Transcriptional and epigenetic mechanisms of cellular reprogramming to induced pluripotency

Enforced ectopic expression of a cocktail of pluripotency-associated genes such as *Oct4*, *Sox2*, *Klf4* and *c-Myc* can reprogram somatic cells into induced pluripotent stem cells (iPSCs). The remarkable proliferation ability of iPSCs and their aptitude to redifferentiate into any cell lineage makes these cells a promising tool for generating a variety of human tissue *in vitro*. Yet, pluripotency induction is an inefficient process, as cells undergoing reprogramming need to overcome developmentally imposed epigenetic barriers. Recent work has shed new light on the molecular mechanisms that drive the reprogramming of somatic cells to iPSCs. Here, we present current knowledge on the transcriptional and epigenetic regulation of pluripotency induction and discuss how variability in epigenetic states impacts iPSCs' inherent biological properties.

First draft submitted: 17 March 2016; Accepted for publication: 20 May 2016; Published online: 15 July 2016

Keywords: chromatin • DNA methylation • epigenetics • histone modifications • induced pluripotent stem cell • iPSC • pluripotency • reprogramming

The transition from stem cells to mature, functionally specialized cells has long been considered a finite phenomenon. Evidence for the reversibility of cell fate commitment came from observations that an adult somatic cell could be returned to an immature state upon transplantation of its nucleus to an enucleated oocyte (somatic cell nuclear transfer) [1–3] or by fusion with an embryonic stem cell (ESC) [4,5]. In 2006, innovative work by Yamanaka and coworkers showed that forced ectopic expression of the pluripotency-associated transcription factors *Oct4*, *Sox2*, *Klf4* and *c-Myc* (collectively termed OSKM) could reprogram differentiated cells, such as skin fibroblasts, to an ESC-like state, yielding induced pluripotent stem cells (iPSCs) [6]. The technology of induced pluripotency has since received considerable interest in the scientific community because it holds great promise for *in vitro* human disease modeling, drug screening and cell-based therapies in regenerative medicine.

iPSCs are very similar to ESCs in their morphology, proliferative capacity and developmental potential, and both pluripotent cell types exhibit highly comparable global gene expression and epigenetic profiles [7–10]. The chromatin landscapes and DNA methylomes of iPSCs are, however, markedly different from those of lineage-committed cells, indicating that reprogramming requires widespread remodeling of somatic cell epigenetic profiles to reflect the ESC epigenome [11–14]. In generating iPSCs from somatic cells, epigenome-modifying molecules may be used to improve the efficiency of the overall process and even limit the number of exogenous pluripotency-inducing factors [15–18], confirming the general notion that epigenetics plays a key role in the regulation of cell fate specification.

Somatic cell reprogramming is a relatively slow and inefficient process, with only a minority of transduced somatic cells becoming fully reprogrammed to *bona fide* iPSCs after several weeks [19–21]. Observations that

Mark van den Hurk^{1,2,3},
Gunter Kenis^{*,1,2}, Cedric
Bardy^{*,3}, Daniel L van den
Hove^{1,2,4}, Fred H Gage³,
Harry W Steinbusch^{*,1,2}
& Bart P Rutten^{1,2}

¹Department of Psychiatry
& Neuropsychology, Division of
Translational Neuroscience, Maastricht
University, Maastricht, 6200 MD,
The Netherlands

²European Graduate School of
Neuroscience (EURON), Maastricht
University, Maastricht, 6200 MD,
The Netherlands

³Laboratory of Genetics, The Salk
Institute for Biological Studies, La Jolla,
CA 92037, USA

⁴Laboratory of Translational
Neuroscience, Department of Psychiatry,
Psychosomatics & Psychotherapy,
University of Wuerzburg,
97080 Wuerzburg, Germany

*Author for correspondence:

Tel.: +31 43 388 2004

Fax: +31 43 388 4086

h.steinbusch@maastrichtuniversity.nl

*Authors contributed equally

Future
Medicine

part of fsg

stem and progenitor cells reprogram with higher efficiency and kinetics than terminally differentiated cells [22–24] suggest that epigenetic barriers established during embryonic differentiation hinder efficient reprogramming to the pluripotent state (for excellent reviews, see [25–27]). Somatic cell types that are developmentally closer to ESCs supposedly require less epigenetic remodeling, potentially facilitating their reprogramming into iPSCs.

Despite major advances in the methods for deriving and culturing iPSCs, the precise molecular mechanisms that drive cells to overcome developmentally imposed epigenetic barriers are only beginning to be elucidated. Most of our current information about the transcriptional and epigenetic events regulating pluripotency and reprogramming has come from studies using murine cells. Yet, strong cross-species conservation of fundamental genetic and epigenetic mechanisms controlling stem cell self-renewal and differentiation has enabled the translation of numerous experimental procedures and insights from mouse to human (Box 1). In this review, we summarize the current knowledge of the transcriptional and epigenetic regulation of pluripotency induction, and discuss the sources and functional biological consequences of epigenetic variability in iPSCs. Though this review mainly focuses on murine somatic cell reprogramming, a greater understanding of the molecular events governing pluripotency induction in mouse provides important insights to improve human cell reprogramming methods and guide safe and large-scale iPSC production for therapeutic use in human [28].

Molecular & transcriptional routes to induced pluripotency

Reprogramming of somatic cells into iPSCs requires silencing of the somatic cell transcriptional profile and the induction of an ESC-like expression program. This change in global gene expression has been demonstrated to occur gradually and through a defined sequence of molecular events (Figure 1). Upon reprogramming factor expression, mouse fibroblasts initially suppress somatic cell-specific genes such as the fibroblast marker *Thy1* before upregulating the early embryonic markers alkaline phosphatase and SSEA-1, and ultimately activating core pluripotency factors such as *Nanog*, *Sall4*, *Oct4* and *Sox2* [43–45]. Fluorescence-activated cell sorting on the basis of those markers enables the isolation of distinct intermediate populations of iPSC generation from the highly heterogeneous reprogramming culture, thereby facilitating the study of the molecular mechanisms governing successful pluripotency induction [43–45]. In this way, Polo *et al.* showed that, compared with SSEA-1+ cells, which over time progressed

toward a *bona fide* pluripotent state (*Oct4*-GFP), cells that failed to downregulate *Thy1* shortly after OSKM induction became refractory to the reprogramming process [45]. Gene expression profiling by microarray analysis further demonstrated that refractory *Thy1*+ cells, in contrast to progressing SSEA-1+ intermediates, did not undergo proper mesenchymal-to-epithelial transition (MET) [45], a developmental process by which cells acquire epithelial characteristics and lose mesenchymal features. MET represents a key initiating event during iPSC derivation that depends on intrinsic BMP signaling, and coincides with increased proliferation and cell size reduction (Figure 1) [9,46–48].

Following the establishment of MET during the initiation phase of mouse pluripotency induction, cells gradually transit into a maturation phase, which is characterized by the activation of a subset of pluripotency-associated genes, including *Nanog*, *Sall4*, *Esrrb* and endogenous *Oct4* [47,49]. Yet, induction of the complete pluripotency transcriptional network with expression of *Utf1*, *Lin28*, *Dnmt3l*, *Dppa2/3/4*, *Pecam* and endogenous *Sox2*, does not occur until the later stabilization phase [47,49]. Studies that used doxycycline (Dox)-inducible expression systems to delineate the temporal order of molecular events in mouse iPSC formation showed that the initiation phase was not self-sustainable and that transition to the intermediate maturation phase was dependent upon continuous expression of the OSKM transgenes [43,44,47]. However, sustained expression of reprogramming factors suppressed the upregulation of stabilization-phase pluripotency markers late in maturation, indicating that transgene silencing was required for cells to transition from maturation to stabilization and become fully pluripotent [49]. Through comparative gene expression profiling of clonally cultured cells that had gained competency to progress to the stabilization phase versus those that had not, Golipour *et al.* identified a distinctive transcriptomic signature associated with successful acquisition of the pluripotent state [49]. Notably, genes governing maturation-to-stabilization transition were distinct from those regulating pluripotency, and bioinformatics analyses suggested their involvement in signaling pathways, cell cycle, cytoskeletal reorganization, chromosome segregation and chromosome stability. Hence, it has been suggested that late-phase progression to transgene-independent *bona fide* iPSCs and maintenance of the pluripotent state are under the control of distinct molecular regulatory pathways [49].

Paralleling the occurrence of the initiation, maturation and stabilization phases, during reprogramming, cells destined to become iPSCs showed two major waves of gene activation that were separated by an intermediary period of less profound transcriptional

Box 1. Conservation and divergence in human and murine (induced) pluripotency.

