Mechanisms and models of somatic cell reprogramming

Yosef Buganim[1], Dina A. Faddah[1,2] and Rudolf Jaenisch[1,2]

Abstract | Conversion of somatic cells to pluripotency by defined factors is a long and complex process that yields embryonic-stem-cell-like cells that vary in their developmental potential. To improve the quality of resulting induced pluripotent stem cells (iPSCs), which is important for potential therapeutic applications, and to address fundamental questions about control of cell identity, molecular mechanisms of the reprogramming process must be understood. Here we discuss recent discoveries regarding the role of reprogramming factors in remodelling the genome, including new insights into the function of MYC, and describe the different phases, markers and emerging models of reprogramming.

Epigenome
Heritable changes in chromatin (such as histone post-translational modifications and DNA methylation) that affect gene expression.

Reprogramming
Conversion of one cell type to another cell type by transcription factors or chemically defined media.

Cell plasticity
The ability of a cell to acquire a new identity and to adopt an alternative fate when exposed to different conditions.

[1]*Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA.*
[2]*Department of Biology, Massachusetts Institute of Technology, 31 Ames Street, Cambridge, Massachusetts 02139, USA.*
Correspondence to R.J.
e-mail: jaenisch@wi.mit.edu
doi:10.1038/nrg3473

Resetting the epigenome of a somatic cell to a pluripotent state has been achieved by somatic cell nuclear transfer (SCNT), cell fusion and ectopic expression of defined factors such as OCT4 (also known POU5F1), SOX2, Krüppel-like factor 4 (KLF4) and MYC (also known as c-MYC); collectively, these are known as OSKM factors[1–3]. Understanding the molecular mechanisms that underlie somatic cell reprogramming to pluripotency is crucial for the creation of high-quality pluripotent cells and may be useful for therapeutic applications. Moreover, insights gained from *in vitro* reprogramming approaches may yield relevant information for SCNT or cell-fusion-mediated reprogramming and may broaden our understanding of fundamental questions regarding cell plasticity, cell identity and cell fate decisions[4–6].

Reprogramming by SCNT is rapid, thought to be deterministic and yields embryonic stem cells (ESCs) from the cloned embryo that are similar to ESCs derived from the fertilized embryo[7,8]. However, the investigation of SCNT and cell fusion is difficult because oocytes and ESCs contain multiple gene products that may be involved in reprogramming. By contrast, in the transcription-factor-mediated reprogramming method, the factors that initiate the process are known and can easily be modulated, which makes examination of the process less complicated and easier to follow. However, the process is long, inefficient and generates induced pluripotent stem cells (iPSCs) that vary widely in their developmental potential[1,2,9,10].

In this Review, we focus on recent studies and technologies aimed at understanding the molecular mechanisms of cellular reprogramming mediated by transcription factors. For example, insights have been gained from methods to study single cells as well as studies of populations of cells undergoing reprogramming. We describe current views of the phases of transcriptional and epigenetic changes that occur and discuss new concepts regarding the role of OSKM in driving the conversion to pluripotency. We then consider markers of cells progressing through reprogramming and emerging models of the process. Finally, we summarize criteria that allow assessment of iPSC quality.

Phases of reprogramming

Insights gained from population-based studies. After the first demonstration of reprogramming to pluripotency by defined factors[11,12], many groups raced to study the reprogramming process by analysing transcriptional and epigenetic changes in cell populations at different time points after factor induction. These are the most straightforward experiments to carry out for unravelling the molecular mechanism of this complicated process. Most studies analysing cellular changes during the reprogramming process have been done using populations of mouse embryonic fibroblasts (MEFs).

Microarray data at defined time points during the reprogramming process[13] showed that the immediate response to OSKM is characterized by de-differentiation of MEFs and upregulation of proliferation genes; this is consistent with the expression of MYC. Gene expression profiling and RNA interference (RNAi) screening in fibroblasts revealed three phases of reprogramming termed initiation, maturation and stabilization; the initiation phase is marked by a mesenchymal-to-epithelial transition (MET)[14,15]. Also, bone morphogenic protein (BMP) signalling has been shown to synergize with

OSKM to stimulate a microRNA (miRNA) expression signature associated with MET-promoting progression through the initiation phase[15].

The late maturation and stabilization phases have been studied by tracing clonally derived cells[16]. This study showed that repression of the OSKM transgenes is required for the transition from the maturation to the stabilization phase. By comparing the expression profiles of clones that could transit from the maturation to the stabilization phase to those that could not, the authors found a unique signature associated with pluripotent competency. Surprisingly, few pluripotency regulators had a role in the maturation–stabilization transition. Rather, genes that are associated with gonads, gametes, cytoskeletal dynamics and signalling pathways were upregulated during this phase[16] (FIG. 1). The authors also found that genes that are induced on transgene inhibition (for example, ESC-expressed Ras (*Eras*) and left–right determination factor 2 (*Lefty2*)) tend to be important for ESC maintenance, whereas genes that retain a similar expression level before and after transgene silencing (for example, AT-rich interactive domain 3B (*Arid3b*) and Sal-like 1 (*Sall1*)) tend to be involved in regulating the maturation–stabilization transition. This study suggests that the transition to the stabilization phase on transgene removal is dependent on regulatory pathways that are distinct from those controlling ESC pluripotency[16].

Another study used genome-wide analyses to examine intermediate cell populations poised to become iPSCs[17]. This study revealed two distinct waves of major gene activity: the first wave occurred between days 0 and 3, and the second wave started after day 9, which is towards the end of the process (day 12). The number of differentially expressed genes between progressing cells and cells that are refractory to reprogramming at each time point was gradually increased, reaching 1,500 genes by the end of the process[17]. The first wave was characterized by the activation of genes responsible for proliferation, metabolism, cytoskeleton organization and downregulation of genes associated with development (FIG. 1). This step occurred in most cells and is equivalent to the initiation phase described above. Several early pluripotency-associated genes were gradually upregulated, and some developmental and cell-type-specific genes were transiently regulated during the process. The second wave was characterized by the expression of genes responsible for embryonic development and stem cell maintenance. Genes from this step facilitate the activation of the core pluripotency network and mark the acquisition of a stable pluripotent state. By contrast, genes related to extracellular space or matrix, plasma membrane, retinoic acid binding and immune response processes were aberrantly expressed in cells refractory to reprogramming[17].

In agreement with these findings, quantitative proteomic analysis during the course of reprogramming of fibroblasts to iPSCs revealed a two-step resetting of the proteome during the first 3 days and last 3 days of reprogramming[18]. Proteins related to regulation of gene expression, RNA processing, chromatin organization, mitochondria, metabolism, cell cycle and DNA repair were strongly induced at an early stage, and proteins related to the electron transport system were downregulated. In contrast to these processes, glycolytic enzymes exhibited a slow increase in the intermediate phase, suggesting a gradual transformation of energy metabolism[19]. Proteins involved in vesicle-mediated transport, extracellular matrix, cell adhesion and EMT were downregulated in the early phase, retained low levels during the intermediate step and became upregulated in the final stage[18]. These data suggest that reprogramming is a multi-step process characterized by two waves of transcriptome and proteome resetting[20].

Insights gained from single-cell studies. Knowledge gained from population-based studies is essential for understanding the global changes that occur in cells during the reprogramming process. A challenge for gaining mechanistic insights of reprogramming by the analysis of cell populations is cell heterogeneity. Because only a small fraction of the induced cells becomes reprogrammed, gene expression profiles of cell populations at different time points after factor induction will not detect changes in rare cells destined to become iPSCs. In an attempt to overcome the problem of cell heterogeneity, reprogramming has been traced at single-cell resolution using time-lapse microscopy[21,22]. Single-cell tracking by real-time microscopy has given insights into morphological changes during reprogramming, but the approach has not provided information on molecular events driving the process at the single-cell level. These studies showed that the cells underwent a shift in their proliferation rate and reduction in cell size soon after factor induction. These events occurred within the first cell division and with the same kinetics in all cells that give rise to iPSCs.

As a complementary approach to the population-based studies, two single-cell techniques have been used to quantify gene expression in the rare cells that undergo reprogramming[23]: Fluidigm BioMark, which allows quantitative analysis of 48 genes in duplicate in 96 single cells[24–27]; and single-molecule mRNA fluorescent *in situ* hybridization (sm-mRNA-FISH), which allows quantification of mRNA transcripts of up to three genes in hundreds to thousands of cells[28]. The 48 genes in the BioMark system included those known to be involved in major events that occur during reprogramming (for example, proliferation, epigenetic modification, ESC-supporting pathways, pluripotency markers and MEF markers). In the first 6 days after factor induction, there was high variation among cells in expression of the 48 genes[23]. This suggests that early in the reprogramming process OSKM factors induce stochastic gene expression changes in a subset of pluripotency genes that is crucial for instigation of the second phase (FIG. 1). These stochastic changes are in addition to the alterations in the expression of genes that control MET, proliferation and metabolism, which are global changes that must occur during reprogramming but are not restricted to cells that are destined to become iPSCs[15–17]. Single-cell analyses of clonally derived cell populations revealed that the stochastic gene expression phase is long and

Deterministic
A collection of actions during the reprogramming process that must occur in a particular order (that is, activation or silencing of different combinations of genes) before induced pluripotent stem cell formation.

Transcription-factor-mediated reprogramming
Conversion of a somatic cell to a pluripotent cell using defined transcription factors.

Developmental potential
The sum of all possible fates that a cell can undergo under any experimental condition.

Refractory
Unresponsive to a stimulus or unable to bind a transcription factor.

Cell heterogeneity
Variation among cells that occurs owing to gene expression differences.

