Navigating the epigenetic landscape of pluripotent stem cells

Mo Li[1,*], *Guang-Hui Liu*[1,2,*] *and Juan Carlos Izpisua Belmonte*[1,3]

Abstract | Pluripotent stem cells, which include embryonic stem cells and induced pluripotent stem cells, use a complex network of genetic and epigenetic pathways to maintain a delicate balance between self-renewal and multilineage differentiation. Recently developed high-throughput genomic tools greatly facilitate the study of epigenetic regulation in pluripotent stem cells. Increasing evidence suggests the existence of extensive crosstalk among epigenetic pathways that modify DNA, histones and nucleosomes. Novel methods of mapping higher-order chromatin structure and chromatin–nuclear matrix interactions also provide the first insight into the three-dimensional organization of the genome and a framework in which existing genomic data of epigenetic regulation can be integrated to discover new rules of gene regulation.

All cells in an animal originate from a singe cell — the fertilized egg. These cells share the same genome yet display different phenotypes and perform diverse functions. Individual cellular identity is established during development mainly by epigenetic means, as the genetic material remains mostly unchanged. Cells in the early embryo have the potential to differentiate into all cell types of the animal and are thus termed pluripotent. However, to gain specialized cellular functions, cells need to differentiate and lose their potency. In this process, they acquire stable changes to their epigenomes, and these changes ensure that the differentiation process is unidirectional. Studying the epigenetic mechanisms governing pluripotency and differentiation *in vivo* remains difficult because pluripotent cells only exist transiently and in small numbers. The isolation of embryonic stem (ES) cells, which are pluripotent and can self-renew indefinitely *in vitro*, offers an exciting opportunity to study the interplay between epigenetic regulation and cellular identity[1–3]. In addition, the recent discovery of induced pluripotent stem cells (iPS cells) has uncovered an incredible level of plasticity hidden beneath the layers of epigenetic changes, which were previously thought to be irreversible[4–6].

Epigenetic regulation of gene expression consists of several layers, including DNA methylation, histone modifications, nucleosome packaging and rearrangement, higher-order chromatin structures and the dynamic interplay between chromatin and the nuclear lamina. Perturbation to each of these regulatory layers may cause alterations in the identity of pluripotent stem cells, such as their self-renewal and differentiation potentials, and this might ultimately lead to diseases. Studies on the differentiation of pluripotent stem cells and somatic reprogramming revealed a complex and dynamic interplay among different epigenetic pathways. Gaining insights into the spatiotemporal regulation of epigenetic pathways in pluripotent stem cells will not only further our understanding of human development and diseases but also guide our efforts to better manipulate these cells in the clinical setting.

Here, we first discuss epigenetic regulations in stem cells that primarily pertain to the linear configuration of the genome at the levels of DNA, histones and nucleosomes. We then dedicate two sections to spatial chromatin organization in the three-dimensional nuclear space of pluripotent stem cells, which is the genuine context in which epigenetic regulation occurs. Last, we examine the connection between aberrant epigenetic pathways and human diseases and summarize recent research on modelling such diseases using iPS cells. Furthermore, we share our views on how novel genomic tools may accelerate epigenetic research and how a deeper understanding of epigenetic mechanisms may transform the way we manipulate cell fate conversions, such as differentiation, reprogramming and transdifferentiation.

Changes in DNA methylation

The methylation status of DNA constitutes the first layer of the epigenetic code that is important for lineage specification and reprogramming to pluripotency.

[1]*Gene Expression Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.*
[2]*National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China.*
[3]*Center for Regenerative Medicine in Barcelona, Dr. Aiguader 88, 08003 Barcelona, Spain.*
**These authors contributed equally to this work.*
Correspondence to J.C.I.B
e-mails: belmonte@salk.edu; izpisua@cmrb.eu
doi:10.1038/nrm3393

Box 1 | **Histone modifications by Polycomb and Trithorax group proteins**

Methylation of Lys27 on histone 3 (H3K27) is a major repressive histone modification and a hallmark of facultative heterochromatin. H3K27 methylation is catalysed by enhancer of zeste homologue 2 (EZH2) of Polycomb repressive complex 2 (PRC2). In embryonic stem (ES) cells, H3K27me3 is found at promoters of many key developmental regulators[94]. The mechanism of H3K27me3-mediated repression involves both PRC1 and PRC2. PRC1 is recruited by H3K27me3 and catalyses monoubiquitylation at Lys119 of H2A, which imposes a poised state of RNA polymerase II at repressed promoters[95]. PRC1 directly mediates compaction of chromatin marked by H3K27me3 through a mechanism that is independent of its enzymatic activity[96]. H3K27me3 also recruits PRC2, providing a mechanism of maintaining and propagating H3K27me3-marked domains. Loss of these repressive histone marks increases the propensity of ES cells to differentiate spontaneously, which is consistent with the idea that these epigenetic marks act to prevent precocious lineage commitment by silencing genes involved in differentiation[97]. However, pluripotency per se is not affected by the loss of PRC1 or PRC2 (REFS 97–99).

Interestingly, H3K27me3 modifications are often found at developmental gene promoters that are also marked with H3K4me3, an activating epigenetic mark that is catalysed by Trithorax group (TrxG) proteins. The activity of TrxG proteins opposes the action of Polycomb group (PcG) proteins during development[100–103]. The presence of opposing epigenetic marks, termed 'bivalency', is thought to keep developmental genes silenced yet poised for activation. Bivalent domains could resolve into active or repressive chromatin depending on whether TrxG or PcG proteins prevail when cells commit to a specific lineage.

Induced pluripotent stem cells
(iPS cells). Somatic cells that have been reprogrammed to a pluripotent state, which is highly similar to that of embryonic stem cells. iPS cells were first generated by the Yamanaka group in 2006 from mouse somatic cells by enforced expression of OCT4, SOX2, KLF4 (Krüppel-like factor 4) and c-MYC. iPS cells have been successfully derived from somatic cells of different species through overexpression of various combinations of factors.

Somatic reprogramming
Specifically referred to as reprogramming of somatic cells towards pluripotency, which entails the erasure of epigenetic marks of somatic cell origin and re-establishment of pluripotency-specific transcriptional and epigenetic programmes. The three major approaches for somatic reprogramming are somatic cell nuclear transfer, cell fusion-based reprogramming and transcription factor-based reprogramming.

DNA methylation of CpG dinucleotides. Methylation of cytosine at CpG dinucleotides is the predominant form of epigenetic modification at the level of DNA. The DNA methylation pattern of the mammalian genome is maintained by DNA methyltransferase 1 (DNMT1). DNMT1 preferentially methylates hemi-methylated DNA, thereby maintaining DNA methylation patterns during DNA replication. By contrast, DNMT3A and DNMT3B have no preference for hemi-methylated DNA and can methylate fully unmethylated DNA *de novo*. They are responsible for restoring DNA methylation levels in pluripotent inner cell mass (ICM) cells after fertilization[7].

Both maintenance and *de novo* methyltransferases are highly expressed in ES cells. Deficiencies in these enzymes cause hypomethylation and a progressive loss of the differentiation potential in ES cells[8]. Interestingly, the ability of ES cells to self-renew is not affected by global loss of DNA methylation or deficiency in DNMTs[9,10]. These data demonstrate the importance of DNA methylation in lineage specification but not in self-renewal of ES cells.

DNA demethylation in development and reprogramming. DNA demethylation may occur passively during DNA replication when the fully methylated state fails to be restored on the newly synthesized hemi-methylated DNA, or it may occur actively through enzymatic activities. The genome-wide DNA methylation pattern is reset during early development, as global demethylation first occurs during pre-implantation development in zygotes and blastomeres and later in primordial germ cells (PGCs)[11].

Recent data indicate two pathways in active DNA demethylation. Activation-induced cytidine deaminase (AID) has been shown to be required for genome-wide demethylation of PGCs and for demethylation of the promoters of two master regulators of pluripotency, namely *OCT4* (also known as *POU5F1*) and *NANOG*, during heterokaryon-based reprogramming of human fibroblasts[12,13]. However, the role of AID in reprogramming has recently been challenged because of the lack of endogenous AID expression in certain mouse ES cell lines, and because of the absence of *Oct4* promoter demethylation in somatic fusion partners when AID is overexpressed in the ES cells that are used for the fusion[14].

Another demethylation pathway entails iterative oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by the ten-eleven translocation (TET) family of enzymes followed by base excision repair[15]. After fertilization, the paternal genome is quickly and actively demethylated, and it has recently been shown that this process requires the hydroxylation of 5mC by TET3 (REF. 16). TET3 is also important for *Oct4* activation during somatic cell nuclear reprogramming by oocytes[16]. In somatic reprogramming, demethylation of pluripotency gene promoters is a rate-limiting step towards the fully reprogrammed iPS cell state; however, this process is poorly understood. The elucidation of the mechanisms that control promoter demethylation in this context is likely to result in novel strategies to enhance reprogramming in the future.