- Mammalian pluripotency is conferred by a unique and highly conserved network of pluripotency transcription factors, of which Oct4, Sox2 and Nanog constitute key regulators [29–31]. Comparisons of mouse and human ESCs have, however, revealed important interspecies differences in the target genes controlled by these pluripotency regulators [30] and specific molecular signaling pathways activated [32]. For instance, while mouse ESCs require LIF-Stat3 signaling for self-renewal and maintenance of pluripotency, human ESCs are insensitive to LIF and show elevated expression of SOCS-1, an inhibitor of STAT3 signaling [32,33]. Despite these differences, and differences in cell culture requirements, expression of cell-surface antigens (mouse: SSEA-1; human: SSEA-3, SSEA-4, TRA-1-60 and TRA-1-81 [34]) and developmental potential (e.g., the inability of mouse ESCs to differentiate to trophoblasts [35]), there is also a substantial overlap in gene expression and pathway activation between both species [32]. The high evolutionary conservation of core pluripotency transcriptional and epigenetic mechanisms has thus enabled many insights from studies conducted in mice to be translated to the human situation. Ectopic expression of the same set of pluripotency-associated transcription factors (Oct4, Sox2, Klf4 and c-Myc), for example, induces pluripotency in somatic cells of mouse and human origin [6,36–38]. Likewise, a highly conserved miRNA cluster (miR-302/367) can efficiently reprogram mouse and human somatic cells to iPSCs, even in the complete absence of exogenous pluripotent factors [39]. The miR-302/367 cluster is specifically expressed in human and mouse ESCs [40], and has been identified as a direct target of the Oct4 and Sox2 pluripotency transcription factors [41], thus providing evidence for a conserved function of this specific miRNA cluster in the regulation and maintenance of the undifferentiated stem cell state. All in all, we can conclude that core members of the pluripotency regulatory network appear to be well conserved between mice and humans, enabling us to use the murine system to study human cell reprogramming mechanisms. Their downstream targets and signaling pathways, however, seem more evolutionarily divergent, which suggests the need for caution in using mouse pluripotent stem cells as a model for the differentiation of human pluripotent stem cells in the context of therapeutic applications [42]

ESC: Embryonic stem cell.

alterations (Figure 1, pink shade) [45]. The first transcriptional wave was initiated in the majority of cells and occurred between days 0 and 3, reflecting a common early response to ectopic OSKM expression. The second wave was activated late in reprogramming (after day 9) and only in cell populations progressing to *bona fide* pluripotency. During the first wave, genes related to proliferation, metabolism and cytoskeletal organization showed activation, whereas developmental genes were downregulated [45]. This finding is consistent with observations that most murine fibroblasts repress the somatic cell transcriptional program immediately following reprogramming factor transduction, increase their proliferation rate and reduce their size [9,48]. Of the OSKM reprogramming factors, c-Myc has been implicated as a major inducer of these initial reprogramming events, in part because of its potentiating effects on cell cycle progression [45,50]. c-Myc targets, including the cell cycle regulatory genes *Ccnb1*, *Cdk1* and *Aurka*, are robustly upregulated within the first few days after OSKM induction and only modestly change expression until the end of reprogramming (day 12; see also Figure 1) [45]. The combined activity of Oct4 and Sox2 seems to result in the gradual upregulation of core pluripotency genes such as *Nanog* that hallmark the second transcriptional wave and reinforce the transcriptional program to establish a stable pluripotent state [45]. Klf4, the other reprogramming factor, presumably serves a binary sup-

porting role in iPSC generation by repressing somatic genes and activating pluripotency genes during the first and second transcriptional waves, respectively [45].

A shortcoming of the Thy1/SSEA-1/Oct4-GFP route map to study the molecular mechanisms of iPSC generation is that the SSEA-1 surface antigen is a poor predictor of complete reprogramming. O'Malley *et al.* found that Nanog-eGFP expression emerged in both SSEA-1+ and SSEA-1- cell populations after 8 days of Dox-induced reprogramming, and SSEA-1 expression was highly heterogeneous in established iPSCs [51]. The authors identified a new murine fibroblast-to-iPSC reprogramming route map contingent on the loss of CD44 expression and a subsequent rapid upregulation of ICAM1, events that closely paralleled the gain of Nanog-eGFP. ICAM1 was expressed, though to variable extent, in fibroblasts and early reprogramming intermediates, but most cells became ICAM1- by day 6 of Dox-induced OSKM expression. 2 days later, CD44 was downregulated in the vast majority of cells and, toward the end of reprogramming (day 12), the cells appeared enriched for an ESC-/iPSC-like phenotype (CD44-/ICAM1+/Nanog-eGFP+). Nanog-eGFP was, however, already expressed in some cells as early as day 6, prior to the disappearance of CD44 and reupregulation of ICAM1. Yet, Nanog-eGFP+ cells had a higher probability of transitioning to fully reprogrammed iPSCs than Nanog-eGFP-cells [51], underscoring the important role of Nanog in driving pluripotency acquisition.

Figure 1. Cellular, transcriptional and epigenetic changes during mouse somatic cell reprogramming to induced pluripotency. Markers upregulated before and during the different phases of reprogramming are shown at the top. The two major waves of gene activity that occur during pluripotency induction are depicted in pink. Gray intensities of the bars denote the approximate magnitude of the indicated process over the time course of reprogramming.
AP: Alkaline phosphatase; iPSC: Induced pluripotent stem cell; MET: Mesenchymal-to-epithelial transition; OSKM: Oct4, Sox2, Klf4, c-Myc.

RNA-sequencing analysis of reprogramming intermediates purified by CD44/ICAM1/Nanog-eGFP fluorescence-activated cell sorting status subsequently indicated that pluripotency gene activation occurred in two transcriptional waves. Several pluripotency genes, including endogenous *Oct4* and *Sall4*, already showed marked upregulation very early in the reprogramming process [51]. Other pluripotency gene markers, such as endogenous *Sox2*, *Esrrb* and *Dppa2*, demonstrated more gradual activation at later reprogramming stages. This finding was corroborated by an earlier study that used single-cell expression profiling to characterize

various stages and transitions on the road to pluripotency [52]. Notably, in that study, expression of endogenous *Oct4* occurred very early in the reprogramming process and could be detected in some partially reprogrammed colonies, suggesting that *Oct4* activation is a poor predictor of successful pluripotency induction. By contrast, endogenous *Sox2* expression appeared late, occurred in intermediate cells to a very low extent and was predicted as an upstream regulator of activation of many genes involved in pluripotency [52].
In addition to the two waves of pluripotency gene expression, O'Malley *et al.* identified a group of

epidermis-related genes that were transiently upregulated in intermediate reprogramming stages [51]. Interestingly, as determined by single-cell PCR, upregulation of these epidermis genes concurred with the first wave of early pluripotency gene activation, whereas their downregulation paralleled the activation of the late-stage pluripotency genes of the second wave. Given the occurrence of a similar transient epidermal gene expression pattern in three previously published datasets [45,47,50], transient ectodermal gene activation has been pointed out as a common hallmark of reprogramming [51]. Future research aimed at thoroughly investigating this phenomenon of temporary epidermal gene expression may provide important novel mechanistic insights into the molecular mechanisms of cellular reprogramming.

Epigenetic control of cellular reprogramming

The inefficiency and slow kinetics by which pluripotency is established in somatic cells after ectopically enforced OSKM expression suggests a critical role for additional regulators, including epigenetic modulators, in the induction and/or maintenance of the pluripotent state. Close interrelationships are known to exist between the transcriptional regulatory circuitry controlling pluripotency and modifiers of chromatin structure [53,54]. The pluripotency factors Oct4, Sox2 and Nanog have been found to physically associate with, as well as regulate the expression of, various regulators of chromatin state [29,55–57]. Furthermore, in iPSC generation from mouse and human somatic cells, exogenously applied epigenome-modifying molecules can be used to functionally replace or complement one or more of the OSKM reprogramming factors [15–18]. In this section, we discuss in detail the various epigenetic regulators and their associated marks involved in OSKM-induced iPSC generation.

Histone modifications in reprogramming

The global chromatin state of pluripotent and differentiated cells is markedly different. Unlike differentiated cells, which are characterized by highly condensed domains of heterochromatin and the accumulation of repressive histone modifications including methylation of H3K9 and H3K27 [11,58], pluripotent stem cells are distinguished by globally decompacted, euchromatic regions of chromatin associated with activating histone marks such as methylation on H3K4 and hyperacetylation of H3/H4ac [59,60]. In ESCs, architectural chromatin proteins such as HP1 bind loosely and in a hyperdynamic fashion to chromatin whereas immobilization of such proteins occurs upon lineage commitment [60]. Chromatin reorganization during differentiation serves to silence lineage-inappropriate

genes and thus constitutes an important mechanism by which cells establish and maintain a stable cell fate-specific gene expression program [61]. It follows that transcription factor-induced reprogramming requires global remodeling of the somatic cell chromatin back to a more dispersed conformation typical of ESCs, a process associated with substantial alterations in histone modification patterns and DNA methylation status [11–14,62].