Single-molecule mRNA fluorescent *in situ* hybridization
(sm-mRNA-FISH). An *in situ* hybridization method capable of detecting individual mRNA molecules, thus permitting the precise quantification and localization of mRNA within a single cell.

Stochastic
In this context, this term refers to an unpredictable and random action that leads at some point to the activation or repression of genes that will then set a cell on the path to becoming an induced pluripotent stem cell.

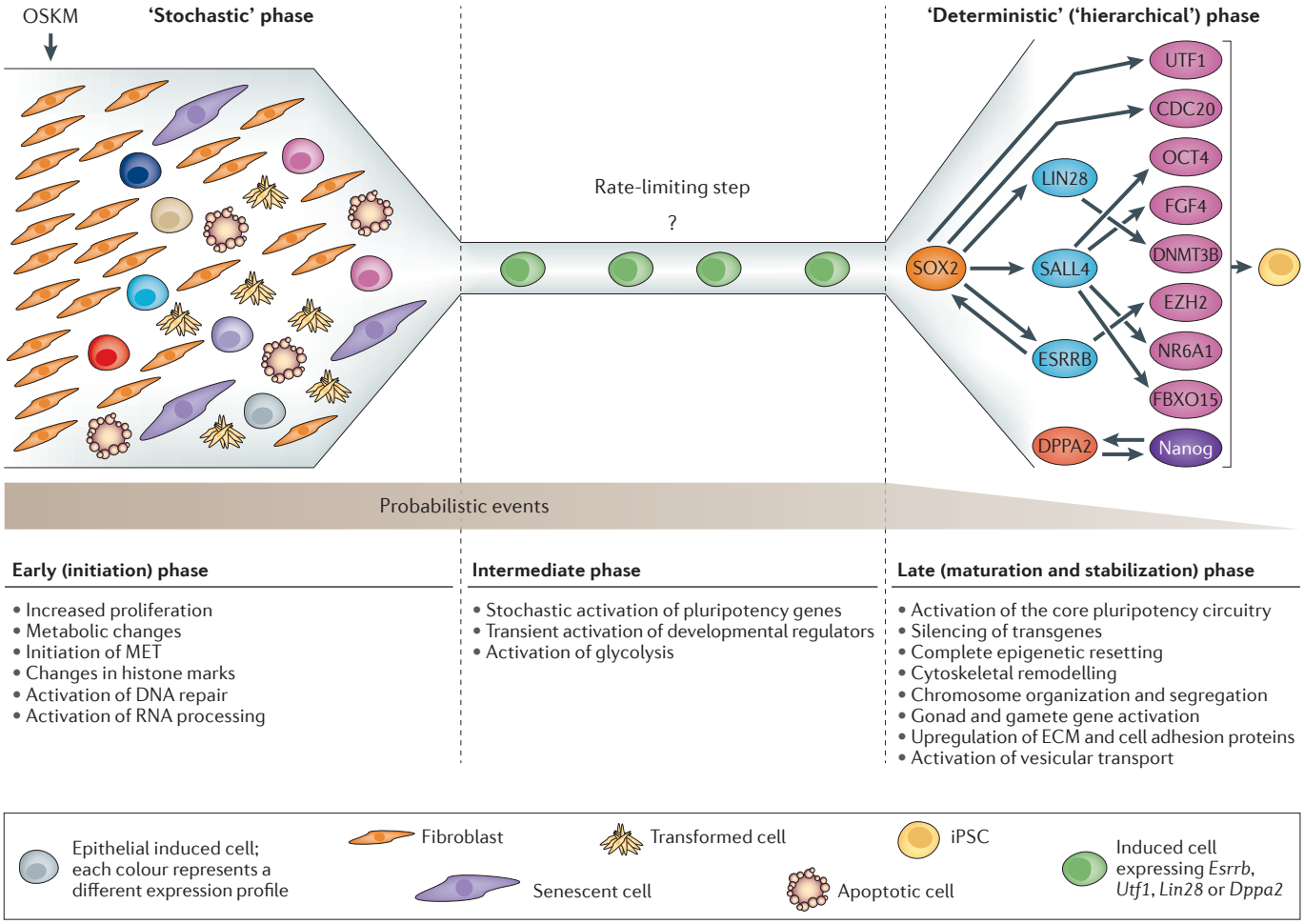

Figure 1 | **Phases of the reprogramming process.** In the model we discuss in this Review, the reprogramming process can broadly be divided into two phases: first, a long stochastic phase of gene activation; and second, a shorter, hierarchical, more deterministic phase of gene activation that begins with the activation of the *Sox2* locus. After a fibroblast is induced with OSKM (OCT4, SOX2, KLF4 and MYC), it will initiate stochastic gene expression and assume one of several possible fates (such as apoptosis, senescence, transformation, transdifferentiation or reprogramming). In the early phase, reprogrammable cells will increase proliferation, undergo changes in histone modifications at somatic genes, initiate mesenchymal-to-epithelial transition (MET) and activate DNA repair and RNA processing. The reprogrammable cells will then enter an intermediate phase with an unknown rate-limiting step that delays the conversion to induced pluripotent stem cells (iPSCs) and contributes to the long latency of the process. In this phase, cells undergo a stochastic activation of pluripotency markers[23], a transient activation of developmental regulators[17] and activation of glycolysis[18]. In general, the transcriptional changes in this phase are small. In some rare cases, the stochastic gene expression will lead to the activation of predictive markers, such as undifferentiated embryonic cell transcription factor 1 (*Utf1*), oestrogen-related receptor beta (*Esrrb*), developmental pluripotency associated 2 (*Dppa2*) and *Lin28*, which will then instigate the second phase, starting with the activation of *Sox2*. Activation of *Sox2* by the predictive markers can be direct or indirect and will trigger a series of deterministic events that will lead to an iPSC. In this late phase, the cells eventually stabilize into the pluripotent state, in which the transgenes are silenced, the cytoskeleton is remodelled to an embryonic stem cell (ESC)-like state, the epigenome is reset and the core pluripotency circuitry is activated[16–18,23]. In this model, probabilistic events decrease and hierarchical events increase as the cell progresses from a fibroblast to an iPSC. DNMT3B, DNA methyltransferase 3B; ECM, extracellular matrix; FBXO15, F box only protein 15; FGF4, fibroblast growth factor 4; NR6A1, nuclear receptor subfamily 6 group A member 1.

Rate-limiting
In this context, this term refers to a step that is responsible for the low efficiency of the reprogramming process. Reprogrammable cells must pass this step to instigate the late hierarchical phase and to become fully reprogramed. This step determines the length of the reprogramming process.

Hierarchical
An arrangement of items that are directly or indirectly linked. For reprogramming, this is a predictable sequence of gene activations or repressions.

variable[23]. Although cells with an ESC-like morphology appear early, they must pass through a bottleneck — probably a rate-limiting stochastic event — before transiting into stable iPSCs[23,29]. At a later stage, when the cells start to express *Nanog*, the variation between individual cells dramatically decreases, which is consistent with a model in which the early 'stochastic' phase of gene expression is followed by a 'deterministic' or more 'hierarchical' phase that leads to activation of the pluripotency circuitry. This deterministic or hierarchical phase is discussed further below in the context of models of reprogramming.

Chromatin modifiers
Proteins that can modify chromatin architecture and thereby control gene expression.

Epigenetic changes

The studies discussed above characterized phases of transcriptional changes during reprogramming; therefore, what are the epigenetic alterations that underlie these changes and what might drive them? The epigenetic signature of the somatic cell must be erased during the conversion in order to adopt a stem-cell-like epigenome. These changes include chromatin reorganization, DNA demethylation of promoter regions of pluripotency genes such as *Nanog, Sox2* and *Oct4*, reactivation of the somatically silenced X chromosome and genome-wide resetting of histone post-translational modifications[11,30–32]. There are more than 100 different histone post-translational modifications, and lysine methylation and acetylation are the ones that are most frequently studied[33]. Changes in histone marks and the role of various chromatin modifiers during reprogramming have been extensively reviewed elsewhere[4,34,35], so here we briefly summarize the key points. The roles of the relevant histone marks and of chromatin modifiers are summarized in TABLE 1 and TABLE 2, respectively.

DNA demethylation and X-chromosome reactivation occur late in the reprogramming process[17], whereas changes in histone modifications can be seen immediately after factor induction[36], suggesting that changes in histone marks are an early event that is associated with initiation of the reprogramming process. Immediately after factor induction, a peak of *de novo* deposition of the histone H3 dimethylated at lysine 4 (H3K4me2) mark is observed at promoter and enhancer regions. At this time, H3K4me2 accumulates at the promoters of many pluripotency genes, such as *Sall4* and fibroblast growth factor 4 (*Fgf4*), which are enriched for OCT4 and SOX2 binding sites and lack H3K4me1 or H3K4me3 marks[36]. This stage is also associated with a gradual depletion of H3K27me3 and promoter hypomethylation in regions that are important for the conversion[17]. However, at early time points, H3K4me2 does not correlate with the transcription-associated histone mark H3K36me3, occupancy of RNA polymerase II (RNA Pol II) or transcriptional activity, suggesting that these loci have not completed chromatin remodelling at early time points, and an additional step is required to achieve full activation of these genes[36]. At the beginning of the reprogramming process, changes in these modifications are almost exclusively restricted to CpG islands, as these regions are more responsive to transcription factor activity and permissive to change[37]. In parallel, the promoters of somatic genes begin to lose H3K4me2, which is consistent with early downregulation of MEF markers, such as thymus cell antigen 1 theta (*Thy1*) and periostin, osteoblast-specific factor (*Postn*)[38,39]. A large number of somatic gene enhancers also lose H3K4me2; this change leads to hypermethylation and silencing at later stages. Thus, epigenetic modifications of key MEF identity factors and early pluripotency genes that result in changes in their expression may represent one of the first steps in the conversion of a somatic cell to a pluripotent state.