Histone modifications: dynamic codes

Rather than being the 'first responder' to lineage specifying cues, DNA methylation acts as a consolidator to stabilize previously established repression to provide enduring silencing of unnecessary genes in a specific lineage. In ES cells, most differentiation-related genes are regulated by chromatin-mediated mechanisms rather than by DNA methylation. Regulation of gene expression at the chromatin level involves histone modifications and rearrangement of nucleosomes and higher-order chromatin structures. Compared to DNA methylation, these epigenetic modifications are more dynamically regulated during development.

Various histone modifications have been extensively studied in pluripotent stem cells. Polycomb group (PcG) and Trithorax group (TrxG) proteins are the major effectors of histone modifications. PcG proteins are thought to prevent precocious lineage commitment of pluripotent stem cells but are not required for self-renewal and pluripotency (BOX 1). The requirement of TrxG proteins in pluripotent stem cell biology is less investigated, although a recent study indicates a crucial role of the TrxG protein WD repeat-containing protein 5 (WDR5) in ES cell self-renewal. As the topic of histone modifications in pluripotency has been recently reviewed[17], we limit our discussion to the latest findings.

Mechanisms that ensure target specificity of TrxG and PcG proteins. In pluripotent stem cells, PcG and TrxG proteins are essential for maintaining the balance between self-renewal and differentiation (BOX 1). As PcG and TrxG proteins do not bind to specific DNA sequences, a major question is how genes are

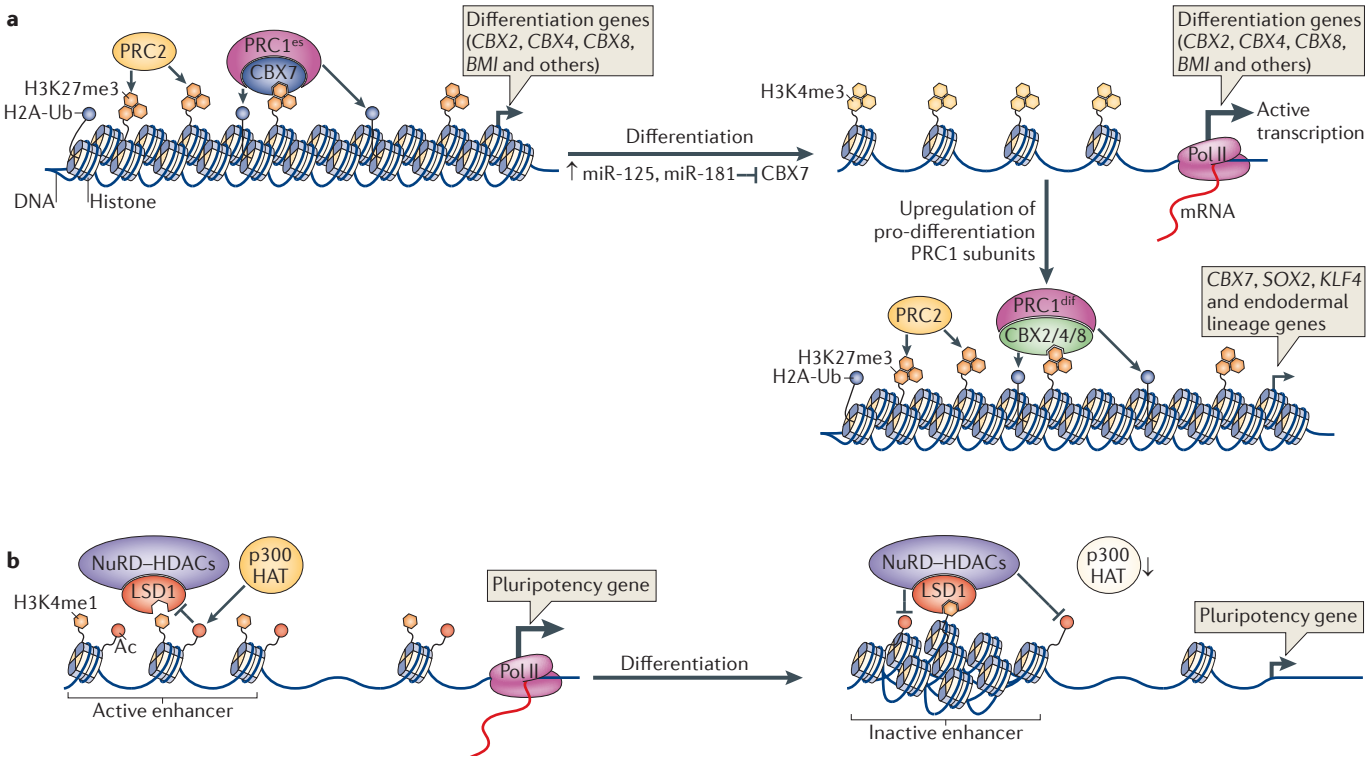


Figure 1 | **Mechanisms by which epigenetic regulators control the switch from pluripotency to differentiation.** **a** | In embryonic stem (ES) cells, Polycomb repressive complex 1 (PRC1) preferentially utilizes CBX7, which is a subunit of the chromobox family of PRC1 components. CBX7 binds trimethylated Lys27 on histone H3 (H3K27me3) through its chromodomain. This ES cell-specific PRC1 (called PRC1^{es}) complex auto-regulates differentiation-associated CBX proteins and represses developmental regulators. CBX7 is downregulated during differentiation, possibly through a microRNA (miRNA)-mediated mechanism. CBX2, CBX4 and CBX8 are subsequently derepressed and form a differentiation-specific PRC1 (termed PRC1^{dif}) complex that has important roles in lineage commitment. **b** | The Lys-specific demethylase 1 (LSD1)–nucleosome remodelling and deacetylase (NuRD) complex, which contains histone deacetylases (HDACs), is found at active enhancers in pluripotent stem cells, but is thought to be inhibited by the activity of co-activators such as p300 (which contains a histone acetyltransferase (HAT) domain). p300 levels decrease during differentiation, which allows LSD1 to demethylate H3K4 and inactivate the enhancer. Long black arrows depict actively transcribed genes. Short black arrows depict silent genes. Ac, acetylation; *BMI*, B lymphoma Mo-MLV insertion region 1 homologue; Pol II, RNA polymerase II; Ub, ubiquitin.

differentially targeted by these proteins in pluripotent and differentiated cells. WDR5, a major component of the MLL (mixed lineage leukaemia) complex, which maintains histone 3 Lys4 trimethylation (H3K4me3) levels, is specifically downregulated during differentiation and upregulated when pluripotency is regained through reprogramming[18]. Similarly, ES cell-specific subunits of Polycomb repressive complex 2 (PRC2) are also rapidly downregulated during differentiation[19]. PRC1 consists of multiple subunits that are maintained at stoichiometric ratios, and these subunits have diversified in mammals. Consequently, functionally distinct PRC1 complexes could result from the incorporation of alternative subunits. Indeed, the members of the chromobox family of PRC1 components (termed CBX2, CBX4, CBX6, CBX7 and CBX8), which detect H3K27me3, are incorporated differentially into PRC1 in ES cells and differentiated cells[20,21]. In ES cells, PRC1 predominantly contains CBX7, which has crucial roles in preventing precocious differentiation[20,21]. CBX7 also negatively regulates other CBX proteins in pluripotency. Conversely, CBX2, CBX4 and CBX8 are upregulated in ES cells that undergo differentiation when CDX7 is downregulated[20,21]. Interestingly, this PRC1 subunit switch could be aided by microRNAs[21] (FIG. 1a). Differential expression of alternative subunits is likely to represent a common mechanism of retargeting PcG and TrxG proteins during development.

H3K27 and H3K4 methylation marks can be removed by specific histone demethylases, providing a means to remove epigenetic memory in daughter cells[22–26]. Recently several groups examined the roles of the H3K4 demethylase LSD1 (Lys-specific demethylase 1) in pluripotency and differentiation in ES cells. LSD1 binds to enhancers of active and bivalent genes (BOX 1) together with pluripotency transcription factors and co-activators. However, H3K4 demethylation by LSD1 at active enhancers is held in check in pluripotency, possibly by co-occupying transcriptional activators. During mouse ES cell differentiation, the levels of the activators decrease, thereby permitting LSD1 to demethylate and inactivate these enhancers[27] (FIG. 1b). This suggests that

Inner cell mass
(ICM). Refers to a population of cells inside the early embryo (the blastocyst), which ultimately give rise to all fetal tissues. The ICM is located at the embryonic pole of the blastocyst and is surrounded by a monolayer of trophoblast cells. Mouse embryonic stem cells were initially isolated from ICM.

Blastomeres
Cells formed by cleavage (which is the initial rapid cell divisions after fertilization) during early embryonic development.

LSD1 has a key role in shutting down the pluripotency programme during differentiation. Interestingly, our own data show that knockdown of LSD1 in human ES cells affects the expression of a rather small group of direct target genes, which is consistent with mouse studies. However, many LSD1 target genes are bivalently marked developmental regulators, which become precociously expressed following LSD1 knockdown[26]. Therefore, our data suggest a role of LSD1 in balancing the level of the H3K4me3 mark in bivalent domains of developmental regulators.