OSKM-resistant chromatin

Understanding at the molecular level exactly how the reprogramming factors change the somatic cell epigenome to reset cell identity represents an important goal for the reprogramming field. In iPSC generation, a major challenge for the reprogramming factors is to gain access to and reactivate the endogenous pluripotency genes buried in densely compacted heterochromatin. A study investigating global gene expression and epigenetic changes early in mouse somatic reprogramming identified acquisition of euchromatic H3K4 dimethylation (H3K4me2) at numerous pluripotency-associated loci as an immediate primary response to ectopic reprogramming factor induction (see also [Figure 1](#)) [13]. Notably, gain of H3K4 methylation at these sites did not lead to changes in gene transcription. In fact, genes that were upregulated early in reprogramming generally exhibited activating H3K4 trimethylation (H3K4me3) at their promoters in fibroblasts (i.e., prior to reprogramming factor induction) [13], suggesting that somatic cell chromatin status may limit reprogramming factor action to readily accessible promoter DNA. Supporting this idea, Soufi *et al.* identified megabase-long chromatin domains in human fibroblasts, called ‘differentially bound regions’ that blocked initial access of the OSKM reprogramming factors to the genome [63]. These OSKM refractory regions contained many genes necessary for establishing pluripotency, including *SOX2* and *NANOG*, and were specifically enriched for heterochromatic H3K9 trimethylation (H3K9me3). Transient knockdown of the H3K9 methyltransferases *SUV39H1*, *SUV39H2* or *SETDB1* by small interfering RNA technology enabled Oct4 and Sox2 reprogramming factors to access the differentially bound regions and resulted in more efficient and accelerated pluripotency induction [63]. Consistent with the notion that H3K9 methylation constitutes a critical epigenetic impediment to cellular reprogramming ([Figure 2A](#)), Chen *et al.* demonstrated that its specific removal could fully convert mouse partially reprogrammed cells to *bona fide* iPSCs expressing core pluripotency markers [64]. Specifically, knockdown of the H3K9 methyltransferase *Setdb1* or overexpression of the H3K9 demethylase *Kdm4b* promoted the

Figure 2. Roles of histone-modifying proteins and DNA methylation in pluripotency and reprogramming. (A) In somatic cells, pluripotency gene promoters are methylated and enriched for H3K9me3, a heterochromatin-associated histone mark that blocks access of the OSM factors to the DNA, hindering reprogramming. Reprogramming of somatic cells to iPSCs is associated with DNA demethylation and acquisition of H3K4me3 at promoters of pluripotency genes. The establishment of this permissive H3K4me3 chromatin state is brought about by Trithorax group (trxG) proteins, whose core component Wdr5 was found in ESCs to be directed to pluripotency loci by Oct4. The occupancy of Oct4 (and other pluripotency factors) with trxG drives strong activation of the pluripotency transcriptional network. In apparent contrast, somatic genes become silenced during reprogramming due to *de novo* promoter DNA methylation and the action of PRC2 that imposes a transcriptionally silent state of chromatin characterized by H3K27me3 modification. (B) Bivalent H3K4me3/H3K27me3 chromatin domains at promoter regions of developmental genes constitute a hallmark of ESC and iPSC pluripotency. The bivalent chromatin-associated protein Utf1, a target of Oct4 and Sox2, enforces a 'poised' state of gene expression in ESCs by preventing excessive PRC2 binding and H3K27me3, and facilitating the loading of Dcp1a to mRNAs generated from leaky transcription for cytoplasmic degradation. These contrasting functions of Utf1 prevent both the oversilencing and insufficient repression of bivalent genes, thereby properly coordinating developmental processes. 5mC: 5-methylcytosine; C: (Nonmethylated) Cytosine; ESC: Embryonic stem cell; iPSC: Induced pluripotent stem cell.

further reprogramming of pre-iPSC intermediates into Oct4-GFP-expressing colonies under reprogramming-inhibiting culture conditions. The authors also found that vitamin C-mediated enhancement of iPSC induction was profoundly inhibited upon treatment with small interfering RNAs targeting various members of the Kdm family of H3K9 demethylases [64]. These data indicate that specific constituents of the cell culture environment, including the small-molecule supplement vitamin C, can affect chromatin structure by regulating H3K9 methylation levels and thereby influence the efficiency and kinetics of somatic cell reprogramming.

Immunofluorescence analysis of mouse embryonic fibroblasts that underwent reprogramming revealed that the redistribution of the heterochromatin-associated histone mark H3K9me3 and its binding protein HP1, from confined foci with well-defined boundaries to a more diffuse organization with more prominent nucleoplasmic staining, is a rather early event during iPSC generation [14]. This global change in heterochromatin organization could already be observed in small colonies of OSKM-transduced cells by day 6 of the reprogramming process, prior to the expression of *Nanog* and the establishment of euchromatic chromatin features such as histone acetylation and H3K4 trimethylation (Figure 1). Euchromatic histone marking paralleled the expression of *Nanog*; hence establishment of an open chromatin configuration typical of ESCs does not seem to occur until relatively late in reprogramming [14]. Supporting this notion, conversion of mouse partial iPSCs into full iPSCs by dual chemical inhibition of MEK and GSK3 signaling (known as ‘2i’ culture [65]) resulted in the decondensation of highly compacted heterochromatin into dispersed, open 10-nm fibers concurrently with transgene silencing and pluripotent gene activation [62], both of which are late events in cellular reprogramming.

Role of Polycomb & Trithorax group proteins

The Polycomb group (PcG) and Trithorax group (trxG) protein complexes, which catalyze methylation of different lysines of histone H3, play an important role in controlling pluripotency-related gene expression. Whereas PcG proteins assist in chromatin compaction and thus gene silencing by mediating repressive trimethylation of H3 at lysine 27 (H3K27me3), trxG proteins favor a transcriptionally permissive chromatin structure through establishment of H3K4 trimethylation (H3K4me3) marks at gene promoters [66].

In ESCs, genes required for pluripotency and self-renewal, such as *Oct4*, *Sox2* and *Nanog*, are actively transcribed, and their promoters are typically associated with activating H3K4me3 histone marks [67–70].

Wdr5, a key component of the trxG protein complex in mammals, co-occupies numerous pluripotency genes together with Oct4 in mouse ESCs to positively regulate their expression [71]. Reduction in H3K4me3 levels in these ESCs as a result of short-hairpin RNA-induced knockdown of *Wdr5* causes downregulation of pluripotency-associated genes including *Oct4*, *Sox2*, *Klf4*, *Nanog* and *Esrrb*, and induced differentiation [71]. Interestingly, *Wdr5* expression is upregulated during iPSC generation and loss of *Wdr5* function markedly reduces reprogramming efficiency [71]. These findings highlight the importance of H3K4me3 marking in controlling pluripotent cell identity and suggest a requirement for trxG activity by the OSKM reprogramming factors to reset the epigenome during iPSC generation (Figure 2A).

In addition to the trxG proteins, PcG protein complexes play an indispensable role in specifying and maintaining pluripotent cell identity, as no ESC lines can be established from mouse blastocysts lacking the PRC2 subunit *Ezh2* [72] and ESCs deficient in the PRC2 component *Eed* differentiate spontaneously [73]. PcG proteins exert their preventive effect on ESC differentiation by H3K27me3-mediated silencing of key developmental regulators [73,74]. Accordingly, ESCs lacking the PRC2 component *Suz12* have globally reduced H3K27me3 levels and show increased expression of differentiation-specific markers [75]. PRC2 complex function is critical for transcription factor-induced reprogramming, as functional ablation of any of the three core PRC2 components, *Eed*, *Ezh2* or *Suz12*, impairs iPSC generation significantly [52,76,77]. Overexpression of *Ezh2*, by contrast, enhances efficiency of iPSC formation, as determined by an increased number of *Nanog*-GFP+ cells relative to control after 7 days of Dox induction of the OSKM reprogramming factors [52].

