistic barrier to transcription factor-induced reprogramming. The evidence comes mainly from studies in which the manipulation of repressive chromatin states (for example, inhibiting chromatin-modifying enzymes) enhances reprogramming[6,25–29,90] (BOX 3).

Repressive chromatin and activation of pluripotency genes. The regulatory regions of some pluripotency-related genes, such as *Oct4*, *Nanog*, undifferentiated embryonic cell transcription factor 1 (*Utf1*), developmental pluripotency associated 5 (*Dppa5*), zinc finger protein 42 (*Zfp42*; also known as *Rex1*) and *Dppa3*, are hypermethylated at the DNA level in somatic cells and pre-iPSCs, and lack the activating mark histone H3 lysine 4 trimethylation (H3K4me3)[1,6]. Many pluripotency genes are enriched for repressive H3K27 and/or H3K9 methylation in somatic cells[4,6,23,89]. DNA demethylation and the loss of repressive histone methylation marks at the pluripotency genes probably occur at the end of the reprogramming process, concomitant with the binding of the reprogramming factors OCT4, SOX2 and KLF4 and the transcriptional upregulation of these genes[6,23] (FIG. 2). These findings support the idea that repressive chromatin at promoters and enhancers of pluripotency-related genes may initially block engagement of the reprogramming factors.

Intriguingly, a recent report demonstrated that *Nanog* overexpression and inhibition of DNA methylation synergistically enhance the final phase of reprogramming; this indicates that both models proposed above may be involved in the activation of pluripotency-related genes[88]. In any case, the activation of *Nanog* or changes in repressive chromatin structure at pluripotency-related genes (or elsewhere) occur through unknown mechanisms during reprogramming.

A very recent twist is that specific chromatin changes precede the activation of pluripotency-related genes[23,66]. For example, many pluripotency-related genes with CpG-dense promoter and enhancer elements that are hypomethylated in fibroblasts gain a histone mark associated with active transcription — H3K4me2 — in the

early phase of reprogramming, despite the fact that they are upregulated only much later in the reprogramming process[66] (FIG. 1). Gain of H3K4me2 at the CpG island does not alter the repressive chromatin character in surrounding regions, so the silent state of these genes is maintained at the early phase of reprogramming[66]. It remains to be tested whether H3K4me2 at pluripotency gene promoters is required for their subsequent activation, but changes in H3K4me2 apparently occur in the majority of fibroblasts in response to reprogramming factor expression even before the first cell division is initiated[66]. This means that the reprogramming factors are not only inducing major transcriptional changes early on in the reprogramming process but also affecting the chromatin landscape in a global manner without cell division, perhaps by altering the activity or levels of chromatin remodellers or modifiers. In this context it should be noted that chromatin remodelling is crucial for efficient reprogramming[91,92] (FIG. 2).

Somatic chromatin and the initial transcriptional response. Chromatin states influence reprogramming at various stages. For example, they also seem to determine where initial transcriptional responses to reprogramming factors occur in somatic cells. By comparing the transcriptional response with genome-wide maps for histone modifications and DNA methylation in the early phase of reprogramming, it was found that transcriptional changes are limited to those promoters that carry histone H3K4me3, a histone modification that is strongly associated with transcriptional activation[66]. Although the binding targets of the reprogramming factors at this early phase of reprogramming are not yet mapped, it is likely that the factors can only access their target binding sites in pre-existing open chromatin. This could explain why, early in reprogramming, the reprogramming factors are more likely to enhance the transcription of proliferation genes and silence somatic genes than activate pluripotency genes.

Somatic gene silencing and irreversibility of reprogramming. During reprogramming, silencing of somatic genes is associated with a change in chromatin structure at their enhancers and promoters, and in particular with a rapid loss of histone H3K4me2 (REF. 66). Interestingly, many fibroblast-specific enhancers need to gain DNA methylation during reprogramming (they are hypermethylated in ESCs), but seem to do so only towards the end of the process[66]. This finding might explain at least partially why cells on the reprogramming path that have not yet induced pluripotency can return to a fibroblast-like morphology upon withdrawal of the reprogramming factors[19,20], because DNA hypermethylation, among other mechanisms, might be required to 'lock' the silent state of somatic genes upon reprogramming.

X chromosome inactivation and reprogramming

Female mammalian cells silence one of the two X chromosomes in a process called X chromosome inactivation (XCI; for a review, see REF. 93). XCI is initiated early during female embryonic development when pluripotent cells of the blastocyst differentiate. Thus, female mouse ESCs carry two active X chromosomes (XaXa) and initiate XCI upon differentiation by upregulating the large non-coding RNA *Xist* on the future inactive X chromosome (Xi) and inducing a cascade of events that leads to a heritable heterochromatic state (FIG. 3a). Given that the X chromosome represents the largest continuous DNA segment that is subject to epigenetic silencing when pluripotent cells differentiate, a key question has been whether the Xi reactivates during reprogramming.

Xi reactivation in mouse iPSCs. As expected from the XaXa pattern in mouse ESCs, we have shown that the Xi is reactivated in female mouse iPSCs and its heterochromatic state is reset to that of the Xa, enabling random XCI upon induction of differentiation[5] (FIG. 3a). Xi reactivation occurs very late in the reprogramming process, around the time when the pluripotency network is activated[19], emphasizing the tight link between pluripotency and the XaXa state. The pluripotency network might need to be established to allow the down-regulation of *Xist* and reactivation of the Xi, as it has been suggested that pluripotency transcription factors regulate *Xist* expression[94–96] (FIG. 3b). Because XCI can be studied at the single-cell level by fluorescent imaging approaches, it provides an attractive model to study changes in transcription and repressive chromatin in the context of reprogramming.

Xi in human cell reprogramming. For human iPSCs the picture appears to be different: our data show that reactivation of the Xi does not occur when female human cells are reprogrammed[97]. Because of their clonality and lack of Xi reactivation, the cells of a given iPSC line all have the same X chromosome silenced (FIG. 3c). The differences between human and mouse reprogramming are probably due to the cells not being developmentally equivalent rather than reflecting a difference in the way XCI is regulated. Whereas mouse ESCs are in a naive pluripotent state, human ESCs and iPSCs are thought to be in a primed pluripotent state similar to that of mouse epiblast stem cells (EpiSCs)[98]. In support of this notion, female mouse EpiSCs derived from post-implantation embryos are XiXa[99]. The transition of mouse EpiSCs to the naive ESC-like state is accompanied by the reactivation of the Xi[35,99,100]. Similarly, overexpression of KLF4 in human ESCs or iPSCs in combination with a small molecule cocktail that supports growth of mouse ESCs leads to the establishment of XaXa cells (that is, a mouse ESC-like state)[101] (FIG. 3c).

Notably, although most classical (mouse EpiSC-like) female human ESC lines are (like iPSCs) XiXa, XaXa ESCs have been generated and maintained in some circumstances, particularly when derived under hypoxic conditions to more accurately model the *in vivo* environment of the developing embryo[102–104]. However, using standard reprogramming methods, even under hypoxic conditions we have been unable to generate XaXa iPSCs[97]. This discrepancy could be because of inherent differences between human ESCs and iPSCs, and it will take further work to understand what the molecular

X chromosome inactivation
Transcriptional silencing of one of the two X chromosomes in female mammalian cells, initiated during development when epiblast cells of the blastocyst differentiate.

Naive pluripotent state
This stage of pluripotency is captured *in vitro* in the form of mouse embryonic stem cells or induced pluripotent stem cells. These cells can differentiate *in vitro* into many different cell types and, upon injection into blastocysts, can give rise to all tissues of the mouse, including the germ line.

Primed pluripotent state
This stage of pluripotency is captured *in vitro* in the form of mouse epiblast stem cells and is considered developmentally more advanced than naive pluripotency, with respect to X-inactivation, signalling dependence, gene expression and the inability to contribute to chimeric animals. Human embryonic stem cells are more similar to mouse epiblast stem cells.

Epiblast stem cells
Primed pluripotent cells derived from the post-implantation mouse epiblast of day 5.5–6.5 embryos.

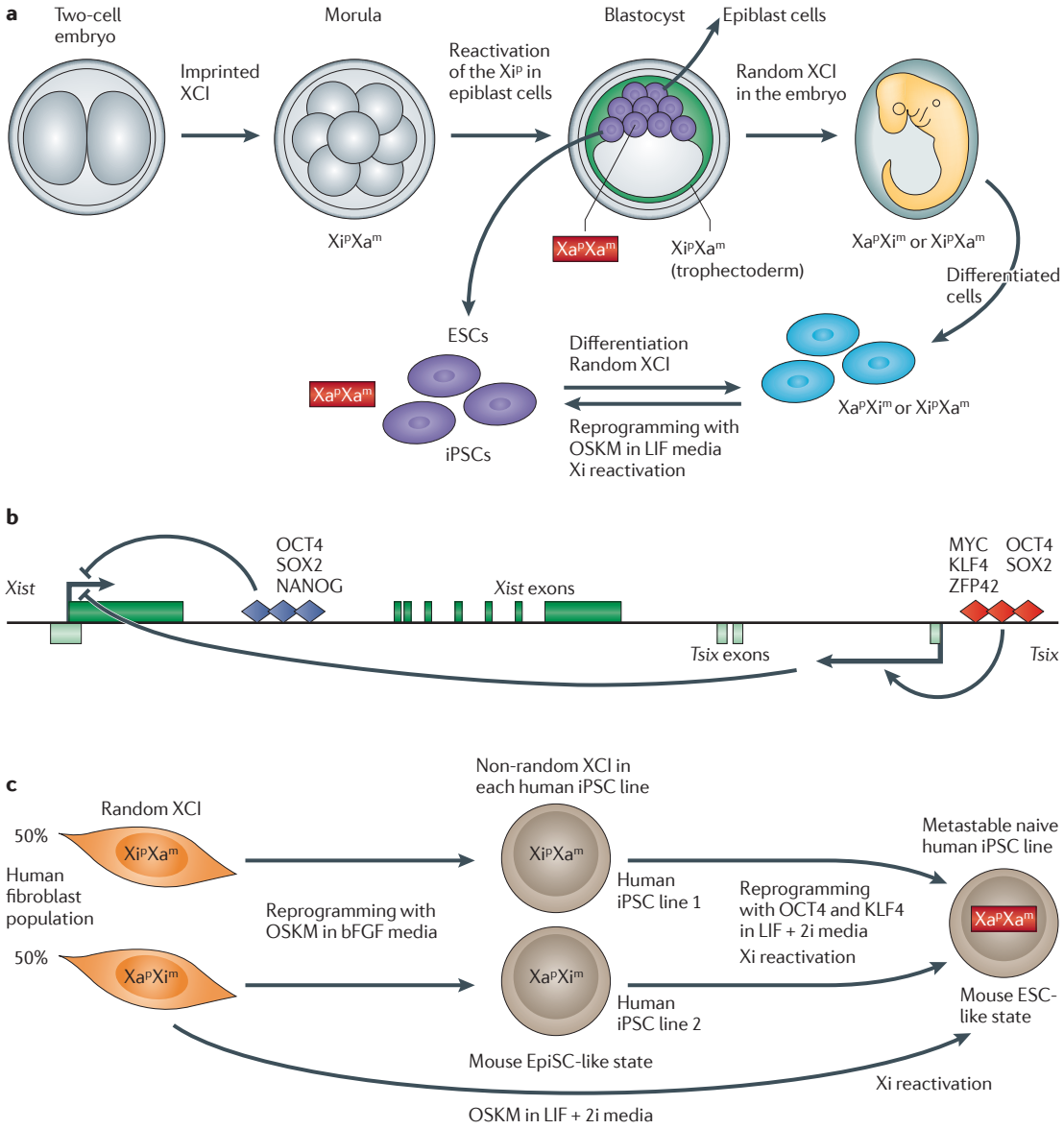


Figure 3 | **X chromosome inactivation and reprogramming. a** | Reactivation of the inactive X chromosome (Xi) is observed during female mouse pre-implantation development. Here, X chromosome inactivation (XCI) occurs first in an imprinted fashion that exclusively inactivates the paternally inherited X chromosome (X^p) in all cells of the pre-implantation embryo. Reactivation of the Xi^p in epiblast cells of the blastocyst allows a second round of XCI, in which each cell, upon differentiation, has a random chance of inactivating the X^p or maternally inherited X chromosome (X^m). Mouse embryonic stem cells (ESCs), which are derived from epiblast cells of the blastocyst, therefore carry two active X chromosomes (XaXa). Reprogramming of female mouse cells with OSKM (OCT4 (also known as POU5F1), SOX2, Krüppel-like factor 4 (KLF4) and MYC (also known as c-MYC)) in media containing leukaemia inhibitory factor (LIF) leads to reactivation of the Xi such that mouse induced pluripotent stem cells (iPSCs), like female mouse ESCs, have XaXa[5], but the underlying mechanism remains unknown. **b** | *Xist* encodes a non-coding RNA that is the key mediator of XCI. In mice, XCI is closely coupled to pluripotency, as pluripotency transcription factors are involved in *Xist* regulation. It has been suggested that to maintain XaXa status in mouse ESCs, different transcription factors may directly repress *Xist* by binding to its intronic region (purple shapes) and activate *Tsix* (*Xist*'s antisense transcript) by binding to its regulatory regions (red shapes); *Tsix* transcription also blocks *Xist* accumulation[94–96]. These findings may explain why Xi reactivation during mouse somatic cell reprogramming appears to occur simultaneously with establishment of the pluripotency network. The exact timing of Xi reactivation relative to the pluripotency network remains to be established. **c** | Reprogramming of female human fibroblasts to the typical human iPSC state with basic fibroblast growth factor (bFGF) culture conditions does not lead to Xi reactivation[97]. In each female human iPSC line, the XCI status of the single fibroblast cell it originated from is therefore propagated, resulting in non-random XCI (centre). It is thought that human iPSCs represent the mouse epiblast stem cell (EpiSC) state. Reprogramming under LIF cell culture conditions with 2i (a small molecule cocktail that inhibits mitogen-activated protein kinases and glycogen synthase kinase 3) leads to the establishment of metastable mouse ESC-like human iPSCs with XaXa[101]. ZFP42, zinc finger protein 42 (also known as REX1).

Table 1 | **Similarities and differences between iPSCs and ESCs**

Characteristic	ESCs vs. iPSCs (mouse)	ESCs vs. iPSCs (human)
mRNA expression	Early-passage iPSCs are distinct from ESCs, reflecting expression from the cell of origin[37,38]; late-passage iPSCs are nearly identical to ESCs[13,37]	Early-passage iPSCs are distinct[2,3] from ESCs, reflecting expression from the target cell[105,106]; late-passage iPSCs are closer to ESCs[2,3,111]
miRNA expression	The imprinted *Dlk1–Dio3* cluster is not expressed in most iPSC lines[13]	Some differences have been described[2,118], but no consistent differences have been found across multiple ESC and iPSC lines[119]
lncRNA expression	Not determined	Differences have been described; some have functional roles in reprogramming[109]
Histone modifications	Those modifications tested (H3K4me3 and H3K27me3)[5,6] seem to be indistinguishable between ESCs and iPSCs	Two modifications (H3K4me3 and H3K27me3) seem to be identical; H3K9me3 is different[2,4,111]
DNA methylation	Distinct at early passage, reflecting the pattern of target cells[37,38]; late-passage cells are nearly identical[37]	Some differences have been described[107,108,110]
X chromosome activation status	Both iPSCs and ESCs are XaXa[5]	Human ESCs are mostly XiXa but can be XaXa depending on culture conditions[102–104]; human iPSCs are XaXi[97]
Metabolism	Not determined	Identical or nearly identical[120,121]

ESC, embryonic stem cell; H3K4me3, histone H3 lysine 4 trimethylation; iPSC, induced pluripotent stem cell; lncRNA, long non-coding RNA; miRNA, microRNA; Xa, active X chromosome; Xi, inactive X chromosome.

differences between these cell types can tell us about the process of reprogramming.

Insights from the iPSC state

Reprogramming to the iPSC state by the introduction of pluripotency transcription factors seems to generate pluripotent stem cells that are superficially indistinguishable from ESCs[5,7–9,14–16]. However, numerous studies have now described molecular differences between iPSCs and ESCs in both mouse and human systems[2,3,13,37,38,105–110], whereas others argue that there are no fundamental differences between them[111]. Future research will provide clarity on this issue and in the context of this Review we consider these studies with an eye towards using their findings to understand the mechanisms underlying the reprogramming process (TABLE 1).

Evaluating the pluripotent state by transcription. So far, iPSCs and ESCs have been compared at the epigenetic, transcriptional, proteomic and metabolic levels. Our group and others have performed several analyses of human iPSCs and ESCs and suggested that these two cell types, although very similar, can still be distinguished by their expression of protein-coding RNAs[2,3,105,106]. A substantial portion of the gene expression differences between human ESCs and iPSCs are due to residual expression of somatic genes[108,111], and many of these differences seem to dissipate upon extended passaging[2,3]. There are several possible explanations that are not mutually exclusive: that reprogramming is not immediately complete upon induction of the endogenous pluripotency network; that there is selection of authentic pluripotent cells within a heterogenous culture over time; or perhaps that the exogenous versions of the reprogramming factors need to be silenced completely to complete the process.

One group has shown that repression of a small group of non-coding RNAs encoded in the *Dlk1–Dio3* imprinted gene cluster may distinguish mouse iPSCs from ESCs at a functional level[13]. The pluripotent stem cell lines (ESC or iPSC) that show normal expression of these genes are able to contribute to animals entirely derived from these cells in the tetraploid complementation assay, whereas those iPSCs that do not show expression at this locus are able to generate normal chimaeras but are incapable of satisfying this 'gold standard' assay for mouse pluripotency. Therefore, these non-coding RNAs might serve as a 'landmark' for reprogramming. These experiments suggest that reprogramming is complete in some cases. However, it should be noted that the lines were analysed at a later passage, when many of the expression differences often observed between iPSCs and ESCs have disappeared (K. Hochedlinger, personal communication). Because the mechanisms by which these imprinted non-coding RNAs are regulated have only begun to be explored, it is difficult to link the mechanism by which these RNAs are misexpressed and the process of reprogramming; further work should shed light on this issue.

Recently, it was shown that there are ten large intergenic non-coding RNAs (lincRNAs) that are differentially expressed between human iPSCs and ESCs, and that at least one of these can play a part in the reprogramming process as its overexpression enhances and its depletion inhibits this process[109]. The fact that at least some of the misregulated lincRNAs are targets of OCT4 and SOX2 in pluripotent cells indicates that they could be deregulated during reprogramming owing to aberrant binding of the reprogramming factors.

Epigenetic analysis to assess the extent of reprogramming. Extensive examination of the chromatin state of iPSCs and ESCs has also shown that although these two cell types are clearly very similar, consistent differences can be observed, and some differences have been shown to be functionally relevant. As described above, based on X inactivation status it could be argued that at least some human ESC lines are more epigenetically 'pristine' than human iPSCs (even when apparently at the same developmental stage), but it is unknown whether X inactivation status simply reflects the biology of this chromosome or if it is a clue to more profound genome-wide epigenetic variability.