Heterochromatin organization in pluripotent stem cells. H3K9 methylation marks constitutive heterochromatin in pericentric and telomeric regions. In ES cells, staining of H3K9me3 and heterochromatin protein 1 (HP1), which is a major structural protein of heterochromatin, as well as direct visualization of heterochromatin DNA by fluorescence *in situ* hybridization (FISH) show a more diffused and less compartmentalized pattern than observed in differentiated cells, indicating reorganization of heterochromatin during differentiation[28]. At the ultrastructural level, heterochromatin in pluripotent stem cells is comprised exclusively of dispersed 10 nm fibres, as opposed to the highly compact 10 nm fibres found in mouse embryonic fibroblasts (MEFs) and partially reprogrammed iPS cells[29]. Consistently, telomeric chromatin loosens up in pluripotent stem cells and becomes more amenable to transgene expression[30].

Heterochromatin compaction in somatic cells is likely to represent an epigenetic barrier to somatic reprogramming, as inhibiting H3K9 methylation by chemical inhibitors of the H3K9 methyltransferase G9A (also known as EHMT2), knocking down G9A or overexpressing the H3K9me3 demethylase KDM3A (also known as JHDM2A) all increase the efficiency of iPS cell generation[31–33]. Interestingly, recent evidence has shown that global H3K9me2 levels are similar between ES cells and terminally differentiated cells in mice. However, local changes of H3K9me2 are detected during differentiation of mouse ES cells[34]. These data suggest that heterochromatin reorganization during differentiation is not likely to be due to a massive increase in repressive histone marks but rather due to a reorganization of such marks, and that mechanisms in addition to histone modification must be involved in the remodelling of heterochromatin in pluripotent stem cells.

Chromatin remodelling in pluripotent stem cells

With the exception of acetylation, which directly acts to loosen histone–DNA bonds and opens nucleosomes via charge neutralization, most histone modifications do not impose changes to the chromatin conformation[35]. They often act as a 'homing beacon' for other chromatin-associated factors including chromatin-remodelling factors, which utilize energy from ATP hydrolysis to exchange histones and reposition or evict nucleosomes. Because the involvement of different chromatin-remodelling factors in pluripotency and differentiation has been extensively reviewed[17], we only discuss the most recent findings.

Several subunits of the BRG- or Brahma-associated factor (BAF) complex have been shown to be important for ES cell self-renewal and pluripotency[36–38]. Interestingly, the BAF complex not only suppresses developmental genes but also fine-tunes the level of expression of ES cell-specific genes. A recent study has shed light on the mechanism underlying the dual role of BAF. BRG1 (also known as SMARCA4), which is a subunit of the ES cell-specific BAF complex (esBAF), is found to be essential for chromatin accessibility of the targets of the LIF–STAT3 signalling pathway. PcG proteins ectopically silence these LIF–STAT3 target loci in the absence of BRG1, indicating an antagonizing relationship between esBAF and PcG proteins in LIF–STAT3 signalling[39]. Surprisingly, despite its classification as a TrxG member, BRG1 synergizes with PcG proteins to silence *HOX* genes in ES cells[39]. The molecular mechanisms underlying the context-dependent modulation of PcG protein function by esBAF await further investigation. It would also be interesting to see whether this is a phenomenon unique to esBAF or a property of other forms of the BAF complex as well.

The nucleosome remodelling and deacetylase (NuRD) complex contains both a nucleosome remodelling ATPase (called MI2) and histone deacetylases (termed HDAC1 and HDAC2). Loss of NuRD components de-represses trophectoderm markers and impairs ES cell differentiation[40]. From several recent publications, a common theme emerged for the role of the NuRD complex in coordinating different epigenetic pathways. NuRD physically associates with LSD1 and co-occupies most of the LSD1-bound enhancers[27]. Thus, NuRD may act synergistically with LSD1 in silencing enhancers during ES cell differentiation. Conversely, the NuRD subunit MBD3 (methyl-CpG-binding domain protein 3) and the BAF subunit BRG1 bind cooperatively to common targets but antagonistically regulate nucleosome occupancy at these sites[41]. In a recent study, analysis of 38 sets of genome-wide binding data uncovered crosstalk between the MBD3 and 5hmC pathways. The data show that MBD3-binding sites overlap with 5hmC and with the binding sites of TET1, the enzyme that catalyses the hydroxylation of 5mC. Genes misregulated in *Mbd3*-knockdown cells show higher levels of TET1 binding and 5hmC. Interestingly, both MBD3 and BRG1 are required for maintaining the global levels of 5hmC in ES cells[41]. These studies demonstrate the exquisite interplay of different layers of epigenetic regulation.

Higher-order chromatin structures

Another fascinating yet underexplored question of epigenetic mechanisms is how higher-order chromatin structures beyond the 'nucleosomes-on-a-string' level take part in the regulation of pluripotency and differentiation.

Impact on gene regulation in development and pluripotency. During development, local chromatin domain organization of lineage-specific genes displays tissue and temporal specificity. For example, by using an experimental strategy named chromosome conformation

Primordial germ cells
(PGCs). The precursors of sperms and eggs. During development, PGCs are specified far from their somatic niche and have to actively migrate to the gonadal ridge to become mature germ cells.

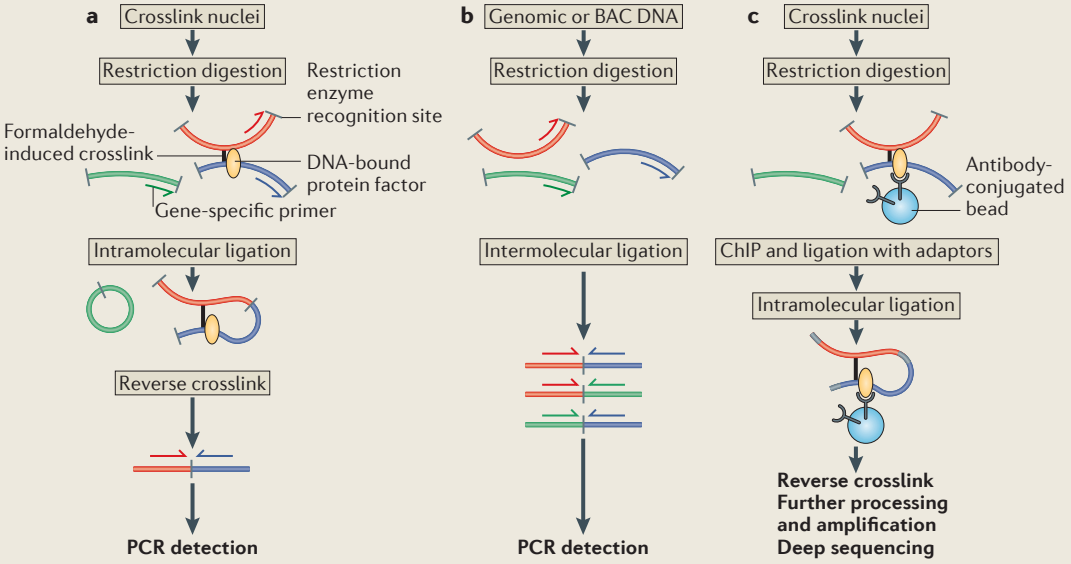
Activation-induced cytidine deaminase
(AID). A factor required for generating antibody diversity by introducing mutations into the immunoglobulin loci in B cells. AID enzymatically converts cytidines into uracils, thus creating mismatches that initiate downstream repair pathways, which produce diversified immunoglobulin sequences. AID can also convert 5mC into thymidine through deamination.

Heterokaryon-based reprogramming
A reprogramming strategy in which somatic cells are reprogrammed by fusion with pluripotent stem cells to create hybrid cells (also known as heterokaryons). Heterokaryon-based reprogramming is rapid, efficient and independent of cell division.

LIF–STAT3 signalling
(leukaemia inhibitory factor – signal transducer and activator of transcription 3 signalling). In mouse embryonic stem cells, the binding of LIF to its receptor leads to the activation of the transcription factor STAT3, which is important for the maintenance of self-renewal. The LIF–STAT3 pathway activates many downstream targets, including the pluripotency factors c-MYC, SALL4 and KLF4 (Krüppel-like factor 4), to form a pluripotency transcriptional network.

Trophectoderm
The cell layer from which the trophoblast differentiates. Trophoblasts are the peripheral cells of the blastocyst that develop into a large part of the placenta and the membranes that nourish and protect the developing embryo.

Box 2 | **Chromosome conformation capture and its derivatives**



Chromosome conformation capture (3C) was invented a decade ago by Dekker *et al.* as a means to study the spatial organization of the genome[104]. The 3C technology and its later variants remain the methods of choice for systematic analysis of genomic interactions and chromatin topography. The figure shows the 3C technique (part **a**) and the procedures for generating a reference sample for 3C in which every DNA fragment has the same chance of ligating to every other DNA fragment (part **b**).