Many developmental genes in ESCs are enriched for both H3K27me3 and H3K4me3 modifications in their promoter regions. This concurrence of both repressive and activating histone marks at lineage-specific gene promoters establishes a bivalent (‘poised’) state of chromatin activation in response to differentiation-inducing signals [69,78]. Upon lineage commitment, most bivalent chromatin domains become resolved to the monovalent state of either H3K4me3 or H3K27me3. Genes activated upon differentiation acquire H3K4me3 and lose H3K27me3, whereas nonactivated genes lose H3K4me3 but retain H3K27me3 [68,69]. For successful iPSC generation to occur, H3K4 and H3K27 methylation patterns thus need to be reset to the ESC-like bivalent state. Indeed, bivalent histone marks are typically re-established at developmental gene promoters during mouse cellular reprogramming [9,12]. The chromatin-associated

transcription factor Utl1, a direct downstream target of Oct4 and Sox2, is strongly enriched on bivalent genes in ESCs, where it tightly controls their poised state by preventing excessive PRC2 binding and promoting the tagging of mRNAs generated from 'leaky' transcription for subsequent cytoplasmic degradation (a process referred to as mRNA pruning) (Figure 2B) [79]. Utl1 ensures rapid proliferation of ESCs by mRNA pruning-mediated inhibition of *Arf* expression [79], which might explain why Utl1 markedly enhances the efficiency of iPSC generation even in the absence of c-Myc [80]. Collectively, these results highlight the complex interplay of the OSKM reprogramming factors with regulators of chromatin structure in the establishment and maintenance of a ground pluripotent state.

DNA methylation in reprogramming

DNA methylation is the covalent attachment of a methyl (CH₃) group to the carbon-5 position of a cytosine pyrimidine ring, forming 5-methylcytosine (5mC). This reversible epigenetic modification is carried out by specific enzymes, called DNA methyltransferases and occurs predominantly on cytosine residues at cytosine-guanine dinucleotide (CpG) sites in the DNA. There is, in general, a strong correlation between CpG density at gene promoter regions, the methylation status of those CpGs and gene function. Promoter regions with high-CpG content are mostly hypomethylated in ESCs and relate to both ubiquitous housekeeping genes and genes subject to complex regulation during development [70,81]. Whereas the CpG-rich promoters of housekeeping genes are actively expressed and associated with monovalent H3K4me3 marks in ESCs, those of developmental regulators are generally bivalently marked (H3K4me3/H3K27me3) and transcriptionally silent [68]. Notably, H3K4me3 marks are much less prominent at low-CpG promoters, which tend to be hypermethylated and are associated with genes with tissue-specific functions [68,70]. Characterization of promoter DNA methylation profiles in ESCs has shown that, while methylated genes are typically silenced and involved in differentiation, pluripotency genes including *Oct4*, *Sox2* and *Nanog* are generally active and distinguished by promoter hypomethylation [82,83]. Upon lineage commitment, many pluripotency gene promoters acquire methylation, contributing to their transcriptional silencing [84]. During the generation of iPSCs from murine fibroblasts, key pluripotency gene promoters become demethylated whereas genes associated with differentiation (e.g., *HoxA10* and *Gja8*) acquire *de novo* promoter methylation; these changes in DNA methylation patterns have been observed to take place primarily toward the end of reprogramming (Figure 1) [45].

In ESCs, but not in differentiated cells, the genes encoding the Dnmt3a and Dnmt3b methyltransferases are highly expressed, suggesting an important link between *de novo* DNA methyltransferases function and pluripotency [85]. Yet, in the absence of DNA methyltransferase activity, mouse ESCs preserve stem cell properties such as *Oct4* expression and self-renewal but demonstrate little or no ability to differentiate [86,87], which suggests that DNA methylation is essential for differentiation but not important for ground state pluripotency. Consistent with this notion, *de novo* DNA methylation is dispensable for transcription factor-mediated reprogramming, as iPSCs can be derived from mouse embryonic fibroblasts with conditionally deleted *Dnmt3a* and *Dnmt3b* [88].

A substantial degree (up to ~25%) of all DNA methylation in mammalian ESCs occurs at non-CpG sites (CpH, where H = A, C or T) [89,90]. As of today, little is known about the importance of this type of DNA methylation. Methylation in non-CpG contexts is much less abundant in differentiated cells compared with ESCs, despite similar levels of CpG methylation in both cell types [89,90]. Interestingly, CpH methylation is lost upon differentiation of human ESCs and is re-established during reprogramming into iPSCs [89,91,92], suggesting that it is specific to a stem cell state. Ziller *et al.*, however, showed that knock-down of *DNMT3A* and *DNMT3B* in human ESCs globally reduced non-CpG methylation whereas pluripotency remained unaffected [92]. Additional support for the notion that CpH methylation might not be directly related to the pluripotent state is provided by Guo *et al.*'s recent observation that CpH methylation abundantly occurs in the mouse and human brain, where it may function in the silencing of neuronal gene expression [93]. Notably, CpH but not CpG methylation levels were found to gradually increase during postnatal neuronal maturation and to be actively preserved by DNMT3A. Further studies investigating the specific role of non-CpG methylation in the regulation of pluripotency and differentiation are likely to offer new insights into the mechanisms by which cell fate is controlled and may contribute to a better molecular understanding of the reprogramming process.

DNA hydroxymethylation in reprogramming

The TET family of oxygenases plays a particularly important and active role in DNA demethylation by hydroxylating 5mC to 5-hydroxymethylcytosine (5hmC) [94,95]. 5hmC represents a first and principal intermediate of the complete demethylation reaction (Box 2) and can be further oxidized by TET enzymes to subsequently generate 5-formylcytosine and 5-carboxylcytosine, whose exact importance as functional

epigenetic marks requires further investigation [96]. In mammals, the TET family of 5mC hydroxylases consists of three proteins of which two members, Tet1 and Tet2, are specifically abundant in ESCs [97,98]. In accordance, levels of 5hmC in ESCs are high but decline following induction of differentiation [94,99,100]. Interestingly, reprogramming somatic cells to iPSCs induces *Tet1* and *Tet2* expression and enriches 5hmC to levels comparable to those in ESCs [100,101]. Together, these results suggest an essential role for TET proteins and 5hmC in epigenetic regulation of the pluripotent state and differentiation. However, whereas some studies have shown that knockdown of *Tet1* causes

loss of ESC identity [97,102], others have not found a detrimental effect of Tet1 depletion on maintenance of the ESC state [101,103,104]. Yet, deficiency of Tet1 in ESCs is accompanied by a biased differentiation propensity, in particular toward the extraembryonic trophoblast lineage [97,101,103–105]. Several studies that profiled Tet1 binding and 5hmC distribution in mouse ESCs on a genome-wide scale have provided insight into the mechanisms with which TET proteins and 5hmC could regulate lineage differentiation potential (Figure 3). In particular, these studies revealed enrichment of Tet1 and 5hmC at transcription start sites of genes whose promoters were CpG-rich and

Box 2. Possible pathways of DNA demethylation.

- Patterns of DNA methylation (5mC) are established and maintained by the DNA methyltransferase family of enzymes. 5mC can be oxidized in a stepwise manner to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by the action of TET oxygenases. 5hmC may passively become diluted out during cell multiplication as hemihydroxylated DNA is not efficiently recognized by the maintenance DNA methyltransferase Dnmt1 [113–115]. Both the 5fC and 5caC oxidation derivatives are subject to removal by thymine DNA glycosylase, leaving an abasic site that can be repaired back to unmodified cytosine (C) by the base excision DNA repair (BER) machinery [95,116,117]. 5fC and 5caC may also be directly converted back to C by the action of putative deformylases or decarboxylases (dashed arrows) [118,119]. In an alternative pathway, 5mC and 5hmC can be deaminated by the AID or APOBEC family of cytidine deaminases to generate thymine (T) and 5-hydroxymethyluracil (5hmU), respectively. T and 5hmU are then excised by thymine DNA glycosylase or uracil DNA glycosylase and finally repaired by BER machinery [95]. A novel interaction between the deamination and oxidation pathways is highlighted by the finding that T can serve as a substrate for TET-induced oxidation to 5hmU (dark blue arrow) [120]

were associated with bivalent chromatin states [106–109]. Therefore, it seems plausible that Tet1 and/or 5hmC have a function in controlling the pluripotent state and preventing lineage-specific gene expression by regulating bivalent chromatin marking. Supporting this notion, it was found that Tet1 bound to promoters of developmental regulators and recruited PRC2 to deposit H3K27me₃, thereby inducing bivalency-associated transcriptional silencing (Figure 3A) [109]. Interestingly, Mbd3, a methyl-CpG-binding domain protein that specifically binds 5hmC- but not 5mC-marked DNA, was also found to be enriched at high-CpG promoters of PcG-targeted bivalent genes and was physically associated with Tet1 in electrophoretic mobility shift assay experiments [110]. Mbd3, being a component of the NuRD complex, recruits several histone deacetylases (HDACs), chromatin remodeling complexes and transcriptional corepressors, resulting in the imposition of a transcriptionally repressive chromatin structure [111,112]; thus, 5hmC might play a role in the transcriptional silencing of developmentally regulated genes via its interaction with Mbd3.