Genome-wide approaches to identify sites enriched in H3K4me3 and H3K27me3 have suggested that human iPSCs and ESCs have identical patterns for these marks, even for promoters of genes that are differentially expressed between the two cell types[2,4,111]. However, the pattern of H3K9me3 within promoter regions was found to be different, and this mark is overrepresented among genes that were differentially expressed between human ESCs and iPSCs[4]. There is of course a panoply of other histone modifications that have yet to be probed and it is challenging to demonstrate a functional role for these marks at particular genomic sites, so it is difficult to use these differences to elucidate mechanisms of reprogramming. Most of the transcriptional and chromatin differences described to date seem to reflect the state found in the cell type that was reprogrammed, suggesting a form of 'epigenetic memory' that might indicate incomplete reprogramming[2,105,106].

In fact, two groups showed that the DNA methylation pattern of the original cell persists in mouse iPSCs and demonstrated that this residual DNA methylation pattern affects their differentiation potential[37,38]. For instance, iPSCs from the blood more easily differentiate towards blood lineages than iPSCs made from fibroblasts[37,38]. Importantly, many blood markers are hypermethylated at the DNA level in fibroblast-derived iPSCs, probably preventing their efficient upregulation upon induction of differentiation towards the blood lineage. Furthermore, treatment of iPSCs generated from non-blood lineages with histone deacetylase and DNA methylation inhibitors seemed to allow for more efficient blood differentiation. The fact that residual DNA methylation within lineage-specific genes is found in iPSCs provides tangible evidence that resetting this mark is fundamental to reprogramming, and that failure to do so has a functional consequence. Notably, one of these studies also showed that continued passaging of the iPSCs appeared to erase this epigenetic memory[37], a finding reminiscent of work in human iPSCs which showed that continued passaging abrogated transcriptional differences between iPSCs and ESCs[2,3].

Recent work has also uncovered an epigenetic memory in human iPSCs at the level of DNA methylation by generating single-base, whole-genome DNA methylation maps[110]. This study also argued that as well as failing to properly erase parts of the somatic DNA methylome, which leads to an epigenetic memory of the somatic DNA methylation pattern, reprogramming often induces aberrant methylation that seems to be specific to the iPSC state, and that some iPSCs are unable to re-establish ESC-like methylation, particularly non-CpG methylation. These methylation differences between ESCs and iPSCs are associated with differences at the transcriptional level that can be found after many passages and might affect the differentiation behaviour of these cells.

The difficulty with all of these molecular comparisons is that both iPSCs and ESCs show significant variability among individual lines. To quantify such variability, a recent study profiled 20 human ESC and 12 iPSC lines and generated a 'scorecard' to measure the fidelity and utility of reprogrammed lines versus a set of standard ESC lines[108]. This effort included DNA methylome, transcriptome and differentiation studies to determine whether quantification of molecular similarity to gold standard pluripotent cells could be predictive of their ability to differentiate down various lineages. The study concluded that although ESCs exhibit significant variability across individual lines, and that some iPSCs fall within the variability of ESCs, iPSCs were more variable at the molecular level than ESCs. It is imperative that any differences between iPSCs and ESCs be tested experimentally to determine whether or not they are functionally significant[13,109], as these experiments will yield mechanistic insights into reprogramming and the pluripotent state, as well as whether one of these pluripotent cell types is more suitable for a desired application.

Prospective approaches to study reprogramming

Currently, the compendium of differences described between iPSCs and ESCs is further evidence that the reprogramming process requires a wide variety of molecular changes and that cells can either get trapped (partial reprogramming), get close to the final destination (reprogrammed state with epigenetic memory) or reach a bona fide pluripotent state. Perhaps the only way to truly understand the reprogramming process will be to extend the recent studies that combined single-cell analysis with fine temporal resolution[18,65].

To understand the mechanisms of the reprogramming factors, several groups have used tetracycline-inducible expression of the reprogramming factors (BOX 2). An inducible system facilitates the identification of broad landmarks of reprogramming, such as the suppression of somatic genes and induction of epithelial and pluripotency genes, as discussed above. Strikingly, even with robust expression of all reprogramming factors by polycistronic methods, typically only a few cells undergo complete reprogramming, suggesting that formidable barriers to the process exist beyond expression of the Yamanaka factors. As we can currently only identify faithful reprogramming events when the pluripotency network is expressed, it is nearly impossible to determine which earlier molecular changes occur in cells that are destined to make it to a pluripotent state versus those that will end up lost along the way. To overcome this obstacle, we either need a technique that enables a very high reprogramming efficiency or a set of early epigenetic landmarks that reliably mark cells that will proceed towards complete pluripotency. Similarly, inducible reprogramming factor expression and single-cell approaches need to be combined with genome-wide approaches such as transcriptome analysis and chromatin immunoprecipitation followed by high-throughput sequencing (ChIP–seq). Currently, merging these technologies is still very challenging, if not prohibitive, but it will be essential to understand the molecular steps underlying reprogramming. If someday all of these issues can be adequately addressed, we may be able to gain a clear understanding of reprogramming. Until then, we will have to rely on studies that use transcriptional or epigenetic manipulations that drive or impede the process to shed light on this fascinating phenomenon.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006). **The first demonstration that expression of four pluripotency-related transcription factors can convert somatic cells to a pluripotent state (now known as the iPSC state).**
2. Chin, M. H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111–123 (2009).
3. Chin, M. H., Pellegrini, M., Plath, K. & Lowry, W. E. Molecular analyses of human induced pluripotent stem cells and embryonic stem cells. *Cell Stem Cell* **7**, 263–269 (2010).
4. Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
5. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007). **The first characterization of XCI status in iPSC reprogramming. It demonstrated that mouse iPSCs are XaXa and can undergo random X-inactivation, indicating complete erasure of the memory of the prior inactive X chromosome.**
6. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
7. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
8. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
9. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
10. Boland, M. J. *et al.* Adult mice generated from induced pluripotent stem cells. *Nature* **461**, 91–94 (2009).
11. Kang, L., Wang, J., Zhang, Y., Kou, Z. & Gao, S. iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell Stem Cell* **5**, 135–138 (2009).
12. Zhao, X. Y. *et al.* iPS cells produce viable mice through tetraploid complementation. *Nature* **461**, 86–90 (2009).
13. Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175–181 (2010).
14. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
15. Park, I. H., Lerou, P. H., Zhao, R., Huo, H. & Daley, G. Q. Generation of human-induced pluripotent stem cells. *Nature Protoc.* **3**, 1180–1186 (2008).
16. Lowry, W. E. *et al.* Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl Acad. Sci. USA* **105**, 2883–2888 (2008).
17. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009). **This study addressed the inefficiency of reprogramming, argued for the first time that all cells of the starting population have the potential to be reprogrammed if given enough time in culture, and provided the first quantitative data on the actions of the reprogramming factors.**
18. Smith, Z. D., Nachman, I., Regev, A. & Meissner, A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nature Biotech.* **28**, 521–526 (2010). **The first high-resolution time-lapse imaging approach that enabled the retroactive tracking of faithful reprogramming events, and recognized that the path to the iPSC state begins with a transition to faster proliferation and a decrease in cell size immediately upon induction of the reprogramming factors.**
19. 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).
20. Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**, 151–159 (2008). **References 19 and 20 represent some of the first efforts to determine steps of reprogramming and define reprogramming factor dependence using an inducible system for reprogramming factor expression.**
21. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51–63 (2010).
22. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010). **References 21 and 22 identified the MET as an important step in the reprogramming of fibroblasts and showed that modulation of signalling pathways that affect the MET alter reprogramming efficiency.**
23. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364–377 (2009). **The first detailed description of the action of the reprogramming factors during induction of pluripotency, demonstrating differences between MYC and the other three reprogramming factors and defining barriers to the final step of reprogramming.**
24. Silva, J. *et al.* Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6**, e253 (2008).
25. Mali, P. *et al.* Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells* **28**, 713–720 (2010).
26. Shi, Y. *et al.* Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* **3**, 568–574 (2008).
27. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotech.* **26**, 795–797 (2008).
28. Huangfu, D. *et al.* Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nature Biotech.* **26**, 1269–1275 (2008).
29. Liang, G., Taranova, O., Xia, K. & Zhang, Y. Butyrate promotes induced pluripotent stem cell generation. *J. Biol. Chem.* **285**, 25516–25521 (2010).
30. Han, D. W. *et al.* Direct reprogramming of fibroblasts into epiblast stem cells. *Nature Cell Biol.* **13**, 66–71 (2011).
31. Ichida, J. K. *et al.* A small-molecule inhibitor of Tgf-β signaling replaces *Sox2* in reprogramming by inducing *nanog*. *Cell Stem Cell* **5**, 491–503 (2009).
32. Maherali, N. & Hochedlinger, K. Tgfβ signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr. Biol.* **19**, 1718–1723 (2009).
33. Marson, A. *et al.* Wnt signaling promotes reprogramming of somatic cells to pluripotency. *Cell Stem Cell* **3**, 132–135 (2008).
34. Esteban, M. A. *et al.* Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6**, 71–79 (2010).
35. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722–737 (2009). **A detailed analysis of the role of NANOG in establishing pluripotency, particularly in the reprogramming process. This paper demonstrated that NANOG is essential for generating iPSCs and is required during the final step of reprogramming, in which its overexpression also acts to enhance the process.**
36. Ruiz, S. *et al.* A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr. Biol.* **21**, 45–52 (2010).
37. Polo, J. M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature Biotech.* **28**, 848–855 (2010).
38. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285–290 (2010). **References 37 and 38 provide functional evidence showing that an epigenetic memory of the target cell is present in mouse iPSCs, which can influence the differentiation behaviour of these cells and may be linked to DNA methylation.**
39. Feng, B., Ng, J. H., Heng, J. C. & Ng, H. H. Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. *Cell Stem Cell* **4**, 301–312 (2009).
40. Shao, L. *et al.* Generation of iPS cells using defined factors linked via the self-cleaving 2A sequences in a single open reading frame. *Cell Res.* **19**, 296–306 (2009).
41. Gonzalez, F. *et al.* Generation of mouse-induced pluripotent stem cells by transient expression of a single nonviral polycistronic vector. *Proc. Natl Acad. Sci. USA* **106**, 8918–8922 (2009).
42. Chang, C. W. *et al.* Polycistronic lentiviral vector for "hit and run" reprogramming of adult skin fibroblasts to induced pluripotent stem cells. *Stem Cells* **27**, 1042–1049 (2009).
43. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nature Methods* **7**, 53–55 (2010).
44. Woltjen, K. *et al. piggyBac* transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766–770 (2009).
45. Carey, B. W. *et al.* Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proc. Natl Acad. Sci. USA* **106**, 157–162 (2009).
46. Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotech.* **26**, 916–924 (2008).
47. Hockemeyer, D. *et al.* A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* **3**, 346–353 (2008).
48. Sommer, C. A. *et al.* Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* **27**, 543–549 (2009).
49. 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).
50. Carey, B. W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nature Methods* **7**, 56–59 (2009).
51. Stadtfeld, M., Brennand, K. & Hochedlinger, K. Reprogramming of pancreatic β cells into induced pluripotent stem cells. *Curr. Biol.* **18**, 890–894 (2008).
52. Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nature Genet.* **41**, 968–976 (2009).
53. Loh, Y. H. *et al.* Reprogramming of T cells from human peripheral blood. *Cell Stem Cell* **7**, 15–19 (2010).
54. Seki, T. *et al.* Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. *Cell Stem Cell* **7**, 11–14 (2010).
55. Staerk, J. *et al.* Reprogramming of human peripheral blood cells to induced pluripotent stem cells. *Cell Stem Cell* **7**, 20–24 (2010).
56. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250–264 (2008).
57. Aoi, T. *et al.* Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321**, 699–702 (2008).
58. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced pluripotent stem cells generated without viral integration. *Science* **322**, 945–949 (2008).
59. 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).
60. Yu, J. *et al.* Human induced pluripotent stem cells free of vector and transgene sequences. *Science* **324**, 797–801 (2009).
61. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**, 618–630 (2010).
62. González, F., Boué, S. & Belmonte, J. C. I. Methods for making induced pluripotent stem cells: reprogramming à la carte. *Nature Rev. Genet.* **12**, 231–242 (2011).
63. Winkler, T. *et al.* No evidence for clonal selection due to lentiviral integration sites in human induced pluripotent stem cells. *Stem Cells* **28**, 687–694 (2010).
64. Varas, F. *et al.* Fibroblast-derived induced pluripotent stem cells show no common retroviral vector insertions. *Stem Cells* **27**, 300–306 (2009).
65. Chan, E. M. *et al.* Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nature Biotech.* **27**, 1033–1037 (2009).
66. Koche, R. P. *et al.* Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8**, 96–105 (2011).
67. Banito, A. *et al.* Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev.* **23**, 2134–2139 (2009).

68. Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460**, 1140–1144 (2009).
69. Li, H. *et al.* The *Ink4/Arf* locus is a barrier for iPS cell reprogramming. *Nature* **460**, 1136–1139 (2009).
70. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53–p21 pathway. *Nature* **460**, 1132–1135 (2009).
71. Marion, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**, 1149–1153 (2009).
72. Utikal, J. *et al.* Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* **460**, 1145–1148 (2009).
73. Krizhanovsky, V. & Lowe, S. W. Stem cells: the promises and perils of p53. *Nature* **460**, 1085–1086 (2009).
74. Pereira, C. F. *et al.* ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. *Cell Stem Cell* **6**, 547–556 (2010).
75. Chou, Y. F. *et al.* The growth factor environment defines distinct pluripotent ground states in novel blastocyst-derived stem cells. *Cell* **135**, 449–461 (2008).
76. Aasen, T. *et al.* Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. *Nature Biotech.* **26**, 1276–1284 (2008).
77. Markoulaki, S. *et al.* Transgenic mice with defined combinations of drug-inducible reprogramming factors. *Nature Biotech.* **27**, 169–171 (2009).
78. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotech.* **26**, 101–106 (2008).
79. 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).
80. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
81. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061 (2008).
82. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
83. Jiang, J. *et al.* A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nature Cell Biol.* **10**, 353–360 (2008).
84. Kim, J. *et al.* A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* **143**, 313–324 (2010).
85. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
86. Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364–368 (2006).
87. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
88. Theunissen, T. W. *et al.* Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. *Curr. Biol.* **21**, 65–71 (2011).
89. Feldman, N. *et al.* G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nature Cell Biol.* **8**, 188–194 (2006).
90. Li, W. *et al.* Generation of human induced pluripotent stem cells in the absence of exogenous *Sox2*. *Stem Cells* **27**, 2992–3000 (2009).
91. Singhal, N. *et al.* Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* **141**, 943–955 (2010).
92. Gaspar-Maia, A. *et al.* Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460**, 863–868 (2009).
93. Payer, B. & Lee, J. T. X chromosome dosage compensation: how mammals keep the balance. *Annu. Rev. Genet.* **42**, 733–772 (2008).
94. Navarro, P. *et al.* Molecular coupling of *Xist* regulation and pluripotency. *Science* **321**, 1693–1695 (2008).
95. Navarro, P. *et al.* Molecular coupling of *Tsix* regulation and pluripotency. *Nature* **468**, 457–460 (2010).
96. Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N. & Lee, J. T. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128–132 (2009).
97. Tchieu, J. *et al.* Female human iPS cells retain an inactive X-chromosome. *Cell Stem Cell* **7**, 329–342 (2010).
98. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).
99. Guo, G. *et al. Klf4* reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063–1069 (2009).
100. Bao, S. *et al.* Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* **461**, 1292–1295 (2009).
101. Hanna, J. *et al.* Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc. Natl Acad. Sci. USA* **107**, 9222–9227 (2010).
The first demonstration that human fibroblasts as well as human ESCs and iPSCs can be driven to a more primitive pluripotent state known as naive pluripotency. Based on X-inactivation status, morphology, gene expression and signalling dependence, this state is very similar to that of mouse ESCs.
102. Shen, Y. *et al.* X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. *Proc. Natl Acad. Sci. USA* **105**, 4709–4714 (2008).
103. Silva, S. S., Rowntree, R. K., Mekhoubad, S. & Lee, J. T. X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 4820–4825 (2008).
104. Lengner, C. J. *et al.* Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell* **141**, 872–883 (2010).
105. Ghosh, Z. *et al.* Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PLoS ONE* **5**, e8975 (2010).
106. Marchetto, M. C. *et al.* Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS ONE* **4**, e7076 (2009).
107. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genet.* **41**, 1350–1353 (2009).
108. Bock, C. *et al.* Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452 (2011).
109. Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature Genet.* **42**, 1113–1117 (2010).
The authors uncovered large non-coding RNAs that are differentially expressed between human iPSCs and ESCs and demonstrated a role in reprogramming for one of these RNAs.
110. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 2 Feb 2011 (doi:10.1038/nature09798).
111. Guenther, M. G. *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* **7**, 249–257 (2010).
112. Sommer, C. A. *et al.* Excision of reprogramming transgenes improves the differentiation potential of iPS cells generated with a single excisable vector. *Stem Cells* **28**, 64–74 (2010).
113. Roth, S. Y., Denu, J. M. & Allis, C. D. Histone acetyltransferases. *Annu. Rev. Biochem.* **70**, 81–120 (2001).
114. Zhou, H. *et al.* Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* **4**, 381–384 (2009).
115. Bolden, J. E., Peart, M. J. & Johnstone, R. W. Anticancer activities of histone deacetylase inhibitors. *Nature Rev. Drug Discov.* **5**, 769–784 (2006).
116. Chang, Y. *et al.* Structural basis for G9a-like protein lysine methyltransferase inhibition by BIX-01294. *Nature Struct. Mol. Biol.* **16**, 312–317 (2009).
117. Shi, Y. *et al.* A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* **2**, 525–528 (2008).
118. Wilson, K. D. *et al.* MicroRNA profiling of human-induced pluripotent stem cells. *Stem Cells Dev.* **18**, 749–758 (2009).
119. Neveu, P. *et al.* MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states. *Cell Stem Cell* **7**, 671–681 (2010).
120. Prigione, A., Fauler, B., Lurz, R., Lehrach, H. & Adjaye, J. The senescence-related mitochondrial/oxidative stress pathway is repressed in human induced pluripotent stem cells. *Stem Cells* **28**, 721–733 (2010).
121. Armstrong, L. *et al.* Human induced pluripotent stem cell lines show stress defense mechanisms and mitochondrial regulation similar to those of human embryonic stem cells. *Stem Cells* **28**, 661–673 (2010).

Acknowledgements
K.P. is supported by the US National Institutes of Health (NIH) Director's Young Innovator Award (DP2OD001686) and a California Institute for Regenerative Medicine Young Investigator Award (RN1-00564). W.E.L. is the Maria Rowena Ross Professor of Cell Biology and Biochemistry and is supported by the NIH, The March of Dimes, and the Fuller Foundation. K.P. and W.E.L. are supported by the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at the University of California Los Angeles.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Kathrin Plath's homepage: http://www.biolchem.ucla.edu/labs/plath
William E. Lowry's homepage: http://lowrylab.com
***Nature Reviews Genetics* Focus issue on Stem cells:** http://www.nature.com/focus/stemcells/index.html
UCLA Broad Stem Cell Research Center: https://stemcell.ucla.edu

ALL LINKS ARE ACTIVE IN THE ONLINE PDF