In essence 3C converts the three-dimensional information on spatial proximity of DNA into biochemical events that are quantifiable by various molecular biology tools (for example, PCR, microarray and next-generation sequencing). This technique uses formaldehyde-mediated crosslinking to capture spatial juxtaposition of DNA. Crosslinked DNA is first digested with a restriction enzyme and subsequently subjected to intramolecular ligation. Every ligation event captures the physical interaction of two DNA fragments that may or may not be separated by large linear distances. Once captured, the frequency of these events can be quantified and compared, providing insights into three-dimensional chromatin structure. The same information stored in the ligations is analysed differently in various 3C derivatives. The original 3C technique detects one pair of interactions at a time and is therefore highly inefficient.

Genomic adaptations of 3C (for example, circularized chromosome conformation capture (4C), chromosome conformation capture carbon copy (5C) and high-throughput chromosome capture (Hi-C); reviewed in REF. 62) combine the technique with high-throughput genomic tools such as next-generation sequencing, thereby greatly enhancing the power of discovery. Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) incorporates a chromatin immunoprecipitation (ChIP) step into 3C procedures to allow for enrichment of interactions bound by (or maybe organized by) specific proteins (see the figure, part **c**). It should be noted that 3C-based methods can only measure the population average steady-state of DNA–DNA interactions. Other microscopy-based methods (such as fluorescence *in situ* hybridization (FISH)) should be used to confirm the observations at the single cell level and to study the dynamics of such interactions. BAC, bacterial artificial chromosome.

capture (3C) (BOX 2), it has been shown that the murine β-globin locus control region enhancer interacts with active β-globin genes through long-distance chromatin looping in erythroid lineages but not in neural cells[42]. The chromatin loops in erythroid progenitors undergo a developmental switch from favouring the expression of embryonic globin genes to favouring the expression of the adult globin genes as definitive erythroid cells emerge[43]. Also by using 3C, it has been shown that OCT4 has an important role in the organization of higher-order chromatin structure. In mouse ES cells, the extended *Nanog* locus (from −151 to +9 kb) harbours extensive long-range interactions between regulatory elements marked by DNaseI hypersensitive sites. These chromatin loops are postulated to bring distant regulatory elements together to form an active chromatin domain, which could ensure the high expression level of the genes within this domain. The formation of these active chromatin loops is dependent on OCT4, as acute loss of OCT4 causes the collapse of the higher-order chromatin structures[44] (FIG. 2a).

Factors organizing higher-order chromatin structure. How does OCT4 organize higher-order chromatin structure? The insulator protein CCCTC-binding factor (CTCF) is one of the best known organizers of chromatin domains. This protein has been shown to mediate long-range interactions between distant regulatory elements. OCT4 may modulate the binding of CTCF, as the binding site of CTCF in the *Nanog* locus is flanked by OCT4-binding sites. In support of this, OCT4 and CTCF have been shown to interact during X chromosome inactivation[45]. Alternatively, OCT4 may affect other factors that are important for chromatin looping.

Insulator protein
Regulatory protein that binds to insulator elements in the DNA. DNA-bound insulators can block the communication between enhancers and gene promoters when situated between them. They can act as a barrier to the spread of heterochromatin.

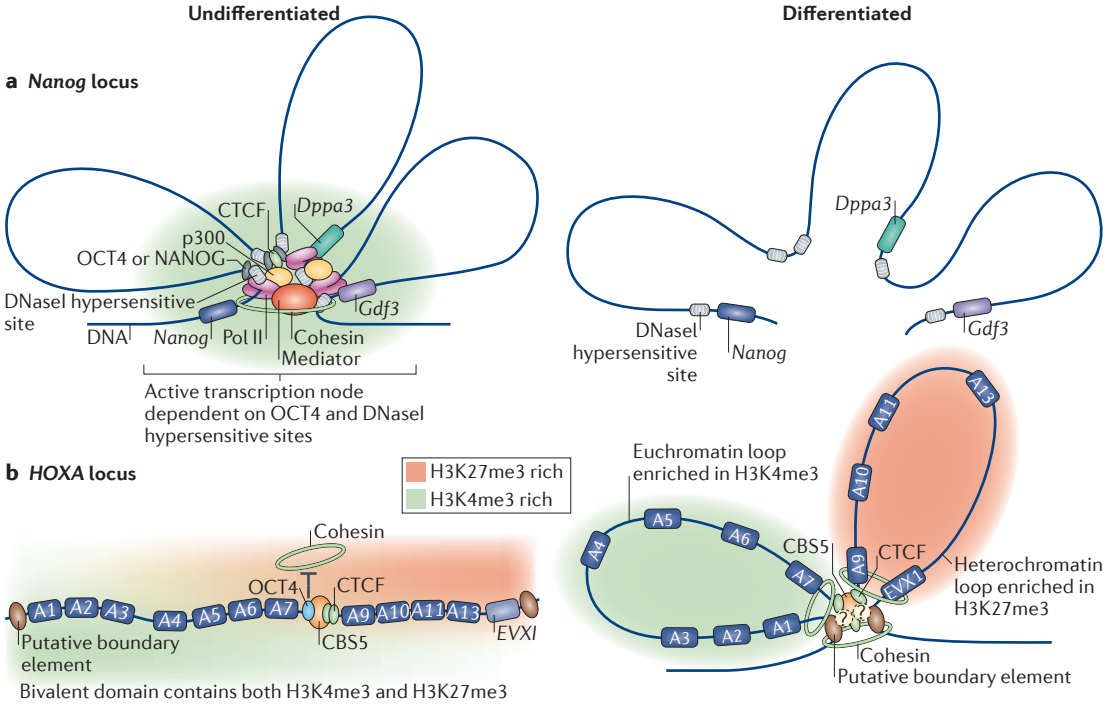


Figure 2 | **OCT4 and CTCF organize three-dimensional chromatin loops in ES cells. a** | Models depicting the higher-order chromatin structure in the *Nanog* locus. Distant regulatory elements (which are DNaseI hypersensitive sites) in the *Nanog* locus interact extensively with each other in mouse embryonic stem (ES) cells. These interactions could serve to bring the genes that encode NANOG, DPPA3 (developmental pluripotency-associated protein 3) and GDF3 (growth/differentiation factor 3) into an active transcription node, where RNA polymerase II (Pol II), p300 and mediator are present. The formation of this higher-order structure is dependent on OCT4 and possibly involves CTCF (CCCTC-binding factor) and cohesin. The downregulation of OCT4 in differentiated cells leads to the collapse of the higher-order chromatin structure. **b** | Models depicting the higher-order chromatin structure in the *HOXA* locus. In undifferentiated cells, the conserved CTCF-binding site (termed CBS5) is co-occupied by CTCF and OCT4, and OCT4 prevents cohesin loading to chromatin at CBS5. As a result, chromatin loops are not formed. Upon differentiation, cohesin can bind CTCF at CBS5 in the absence of OCT4 and partitions the *HOXA* locus into a euchromatin loop (green) and a heterochromatin loop (red), enriched in trimethylated Lys4 on histone 3 (H3K4me3) and H3K27me3, respectively. The formation of the chromatin loops may involve putative boundary elements (brown ovals) and unidentified boundary proteins (yellow circles with a question mark). Genes in the *HOXA* locus (*HOXA1* to *HOXA7*, *HOXA9* to *HOXA11*, *HOXA13* and *EVX1* (even skipped homeobox 1)) are depicted as blue boxes. Images in part **a** are modified, with permission, from REF. 44 © (2008) Cold Spring Harbor Laboratory Press.

A recent study on higher-order chromatin structure of the mouse and human *HOXA* locus shows that OCT4 prevents the formation of chromatin loop domains by antagonizing chromatin loading of cohesin, which is a key partner for CTCF and a mediator of chromatin looping[46]. In human lung fibroblasts, cohesin binds together with CTCF at a conserved CTCF-binding site (named CBS5) in the *HOXA* locus and partitions the active and silent genes in this locus into a euchromatin loop and a heterochromatin loop, respectively (FIG. 2b). Such loop formation is not detected in ES cells, where OCT4, but not cohesin, is found to colocalize with CTCF (FIG. 2b). Following differentiation of mouse ES cells into neural progenitors, cohesin binds to CBS5 in the absence of OCT4 and re-establishes the heterochromatin loop[46]. In addition to providing a mechanism by which OCT4 regulates higher-order chromatin structure, the most salient point of this study is the finding that perturbations of higher-order chromatin structures result in dramatic locus-wide gene expression changes. Although the *in vivo* significance of these findings remains to be elucidated, they nonetheless bolster the idea that higher-order chromatin structure has an important role in regulating pluripotency and differentiation.

A recent survey of genomic binding of cohesin in ES cells has revealed overlaps with not only CTCF but also the co-activator mediator at active enhancers and promoters[47]. Furthermore, mediator, cohesin and the cohesin-loading factor NIPBL (nipped-B-like protein) physically interact and co-occupy active genes. These data suggest a model in which the mediator–cohesin–NIPBL complex contributes to chromatin looping between enhancers and promoters. Consistently, interactions between enhancers and core promoters of ES cell-specific genes are detected by 3C only in ES cells but not in MEFs[47].