Genome-wide and functional analyses have also demonstrated a role for TET proteins and 5hmC in promoting active gene expression in ESCs (Figure 3B). In particular, Tet1 and 5hmC are also represented at CpG-rich promoters of transcriptionally active genes, and knockdown of *Tet1* alone or in combination with *Tet2* can result in hypermethylation and downregulation of several pluripotency-related genes in conjunction with a biased differentiation potential [105,109]. Thus, in ESCs, TET-induced conversion of 5mC to 5hmC at pluripotency genes possibly contributes to their transcriptional activation via abrogation of the repressive effect of 5mC. Consistent with this notion, the methyl-CpG-binding transcriptional repressor MeCP2, which associates with corepressor complexes containing Sin3a and HDACs to induce a condensed, transcriptionally incompetent chromatin state at methylated gene promoters [121], fails to recognize 5hmC [122]. Furthermore, 5hmC, and potentially TET itself, may hinder Uhrf1-mediated recruitment of Dnmt1 to the DNA [113–115], resulting in the progressive loss of inhibitory methylation with each DNA replication cycle. This may explain how these rapidly dividing stem cells can maintain low levels of DNA methylation at promoters of constitutively active pluripotency genes. In addition, as Dnmt1 itself associates with transcriptional repressors, such as HDACs [123] and H3K9 methyltransferases [124,125], a failure to recruit Dnmt1 to hydroxymethylated DNA may further prevent the establishment of a repressed chromatin state at these loci. By contrast, the unrestricted binding of Dnmt1 to the heavily methylated promoter regions

of somatic genes in ESCs may ensure that these genes remain transcriptionally silent through cell division (Figure 3C).

Together, these results suggest that TET enzymes and/or 5hmC have a crucial function in finely controlling pluripotency gene expression and preventing differentiation. Accordingly, Tet1 facilitates reprogramming of mouse embryonic fibroblasts toward iPSCs and can even substitute for Oct4 in the original OSKM reprogramming factor cocktail [126]. Tet2 has also been found to be essential for establishing a transcriptionally permissive chromatin environment at stem cell loci during reprogramming of somatic cells [127]. It should be pointed out, however, that TET-induced 5hmC deposition might also recruit Mbd3 along with transcriptional repression machinery, potentially resulting in repression of pluripotency gene expression. Accordingly, Mbd3 has been identified as a prime impediment to the reprogramming process, and ablating Mbd3 greatly improved reprogramming efficiency [128].

Transcriptional & epigenetic variability in cellular reprogramming

Successful reprogramming to iPSCs is largely dependent on faithful remodeling of the cell's epigenetic landscape to shut down somatic gene expression and activate a transcriptional program characteristic of pluripotent cells. Incomplete reprogramming, however, may result in improper epigenetic states that could contribute to variability in gene expression and biological function among iPSC lines, as well as between iPSC and ESC lines. Various studies have indicated distinguishable epigenetic differences between iPSCs and ESCs [91,129–138], many of which have been extensively reviewed elsewhere [27]. In summary, dissimilarities between ESCs and iPSCs have particularly been observed at the level of DNA methylation and are generally due to a failure to reset methylation signatures from the somatic cell of origin or the acquisition of *de novo* aberrant methylation during the reprogramming process. The observation that inhibiting DNA methyltransferases by 5-aza-cytidine could efficiently and rapidly convert partially reprogrammed cells to full iPSCs implies that persistent DNA methylation or insufficient DNA demethylation, in particular at pluripotency-related genes, constitutes a barrier to full reprogramming [9]. Incomplete promoter DNA methylation represents another mechanism underlying the failure to correctly reset the DNA methylome during iPSC formation and has been shown to contribute to residual expression of somatic genes in human iPSCs [91,132]. Importantly, retention of an epigenetic memory of the somatic tissue of origin has been shown to affect the *in vitro* differentiation potential of

Figure 3. Roles of TET enzymes and hydroxymethylation in the control of pluripotency. (A) In embryonic stem cells, TET enzymes induce a poised state of chromatin at promoters of lineage-specific genes by recruiting PRC2, whose Ezh2 subunit catalyzes the methylation of nearby nucleosomes at H3K27. Mbd3, a transcriptional repressor of the NuRD complex that specifically binds to hydroxymethylated cytosines (5hmC), reinforces the repressed chromatin state via associated HDACs and chromatin remodelers. (B) At pluripotency gene promoters, 5hmC and possibly TET itself may hinder Uhrf1-mediated recruitment of the maintenance DNA methyltransferase Dnmt1 and associated transcriptional repressors, thereby keeping inhibitory DNA methylation (5mC) at a low level and safeguarding a transcriptionally permissive state of chromatin. Furthermore, MeCP2, a transcriptional repressor associated with Sin3a and HDAC-containing corepressor complexes, fails to recognize 5hmC. (C) The unhindered binding of Dnmt1 and MeCP2 at heavily methylated promoter regions of somatic genes contributes to gene repression by recruitment of chromatin-modifying transcriptional repressors.

resulting iPSCs, favoring differentiation along donor cell-related lineages while restricting commitment toward other lineages [134–136].

Comparing genetically matched ESCs & iPSCs

To more reliably assess whether or not iPSCs are identical to ESCs on the molecular and functional level, one needs to rule out genetic background effects and compare genetically matched iPSC and ESC lines. Stadtfeld *et al.* used ESCs containing Dox-inducible OKSM transgenes to engineer ‘reprogrammable’ mice from which various somatic cell types were collected for the generation of genetically identical iPSCs [139]. Genetically matched iPSCs and ESCs were found to be transcriptionally very similar and could only be distinguished by a minimal set of two transcripts (*Gtl2* and *Rian*), suggesting that earlier observations of differential gene expression between ESCs and iPSCs [140–142] may be based on differences in genetic background or on other hard-to-control variables inherent to the reprogramming process. Despite the overall similarity in transcriptional profiles, however, iPSCs generated using this transgenic reprogramming system were unable to give rise to viable and germline-competent ‘all-iPSC’ mice (i.e., mice produced exclusively from iPSCs) in the tetraploid complementation assay [139], the golden standard for testing pluripotency. This developmental deficiency of iPSCs was attributed to aberrant epigenetic repression of the *Dlk1–Dio3* imprinted region on chromosome 12qF1, a finding that corroborates another study revealing a positive correlation between stem cell pluripotency level and the degree of activation of this particular gene cluster [143]. Bisulfite pyrosequencing and chromatin immunoprecipitation analyses indicated that the abnormal silencing of normally maternally expressed genes of this locus is caused by DNA hypermethylation as well as reduction of histone modifications associated with active transcription (H3ac, H4ac and H3K4me) [139]. Interestingly, the *Dlk1–Dio3* gene cluster was found to be silenced in a predominant number of iPSC clones derived from various somatic cell types at different differentiation stages, leading the authors to propose that loss of imprinting at this locus and incomplete pluripotency induction are common consequences of transcription factor-mediated reprogramming. In apparent contrast to those findings, Carey *et al.* [21] demonstrated that most iPSCs generated using a highly similar transgenic reprogramming system preserved normal *Dlk1–Dio3* imprinting and were able to produce all-iPSC mice via tetraploid blastocyst complementation. The discordance in findings was attributed to subtle differences in design of the polycistronic expression vectors between the two studies, most notably a switch in the

order of reprogramming factors that resulted in altered levels and stoichiometry of reprogramming factor proteins. iPSCs supporting the development of all-iPSC mice were characterized by 15- and 5-fold higher Klf4 and Oct4, and 2-fold lower Sox2 and c-Myc protein levels than those with impaired developmental potential [21]. Interestingly, increased ectopic expression of Klf4 and Oct4, or treatment with the HDAC inhibitor valproic acid, could restore functional developmental competence of the poor-quality iPSCs [21,139].

