The epigenetic basis for embryonic stem cell pluripotency

Henrietta Szutorisz and Niall Dillon*

Summary
As well as having the remarkable ability to differentiate into all of the cell types in the embryo, embryonic stem (ES) cells also have the capacity to divide and self-renew. Maintenance of pluripotency through repeated cell divisions indicates that the developmental plasticity of ES cells has a specific epigenetic basis. We propose that tightly localised regions of histone modification are formed in ES cells by binding of sequence-specific transcription factors at genes that are destined for expression at later stages of differentiation. These ‘early transcription competence marks’ would help to maintain pluripotency by preventing the spread of repressive chromatin modifications. We further propose that the presence of discrete histone modification marks in pluripotent cells facilitates the binding of lineage-specific and general transcription factors to the marked regions as ES cells commit to different fates. By helping to organise the precisely timed responses of genes to the signals that determine lineage choice, the gene-specific localised epigenetic marks would play a key role in the establishment of complex gene expression programmes in differentiating cells. BioEssays 27:1286–1293, 2005. © 2005 Wiley Periodicals, Inc.

Introduction
The diverse cell types that are found in a mammal are generated from a single fertilised egg which contains copies of the paternal and maternal genomes derived from the haploid sperm and oocyte. Development occurs through the establishment of specialised gene expression programmes as cells acquire very different properties and morphologies, depending on which differentiation pathway is followed. The essential components of cell differentiation are the expansion of primordial cells through sequential divisions and a gradual restriction of differentiation potential as cells commit to various fates (for review, see Ref. 1). Thus, the fertilised egg and the cells of the 2- and 4-cell embryo are totipotent and can give rise to every cell type in the body (Fig. 1). By the time that the blastocyst stage is reached, some restriction has already occurred. The cells of the inner cell mass (ICM) are still able to give rise to all somatic lineages, but lack the ability to differentiate into the placenta and the extraembryonic membranes. The latter tissues are derived from the cells of the hollow sphere that forms the bulk of the blastocyst (Fig. 1).

A critical change occurs in the period between fertilisation of the egg and formation of the blastocyst. The germ cells (the sperm and the oocyte) and the fertilised egg, are single cell transitional stages that lack the ability to self-renew. In other words, an egg cannot divide to give rise to more eggs. After fertilisation, totipotency (see Box 1) is maintained for the first few cleavage divisions, but the embryo must then progress to the next developmental stage. In contrast, cells can be isolated from the inner cell mass of mouse blastocysts that are able to divide indefinitely while maintaining a consistent phenotype that includes the ability to differentiate into all somatic cell types. These pluripotent cells are called embryonic stem (ES) cells (see Box 1). Their ability to give rise to multiple differentiated cell types has made them the subject of enormous scientific and clinical interest but the basis for this capability remains enigmatic.

Here we examine the different types of model that can explain pluripotency at the level of the gene. We propose a novel epigenetic model to explain how genes are primed in ES cells to express at later stages of development. According to this model, transcription competence of lineage-specific genes is determined by localised marks of histone modification established through binding of sequence-specific factors to discrete regions at the ES cell stage. We suggest that these gene-specific marks act as recruitment centres for transcriptional activators as ES cells commit to different cell lineages.

Is differentiation of ES cells regulated solely by a hierarchical network of transcription factors? In contrast to the specific gene expression programmes that are observed in differentiated cells, pluripotent embryonic stem cells are defined by their potential to activate all of the gene expression programmes that are found in embryonic and adult cell lineages (reviewed by 4). The primary determinants...
Figure 1. Differentiation potential of cells in the mouse embryo at successive stages of pre-implantation development. Dark red shading of the cells corresponds to high levels of DNA methylation, fading red to decreased methylation, and light yellow to the lowest levels. The earliest stages of development involve sequential cleavage divisions during which totipotency is maintained. The differentiation potential of cells in the 8-cell embryo and the morula has not been fully defined. A global demethylation of DNA is observed during the early cleavage stages of embryonic development and methylation reaches the lowest level in the morula. At the blastocyst stage, the DNA methylation state of the trophectoderm is similar to that of the morula, whereas a dramatic increase in methylation is detectable in the cells of the inner cell mass (ICM). In contrast to the totipotent cells at the earlier transitional cleavage stages of pre-implantation development, embryonic stem (ES) cells isolated from the ICM can continue to divide and self-renew while maintaining a relatively stable, pluripotent phenotype.
Hypotheses

Box 1
Histone modifications: Covalent modifications of the amino-terminal tails of the core histones that form the nucleosomes are thought to play a major part in the epigenetic regulation of gene expression.\(^{25–27}\) Acetylation of multiple lysine (K) residues in histone H3 and H4 and methylation of H3 lysine 4 (K4) are known to be associated with transcriptionally active states. Methylation of H3 lysine 9 (K9) and lysine 27 (K27) have been shown to be associated with gene silencing. Modifications of specific histone tail residues are recognised by transcription factors thereby facilitating the binding of activator and repressor proteins to regions of chromatin that contain these marks.\(^{28–30}\)

Totipotency: Describes the capacity of the fertilised egg and the cells of the 2- and 4-cell embryo to give rise to every cell type including the embryonic (somatic) lineages and the extra-embryonic lineages that give rise to the placenta.