Protein factors are not the only regulators of higher-order chromatin structure. Recently, long non-coding RNAs (lncRNAs) have emerged as regulators of chromatin conformation in *Hox* gene clusters during mouse

Cohesin
A multiprotein complex that mediates sister chromatid cohesion during cell division. Cohesin is also involved in other processes including DNA double-strand break repair and transcription. Recently, cohesin has been implicated in organizing higher-order chromatin structure.

Box 3 | **Organizing chromatin at the nuclear lamina**

The nuclear lamina is a protein meshwork underneath the nuclear envelope. In somatic cells, it is composed of A- and B-type lamins and a number of inner nuclear membrane proteins[105,106]. In somatic cells, the nuclear lamina is particularly important in organizing heterochromatin at the nuclear periphery and for the positioning of chromosome territories[82,107,108]. The interaction between the nuclear lamina and chromatin is either mediated by intermediate proteins, including LBR (lamin B receptor), LAP2β (lamina-associated protein 2β), BAF (BRG- or BRM-associated factor) and others, or via direct interactions among lamins, histones and DNA[105,106]. DamID (DNA adenine methyltransferase identification) is especially useful in mapping lamina–chromatin interactions. DamID relies on the ability of the fusion protein between a factor of interest and the *Escherichia coli* DNA adenine methyltransferase (Dam) to mark genomic binding sites of the Dam fusion partner with adenine methylation, a modification absent in eukaryotes. When applied to nuclear lamina components, DamID reveals large chromatin domains (termed lamina-associated domains (LADs)) that associate with the nuclear lamina across the genome[109]. Compared to regions that fall between LADs, LADs are enriched in trimethylated Lys27 on histone 3 (H3K27me3), which is a silent chromatin mark, and are depleted of the active chromatin mark H3K4me3 (REF. 109). Thus, LADs are usually transcriptionally silenced.

ES cell differentiation and development[48,49]. Common themes are emerging from published data. First, unlike the global effects on chromatin domains by CTCF, lncRNA-mediated chromatin looping is confined to a specific locus[48,49]. Second, lncRNAs act in concert with chromatin modifiers, such as MLL and PRC2, to influence expression of distant genes that are brought together by chromatin looping[48–51]. Whether lncRNAs coordinate chromatin looping and chromatin modifications in a similar fashion in loci other than the *Hox* gene clusters remains to be determined.

Influence of nuclear localization on gene transcription. Beyond the level of individual gene loci, chromatin organization also influences gene expression at a more global level. Studies on *Hox* gene clusters in differentiating ES cells and developing embryos have revealed that colinear expression of *Hox* genes is linked to chromatin decondensation and to large-scale movement of genes with respect to specific chromosome territories[52–54]. In human ES cells, pluripotency gene loci, such as *NANOG* and *OCT4*, are more likely to occupy nuclear space that is more amenable to active transcription[55,56]. Consistent with the aforementioned fluid and open nature of chromatin in pluripotent stem cells, centromeres in ES cells are less frequently found at the nuclear periphery (a site often associated with repressed chromatin) compared to centromeres in differentiated cells[56]. This observation raises the enticing idea that nuclear localization of centromeres may influence gene silencing at a global level. It should be noted that establishing causality between nuclear relocalization of genes and their expression status remains challenging. Using balanced chromosome translocation as a model, it was revealed that chromosomal rearrangements caused large-scale gene expression changes, which are thought to be consequences of altered nuclear organization[57]. However, other studies have uncoupled nuclear repositioning of genes and transcription, painting a more complex picture of the relationship between these two processes[58,59].

Centromeres
The sites of sister chromatid cohesion on the chromosome after DNA replication, and the sites of chromosome binding to the mitotic spindle. Eukaryotic centromeres mainly consist of repetitive DNA and are in a heterochromatin state.

Genome-wide mapping of higher-order chromatin structure. With the advent of high-throughput adaptations of 3C (BOX 2) and next generation sequencing technology, it is now possible to map higher-order chromatin structures at the genome level and integrate this knowledge with existing data of histone modifications, transcription factor binding and gene expression[60–62]. Genome-wide CTCF-mediated chromatin interactions in mouse ES cells were mapped using ChIA-PET (chromatin interaction analysis by paired-end tag sequencing), which is a technique combining chromatin immunoprecipitation (ChIP), 3C and deep sequencing (BOX 2). By comparing the map of CTCF–chromatin interactions with the localization of various histone modifications, RNA polymerase II (Pol II) binding, p300 binding and lamina-associated domains (LADs) (BOX 3), the authors uncovered surprising new rules of chromatin organization, which would be hard to appreciate from analysing any data set alone or from studying single gene loci[61] (FIG. 3). For example, CTCF was not only found to demarcate repressive chromatin domains such as LADs, but was also frequently (in 19% of the cases) found to perform an unexpected 'enhancer-bridging' function. The strength of genome-wide analysis of three-dimensional chromatin configurations is that it brings in the missing third dimension, therefore providing a framework in which independent genomic data (for example, p300 and H3K4me3 ChIP-sequencing data) can be integrated to gain novel insights into gene regulation. However, there is at least one caveat. ChIA-PET only surveys spatial genomic interactions that are organized by a predetermined factor (for example, CTCF).

To have a more complete and unbiased view of the spatial organization of the genome, two recent studies used two approaches that couple 3C derivatives and deep sequencing, therefore affording unprecedented coverage and resolution. One group investigated the spatial genome organization of a 4.5 Mb region encompassing the X chromosome inactivation centre (XIC) by using chromosome conformation capture carbon copy (5C)[62] (BOX 2) in mouse ES cells, neuronal progenitors and MEFs. The parallel analysis of 250,000 possible genomic interactions revealed a series of topologically distinct domains, termed topologically associating domains (TADs)[63]. The data suggest that TADs represent a more fundamental organization framework of the genome, as they are largely unchanged during differentiation and X chromosome inactivation and are insensitive to perturbation of histone modifications[63]. The boundaries of TADs play an active part in organizing TADs, as the deletion of one such boundary leads to ectopic interactions between TADs and misregulation of gene expression[63]. Importantly, TADs are enriched for genes that are coordinately regulated in development. In particular, the regulatory sequences of the *Xist* locus and its antisense transcript *Tsix* are partitioned into two adjacent TADs, which may ensure transcriptional independence of these loci.

Unlike 5C, the high-throughput chromosome capture (Hi-C) approach is not confined to a predetermined locus and in principle captures genome-wide chromatin interactions[63–65] (BOX 2). Due to the enormous complexity

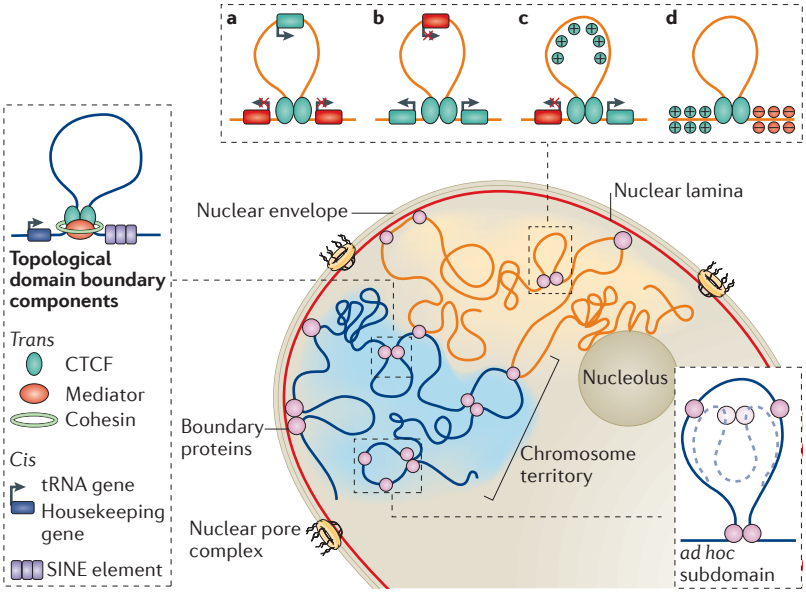


Figure 3 | **Model for the organization of topological domains in the genome.** A conceptual schematic showing two chromosomes (shown as orange and blue lines) compartmentalized in their respective chromosome territories (shown as orange and blue shaded areas). Chromatin boundary proteins partition the genome into topologically distinct domains by organizing intra- and inter-chromosomal interactions, as well as interactions between chromatin and the nuclear lamina. The panel on the left shows a schematic summary of the *cis* and *trans* components enriched in topological domain boundaries as identified in recent studies[47,63,66]. It is not necessary for a boundary to contain all components. CCCTC-binding factor (CTCF) is found in most boundaries of topological domains. The panel at the top depicts four types of CTCF-mediated topological domains, grouped on the basis of histone modification profiles in the topological domains and its neighbouring regions[67]. The observations suggest a common role of CTCF-mediated topological domains in ensuring regulatory independence, including independent transcription of neighbouring genes (**a**–**c**) and separation of open and repressive chromatin (**d**). Active and silent genes are depicted in green and red, respectively. Plus and minus signs represent open and repressive chromatin, respectively. The panel on the right shows a model of an ad hoc subdomain that is organized by internal boundaries. Images in parts **a**–**d** are modified, with permission, from REF. 61 © (2011) Macmillan Publishers Ltd. All rights reserved. SINE, short interspersed element.