Epigenetics & alternative pluripotent states

Further supporting the notion that the expression level of the four reprogramming transcription factors greatly influences the epigenetic state and biological functionality of generated iPSCs, Tonge *et al.* [144–148] described, in a collection of five manuscripts, a novel stable state of induced pluripotency that is distinct from the ESC-like state and requires constitutively high OSKM transgene expression. These so-called F-class cells did not form colonies with typical ESC-like morphology but instead appeared ‘fuzzy’ due to their low intercellular adhesion. Yet, F-class cells displayed high proliferation, expressed *Nanog*, endogenous *Oct4* and several other pluripotency-related genes at ESC levels and were able to give rise to teratomas containing tissues representing all three germ layers, thereby effectively fulfilling some commonly used criteria for defining pluripotency [144]. However, the F-class cells were incapable of contributing to the formation of chimaeric mice, indicating impaired embryonic developmental potential of these cells. By global gene expression profiling of mouse embryonic fibroblasts, traditional (ESC-like) iPSCs and F-class cells, the authors revealed a transcriptional signature unique to the F-class state that persisted in long-term culture and exhibited low variability between independent subclones. Thus, F-class cells do not seem to represent an intermediate reprogramming state but rather a pluripotent alternative to iPSCs. In another paper, Hussein *et al.* [145] provide the results of an extensive multiomic analysis that profiled the transcriptome, proteome and epigenome of the F-class and ESC-like states to better understand, at the molecular level, their emergence. They found that a subset of genes typically expressed in ESCs and iPSCs are repressed in F cells. Although high expression levels of the reprogramming factors were initially required for a genome-wide loss of repressive H3K27me3 and concomitant general opening of somatic cell chromatin, maintenance of high transgene expression induced an F-class state characterized by reacquisition of this repressive histone modification [145]. By contrast, lower transgene levels resulted in the acquisition of activating H3K4me3 and endowed reprogramming cells with an ESC-like phenotype. In

addition, whereas F-class cells failed to lose some parental fibroblast-inherited DNA methylation patterns, ESC-like cells showed proper demethylation of somatic cell-originating signatures. Taken together, these data suggest that active epigenetic mechanisms play a crucial role in directing cell identity toward either the iPSC or F-class state during cellular reprogramming. Further corroborating this notion, Tonge *et al.* [144] showed that treatment with an inhibitor of HDACs converted F-class cells to a transgene-independent, ESC-like state with embryonic developmental potential.

Lee *et al.* [146] investigated the dynamics of DNA (de)methylation throughout reprogramming in more detail and showed that the promoter methylation status of genes at the start of the process controlled their regulation by histone modification alterations. The authors indicated that DNA methylation represented a key epigenetic barrier *en route* to ESC-like induced pluripotency and served as an important determinant of pluripotent outcome (that is, whether cells transitioned toward the ESC-like or F-class state). Yet, epigenetic mechanisms other than DNA meth-

Executive summary

Molecular & transcriptional routes to induced pluripotency

- Gene expression profiling analyses have identified three phases of the murine somatic cell reprogramming process. During the initiation phase, transcription factor-induced cells suppress somatic cell-specific genes and undergo mesenchymal-to-epithelial transition. Cells then gradually transition into maturation, in which a subset of pluripotency-associated genes (e.g., *Nanog*, *Sall4*, *Oct4*) becomes activated. Ultimately, cells transition to the stabilization phase, characterized by induction of the complete pluripotency transcriptional network with expression of endogenous *Sox2*.
- Two major waves of gene activation occur during reprogramming. The first transcriptional wave occurs early and in the majority of cells, and genes related to proliferation, metabolism and cytoskeletal organization are activated. The second wave occurs late and is characterized by upregulation of core pluripotency genes in cells progressing to *bona fide* pluripotency. Transient activation of epidermis-related genes occurs at intermediate reprogramming stages and may represent a common hallmark of reprogramming.

Epigenetic control of cellular reprogramming

- Reprogramming requires global remodeling of somatic cell chromatin from a highly condensed state (heterochromatin) to a more dispersed conformation (euchromatin). Whereas heterochromatin reorganization occurs early in reprogramming, the establishment of euchromatin features constitutes a relatively late event. Heterochromatin-associated H3K9 methylation constitutes an important epigenetic barrier to cellular reprogramming.
- Trithorax and Polycomb group proteins play an indispensable role in specifying and maintaining pluripotent cell identity. During reprogramming, monovalent histone methylation marks at lineage-specific gene promoters are reset to the bivalent state (H3K27me3/H3K4me3) characteristic of stem cells.
- Changes in DNA methylation take place primarily toward the end of reprogramming. Whereas the promoters of differentiation-associated genes acquire *de novo* methylation, pluripotency gene promoters become demethylated. The functional importance of the re-establishment of DNA methylation in non-CpG contexts in induced pluripotent stem cells (iPSCs) requires further investigation.
- iPSC generation induces the expression of Tet1 and Tet2, two members of the ten-eleven translocation family of enzymes that hydroxylate 5-methylcytosine to 5-hydroxymethylcytosine to initiate a process of DNA demethylation. These ten-eleven translocation proteins, along with the 5-hydroxymethylcytosine mark itself, control the pluripotent state and prevent lineage-specific differentiation through their association with both active (pluripotency) and bivalent (developmental) gene promoters.

Transcriptional & epigenetic variability in cellular reprogramming

- Incomplete reprogramming may result in improper epigenetic states that contribute to variability in gene expression and biological function among iPSC lines. Epigenetic aberrations are mostly observed at the level of DNA methylation and include remnant methylation signatures originating from the somatic cell as well as reprogramming-induced *de novo* aberrant methylation patterns.
- Faithful reprogramming toward high-quality iPSCs with full developmental competence is very dependent on various technological parameters that influence epigenetic status, including the levels and stoichiometry of reprogramming factor proteins. Suboptimal reprogramming conditions can cause the emergence of an alternative pluripotent state (F-class cells) with unique transcriptional and epigenetic profiles and altered developmental potential.

Future perspective

- Parallel profiling of single-cell methylomes and transcriptomes will enhance our understanding of the reprogramming process and of the mechanisms contributing to epigenetic, transcriptional and functional heterogeneity of iPSCs.

ylation are also involved in regulating ESC-like versus F-class identity. Notably, Clancy *et al.* [147] revealed the existence of a novel group of differentially expressed miRNAs that specifically control pluripotency in the F-class cells. Epigenetic changes such as these are likely to contribute to the establishment of cell state-specific patterns of protein expression. In support of this idea, Benevento *et al.* [148] showed, using mass spectrometry-based proteomics, that the somatic cell proteome was reorganized in two distinct waves during pluripotency induction and that the F-class proteome was different from the iPSC proteome. As these studies have demonstrated, understanding how the epigenome changes during reprogramming to induced pluripotency will enhance our molecular understanding of the mechanisms of cell fate change and of somatic cell reprogramming in particular.

Conclusion & future perspective

Reprogramming of somatic cells to induced pluripotency using defined transcription factors is predominantly an epigenetic phenomenon. Epigenetic as well as (single-cell) transcriptomic analyses on isolated reprogramming intermediates *en route* to induced pluripotency have generated invaluable mechanistic insights into the process of cell fate specification and transition. It has become evident that faithful reprogramming is very dependent on various technological parameters that can influence cellular epigenetic status, such as the

levels and stoichiometric ratio of reprogramming factors and the composition of the culture environment. In the next few years, parallel profiling of the methylome and transcriptome of single cells (e.g., scM&T-seq [149]) should further advance our understanding of how the OSKM reprogramming factors, and other factors in the reprogramming milieu, impact the epigenome, gene expression pattern and functionality of reprogramming cells and especially those of human origin. This new information will bring us one step closer to the development of more efficient reprogramming methods and, ultimately, high-fidelity generation of therapeutic-grade iPSCs with limited heterogeneity for use in clinical applications [28].

Acknowledgements

The authors would like to thank ML Gage for editing the manuscript.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:

• of interest; •• of considerable interest

- Gurdon JB. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.* 10, 622–640 (1962).
- Wilmot I, Schnieke AE, McWhir J, Kind AJ, Campbell KH. Viable offspring derived from fetal and adult mammalian cells. *Nature* 385(6619), 810–813 (1997).
- Markoulaki S, Meissner A, Jaenisch R. Somatic cell nuclear transfer and derivation of embryonic stem cells in the mouse. *Methods* 45(2), 101–114 (2008).
- Cowan CA, Atienza J, Melton DA, Eggan K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* 309(5739), 1369–1373 (2005).
- Tada M, Takahama Y, Abe K, Nakatsuji N, Tada T. Nuclear reprogramming of somatic cells by *in vitro* hybridization with ES cells. *Curr. Biol.* 11(19), 1553–1558 (2001).
- Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4), 663–676 (2006).
- Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature* 448(7151), 313–317 (2007).
- Wernig M, Meissner A, Foreman R *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 448(7151), 318–324 (2007).
- Mikkelsen TS, Hanna J, Zhang X *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454(7200), 49–55 (2008).
- Guenther MG, Frampton GM, Soldner F *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* 7(2), 249–257 (2010).
- Hawkins RD, Hon GC, Lee LK *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6(5), 479–491 (2010).
- Maherali N, Sridharan R, Xie W *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 1(1), 55–70 (2007).
- Koche RP, Smith ZD, Adli M *et al.* Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* 8(1), 96–105 (2011).
- Mattout A, Biran A, Meshorer E. Global epigenetic changes during somatic cell reprogramming to iPS cells. *J. Mol. Cell. Biol.* 3(6), 341–350 (2011).
- Huangfu D, Maehr R, Guo W *et al.* Induction of pluripotent stem cells by defined factors is greatly improved