Pluripotency: Refers to the ability of ES cells to differentiate into all somatic cell-types but not into the extra-embryonic (trophectoderm) lineages. ES cells can self-renew, which means that they are able to undergo a large number of divisions while maintaining their pluripotency.\(^{4}\)

Multipotency: Multipotent stem cells have the capacity to self-renew but their differentiation capacity is restricted to a subset of cell lineages. For example, haematopoietic stem cells can develop into various types of blood cell but not into nerve or muscle cells (reviewed in Ref. 5).

of cellular identity are the transcription factors that are expressed in a cell and define its gene expression programme by activating or silencing the expression of target genes. Transcription factors that are expressed specifically in ES cells and are required for the maintenance of pluripotency include Oct-4, Nanog and Sox-2 and it seems likely that more factors will be added to this list as the ES cell proteome is characterised in greater detail.

The role of ES-cell-specific transcription factors in activating multiple gene expression programmes during cell differentiation could be explained by a straightforward hierarchical gene activation model (Fig. 2A). This type of model proposes that tissue-specific genes (including those that code for lineage-specific transcription factors) are kept in a silent state in ES cells and that this transcriptionally inactive state is reinforced by repressive chromatin modifications. Different types of signals that initiate commitment of cells along a particular cell lineage would allow ES cell-specific factors to trigger a cascade of gene activation events. Expression of one lineage-specific factor would lead to sequential activation of further sets of factors at each stage of differentiation. According to such a hierarchical activation model, the differentiation potential of ES cells is essentially a programming phenomenon that involves a tightly determined order of expression of transcription factors interacting with external signals at each developmental stage. Sequential expression of lineage-specific factors would be necessary and sufficient for activating further tissue-specific genes without the need for positive marking of transcriptional competence at the epigenetic level.

Hierarchical models are already part of a well-established paradigm for lineage choice during differentiation of multipotent stem cells (e.g. haematopoietic stem cells, see Ref. 5 for review). They also fit well with ideas on genetic networks acting during development that have been established in model organisms such as Drosophila and C. elegans (reviewed in 6). However, there is a significant amount of evidence that ES cell pluripotency is not simply a consequence of the cell being at one end of a differentiation and gene expression cascade. Nuclear transfer experiments have shown that ES cell nuclei give much higher cloning efficiencies than differentiated cells,\(^{7}\) implying that ES cell-specific genes have epigenetic properties that make them easier to reprogramme. This idea is reinforced by the observation that endogenous genes in differentiated cells are generally difficult to reactivate even when appropriate transcription factors are expressed. Taken together, these observations suggest that there are additional mechanisms that make genes in ES cells more amenable to transcriptional activation.

Early gene marking models
Epigenetic marking of tissue-specific genes at the ES cell stage offers a potential explanation for the observation that the ES cell genome is more amenable to reprogramming than genes in differentiated cells (Fig. 2B,C). Candidates for the generation of these ‘positive’ epigenetic marks include modifications of the histones that form the nucleosomal core (see Box 1). In addition, the developmental plasticity of ES cells is a stable phenotype that must be maintained through repeated cell cycles as the cells divide and self-renew. The presence of early (ES cell-specific) epigenetic marks would offer a solution to the problem of how ES cell plasticity is propagated.

One possible early epigenetic modification model would involve a widespread and non-specific increase in activating histone modifications in ES cell chromatin (Fig. 2B). This would make the chromatin generally accessible to transcription factors and would facilitate activation of the genes. Although a global increase in histone acetylation and H3 K4 methylation has been observed by immunofluorescence and western blotting in ES cells, it has been suggested that this could be due to increased transcription of B2 repetitive elements.\(^{8}\) Studies showing that ES cell chromatin contains significant
amounts of repressive epigenetic modifications including DNA methylation and histone H3 K9 and K27 methylation\(^{(9–11)}\) also raise the question of how active and repressive modifications are able to co-exist in the ES cell genome. This problem is further highlighted by a recent analysis of replication timing in ES cells that showed that 15 out of 43 genes analysed were late replicating and presumably have a chromatin configuration that is at least partially repressed.\(^{(12)}\)
**Localised epigenetic marking as a determinant of transcriptional competence of lineage-specific genes in ES cells**

As an alternative to hierarchical activation and general accessibility models, we propose a gene-specific early marking model (illustrated in Figs. 2C and 4). The essential feature of this model is that, instead of ES cell factors acting on a small subset of genes to initiate a sequence of activation events that take place as cells progress through differentiation, most or all tissue-specific genes would be targets for binding of sequence-specific factors at the ES cell stage. This binding would lead to recruitment of histone-modifying enzymes that would generate tightly localised marks and prime the genes for expression at later stages. It is already well established from studies in other cell types that sequence-specific factors can recruit histone-modifying enzymes to generate epigenetic marks in the vicinity of genes (reviewed in 13), and some ES cell-specific transcription factors (including factors that have not yet been identified) could potentially have similar roles. As ES cells differentiate, binding of further lineage-specific factors to the marked genes would lead to recruitment of chromatin-modifying and other transcriptional activator complexes. As a consequence of the activity of these factors, a larger epigenetically modified region would be established in multipotent stem cells, expanding the mark at genes that have the potential to be expressed in lineages arising from those cells, while the marks would disappear from genes that are destined for long-term silencing.

Evidence that genes that are destined for expression at later stages of development are subject to specific epigenetic marking in ES cells comes from a recent study of the chromatin structure of the mouse λ.5-VpreB1 locus (Fig. 3A). The λ.5 and VpreB1 genes are expressed in pro- and pre-B cells where they encode the components of the surrogate light-chain. Chromatin immunoprecipitation (ChIP) analysis of ES cell chromatin showed no evidence of general acetylation across the 19 kb λ.5-VpreB1 domain. However, the results did show that an enhancer element located between the λ.5 and VpreB1 genes is already marked by a localised region of histone H3 acetylation and H3 K4 methylation in ES cells (Fig. 3A). The modified region, which we have called the Early Transcription Competence Mark (ETCM), expands and becomes hypersensitive to DNase I digestion in early pro-B cells before the λ.5-VpreB1 locus is activated. In pre-B cells where the genes

![Figure 3](image-url)

*Figure 3.* Tissue-specific gene loci where localised epigenetic marks have been identified at the ES cell stage. Red arrowheads show the positions of the histone modification marks. Vertical arrows indicate the positions of DNase I hypersensitive sites (HS). **A:** The λ.5-VpreB1 domain. The ETCM region co-localises with HS8 and forms a centre for recruitment of tissue-specific and general transcription factors at later stages of B cell differentiation. **B:** The mouse β-globin locus. LCR = locus control region. Blue arrow indicates the direction of activation of the genes during development. **C:** The Hoxb cluster. The genes are activated 3’ to 5’ (blue arrow) during gastrulation in vertebrate development. It should be noted that the relatively low resolution ChIP assays used to analyse epigenetic marking in the β-globin and Hoxb loci could mean that there are additional undetected marks elsewhere in these loci.
are expressed, the region of histone modification expands to include the entire λ5-VpreB1 domain. The ETCM region is also a centre for recruitment of general transcription factors and RNA polymerase II from the ES cell stage through to pre-B cells, suggesting that the mark facilitates binding of factors to a specific regulatory element.

Localised marking has also been observed in ES cells at other gene loci. In the locus control region (LCR) of the mouse β-globin locus, histone acetylation, H3 K4 methylation and RNA Pol II binding have been detected in ES cells at an enhancer located at DNase I hypersensitive site 2 (HS2)\(^{(15,16)}\) (Fig. 3B). It has been shown that Pol II and the basal transcription factors TBP and TFII B are specifically recruited to the marked HS2 region prior to erythroid differentiation.\(^{(16)}\) Analysis of histone modifications in the mouse Hox locus has also detected the presence of a discrete region of histone H3 K4 methylation in undifferentiated ES cells before the genes are expressed (Fig. 3C). The marked region is located in the promoter and the first exon of the Hoxb9 gene, which is expressed at a relatively late stage of differentiation.\(^{(17)}\)

We hypothesise that localised epigenetic marks of the type described above are involved in priming genes in ES cells for expression at later stages of development (Fig. 4). By determining which gene expression programmes can be activated during differentiation, this gene-specific marking would effectively define the differentiation potential of a stem cell.

**Localised marking and ES cell pluripotency**

Localised epigenetic marking (Fig. 4) offers a potential solution to the conundrum of how rapidly dividing ES cells are able to keep genes in a transcriptionally competent state and remain in a pluripotent state for long periods. Acquisition of the marks would occur at the time when the cells are becoming committed to a somatic embryonic fate and would be mediated by binding of sequence-specific transcription factors to the region. Specific epigenetic marking by histone modification is already known to occur in multipotent stem cells and to be a consequence of the binding of transcription factors that are involved in lineage choice.\(^{(18–21)}\) According to our model, transcription factors that are expressed in ES cells (gene C in Fig. 4) would have a similar role in establishing epigenetic marks. These factors could be specific to ES cells, or they could be ubiquitous, or a combination of both. In this context, it is notable that binding of the sequence-specific pluripotency factor Oct-4 is essential for enhancer activity in the Sox-2 locus in ES cells.\(^{(22)}\) The Sox-2 enhancer is also active in neural cells and it has been suggested that Oct-4 initiates Sox2 transcription in ES cells and that differentiation of ES cells into neural

![Figure 4. Epigenetic basis for ES cell pluripotency proposed by the localised marking model. Formation of Early Transcription Competence Marks (ETCMs) in ES cells at tissue-specific genes A and B is the result of sequence-specific binding of a transcription factor (green oval, product of transcribed gene C) which in turn recruits histone-modifying enzymes (not shown). Where the marked regions are located at promoters, their activating effects could be counteracted by binding of repressors\(^{(24)}\) (not shown). The chromatin around the marks (light red) remains condensed and may be subject to modifications that help to keep genes inactive at this early stage. Maintenance and expansion of the marked regions would allow them to form centres for recruitment of lineage-specific and general transcription factors (coloured ovals). Silencing of genes that are not expressed in that lineage would occur through loss of the mark, which would allow spreading of additional repressive chromatin modifications (dark red zigzag). Genes that are specific for ES cells (gene C) would also be subject to long-term silencing in multipotent stem cells and differentiated cells through the acquisition of repressive chromatin marks.](image-url)
cells is accompanied by a transition to neural-specific stimulation of transcription by POU factors binding to the same sites.\(^{(22)}\)

Once they have been established by specific factors, the histone modification marks would have several different effects. One potential effect would be to maintain a local region of open chromatin in the immediate vicinity of the marks (Fig. 4, genes A and B). This could have a wider impact on the higher order chromatin structure of neighbouring sequences, thereby affecting the accessibility of marked genes to lineage-specific factors when differentiation begins. It has already been shown that a relatively short region of accessible chromatin at a DNase I hypersensitive site can disrupt the higher order organisation of nearby heterochromatin.\(^{(23)}\)

The marks would also prevent the spreading of repressive modifications to regulatory elements that are important for the initiation of gene activation as differentiation progresses. Differentiation of pluripotent ES cells into multipotent stem cells would be accompanied by a decrease in the number of genes carrying activating marks (Figs. 2C and 4, gene B).

In addition to effects on chromatin, we propose that ETCMs act as nucleation centres for factor binding as ES cells differentiate into multipotent stem cells (Fig. 4, gene A). This would be accompanied by expansion of the histone modification marks as binding of lineage-specific factors causes recruitment of histone-modifying enzymes and chromatin-remodelling complexes. The marks would facilitate specific, rapid and ordered activation of gene loci during lineage commitment. A particularly important role for localised marks could be to facilitate the regulation of gene expression programmes that require the precisely timed activation of large numbers of genes at different stages of development. The ability to achieve this may well be essential for correct differentiation of multiple cell lineages from a single pluripotent stem cell.

Conclusions
The ETCM model makes a specific testable prediction, namely that tightly localised regions of H3 acetylation and H3 lysine 4 methylation will be found at or near a high proportion of tissue-specific genes in pluripotent ES cells. The model also predicts that ETCMs will be associated with specific factor binding in ES cells and that they will often coincide with functionally important sequences such as promoters and enhancers. The data from the \(5\text{-VpreB}1\) locus suggest that recruitment of factors involved in lineage choice will result in expansion of the regions covered by the marks in multipotent progenitor cells. Because they are likely to occupy regions of as little as 1 kb, ETCMs will only be detected by high-resolution ChIP across large genomic regions. Tiling path microarrays offer the best prospect for achieving this. Functional analysis of the marked sequences by deletion and mutation is likely to provide information on the significance of the marked elements in the activation of the corresponding tissue-specific genes during cell differentiation. Identification of protein factors that mediate establishment of specific epigenetic marks in ES cells, together with the characterisation of interactions between the marked chromatin regions and the transcription machinery at later stages of differentiation, will also be essential to understand how putative ETCMs contribute to transcriptional competence of genes and the differentiation potential of cells.

Acknowledgments
We thank Pierangela Sabbatini for comments on the manuscript.

References
alterations at the murine c-fms (CSF-1 receptor) locus during maturation of myeloid precursor cells. Genes Dev 16:1721–1737.