Chromatin modifiers involved in reprogramming. Although histone marks are robustly modified during reprogramming, it is not clear which chromatin modifiers participate in reshaping the epigenomic landscape of the somatic cells and how they are targeted to genes with an altered expression that is crucial for the conversion. It is reasonable to assume that OSKM binding sites throughout the genome mark regions that will eventually be epigenetically modified. Consistent with this notion is the finding that OCT4 interacts with the WD repeat protein 5 (WDR5), which is a core member of the mammalian Trithorax complex, on pluripotency gene promoters, and this maintains global and localized H3K4me3 distribution[40]. The H3K27 demethylase enzyme UTX physically interacts with OSK (that is, OCT4, SOX2 and KLF4) to remove the repressive mark H3K27me3 from early activated pluripotency genes such as *Fgf4*, *Sall4*, *Sall1* and undifferentiated embryonic cell transcription factor 1 (*Utf1*)[41]. Loss of UTX is associated with aberrant

Table 1 | **Roles of various histone marks during reprogramming**

Histone mark	Function	Phase of reprogramming in which change occurs	Example of change	Refs
H3K4me2	Marks promoters and enhancers	Early phase	Decrease at MEF and EMT genes. Increase at proliferation, metabolism, pluripotency and MET genes	34,36, 38,50
H3K4me3	Marks active loci	Early phase	Increase at proliferation and metabolism genes	34,36,38
H3K27me3	Marks repressed loci	Early phase	Increase at MEF and EMT genes	34,36,38
H3K4me1	Marks enhancers	Early phase	Increase at proliferation and metabolism genes	36
H3K36me3	Marks transcriptionally active regions	Early to middle phase	Increase at early and late pluripotency genes	36
H3K9me3	Marks heterochromatin regions	Late phase	Decrease at late pluripotency genes	50,93
H3K36me2	Marks potential regulatory regions (such as newly transcribed genes)	Early phase	Increase at early pluripotency genes	46,47
H3K79me2	Marks transcriptionally active regions	Early to middle phase	Decrease at MEF and EMT genes	48
H3K27ac	Marks open chromatin and active enhancers			

EMT, epithelial-to-mesenchymal transition; MEF, mouse embryonic fibroblast; MET, mesenchymal-to-epithelial transition; H3K4me2, histone H3 dimethylated at lysine 4; H3K27ac, histone H3 acetylated at lysine 27.

H3K27me3 distribution throughout the genome and with inhibition of reprogramming[41]. TET1 and TET2 — two methylcytosine hydroxylase family members that are important for the early generation of 5-hydroxymethylcytosine (5hmC) during reprogramming — can be recruited by Nanog to enhance the expression of a subset of key reprogramming target genes, such as *Nanog* itself, oestrogen-related receptor beta (*Esrrb*) and *Oct4*. TET1 and TET2 thus appear to be involved in the demethylation and reactivation of genes and regulatory regions that are important for pluripotency[42–44]. The poly(ADP-ribose) polymerase 1 (PARP1) has a complementary role in the establishment of early epigenetic marks during somatic cell reprogramming by regulating 5-methylcytosine (5mC) modification[43]. BRG1 (also known as SMARCA4) and BAF155 (also known as SMARCC1), two components of the BAF chromatin-remodelling complex, enhance reprogramming by establishing a euchromatic chromatin state and enhancing binding of reprogramming factors to key reprogramming gene promoters[45]. Overexpression of BRG1 and BAF155 induces OSKM-mediated demethylation of pluripotency genes such as *Oct4*, *Nanog* and *Rex1* (also known as *Zfp42*) and enhances conversion to iPSCs.

Many other chromatin modifiers have been shown to have a role in resetting the epigenome of reprogrammable cells (summarized in TABLE 2). For example, KDM2A and KDM2B — two H3K36me2 demethylases — cooperate with OCT4 and have roles in facilitating the reprogramming process by regulating H3K36me2 levels at the promoters of early activated genes: mainly epithelial-cell-associated genes, the *miR-302–367* cluster and early pluripotency genes[46,47]. In the conversion of human fibroblasts to iPSCs, the H3K9 methyltransferases EHMT1 and SETDB1 and five components of the Polycomb repressive complexes (PRCs; namely, BMI1 and RING1 from PRC1, and EZH2, EED and SUZ12 from PRC2) are required to reset the epigenome of the somatic cells. Loss of these genes substantially reduces iPSC formation[48].

Another H3K9 methyltransferase, SUV39H, which contributes to heterochromatin formation[49], hinders the reprogramming process. This suggests that loss of SUV39H may have a global effect on chromatin organization that leads to aberrant transcriptional regulation or that H3K9 methyltransferases have different specificities: some target somatic-state-associated genes and others target pluripotency-associated genes. Similarly, the histone H3 lysine 79 (H3K79me2) methyltransferase DOT1L inhibits the reprogramming process in the early to middle phase. Loss of DOT1L increases reprogramming efficiency by facilitating loss of H3K79me2 from fibroblast-associated genes, such as the mesenchymal master regulators snail 1 (*SNAI1*), *SNAI2*, zinc finger E-box-binding homeobox 1 (*ZEB1*) and transforming growth factor beta 2 (*TGFB2*). Silencing of these genes is essential for proper reprogramming and indirectly increases the expression of the pluripotency genes *NANOG* and *LIN28* (REF. 48).

Table 2 | **Roles of example chromatin modifiers in reprogramming**

Chromatin modifier factor	Enzymatic function	Role in reprogramming	Refs
UTX	H3K27 demethylase	Physically interacts with OSK proteins to remove the repressive mark H3K27 from early pluripotency genes	41
KDM2A and KDM2B	H3K36 demethylases	Initiation of the reprogramming process by regulating H3K36me2 levels at the promoters of early-activated genes	46,47
EHMT1 and SETDB1	H3K9 methyltransferases	Required to reset the epigenome of somatic cells	48
BMI1, RING1, EZH2, EED and SUZ12	H3K27 methyltransferases	Involved in maintaining the transcriptional repressive state of genes	48
SUV39H	H3K9 methyltransferase	Contributes to heterochromatin formation, hinders the reprogramming process	48
DOT1L	H3K79 methyltransferase	Inhibits the reprogramming process in the early to middle phase by maintaining the expression of EMT genes such as *SNAI1*, *SNAI2*, *ZEB1* and *TGFB2*	48
PARP1	Chromatin-associated enzyme poly(ADP-ribosyl) transferase, which modifies various nuclear proteins by poly(ADP-ribosyl)ation	Functions in the regulation of 5mC, targets *Nanog* and *Esrrb*	43
SWI/SNF (also known as BAF) complex	Chromatin-remodelling complex	Induces demethylation of pluripotency genes such as *Oct4*, *Nanog* and *Rex1*	45
TET1 and TET2	Methylcytosine dioxygenase that catalyses the conversion of 5mC to 5hmC	Important for the early generation of 5hmC by oxidation of 5mC, target *Nanog*, *Esrrb* and *Oct4* through physical interaction with Nanog	42–44
WDR5 complex	A core member of the mammalian Trithorax complex. An 'effector' of H3K4 methylation	Interacts with OCT4 on pluripotency gene promoters and facilitates their activation	40

5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; EMT, epithelial-to-mesenchymal transition; *Esrrb*, oestrogen-related receptor beta; H3K36me2, histone H3 dimethylated at lysine 36; KDM2A, lysine-specific demethylase 2A; OSK, OCT4, SOX2, KLF4; PARP1, poly(ADP-ribose) polymerase 1; *SNAI1*, snail 1; WDR5, WD repeat protein 5; *TGFB2*, transforming growth factor beta 2; *ZEB1*, zinc finger E-box binding homeobox 1.

It will be interesting to explore whether specific combinations of chromatin modifiers are able to reset the epigenome of a somatic cell and to reprogram it to pluripotency in the absence of pluripotency factors. In addition, these data raise the question of whether the four factors themselves act as pioneer factors that direct conversion by physical interaction with epigenetic and transcriptional regulators.

Roles of the OSKM factors

OSK factors as pioneer factors. Little is known about how ectopic expression of OSKM drives the conversion of somatic cells to the pluripotent state. It has been shown that the first transcriptional wave is mostly mediated by MYC and occurs in all cells, whereas the second wave is more restricted to reprogrammable cells and involves a gradual increase in the expression of OCT4 and SOX2 targets, leading to the activation of other pluripotency genes that aid in the activation of the pluripotency network. KLF4 seems to support both phases by repressing somatic genes during the first phase and facilitating the expression of pluripotency genes in the second phase[17].

In mouse and human fibroblasts, immediately after factor induction, OSKM factors occupy accessible chromatin, binding promoters of genes that are active or repressed[34,36,38,50]. In addition, OSK proteins become associated with distal elements of many genes throughout the genome that display minimal, if any, pre-existing histone modifications or DNase I hypersensitivity[50] (FIG. 2). Thus, the multiple distal genomic sites initially occupied by OSK do not correspond to the distal genomic regions that are bound by these pluripotency factors in ESCs; we will refer to this atypical binding of ectopic OSK in somatic cells as 'promiscuous binding' throughout this article. On the basis of these observations, it has been suggested that OSK factors may act as pioneer factors that open chromatin regions and allow the activation of those genes that are essential for establishment and maintenance of the pluripotent state[50], whereas MYC only facilitates this process (the mode of action by which MYC aids in the conversion is extensively discussed in the next section).