- by small-molecule compounds. *Nat. Biotechnol.* 26(7), 795–797 (2008).
- 16 Huangfu D, Osafune K, Maehr R *et al.* Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat. Biotechnol.* 26(11), 1269–1275 (2008).
 - 17 Shi Y, Despons C, Do JT, Hahm HS, Scholer HR, Ding S. Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* 3(5), 568–574 (2008).
 - 18 Hou P, Li Y, Zhang X *et al.* Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science* 341(6146), 651–654 (2013).
 - 19 Hanna J, Saha K, Pando B *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* 462(7273), 595–601 (2009).
 - 20 Papapetrou EP, Tomishima MJ, Chambers SM *et al.* Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proc. Natl Acad. Sci. USA* 106(31), 12759–12764 (2009).
 - 21 Carey BW, Markoulaki S, Hanna JH *et al.* Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell* 9(6), 588–598 (2011).
 - **Indicates that the levels and stoichiometry of reprogramming factor proteins during reprogramming can critically impact epigenetic status and developmental potential of induced pluripotent stem cells.**
 - 22 Eminli S, Foudi A, Stadtfeld M *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nat. Genet.* 41(9), 968–976 (2009).
 - 23 Kim JB, Zaehres H, Wu G *et al.* Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* 454(7204), 646–650 (2008).
 - 24 Bar-Nur O, Brumbaugh J, Verheul C *et al.* Small molecules facilitate rapid and synchronous iPSC generation. *Nat. Methods* 11(11), 1170–1176 (2014).
 - 25 Hochedlinger K, Plath K. Epigenetic reprogramming and induced pluripotency. *Development* 136(4), 509–523 (2009).
 - 26 Papp B, Plath K. Reprogramming to pluripotency: stepwise resetting of the epigenetic landscape. *Cell Res.* 21(3), 486–501 (2011).
 - 27 Liang G, Zhang Y. Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective. *Cell Res.* 23(1), 49–69 (2013).
 - 28 Bhutani K, Nazor KL, Williams R *et al.* Whole-genome mutational burden analysis of three pluripotency induction methods. *Nat. Commun.* 7, 10536 (2016).
 - 29 Boyer LA, Lee TI, Cole MF *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122(6), 947–956 (2005).
 - 30 Loh YH, Wu Q, Chew JL *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* 38(4), 431–440 (2006).
 - 31 Babaie Y, Herwig R, Greber B *et al.* Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem Cells* 25(2), 500–510 (2007).
 - 32 Sato N, Sanjuan IM, Heke M, Uchida M, Naef F, Brivanlou AH. Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev. Biol.* 260(2), 404–413 (2003).
 - 33 Daheron L, Opitz SL, Zaehres H *et al.* LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* 22(5), 770–778 (2004).
 - 34 Thomson JA, Itskovitz-Eldor J, Shapiro SS *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* 282(5391), 1145–1147 (1998).
 - 35 Xu RH, Chen X, Li DS *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat. Biotechnol.* 20(12), 1261–1264 (2002).
 - 36 Takahashi K, Tanabe K, Ohnuki M *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131(5), 861–872 (2007).
 - 37 Lowry WE, Richter L, Yachechko R *et al.* Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl Acad. Sci. USA* 105(8), 2883–2888 (2008).
 - 38 Park IH, Zhao R, West JA *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 451(7175), 141–146 (2008).
 - 39 Anokye-Danso F, Trivedi CM, Juhr D *et al.* Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* 8(4), 376–388 (2011).
 - 40 Suh MR, Lee Y, Kim JY *et al.* Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.* 270(2), 488–498 (2004).
 - 41 Card DA, Hebbard PB, Li L *et al.* Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol. Cell. Biol.* 28(20), 6426–6438 (2008).
 - 42 Schnierch A, Cerdan C, Bhatia M. Distinguishing between mouse and human pluripotent stem cell regulation: the best laid plans of mice and men. *Stem Cells* 28(3), 419–430 (2010).
 - 43 Brambrink T, Foreman R, Welstead GG *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* 2(2), 151–159 (2008).
 - 44 Stadtfeld M, Maherali N, Breault DT, Hochedlinger K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* 2(3), 230–240 (2008).
 - 45 Polo JM, Anderssen E, Walsh RM *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 151(7), 1617–1632 (2012).
 - **Genome-wide profiling of the transcriptome, chromatin status and DNA methylation in reprogramming intermediates en route to induced pluripotency. Reveals a c-Myc-driven early and Oct4/Sox2-driven late wave of transcriptional activation, and identifies DNA methylation changes as a late event in pluripotency induction.**

- 46 Li R, Liang J, Ni S *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* 7(1), 51–63 (2010).
- 47 Samavarchi-Tehrani P, Golipour A, David L *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 7(1), 64–77 (2010).
- 48 Smith ZD, Nachman I, Regev A, Meissner A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat. Biotechnol.* 28(5), 521–526 (2010).
- 49 Golipour A, David L, Liu Y *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* 11(6), 769–782 (2012).
- Indicates that silencing of OSKM transgenes is required for cells to become fully pluripotent. A transcriptomic signature associated with successful acquisition of the pluripotent state was identified that was distinct from the network of genes regulating maintenance of pluripotency.
- 50 Sridharan R, Tchieu J, Mason MJ *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* 136(2), 364–377 (2009).
- 51 O'Malley J, Skylaki S, Iwabuchi KA *et al.* High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* 499(7456), 88–91 (2013).
- Identifies a reprogramming route map contingent on the loss and upregulation of, respectively, CD44 and ICAM1 cell surface markers. RNA-sequencing analysis of intermediate reprogramming populations demonstrated upregulation of pluripotency genes in two distinct waves, as well as transient activation of epidermis-related genes.
- 52 Buganim Y, Faddah DA, Cheng AW *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 150(6), 1209–1222 (2012).
- Single-cell gene expression profiling to characterize various stages and transitions during reprogramming. Reveals substantial variation in gene expression early in reprogramming and identifies Sox2 as an upstream regulator of activation of many genes involved in pluripotency.
- 53 Orkin SH, Hochedlinger K. Chromatin connections to pluripotency and cellular reprogramming. *Cell* 145(6), 835–850 (2011).
- 54 Apostolou E, Hochedlinger K. Chromatin dynamics during cellular reprogramming. *Nature* 502(7472), 462–471 (2013).
- 55 Ho L, Ronan JL, Wu J *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc. Natl Acad. Sci. USA* 106(13), 5181–5186 (2009).
- 56 Liang J, Wan M, Zhang Y *et al.* Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat. Cell Biol.* 10(6), 731–739 (2008).
- 57 Wang J, Rao S, Chu J *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444(7117), 364–368 (2006).
- 58 Wen B, Wu H, Shinkai Y, Izarrry RA, Feinberg AP. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat. Genet.* 41(2), 246–250 (2009).
- 59 Krejci J, Uhlirova R, Galiova G, Kozubek S, Smigova J, Bartova E. Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation. *J. Cell. Physiol.* 219(3), 677–687 (2009).
- 60 Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* 10(1), 105–116 (2006).
- 61 Gaetz J, Clift KL, Fernandes CJ *et al.* Evidence for a critical role of gene occlusion in cell fate restriction. *Cell Res.* 22(5), 848–858 (2012).
- 62 Fussner E, Djuric U, Strauss M *et al.* Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.* 30(9), 1778–1789 (2011).
- 63 Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* 151(5), 994–1004 (2012).
- Identifies megabase-long chromatin domains in fibroblasts, enriched for heterochromatic H3K9me3 and including many genes required for establishing pluripotency, that block binding of the OSKM factors and hinder reprogramming.
- 64 Chen J, Liu H, Liu J *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* 45(1), 34–42 (2013).
- 65 Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* 6(10), e253 (2008).
- 66 Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. *Cell* 128(4), 735–745 (2007).
- 67 Pan G, Tian S, Nie J *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1(3), 299–312 (2007).
- 68 Mikkelsen TS, Ku M, Jaffe DB *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448(7153), 553–560 (2007).
- 69 Bernstein BE, Mikkelsen TS, Xie X *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125(2), 315–326 (2006).
- 70 Meissner A, Mikkelsen TS, Gu H *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454(7205), 766–770 (2008).
- 71 Ang YS, Tsai SY, Lee DF *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* 145(2), 183–197 (2011).
- 72 O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, Jenuwein T. The polycomb-group gene Ezh2 is required for early mouse development. *Mol. Cell. Biol.* 21(13), 4330–4336 (2001).
- 73 Boyer LA, Plath K, Zeitlinger J *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441(7091), 349–353 (2006).