of all possible genomic interactions, Hi-C requires extremely high sequencing depth to reach a satisfactory resolution. Recently, a group performed Hi-C analysis in ES cells and differentiated cells with an unprecedented approximately 100-fold increase in sequencing depth[66]. The results revealed that the genome is organized into large, discrete, self-interacting domains (termed topological domains) that are reminiscent of TADs (FIG. 3). Topological domains are invariant between ES cells and differentiated cells, a property that is also shared with TADs. However, dynamic tissue-specific chromatin interactions are observed within topological domains and preferentially involve differentially regulated genes[66]. Topological domains are highly conserved across species, indicating that these higher-order chromatin structures represent a fundamental organizing principle of the mammalian genome. This is especially surprising considering the lack of evolutionary conservation in other epigenetic features in the linear genome (for example, histone modifications). Topological domains also provide a reference map to which known genomic domains can be compared. For example, topological domains seem to demarcate heterochromatin domains and LADs in some cases, but they do not always overlap with these and other previously identified domain-like structures[66]. Not surprisingly, the boundaries of topological domains are enriched for CTCF. In fact, the data validated an experimentally defined insulator, termed CBS5 (FIG. 2b), as a bona fide boundary element that partitions two topological domains. However, as in the case of TADs, many CTCF-binding sites are located within the topological domains, which suggests two possibilities: CTCF alone is not sufficient to establish a boundary; or these internal CTCF-binding sites could organize *ad hoc* subdomains that exist transiently under certain circumstances, which may not be detectable by Hi-C (FIG. 3). Other than CTCF binding, the boundary regions show enrichment of additional genomic features such as housekeeping genes, tRNA genes and short interspersed element (SINE) retrotransposons (FIG. 3). Future studies on these sequences may shed new light on the organizing principles of the genome. The ongoing technological innovation in high-resolution mapping of higher-order chromatin structure will certainly provide novel insights into how three-dimensional chromatin fibres are woven into the programme of cellular differentiation.

Chromatin and nuclear lamina interplay

Both local chromatin looping and large-scale nuclear reorganization entail an interaction between chromatin and nuclear matrix. The nuclear lamina (BOX 3), a major part of the nuclear matrix, regulates gene expression by either providing an inhibitory chromatin environment or by interacting with DNA at the nuclear periphery. The maintenance of pluripotency and the initiation of differentiation involve an orchestrated change of the composition, localization and the activity of various nuclear lamina components and requires a highly dynamic interaction between the nuclear envelope and the genome (FIG. 4).

Nuclear lamina: a hub of chromatin organization. Using DamID (DNA adenine methyltransferase identification) (BOX 3), a group recently mapped genome-wide LADs in mouse ES cells and their neural progenitor and astrocyte derivatives, as well as in MEFs[67]. Although most lamina-associated genes remain stably associated with the nuclear lamina during differentiation, specific organization patterns change during lineage switch. For example, following differentiation into neural progenitor cells, induction of neural genes is accompanied by their relocation away from the nuclear envelope, whereas silencing of pluripotency genes correlates with a tight association with the nuclear lamina. In addition, cell cycle genes become increasingly associated with the nuclear envelope when they are silenced in astrocytes that exit mitosis[67]. Similarly, the binding of lamin B to promoters negatively correlates with gene expression levels in the context of trophectoderm differentiation of mouse ES cells[68]. Thus, dynamic interactions between LADs and the nuclear lamina provide a spatial layer of control of gene transcription in lineage commitment (FIG. 4).

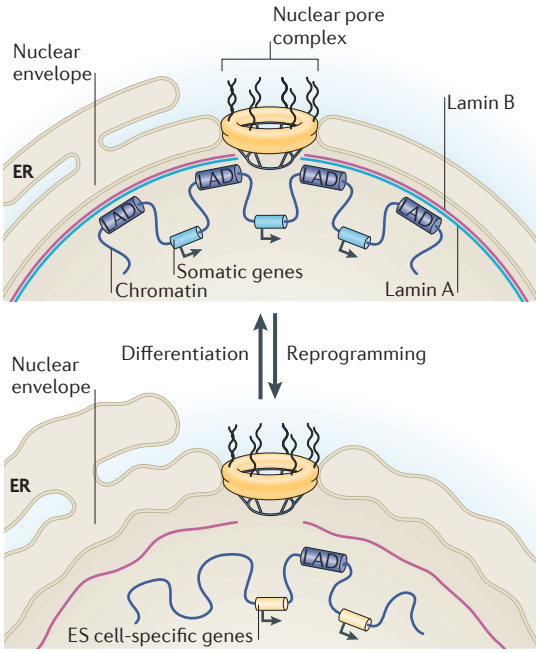


Figure 4 | **Dynamic interplay between the nuclear envelope and chromatin during reprogramming and differentiation.** In differentiated cells, the nuclear envelope is regular and round. Nuclear lamina proteins, including A-type and B-type lamins, are tightly arranged along the inner nuclear membrane. Accordingly, the genome is organized into lamina-associated domains (LADs), which constitute the majority of heterochromatin at the nuclear periphery. Most differentiation-associated genes are located in inter-LADs away from the inner nuclear membrane and remain transcriptionally active. By contrast, pluripotent stem cells do not express A-type lamins, and the association of B-type lamins to chromatin is dynamic. This is associated with a pliable and deformed nuclear envelope morphology. Moreover, the intermembrane space of the nuclear envelope is wider, and irregular. As a result, the genome is loosely organized and heterochromatin is seldom observed at the nuclear periphery. In addition, the stem cell-specific genes are not localized in LADs and become activated. ER, endoplasmic reticulum; ES cell, embryonic stem cell.

Nuclear envelope composition and stemness regulation. Dynamic morphological and compositional changes of the nuclear envelope during differentiation and reprogramming suggest that the mammalian nuclear envelope is highly plastic[69–75]. Pluripotent stem cells have a different nuclear lamina protein composition than somatic cells. Undifferentiated human ES cells express B-type but not A-type lamins[74,76]. Expression of *LMNA*, the gene encoding A-type lamins, is also silenced in iPS cells that are reprogrammed from human skin fibroblasts[74,77,78]. During ES cell differentiation, *LMNA* induction occurs before silencing of the pluripotency gene encoding OCT4 (REF. 76), suggesting that nuclear architecture change precedes the loss of pluripotency. By contrast, B-type lamins are constitutively expressed in pluripotent stem cells. Interestingly, the association between lamin B and chromatin is more dynamic in ES cells than in differentiated cells[79]. The absence of A-type lamins and increased mobility of B-type lamins might contribute to a less rigid nuclear structure. Physically, ES cell nuclei are highly pliable, but become stiff upon terminal differentiation[80]. When grown as colonies, human iPS cells exhibit a deformed nuclear lamina compared to their fibroblast counterparts[74]. Ultrastructural analysis shows that the intermembrane space of the nuclear envelope is more irregular and wider in ES cells than in differentiated cells[81]. These pluripotent stem cell-specific characteristics of the nuclear envelope could provide a structural basis at the nuclear periphery for an 'open' chromatin conformation that is conducive to transcription and facilitates the maintenance of pluripotency[80].

Consistent with a role of the nuclear lamina in maintaining chromatin conformation, nuclear lamina components are known to be crucial regulators of stem cell differentiation and development. Although *LMNA* is not required for ES cell identity, its expression is crucial for the differentiation towards various lineages including bone and muscle[82,83]. Interestingly, a recent study has indicated that lamins are dispensable for mouse ES cell self-renewal and pluripotency[68]. In *Lmnb1*$^{-/-}$ and *Lmnb2*$^{-/-}$ mouse ES cells, which lack lamin proteins, nuclear pores and the nuclear envelope seem normal. These engineered ES cells exhibit normal morphology, growth rate and ploidy and differentiate into trophectoderm cells at normal efficiency. In contrast to the lack of phenotype in ES cells, mice lacking lamin B1 and lamin B2 have defects in the diaphragm, lungs and brain, and they die after birth[68]. These observations highlight the distinct requirements of lamin-based nuclear architecture in different cellular programmes and are consistent with the dynamic changes of LADs during development[61,67]. Considering the hyperdynamic exchange of chromatin components and the fluid nature of the nuclear envelope in pluripotent stem cells, it is not entirely surprising that anchoring to the nuclear lamina is not the predominant mechanism of gene silencing. On the other hand, terminally differentiated cells may rely on LADs to stably switch off gene expression programmes of alternative lineages by forming repressive chromatin domains.

Epigenetic pathways and human diseases

Many human diseases are caused by epigenetic pathways that have gone awry. Pluripotent stem cells have proved a useful platform for modelling these diseases *in vitro*. The autism spectrum disorder Rett syndrome (RTT) is an example for the successful modelling of a disease by using patient iPS cells[84]. The genetic defect of RTT lies in methyl-CpG-binding protein 2 (MeCP2), which binds to 5mC and exerts a variety of effects on transcription[85]. A group recently generated iPS cells from fibroblasts of patients with RTT, differentiated the iPS cells into functional neurons and uncovered many neuron-specific defects[84]. Interestingly, the same group also found an increased susceptibility to LINE 1 (long interspersed element 1) retrotransposition in mutant MeCP2 neurons, providing novel insights into RTT pathology[86].