The initial promiscuous binding of OSKM, when expressed in fibroblasts, to target sequences present in many genomic regions raises the question of their molecular role in the conversion of somatic cells to pluripotent cells. Vector-transduction-mediated or doxycycline-induced expression of the reprogramming factors in fibroblasts probably does not mimic the expression mode of the endogenous genes in ESCs, in terms of expression levels and factor stoichiometry. This may result in the widespread and seemingly promiscuous binding of OSKM to multiple regions in the genome, many of which are not occupied by these factors in ESCs. Possibly, OSKM can interact with the Mediator or Cohesin complexes or with RNA Pol II elongation factor ELL3 and initially recruit them to atypical distal enhancers to aid in the opening of these 'closed' regions[51,52]. Mediator bridges interactions between transcription factors at enhancers and the transcription initiation apparatus at core promoters and in combination with RNA Pol II and TATA-binding protein (TBP) may gradually initiate transcription from those 'blocked' regions[51]. Binding of the pioneer factors OSK to 'super enhancers' and the recruitment of the Mediator complex may provide cell type specificity[53] at later stages in the reprogramming process. Supporting the notion that OSKM factors are capable of 'loosening' chromatin and inducing cell plasticity early in reprogramming is the observation that transient expression of the factors is sufficient to open the chromatin and to induce transdifferentiation of fibroblasts to other somatic cells, such as cardiomyocytes and neural progenitor cells[54,55].

Although the four factors often jointly bind to their targets, subsets and different combinations of the factors frequently occupy non-overlapping genomic regions. For example, KLF4 and MYC frequently jointly bind to promoters, whereas all of the other OSKM combinations predominantly occupy distal elements at sites conserved between humans and mice[50]. OSKM factors bind together at gene regions that initiate and support the conversion to pluripotency, such as GLIS family zinc finger 1 (*Glis1*), *mir-302–367* cluster, F box only protein 15 (*Fbxo15*), *Fgf4*, *Sall4* and *Lin28*, and factors that promote MET[14,23,50,56–59]. However, only half of the enhancers that acquire H3K4me2 in the induced cells are shared enhancers with ESCs[36]. The other half represents enhancers that are not ESC-specific, supporting the promiscuous binding of OSKM factors to various genomic regions that aid in the conversation process (FIG. 2). Also, in addition to the four factors, activation of other genes early in the reprogramming process may affect the efficiency and specificity of OSKM binding. Binding of the pioneer factors OSK, in combination with MYC, to enhancer regions that are not ESC-specific results in ectopic gene expression. This may render the initial cells susceptible to other gene expression changes, such as activation of apoptotic genes, metabolic genes and MET-inducing genes, silencing of MEF-specific genes and eventually activation of pluripotency genes[17] (FIG. 2).

Revisiting the function of MYC in reprogramming. Because MYC enhances the transcription of proliferation-associated genes[60–62], its role in cellular reprogramming was initially attributed to its ability to promote proliferation and to activate a set of pluripotency genes and miRNAs. MYC is a basic helix–loop–helix (bHLH) transcription factor that at basal levels interacts with MAX on actively transcribed genes via E box sequences[63]. It has been shown to be dispensable for reprogramming but facilitates the emergence of rare reprogrammed cells[64,65]. Supporting this observation is the finding that MYC does not greatly contribute to the activation of pluripotency regulators in partially reprogrammed cells and that its expression is essential only for the first 5 days[38]. However, in ESCs, MYC augments the transcription elongation of many actively transcribed genes via their core promoter regions and by these means maintains pluripotency[66].

Recently, the role of MYC during transcription has been revisited, and it has been demonstrated that MYC does not regulate a unique set of target genes but

Pioneer factors
A subset of transcription factors that initially accesses silent chromatin and directs the binding of other transcription factors during embryonic development. Pioneer factors (OSK proteins during reprogramming) create a hyperdynamic chromatin state.

Promiscuous binding
In this context, the multiple distal genomic sites initially occupied by OSK proteins that do not correspond to the distal genomic regions that are bound by these pluripotency factors in embryonic stem cells.

Factor stoichiometry
Different levels of and the ratios between reprogramming factors (OSKM) in single cells.

Mediator
A complex comprised of multiple protein subunits that function as a transcriptional co-activator to increase gene expression.

Super enhancers
Expansive regions of DNA that are bound by large amounts of Mediator and other proteins to enhance the transcription of genes.

rather acts as a general amplifier of gene expression, increasing the transcription at all active promoters[67,68]. In contrast to many other transcription factors that activate genes in a binary switch way[69], MYC binding resembles a continuous, analogue process[67]: MYC binding to promoter regions is associated with open chromatin marks, including H3K4me3 and acetylated H3K27 (H3K27ac) and is correlated with the amount of RNA polymerase recruited at those promoters[67,68]. MYC recruits the pause release factor PTEFB, increases