- 74 Lee TI, Jenner RG, Boyer LA *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125(2), 301–313 (2006).
- 75 Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* 27(10), 3769–3779 (2007).
- 76 Ding X, Wang X, Sontag S *et al.* The polycomb protein Ezh2 impacts on induced pluripotent stem cell generation. *Stem Cells Dev* 23(9), 931–940 (2014).
- 77 Onder TT, Kara N, Cherry A *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* 483(7391), 598–602 (2012).
- 78 Azuara V, Perry P, Sauer S *et al.* Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* 8(5), 532–538 (2006).
- 79 Jia J, Zheng X, Hu G *et al.* Regulation of pluripotency and self-renewal of ESCs through epigenetic-threshold modulation and mRNA pruning. *Cell* 151(3), 576–589 (2012).
- 80 Zhao Y, Yin X, Qin H *et al.* Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell* 3(5), 475–479 (2008).
- 81 Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA* 103(5), 1412–1417 (2006).
- 82 Fouse SD, Shen Y, Pellegrini M *et al.* Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* 2(2), 160–169 (2008).
- 83 Imamura M, Miura K, Iwabuchi K *et al.* Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.* 6, 34 (2006).
- 84 Mohn F, Weber M, Rebhan M *et al.* Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* 30(6), 755–766 (2008).
- 85 Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* 19(3), 219–220 (1998).
- 86 Jackson M, Krassowska A, Gilbert N *et al.* Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol. Cell. Biol.* 24(20), 8862–8871 (2004).
- 87 Tsumura A, Hayakawa T, Kumaki Y *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* 11(7), 805–814 (2006).
- 88 Pawlak M, Jaenisch R. *De novo* DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state. *Genes Dev.* 25(10), 1035–1040 (2011).
- 89 Lister R, Pelizzola M, Downen RH *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271), 315–322 (2009).
- 90 Ramsahoye BH, Biniszkiwicz D, Lyko F, Clark V, Bird AP, Jaenisch R. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl Acad. Sci. USA* 97(10), 5237–5242 (2000).
- 91 Lister R, Pelizzola M, Kida YS *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471(7336), 68–73 (2011).
- 92 Ziller MJ, Muller F, Liao J *et al.* Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet* 7(12), e1002389 (2011).
- 93 Guo JU, Su Y, Shin JH *et al.* Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* 17(2), 215–222 (2014).
- 94 Tahiliani M, Koh KP, Shen Y *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324(5929), 930–935 (2009).
- 95 Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145(3), 423–434 (2011).
- 96 Ito S, Shen L, Dai Q *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333(6047), 1300–1303 (2011).
- 97 Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466(7310), 1129–1133 (2010).
- 98 Huang Y, Chavez L, Chang X *et al.* Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* 111(4), 1361–1366 (2014).
- 99 Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* 38(19), e181 (2010).
- 100 Ruzov A, Tsenkina Y, Serio A *et al.* Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res.* 21(9), 1332–1342 (2011).
- 101 Koh KP, Yabuuchi A, Rao S *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 8(2), 200–213 (2011).
- 102 Freudenberg JM, Ghosh S, Lackford BL *et al.* Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res.* 40(8), 3364–3377 (2012).
- 103 Dawlaty MM, Ganz K, Powell BE *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* 9(2), 166–175 (2011).
- 104 Dawlaty MM, Breiling A, Le T *et al.* Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* 24(3), 310–323 (2013).

- 105 Ficiz G, Branco MR, Seisenberger S *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473(7347), 398–402 (2011).
- 106 Williams K, Christensen J, Pedersen MT *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473(7347), 343–348 (2011).
- 107 Pastor WA, Pape UJ, Huang Y *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473(7347), 394–397 (2011).
- 108 Xu Y, Wu F, Tan L *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* 42(4), 451–464 (2011).
- 109 Wu H, D'Alessio AC, Ito S *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473(7347), 389–393 (2011).
- 110 Yildirim O, Li R, Hung JH *et al.* Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147(7), 1498–1510 (2011).
- 111 Morey L, Brenner C, Fazi F *et al.* MBD3, a component of the NuRD complex, facilitates chromatin alteration and deposition of epigenetic marks. *Mol. Cell. Biol.* 28(19), 5912–5923 (2008).
- 112 Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* 13(15), 1924–1935 (1999).
- 113 Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* 67(3), 946–950 (2007).
- 114 Hashimoto H, Liu Y, Upadhyay AK *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* 40(11), 4841–4849 (2012).
- 115 Otani J, Kimura H, Sharif J *et al.* Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. *PLoS ONE* 8(12), e82961 (2013).
- 116 He YF, Li BZ, Li Z *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333(6047), 1303–1307 (2011).
- 117 Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286(41), 35334–35338 (2011).
- 118 Xu S, Li W, Zhu J *et al.* Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. *Cell Res.* 23(11), 1296–1309 (2013).
- 119 Schiesser S, Hackner B, Pfaffeneder T *et al.* Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed. Engl.* 51(26), 6516–6520 (2012).
- 120 Pfaffeneder T, Spada F, Wagner M *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* 10(7), 574–581 (2014).
- 121 Nan X, Ng HH, Johnson CA *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393(6683), 386–389 (1998).
- 122 Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* 32(14), 4100–4108 (2004).
- 123 Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat. Genet.* 24(1), 88–91 (2000).
- 124 Fuks F, Hurd PJ, Deplus R, Kouzarides T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* 31(9), 2305–2312 (2003).
- 125 Esteve PO, Chin HG, Smallwood A *et al.* Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* 20(22), 3089–3103 (2006).
- 126 Gao Y, Chen J, Li K *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(4), 453–469 (2013).
- 127 Doege CA, Inoue K, Yamashita T *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* 488(7413), 652–655 (2012).
- 128 Rais Y, Zviran A, Geula S *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 502(7469), 65–70 (2013).
- 129 Bock C, Kiskinis E, Verstappen G *et al.* Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* 144(3), 439–452 (2011).
- 130 Deng J, Shoemaker R, Xie B *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* 27(4), 353–360 (2009).
- 131 Doi A, Park IH, Wen B *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41(12), 1350–1353 (2009).
- 132 Ohi Y, Qin H, Hong C *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nat. Cell Biol.* 13(5), 541–549 (2011).
- 133 Bar-Nur O, Russ HA, 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(1), 17–23 (2011).
- 134 Polo JM, Liu S, Figueroa ME *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat. Biotechnol.* 28(8), 848–855 (2010).
- 135 Kim K, Doi A, Wen B *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* 467(7313), 285–290 (2010).

- 136 Kim K, Zhao R, Doi A *et al.* Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat. Biotechnol.* 29(12), 1117–1119 (2011).
- 137 Ruiz S, Diep D, Gore A *et al.* Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. *Proc. Natl Acad. Sci. USA* 109(40), 16196–16201 (2012).
- 138 Nishino K, Toyoda M, Yamazaki-Inoue M *et al.* DNA methylation dynamics in human induced pluripotent stem cells over time. *PLoS Genet* 7(5), e1002085 (2011).
- 139 Stadtfeld M, Apostolou E, Akutsu H *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* 465(7295), 175–181 (2010).
- 140 Chin MH, Mason MJ, Xie W *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* 5(1), 111–123 (2009).
- 141 Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR. Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS ONE* 4(9), e7076 (2009).
- 142 Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC. Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PLoS ONE* 5(2), e8975 (2010).
- 143 Liu L, Luo GZ, Yang W *et al.* Activation of the imprinted *Dlk1-Dio3* region correlates with pluripotency levels of mouse stem cells. *J. Biol. Chem.* 285(25), 19483–19490 (2010).
- 144 Tonge PD, Corso AJ, Monetti C *et al.* Divergent reprogramming routes lead to alternative stem-cell states. *Nature* 516(7530), 192–197 (2014).
- Identifies epigenetic modifications that actively direct reprogramming cells toward an alternative stable state of pluripotency (F-class cells) that is characterized by a unique transcriptional signature and is dependent on high reprogramming factor expression.
- 145 Hussein SM, Puri MC, Tonge PD *et al.* Genome-wide characterization of the routes to pluripotency. *Nature* 516(7530), 198–206 (2014).
- 146 Lee DS, Shin JY, Tonge PD *et al.* An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. *Nat. Commun.* 5, 5619 (2014).
- 147 Clancy JL, Patel HR, Hussein SM *et al.* Small RNA changes en route to distinct cellular states of induced pluripotency. *Nat. Commun.* 5, 5522 (2014).
- 148 Benevento M, Tonge PD, Puri MC *et al.* Proteome adaptation in cell reprogramming proceeds via distinct transcriptional networks. *Nat. Commun.* 5, 5613 (2014).
- 149 Angermueller C, Clark SJ, Lee HJ *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* 13(3), 229–232 (2016).