One of the most dramatic examples is Hutchinson–Gilford progeria syndrome (HGPS), which is a premature ageing disorder caused by a point mutation in *LMNA*. This mutation creates a cryptic splicing site that leads to the generation of progerin, which is a truncated form of lamin A. Accumulated expression of progerin in mesodermal cells results in progressive disorganization of the nuclear lamina and loss of heterochromatin marks such as H3K9me3 (REF. 74). Recently, three groups, including ours, have generated iPS cells from fibroblasts from patients with HGPS[74,78,87]. We observe that the nuclear defects, including disorganized nuclear lamina and epigenetic aberrance manifested in HGPS fibroblasts, are erased once the cells have been reprogrammed into iPS cells (in which *LMNA* is not expressed) but reappear in differentiated iPS cells[74,78]. Thus, HGPS–iPS cells present a good platform to study the link among chromatin organization, nuclear lamina and accelerated human ageing.

Recently, there have been rapid advances in targeted gene-editing technology in human pluripotent stem cells. The ability to efficiently and precisely modify any region of the genome will greatly improve the precision with which epigenetic pathways can be manipulated and studied. For disease modelling, engineered isogenic pluripotent stem cell lines that differ only at a specific pathogenic mutation are invaluable in filtering out the noise introduced by complex genetic backgrounds when analysing disease phenotypes[88]. Combining targeted gene correction with patient-specific iPS cell technology also promises future cell-based therapies for genetic diseases[89–91].

Conclusion and perspectives

In summary, pluripotent stem cells use a complex network of genetic and epigenetic pathways to maintain a delicate balance between self-renewal and multilineage differentiation. A recurring theme is the interconnectivity of the network, such as the regulation of common targets by opposing epigenetic activities (for example, PcG proteins versus TrxG proteins at bivalent genes, NuRD versus BAF at 5hmC marked genes and LSD1–NuRD versus transcriptional co-activators at active enhancers). Perhaps we can imagine these competing epigenetic influences as thousands of binary switches occurring on genes that can be flipped on or off to programme any cell fate according to the intrinsic transcriptional state and extrinsic cellular environment. This precarious balance between activation and repression may prime pluripotent stem cells for differentiation cues. Somatic reprogramming shows that a few pluripotency-associated transcription factors are sufficient to instigate the pluripotency epigenetic network, although the reprogramming process remains poorly understood. With the abundant data accumulated on epigenetic regulation from the level of naked DNA to that of the chromosome, the next major challenge will be to decode how different epigenetic pathways are choreographed into a coherent programme of cellular identity.

Our current understanding of epigenetics mainly stems from analysing individual pathways by focusing on single genes or single epigenetic marks. In future, we are likely to discover new principles of epigenetic regulation by genome-wide approaches. Thanks to the technological advances in genomic tools, recent years have seen an explosion of genome-wide data on epigenetic marks and transcription. However, extracting novel principles of gene regulation from the vast amount of data remains a tall order. As discussed above, by integrating chromosome interacting maps with existing maps of epigenetic marks and factor binding, new rules of genome organization have emerged. Clearly, any correlative observations should be rigorously examined by independent methodologies. For instance, sequential ChIP and biochemical analysis should be carried out to determine whether the apparent overlap in genomic targets of given factors is due to simultaneous binding or heterogeneity in the population.

Another gap in our knowledge is how three-dimensional chromatin and nuclear organization affect gene regulation. There has been ample evidence of dramatic changes in chromatin and nuclear architecture during ES cell differentiation, yet we know very little about the mechanism and significance of this phenomenon. Systematic mapping of higher-order chromatin interactions and chromatin–matrix interactions using novel technologies such as Hi-C and ChIA-PET are starting to claim this new frontier. With such maps in hand, we could start to study the *cis* and *trans* factors that organize higher-order chromatin structures. More importantly, future efforts should be focused on understanding the significance of higher-order chromatin structure through disrupting the normal chromatin domains. In this regard, the recently developed gene-editing technologies in pluripotent stem cells may be used for targeted knockout of the *cis* or *trans* organizers of higher-order chromatin structure[78,92,93].

A deeper understanding of the mechanisms of epigenetic regulation will also allow us to actively manipulate cell fate conversion, such as differentiation, reprogramming and transdifferentiation, all of which are fundamentally epigenetic processes. There is already a catalogue of epigenetic inhibitors that enhance the generation of iPS cells (Supplementary information S1 (table)). With new insights into epigenetic mechanisms, we could devise more precise ways to coax cells into a desired cell fate. In turn, these insights could provide better tools to unravel the dynamics of the epigenome during cell fate conversion.

1. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
2. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
3. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
4. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
5. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
6. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
7. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**, 247–257 (1999).

8. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
9. Tsumura, A. *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**, 805–814 (2006).
10. Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. *Nature Biotech.* **28**, 1079–1088 (2010).
11. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293**, 1089–1093 (2001).
12. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042–1047 (2010).
13. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
14. Foshay, K. M. *et al.* Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming. *Mol. Cell* **46**, 159–170 (2012).
15. He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
16. Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
17. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nature Rev. Mol. Cell Biol.* **12**, 36–47 (2011).
18. Ang, Y. S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183–197 (2011).
19. Jones, A. & Wang, H. Polycomb repressive complex 2 in embryonic stem cells: an overview. *Protein Cell* **1**, 1056–1062 (2010).
20. Morey, L. *et al.* Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47–62 (2012).
21. O'Loghlen, A. *et al.* MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell* **10**, 33–46 (2012).
Shows, together with reference 20, that differential incorporation of alternative PRC1 subunits allows PRC1 to target different genes in ES cells and differentiated cells.
22. Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**, 447–450 (2007).
23. Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731–734 (2007).
24. Christensen, J. *et al.* RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. *Cell* **128**, 1063–1076 (2007).
25. Pasini, D. *et al.* Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-repressive complex 2. *Genes Dev.* **22**, 1345–1355 (2008).
26. Adamo, A. *et al.* LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nature Cell Biol.* **13**, 652–659 (2011).
27. Whyte, W. A. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221–225 (2012).
28. Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116 (2006).
29. Fussner, E. *et al.* Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.* **30**, 1778–1789 (2011).
30. Gao, Q. *et al.* Telomeric transgenes are silenced in adult mouse tissues and embryo fibroblasts but are expressed in embryonic stem cells. *Stem Cells* **25**, 3085–3092 (2007).
31. Ma, D. K., Chiang, C. H., Ponnusamy, K., Ming, G. L. & Song, H. G9a and Jhdm2a regulate embryonic stem cell fusion-induced reprogramming of adult neural stem cells. *Stem Cells* **26**, 2131–2141 (2008).
32. Shi, Y. *et al.* Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* **3**, 568–574 (2008).
33. Epsztejn-Litman, S. *et al. De novo* DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature Struct. Mol. Biol.* **15**, 1176–1183 (2008).
34. Lienert, F. *et al.* Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS Genet.* **7**, e1002090 (2011).
35. Davie, J. R. & Chadee, D. N. Regulation and regulatory parameters of histone modifications. *J. Cell Biochem. Suppl.* **30–31**, 203–213 (1998).
36. Kidder, B. L., Palmer, S. & Knott, J. G. SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells* **27**, 317–328 (2009).
37. Yan, Z. *et al.* BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. *Stem Cells* **26**, 1155–1165 (2008).
38. Gao, X. *et al.* ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc. Natl Acad. Sci. USA* **105**, 6656–6661 (2008).
39. Ho, L. *et al.* esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nature Cell Biol.* **13**, 903–913 (2011).
40. Fazzio, T. G. & Panning, B. Control of embryonic stem cell identity by nucleosome remodeling enzymes. *Curr. Opin. Genet. Dev.* **20**, 500–504 (2010).
41. Yildirim, O. *et al.* Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* **147**, 1498–1510 (2011).
42. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* **10**, 1453–1465 (2002).
43. Palstra, R. J. *et al.* The β-globin nuclear compartment in development and erythroid differentiation. *Nature Genet.* **35**, 190–194 (2003).
44. Levasseur, D. N., Wang, J., Dorschner, M. O., Stamatoyannopoulos, J. A. & Orkin, S. H. Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev.* **22**, 575–580 (2008).
45. Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N. & Lee, J. T. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128–132 (2009).
46. Kim, Y. J., Cecchini, K. R. & Kim, T. H. Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl Acad. Sci. USA* **108**, 7391–7396 (2011).
Finds that OCT4 regulates the formation of chromatin loop domains by modulating the binding of cohesin, and that perturbation of higher-order chromatin structure results in dramatic locus-wide gene expression changes.
47. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
48. Bertani, S., Sauer, S., Bolotin, E. & Sauer, F. The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol. Cell* **43**, 1040–1046 (2011).
49. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
50. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* **106**, 11667–11672 (2009).
51. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
52. Morey, C., Da Silva, N. R., Perry, P. & Bickmore, W. A. Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development* **134**, 909–919 (2007).
53. Chambeyron, S., Da Silva, N. R., Lawson, K. A. & Bickmore, W. A. Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development* **132**, 2215–2223 (2005).
54. Chambeyron, S. & Bickmore, W. A. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130 (2004).
55. Bartova, E., Krejci, J., Harnicarova, A. & Kozubek, S. Differentiation of human embryonic stem cells induces condensation of chromosome territories and formation of heterochromatin protein 1 foci. *Differentiation* **76**, 24–32 (2008).
56. Wiblin, A. E., Cui, W., Clark, A. J. & Bickmore, W. A. Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. *J. Cell Sci.* **118**, 3861–3868 (2005).
57. Harewood, L. *et al.* The effect of translocation-induced nuclear reorganization on gene expression. *Genome Res.* **20**, 554–564 (2010).
58. Morey, C., Kress, C. & Bickmore, W. A. Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. *Genome Res.* **19**, 1184–1194 (2009).
59. Morey, C., Da Silva, N. R., Kmita, M., Duboule, D. & Bickmore, W. A. Ectopic nuclear reorganisation driven by a Hoxb1 transgene transposed into Hoxd. *J. Cell Sci.* **121**, 571–577 (2008).
60. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nature Methods* **4**, 895–901 (2007).
61. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genet.* **43**, 630–638 (2011).
Maps genome-wide CTCF-mediated chromatin interactions in mouse ES cells by using ChIA-PET. Identifies five types of distinct CTCF-associated chromatin domains and a surprising role of CTCF in bridging enhancers and promoters.
62. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11–24 (2012).
63. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
Investigates the spatial genome organization of the XIC using 5C in mouse ES cells, neuronal progenitors and MEFs. Reveals several of topologically distinct domains, termed topologically associating domains.
64. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
65. Dostie, J. *et al.* Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
66. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
Reveals that the genome is organized into large, discrete, self-interacting topological domains by using Hi-C analysis in ES cells and differentiated cells. Topological domains are evolutionarily conserved.
67. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38**, 603–613 (2010).
Shows that expression levels of pluripotency or lineage-specific genes are negatively correlated to the interaction strengths between these gene loci and the nuclear lamina during mouse ES cell differentiation.
68. Kim, Y. *et al.* Mouse B-type lamins are required for proper organogenesis but not by embryonic stem cells. *Science* **334**, 1706–1710 (2011).
Shows that lamin proteins are not required for self-renewal and pluripotency of mouse ES cells.
69. Meshorer, E. & Misteli, T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nature Rev. Mol. Cell Biol.* **7**, 540–546 (2006).
70. Tilgner, K., Wojciechowicz, K., Jahoda, C., Hutchison, C. & Markiewicz, E. Dynamic complexes of A-type lamins and emerin influence adipogenic capacity of the cell via nucleocytoplasmic distribution of β-catenin. *J. Cell Sci.* **122**, 401–413 (2009).
71. Zhang, X. *et al.* SUN1/2 and Syne/Nesprin-1/2 complexes connect centrosome to the nucleus during neurogenesis and neuronal migration in mice. *Neuron* **64**, 173–187 (2009).
72. Dauer, W. T. & Worman, H. J. The nuclear envelope as a signaling node in development and disease. *Dev. Cell* **17**, 626–638 (2009).
73. Asally, M. *et al.* Nup358, a nucleoporin, functions as a key determinant of the nuclear pore complex structure remodeling during skeletal myogenesis. *FEBS J.* **278**, 610–621 (2011).
74. Liu, G. H. *et al.* Recapitulation of premature ageing with iPSCs from Hutchinson–Gilford progeria syndrome. *Nature* **472**, 221–225 (2011).
Shows that premature ageing-associated nuclear envelope defects and epigenetic abnormalities are reset during reprogramming of HGPS fibroblasts towards iPS cells, and that they reoccur after differentiation towards mesodermal cells.