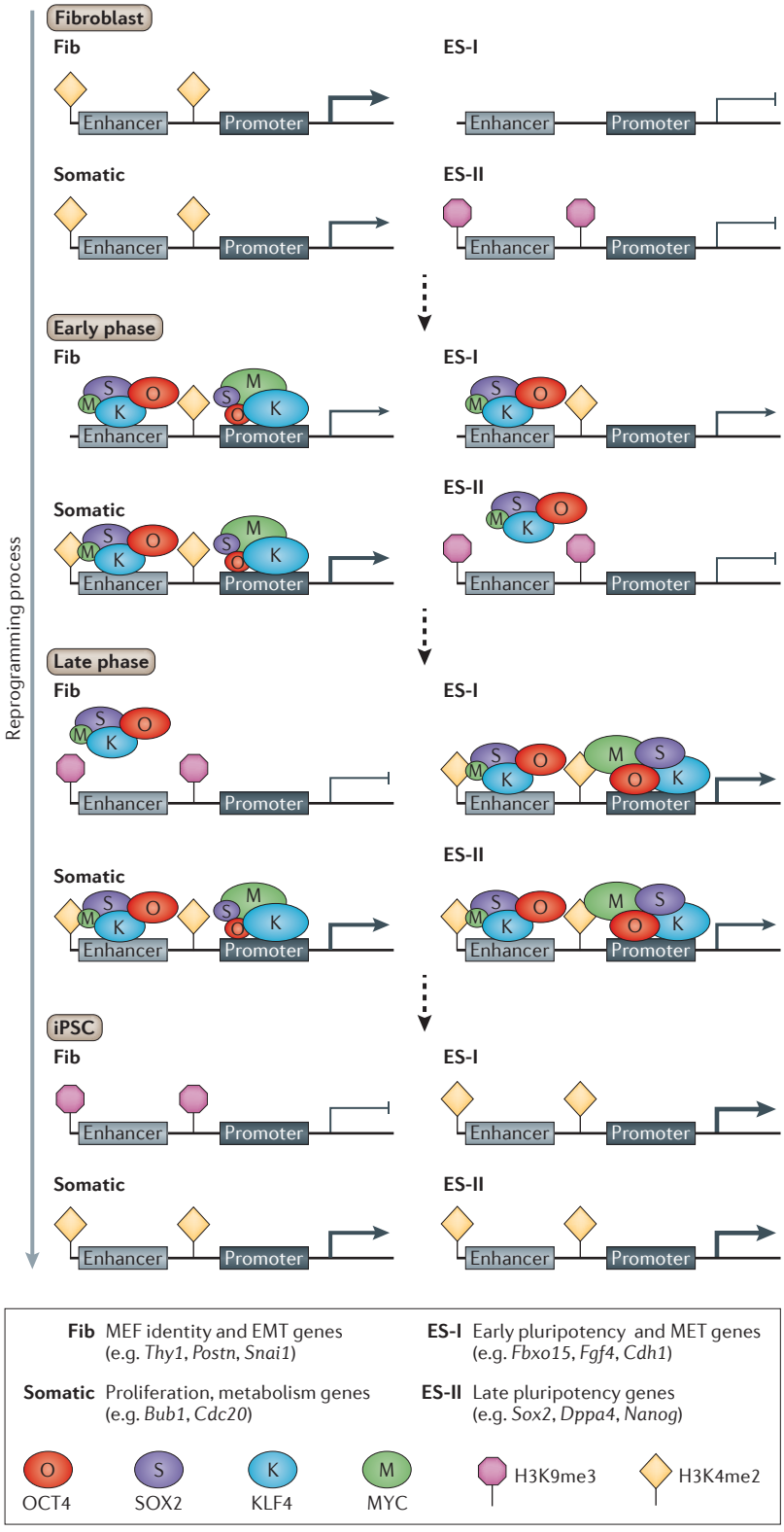

Figure 2 | **OSKM factors as pioneer factors for remodelling the epigenome.** During reprogramming, exogenous OSKM (OCT4, SOX2, KLF4 and MYC) bind enhancers and promoters of fibroblast and embryonic stem cell (ESC) genes along with regions that are not occupied by OSKM in ESCs and that are not specific to fibroblasts (here called 'somatic'). The factors mark the loci that will eventually be epigenetically modified. In general, OSKM factors bind four different classes of genes. The first class (Fib) contains genes such as thymus cell antigen 1 theta (*Thy1*), periostin, osteoblast-specific factor (*Postn*), collagen, type V, alpha 2 (*Col5a2*) that are important for the identity of the fibroblasts and epithelial-to-mesenchymal (EMT) genes such as snail 1 (*Snai1*), *Snai2* and twist basic helix–loop–helix transcription factor 1 (*Twist1*). The second class (somatic) contains genes that are bound by OSKM in somatic cells but not in ESCs and are not specific to fibroblasts. This includes apoptotic genes, such as *Tp53*, genes that are important for proliferative cells, such as cell cycle genes (for example, budding uninhibited by benzimidazoles 1 (*Bub1*), *Cdc20* and *Cdc25c*) and metabolic genes such as phosphofructokinase, liver, B-type (*Pfkl*) and glucose phosphate isomerase (*Gpi*). The third class (ES-I) contains ESC genes such as F box only protein 15 (*Fbxo15*), fibroblast growth factor 4 (*Fgf4*) and *Sall4* that are activated early in the process. The fourth class (ES-II) contains genes such as *Sox2*, *Nanog* and developmental pluripotency associated 4 (*Dppa4*) that are activated late in the reprogramming process. During the early phase of reprogramming, OSKM factors occupy the enhancers of all classes except enhancers of ES-II genes that contain the heterochromatin mark histone H3 trimethylated at lysine 9 (H3K9me3) and are refractory to the four factors. MYC and Krüppel-like factor 4 (KLF4) bind promoters of Fib genes and repress their activity while increasing the activation of genes from the somatic class (shown by the weight of the arrow). As a result, enhancers and promoters from the Fib class start to lose H3K4me2, whereas genes from the somatic class maintain high levels of H3K4me2. OSK proteins act as pioneer factors and occupy the distal enhancer of ES-I genes, which gain *de novo* H3K4me2 marks and will initiate expression a few days later. The late phase is less well understood, but it can be speculated that Fib genes become heterochromatic and are silenced, whereas the genes from the somatic class are highly activated. ES-I genes are highly activated and contain high levels of H3K4me2, and ES-II genes start to lose the H3K9me3 mark, to gain H3K4me2 marks and to initiate expression. It is reasonable to assume that more ES-II class factors that are switched on late in reprogramming are needed to open those blocked regions. After the silencing of the exogenous factors, all groups are highly expressed except Fib, which remains silenced. The sizes of the ovals that represent OSKM indicate their binding preference. For example, MYC is a global amplifier of gene expression increasing the transcription at all active promoters; therefore, the oval 'M' is larger on promoters.

transcriptional elongation and transcription levels[66,70,71], and when overexpressed, its localization to the enhancers of active genes is substantially increased through binding to a variant E box motif. When OSK factors are overexpressed together with MYC, OSK factors act as pioneer factors to enable MYC to bind to regions that are in inaccessible chromatin. In parallel, driven in part by a variant MYC binding site[50], MYC also cooperatively enhances the initial OSK engagement with chromatin. Continuous binding of the factors to those blocked distal elements leads to binding at the promoters of genes that acquire a *de novo* H3K4me2 and eventually leads to the transcription of those genes.

It will be interesting to examine whether in cancer cells other pioneer factors recruit MYC to specific blocked regions through the variant E box motif. Given this notion, MYC expression should enhance any given transdifferentiation or cellular reprogramming process. However, expression of MYC in combination with transcription factors that generate iPSCs but that lack OCT4 (such as SALL4, Nanog, ESRRB and LIN28) only slightly enhanced the reprogramming process[23], suggesting that different key factors have a different affinity for MYC. Future studies should address how different key factors cooperate with this master transcriptional amplifier.

Factor stoichiometry. The number of proviruses in iPSCs widely differs among the individual factors, suggesting that reprogramming requires different expression levels of OSKM[23,31]. Indeed, factor stoichiometry can profoundly influence the epigenetic and biological properties of iPSCs, as was demonstrated by comparing two genetically well-defined doxycycline-inducible transgenic 'reprogrammable' mouse strains[72,73]. The authors showed that, although a high number of iPSC colonies could be obtained, ~95% exhibited aberrant methylation of the delta-like 1 (*Dlk1*)–deiodinase, iodothyronine type III (*Dio3*) locus and were unable to generate mice derived entirely from iPSCs (that is, 'all-iPSC' mice) by tetraploid complementation, which is the most stringent test for pluripotency[73]. By contrast, another study using an almost identical reprogrammable transgenic donor mouse strain showed that most iPSCs had retained normal imprinting at the *Dlk1–Dio3* locus and were competent to generate all-iPSC mice by tetraploid complementation[72]. The only difference between the two transgenic systems was a different stoichiometry of the reprogramming factors: high-quality iPSCs resulted from the donor strain that generated 10- to 20-fold higher levels of OCT4 and KLF4 protein and lower levels of SOX2 and MYC[72] than the donor strain that produced low-quality iPSCs[73]. Consistent with this notion, two other studies concluded that high levels of OCT4 and low levels of SOX2 are preferable for iPSC generation[74,75].

The levels of transgene expression also have a role in the formation of partially reprogrammed iPSCs. It has been shown that partially reprogrammed colonies express a unique set of genes that are often bound by more reprogramming factors in the intermediate state than in ESCs[38] (for example, promoter or enhancer regions that are bound only by OCT4, and SOX2 in ESCs are bound by OSKM in intermediate stage cells). By contrast, genes that are highly expressed in ESCs are bound by fewer reprogramming factors in the partially reprogrammed cells. Promoter regions bound by OSKM in partially reprogrammed cells often contain known DNA-binding sites for the bound factors, indicating that the factors might bind those sites when the factors are present at high levels. These observations are consistent with the notion that excess levels of transgenes or different factor stoichiometry can cause binding of the four factors in a manner that differs from that seen in ESCs. Therefore, the promiscuous binding of OSKM may be influenced by the stoichiometry of the four factors and can either facilitate or block reprogramming.

Other parameters known to affect the characteristics of pluripotent cells are the culture conditions and supplements used to derive the cells[76]. For example, addition of small molecules and supplements such as vitamin C, valproic acid (VPA) and transforming growth factor-β (TGFβ) inhibitors to the medium lead to more efficient derivation of iPSCs[77–80]. More importantly, derivation of iPSCs in the absence of serum and in the presence of vitamin C produced high-quality tetraploid complementation-competent iPSCs even when a suboptimal factor stoichiometry was used for inducing pluripotency[81,82]. In addition, use of physiological oxygen levels during the isolation of human ESCs led to human ESCs with two active X chromosomes, whereas X-chromosome inactivation occurs if conventional conditions are used[83]. Thus, the available evidence suggests that factor stoichiometry as well as specific culture conditions strongly affect the quality and the efficiency of iPSC generation (summarized in TABLE 3).

Markers of reprogramming

Ectopic expression of the reprogramming factors induces a heterogeneous population of cells with individual cells embarking on different fates such as cell death, cell cycle arrest (senescence), uncontrolled proliferation (malignant transformation), transdifferentiation and partial or full reprogramming (FIG. 1). Although it is easy to differentiate between non-reprogrammed and reprogrammed cells, it is more challenging to distinguish partially reprogrammed cells from fully reprogrammed cells. This is because partially reprogrammed cells can be morphologically identical to ESCs and can express many pluripotency genes[23]. Also, owing to the stochastic nature of reprogramming[29], no molecular markers have been identified that would predict whether a given cell early in the process will generate an iPSC daughter. Changes including loss of MEF markers, activation of the MET programme or appearance of markers such as stage-specific embryonic antigen 1 (SSEA1) or alkaline phosphatase must occur in the reprogramming process, but these are not restricted to cells destined to become iPSCs[23,18,59].

To define molecularly the various phases of the reprogramming process, global gene expression and proteomic patterns of clonal cell populations or

Transcriptional amplifiers
Proteins such as MYC that can increase expression from any active promoter.

Table 3 | **Parameters that influence the quality of iPSCs**

Parameter	Reprogramming cocktail or conditions	Effect on the quality of iPSCs	Refs
Stoichiometry	High OCT4, high KLF4, low SOX2, low MYC	Low reprogramming efficiency, normal *Dlk1–Dio3** methylation, no tumours in mice, improved efficiency to produce 4n mice‡	72
	High SOX2, high MYC, low OCT4, low KLF4	High reprogramming efficiency, aberrant methylation of *Dlk1–Dio3*, tumours in mice, low efficiency to produce 4n mice	73
Other factors	TBX3§, ZSCAN4‖	Improve reprogramming efficiency and/or improved efficiency to produce 4n mice	125,126
Culture conditions	Knockout DMEM, 20% KSR	Efficient generation of iPSCs from MEFs and tail tip fibroblasts, improved efficiency to produce 4n mice	127
	Oxygen levels	Hypoxia conditions improve iPSC generation and aid X reactivation	83
Supplement	Vitamin C	Activates *Dlk1–Dio3* locus, improved efficiency to produce 4n mice	82
	Histone deacetylase inhibitor	Activates *Dlk1–Dio3* locus, improved efficiency to produce 4n mice	73
	Dual inhibition of GSK3β and MEK proteins (2i) and LIF	Upregulation of OCT4 and Nanog, competence for somatic and germline chimerism	128
	Protein arginine methyltransferase inhibitor AMI5 and TGFβ inhibitor A-83-01	Improved efficiency to produce 4n mice	129
Genetic and epigenetic background	Not applicable	Unknown	

Dio3, deiodinase, iodothyronine type III; *Dlk1*, delta-like 1; DMEM, Dulbecco's modified eagle medium; GSK3β, glycogen synthase kinase 3β; iPSC, induced pluripotent stem cell; KSR, knockout serum replacement; LIF, leukaemia inhibitory factor; MEF, mouse embryonic fibroblast; MEK, also known as MAP2K; TGFβ, transforming growth factor-β. *Imprinted control domain that contains the paternally expressed imprinted genes *Dlk1*, *Rtl1* and *Dio3* and the maternally expressed imprinted genes *Meg3* (also known as *Gtl2*), *Meg8* (also known as *Rian*) and antisense *Rtl1* (*asRtl1*). This locus is reported to distinguish 'good' iPSCs (those that generate all-iPSC mice and contribute to chimaeras) from 'bad' iPSCs (those that do not generate all-iPSC mice and contribute to chimaeras) in REF. 73. Carey *et al.*[72] found that loss of imprinting at the *Dlk1–Dio3* locus did not strictly correlate with reduced pluripotency. ‡4n mice are mice produced through tetraploid complementation. §TBX3 is a transcriptional repressor involved in developmental processes. ‖ZSCAN4 is a protein involved in telomere maintenance, specifically aiding cell in escaping senescence. It also has a role as a pluripotency factor.

enriched populations were established at different stages after factor induction[15–18]. These analyses suggested that: genes such as *Fbxo15*, *Fgf4*, *Sall1*, fucosyltransferase 9 (*Fut9*), chromodomain helicase DNA binding protein 7 (*Chd7*) and cadherin 1 (*Cdh1*) mark the initiation phase; genes including *Sall4*, *Oct4*, *Nanog*, *Eras*, *Nodal*, *Sox2* and *Esrrb* are activated during the intermediate or maturation phase; and genes such as *Rex1*, growth differentiation factor 3 (*Gdf3*), developmental pluripotency associated 2 (*Dppa2*), *Dppa3* and *Utf1* might define the late or stabilization phase. However, the information from gene expression or proteomic analyses of heterogeneous populations is limited because the rare cells destined to become iPSCs are masked.

Single-cell expression analyses of intermediate SSEA1-positive cells identified early, intermediate and late markers. These included the early epithelial cell adhesion molecule (EPCAM), the intermediate KIT receptor and the late platelet endothelial cell adhesion molecule (PECAM1)[17]. Sorting SSEA1-positive, EPCAM-positive early cells showed modest increase in reprogramming efficiency but could not predict which cells would eventually become fully reprogrammed[17]. Pluripotency genes such as *Utf1*, *Esrrb*, *Lin28* and *Dppa2* were identified as potential 'predictive' indicators that were activated in a small subset of cells and might mark cells early in the process that are destined to become iPSCs[23]. Some of these markers were also detected in the population-based studies but, in contrast to single cell analyses, were detected only at late stages of the process and thus could not identify potential genes for which activation may constitute early markers for cells destined to become iPSCs. The question remains unresolved regarding whether these genes execute a crucial role in the conversion to fully reprogrammed cells or only mark those rare cells.

The endogenous key reprogramming factor genes *Oct4* and *Sall4* are activated early in rare cells but are also activated in partially reprogrammed cells and thus do not represent predictive early markers for iPSC generation[23]; this was confirmed in a study using an inducible *Oct4* lineage label[84]. In agreement with these observations, *Sall4* and endogenous *Oct4* have been found to be poor predictors of reprogramming competency[16].

Predictive early markers
Genes that are activated early in the reprogramming process in rare cells that have a higher probability of activating the *Sox2* locus and to become fully reprogrammed induced pluripotent stem cells.

Models of reprogramming

Somatic stem cells versus differentiated donor cells. Because the generation of cloned animals by SCNT is so inefficient, it was hypothesized that cloned animals such as Dolly the sheep may not have been derived from differentiated cells as assumed but rather from rare somatic stem cells present in the heterogeneous donor cell population[85]. This issue was resolved when mature B and T cells were used as donors to create monoclonal mice that carried in all tissues the immunoglobulin and T cell receptor rearrangements of the B and T cell donors, respectively, thus proving a terminally differentiated donor cell[86]. Similarly, because reprogramming by transcription factors is inefficient, it appeared possible that only a fraction of cells are able to generate iPSCs, which is consistent with an 'elite model' in which only rare somatic stem cells present in the donor population could generate iPSCs, whereas

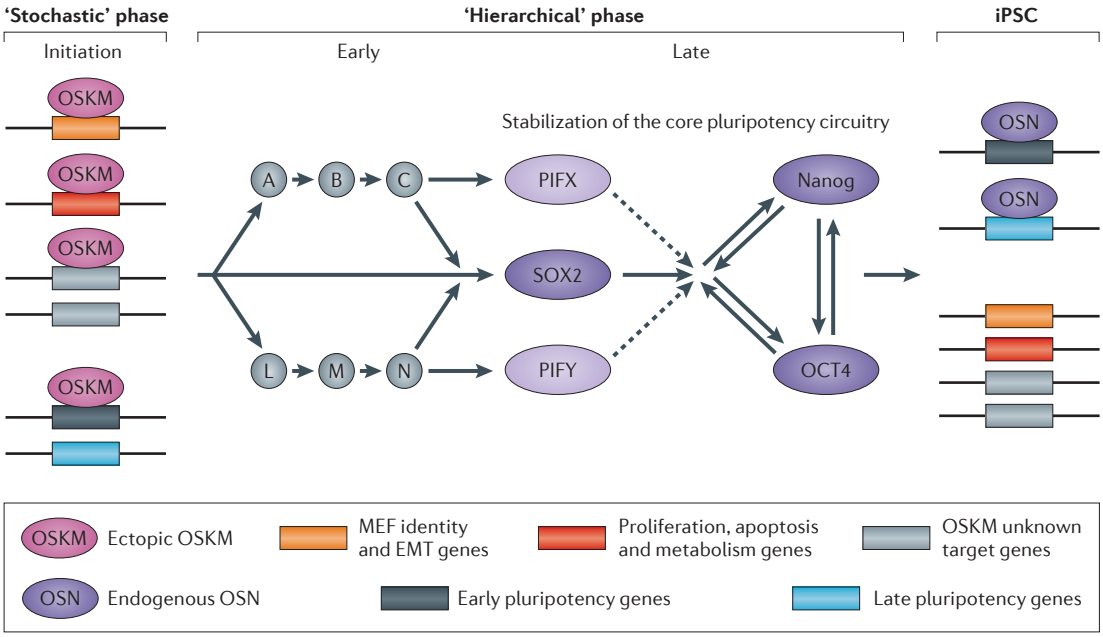

Figure 3 | **Model of molecular events that precede iPSC formation.** In the early phase, ectopic OSKM (OCT4, SOX2, KLF4 and MYC) factors act as pioneer factors and occupy many genomic regions and help to generate a hyperdynamic chromatin state. OSKM factors will bind many regions throughout the genome of the fibroblast that are not OSKM targets in embryonic stem cells (ESCs). Among these regions are: genes that determine the identity of the fibroblast, such as extracellular components and mesenchymal-to-epithelial transition (MET) identity and epithelial-to-mesenchymal transition (EMT) genes (orange box); genes that promote proliferation, apoptosis and increase metabolism (red box); and unknown target genes that facilitate genomic fluidity (that is, a state that allows rapid changes in transcription; light grey box). In addition, OSKM factors will occupy distal regions of early pluripotency genes (dark grey box); this binding will aid in activating those loci at later stages. A group of late pluripotency genes (blue box) is refractory to OSKM binding in this early phase. In the early hierarchical phase (which is more speculative), early pluripotency genes become activated in rare individual cells and will either directly or in a hierarchical manner instigate a more deterministic process that eventually leads to the activation of *Sox2*. *Sox2* represents one gene of a group of late pluripotency initiating factors (PIFs) that are essential for the activation of the core pluripotency circuitry. After they have been activated, the endogenous pluripotency proteins OCT4, SOX2 and Nanog (OSN) occupy their target genes[94] and maintain the induced pluripotent stem cell (iPSC) state in the absence of the exogenous factors.

the differentiated cells would be refractory to reprogramming[87,88]. Several lines of evidence rule out the elite model and argue that all cells, including terminally differentiated cells, have the potential to generate iPSC daughters. First, iPSC colonies have been derived from terminally differentiated cells, such as B cells, T cells, liver and spleen cells[82,89–91]. As with SCNT, specific genomic rearrangement of the immunoglobulin locus or the T cell receptor in iPSC clones unambiguously proved that the cells were indeed derived from mature B or T cells and excluded the possibility of mesenchymal stem cell contamination[90]. Second, clonal analysis of single B cells indicated that >90% have the potential to generate daughter cells that at some point become iPSCs[29].

The stochastic and deterministic modes of reprogramming. In principle, reprogramming of somatic cells could occur by two mechanisms: a stochastic mode, in which iPSCs appear with variable latencies; or a deterministic mode, in which reprogrammed cells would be generated with a fixed latency. In the stochastic model, it cannot be predicted whether or when a given cell would generate an iPSC daughter. Strong support for the stochastic model comes from single-cell cloning experiments demonstrating that sister cells from an early colony generate iPSCs with variable latency and with some sister cells never giving rise to iPSCs[23,92]. Although it cannot be predicted whether or when a given cell will generate an induced pluripotent stem daughter cell, activation of some genes, such as *Esrrb* or *Utf1* (as discussed above), may mark rare early cells that are on their path to iPSCs (FIG. 3). Activation of these genes early in the process suggests that their promoter regions are accessible for OSKM[15–17,23] (FIG. 2). By contrast, late activated loci are marked by H3K9me3 and are refractory to OSKM binding at early stages, and activation of these loci appears to be a crucial step for the proposed transition from a stochastic to a deterministic phase[50,93] (FIGS 1,3). Indeed, several essential pluripotency loci that are marked by H3K9me3, such as *Nanog*, *Dppa4*, *Gdf3* and *Sox2*, are activated later in reprogramming and are refractory to activation by the reprogramming factors during early stages[13,15,16,23,38,50] (FIG. 1,2). Thus, the removal of H3K9me3 may represent another primary epigenetic barrier to complete reprogramming[93].

The key event initiating the late hierarchical phase appears to involve activation of the endogenous *Sox2* gene, which then triggers a series of steps of gene activation that allow the cells to enter the pluripotent state[23] (FIG. 1,3). *Sox2* represents one of a group of pluripotency initiating factors (PIFs) that are crucial and indispensable for the instigation of the deterministic phase[16,23]. The hierarchical network displayed in FIG. 1 predicts that factors other than the canonical Yamanaka factors OCT4, SOX2, KLF4, MYC or Nanog should be able to induce pluripotency. Indeed, downstream factors such as ESRRB, LIN28, DPPA2 and SALL4 were sufficient to induce iPSCs from MEFs[23].

It has been suggested that the initial response to ectopic expression of OSKM in somatic cells may be an orchestrated and possibly deterministic response involving epigenetically definable events that activate loci crucial for pluripotency[17,22]. Here we suggest an alternative view of the initial interaction of OSKM with the genome. As outlined in FIG. 3, initial stochastic gene activation may render the cells susceptible to other gene expression changes (such as activation of apoptotic genes, metabolic genes, MET-inducing genes, silencing of MEF-specific genes and eventually activation of pluripotency genes)[17]. During this initial phase, stochastic OSKM–genome interactions could also instigate the activation of early PIFs, such as *Esrrb* or *Utf1* (REF. 23), in rare cells (FIG. 3), and these would eventually lead to the expression of the late pluripotency genes *Sox2* and *Nanog* and stabilization of the core pluripotency circuitry. At this later stage, the endogenous pluripotency factors (namely, OCT4, SOX2 and Nanog (collectively referred to as OSN proteins)) will, in contrast to the exogenous OSKM factors, occupy only ESC-specific target regions[94].

The initial promiscuous interaction of OSKM with the genome might be initiated by any factor that destabilizes the compacted chromatin typical of somatic cells. It is this destabilization that may render the somatic chromatin susceptible to becoming 'hyperdynamic', which is the hallmark of the ESC epigenetic state[95,96]. Consistent with this notion are the findings that general chromatin-remodelling complexes, such as BAF[45,97], or global basal transcription machinery components such as the transcription factor IID (TFIID) complex[98] or exposure of cells to general DNA methyltransferase and histone deacetylase inhibitors such as 5-azacytidine[13] and valporic acid[78] can substantially enhance reprogramming in cooperation with OSKM. Also, in fibroblasts, down-regulation of the global chromatin organization modulator lamin A, which is not expressed in ESCs[99], has been reported to increase reprogramming efficiency[100]. Thus, although OSKM factors are highly efficient in inducing pluripotency, any chromatin remodeller or transcription factor — even those that do not normally function in ESCs — might be able to initiate the process that leads to pluripotency, albeit with an efficiency that might be too low to be detected in standard reprogramming assays.

It has been suggested that reprogramming by SCNT or by somatic cell–ESC fusion is deterministic, as it leads to activation of the somatic *Oct4* within two cell divisions (in the case of SCNT) or in the absence of DNA replication (in the case of fusion)[1,2]. However, defining pluripotency functionally in cloned embryos or in heterokaryons has been difficult, so it remains to be determined whether these methods activate the pluripotency circuitry by deterministic or stochastic mechanisms. Both types of mechanism might be involved in the various forms of reprogramming.

How similar are ESCs and iPSCs?

Although ESCs and iPSCs are similar in morphology, in the characteristics of age-affected cellular systems (such as telomeres and mitochondria)[101,102] and surface markers, and in the amount of overall gene expression, several studies have identified biological and epigenetic differences between ESCs and iPSCs, as well as among individual ESC and iPSC lines[103–115]. For example, genetic alterations and differences in the transcriptome, proteome and epigenome were detected when ESCs and iPSCs were compared; this led to concerns being raised about the safety of iPSCs for therapeutic applications. However, other studies have failed to find epigenetic and genetic abnormalities that consistently distinguish iPSCs from ESCs[105,116–119]. Rather, these data suggested that the extent of variations seen between ESCs and iPSCs were similar to variations seen within different ESC lines or within different iPSC lines[120].

Recently, it has been suggested that the genetic abnormalities seen in iPSCs might be a result of oncogenic stress induced by the four reprogramming factors[121]. A substantially higher level of phosphorylated histone H2A.X — one of the earliest cellular responses to DNA double-strand breaks (DSBs) — was detected in cells exposed to OSKM or OSK. The authors also linked the homologous recombination pathway (which is essential for error-free repair of DNA DSBs) to the reprogramming process and suggested a direct role for this pathway in maintaining genomic integrity[121]. In summary, the available evidence has not settled whether the alterations seen in iPSCs are the result of the reprogramming process per se or whether they are due to pre-existing genetic and epigenetic differences within individual parental fibroblasts[119,122].

Much evidence indicates that the biological properties, such as *in vitro* differentiation, differ among individual ESC and iPSC lines, raising the concern that the unpredictable variation among cell lines could pose a potentially serious problem for iPSC-based disease research. That is, a subtle phenotype seen between a disease-specific iPSC and a control iPSC line might not be relevant to the disease but may rather reflect a system-immanent difference[123]. Efforts have been directed towards defining experimental conditions of iPSC and ESC derivation that affect the developmental potential of the cells (summarized in TABLE 3).

Perspective

The 2012 Nobel Prize in Physiology and Medicine was awarded to Shinya Yamanaka and John Gurdon for their discoveries on reprogramming somatic cells to pluripotency[124]. The 7 years since Yamanaka's first demonstration of somatic reprogramming using defined

Pluripotency initiating factors
(PIFs). Protein factors that are responsible for triggering the late deterministic phase responsible for transitioning to the pluripotent state.

Hyperdynamic
This term describes a state of dynamic chromatin characterized by hypermobility of chromatin-associated proteins in pluripotent cells.

factors[12] have witnessed much progress in understanding this complex process, and the most straightforward experiments have been done. However, many questions pertaining to the molecular mechanism of reprogramming remain unsolved. For example: how do OSKM factors convert chromatin to a hyperdynamic state; how does the promiscuous binding of OSKM in somatic cells contribute to the reprogramming process; what defines the rate-limiting step; and what are the criteria for and the most effective methods for producing high-quality iPSCs? Addressing these questions will be essential for a deeper understanding of reprogramming and will require the development of new technologies that allow genome-wide epigenetic analyses of single cells.

1. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
2. Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).
3. Pasque, V., Miyamoto, K. & Gurdon, J. B. Efficiencies and mechanisms of nuclear reprogramming. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 189–200 (2010).
4. Vierbuchen, T. & Wernig, M. Molecular roadblocks for cellular reprogramming. *Mol. Cell* **47**, 827–838 (2012).
5. Buganim, Y. & Jaenisch, R. Transdifferentiation by defined factors as a powerful research tool to address basic biological questions. *Cell Cycle* **11**, 4485–4486 (2012).
6. Stadtfeld, M. & Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* **24**, 2239–2263 (2010).
7. Wakayama, S. *et al.* Equivalency of nuclear transfer-derived embryonic stem cells to those derived from fertilized mouse blastocysts. *Stem Cells* **24**, 2023–2033 (2006).
8. Brambrink, T., Hochedlinger, K., Bell, G. & Jaenisch, R. ES cells derived from cloned and fertilized blastocysts are transcriptionally and functionally indistinguishable. *Proc. Natl Acad. Sci. USA* **103**, 933–938 (2006).
9. Zhao, X. Y. *et al.* iPS cells produce viable mice through tetraploid complementation. *Nature* **461**, 86–90 (2009).
10. 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).
11. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
12. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
13. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
14. 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).
15. 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).
This paper describes the various phases (namely, initiation, maturation and stabilization) during the reprogramming process and sheds lights on the role of BMP signalling and the induction of MET during the initiation phase.
16. Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11**, 769–782 (2012).
17. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012).
This paper demonstrates the various pathways and genes involved in early, intermediate and late phases using enriched populations and single-cell approaches. It also shows the various epigenetic modifications that occur at early and late time points and the rescue of partially reprogramed cells by the addition of external factors.
18. Hansson, J. *et al.* Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Rep.* **2**, 1579–1592 (2012).
This paper demonstrates the various pathways and proteins involved in early, intermediate and late phases using enriched populations.
19. Zhang, J., Nuebel, E., Daley, G. Q., Koehler, C. M. & Teitell, M. A. Metabolic regulation in pluripotent stem cells during reprogramming and self-renewal. *Cell Stem Cell* **11**, 589–595 (2012).
20. Sancho-Martinez, I. & Izpisua Belmonte, J. C. Stem cells: surf the waves of reprogramming. *Nature* **493**, 310–311 (2013).
21. Araki, R. *et al.* Conversion of ancestral fibroblasts to induced pluripotent stem cells. *Stem Cells* **28**, 213–220 (2010).
22. 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).
This paper uses single-cell imaging to trace the rare cells that undergo reprogramming based on morphology.
23. Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
Using two complementary single-cell approaches, this study discovered early markers for reprogramming, the late deterministic phase that starts with the activation of *Sox2* and shows iPSC formation without 'Yamanaka' factors.
24. Guo, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18**, 675–685 (2010).
25. Diehn, M. *et al.* Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature* **458**, 780–783 (2009).
26. Citri, A., Pang, Z. P., Sudhof, T. C., Wernig, M. & Malenka, R. C. Comprehensive qPCR profiling of gene expression in single neuronal cells. *Nature Protoc.* **7**, 118–127 (2012).
27. Narsinh, K. H. *et al.* Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells. *J. Clin. Invest.* **121**, 1217–1221 (2011).
28. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5**, 877–879 (2008).
29. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009).
This paper shows that all somatic cells have the potential to generate iPSCs by a process that involves stochastic events.
30. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).
31. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
32. Fussner, E. *et al.* Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.* **30**, 1778–1789 (2011).
33. Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128**, 669–681 (2007).
34. Schmidt, R. & Plath, K. The roles of the reprogramming factors Oct4, Sox2 and Klf4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. *Genome Biol.* **13**, 251 (2012).
35. Liang, G. & Zhang, Y. Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective. *Cell Res.* **23**, 49–69 (2013).
36. Koche, R. P. *et al.* Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8**, 96–105 (2011).
37. Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114–128 (2009).
38. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364–377 (2009).
39. 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).
40. Ang, Y. S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183–197 (2011).
41. Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488**, 409–413 (2012).
42. Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**, 370–374 (2013).
43. Doege, C. A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652–655 (2012).
44. Gao, Y. *et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453–469 (2013).
45. Singhal, N. *et al.* Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* **141**, 943–955 (2010).
46. Liang, G., He, J. & Zhang, Y. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nature Cell Biol.* **14**, 457–466 (2012).
47. Wang, T. *et al.* The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* **9**, 575–587 (2011).
48. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483**, 598–602 (2012).
49. Schotta, G., Ebert, A. & Reuter, G. SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing. *Genetica* **117**, 149–158 (2003).
50. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
This paper raises the notion that OSK proteins are pioneer factors that guide reprogramming through promiscuous binding to distal elements that lack pre-existing histone modifications.
51. Kagey, M. H. *et al.* Mediator and Cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
52. Lin, C., Garruss, A. S., Luo, Z., Guo, F. & Shilatifard, A. The RNA Pol II elongation factor Ell3 marks enhancers in ES cells and primes future gene activation. *Cell* **152**, 144–156 (2012).
53. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
54. Efe, J. A. *et al.* Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nature Cell Biol.* **13**, 215–222 (2011).
55. Kim, J. *et al.* Direct reprogramming of mouse fibroblasts to neural progenitors. *Proc. Natl Acad. Sci. USA* **108**, 7838–7843 (2011).
56. Anokye-Danso, F. *et al.* Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* **8**, 376–388 (2011).
57. Maekawa, M. *et al.* Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature* **474**, 225–229 (2011).
58. Liao, B. *et al.* MicroRNA cluster 302–367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. *J. Biol. Chem.* **286**, 17359–17364 (2011).

59. Subramanyam, D. *et al.* Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nature Biotech.* **29**, 443–448 (2011).
60. Dang, C. V. MYC on the path to cancer. *Cell* **149**, 22–35 (2012).
61. Eilers, M. & Eisenman, R. N. Myc's broad reach. *Genes Dev.* **22**, 2755–2766 (2008).
62. Meyer, N. & Penn, L. Z. Reflecting on 25 years with MYC. *Nature Rev. Cancer* **8**, 976–990 (2008).
63. Blackwood, E. M. & Eisenman, R. N. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* **251**, 1211–1217 (1991).
64. 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).
65. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotech.* **26**, 101–106 (2008).
66. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
67. Nie, Z. *et al.* c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* **151**, 68–79 (2012).
68. Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).
References 67 and 68 revisit the role of MYC during transcription regulation and show that MYC is a global gene amplifier.
69. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
70. Bouchard, C., Marquardt, J., Bras, A., Medema, R. H. & Eilers, M. Myc-induced proliferation and transformation require Akt-mediated phosphorylation of FoxO proteins. *EMBO J.* **23**, 2830–2840 (2004).
71. Eberhardy, S. R. & Farnham, P. J. c-Myc mediates activation of the cad promoter via a post-RNA polymerase II recruitment mechanism. *J. Biol. Chem.* **276**, 48562–48571 (2001).
72. Carey, B. W. *et al.* Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell* **9**, 588–598 (2011).
73. Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175–181 (2010).
74. Tiemann, U. *et al.* Optimal reprogramming factor stoichiometry increases colony numbers and affects molecular characteristics of murine induced pluripotent stem cells. *Cytometry A* **79**, 426–435 (2011).
75. Yamaguchi, S., Hirano, K., Nagata, S. & Tada, T. Sox2 expression effects on direct reprogramming efficiency as determined by alternative somatic cell fate. *Stem Cell Res.* **6**, 177–186 (2011).
76. Chen, J. *et al.* Rational optimization of reprogramming culture conditions for the generation of induced pluripotent stem cells with ultra-high efficiency and fast kinetics. *Cell Res.* **21**, 884–894 (2011).
77. 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).
78. 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).
79. Ichida, J. K. *et al.* A small-molecule inhibitor of TGF-βsignaling replaces SOX2 in reprogramming by inducing nanog. *Cell Stem Cell* **5**, 491–503 (2009).
80. Maherali, N. & Hochedlinger, K. Tgfβ signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr. Biol.* **19**, 1718–1723 (2009).
81. Esteban, M. A. & Pei, D. Vitamin C improves the quality of somatic cell reprogramming. *Nature Genet.* **44**, 366–367 (2012).
82. Stadtfeld, M. *et al.* Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. *Nature Genet.* **44**, 398–405 (2012).
83. Lengner, C. J. *et al.* Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell* **141**, 872–883 (2010).
84. Greder, L. V. *et al.* Brief report: analysis of endogenous Oct4 activation during induced pluripotent stem cell reprogramming using an inducible Oct4 lineage label. *Stem Cells* **30**, 2596–2601 (2012).
85. Pennisi, E. & Williams, N. Will Dolly send in the clones? *Science* **275**, 1415–1416 (1997).
86. Hochedlinger, K. & Jaenisch, R. Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature* **415**, 1035–1038 (2002).
87. Wakao, S., Kitada, M. & Dezawa, M. The elite and stochastic model for iPS cell generation: multilineage-differentiating stress enduring (Muse) cells are readily reprogrammable into iPS cells. *Cytometry A* **83**, 18–26 (2013).
88. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**, 49–52 (2009).
89. Aoi, T. *et al.* Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321**, 699–702 (2008).
90. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250–264 (2008).
91. Seki, T. *et al.* Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. *Cell Stem Cell* **7**, 11–14 (2010).
92. Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nature Biotech.* **25**, 1177–1181 (2007).
93. Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature Genet.* **45**, 34–42 (2012).
94. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
95. Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116 (2006).
96. Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642–654 (2013).
97. Takeuchi, J. K. & Bruneau, B. G. Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* **459**, 708–711 (2009).
This paper describes the fluidity of chromatin in ESCs.
98. Pijnappel, W. W. *et al.* A central role for TFIID in the pluripotent transcription circuitry. *Nature* **495**, 516–519 (2013).
99. Mattout, A., Biran, A. & Meshorer, E. Global epigenetic changes during somatic cell reprogramming to iPS cells. *J. Mol. Cell. Biol.* **3**, 341–350 (2011).
100. Zuo, B. *et al.* Influences of lamin A levels on induction of pluripotent stem cells. *Biol. Open* **1**, 1118–1127 (2012).
101. Suhr, S. T. *et al.* Mitochondrial rejuvenation after induced pluripotency. *PLoS ONE* **5**, e14095 (2010).
102. Van Haute, L., Spits, C., Geens, M., Seneca, S. & Sermon, K. Human embryonic stem cells commonly display large mitochondrial DNA deletions. *Nature Biotech.* **31**, 20–23 (2013).
103. Hussein, S. M. *et al.* Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–62 (2011).
104. Laurent, L. C. *et al.* Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* **8**, 106–118 (2011).
105. Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63–67 (2011).
106. Mayshar, Y. *et al.* Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7**, 521–531 (2010).
107. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68–73 (2011).
108. 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).
109. Ohi, Y. *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature Cell Biol.* **13**, 541–549 (2011).
110. Bar-Nur, O., Russ, H. A., Efrat, S. & Benvenisty, N. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. *Cell Stem Cell* **9**, 17–23 (2011).
111. 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).
112. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285–290 (2010).
113. Kim, K. *et al.* Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nature Biotech.* **29**, 1117–1119 (2011).
114. 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).
115. Phanstiel, D. H. *et al.* Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nature Methods* **8**, 821–827 (2011).
116. Newman, A. M. & Cooper, J. B. Lab-specific gene expression signatures in pluripotent stem cells. *Cell Stem Cell* **7**, 258–262 (2010).
117. 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).
118. 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).
119. Cheng, L. *et al.* Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell* **10**, 337–344 (2012).
120. Vitale, A. M. *et al.* Variability in the generation of induced pluripotent stem cells: importance for disease modeling. *Stem Cells Transl. Med.* **1**, 641–650 (2012).
121. Gonzalez, F. *et al.* Homologous recombination DNA repair genes play a critical role in reprogramming to a pluripotent state. *Cell Rep.* **3**, 651–660 (2013).
122. Abyzov, A. *et al.* Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492**, 438–442 (2012).
123. Soldner, F. & Jaenisch, R. Medicine. iPSC disease modeling. *Science* **338**, 1155–1156 (2012).
124. Jaenisch, R. Nuclear cloning and direct reprogramming: the long and the short path to stockholm. *Cell Stem Cell* **11**, 744–747 (2012).
125. Han, J. *et al.* Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature* **463**, 1096–1100 (2010).
126. Jiang, J. *et al.* Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPS cells as demonstrated by tetraploid complementation. *Cell Res.* **23**, 92–106 (2012).
127. Zhao, X. Y., Lv, Z., Li, W., Zeng, F. & Zhou, Q. Production of mice using iPS cells and tetraploid complementation. *Nature Protoc.* **5**, 963–971 (2010).
128. Silva, J. *et al.* Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6**, e253 (2008).
129. Yuan, X. *et al.* Brief report: combined chemical treatment enables Oct4-induced reprogramming from mouse embryonic fibroblasts. *Stem Cells* **29**, 549–553 (2011).

Acknowledgements
We thank M. Dawlaty, A. Soufi and K. Zaret for insightful comments on the manuscript. Y.B. is supported by a US National Institutes of Health (NIH) Kirschstein National Research Service Award (1 F32 GM099153-01A1). D.A.F. is a Vertex Scholar and was supported by a US National Science Foundation Graduate Research Fellowship and Jerome and Florence Brill Graduate Student Fellowship. R.J. is supported by US NIH grants R37-CA084198 and RO1-CA087869.

Competing interests statement
The authors declare competing financial interests: see Web version for details.