75. Verstraeten, V. L. *et al.* Reorganization of the nuclear lamina and cytoskeleton in adipogenesis. *Histochem. Cell Biol.* **135**, 251–261 (2011).
76. Constantinescu, D., Gray, H. L., Sammak, P. J., Schatten, G. P. & Csoka, A. B. Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. *Stem Cells* **24**, 177–185 (2006).
77. Zhang, J. *et al.* A human iPSC model of Hutchinson Gilford Progeria reveals vascular smooth muscle and mesenchymal stem cell defects. *Cell Stem Cell* **8**, 31–45 (2011).
78. Liu, G. H. *et al.* Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs. *Cell Stem Cell* **8**, 688–694 (2011).
79. Bhattacharya, D., Talwar, S., Mazumder, A. & Shivashankar, G. V. Spatio-temporal plasticity in chromatin organization in mouse cell differentiation and during *Drosophila* embryogenesis. *Biophys. J.* **96**, 3832–3839 (2009).
80. Pajerowski, J. D., Dahl, K. N., Zhong, F. L., Sammak, P. J. & Discher, D. E. Physical plasticity of the nucleus in stem cell differentiation. *Proc. Natl Acad. Sci. USA* **104**, 15619–15624 (2007).
81. Smith, E. R., Zhang, X. Y., Capo-Chichi, C. D., Chen, X. & Xu, X. X. Increased expression of Syne1/nesprin-1 facilitates nuclear envelope structure changes in embryonic stem cell differentiation. *Dev. Dyn.* **240**, 2245–2255 (2011).
82. Sullivan, T. *et al.* Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J. Cell Biol.* **147**, 913–920 (1999).
83. Li, W. *et al.* Decreased bone formation and osteopenia in lamin a/c-deficient mice. *PLoS ONE* **6**, e19313 (2011).
84. Marchetto, M. C. *et al.* A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
85. Samaco, R. C. & Neul, J. L. Complexities of Rett syndrome and MeCP2. *J. Neurosci.* **31**, 7951–7959 (2011).
86. Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468**, 443–446 (2010).
87. Ho, J. C. *et al.* Generation of induced pluripotent stem cell lines from 3 distinct laminopathies bearing heterogeneous mutations in lamin A/C. *Aging (Albany NY)* **3**, 380–390 (2011).
88. Soldner, F. *et al.* Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* **146**, 318–331 (2011).
89. Hockemeyer, D. & Jaenisch, R. Gene targeting in human pluripotent cells. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 201–209 (2010).
90. Liu, G. H., Sancho-Martinez, I. & Izpisua Belmonte, J. C. Cut and paste: restoring cellular function by gene correction. *Cell Res.* **22**, 283–284 (2012).
91. Sancho-Martinez, I., Li, M. & Izpisua Belmonte, J. C. Disease correction the iPSC way: advances in iPSC-based therapy. *Clin. Pharmacol. Ther.* **89**, 746–749 (2011).
92. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nature Biotech.* **29**, 731–734 (2011).
93. Li, M. *et al.* Efficient correction of hemoglobinopathy-causing mutations by homologous recombination in integration-free patient iPSCs. *Cell Res.* **21**, 1740–1744 (2011).
Describes, together with references 80 and 94, two novel gene-editing technologies, which can be used to efficiently engineer human pluripotent stem cells to manipulate and study epigenetic pathways.
94. 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).
95. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature Cell Biol.* **9**, 1428–1435 (2007).
96. Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
97. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24**, 265–276 (2010).
98. Chamberlain, S. J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **26**, 1496–1505 (2008).
99. Pasini, D. *et al.* Regulation of stem cell differentiation by histone methyltransferases and demethylases. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 253–263 (2008).
100. Zhao, X. D. *et al.* Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286–298 (2007).
101. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
102. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
103. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
104. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
105. Schneider, R. & Grosschedl, R. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev.* **21**, 3027–3043 (2007).
106. Dechat, T. *et al.* Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. *Genes Dev.* **22**, 832–853 (2008).
107. Masny, P. S. *et al.* Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum. Mol. Genet.* **13**, 1857–1871 (2004).
108. Nikolova, V. *et al.* Defects in nuclear structure and function promote dilated cardiomyopathy in lamin A/C-deficient mice. *J. Clin. Invest.* **113**, 357–369 (2004).
109. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).

Acknowledgements
The authors apologize to those colleagues whose work could not be cited due to space constraint. We would like to thank M. J. Barrero, N.Y. Kim, M. Schwarz, P. Schwarz and I. Dubova for critical reading of the manuscript. We thank N.Y. Kim for rendering the schematic model of topological domains. G.H.L. was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences. This study was supported by grants from the G. Harold and Leila Y. Mathers Charitable Foundation, Sanofi, Ellison Medical Foundation, The Leona M. and Harry B. Helmsley Charitable Trust, the Glenn Foundation for Medical Research, Ministerio de Economía y Competitividad (MINECO) and Fundacion Cellex (J.C.I.B.).

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Juan Carlos Izpisua Belmonte's homepage:
http://www.salk.edu/faculty/belmonte.html

SUPPLEMENTARY INFORMATION
See online article: S1 (table 1)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF