Epigenome plasticity during cellular differentiation

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Fabio Mohn
aus Märstetten, TG

Basel, 2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Susan Gasser, Dr. Irina Stancheva, and Dr. Dirk Schübeler.

Basel, den 23. Juni 2009

Professor Dr. Eberhard Parlow
Dekan

creative commons

Attribution-Noncommercial-No Derivative Works 2.5 Switzerland

---

You are free:

to Share — to copy, distribute and transmit the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

Your fair dealing and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license) available in German:
http://creativecommons.org/licenses/by-nc-nd/2.5/ch/legalcode.de

Disclaimer:
The Commons Deed is not a license. It is simply a handy reference for understanding the Legal Code (the full license) — it is a human-readable expression of some of its key terms. Think of it as the user-friendly interface to the Legal Code beneath. This Deed itself has no legal value, and its contents do not appear in the actual license. Creative Commons is not a law firm and does not provide legal services. Distributing of, displaying of, or linking to this Commons Deed does not create an attorney-client relationship.

Quelle: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/deed.en Datum: 3.4.2009

Table of contents

Abbreviations and nomenclature

ES	Embryonic stem cell
DNMT	DNA methyltransferase
MBD	Methyl-CpG-binding domain protein
Pol II	RNA polymerase II
TF	DNA-sequence specific transcription factor
HDAC	Histone deacetylase
KAT	Histone acetyltransferase
KMT	Histone lysine methyltransferase
KDM	Histone lysine demethylase
PRMT	Protein arginine methyltransferase
PcG	Polycomb group protein
PRC	Polycomb repressive complex
PTM	Post-translational histone protein modification
me	any methylation state of an arginine or lysine
me1	mono-methylation
me2	di-methylation
me3	tri-methylation
me2/3	colocalization of di- and tri-methylation of the same lysine
piRNA	Piwi-interacting short RNA species
dpc	days post coitum of mouse embryonic development
ERV	Endogenous retroviral element
IAP	Intracisternal A particle (low copy retroviral-like element)

Protein names	in capitals irrespective of mouse or human origin
Gene names	italic

1. Summary

Tight control of gene expression programs is crucial to govern cell function and identity at any developmental stage. Epigenetic modifications of chromatin have emerged as important determinants for chromatin structure and gene expression. It has been hypothesized that epigenetic mechanisms contribute to the establishment and maintenance of cell type specific gene expression patterns and could stabilize the restriction of developmental potential via repression of genes from distinct cellular lineages. We set out to test this hypothesis by defining genome-wide dynamics and targets of epigenetic reprogramming in a neuronal differentiation system.

DNA methylation, which is a potent and stable repressive modification, is increasing during differentiation of embryonic stem cells into neurons. Many *de novo* methylation targets encode pluripotency-associated and germline specific genes and only few appear to be specific for alternative lineages. Coinciding with *de novo* DNA methylation at CpG-rich promoters we detect a loss of H3K4 methylation, a hallmark of transcriptionally permissive chromatin. This suggests that DNA methylation might lock-in a silent state of early embryonic genes which need to be stably repressed in somatic cells. The unidirectionality of DNA methylation dynamics is furthermore in agreement with the epigenetic restriction model depicted above which predicts an increase of epigenetic repression during cellular differentiation.

To gain further insight into the dynamics of epigenetic repression we subsequently expanded the analysis towards the Polycomb pathway employing the same neuronal differentiation model. Polycomb-mediated repression was previously shown to be essential for embryonic patterning and maintaining developmental potential in stem cells via repression of key developmental transcription factors. During neuronal differentiation, Polycomb repression is resolved at activated genes while novel targets appear at both the multipotential neuronal progenitor state and the terminally differentiated neuron state. Such plasticity was unexpected since current models predicted that Polycomb repression is set up early in development and then either lost upon gene activation or maintained if the gene is not induced. In contrast to this model we find that the number of targets at each state is comparable and progenitor specific targets are enriched for genes expressed in post-mitotic neurons. This suggests that Polycomb is not only important in stem cells and early embryonic development for safe-guarding genes which can be activated upon induction of differentiation. Our results indicate that Polycomb could have a similar role in progenitor cells, which can still develop into a limited number of terminal subtypes.

The genome-wide analysis further revealed that stem cell specific Polycomb targets undergo preferential *de novo* DNA methylation during differentiation, which hints towards a potential interplay between Polycomb and DNA methylation.

In a collaborative effort with the laboratory of Matthew Lorincz we further aimed at addressing a potential interplay between DNA methylation and H3K9 methylation. Using G9a (a KMT mediating euchromatic H3K9me2) knock-out stem cells we observe that DNA methylation is significantly reduced at endogenous retroviral elements (ERVs) and densely methylated CpG-rich promoters. This reduction in methylation did however not lead to transcriptional reactivation of ERVs and suggests that an additional mechanism must exist to silence these repeats. Upon introduction of a catalytic mutant G9a protein into the knock-out stem cells, the DNA methylation defect was partially restored. This indicats that G9a but not its catalytic activity is necessary for efficient DNA methylation at ERVs and promoters. These data reveal a crosstalk between the H3K9 methylation and DNA methylation pathways, which however awaits further characterization.

Our data demonstrate that there are at least two distinct epigenetic modes of repression which nonetheless might crosstalk for target specification. Stable repression is conferred by DNA methylation and/or H3K9 methylation. It is essential for permanent repeat silencing and locking-in silent states throughout further development. In turn a more transient pathway with cell type and developmental stage specific targets is mediated by Polycomb. This mode of repression might confer robustness to specific cellular states via repressing genes that can be activated upon further differentiation but which need to be tightly controlled to avoid precocious activation.

2. Introduction

Development involves tightly regulated differentiation processes during which pluripotent cells in the early embryo give rise to the many distinct cell types of the adult organism. During this phase, developmental potential of the differentiating cells is gradually decreasing and specificity towards a single terminal fate is acquired. This entails dramatic changes of gene expression programs, which need to be adjusted to the respective developmental state and cell function. Importantly, transcriptional programs need to be robust to govern cell identity and function and to avoid dedifferentiation, trans-differentiation or transformation at any state. Conversely in pluripotent stem cells, multipotential progenitor cells or tissue-specific stem cells gene expression patterns also need to be plastic in order to allow changes upon further differentiation. This balance of stability versus plasticity presents an inherent regulatory challenge for developing organisms (Reik, 2007).

The major regulators of gene activity are DNA sequence specific transcription factors (TF). Every cell type has a characteristic set of TFs which orchestrate the cell type specific transcriptional program (Vaquerizas et al., 2009). However, the eukaryotic packaging of DNA into chromatin imposes a steric barrier for TF access to DNA. Therefore it is conceivable that molecular mechanisms which shape or modify chromatin structure reside upstream of DNA sequence based activities and present an opportunity for regulating and stabilizing DNA-templated events such as transcription, replication and repair.

Over the last decade, an increasing body of evidence suggests that chromatin-based regulatory mechanisms in conjunction with sequence specific transcription factors could contribute to establishment and maintenance of cell type specific gene expression programs. The following paragraphs will summarize current knowledge on chromatin structure, chromatin remodeling, transcription initiation and promoter structure, epigenetic modifications of chromatin and their potential interplay during cellular differentiation in relation to my PhD thesis project.

2.1. Mammalian genome structure and its impact on transcription initiation

Over evolutionary time, genome sizes increased in parallel to organismal complexity. Remarkably, gene number increased less dramatically and appears decoupled from total genome size, a phenomenon termed the C-value enigma (Gregory, 2001). For example the roundworm *C. elegans* has a 30 times smaller genome than humans but contains roughly the same number of genes. This illustrates that vertebrate and in particular mammalian genomes accumulated non-coding DNA consisting mostly of remnants of transposable elements, inactive retroviruses and repetitive elements which make up most of the total genome while only about 1.5% of the sequence is protein-coding (Lander et al., 2001). It is still a matter of debate if and how this accumulation might be beneficial for evolution. It was suggested that these apparently parasitic genomic elements must present an evolutionary advantage otherwise they should have been eliminated by selection. For example transposons were implicated in shuffling of transcription factor binding sites conferring rapid evolution of novel regulation and functions to existing genes. Further transposons were suggested to enhance genome plasticity via elevating the frequency of genomic rearrangements (Britten, 1997; Deragon and Capy, 2000). In contrast, it was proposed that the self-replicating nature of transposons is sufficient to explain their existence and maintenance by natural selection without apparent benefits for the host (Doolittle and Sapienza, 1980). In support of that transposon and even a plasmid which reduces fitness of the host became fixed in yeast populations (Futcher et al., 1988; Zeyl et al., 1996). Interestingly, the likelihood for becoming fixed is much higher in sexually reproducing yeast than in asexual populations, which strongly supports the hypothesis that transposon accumulation might simply be a consequence of obligate sexual reproduction (Bestor, 2003). This hypothesis further predicts that sexually reproducing species should have a larger number of transposons than asexually reproducing species. This seems plausible considering that 45% of the human genome is derived form transposons (Lander et al., 2001).

Irrespective of this, the unfavorable ratio between coding and non-coding DNA leads to an increase in transcriptional noise, unspecific initiation and eventual dilution of transcription factor pools due to randomly occurring binding sites. In order assure correct gene expression, this needs to be counteracted. Such a concept of noise suppression was already suggested 50 years ago by C.H. Waddington in an attempt to explain phenotypic stability of species over time despite genotype or environmental variations (Waddington, 1957). Waddington moreover hypothesized that determinants on top of DNA sequence, which he termed “epigenetic”, might confer the observed “buffering” and robustness of phenotypes. A similar idea was later entertained to explain how evolution can cope with

increasing gene number and genome size especially during the two marked changes in organismal complexity: the origin of eukaryotes and later at the origin of vertebrates. It was hypothesized that these two major transitions were only possible due to the parallel evolution of new mechanisms to control transcriptional noise, an unavoidable byproduct of increasing genome complexity (Bestor, 1990; Bird, 1995). The prokaryote-to-eukaryote transition was paralleled by the appearance of nucleosomes, which already greatly lower the chance of aberrant transcription initiation as compared to naked DNA. The later invertebrate-to-vertebrate step was accompanied by genome-wide DNA methylation, which has been shown to be an efficient repression mechanism (Bird, 2002).

Genome-wide DNA methylation had a dramatic impact on the overall genomic sequence since methylated cytosines are prone to spontaneous deamination resulting in the generation of a tymidine. Hence, DNA methylation imposes a much higher mutation rate to CG dinucleotides as compared to any other dinucleotide. This manifests in a global depletion of CpG dinucleotides in vertebrate genomes over time (Bestor and Coxon, 1993). Remarkably, several ten thousands of genomic regions nearly maintained their original CpG content, indicating that they were somehow protected from DNA methylation in germ cells. These CpG-rich sequences were termed CpG islands and remarkably many (up to 70%) mammalian gene promoters contain such CpG-rich sequences [Fig. 1; (Antequera and Bird, 1993; Gardiner-Garden and Frommer, 1987)].

Ubiquitously expressed “house-keeping” genes are in their majority under the control of CpG island promoters, while CpG-poor promoters mostly drive tissue-specific genes (Schug et al., 2005). In addition, large scale CAGE-tag based mapping of transcription start sites (TSS) demonstrated that CpG-rich promoters do rarely have a uniquely defined TSS but transcription rather initiates over an interval of 10-60bp. CpG-poor promoters on the other hand usually have a precisely defined TSS and many rely on TATA-binding protein mediated transcriptional initiation (Bajic et al., 2006; Carninci et al., 2006; Schug et al., 2005; Vaquerizas et al., 2009); reminiscent of TSS definition in invertebrates.

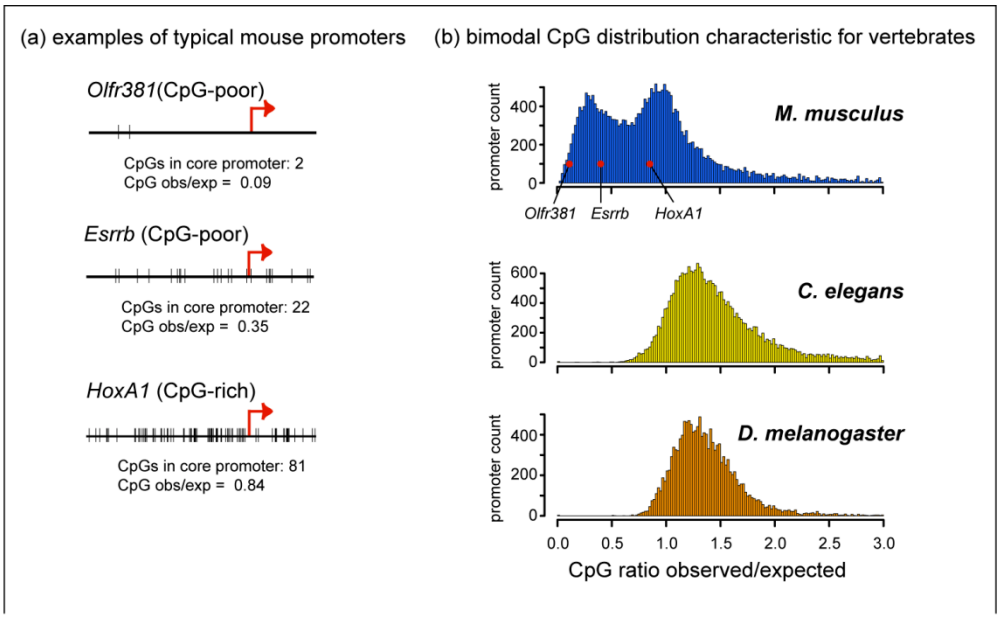


Figure 1. CpG distribution is markedly different in the genomes of vertebrates and invertebrates.
Invertebrate genomes show homogenous distribution and no depletion of CpG dinucleotides. Vertebrate genomes, by contrast, are globally depleted for CpG dinucleotides, except at CpG-islands, which often mark regulatory elements such as enhancers and promoters. (a) Representative examples of mouse promoters with up to 40-fold different abundance of CpG dinucleotides. CpG-poor: olfactory receptor 381 (Olfr381) and estrogen-related receptor b (Esrrb); CpG-rich: HoxA1. In each case, a 1.3-kb window around the transcription start site (red arrow) is shown. Each vertical line represents an individual CpG dinucleotide. (b) CpG dinucleotide distribution at promoter sequences in a vertebrate (Mus musculus) and two invertebrate (C. elegans and D. melanogaster) genomes. The ratio of observed versus expected CpGs is plotted as a histogram for all promoters. Vertebrates (here, mouse) show a bimodal distribution of promoter CpGs with two major peaks corresponding to CpG-poor (left peak) and CpG-rich (right peak) promoters. Invertebrates, illustrated by D. melanogaster and C. elegans, do not show such a distribution; instead, the major peak of CpG content is close to 1, indicating that invertebrates contain the expected number of CpGs in their promoters.

Similar to gene proximal sequences, CpG islands outside of promoters are highly conserved. This positive selection for maintaining CpGs suggests that CpG islands in general have regulatory and/or functional relevance in the vertebrate genomes (Bernstein et al., 2006; Tanay et al., 2007). Indeed, many non-promoter CpG islands might function as enhancers (Heintzman et al., 2007).

2.2. Chromatin

Prokaryotic genomes are much smaller than those of eukaryotes and usually organized in one circular DNA molecule which is compacted by supercoilling with the help of specific scaffold proteins (Dame, 2005). Because prokaryotes do not possess membrane-bound cellular organelles their DNA localizes to a distinct territory in the cell termed the nucleoid. The evolution of eukaryotes led to a dramatic increase in DNA content per cell and to the occurrence of two features which are tightly linked to the increase in genome size: the nucleus (and other membrane-bound organelles), and a distinct form of organizing genomic DNA termed chromatin, which is a complex structure of the DNA, proteins and RNA [Fig. 2; (Felsenfeld and Groudine, 2003)].

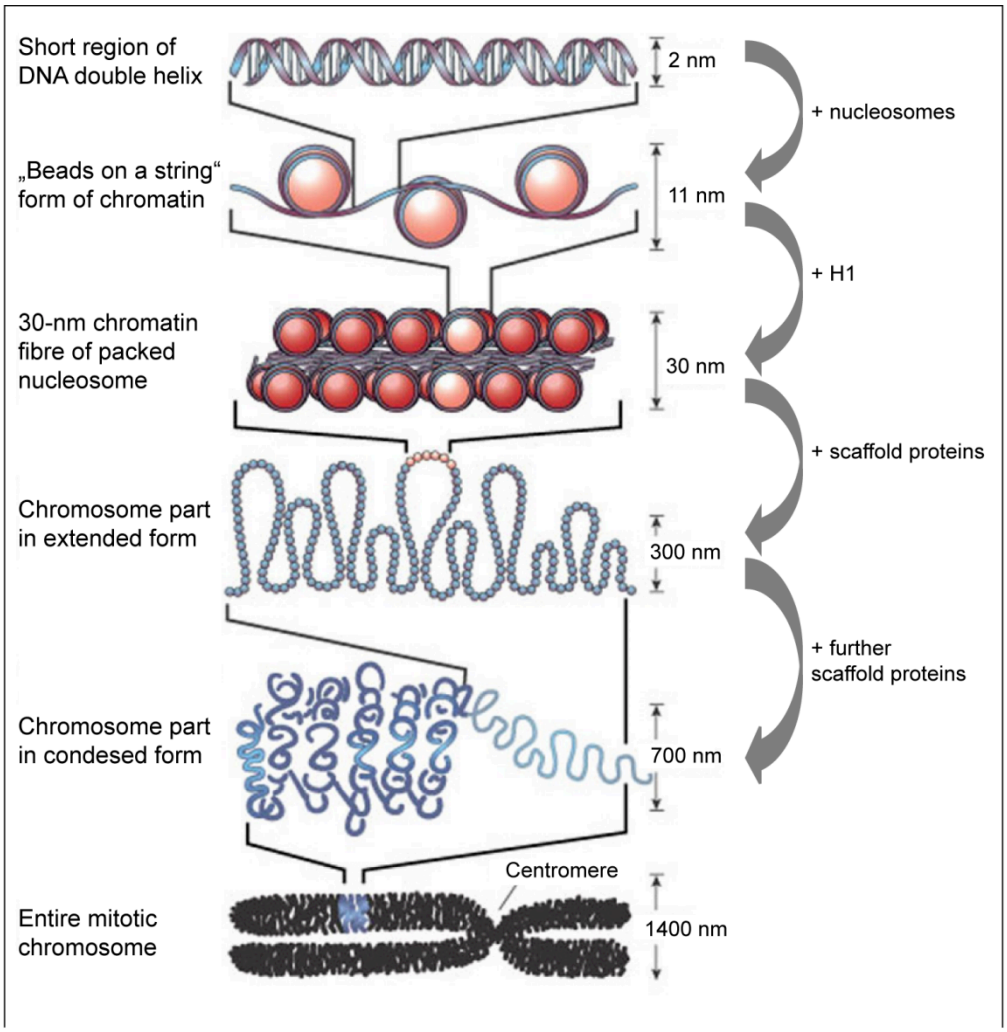


Figure 2. Model of chromatin compaction (adapted from Felsenfeld and Groudine, 2003)

2.2.1. The nucleosome – basic scaffold of chromatin structure

The fundamental building blocks of chromatin are the nucleosomes, which consist of globular histone protein cores around which the DNA is wrapped 1.65 times, corresponding to 147bp of DNA. These histone cores are composed of an octamer with two copies of each of the four highly conserved canonical histones H2A, H2B, H3 and H4 (Luger et al., 1997).

The individual histones have a positively charged globular histone-fold domain which binds to the negatively charged DNA via electrostatic interactions. The unstructured C- and N-terminal histone tails, some of which protrude outside of the nucleosomes, are subject to various post-translational modifications. These modifications are thought to directly impact the affinity to DNA, the interaction with neighboring nucleosomes and/or the recruitment of chromatin binding proteins (see below). If not further compacted, the nucleosomes are arranged in a linear fashion along the DNA molecule, leading to a so-called “beads-on-a-string” structure with 10-60 bp of “free” linker DNA between the individual nucleosomes (Fig. 2). This form of chromatin is generally found at active genes. Upon incorporation of histone H1 the chromatin structure can be further compacted. H1 binds to the linker DNA between the nucleosomes and leads to a transcriptionally inert higher order structure termed 30nm fiber [Fig. 2; (Robinson and Rhodes, 2006)]. Of note, such 30 nm fiber could not yet be detected *in vivo* and its existence is controversial (Tremethick, 2007). Microscopycal studies instead identified larger fiber-like structures in mammalian nuclei ranging form 60-80 nm in interphase up to 750 nm in metaphase when chromosomes are highly condensed (Kireeva et al., 2004). This could reflect further compaction of a 30 nm fiber with the help of yet to be discovered scaffold proteins (Fig. 2). Interestingly, in yeast the weakly conserved H1 homolog HHO1 is not essential (Ushinsky et al., 1997) and deletion does not affect global chromatin organization (Downs et al., 2003), suggesting that it could serve other functions than chromatin organization. Indeed, it was shown that HHO1 can block DNA repair via homologous recombination, which implicates HHO1 rather in genome surveillance than in chromatin organization (Downs et al., 2003). In contrast, in Drosophila H1 is essential for development and correct heterochromatin formation (Lu et al., 2009). Mammals contain at least six somatic H1 isoforms and mice do not display a phenotype upon deletion of one of these, which is probably due to partial redundancy among the H1 isoforms. Only mice lacking three isoforms have a significantly reduced total H1 level. These mice are not viable anymore and display a broad range of defects (Fan et al., 2005). Together, the data from mouse and fly indicate that H1 and accordingly proper higher order chromatin organization are essential for metazoan development.

2.2.2. Histone variants

In addition to the canonical core histones a variety of histone variants exist for H2A, H2B and H3. Only for few of these function and assembly factors are known. For example histone H3.3 is incorporated in actively transcribed regions to replace canonical H3 (Mito et al., 2005; Wirbelauer et al., 2005). Further, the H3 variant CenH3 (in mouse/human termed CENPA, in Drosophila Cid) exclusively localizes to centromeres and is essential for proper centromer functioning and chromosome segregation (Black and Bassett, 2008). Other variants, such as H2A.Z are less well understood and have been implicated in both gene activation and transcription as well as repression of gene activity (Draker and Cheung, 2009).

Interestingly, most protein sequence alterations between histone variants and their canonical counterparts occurred at the exposed surface. Moreover, many of these histone variants are constitutively expressed throughout the cell cycle as opposed to the canonical histones, which are only expressed during S phase when DNA replication takes place. Together, this led to the hypothesis that histone variants provide a means to modulate nucleosomal stability and local chromatin structure independent of DNA replication. They might thus contribute to cell type specific gene regulation and chromatin organization [for a more detailed overview of histone variants, their assembly and putative functions see (Kamakaka and Biggins, 2005; Malik and Henikoff, 2003; Sarma and Reinberg, 2005)].

2.2.3. Chromatin dynamics and remodeling

Chromatin decreases DNA accessibility for transcription factors and the transcriptional machinery when compared to naked DNA (Struhl, 1999; Workman and Kingston, 1998). As a consequence, chromatin imposes basic repression to genes by preventing spurious binding of transcription factors and transcriptional initiation without appropriate stimuli. In contrast, in prokaryotes, which do not package their DNA into chromatin, most genes are in a transcriptionally active ground state (Struhl, 1999). The evolution of chromatin led to a shift in gene-regulatory paradigms from a basal active state to a basal repressed state. Hence, transcription in eukaryotes generally requires activators (Struhl, 1999) and probably chromatin remodeling to increase accessibility of transcription factors to the DNA template (Clapier and Cairns, 2009; Workman and Kingston, 1998). In an elegant study, Lam and colleagues demonstrated that the amount of physiological stimulation required to activate a promoter indeed depends on the position of transcription factor binding motifs within the nucleosomes (Lam et al., 2008). Moreover, these results provide evidence that chromatin remodeling and nucleosome sliding can significantly impact the level of gene activation by occluding or exposing regulatory sequences. In the past years, two mechanisms have been proposed to potentially mediate chromatin remodeling and modify accessibility of chromatinized DNA: Widom and collegues proposed that non-

catalyzed thermodynamic fluctuations of nucleosomes could lead to transient exposure of regulatory sites and allow transcription factor binding. Upon binding of one factor the dislocation of the nucleosomes would be stabilized and further activators including the RNA polymerase II holoenzyme could be recruited to initiate transcription (Polach and Widom, 1996).

The second and probably major pathway is mediated by ATP-dependent chromatin remodeling complexes. These have been shown to slide, replace and/or evict nucleosomes in order to modulate DNA sequence accessibility for transcription factors and transcriptional regulation [reviewed in (Clapier and Cairns, 2009)]. Eukaryotes contain at least four families of chromatin remodeling complexes with partially distinct functions: SWI/SNF, ISWI, CHD, and INO80 family remodeling complexes (Clapier and Cairns, 2009; Jiang and Pugh, 2009). All these remodeler families are conserved from yeast to humans and all share an ATPase domain. Nonetheless the flanking domains involved in substrate recognition and interaction with other proteins are highly divergent.

SWI/SNF family remodelers have a broad range of activities and slide and/or eject nucleosomes at many loci in the genome. This class of remodelers is further characterized by containing a Bromo domain that specifically recognizes acetylated histone tails (a hallmark of active genes; see below) and other acetylated proteins. Therefore it is not surprising that SWI/SNF complexes were implicated in transcription but not in chromatin assembly.

ISWI type remodelers have been implicated in activating as well as in repressive functions, which entirely depends on the other complex members. For example the CHRAC and ACF complexes assist regular nucleosomal spacing and thereby facilitate chromatin assembly and transcriptional repression (Fyodorov et al., 2004). In contrast, the NURF complex randomizes spacing and facilitates RNA polymerase II transcription initiation (Mizuguchi et al., 1997). ISWI family remodelers have a characteristic SANT and SLIDE domains which form a nucleosomes recognition module. Interestingly, they preferably bind unacetylated histone tails suggesting an involvement in specific targeting (Boyer et al., 2004).

CHD family remodelers have characteristic chromodomains which are crucial for recruitment and remodeling activity (Brehm et al., 2004). Many chromodomains have specific affinity for methylated histones (see below). Similar to ISWI remodelers, CHD family remodelers have also been shown to assist both, gene activation and repression, depending on the overall complex composition.

INO80 family remodelers function in transcriptional activation, DNA repair and in case of the SWR1/SRCAP containing complexes they also mediate incorporation of the histone variant H2A.Z (Kobor et al., 2004).

Despite the numerous purified remodeling complexes the many shared subunits make it inherently difficult to assign unique functions to each of these complexes. It is emerging that not the remodeling activities but rather the subunits of the remodeling complexes confer functional and target specificity, which adds another level of complexity and regulatory potential. Supporting this, genetic studies revealed that certain remodelers and/or subunits are essential to animal development, while others caused only mild phenotypes or tissue-specific defects (Clapier and Cairns, 2009). Such tissue-specificity is further supported by evidence that remodeling complexes have tissue-specific subunit composition which might customizes the function and targeting towards the respective cell type. For example, the neuronal BAF chromatin remodeling complex (SWI/SNF-like) displays developmental stage specific subunit composition which is essential for normal neuronal development. In dividing neuronal stem/progenitor cells, it contains BAF45a and BAF53a subunits which are essential for progenitor self-renewal. Upon differentiation and mitotic exit, these two progenitor-specific subunits are replaced by the post-mitotic neuron-specific BAF45b, BAF45c and BAF53b subunits (Lessard et al., 2007).

In contrast to such cell type specific activities, chromatin remodelers are generally implicated in transcriptional initiation and elongation due to their ability to move, reconstruct, or eject nucleosomes and thereby exposing the underlying DNA for regulatory factors. Genetic evidence form Drosophila suggests that the dBrahma complex, a SWI/SNF family remodeler, generally facilitates transcription via RNA polymerase II (Pol II). Upon reduction of dBrahma, Pol II was greatly reduced at polytene chromosomes (Armstrong et al., 2002), implying a direct connection between remodeling activity and Pol II recruitment. However, currently it is unclear if chromatin remodeling is a prerequisite for, coinciding with or a consequence of transcription initiation and further work addressing this is required.

2.3. Epigenetic modifications of chromatin

As eluted above chromatin is essential for DNA organization however it also presents a barrier to any DNA-templated event, which can be alleviated by chromatin remodeling processes. Superimposed upon the level of genomic DNA sequence and physical chromatin organization is a layer of "epigenetic" information, which we have only just begun to explore and understand. The term epigenetic was originally coined by Conrad Waddington (Waddington, 1942), which he described as "the causal interaction of genes and their products, which bring the phenotype into being". Over the years epigenetics became a collection of phenomena which could not be explained by pure genetics. As a consequence the definitions of epigenetics vary considerably. While there is a general agreement that epigenetics refers to phenomena that change the expression state of a genetic locus without changing the underlying DNA sequence, the discrepancy of definitions is largely about the heritability of these phenomena. Strictly, to be epigenetic, a phenomenon has to be inherited during mitosis without relying on the initial trigger (Ptashne, 2007). Due to the inherent difficulty to distinguish between sequence-independent self-propagation of epigenetic states and a re-establishment after cell division mediated by sequence-dependent recruitment of enzymatic activities, most chromatin modifications are termed epigenetic without knowing the molecular mechanisms that mediate propagation (Bird, 2007). Notably, DNA methylation presents the only epigenetic modification for which a sequence independent propagation mechanism has been identified.

2.3.1. Post-translational modifications of core histones

DNA-bound histone proteins contain over 60 sites which are subject to post-translational modifications (PTMs) such as acetylation, methylation, ubiquitination, phosphorylation, sumoylation and others (Kouzarides, 2007). These modifications mostly take place at the N-terminal tails of histone H3 and H4, which protrude out of the nucleosomes cores (Fig. 3).

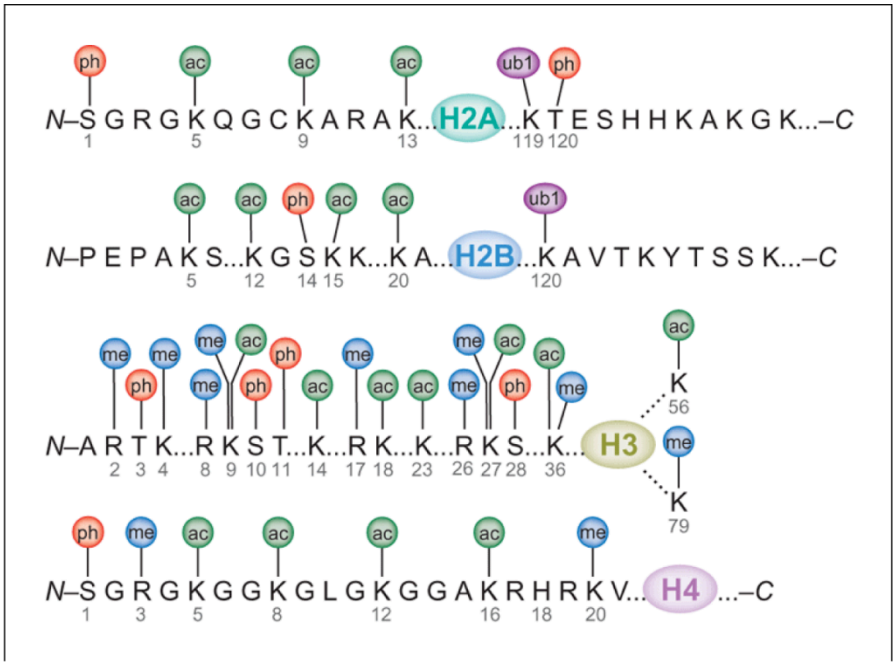


Figure 3. Overview of known sites for post-translational modifications on the N- and C-terminal tails of canonical histones. ph: phosphorylation, ac: acetylation, me: methylation, ub1: monoubiquitination (adapted from Bhaumik et al, 2007).

As a result the PTMs are accessible to non-histone proteins and present a way to specifically recruit non-histone protein complexes to chromatin. A second possible regulatory role of PTMs is direct interference with binding of DNA around nucleosomes via altering the electric charge of histones. At first, these chromatin modifications and the overall chromatin structure were thought to be stable and simply providing a structural scaffold. However, the discoveries that the yeast transcriptional coactivator Gcn5 bears histone acetylation activity (Brownell et al., 1996), while the corepressor Rpd3 mediates histone deacetlyation (Taunton et al., 1996) for the first time directly linked transcriptional regulation to PTMs. A few years later Suv39h, a previously identified transcriptional regulator and suppressor of variegation in Drosophila (Tschiersch et al., 1994), and its yeast homolog Clr4 were shown to specifically methylated lysine 9 at histone H3 [H3K9; (Rea et al., 2000)]. This subsequently led to a change of paradigms towards a regulatory role of chromatin and PTMs of histones in DNA-templated processes. Ultimately, the discovery of histone demethylases indicated that PTMs are much more dynamic than previously anticipated (Shi et al., 2004).

Although many of these PTMs are considered to be epigenetic and are thought to be inherited during mitotic cell divisions, so far the propagation mechanisms are unknown for all of these modifications. Below I will restrict the discussion of PTMs towards methylation and acetylation of lysine residues on histone H3 and H4 which are directly related to my PhD project.

2.3.1.1. Histone acetylation

Lysines are the major source of the net positive charge of histone octamers and therefore crucial for binding the negatively charged phosphate backbone of DNA. For transcription, replication or DNA repair it is important that histone-DNA interactions can be modified in order to facilitate passage of polymerases. All four core histones bear lysine residues which are subject to acetylation (Fig. 3) and deacetylation by specific histone acetyltransferases (KAT) and deacetylases (HDAC) (Kouzarides, 2007). Generally, acetylated lysines on histones H2B, H3 and H4 are highly correlated with active transcribed regions (Pokholok et al., 2005; Schubeler et al., 2004; Wang et al., 2008). This is thought to be due to neutralization of the positive charge of lysines upon acetylation, which lowers the electrostatic interactions with the negatively charged phosphate backbone of DNA and consequently weakens the DNA-histone interaction. Indeed, acetylation increases DNA accessibility, destabilizes nucleosomes and leads to an increase of non-histone protein binding to DNA *in vitro* (Lee et al., 1993; Vettese-Dadey et al., 1996; Wolffe and Hayes, 1999). Thus, it is conceivable that acetylation of individual lysines conveys little specificity, but rather the cumulative effect of acetyl groups at multiple lysines would be important for regulating DNA accessibility.

One marked exception is acetylation of lysine 16 of histone H4 (H4K16), which was shown to directly interfere with higher order chromatin structure formation via preventing interactions between neighboring nucleosomes (Robinson et al., 2008; Shogren-Knaak et al., 2006). Moreover, H4K16 acetylation plays a specific role in Drosophila dosage compensation (Bell et al., 2008; Kind et al., 2008).

Many proteins in chromatin-associated complexes contain highly conserved Bromo-domains which specifically bind to acetylated lysines. This indicates that there might be more specific regulatory potential to acetylation than previously anticipated (Taverna et al., 2007). However, an “acetylation-code” (Kurdistani et al., 2004) appears unlikely. Global distributions of histone H3 and H4 lysine acetylation marks are largely overlapping and highly correlated with transcription (Wang et al., 2008). Further, genetic evidence from yeast revealed that substitution of individual lysines with the exception of H4K16 did not produce specific phenotypes (Dion et al., 2005). These data favor a transcription-coupled “cumulative” acetylation model with additive functions for acetylation marks. Importantly, this does not exclude additional specific functions for a few individual acetylated lysines such as H4K16ac in higher order chromatin compaction (Shogren-Knaak et al., 2006) and H3K56ac in nucleosome assembly during DNA repair (Das et al., 2009).

2.3.1.2. Histone methylation

Methylation of histones can either occur at lysine or arginine residues (Fig. 3). In contrast to acetylation, mono- (me1), di- (me2) and tri-methylation (me3) states of the same residue are observed. These differential methylation states present another level of regulatory potential which indeed appears to be exploited. Several lysines display diverging functions and localization in the genome depending on their methylation state (Barski et al., 2007; Peters and Schubeler, 2005). Arginine methylation is performed by protein arginine methyltransferases (PRMTs) and is antagonized by PADI4 (Klose et al., 2006; Zhang and Reinberg, 2001). Lysine methylation is carried out by specific lysine methyltransferases (KTMs), which all contain a conserved SET-domain with the exception of Dot1/KTM4 (Zhang and Reinberg, 2001). Lysine methylation can be removed by two distinct classes of histone demethylases (KDMs): the LSD1 enzyme and the JmjC protein family (Klose et al., 2006). In contrast to acetylation, methylation does not neutralize the charge of nucleosomes and is therefore thought to serve as a signal or recruitment platform for chromatin modifying and regulatory proteins. So far three protein domains have been found to specifically recognize methylated lysines: Tudor domains, chromodomains and PHD-finger domains. Each domain has characteristic affinities for different lysines and methylation states which further depend on other domains of the respective protein and its interaction partners [review in (Martin and Zhang, 2005; Taverna et al., 2007)].

Chromatin immunoprecipitation experiments have revealed that active genes are methylated at lysine 4 of histone H3 (H3K4), H3K36 and H3K79 [Fig. 4; (Barski et al., 2007; Pokholok et al., 2005; Saunders et al., 2006; Schubeler et al., 2004)]. These modifications are thus thought to have a role in transcription. This is supported by data from yeast indicating that Set1 and Set2, which methylate H3K4 and H3K36, directly interact with factors bound to the Pol II complex (Krogan et al., 2003a; Krogan et al., 2003b). Genetic evidence also predicts a recruitment of Dot1 (a H3K79 KMT) to chromatin via elongating Pol II (van Leeuwen et al., 2002). H3K4me peaks around the transcription start site and is gradually diminished further 3'. H3K36me and H3K79me display a broader distribution within the gene body, starting just downstream of the H3K4me peak [Fig. 4; (Bell et al., 2007; Wirbelauer et al., 2005)]. Consistent with a role for H3K36me in transcription, data from yeast denote that H3K36me prevents cryptic initiation via recruiting a histone deacetlyase to the body of genes, which presumably leads to a less accessible chromatin structure (Carrozza et al., 2005).

H3K4 methylation has been implicated in transcriptional activation pathways since many chromatin remodeling and co-activator complexes bear a module which specifically recognizes H3K4me2/3. For example a PHD-domain in the NURF chromatin remodeling complex specifically recognizes H3K4me3 and might facilitate transcriptional activation via

opening the chromatin structure around H3K4me2/3 modified promoters (Wysocka et al., 2006).

More recent data from mammalian systems indicates that in contrast to invertebrates H3K4me2/3 are not exclusively marking actively transcribed regions [Fig. 4; (Bernstein et al., 2006; Guenther et al., 2007; Mikkelsen et al., 2007; Roh et al., 2006; Weber et al., 2007)]. Interestingly, these loci are CpG-rich sequences and many bear low but detectable levels of Pol II and acetylated histone H3 (Guenther et al., 2007; Roh et al., 2006). As a consequence virtually all CpG-rich promoters reside in chromatin carrying H3K4 di-/tri-methylation independent of transcriptional activity. This is in sharp contrast to CpG-poor promoters. These are only H3K4 methylated when transcribed, which is reminiscent of the situation in invertebrates such as Drosophila and yeast (Pokholok et al., 2005; Schubeler et al., 2004).

Inactive loci display a different set of methylation marks mainly consisting of methylation of H3K9, H4K20, and H3K27. H3K9 and H4K20 di- and tri-methylation play essential roles in heterochromatin maintenance at pericentromeric repeat regions and are further present at repetitive, transposable and retroviral elements in mammalian genomes (Lehnertz et al., 2003; Mikkelsen et al., 2007; Peters et al., 2003). Only few regulatory regions have so far been identified to be methylated at H3K9 and/or H4K20. These are mostly CpG-poor promoters of large gene families such as the olfactory receptor clusters or zinc finger proteins (Mikkelsen et al., 2007; Vogel et al., 2006), suggesting that recognition is based on the repetitive nature of their genomic organization. H3K9 methylation is carried out by 5 known KMTs with distinct specificities: Suv39h1 and Suv39h2 are mostly responsible for methylation in constitutive heterochromatin in pericentric and telomeric regions. G9a, GLP and Setdb1 rather localize to euchromatin and have roles in silencing of repetitive and retroviral elements (Kouzarides, 2007).

H3K9me2/3 was further shown to recruit HP1 to chromatin, which is an integral component of heterochromatin and essential for repression (Lachner et al., 2001). Further evidence indicates that H3K9 methylation crosstalks to DNA methylation. For example direct interactions between DNMT1 and the H3K9 KMT G9a, and between DNMT3b and HP1 have been reported (Esteve et al., 2006; Lehnertz et al., 2003). The latter is important for correct establishment of pericentric heterochromatin (Lehnertz et al., 2003). Feldman and coworkers suggested a stepwise model for gene silencing initiated by a G9a complex which methylates H3K9. This methylation mark subsequently recruits HP1, which further recruits DNMTs to mediate stable repression by DNA methylation (Cedar and Bergman, 2009; Feldman et al., 2006). While this appears as an attractive model for a step-wise establishment of facultative heterochromatin, it awaits further experimental prove.

H3K27 di- and tri-methylation in turn is excluded from regions carrying H3K9 methylation and predominantly localizes to CpG-rich regions, which strongly implies different functions of these two repressive histone methylation marks (see below).

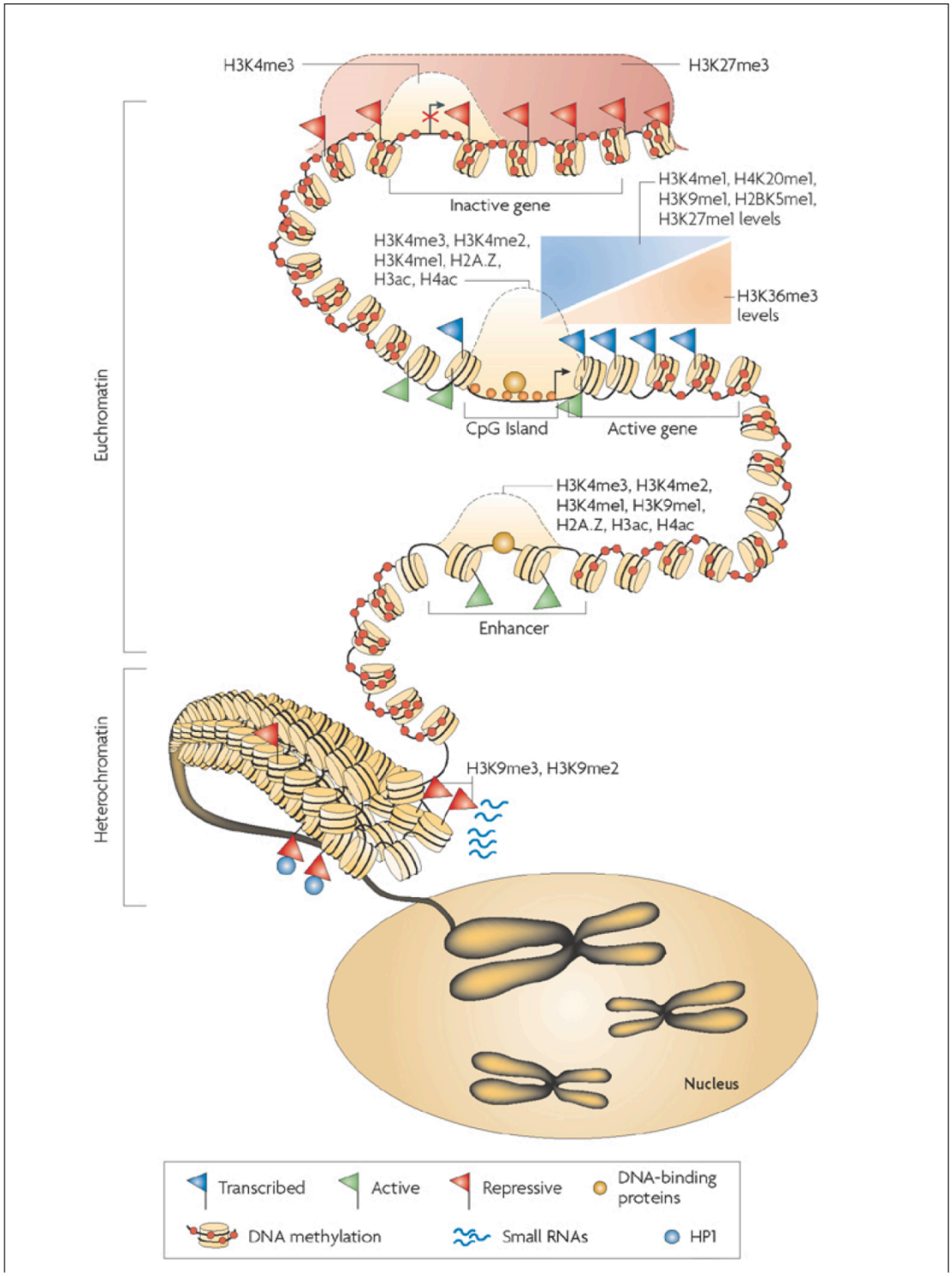


Figure 4. Chromatin structure and modifications.
Chromosomes are divided into accessible regions of euchromatin and poorly accessible regions of heterochromatin. Heterochromatic regions are marked with H3K9me2 and H3K9me3, which serve as a platform for HP1 binding. Small RNAs have been implicated in the maintenance of heterochromatin. DNA methylation is persistent throughout genomes, and is missing only in regions such as CpG islands, promoters and possibly enhancers. The H3K27me3 modification is present in broad domains that encompass inactive genes and usually coincides with H3K4me2/3 to from so-called bivalent domains. H3K4me3, H3K4me2, H3K4me1 as well as histone acetylation and histone variant H2A.Z mark the transcription start site regions of active genes. The monomethylations of H3K4, H3K9, H3K27, H4K20 and H2BK5 mark actively transcribed regions, peaking near the 5' end of genes. The trimethylation of H3K36 also marks actively transcribed regions, but peaks near the 3' end of genes. High levels of H3 and H4 acetylation are typically detected at the promoter and 5' regions (adapted from Schones and Zhao, 2008).

2.3.2. Polycomb-mediated repression

The antagonistic trithorax-group (TrxG) and Polycomb-group (PcG) proteins are well characterized transcriptional gene regulators which have been first identified as Hox gene regulators guiding embryonic patterning in Drosophila (Bantignies and Cavalli, 2006; Ringrose and Paro, 2007; Schuettengruber et al., 2007; Schwartz and Pirrotta, 2007; Sparmann and van Lohuizen, 2006). Remarkably, PcG proteins underwent an expansion during evolution of vertebrates. Many paralogs arose, which has been suggested to contribute to cell type specific gene regulation; an essential requirement for the observed increase in organismal complexity (Whitcomb et al., 2007).

Beyond their role in embryonic development, PcG proteins have been implicated in maintaining pluripotency and cell identity via repression of key developmental regulators in embryos and ES cells (Bernstein et al., 2006; Boyer et al., 2006; Lee et al., 2006). PcG proteins have further been shown to play a role in cell differentiation, cell fate plasticity (Caretti et al., 2004; Ezhkova et al., 2009; Klebes et al., 2005; Lee et al., 2005) and proliferation (Blais et al., 2007; Martinez et al., 2006). In addition, their mis-regulation is associated with neoplastic development (Sparmann and van Lohuizen, 2006). ES cell lacking Ezh2, Eed or Suz12 can be maintained in culture, suggesting that Polycomb is not essential for propagation of pluripotent stem cells. However, all mutant cells show a modest upregulation of a number of differentiation genes which are normally under the control of Polycomb. Upon induction of differentiation the mutant cells either die or do not differentiate to term, which is in line with the early embryonic lethal phenotypes *in vivo* (Chamberlain et al., 2008; Erhardt et al., 2003; O'Carroll et al., 2001; Pasini et al., 2007).

Biochemical purification experiments revealed that at least two separate multimeric PcG complexes or Polycomb repressive complexes (PRC) with distinct enzymatic activities exist [Fig. 5; reviewed in (Bantignies and Cavalli, 2006; Schwartz and Pirrotta, 2007; Sparmann and van Lohuizen, 2006)]. PRC2 contains Ezh2 which mediates di/tri-methylation of lysine 27 of Histone H3 [H3K27me2/3; (Czermin et al., 2002; Muller et al., 2002)], the hallmark of Polycomb repressed genes. PRC1 on the other hand contains four conserved core components and mediates H2A lysine 119 mono-ubiquitination [H2AK119u1; (Wang et al., 2004)]. PRC1 is thought to cooperate with PRC2 for binding at target genes enriched in H3K27me3, and to mediate repression by inhibiting chromatin remodeling, blocking transcription and/or by mediating chromatin compaction [Fig. 5; (Levine et al., 2004; Margueron et al., 2008)]. The precise mechanisms by which PRC complexes regulate gene expression are unclear but their overlapping genomic distributions suggest concerted actions (Boyer et al., 2006; Ku et al., 2008). At the level of higher-order chromatin organization, a clustering of silenced genes into sub-nuclear compartments termed PcG-bodies seems to be important and could provide a means to coordinate repression of multiple genes (Bantignies

et al., 2003; Grimaud et al., 2006; Lanzuolo et al., 2007; Terranova et al., 2008; Tiwari et al., 2008).

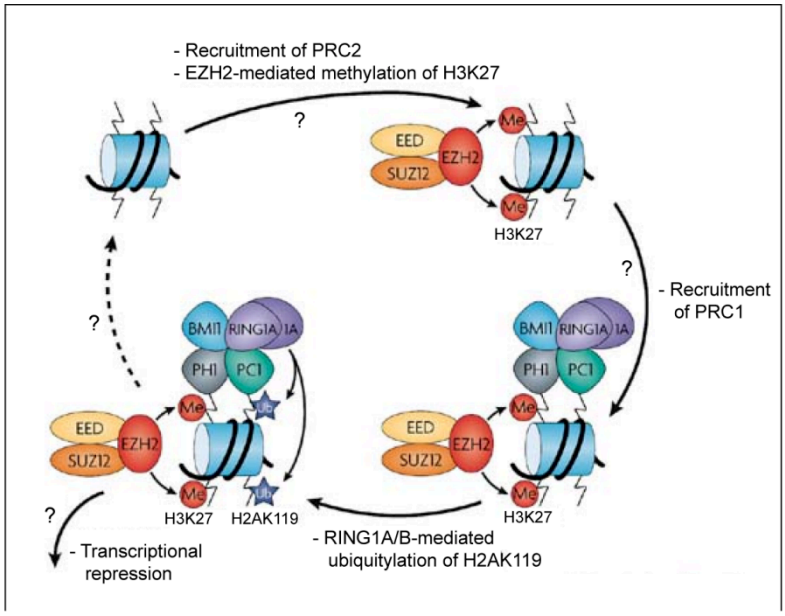


Figure 5. Classical model of Polycomb-mediated repression.
Polycomb repressive complex 2 (PRC2), which consits of enhancer of zeste homologue 2 (EZH2), embryonic ectoderm development (EED) and suppressor of zeste homologue 12 (SUZ12), is recruited via unknown sequence specific factors and tri-methylates H3K27. H3K27me3 subsequently attracts PRC1, which consists of BMI1, RING1A or RING1B, which work as ubiquitin (Ub) ligases for H2AK119, Polycomb homologue 1 (PC1) and Polyhomeotic homologue 1 (PH1). Prc1 is thought to further stabilize repression after the initial set up and depostition of H3K27me3 by PRC2. Questions marks indicate elusive molecular mechanisms (adapted form Spivakov and Fischer, 2007).

Interestingly, in mammals Polycomb mediated H3K27me3 very often coincides with H3K4me2/3 in close proximity, a configuration termed bivalent chromatin due to the opposing effects of the two modifications (Azuara et al., 2006; Bernstein et al., 2006; Bracken et al., 2006; Mikkelsen et al., 2007). While it has been put forward that such a bivalent configuration could “poise” genes for activation at a later time point (Bernstein et al., 2006; Spivakov and Fisher, 2007), no experimental evidence for such a scenario exists.

At some Drosophila genes, PcG proteins bind to regulatory elements called PREs (Polycomb Response Elements) to silence nearby genes, suggesting that Polycomb targets are encoded in the DNA sequence (Ringrose and Paro, 2007). In mammals, despite large-scale mapping efforts, no such DNA sequence elements for PcG recruitment could be identified so far. This, together with the low sequence conservation of Drosophila PREs, might reflect that many different sequence-specific transcription factors can recruit Polycomb and therefore identification of a consensus PRE is not possible. For example in cancer cells

the transcription factor SNAIL was shown to recruit Polycomb proteins to the E-cadherin promoter coinciding with its transcriptional silencing (Herranz et al., 2008).

In addition to the elusive targeting, propagation of H3K27me3 and Polycomb complexes during replication is poorly understood. A recent report suggested that PRC2 directly binds to H3K27me3 and thereby ensures propagation of the mark during mitosis (Hansen et al., 2008). Another report suggested that PRC1 remains bound to the DNA during replication in an *in vitro* system (Francis et al., 2009). However *in vivo* propagation remains to be addressed in more detail.

Several recent reports suggest interplay between the Polycomb pathway and DNA methylation. Direct protein-protein interaction was reported for Ezh2 and DNMT1 in cancer cell lines (Vire et al., 2006), which however is controversial and has not been reproduced in non-transformed cells. In addition, several recent studies suggested preferential aberrant DNA methylation in human cancer cell lines and primary cancers at promoters that are Polycomb targets in unrelated human ES cells in culture (Ohm et al., 2007; Schlesinger et al., 2007; Widschwendter et al., 2007). These studies hint towards a potential targeting pathway of DNMTs via Polycomb, however no direct prove exists so far.

2.3.3. DNA methylation

In prokaryotes, cytosine and adenine methylation are involved in multiple processes such as DNA repair and pathogen defense [reviewed in (Reisenauer et al., 1999)]. Eukaryotes contain exclusively cytosine methylation which in mammals is limited to cytosines in CpG dinucleotides. In plants and fungi cytosines in CpNpG or even CpNpN context can be methylated, too. Interestingly, not all eukaryotes methylate their genomes and moreover levels of DNA methylation can vary greatly between members of the same phylogenic class. For example in the class of fungi DNA methylation is absent in yeast (Antequera et al., 1984) but present at moderate levels in *Neurospora crassa* (Selker et al., 2003). Among plants, maize contains high levels of methylation (Rabinowicz et al., 2003) while *Arabidopsis thaliana* displays moderate amounts of methylated cytosines (Zhang et al., 2006). Further, *Drosophila melanogaster* has only low abundant DNA methylation in early larval stages (Phalke et al., 2009) whereas the honey bee *Apis melifera* has substantial levels also at adult stages (Wang et al., 2006). The roundworm *Caenorhabditis elegans* contains no methylated cytosines (Simpson et al., 1986), however it is not known if other worms eventually methylated their DNA. In contrast, all vertebrates analyzed so far display genome-wide DNA methylation with up to 80% of all CpG dinucleotides being methylated (Suzuki and Bird, 2008).

2.3.3.1. Eukaryotic DNA methyltransferases

Four distinct families of DNA methyltransferases (DNMT), grouped according to homologies in their catalytic domains, have been identified in eukaryotes. Below, I will mainly focus on mammalian DNMTs since some methylation pathways and enzymes in plants appear unique with no counterpart in animals. A comprehensive description of DNA methyltransferases in animals, fungi and plants is provided in a review by Goll and Bestor (Goll and Bestor, 2005).

DNA methyltransferase 1 family

DNMT1 was the first eukaryotic DNA methyltransferase to be identified (Bestor et al., 1988). It was subsequently found that it has a preference for hemi-methylated DNA, making it a prime candidate for propagation and stable maintenance of DNA methylation patterns through mitosis (Stein et al., 1982). This is further supported by direct interactions of DNMT1 with Pcna (Chuang et al., 1997) and Uhrf1 (also known as NP95 and ICBP90), which recruits DNMT1 to hemimethylated DNA at replication forks (Arita et al., 2008; Avvakumov et al., 2008; Bostick et al., 2007; Sharif et al., 2007). Upon loss of Uhrf1, global DNA methylation levels are severely reduced and intracisternal A particle (IAP) and long interspersed nuclear element 1 (LINE-1) retrotransposons are transcriptionally activated (Sharif et al., 2007), reminiscent of the phenotype of DNMT1$^{-/-}$ ES cells and embryos [see below; (Okano et al., 1999; Walsh and Bestor, 1999)].

Extensive genetic studies revealed that DNA methylation is essential for normal development. Mice carrying a homozygous deletion for DNMT1 die at 8.5 days post coitus [dpc; (Li et al., 1992)]. At the same time, this argues against a general role of DNA methylation in regulating tissue-specific and developmental genes due to the survival of the embryos up to 8.5 dpc. Furthermore, fibroblasts lacking DNMT1 showed only a handful of genes with strong upregulation upon loss of DNA methylation (Jackson-Grusby et al., 2001). Thus, the lethal phenotype was mostly attributed to loss of imprinting and reactivation of transposons and probably other repetitive elements (Walsh and Bestor, 1999).

In embryonic stem (ES) cells with a homozygous deletion for DNMT1, genomic DNA methylation levels are dramatically reduced to about 5% remaining methylation compared to wild type ES cells (Lei et al., 1996). Moreover, normally imprinted genes lost their monoallelic expression and were either silenced or expressed from both alleles (Li et al., 1993). In addition mutant embryos inactivate both X chromosomes due to reactivation of Xist on the active X upon loss of DNA methylation-mediated Xist repression (Panning and Jaenisch, 1996). While DNMT1$^{-/-}$ ES cells can be maintained over a long period of time, they were reported to undergo cell-autonomous apoptosis upon induction of differentiation, which is consistent with the *in vivo* phenotype (Jackson et al., 2004; Li et al., 1992).

Remarkably, *in vitro* DNMT1 shows even higher activity on unmethylated substrate than the two known *de novo* methyltransferases, implying that DNMT1 might also play a role for *de novo* methylation (Okano et al., 1998a).

DNA methyltransferase 3 family

The members of the DNMT3 family, i.e. DNMT3A and DNMT3B, were identified as *de novo* methyltransferases based on the inability of *DNMT3A*$^{-/-}$ / *DNMT3B*$^{-/-}$ ES cells to methylated newly integrated retroviral DNA (Okano et al., 1999). Remarkably, individual mutants do not show impaired *de novo* methylation of inserted DNA, implying that DNMT3A and DNMT3B might have in part overlapping functions in ES cells and early embryos (see below). *In vitro*, both enzymes catalyze methylation of unmethylated and hemi-methylated substrates without preference for one or the other (Okano et al., 1998a). Homozygous *DNMT3A*$^{-/-}$ mice survive to term but are runted and die within a few weeks after birth with multiple defects including loss of germ cells in males. A *DNMT3B* knock-out is more severe and embryos display growth and neuronal defects already at early stages and die around 9.5 dpc. Embryos lacking both *DNMT3A* and *DNMT3B* display a synergistic effect and die around 8.5 dpc. They are mostly devoid of DNA methylation, although to a lesser extent than *DNMT1*$^{-/-}$ embryos. Further these mutants lack somites, which indicates that growth and development are arrested shortly after gastrulation, reminiscent of *DNMT1* mutants (Okano et al., 1999). Further, endogenous C-type retroviruses and IAPs were hypomethylated, again comparable to *DNMT1*$^{-/-}$ embryos but nevertheless the cause of embryonic lethality is not evident from these genetic experiments.

DNMT3L, the third member of the mammalian DNMT3 family, has no catalytic activity due to non-conservative amino acid substitutions in the catalytic domain. It is exclusively expressed in the germline and early embryonic stages (including embryonic stem cells) where it acts as crucial cofactor of DNMT3A for establishing maternal imprints during oocyte development (Bourc'his et al., 2001). Further, DNMT3L is essential for silencing retrotransposons during early male germline development, successful meiosis and spermatogenesis (Bourc'his and Bestor, 2004). Homozygous *DNMT3L* mutant mice are viable but both male and female mice are infertile due to the afore mentioned germline defects.

Plants sport an additional RNA-directed DNA methylation (RdDM) pathway, which is mediated by the DNMT3 homologs Drm1 and Drm2. Both proteins have been shown to interact with components of the small interfering RNA pathways in Arabidopsis and are targeted to repeats and transposons via small RNAs which arose from transcription of these elements. At their targets, Drm1 and Drm2 confer CpN, CpNpG and CpG methylation. In vertebrates such an RdDM pathway has not been identified yet. Recent advances in

understanding repeat silencing in germ cells indicated that short non-coding RNA pathways are essential for silencing of transposable elements and potentially reside upstream of DNA methylation of these elements, however the mechanistic details have not been worked out yet (see also paragraph “Targeting DNA methylation”).

DNA methyltransferase 2 family

All organisms which contain a DNMT1- and DNMT3-family enzyme also contain a DNMT2 homolog. Strikingly, DNMT2 is the best conserved and most widely distributed eukaryotic methyltransferase and in some species it is the only DNA methyltransferase homolog. However, so far various labs detected either no (Dong et al., 2001; Okano et al., 1998b; Yoder and Bestor, 1998) or only minute (Kuhlmann et al., 2005; Kunert et al., 2003; Tang et al., 2003) DNA methyltransferase activity of DNMT2. Moreover, DNMT2$^{-/-}$ mice are perfectly viable and do not display any noticeable phenotype, suggesting either a different or more specialized function for DNMT2 compared to the other DNMTs. Two recent reports identified DNMT2 of mammals, dipterans and plants as RNA methyltransferase specific to aspartic acid tRNA (Goll et al., 2006; Jurkowski et al., 2008). However, the cellular function of tRNA methylation remains elusive and given that homozygous knock-outs do not have any detectable phenotype in mouse, Drosophila and Arabidopsis under laboratory conditions, it remains to be tested if environmental stress would elicit a detectable response in the mutants. Given that DNMT2 has virtually no DNA methyltransferase activity, this might also explain the diverging methylation patterns in closely related invertebrate species. As eluted above, *S. pombe* has a DNMT2 homolog but does not methylated it’s DNA while *N. Crassa* contains DNMT1 and DNMT3 homologs and methylates DNA. Similarly among insects, *D. melanogaster* only carries a DNMT2 homolog and has only very little genomic methylation while *A. melifera* contains DNMT1 and 3 homologs and substantial amounts of methylated cytosines (Wang et al., 2006).

Interestingly, DNMT2 was suggested to methylate DNA during the first hours of Drosophila embryogenesis. This methylation serves as a trigger to silence retrotransposons and stably maintain certain telomeric repeat sequences via a histone methylation dependent pathway (Phalke et al., 2009). In absence of DNMT2, somatic cells express considerable levels of retrotransposons; however flies lacking DNMT2 are normal and fertile. An explanation for this might be that in germ cells retrotransposons are silenced via RNAi-based mechanisms that is independent of DNMT2 (Klenov et al., 2007; Vagin et al., 2006).

Chromomethylases

Chromomethylases have a characteristic chromodomain and are only found in plants (Henikoff and Comai, 1998). The Arabidopsis genome includes three chromomethylase

homologs: Cmt1, Cmt2 and Cmt3. They have been shown to confer CpNpG methylation and upon loss of chromomethylases retrotransposons become reactivated. However even after several generations of self-fertilization no obvious phenotype was found in plants lacking Cmts. It has been suggested that CpNpG methylation reinforces and replaces CpG methylation in regions with low levels of CpGs. Indeed, reduction of both, CpG and CpNpG methylation leads to a synergistic effect with double mutants displaying a higher increase in transposons reactivation than either single mutant (Kato et al., 2003).

2.3.3.2. DNA methylation patterns and functions in animals

It is emerging that invertebrates mostly contain “mosaic” methylation patterns with domains of dense methylation interspersed by methylation-free regions. Methylation often localizes to gene bodies [e.g. *Ciona intestinalis*; (Suzuki and Bird, 2008; Suzuki et al., 2007)]. Interestingly, genes which display gene body methylation exhibit higher evolutionary conservation. Moreover, these genes are expressed at moderate levels, while genes expressed at very high levels rarely show gene body methylation. Based on this, it was suggested that gene body methylation could be utilized to prevent spurious transcriptional initiation within genes transcribed at low levels (Suzuki et al., 2007). Nevertheless, this model lacks validation and thus the function of mosaic DNA methylation in invertebrates remains elusive.

In contrast to mosaically methylated invertebrate genomes where methylated and unmethylated regions coexist at more or less equal rations, vertebrates methylate up to 98 % of their genome including genes, repeats and transposons (Suzuki and Bird, 2008). The 1-2 % of the genome that is devoid of methylation corresponds to CpG islands, which appear to be protected from DNA methylation (see below). Hence it is not surprising that gene body methylation is also present in vertebrates (Eckhardt et al., 2006; Weber et al., 2005) and it is conceivable that it could prevent spurious transcription initiation, which however needs to be tested. Besides gene body methylation, vertebrates generally methylate repetitive and transposable elements which represent over 60 % of the total genomic DNA in mammals (Lander et al., 2001). DNA methylation is essential for stably silencing certain types of retrotransposons, while others are not reactivated in DNA methylation mutants (Bourc'his and Bestor, 2004; Rollins et al., 2006; Walsh and Bestor, 1999; Walsh et al., 1998). In addition to repeat silencing, DNA methylation is important for X-inactivation (Heard et al., 1997), genomic imprinting (Li et al., 1993) and maintaining chromosome stability and genome integrity. For example mutations in DNMT3B lead to ICF syndrome, which is molecularly characterized by chromosomal rearrangements and hypomethylation at centromeres (Xu et al., 1999). ES cells lacking either DNMT1 or DNMT3A and 3B have an increased frequency of telomere recombination (Gonzalo et al., 2006) and mice with reduced

DNMT1 activity display higher mutation rate and increased loss of heterozygosity due to increased meiotic recombination (Chen et al., 1998).

The role of DNA methylation in tissue specific gene expression is discussed controversially. Early studies involving single genes showed that DNA methylation changes of promoters can indeed follow their transcriptional activity but not in an obligatory fashion (Futscher et al., 2002; Walsh and Bestor, 1999). More recent genome-wide studies revealed that most CpG-rich promoters are unmethylated in ES cells as well as in differentiated cell types and somatic tissues (Farthing et al., 2008; Fouse et al., 2008; Illingworth et al., 2008; Meissner et al., 2008; Mohn et al., 2008; Rollins et al., 2006; Weber et al., 2007). These reports also revealed that 2-5% of CpG island promoters acquire DNA methylation during cellular differentiation and development. Interestingly, the majority of differential DNA methylation between somatic tissues seems to locate outside of promoters at CpG-rich sequences, indicating that DNA methylation might have a role in regulating activity of distal regulatory elements such as enhancers (Eckhardt et al., 2006; Illingworth et al., 2008; Irizarry et al., 2008; Song et al., 2005). In summary, these reports do not support a general role for DNA methylation in regulating tissue-specific gene expression. Nevertheless they indicate that a subset of genes is stably repressed by DNA methylation; maybe to confer robustness to the differentiated state of the cells.

DNA hyper- and hypo-methylation was further proposed to contribute to cellular transformation and progression of several types of cancers. Frequently aberrant methylation of tumor suppressor genes and cell cycle regulators are found in cancer, suggesting that DNA methylation contributes to malignancy and stabilizes the erroneous gene expression program in these cells (Feinberg et al., 2006; Jones and Baylin, 2007). However, also in this case it is debatable if aberrant DNA methylation is an early event and could therefore contribute to transformation and malignancy or if it is a late event, which potentially contributes to progression of the disease.

2.3.3.3. DNA methylation patterns and functions in plants and fungi

Compelling evidence from plants and fungi suggests that DNA methylation is crucial for genome immunity. Several recent reports showed that virtually all transposable elements are DNA methylated in *Arabidopsis thaliana* (Cokus et al., 2008; Lister et al., 2008; Vaughn et al., 2007; Zhang et al., 2006; Zilberman et al., 2007). The fungus *Neurospora crassa* goes even a step further and not only methylates transposons but subsequently deaminates methyl-cytosines. This introduces C-to-T transversions and leads to mutations and functional impairment of transposable elements; a process termed repeat-induced point mutation [RIP; (Galagan and Selker, 2004; Lewis et al., 2009)].

Moreover, studies in plants revealed that gene bodies of basally but not low or very highly expressed genes contain dense methylation, reminiscent of *C. intestinalis*. Upon deletion of Met1, the plant DNMT1 homolog, and subsequent loss of gene body methylation only very little transcriptional changes were observed (Zilberman et al., 2007). Thus, the function of this gene body methylation remains as elusive as in mammals with the best guess being the blocking of spurious initiation. Together this implies that in plants and vertebrates, gene body methylation serves an evolutionary conserved role, or alternatively, that gene body methylation has no functional relevance but is rather a consequence of a default methylation pathway that targets any sequence which is not protected against DNA methylation (see below).

2.3.3.4. Targeting DNA methylation

Even tough heavily studied, targeting of DNA methyltransferases to their specific target sequences remains poorly understood. Recently, two reports demonstrated that DNMT3L specifically interacts with histones devoid of H3K4 methylation, providing a hint how DNA methylation could be targeted and how DNA could be protected from *de novo* methylation (Jia et al., 2007; Ooi et al., 2007). Other reports suggest that DNMTs can be recruited via sequence-specific transcription factors such as Myc (Brenner et al., 2005) or Pu.1 (Suzuki et al., 2006), or by specific post-translational histone modifications and histone modifying enzymes. For example, H3K9 methylation is required for targeting DNA methylation towards pericentric major satellite repeats (Lehnertz et al., 2003) or to the promoter of Oct4 during cellular differentiation (Feldman et al., 2006). Moreover, Suv39h1/2, Setdb1 and G9a (all H3K9 methyltransferases), heterochromatin protein 1 (HP1), and Ezh2 (a H3K27 methyltransferase) have been implicated to recruit DNA methyltransferases in experiments mostly carried out in cancer cell lines (Epsztejn-Litman et al., 2008; Fuks et al., 2003; Vire et al., 2006). Despite all these putative interactions no predictive model about targeting of DNA methylation could be derived and most interactions await further functional characterization in non-transformed cells and *in vivo*.

In plants DNA methyltransferases (Cmt3, Drm1 and Drm2) can be targeted towards repeats and transposons via small non-coding RNA species (Matzke et al., 2007), while in fungi no such pathway has been observed (Freitag et al., 2004b). Due to the absence of an RNA-dependent RNA polymerase in mammals, it was thought that such a pathway is most likely not existent in vertebrates. The recent discovery of short germline specific RNA species, termed piRNAs, stimulated new speculations about the existence of an RNA-based targeting of DNA methylation in order to silence mobile elements. Indeed, upon deletion of the Argonaut homologs Piwil2 or Piwil4 DNA methylation is lost at transposable elements. In line with these data, deletion of DNMT3L causes an increase of short RNAs mapping to

transposons (Aravin et al., 2008; Aravin et al., 2007; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008). Together these results suggest that in the germline the piRNA pathway resides upstream of DNA methylation of transposable elements. However, the recruitment of DNMTs to piRNA targets remains to be determined.

2.3.3.5. DNA demethylation

DNA can be demethylated passively via blocking maintenance during DNA replication and thereby avoiding methylation of the newly synthesized DNA strand. Alternatively it has been proposed that DNA methylation can be removed actively via specific enzymes. While such a pathway has been identified in plants, active DNA demethylation is highly controversial in mammals as there is no compelling biochemical evidence for an efficient active demethylation pathway [reviewed in (Ooi and Bestor, 2008)]. During mammalian development there are two waves of epigenetic reprogramming when DNA methylation and PTMs are partially removed and reset (Reik, 2007): first in the pre-implantation embryo and later during primordial germ cell (PGC) development. The reprogramming of DNA methylation during pre-implantation development occurs very fast at the paternal genome as revealed by immuno-fluorescence staining for 5-methylcytosine. This would argue for an active demethylation process rather than passive “dilution” over subsequent cell divisions as it is probably the case on the maternal genome. However, this model is based on stainings with one antibody specific to methylated cytosine and a handful individual sequences assessed by bisulfite sequencing, thus the extent of demethylation is not clear. Remarkably, methylation of parentally imprinted sequences is protected from demethylation (Hirasawa et al., 2008; Nakamura et al., 2007).

During the second reprogramming event in PGCs, the cells are dividing and distinguishing between active removal or passive loss of DNA methylation is therefore difficult and further investigation is required.

Numerous studies reported active demethylation and proteins such as MBD2 [methyl-binding domain 2; (Bhattacharya et al., 1999)], Tdg [thymidine DNA glycosylase; (Jost, 1993)], Gadd45a (Barreto et al., 2007) and even DNMT3A and DNMT3B (Kangaspeska et al., 2008; Metivier et al., 2008) were suggested to mediate DNA demethylation. However, many characterized demethylase activities were either never followed up or were proven not reproducible by other labs as in the cases of MBD2 and Gadd45a (Jin et al., 2008). Taken together, the current data in vertebrate systems neither proves the existence of active demethylation nor excludes its necessity and/or existence.

In plants, the evidence for active demethylation is much more convincing and exclusively encompasses DNA glycosylases. Mechanistically, methylcytosines are deaminated, resulting in a thymidine which is then excised out of the DNA and replaced by

an unmodified cytosine via base excision repair (Gehring et al., 2006). For example the DNA glycosylase DEMETER was shown to be specifically required for removing DNA methylation from the MEDEA gene in female gametophytes, resulting in a paternally imprinted MEDEA gene after fertilization (Gehring et al., 2006). The other three proteins with demethylase activity (ROS1, DML2, and DML3) belong to the same gene family as DEMETER but are more ubiquitously expressed. Upon mutation of all three of them, DNA methylation significantly increases at several hundred loci (Penterman et al., 2007a; Penterman et al., 2007b; Zhu et al., 2007). Remarkably, the aberrant methylation preferentially takes place at 5' and 3' ends of genes, suggesting that in plants, unmethylated regulatory regions are actively protected against aberrant DNA methylation by a demethylation pathway.

Even though imprinting mechanisms in plants and mammals show remarkable similarities, they evolved independently. This is nicely illustrated by the way differential DNA methylation is set up in the maternal and paternal germline. During PGC development in mammals many epigenetic marks are erased and reset. Accordingly, differential methylation is established by *de novo* methylation mediated by a DNMT3A-DNMT3L complex in a parent of origin specific manner. In plants however, no such epigenetic reprogramming takes place and differential methylation is conveyed by demethylating one allele which is normally fully methylated in somatic cells (Zilberman, 2008).

2.3.3.6. Methyl-CpG binding proteins

Currently, three methyl-CpG binding protein families are known in vertebrates and plants: methyl-CpG-binding domain (MBD) proteins, Kaiso-like proteins and SRA-domain proteins [reviewed in (Clouaire and Stancheva, 2008; Dhasarathy and Wade, 2008; Hendrich and Tweedie, 2003)]. In general, these proteins are believed to bind to methylated DNA and to further recruit chromatin remodeling and modifying complexes, which assist in transcriptional repression. In addition, due to the high abundance and focal presence on chromatin, certain MBDs were implicated in chromatin organization and higher order structure formation (Brero et al., 2005; Clouaire and Stancheva, 2008). Below I will briefly summarize current knowledge about MBD proteins, which is the best characterized family.

Mammals contain five proteins (MeCP2, MBD1, MBD2, MBD3, and MBD4) which share a common and evolutionary conserved MBD sequence motif. In addition, a bioinformatics study identified 6 additional polypeptides which contain such a conserved MBD motif (Roloff et al., 2003). However so far none of these putative methyl-binding proteins was further characterized. *In vitro* and to some extent *in vivo*, the MBD domains of MeCP2, MBD1, MBD2, and MBD4 have been shown to preferentially bind to DNA harboring methylated CpGs. Further, they colocalize with foci of densely methylated satellite DNA,

which together suggested that MBD proteins might be readers of the DNA methylation mark (Hendrich and Bird, 1998).

MeCP2, the first MBD to be biochemically characterized, interacts with the Sin3A histone deacetylation co-repressor complex, which is crucial for MeCP2-mediated repression (Jones et al., 1998; Nan et al., 1998). Moreover, a vast number of other interaction partners have been reported, among them proteins of the repressive CoREST and NCoR/SMRT complexes, H3K9 KMT activity and the chromatin remodeling factors ATRX and Brm1 (reviewed in (Clouaire and Stancheva, 2008). MeCP2 deficiency in humans causes Rett syndrome, a severe autism spectrum disorder. Due to its location on the X-chromosome, male embryos with mutant MeCP2 die perinatally, while females show a mosaic expression of the wild type and mutant MeCP2 alleles. In addition, skewed X-chromosome inactivation can lead to diverse severity of the disorder. Mice carrying mutant MeCP2 recapitulate many human Rett symptoms and are widely used to study the disease (Chahrour and Zoghbi, 2007). Recent studies revealed that, in contrast to previous believes, Rett is not a neurodevelopmental or neurodegenerative disease because the re-expression of MeCP2 in mutant mice completely alleviated the neurological symptoms of Rett (Guy et al., 2007). This indicates that neurons develop normally and are not damaged in the absence of MeCP2, however MeCP2 is essential for their proper function. Recent large scale ChIP-chip data suggested that only a minority of MeCP2 binding sites co-localizes with genes and/or methylated CpG islands while most of the identified binding sites were outside of genes and CpG islands. Moreover, genes bound by MeCP2 were mostly active which suggest that MeCP2 might also have other functions than transcriptional control of methylated genes (Yasui et al., 2007).

MBD1 not only binds methylated CpGs but also unmethylated DNA and was therefore suggested to have repressive functions at both methylated and unmethylated CpG islands (Fujita et al., 2000). Interestingly, the domain responsible for binding to unmethylated DNA contains a CXXC motif, which is thought to specifically recognize unmethylated CpG islands and which is also present in DNMT1 and MLL [an H3K4 KMT; (Voo et al., 2000)]. However, to date it is unclear what the role of MBD1 is. Mice lacking MBD1 are viable and fertile but have reduced neurogenesis and other mild neurological defects (Zhao et al., 2003). The interaction of MBD1 with Setdb1, Suv39 (both H3K9 methyltransferases) and HP1 (which binds to H3K9me2/3) suggests a role in mediating H3K9 methylation in the context of DNA methylated sequences. Moreover it was found that during S-phase the MBD1/Setdb1 complex is displaced from chromatin and associates with Caf1 (chromatin assembly factor 1) at progressing replications forks, where it methylates H3K9 at the time when the new nucleosomes are incorporated into the nascent DNA (Sarraf and Stancheva,

2004). This provides a potential mechanism for propagating H3K9 methylation through mitosis.

MBD2 was found to be a member of the NuRD repressive complex but other than MBD3, its absence does not result in embryonic lethality, suggesting independent functions (Hendrich et al., 2001). MBD2 knock-out mice only display some neuronal defects but are otherwise viable and fertile (Hendrich et al., 2001).

MBD3, which has no methyl-CpG specificity, is a member of the chromatin remodeling and histone deacetylation complex NuRD (Wade et al., 1999; Zhang et al., 1999); however its role in the NuRD corepressor complex is unknown. Strikingly, MBD3 is the only MBD protein with an embryonic lethal phenotype in mutant mice (Hendrich et al., 2001).

MBD4 was more recently found to contain DNA glycosylase activity and rather than reading the DNA methylation it was shown to reduce the mutability of methylated CpGs in the genome. MBD4 recognizes T-G mismatches, which arise due to spontaneous deamination of methylated cytosines, and triggers base excision repair pathways to fix such missmatches before they eventually get propagated during mitosis (Millar et al., 2002; Wong et al., 2002).

SRA proteins such as UHRF1 play an important role in maintenance of DNA methylation (Arita et al., 2008; Avvakumov et al., 2008; Bostick et al., 2007; Sharif et al., 2007), and the diversity in plants suggests that they might also serve other functions. Similar to some MBDs, Kaiso-like proteins were also found to act as transcriptional repressors in conjunction with histone deacetylase complexes (Yoon et al., 2003), however very little is known about this family of methyl-CpG binding proteins.

Even though many interaction partners have been described, the phenotypes of the individual knock-out mice do not support developmentally crucial functions in reading DNA methylation. All DNMTs are essential for proper development, while MBD3, which does not bind to methylated DNA, is the only essential member of the MBD family. This might be due to partial redundancy among MBDs, but nevertheless it demonstrates that functions of methyl-CpG binding proteins still remain elusive.

2.4. Scope of thesis

During mammalian development, the fusion of two highly specific gametes results in a single totipotent cell that ultimately gives rise to an organism comprising over 10^{13} cells. Although all these cells do have the same genetic material, they have very distinct morphological and functional properties based on their characteristic gene expression programs. While gene expression is largely defined by the transcription factors present in a given cell type, it is speculated that epigenetic mechanisms contribute to establishing and maintaining such cell type specific gene expression programs. This model is further supported by the finding that the epigenome is partially reprogrammed during early embryonic development [reviewed in (Reik, 2007)]. Moreover, electron microscopy data suggested that chromatin is more condensed (Francastel et al., 2000) and mobility of heterochromatin protein 1 (HP1) is reduce in differentiated cells compared to undifferentiated cells (Meshorer et al., 2006). However, magnitude, dynamics and targets of this epigenetic reprogramming process remain poorly understood. When I started my PhD project in 2005 it was emerging that mammals do not contain one “epigenome” but rather that every somatic cell type has its own characteristic epigenetic landscape, which is tightly linked to the cell type specific gene expression pattern. This novel concept was greatly stimulated by the discovery of enzymes which can remove acetylation and methylation marks from core histones (Shi et al., 2004; Taunton et al., 1996). This provided direct evidence that epigenetic modifications are reversible and can be dynamically regulated in order to change the epigenetic state of a cell. However, no comprehensive datasets were available to back up this hypothesis. Moreover, given that cell type specific epigenomes exist, it seemed important to ask at what time point during development and cellular differentiation these cell type specific patterns would be established and how this would interplay with gene transcription. In addition, several protein-protein interaction data hinted towards a crosstalk between repressive pathways. These data were merely generated *in vitro* or from cancer cell lines and no functional *in vivo* evidence was available.

Studying epigenetic reprogramming *in vivo* in the early embryo has severe technical limitations due to the low number of cells and only allows immunofluorescence stainings for chromatin modifications, *in situ* hybridization and/or PCR analysis for individual loci. To circumvent this and to allow a more comprehensive analysis, we chose to make use of a recently established *in vitro* ES cell differentiation system (Bibel et al., 2004). This enabled us to generate large amounts of highly pure cell population of three distinct developmental stages. By then, it was accepted that DNA methylation has an important role in silencing of repetitive elements, genomic imprinting, x-inactivation and maintaining genome integrity, however it was still controversial if and to what extent DNA methylation contributes to cell

type specific gene regulation (Walsh and Bestor, 1999). Moreover, based on microscopy data and reduced mobility of chromatin proteins in differentiating cells compared to stem cells it was speculated that cellular differentiation would entail successive increase of epigenetic restriction from a very plastic pluripotent state towards a highly restricted unipotent terminally differentiated state (Meshorer and Misteli, 2006). To test this hypothesis, we first monitored the genome-wide dynamics and targets of DNA methylation during neuronal differentiation. Subsequently, we expanded the analysis towards Polycomb, which is a distinct repressive pathway with known functions in controlling key developmental regulators (Sparmann and van Lohuizen, 2006). According to the model, Polycomb targets are specified early in development and then either maintained if the gene is not activated or lost upon gene activation (Sparmann and van Lohuizen, 2006). However, it was not clear if such early specified Polycomb targets maintain the repression till the terminally differentiated state in case they are never activated. Another immediate question was if Polycomb would target additional genes in differentiating cells and contribute to epigenetic restriction and maintaining cell identity once stem cells lose pluripotency.

In order to relate potential changes in DNA methylation and Polycomb-mediated H3K27me3 to gene activity and a “transcriptionally permissive” chromatin state we further mapped Pol II and H3K4me2 using the same promoter microarrays.

3. Results

3.1. Lineage-Specific Polycomb Targets and De novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors

Mohn F., Weber M., Rebhan M. Roloff T.C., Richter J., Stadler M.B., Bibel M., and Schübeler D.

Summary

Pluripotent embryonic stem (ES) cells isolated from mouse blastocysts represent an early developmental stage before embryonic lineage-specification. This makes them a good model to study early epigenetic reprogramming events which coincide with lineage specification and later terminal differentiation. In this study we made use of a well-defined neuronal differentiation model (Bibel et al., 2004) to investigate the plasticity of the epigenome during the loss of developmental potential and the gain of neuron-specific features. We profiled DNA methylation, Polycomb-mediated H3K27m tri-methylation, RNA polymerase II occupancy and H3K4 di-methylation distribution on over 15'000 promoters in pluripotent ES cells, multipotent neuronal progenitors and unipotent terminally differentiated post-mitotic neurons. This enabled us to monitor two distinct phases of cellular differentiation: the loss of pluripotency during lineage commitment, and the terminal differentiation of the committed progenitor cells. We observed that DNA methylation at gene promoters is increasing during lineage-commitment, which provides experimental support for a long standing hypothesis (Reik, 2007; Weber et al., 2007). This finding is also in agreement with other recent studies investigating DNA methylation during differentiation and comparing different somatic cell types (Farthing et al., 2008; Illingworth et al., 2008; Meissner et al., 2008; Rollins et al., 2006; Weber et al., 2007). Interestingly, DNA methylation appears to preferentially target genes under the control of a weak CpG island promoter. Many of these function in early embryonic development and in maintaining pluripotency. This suggests that DNA methylation could have a role in securing the repression of the pluripotency program and thereby avoiding dedifferentiation and/or trans-differentiation. This is supported by findings from Mikkelsen and colleagues, who reported that the efficiency of reprogramming somatic cells into induced pluripotent stem (iPS) cells is greatly enhanced if cells are treated with the demethylating drug 5'aza-cytidine (Mikkelsen et al., 2008). Further, as the *de novo* DNA methylation takes place almost exclusively during lineage commitment of stem cells to neuronal progenitor cells, the data suggest that DNA methylation at CpG island promoters has little or no regulatory impact on the terminal differentiation and subtype specification.

However, due to detection limitations of our genome-wide approach we cannot rule out that methylation changes at individual CpGs can impact gene expression in a cell type specific manner as it was put forward for Bdnf regulation (Martinowich et al., 2003).

Interestingly, many *de novo* methylated promoters are not active in stem cells, suggesting that DNA methylation can not simply be a consequence of transcriptional shut down but rather needs to be specifically targeted to these regions. While the targeting mechanism remains elusive, one clue comes from the global distribution of H3K4me2. In ES cells virtually all unmethylated CpG-rich (i.e. strong and weak CpG islands) reside in chromatin marked by H3K4me2, irrespective of transcriptional activity. Upon *de novo* DNA methylation, H3K4me2 is lost which is in agreement with the mutual exclusivity of these two antagonistic marks in somatic cells (Weber et al., 2007) and suggests that DNA methylation is stably locking in a silent state of CpG-rich promoters of pluripotency and germline specific genes.

Polycomb-mediated H3K27me3 was previously shown to target many key developmental transcription factors in ES cells (Boyer et al., 2006; Lee et al., 2006) and to be crucial for embryonic development [reviewed in (Schwartz and Pirrotta, 2007)]. In contrast to DNA methylation which displays a unidirectional gain during lineage commitment, we show that H3K27me3 behaves very dynamic. Novel genes are targeted while others are activated and lose the H3K27me3 mark at both stages of differentiation, during lineage commitment of stem cells to neuronal progenitors as well as during terminal neuronal differentiation of the progenitor cells. This plasticity was rather unexpected as the classical model of Polycomb-mediated repression predicts that targets are already specified early in development and then either maintain the repressive H3K27me3 mark if the gene is not activated or lose it upon activation (Ringrose and Paro, 2007; Schuettengruber et al., 2007; Schwartz and Pirrotta, 2007; Sparmann and van Lohuizen, 2006). Strikingly, many genes which are inactive and unmodified in stem cells and which become Polycomb targets upon differentiation into neuronal progenitors have neuronal functions. At first glance this appears counterintuitive. However given that these genes are still inactive at the progenitor stage but eventually become activated upon terminal differentiation, it is reminiscent of the situation in stem cells. Our data suggests that Polycomb could act as a safe-guard mechanism to avoid spurious transcription in absence of a strong activation signal at genes which could be activated at a given cell stage but need to be kept silent such as developmental regulators in ES cells or neuronal genes in neuronal progenitor cells.

Remarkably, H3K27me3 targets are almost entirely CpG-rich promoters, indicating a targeting bias for the still elusive recruiting factors of Polycomb complexes. This preference for CpG-rich sequences furthermore provides a DNA sequence based explanation for the existence of so-called bivalent chromatin domains, where the repressive H3K27me3 and the active H3K4me2/3 marks colocalize (Bernstein et al., 2006) because CpG-rich sequences

are H3K4 methylated by default. Thus any further targeting of CpG-rich sequences by Polycomb will result in the formation of a bivalent chromatin configuration. If true this also predicts that such bivalent chromatin can not be specific to ES cells as originally proposed (Azuara et al., 2006; Bernstein et al., 2006), which we and others indeed have shown (Bracken et al., 2006; Meissner et al., 2008; Pan et al., 2007; Squazzo et al., 2006).

Interestingly, genes which are H3K27me3 modified in ES cells have a 4.5 fold increased likelihood of becoming DNA methylated upon differentiation as compared to other inactive genes in ES cells. Several recent studies suggested preferential aberrant DNA hypermethylation in human cancer cell lines and primary cancers at promoters that are Polycomb targets in unrelated human ES cells (Ohm et al., 2007; Schlesinger et al., 2007; Widschwendter et al., 2007). Thus, our data are compatible with a model in which H3K27me3 can trigger *de novo* DNA methylation during normal somatic differentiation at a subset of promoters, indicating that such crosstalk does occur but is not cancer-specific *per se*. Further, our data argue against a reported direct interaction of Polycomb proteins with DNMTs (Vire et al., 2006). A direct interaction would predict a marked overlap between DNA methylation and H3K27me3 targets, which we and others do not find in non-transformed cells.

Based on our analysis we propose that DNA methylation and Polycomb-mediated repression represent two distinct paradigms of epigenetic repression. DNA methylation is stably maintained and displays only a modest increase during early differentiation with a preference for pluripotency-associated genes, suggesting a function in stabilizing differentiated states. This is further consistent with a minor role in regulating cell type specific gene expression and a dominant role for DNA methylation in repeat and retrotransposons silencing (Walsh and Bestor, 1999). Polycomb on the other hand is a more dynamic and transient repression pathway, which potentially prevents precocious activation of developmental stage specific targets without appropriate stimuli. This might confer robustness to both intermediate and terminal stages of cellular differentiation.

Molecular Cell
Article

Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors

Fabio Mohn,[1] Michael Weber,[1,3] Michael Rebhan,[1,4] Tim C. Roloff,[1] Jens Richter,[2] Michael B. Stadler,[1] Miriam Bibel,[2] and Dirk Schübeler[1,*]
[1]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland
[2]Novartis Institutes for Biomedical Research, Neuroscience Research, Neurodegeneration Department, 4002 Basel, Switzerland
[3]Present address: Institut de Génétique Moléculaire, CNRS UMR 5535, 1919 Route de Mende, 34293 Montpellier Cedex 5, France
[4]Present address: Novartis Pharma AG, 4002 Basel, Switzerland
*Correspondence: dirk@fmi.ch
DOI 10.1016/j.molcel.2008.05.007

SUMMARY

Cellular differentiation entails loss of pluripotency and gain of lineage- and cell-type-specific characteristics. Using a murine system that progresses from stem cells to lineage-committed progenitors to terminally differentiated neurons, we analyzed DNA methylation and Polycomb-mediated histone H3 methylation (H3K27me3). We show that several hundred promoters, including pluripotency and germline-specific genes, become DNA methylated in lineage-committed progenitor cells, suggesting that DNA methylation may already repress pluripotency in progenitor cells. Conversely, we detect loss and acquisition of H3K27me3 at additional targets in both progenitor and terminal states. Surprisingly, many neuron-specific genes that become activated upon terminal differentiation are Polycomb targets only in progenitor cells. Moreover, promoters marked by H3K27me3 in stem cells frequently become DNA methylated during differentiation, suggesting context-dependent crosstalk between Polycomb and DNA methylation. These data suggest a model how de novo DNA methylation and dynamic switches in Polycomb targets restrict pluripotency and define the developmental potential of progenitor cells.

INTRODUCTION

The process of cellular differentiation is generally unidirectional and stably maintained. Reversion to stem cell status or transdifferentiation to alternative lineages is rarely observed without cellular transformation. This reduction of developmental potential was proposed to entail epigenetic restriction such as chromatin or DNA modifications, which can modulate DNA accessibility. These epigenetic mechanisms would in turn stabilize cell-type-specific gene expression patterns and reduce the likelihood that stem cell-specific or lineage-unrelated genes will be reactivated (Reik, 2007).

The repression of developmental gene regulation through histone modifications is best illustrated by the Polycomb group proteins (Ringrose and Paro, 2007). These conserved transcriptional repressors are required for correct body patterning as they control Hox gene expression (Schwartz and Pirrotta, 2007; Sparmann and van Lohuizen, 2006). A hallmark for Polycomb-mediated repression is methylation of lysine 27 of histone H3 (H3K27), which is set up by the Polycomb repressive complex 2 (PRC2) (Czermin et al., 2002; Muller et al., 2002). This mechanism targets genes encoding key transcription factors in mouse and human embryonic stem cells (ESCs) (Boyer et al., 2006; Lee et al., 2006), yet repression can be overcome during differentiation by transcriptional activators. Thus, Polycomb can be viewed as a reversible repression pathway for genes poised to be activated, ensuring that activation is only induced by a strong and specific stimulus.

This model implies that most targets are specified in stem cells and early embryo and then either maintain H3K27me3 or lose it upon transcriptional activation during development (Ringrose and Paro, 2007). However, it remains open whether Polycomb operates on additional targets in multipotent progenitor cells. Unlike embryonic stem cells, progenitors are restricted to a certain lineage but have the potential to differentiate into distinct terminal cell types upon stimulation. In CNS neurogenesis, for example, radial glial cells have been identified as progenitors that can differentiate into diverse neuronal subtypes (Goldman, 2003; Gotz and Barde, 2005). Although well documented, it is not known how such progenitor-specific "multipotency" is established and resolved upon terminal differentiation.

Many promoters of key regulators further bear the "active" lysine 4 methylation mark on histone H3 (H3K4) in addition to the repressive Polycomb-mediated H3K27 methylation (Bernstein et al., 2006; Mikkelsen et al., 2007; Pan et al., 2007; Zhao et al., 2007). While the determinants of this bivalency of active and repressive marks are not known, it has been hypothesized to be an ESC-specific chromatin state involved in creating a poised state amendable to rapid induction (Bernstein et al., 2006).

DNA methylation in the context of CpG dinucleotides is a distinct repression pathway in mammals that is considered to

mediate stable silencing (Goll and Bestor, 2005). It is essential for embryonic development (Okano et al., 1999), and methylation of promoters or regulatory regions is involved in X-inactivation in female mammals (Heard et al., 1997), in genomic imprinting (Li et al., 1993), and in silencing of parasitic elements (Goll and Bestor, 2005). However, the role of DNA methylation in tissue-specific gene expression is controversial: single gene studies showed that changes in DNA methylation at promoters can reflect changes in transcriptional activity but do not follow it in an obligatory fashion (Futscher et al., 2002; Walsh and Bestor, 1999).

It has been widely assumed that promoters in ESCs lack DNA methylation. This is based on the fact that ESCs are derived from blastocystes after a global DNA demethylation event (Howlett and Reik, 1991; Mayer et al., 2000). Furthermore, it was proposed that promoter methylation might be incompatible with the ability of such cells to activate a wide range of tissue-specific genes during subsequent differentiation (Reik, 2007). In a recent study of promoter DNA methylation in human sperm and primary fibroblasts, we observed that several promoters methylated in fibroblasts are unmethylated in sperm (Weber et al., 2007). This suggested that part of the promoter methylation detected in differentiated human somatic cells is set postfertilization. Nevertheless, how this finding translates to global DNA methylation in ESCs and multipotent progenitor cells in relation to their developmental potential remained an open question.

In vitro differentiation of mouse ESCs provides an opportunity to study the characteristics of pluripotency and epigenomic changes that coincide with cellular differentiation. However, most in vitro differentiation systems do not progress through a defined lineage-committed progenitor state toward one defined terminal cell type, although this progression is typical in vivo, e.g., in hematopoiesis or brain development. Here, we take advantage of a robust differentiation model for neurogenesis in which ESCs differentiate first into a highly pure population of Pax6-positive radial-glial neuronal progenitor cells and later into terminally differentiated glutamatergic pyramidal neurons (Bibel et al., 2004). Using this unique system, we track genome-wide epigenetic modification by Polycomb and DNA methylation during both the lineage commitment of ESCs and their terminal differentiation. These experiments present a comprehensive map of promoter DNA methylation during cellular differentiation and lead to a model of progenitor regulation. They reveal gain of DNA methylation during lineage commitment to restrict pluripotency and a parallel acquisition of Polycomb repression to new target genes poised to be activated in subsequent terminal differentiation.

RESULTS

Profiling of Promoter Epigenetic States during Cellular Differentiation

To define changes in DNA methylation and chromatin during cellular differentiation, we exploited a model which encompasses the synchronous generation of multipotent Pax6-positive radial glial cells from mouse ESCs with >90% efficiency (Bibel et al., 2004, 2007; Plachta et al., 2004). They are bona fide neuronal progenitors based on morphology, marker gene expression, and their ability to recapitulate in vitro the differentiation pathway of Pax6-positive radial glial cells during cortical development in vivo (Bibel et al., 2004; Heins et al., 2002; Plachta et al., 2004). Moreover, they are developmentally restricted to certain subtypes of neurons as shown by transplantation experiments in the chick embryo (Plachta et al., 2004). These progenitors can be terminally differentiated in vitro with equally high efficiency (~90%) into postmitotic glutamatergic neurons, characterized by the formation of synaptic connections and defined electrophysiological properties resembling cortical glutamatergic neurons (Figure 1A) (Bibel et al., 2004, 2007).

The formation of such uniform cell populations representing subsequent developmental stages allowed to monitor epigenome changes during two distinct differentiation phases. The first phase, when stem cells differentiate into multipotent neuronal progenitor cells, entails loss of pluripotency and acquisition of lineage specificity. The second phase, when progenitors differentiate into pyramidal neurons, reflects acquisition of a terminally defined identity. At each cellular state, we determined the transcriptome, DNA methylation using the MeDIP technique (Weber et al., 2005), as well as the presence of several histone marks and RNA-polymerase II by chromatin-IP (ChIP). As an experimental read-out for MeDIP and ChIP experiments, we used oligonucleotide tiling microarrays that represent 26,275 putative mouse promoter sequences (Figure 1C). Epigenome and transcriptome analysis were done on replicates of independent differentiations and revealed high reproducibility (Figure 1B and Figure S1).

For determining gene activity, we used both mRNA expression data, as determined by Affymetrix transcriptome analysis, as well as RNA Polymerase II (Pol II) abundance on promoters. We find that presence of Pol II is, in most cases, an excellent predictor of transcript abundance (Figure S2). Nevertheless, in a subset of genes, Pol II is present at the promoter, but no mRNA can be detected (Figure S2), which is in agreement with a recent report showing that a fraction of promoters is bound by stalled Pol II in human cells, ready to be rapidly induced (Guenther et al., 2007).

Data analysis was restricted to 15,100 start sites, which we validated as bona fide promoters (Supplemental Experimental Procedures). We have recently shown that presence of DNA methylation at promoters in human somatic cells is dependent on their CpG density (Weber et al., 2007). Motivated by this finding, we classified the mouse promoter set into CpG-poor promoters and into weak and strong CpG islands (Figure S3 and Experimental Procedures).

Promoter Methylome in Mouse ESCs and during Lineage Commitment and Neuronal Differentiation

The analysis of 15,100 promoters reveals that DNA methylation of CpG islands is largely absent in ESCs: only 0.5% of strong CpG island promoters are hypermethylated in ESCs. CpG-poor promoters, on the other hand, we find to be mostly methylated (Figure S4), consistent with a recent report on DNA methylation in stem cells (Fouse et al., 2008). Intriguingly, of the few methylated CpG islands in ESCs, many control germline-specific genes (e.g., *Dazl*, *Tuba3*, *Piwil1*, and *Spo11*; Figure S4). Using RNA Pol II as a measure for transcriptional activity, we find that

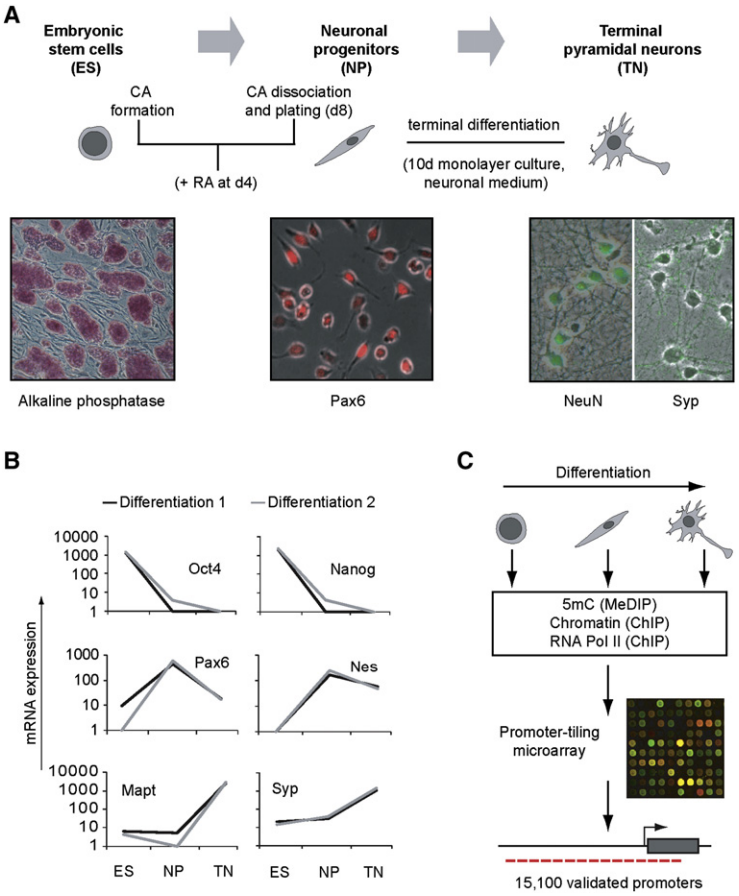


Figure 1. Cellular Differentiation System and Genomics Setup

(A) Mouse embryonic stem cells (ES) are differentiated to neuronal progenitors (NP) and, further, to terminal pyramidal glutamatergic neurons (TN). ESCs are stained for alkaline phosphatase, NPs for Pax6 (marker for radial glia), and TNs for NeuN (marker for postmitotic neurons) and Synaptophysin (Syp; marker for synapses). During differentiation, retinoic acid (RA) is added after cellular aggregate (CA) formation at day 4 after removal of Lif (d4), and cellular aggregates are dissociated by day 8 (NP) and further differentiated into terminal neurons on adherent substrate (see the Experimental Procedures).

(B) mRNA expression profiles of pluripotency factors (*Oct4* and *Nanog*), NP-specific genes (*Pax6* and *Nes*), and TN-specific genes (*Mapt* and *Syp*) from two independent differentiation experiments (gray and black lines).

(C) DNA methylation analysis by MeDIP and protein analysis by ChIP for H3K27me3, H3K4me2, and RNA polymerase II (Pol II) were performed for all three cellular stages. Samples were hybridized to microarrays covering 1.5 kb around the transcription start sites of 26,275 putative promoters, from which 15,100 were validated, classified according to CpG content, and used for further analysis.

this CpG island methylation is incompatible with promoter activity since Pol II is excluded from sites of DNA methylation (Figure S4). In contrast, methylation on CpG-poor promoters is compatible with presence of Pol II and, thus, with promoter activity (Figure S4), which is in line with our recent findings in human somatic cells (Weber et al., 2007). The overall absence of methylated CpG islands in ESCs indicates that DNA methylation cannot be a general mechanism to repress promoter activity in ESCs. Nonetheless, DNA methylation is present in ESCs: CpG-poor promoters in ESCs appear methylated to a similar extent as in somatic cells (data not shown).

To see if promoter methylation changes during differentiation, we generated an analogous profile for terminally differentiated neurons (Figure 2A). We find that the global pattern of promoter methylation is preserved between both developmental stages with CpG-poor promoters being methylated and the majority of CpG island promoters remaining unmethylated. However, at the same time, we observe a gain of DNA methylation during neuronal differentiation on 2.3% (n = 343) and a loss of DNA methylation on 0.1% (n = 22) of all tested promoters (Figure 2A). Single-gene controls by PCR and bisulfite sequencing confirm the microarray predictions for both loss and gain of DNA methylation (Figures 2B and 2C). Demethylation events are frequently linked to gene activation upon terminal neuronal differentiation, and six out of eight of these demethylated and activated genes are brain specific, such as *Lrrtm2* (Figure 2B). This suggests that tissue-specific loss of promoter DNA methylation does occur as previously suggested; however, in the system studied, it is a rare event.

The reciprocal gain of methylation occurs 23 times more frequently, resulting in hypermethylation of several hundred promoters (Figures 2A and 2C). Strikingly, however, we see that the majority of differentiation-coupled de novo methylation events are already present at the neuronal progenitor state with few changes occurring during the subsequent terminal differentiation (Figure 2D). Thus, DNA methylation changes correlate most strongly with commitment to a multipotent progenitor state, i.e., the step through which ESCs lose pluripotency, rather than with terminal differentiation. This is reflected in the biological function of the targeted genes. We find a striking enrichment for genes required for pluripotency of ESCs: seven out of 14 pluripotency-associated genes are de novo methylated; a rate 20 times higher than expected by chance ($p < 2.2 \times 10^{-16}$, χ^2 test). Thus, our unbiased analysis shows that de novo methylation is a prevalent mechanism for the repression of pluripotency-associated genes not limited to *Oct4* and *Nanog* (Deb-Rinker et al., 2005; Gidekel and Bergman, 2002). This pathway is particularly selective for promoters containing weak CpG islands: in this case, six out of seven promoters of pluripotency genes became methylated in neuronal progenitor cells (Figure 2E and Figure S5). Moreover, the target bias for weak CpG islands is not unique to pluripotency genes but is common among all de novo methylated promoters (Figure 2F). We conclude that weak CpG islands are preferentially controlled by DNA methylation during somatic differentiation.

Other promoters targeted for de novo methylation control germline-specific genes, such as *Papolb*, which is essential for spermatogenesis (Kashiwabara et al., 2002); *Dmrtc7*, a gene involved in male meiosis (Kim et al., 2007); and *Zar1*, an

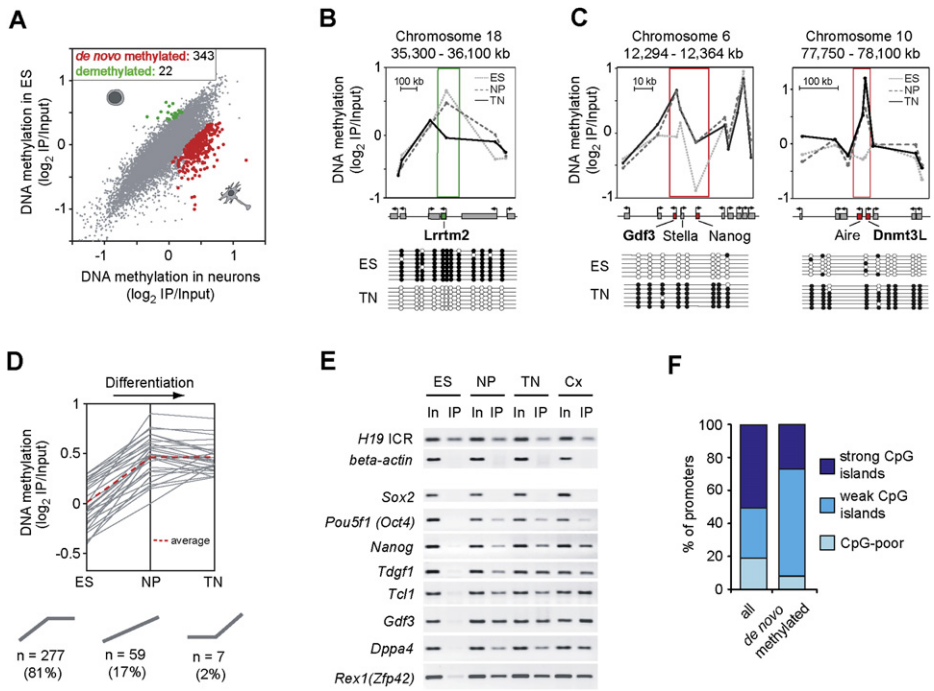


Figure 2. Predominant Gain of DNA Methylation during Neuronal Differentiation

(A) Scatter plot comparing averaged DNA methylation values from replicate microarrays for all 15,000 promoters in ES (y axis) versus TN (x axis). 343 promoters (red) significantly gain DNA methylation during differentiation, and 22 promoters (green) lose DNA methylation (see the Experimental Procedures).

(B) Example of a region containing a demethylated promoter (*Lrrtm2*; green box). The lines display methylation enrichment detected on the microarrays in ES (gray, dotted), NP (gray, dashed), and TN (black). Bisulfite sequencing of the region around the transcription start of *Lrrtm2* confirms the loss of methylation detected by microarray. Each circle represents a CpG either methylated (filled) or unmethylated (open).

(C) Examples of chromosomal regions containing promoters, which become DNA methylated during differentiation (red boxes). Bisulfite sequencing confirms de novo methylation around the transcription start for the promoters in bold (see also Figure 2E and Figure S6).

(D) Time course profiles for a random selection of 26 de novo methylated promoters. The dotted red line indicates the averaged profile for all de novo methylated promoters. Summarized schematic time course profiles (bottom) of all de novo methylated promoters illustrate that 81% gain DNA methylation from ES to NP.

(E) PCR validation of differentiation-coupled hypermethylation of promoters regulating pluripotency genes. PCR amplification of input (In) and immunoprecipitate (IP) after MeDIP was performed on ES, NP, TN, and primary mouse cortex (Cx). For further tissues and bisulfite sequencing, see Figure S5.

(F) Histogram showing promoter class distribution for all and for de novo methylated promoters. De novo methylated promoters are strongly enriched for weak CpG islands ($p < 2.2E\text{-}16$, χ^2 test).

oocyte-specific gene required for the oocyte-to-embryo transition (Wu et al., 2003). Furthermore, targets include *Dnmt3L* (Figure 2C and Figure S6), a germline-specific cofactor for the de novo methyltransferases Dnmt3a and Dnmt3b, which is essential for establishing genomic imprinting marks during germ cell development (Bourc'his et al., 2001).

Finally, many de novo methylation targets are neither pluripotency-related nor germline-specific genes. Among them, we find genes with essential roles in early embryonic development such as *Lefty1* and *Lefty2*, which guide left-right asymmetry in early embryonic development (Meno et al., 1998). Gene ontology analysis of all de novo methylated promoters identified an additional group of somatically expressed tissue-specific genes related to signaling and extracellular matrix (Figure S6). Among these is *Fes* (Feline Sarcoma Oncogene), a protein-tyrosine kinase involved in innate immune response and inflammation (Greer, 2002); *Amn* (Amnionless), a transmembrane protein essential for amnion formation during embryonic development (Kalantry et al., 2001); and *Cldn5* (Claudin 5), a component of tight junctions involved in formation of the blood-brain barrier (Nitta et al., 2003).

It is important to note that all pluripotency-associated targets and germline-specific genes that we found to undergo differentiation-coupled de novo methylation are methylated in vivo in the brain (cortex) and in all other mouse primary tissues tested (Figure 2E and Figures S5 and S6). For other tested targets, hypermethylation was detected in seven of ten cases in mouse cortex even though cortex consists of various neuronal subtypes and other nonneuronal cells. These targets show variable methylation in nonneuronal tissues, suggesting that their de novo methylation is not soma wide, but neuron specific (Figure S6).

Taken together, we find that terminal differentiation of ESCs leads to a predominant gain of DNA methylation at several hundred promoters, which share features with respect to promoter sequence composition and gene function. Surprisingly, we observe relatively few changes between neuronal progenitors and terminal neurons. This indicates that promoter hypermethylation is less dynamic during the transition to a terminally differentiated state and instead characterizes the transition from pluripotent ESCs to a lineage-restricted progenitor state.

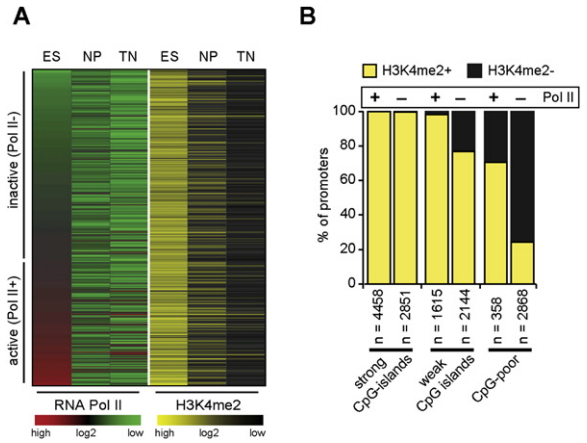


Figure 3. De Novo DNA Methylation Defines Gene Silencing and Loss of Active Chromatin

(A) Heat map illustrating the transcriptional status and chromatin conformation for all 343 de novo DNA-methylated promoters (y axis). 38% are bound by Pol II (red) in ES whereas 62% are not bound by Pol II (green). Upon differentiation, de novo methylated promoters mostly lose Pol II or remain Pol II negative. In contrast, 99% are H3K4me2 positive (yellow) prior to de novo methylation, irrespective of transcriptional state, but lose H3K4me2 when DNA methylated. (B) Bar graph showing that CpG-rich promoters are H3K4 methylated (yellow) irrespective of transcriptional state in ES (for NP and TN see Figure S7). A similar result is obtained for H3K4 trimethylation, which colocalizes with H3K4 dimethylation at almost all promoters (Figure S7 and data not shown).

De Novo Methylation Is Incompatible with Active Chromatin

Next, we assessed the effect of de novo DNA methylation on promoter activity and chromatin structure. We find that de novo methylated promoters are devoid of Pol II in neuronal progenitors and in terminally differentiated neurons, confirming that DNA methylation at CpG island promoters precludes transcription (Figure 3A). Interestingly, more than half (62%) of these targets are already transcriptionally inactive in ESCs prior to de novo DNA methylation (Figure 3A). At these genes, DNA methylation is not simply induced by differentiation-coupled shutdown of transcription. Instead, additional genetic or epigenetic cues may be needed to trigger hypermethylation (see below).

As an indicator for "active chromatin," we monitored dimethylation of lysine 4 of histone H3, which was previously shown to be a marker of active genes in several eukaryotes tested (Barski et al., 2007; Pokholok et al., 2005; Schübeler et al., 2004). Notably, we observe equal presence of H3K4 di- and trimethylation at promoters in agreement with previous global profiling experiments (Figure S7) (Barski et al., 2007; Heintzman et al., 2007).

We detect H3K4me2 at almost all active genes (97%) in ESCs independent of promoter structure (Figure 3B and Figure S7). In addition, H3K4me2 is present on most CpG island promoters of inactive genes (Figure 3B and Figure S7), suggesting that, in mammals, H3K4me2 is not exclusively located to actively transcribed genes and could serve additional functions. Importantly, this is restricted to CpG-rich (weak and strong CpG island) promoters, since we do not find many inactive CpG-poor promoters associated with H3K4me2 (Figure 3B). Thus, CpG-rich promoters reside in a chromatin environment implicated in gene activation, even in the absence of transcription (Guenther et al., 2007; Pan et al., 2007; Zhao et al., 2007). Furthermore, this default H3K4 methylation of CpG-rich promoters is independent of the cellular state as we observe it in ESCs, progenitors, and terminal neurons (Figure S7). This shows that K4 dimethylation is present at CpG islands irrespective of activity, yet those CpG islands that become DNA methylated lose H3K4me2 (Figure 3A and Figure S7), as has been predicted based on studies in human primary cells (Weber et al., 2007).

Plasticity of Polycomb Targets at Progenitor and Terminal State

Polycomb-mediated repression provides the molecular basis of a cellular memory system that propagates transcriptional states through cell division (Ringrose and Paro, 2007). However, while DNA methylation is considered a stable epigenetic silencer, Polycomb-mediated repression is reversible, as targets such as Hox genes are activated in a cell-type-specific manner (Agger et al., 2007; Bracken et al., 2006; Lan et al., 2007; Lee et al., 2007; Mikkelsen et al., 2007). Transcriptional repression by Polycomb entails PRC2-mediated trimethylation of lysine 27 of histone H3 (H3K27me3) (Cao et al., 2002). To ask how Polycomb repression changes during lineage commitment and terminal differentiation and, further, how this reprogramming relates to the observed changes in DNA methylation, we mapped H3K27me3 during neuronal differentiation.

Lineage-specific transcription factors and Hox gene clusters are primary targets of H3K27me3 in ESCs, which is in agreement with work by others (Bernstein et al., 2006; Boyer et al., 2006). As expected, H3K27me3 and Pol II are anticorrelated and consequently most (86%) Polycomb targets are not bound by RNA polymerase (Figure S8). However, a subset of H3K27me3-modified promoters is bound by Pol II, suggesting that they can coincide as previously reported (Bracken et al., 2006; Pan et al., 2007).

Upon lineage commitment, genes with functions in the development of anatomical structures, morphogenesis, and early embryonic development lose H3K27me3 and become activated (Figure 4A and Table S1). H3K27me3 targets that lose the modification during terminal differentiation from progenitor to pyramidal neuron are highly enriched for genes involved in neuronal development, ion transport, and neurotransmitter regulation (Group 1 in Figures 4B and 4C and Table S1), in line with previous findings that many neuronal genes are targeted by Polycomb in ESCs (Boyer et al., 2006; Pan et al., 2007). Most of them (68%) are transcriptionally active in terminal neurons. This group of genes, thus, supports current models that repression by Polycomb is set early in development but is lost in a lineage-specific manner upon gene activation (Ringrose and Paro, 2007; Schwartz and Pirrotta, 2007).

Surprisingly, however, and in parallel to this loss, we observe a coincident gain of Polycomb-mediated H3K27me3 at other targets (Figure 4A). This unexpected plasticity results in a rather constant number of genes that are repressed by Polycomb at any developmental state (Figure 4A). Indeed, we find the majority of K27me3 positive promoters to be cell-state specific. When asking if these are enriched for certain biological functions, we noticed that genes that become H3K27me3 in progenitor cells are involved in neurogenesis and function in differentiated

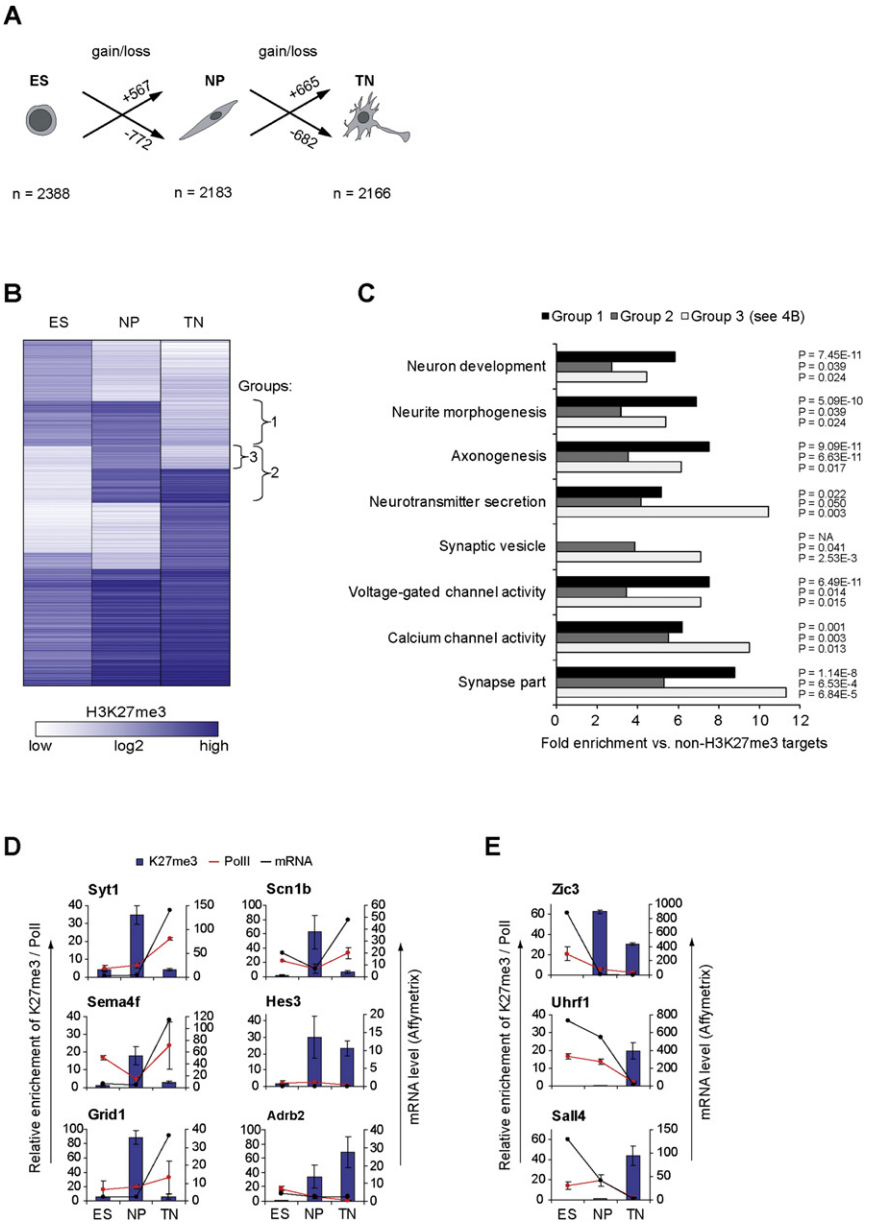


Figure 4. Polycomb Targets Are Highly Dynamic and Stage Specific

(A) Illustration of H3K27me3 target dynamics during neuronal differentiation. Arrows indicate loss (–) and gain (+) of Polycomb targets between the cellular states. "n" indicates the total number of H3K27me3 modified promoters at every individual state.

(B) Heatmap for all promoters that are H3K27me3 positive in at least one cell state. Only 43% remain H3K27me3+ throughout the differentiation, while the majority behaves highly plastic (see text).

(C) GO term analysis for genes that lose H3K27me3 in terminal differentiation to TN (Group 1, black), for genes that become Polycomb targets in NP (Group 2, gray), and for Polycomb targets that are specific for NP and lose H3K27me3 during terminal differentiation (Group 3, white). p values are listed next to bars, while NA indicates no significant enrichment in the respective group.

(D) Validation of microarray results for NP-specific H3K27m3 targets by ChIP and real-time PCR. Blue bars represent H3K27me3 enrichments, and red lines indicate Pol II enrichment (left y axis, numbers normalized to an intergenic control). Black lines indicate mRNA levels (Affymetrix, right y axis). *Syt1*, *Sema4f*, *Grid1*, and *Scn1b* are induced upon terminal differentiation and lose H3K27 methylation. *Hes3* and *Adrb2* are not activated and keep H3K27me3. Error bars indicate ± SEM of averages from at least two independent differentiation experiments.

(E) Examples of genes that become repressed in NP (*Zic3*) or TN (*Sall4* and *Uhrf1*) coinciding with a gain of H3K27me3.

neurons (Group 2 in Figures 4B and 4C and Table S1). Thus, upon commitment to a neuronal lineage, many genes that will be expressed only in terminal neuronal subtypes become novel Polycomb targets in neuronal progenitors.

In agreement with this model, we find that many progenitor-specific targets (54%) become activated and lose Polycomb-repression upon terminal differentiation (Group 3 in Figures 4B and 4C). Importantly, these genes are silent in both stem cells and progenitors, but only H3K27me3 in progenitors as confirmed by real-time PCR (Figure 4D and Figure S8). This class of over 200 genes is enriched for neuronal function, ion transport, and cell motility and includes *Grid1*, a glutamate receptor; *Syt1*, a Ca^{2+} sensor involved in neurotransmitter release (Geppert et al., 1994); *Scn1b*, a neuronal voltage-gated sodium channel; and *Sema4f*, a brain-specific semaphorin potentially involved in axon guidance (Figure 4D and Table S1). Other genes become Polycomb-repressed in neuronal progenitors but keep the H3K27me3 mark and are not induced in pyramidal neurons (Figures 4B and 4D). This group contains genes that define functions characteristic of nonglutamatergic neurons such as *Adrb2* (beta-2-adrenergic receptor) and *Hes3* (hairy and enhancer of split 3) (Figure 4D and Figure S8).

Finally, among the promoters that become Polycomb-repressed only in the postmitotic terminal neurons, we find the cell-cycle regulators cyclinD1 (*Ccnd1*) and *Uhrf1*. *Uhrf1* has further been shown to be essential for maintenance DNA methylation during cell division (Bostick et al., 2007; Sharif et al., 2007), a property that is no longer required in postmitotic cells. Furthermore, we find *Zic3* and *Sall4*, both pluripotency-associated genes that do not undergo de novo DNA methylation but are silenced during differentiation (Figure 4E).

In summary, PRC2 targets appear to be highly plastic during neurogenesis and novel targets of Polycomb-repression surface at both the progenitor and the terminal neuron state, implying

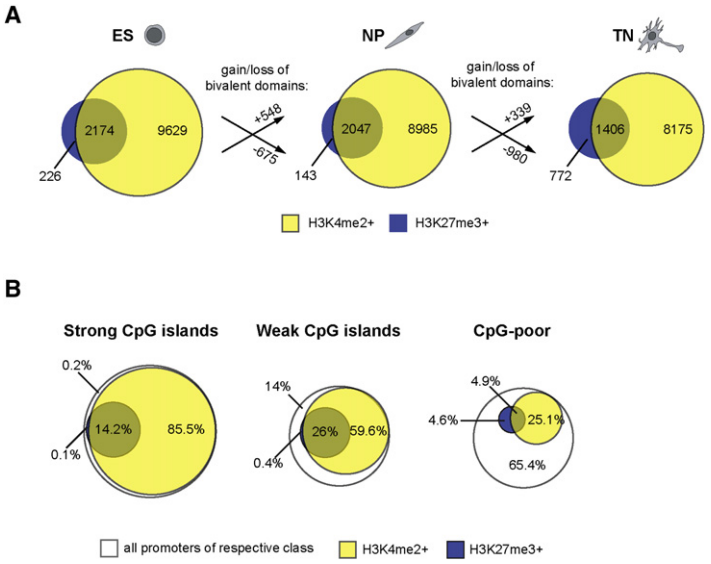


Figure 5. Bivalent Domains during Differentiation and Their Dependency on Promoter Sequence

(A) Venn diagram of H3K27me3 and H3K4me2 showing that the majority of Polycomb targets are also H3K4 methylated and, thus, in a "bivalent" state. This is the case in all three cellular states, yet to a different degree. Note that new bivalent domains form at any differentiation step. (B) Distribution of H3K4me2- and H3K27me3-positive promoters relative to promoter CpG content in stem cells. Venn diagram of H3K27me3 and H3K4me2 showing that H3K27me3 is more frequent in weak and strong CpG islands, as is bivalency, since these promoters are mostly H3K4 methylated.

cell-type specific Polycomb targeting. Progenitor-specific targets include genes that need to be activated upon further terminal differentiation, suggesting an anticipation and regulation of further differentiation choices in multipotent progenitor cells by the Polycomb pathway.

Bivalent Chromatin Domains Are Dynamic and a Function of Promoter Structure

Many Polycomb targets in ESCs have been shown to reside in a chromatin state characterized by the dual presence of "repressive" H3K27 methylation and "active" H3K4 methylation (Bernstein et al., 2006). Since this "bivalency" was hypothesized to be an ESC-specific chromatin state that poises for differentiation-coupled activation (Bernstein et al., 2007), we wondered about the H3K4 methylation of progenitor- and terminal neuron-specific targets of Polycomb. We find that in progenitors, 95% of novel H3K27me3 targets form bivalent chromatin (Figure 5A). This suggests that progenitor-specific Polycomb targets behave similarly to ESC-specific targets, not only in regards to potential activation upon further differentiation (see above), but also in their chromatin state.

We and others have recently shown that H3K4 methylation is present at CpG-rich promoters in the human and mouse genomes even when these promoters are inactive (Barski et al., 2007; Guenther et al., 2007; Weber et al., 2007). Importantly, almost all of the bivalent promoters contain CpG islands (93%), indicating that promoter sequence composition may be the critical parameter that favors the parallel presence of K4 and K27 methylation. This model is further supported by the target preference of PRC2 toward CpG islands. We find that 85% of all PRC2 target sites are CpG-rich, which means they belong to the weak or strong CpG island class (Figure 5B and Figure S9). Again, this occurs irrespective of cellular state. A similar CpG island preference can be deduced from published data sets on PRC2 target genes in ESCs (Figure S9). This excludes the criticism of an experimental bias arising from our focus on stringently filtered promoter elements. Thus, this PRC2 preference for CpG islands, which are H3K4 methylated based on their sequence, indicates that bivalent chromatin is not limited to ESCs.

Polycomb Targets in ESCs Become DNA Methylated during Lineage Commitment

Unexpectedly, we do find that promoters that are marked by H3K27me3 in ESCs are 4.5 times more likely to become de novo DNA methylated during neuronal differentiation than promoters that are not PRC2 targets (Figure 6). This accounts for a significant fraction of promoters that acquire de novo DNA methylation. In particular, two-thirds of de novo methylation targets, which are not transcribed in ESCs (Figure 3A), carry Polycomb-mediated H3K27me3. This establishes that Polycomb targets in stem cells are subject to de novo methylation during normal development, which is compatible with the model that Polycomb repression and de novo DNA methylation are linked.

Since most H3K27me3 targets are in a bivalent chromatin structure (see above), we wondered if DNA methylation could provide a means to resolve bivalent domains. There appear to be two major ways of resolving bivalency during cellular differentiation. One is gene activation, which coincides with a loss of H3K27 methylation. The second involves loss of H3K4 methylation, although the repressive H3K27me3 mark is kept, and the gene is not activated (data not shown). Many of these bivalent promoters, which resolve bivalency by loss of H3K4me2, are indeed de novo DNA methylated (21%). Thus, de novo DNA methylation could lock genes in a silent state, which were poised to be activated in ESCs.

DISCUSSION

Using a well-defined cellular differentiation model, we monitored reprogramming of the epigenome during three consecutive developmental states that represent stem cells, lineage-committed progenitors, and terminally differentiated neurons. Our findings are as follows: methylation of CpG islands occurs during loss of pluripotency, and primarily weak-CpG islands and pluripotency genes become de novo methylated and stably silenced. Surprisingly, we detect little additional promoter DNA hypermethylation as cells terminally differentiate. In contrast, Polycomb-dependent H3K27me3 is found to be present on promoters at all stages of differentiation. Its presence changes as cells pass through the progenitor state, with distinct populations of genes both gaining and losing H3K27 methylation. Strikingly, at the

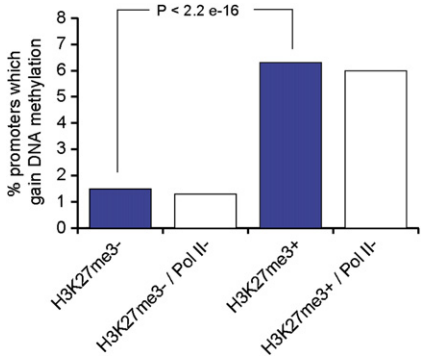


Figure 6. De Novo DNA Methylation of Stem Cell Polycomb Targets
Bar graph illustrating the percentage of promoters, which undergo differentiation-coupled de novo DNA methylation. H3K27-methylated promoters (H3K27me3+) in stem cells are 4.5 times more likely to become de novo DNA methylated than H3K27-unmethylated promoters (H3K27me3−; blue bars). The same preference is observed when we control for any potential bias from housekeeping genes that are constitutively active by only using promoters that are Pol II negative in stem cells (white bars; $p < 2.2E-16$, Wilcox rank-sum test).

progenitor state, the Polycomb mark becomes enriched on neuronal subtype specific genes, including those that will be expressed in the pyramidal lineage studied here and others that would be expressed only in other neuronal subtypes. This argues that Polycomb primes for both activation and inactivation during terminal differentiation in a progenitor-specific fashion. Thus, we present evidence that Polycomb-mediated gene regulation is utilized to define the developmental potential of multipotent progenitor cells, which challenges the view that targets are mostly predetermined in stem cells.

Finally, we see a strong bias of CpG islands for being controlled by H3K27me3, which provides a DNA sequence rationale for a cell-type-independent presence of "bivalent" chromatin. The implications of these findings are discussed below.

Gain of Promoter DNA Methylation during Lineage Commitment

We find ESCs to be mostly free of methylated promoter CpG islands, although CpG-poor promoters are methylated at similar levels as those in differentiated cells. Together with a recent report on DNA methylation in stem cells (Fouse et al., 2008), this provides unbiased experimental support for the popular hypothesis that ESCs specifically lack CpG-rich promoter methylation, possibly to maintain their ability to activate any gene later on. However, it also suggests that promoter DNA methylation is not the prevalent mechanism for transcriptional repression in ESCs. Consistently, ESCs can proliferate, but not differentiate, in absence of the maintenance methyltransferase Dnmt1 (Jackson et al., 2004; Li et al., 1992). Upon commitment to a defined neuronal lineage, only few promoters lose methylation, although many become activated. It remains to be tested if this rare loss of methylation is involved in regulation of these genes or if it is a consequence of transcriptional activation (Lin and Hsieh, 2001). Importantly, as cells achieve the progenitor state, we observe a gain of DNA methylation on several hundred promoters, which is 23-fold more frequent than the reciprocal loss. De novo methylation results in loss of a marker of active chromatin (H3K4me2) and absence of polymerase recruitment, which is maintained upon terminal differentiation. The population of de novo methylated genes is enriched for pluripotency- and germline-specific genes, arguing that DNA methylation has a major role in ensuring the stable repression of transcripts that are required for ESC maintenance. Accordingly, our model predicts that de novo methylation of these genes is not restricted to neurogenesis. Indeed, we find pluripotency- and germline-associated genes methylated in all somatic cells tested. This is likely to be a process conserved throughout mammalian species, since promoter methylation of germ-line specific genes was also reported in human soma (Shen et al., 2007; Weber et al., 2007).

A second class of genes that become DNA methylated in radial-glial progenitors contains genes that are selectively activated in other somatic lineages, even in subregions of the brain (Su et al., 2004). This suggests that DNA methylation could also contribute to lineage choice and/or restriction in neurogenesis. Nonetheless, we do not observe additional de novo methylation upon terminal differentiation. We conclude that de novo methylation has a minor role in the terminal steps of radial glial cell differentiation as compared to its function in lineage commitment and loss of pluripotency. It remains to be tested whether this holds for differentiation in other lineages.

We note a remarkable bias of de novo methylation for weak CpG islands. As this promoter class has also been shown to be more prone to hypermethylation in human primary fibroblasts (Weber et al., 2007), it appears a conserved target between human and mouse. Furthermore, we see a 4.5-fold increased frequency of de novo DNA methylation of promoters that bear the Polycomb H3K27me3 mark (Figure 6). This observation is compatible with a role for Polycomb in targeting DNA methyltransferase activity, which, if indeed the case, could account for 45% of all observed de novo methylated promoters. Nonetheless, this group reflects only a minor fraction of all Polycomb targets (6%). Thus, while H3K27 increases the frequency for de novo methylation, it is clearly not sufficient by itself to target DNA methylation as might be predicted from reported interaction between DNA methyltransferases and the H3K27 methyltransferase EZH2 in human cancer cell lines (Vire et al., 2006). Several recent studies suggested preferential aberrant DNA methylation in human cancer cell lines and primary cancers at promoters that are Polycomb targets in unrelated human ESCs in culture (Ohm et al., 2007; Schlesinger et al., 2007; Widschwendter et al., 2007). Our data are compatible with a model in which H3K27me3 can trigger de novo DNA methylation during normal somatic differentiation at a subset of promoters, indicating that such crosstalk does occur but is not cancer-specific per se. Misregulation of this process could indeed contribute to cancer-specific aberrant promoter methylation.

Importantly, DNA methylation precludes H3K4me2, and hence, its function could be to reduce the risk of reactivating CpG-rich promoters that we show bear this mark of "active" chromatin when DNA unmethylated. Thus, in contrast to bivalent chromatin, which poises promoters for activation, DNA methylation could be employed to lower the chance of spurious activation, which is more likely to occur in a chromatin environment

permissive for transcription. While such stabilization of an off-state has already been suggested for methylation at the *Oct4* promoter (Feldman et al., 2006), we argue that this is likely to be a general mechanism for locking in the silent state of pluripotency- and germline-associated genes in order to prevent reactivation once stem cells commit to a certain lineage. De novo methylation could serve a similar function for those genes, which are already transcriptionally silent in stem cells but nevertheless become methylated upon differentiation.

Polycomb Targets Are Developmental Stage Specific

Our results indicate that only a subset of Polycomb targets is specified in ESCs and that, upon cell-fate commitment, novel lineage-specific genes become H3K27me3 in progenitor cells. Many of these progenitor-specific Polycomb targets are key genes of subsequent developmental fates, and their repression can again be overcome upon terminal differentiation (Figure 4B). It is important to note that many of these genes are not transcribed in ESCs, which could reflect absence of activators or alternative means of repression. This is particularly interesting since a different set of neuronal genes is already targeted by Polycomb in stem cells, implying the need for active repression for some neuronal genes even in stem cells.

Regardless, the progenitor-specific and transient repression by Polycomb suggests that Polycomb functions to ensure that further developmental decisions are firmly controlled by robust induction signals strong enough to overcome the effect of H3K27 methylation. These dynamic Polycomb targets behave very different to Hox genes, which we only observe as actively transcribed or Polycomb repressed, fitting the general model of default Polycomb repression (Ringrose and Paro, 2007; Schwartz and Pirrotta, 2007). The observation of context-dependent targeting might rely on the recruitment of Polycomb by sequence-specific transcription factors. This model is compatible with recent reports of distinct Polycomb targets in transformed human cell lines (Bracken et al., 2006; Squazzo et al., 2006), human T cells (Barski et al., 2007; Roh et al., 2006), and differentiating stem cells (Pasini et al., 2007), which, however, could not be placed in the developmental history of the studied cell types.

Importantly, our findings also suggest a possible function of Polycomb for the regulation of adult stem cells, as the radial glial cells studied here have been delineated as multipotent progenitor cells of in vivo neurogenesis in the adult CNS (Malatesta et al., 2003).

Bivalent Domains Are a Consequence of Promoter Sequence and Polycomb Target Preference

We show that both DNA methylation and Polycomb modify CpG-rich promoters (Figure 2 and Figure S9) that control tissue-specific genes. These CpG island promoters display a mark of active chromatin (H3K4me2), even when not transcribed, as long as they are not DNA methylated. When analyzing H3K27me3 targets, we find that 85% belong to the CpG-rich promoter classes and bear H3K4 methylation by default. Therefore, we propose that a bivalent (K4/K27) chromatin state is a consequence of PRC2 target bias and is not ESC-specific. Indeed, we show that bivalent domains form de novo at progenitor and terminally differentiated states, invariably at unmethylated CpG-rich promoters (Figures 5A and 5B). In addition to this reorganization, 41% of ESC bivalent domains are preserved after differentiation into terminal pyramidal neurons. This agrees with a recent report that 43% of similarly modified ESC domains persist in fibroblasts (Mikkelsen et al., 2007). However, this same study reports many fewer bivalent domains in hyperproliferative neuronal stem cells (Mikkelsen et al., 2007), which are derived by monolayer differentiation and result in a population clearly distinct from the radial-glial cells we analyzed here (Conti et al., 2005). We observe little fluctuation in the number of bivalent promoters, as loss is largely compensated by newly formed bivalent domains at progenitor and terminal states. This result is compatible with the frequency of Polycomb targets in terminally differentiated human T cells and mouse fibroblasts (Mikkelsen et al., 2007; Roh et al., 2006). Since terminal neurons were not profiled by Mikkelsen and colleagues, we can only speculate that cellular heterogeneity of neuronal stem cells, which would dilute cell-type specific targets, might account for this discrepancy (Merkle et al., 2007).

It is noteworthy that the presence of H3K4 methylation at inactive CpG island promoters has so far only been observed in mammals, which show global DNA methylation and resulting diversity in promoter CpG content (Barski et al., 2007; Guenther et al., 2007; Weber et al., 2007). In organisms that lack widespread DNA methylation such as *Drosophila melanogaster*, H3K4me2 appears to mark exclusively active promoters (Schübeler et al., 2004). It thus seems conceivable that bivalent chromatin states and their dependence on Polycomb and unmethylated CpG islands are restricted to vertebrates.

In summary, these results show that mammalian pluripotent ESCs are unique with respect to targets of Polycomb and DNA methylation, but not with respect to the histone modifications established at these sites. Our global analysis of two repressive epigenetic pathways has provided a blueprint of the epigenome during three consecutive stages of mammalian neurogenesis and suggests a model to explain the developmental potential of progenitor cells. They are restricted through silencing of pluripotency-associated genes by DNA methylation, while gain of neuronal plasticity is defined by lineage-specific targets of Polycomb.

EXPERIMENTAL PROCEDURES

Cell Culture and Tissue Samples

Wild-type embryonic stem cells were derived form blastocysts (3.5 PC) of mixed 129-C57Bl/6 background and cultivated on feeder cells (37°C, 7% CO_2). Differentiation was performed essentially as described (Bibel et al., 2007). In brief, ESCs were deprived of feeder cells during 3 to 4 passages, then 4 × 10^6 cells were used for formation of cellular aggregates (CAs). CAs were cultivated in nonadherent bacterial dishes for 8 days. Retinoic acid (5 uM) was added from day 4 to day 8. Subsequently, CAs were dissociated and plated on cationic substrate coated with laminin. Forty-eight hours later, a medium enriched with supplements was added for 8 days of terminal neuronal maturation.

Primary tissue samples were dissected from wild-type mice 4–6 weeks after birth. Samples were homogenized, and genomic DNA was isolated for subsequent MeDIP.

Immunofluorescence

Immunofluorescence stainings were performed as previously described (Bibel et al., 2004) using the following antibodies and dilutions: Pax6, mouse

monoclonal (1:100, Developmental Studies Hybridoma Bank); NeuN, mouse monoclonal (1:200, Chemicon); and Synaptophysin, mouse monoclonal, (1:200, Sigma). For Alkaline Phosphatase stainings, a Kit (Chemicon, Cat. No. SCR004) was used according to the manufacturer's protocol.

DNA Methylation Profiling by MeDIP

MeDIP was performed as previously described (Weber et al., 2007) using 3 μg sonicated (300–1000 bp) genomic DNA as starting material and 10 μg antibody against 5-methylcytidine (Eurogentec, BI-MECY-1000). For PCR, 20 ng sonicated genomic input DNA and 1/40 of a MeDIP reaction were used. For microarray analysis, 7 unamplified MeDIP reactions were pooled and hybridized together with sonicated genomic input DNA as reference. Final promoter methylation $\log_2$ ratios of IP over input signal represent the average of three independent experiments, including one dye swap.

Chromatin-IP

ChIP experiments were done as described in Weber et al. (2007), starting with 70 μg of chromatin and 5 μg of the following antibodies: anti-trimethyl-H3K27 (Upstate, no. 07-449), anti-dimethyl-H3K4 (Upstate, no. 07-030), and anti-RNA Pol II (Santa Cruz Biotechnology, no. SC899). For hybridization to microarrays, samples were amplified by LMPCR. Promoter $\log_2$ values are the averages of at least two biological replicate experiments, including one dye swap microarray hybridization.

LMPCR

For amplification of ChIP samples, we performed ligation-mediated PCR (LMPCR) using an entire ChIP and 30 ng of the corresponding input chromatin according to the protocol by Li et al. (2003).

Selection of Pluripotency Genes

A list of 14 pluripotency genes was collected from literature (for references, see the Supplemental Experimental Procedures). The list consists of genes that were shown to play a role in maintaining pluripotency in embryonic stem cell and/or are essential for integrity and developmental potential of the inner cell mass of mouse blastocysts.

Bisulfite Sequencing

1 μg genomic DNA was bisulfite converted with the EpiTec Bisulfite Kit (QIAGEN). Regions of interest were amplified by PCR and cloned by TOPO-TA cloning (Invitrogen). Primers for PCR amplification are listed in Table S2. Prior to sequencing, plasmids forming individual clones were amplified using the PlasmidAmp Kit (QIAGEN) according to the manufacturer's protocol.

Bioinformatics

Microarray design, GO-term analysis, GNF-Symatlas expression data matching, microarray hybridization and analysis, promoter annotation and filtering, and additional references are described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

Microarray raw data are deposited at GEO (www.ncbi.nlm.nih.gov/geo, accession number GSE11489) and processed, and normalized values can be accessed via the author's website (http://www.fmi.ch/groups/schubeler.d/web/data.html).

SUPPLEMENTAL DATA

The Supplemental Data include nine figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at http://www.molecule.org/cgi/content/full/30/6/755/DC1/.

ACKNOWLEDGMENTS

We thank Antoine Peters, Susan Gasser, Yves Barde, and members of the laboratory for helpful comments on the manuscript. F.M. is supported by a predoctoral fellowship of the Boehringer Ingelheim Fonds. T.R. is supported by the Swiss Cancer League. Furthermore, we thank Edward Oakeley for Affymetrix expression analysis and Maciej Pietrzak for DNA sequencing. Research in the laboratory of D.S is supported by the Novartis Research Foundation, by the European Union (NoE "The Epigenome" [LSHG-CT-2004-503433], and LSHG-CT-2006-037415), and the EMBO Young Investigator program.

Received: December 11, 2007
Revised: May 10, 2008
Accepted: May 16, 2008
Published online: May 29, 2008

REFERENCES

Agger, K., Cloos, P.A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature *449*, 731–734.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669–681.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bibel, M., Richter, J., Lacroix, E., and Barde, Y.A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nat. Protocols *2*, 1034–1043.

Bibel, M., Richter, J., Schrenk, K., Tucker, K.L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nat. Neurosci. *7*, 1003–1009.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760–1764.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536–2539.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20*, 1123–1136.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039–1043.

Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.L., Cattaneo, E., et al. (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. PLoS Biol. *3*, e283. 10.1371/journal.pbio.0030283.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185–196.

Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M., and Walker, P.R. (2005). Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. J. Biol. Chem. *280*, 6257–6260.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat. Cell Biol. *8*, 188–194.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG Methylation Contributes to ES Cell

Gene Regulation in Parallel with Oct4/Nanog, PcG Complex, and Histone H3 K4/K27 Trimethylation. Cell Stem Cell *2*, 160–169.

Futscher, B.W., Oshiro, M.M., Wozniak, R.J., Holtan, N., Hanigan, C.L., Duan, H., and Domann, F.E. (2002). Role for DNA methylation in the control of cell type specific maspin expression. Nat. Genet. *31*, 175–179.

Geppert, M., Goda, Y., Hammer, R.E., Li, C., Rosahl, T.W., Stevens, C.F., and Sudhof, T.C. (1994). Synaptotagmin I: a major Ca2+ sensor for transmitter release at a central synapse. Cell *79*, 717–727.

Gidekel, S., and Bergman, Y. (2002). A unique developmental pattern of Oct-3/4 DNA methylation is controlled by a cis-demodification element. J. Biol. Chem. *277*, 34521–34530.

Goldman, S. (2003). Glia as neural progenitor cells. Trends Neurosci. *26*, 590–596.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Gotz, M., and Barde, Y.A. (2005). Radial glial cells defined and major intermediates between embryonic stem cells and CNS neurons. Neuron *46*, 369–372.

Greer, P. (2002). Closing in on the biological functions of Fps/Fes and Fer. Nat. Rev. Mol. Cell Biol. *3*, 278–289.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

Heard, E., Clerc, P., and Avner, P. (1997). X-chromosome inactivation in mammals. Annu. Rev. Genet. *31*, 571–610.

Heins, N., Malatesta, P., Cecconi, F., Nakafuku, M., Tucker, K.L., Hack, M.A., Chapouton, P., Barde, Y.A., and Gotz, M. (2002). Glial cells generate neurons: the role of the transcription factor Pax6. Nat. Neurosci. *5*, 308–315.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Howlett, S.K., and Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. Development *113*, 119–127.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol. Cell. Biol. *24*, 8862–8871.

Kalantry, S., Manning, S., Haub, O., Tomihara-Newberger, C., Lee, H.G., Fangman, J., Disteche, C.M., Manova, K., and Lacy, E. (2001). The amnionless gene, essential for mouse gastrulation, encodes a visceral-endoderm-specific protein with an extracellular cysteine-rich domain. Nat. Genet. *27*, 412–416.

Kashiwabara, S., Noguchi, J., Zhuang, T., Ohmura, K., Honda, A., Sugiura, S., Miyamoto, K., Takahashi, S., Inoue, K., Ogura, A., et al. (2002). Regulation of spermatogenesis by testis-specific, cytoplasmic poly(A) polymerase TPAP. Science *298*, 1999–2002.

Kim, S., Namekawa, S.H., Niswander, L.M., Ward, J.O., Lee, J.T., Bardwell, V.J., and Zarkower, D. (2007). A mammal-specific Doublesex homolog associates with male sex chromatin and is required for male meiosis. PLoS Genet *3*, e62. 10.1371/journal.pgen.0030062.

Lan, F., Bayliss, P.E., Rinn, J.L., Whetstine, J.R., Wang, J.K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., et al. (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. Nature *449*, 689–694.

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.P., Reinberg, D., Croce, L.D., and Shiekhattar, R. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science *318*, 447–450.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362–365.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q., and Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc. Natl. Acad. Sci. USA *100*, 8164–8169.

Lin, I.G., and Hsieh, C.L. (2001). Chromosomal DNA demethylation specified by protein binding. EMBO Rep. *2*, 108–112.

Malatesta, P., Hack, M.A., Hartfuss, E., Kettenmann, H., Klinkert, W., Kirchhoff, F., and Gotz, M. (2003). Neuronal or glial progeny: regional differences in radial glia fate. Neuron *37*, 751–764.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501–502.

Meno, C., Shimono, A., Saijoh, Y., Yashiro, K., Mochida, K., Ohishi, S., Noji, S., Kondoh, H., and Hamada, H. (1998). lefty-1 is required for left-right determination as a regulator of lefty-2 and nodal. Cell *94*, 287–297.

Merkle, F.T., Mirzadeh, Z., and Alvarez-Buylla, A. (2007). Mosaic organization of neural stem cells in the adult brain. Science *317*, 381–384.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. Cell *111*, 197–208.

Nitta, T., Hata, M., Gotoh, S., Seo, Y., Sasaki, H., Hashimoto, N., Furuse, M., and Tsukita, S. (2003). Size-selective loosening of the blood-brain barrier in claudin-5-deficient mice. J. Cell Biol. *161*, 653–660.

Ohm, J., McGarvey, K., Yu, X., Cheng, L., Schuebel, K., Cope, L., Mohammad, H., Chen, W., Daniel, V., Yu, W., et al. (2007). A stem cell -like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat. Genet. *39*, 237–242.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. Cell Stem Cell *1*, 299–312.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol. Cell. Biol. *27*, 3769–3779.

Plachta, N., Bibel, M., Tucker, K.L., and Barde, Y.A. (2004). Developmental potential of defined neural progenitors derived from mouse embryonic stem cells. Development *131*, 5449–5456.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517–527.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425–432.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223–232.

Roh, T.Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. Proc. Natl. Acad. Sci. USA *103*, 15782–15787.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat. Genet. *39*, 232–236.

Schübeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J.,

et al. (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev. *18*, 1263–1271.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9–22.

Sharif, J., Muto, M., Takebayashi, S.I., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908–912.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A., and Issa, J.P. (2007). Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet *3*, 2023–2036. 10.1371/journal.pgen.0030181.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. Nat. Rev. Cancer *6*, 846–856.

Squazzo, S.L., O'Geen, H., Komashko, V.M., Krig, S.R., Jin, V.X., Jang, S.W., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2006). Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res. *16*, 890–900.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. USA *101*, 6062–6067.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M., et al. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871–874.

Walsh, C.P., and Bestor, T.H. (1999). Cytosine methylation and mammalian development. Genes Dev. *13*, 26–34.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. *37*, 853–862.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. *39*, 457–466.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., et al. (2007). Epigenetic stem cell signature in cancer. Nat. Genet. *39*, 157–158.

Wu, X., Viveiros, M.M., Eppig, J.J., Bai, Y., Fitzpatrick, S.L., and Matzuk, M.M. (2003). Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. Nat. Genet. *33*, 187–191.

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.-K., Shahab, A., Kuznetsov, V.A., et al. (2007). 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.

Molecular Cell, Volume *30*

Supplemental Data

Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors

Fabio Mohn, Michael Weber, Michael Rebhan, Tim C. Roloff, Jens Richter, Michael B. Stadler, Miriam Bibel, and Dirk Schübeler

Supplemental Experimental Procedures

Microarray design. MeDIP and ChIP samples were hybridized to promoter tiling microarrays (MM5_min_promoter_array, NimbleGen Systems Inc.) representing 26'275 putative promoters. On average every probe set spans 1.5kb (+200 to -1300 bp respective to the transcription start site, TSS) with 15 repeat-masked 50-mer oligonucleotides spaced 100 bp apart on average. For microarray hybridization and analysis see Supplemental Experimental Procedures. Final promoter log2 ratios of IP over input signal represent the average of at least two independent experiments, including one dye swap.

Microarray hybridization and analysis. Sample labeling, hybridization and array scanning were performed by NimbleGen Systems Inc. according to standard procedures. For analysis raw fluorescent intensity values were used to calculate $\log_2$ of the bound/input ratios for each individual oligo. DNA methylation arrays were further loess normalized to account for potential labeling dye artifacts in the low signal range. The $\log_2$ ratios of all oligos from one promoter which are located within a 900bp window around the TSS (+200 to -700) were averaged to determine a single $\log_2$ value per promoter. Subsequently, for comparison all arrays were normalized to a median $\log_2 = 0$ and scaled to have the same median absolute deviation using the LIMMA package in R (Smyth, 2004; Smyth and Speed, 2003).

For DNA methylation experiments, we considered weak and strong CpG island promoters with a $\log_2 > 0.4$ to be DNA hypermethylated. This cutoff was verified by PCR and

bisulfite sequencing (Supplementary Fig. 3, data not shown, and ref. (Weber et al., 2007)). To identify differentially methylated promoters between ES cells and neurons, we calculated p-values by applying an empirical Bayes method which takes into account the variances between individual replicate experiments (Smyth, 2004). *De novo* methylated promoters were identified using the following criteria: (i) a p-value <0.01, (ii) a $\log_2$ ratio in ES cells < 0.4, (iii) a delta methylation value ($\log_2$ ratio in neurons – $\log_2$ ratio in ES cells) > 0.3 and (iv) a $\log_2$ ratio in neurons above the running median of DNA methylation ($\log_2$ ratios) versus CpG content. According to these criteria we identified 343 promoters to be significantly *de novo* methylated (Supplementary Fig. 8). The same strategy was employed to find significantly demethylated promoters with a $\log_2$ ratio in ES cells > 0.4, -0.3 as minimal delta methylation value and a $\log_2$ ratio below the running median in neurons. Like this 22 promoters were identified to be significantly demethylated.

In ChIP-on-chip experiments, enriched promoters were defined by assuming a normally distributed background signal. We adapted an approach used previously for identifying enriched probes on oligonucleotide tiling arrays (Li et al., 2003). Briefly, in order to estimate such a background distribution, we fitted two Gaussians, corresponding to background and enriched signals, to promoter $\log_2$ ratios, using R and the Mclust package (Fraley, 2003, 2002). This method will find the most probable parameters of the two normal distributions based on the experimental data, by starting with random values and iterating until convergence. With the estimated background distribution, we can assign a P-value to each promoter $\log_2$ ratio reflecting significance of enrichment. The P-value cutoff used to define significant enrichment can be interpreted as the tolerated false positive rate (Figure S8). For RNA Pol II and H3K4 methylation, a P-value cutoff of 0.05 was used. For H3K27 methylation a more stringent P-value cutoff of 0.001 was applied and four replicates were performed in stem cells to account for the higher variability between replicates. Validation by an independent experimental approach (ChIP-realtime PCR) was performed for all ChIP-chip experiments (Figure S8 and data not shown).

All calculations were done in R using the packages “LIMMA” and “Mclust” (refs. (Fraley, 2003, 2002; Smyth, 2004; Smyth and Speed, 2003) and http://www.bioconductor.org, http://www.r-project.org).

Microarray data has been uploaded to the GEO database and can be accessed via the following link XXXX.

Affymetrix expression analysis. RNA for the 3 differentiation stages from two independent differentiation experiments was extracted using Trizol reagent (Invitrogen) according to the manufacturer’s protocol. Sample preparation and hybridization to Affymetrix MOE430 v2.0 GeneChip arrays was performed as described previously (Sinkkonen et al., 2008).

Promoter annotation and filtering. Promoter filtering was performed using a similar strategy as previously for the human genome (Weber et al., 2007), with modifications: We retrieved from the UCSC genome annotation database (http://genome.ucsc.edu) the RefSeq gene predictions, Ensembl gene predictions, mRNA counts and spliced ESTs counts starting within 150bp of the predicted transcription start site (TSS). The criteria for promoter filtering were adopted as follows: Promoters require either (i) a RefSeq gene and at least one mRNA, (ii) a RefSeq gene and at least 5 spliced ESTs, (iii) an Ensemble gene and at least 3 mRNAs, (iv) an Ensembl gene and at least 2 mRNAs and 5 spliced ESTs, or (v) at least 3 mRNAs and 5 spliced ESTs. All annotations refer to the mouse mm5 genome assembly, which was used for the design of the respective Nimblegen promoter microarray.

Promoter classification. Promoter classes were defined as previously described (Weber et al., 2007), with modifications to account for the different CpG dinucleotide distribution and GC content of the mouse genome (Bajic et al., 2006; Waterston et al., 2002). CpG-poor correspond to LCP, weak CpG islands to ICP and strong CpG islands to HCP. The CpG observed vs. expected ratios were calculated in 500bp windows (sliding 1 bp at every step) within the 1.5kb promoter sequence covered by the oligos (+200bp on each end to account

for resolution limited by DNA fragment size of about 500 bp). The 500bp window with the highest ratio (CpGo/e_500) was then used for promoter classification into 3 groups. Promoters with a CpGo/e_500 > 0.8 were classified as strong CpG island promoters (88% CpG islands according to previously defined criteria (Gardiner-Garden and Frommer, 1987)), promoters with CpGo/e_500 between 0.45-0.8 were islands were classified as weak CpG island promoters (67% CpG islands (Gardiner-Garden and Frommer, 1987)), and promoters with a CpGo/e_500 < 0.45 were termed CpG-poor promoters (6% CpG islands (Gardiner-Garden and Frommer, 1987)) containing promoters.

GO term analysis. Gene Ontology term enrichment analysis was performed using GO Stat, Fatigo and GO::TermFinder in parallel to rule out any bias due to the statistical methods the individual tools use (Al-Shahrour et al., 2006; Beissbarth and Speed, 2004; Boyle et al., 2004). As a background set we applied the full set of probes present on our promoter microarrays. The set of *de novo* methylated promoters was further curated manually to increase sensitivity by only considering the 100 best annotated genes according to Swissprot. Transcripts were mapped to Swissprot Protein IDs and a scoring scheme that emphasizes the length of the text in the CC function field, the number of keywords, and takes into account the level of characterization of other types of information present in Swissprot entries was used for classification (Wu et al., 2006).

Matching to GNF-Symatlas expression data. Expression data from the GNF-Symatlas database (Su et al., 2004) was matched via Ensembl gene IDs. Ubiquitously expressed genes (HK, i.e. “housekeeping” genes) were defined as genes with an expression value of >200 in at least 55 out of 61 tissues. Genes expressed in less than 5 tissues and with at least one expression value of >200 were classified as tissue-specific (TSp) genes. Like that a total number of 4804 promoters could be classified as either HK (2941) or TSp (1863) (see Supplementary Fig. 2). As expected HK are enriched in the strong CpG island promoter class (1.4x), slightly underrepresented in the weak CpG island class (0.8x) and depleted form

the CpG-poor promoters (0.3x) in regard to overall class frequencies. TSp show an opposite distribution as they are depleted in the strong CpG island promoter class (0.6x), close to expected in the weak CpG islands (1.1x) and overrepresented in the CpG-poor promoters (1.8x).

Selection of pluripotency genes. Selected pluripotency genes including references: *Tbn* (Voss et al., 2000), *Foxd3* (Hanna et al., 2002), *Sall4* (Zhang et al., 2006a), *Sox2* (Avilion et al., 2003), *Nanog* (Chambers et al., 2003), *Pou5f1* (also known as *Oct4*) (Nichols et al., 1998), *Tbx3*, *Esrrb*, *Tcl1*, *Dppa4*, (Ivanova et al., 2006), *Gdf3* (Levine and Brivanlou, 2006), *Tdgf1* (also know as *Cripto*) (Koestenbauer et al., 2006; Xu et al., 1999), *Zic3* (Lim et al., 2007), *Zfp42* (also know as *Rex1*) (Zhang et al., 2006b),

Supplemental References

Al-Shahrour, F., Minguez, P., Tarraga, J., Montaner, D., Alloza, E., Vaquerizas, J.M., Conde, L., Blaschke, C., Vera, J., and Dopazo, J. (2006). BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. Nucleic Acids Res *34*, W472-476.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.

Bajic, V.B., Tan, S.L., Christoffels, A., Schonbach, C., Lipovich, L., Yang, L., Hofmann, O., Kruger, A., Hide, W., Kai, C.*, et al.* (2006). Mice and men: their promoter properties. PLoS Genet *2*, e54.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Beissbarth, T., and Speed, T.P. (2004). GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics (Oxford, England) *20*, 1464-1465.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., and Sherlock, G. (2004). GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics (Oxford, England) *20*, 3710-3715.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Fraley, C., and Raftery, A. E. (2003). Enhanced Software for Model-Based Clustering, Density Estimation, and Discriminant Analysis: {MCLUST}. Journal of Classification *20*, 263-286.

Fraley, C., and Raftery, A. E. (2002). Model-Based Clustering, Discriminant Analysis and Density Estimation. Journal of the American Statistical Association *97*, 611-631.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J Mol Biol *196*, 261-282.

Hanna, L.A., Foreman, R.K., Tarasenko, I.A., Kessler, D.S., and Labosky, P.A. (2002). Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. Genes Dev *16*, 2650-2661.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin

signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature *442*, 533-538.

Koestenbauer, S., Zech, N.H., Juch, H., Vanderzwalmen, P., Schoonjans, L., and Dohr, G. (2006). Embryonic stem cells: similarities and differences between human and murine embryonic stem cells. Am J Reprod Immunol *55*, 169-180.

Levine, A.J., and Brivanlou, A.H. (2006). GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. Development *133*, 209-216.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q., and Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc Natl Acad Sci U S A *100*, 8164-8169.

Lim, L.S., Loh, Y.H., Zhang, W., Li, Y., Chen, X., Wang, Y., Bakre, M., Ng, H.H., and Stanton, L.W. (2007). Zic3 is required for maintenance of pluripotency in embryonic stem cells. Molecular biology of the cell *18*, 1348-1358.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P*., et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A *103*, 1412-1417.

Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert, C.J., Jr. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. Genome biology *6*, R33.

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C.G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nature structural & molecular biology *15*, 259-267.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology *3*, Article3.

Smyth, G.K., and Speed, T. (2003). Normalization of cDNA microarray data. Methods (San Diego, Calif *31*, 265-273.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G*., et al.* (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A *101*, 6062-6067.

Vinogradov, A.E. (2005). Dualism of gene GC content and CpG pattern in regard to expression in the human genome: magnitude versus breadth. Trends Genet *21*, 639-643.

Voss, A.K., Thomas, T., Petrou, P., Anastassiadis, K., Scholer, H., and Gruss, P. (2000). Taube nuss is a novel gene essential for the survival of pluripotent cells of early mouse embryos. Development *127*, 5449-5461.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R.*, et al.* (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res *34*, D187-191.

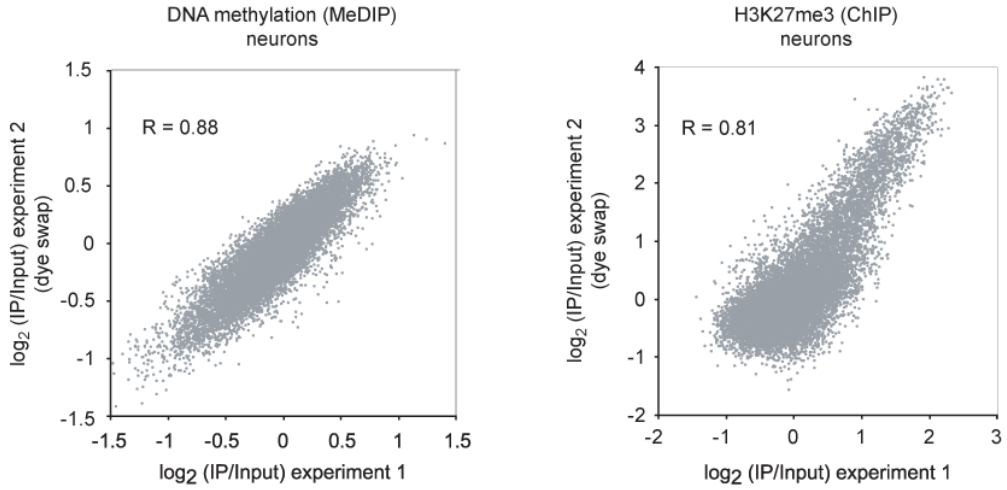
Xu, C., Liguori, G., Persico, M.G., and Adamson, E.D. (1999). Abrogation of the Cripto gene in mouse leads to failure of postgastrulation morphogenesis and lack of differentiation of cardiomyocytes. Development *126*, 483-494.

Zhang, J., Tam, W.L., Tong, G.Q., Wu, Q., Chan, H.Y., Soh, B.S., Lou, Y., Yang, J., Ma, Y., Chai, L.*, et al.* (2006a). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nat Cell Biol *8*, 1114-1123.

Zhang, J.Z., Gao, W., Yang, H.B., Zhang, B., Zhu, Z.Y., and Xue, Y.F. (2006b). Screening for genes essential for mouse embryonic stem cell self-renewal using a subtractive RNA interference library. Stem cells (Dayton, Ohio) *24*, 2661-2668.

Figure S1

A



B

R of replicates	ES	NP	TN
mRNA expression	0.99	0.99	0.98
5mC	0.84**	0.90**	0.88**
H3K27me3	0.63*	0.90	0.81
H3K4me2	0.98	0.97	0.90
RNA Pol II	0.74	0.93	0.83

Figure S1. Reproducibility of transcriptome (mRNA expression), MeDIP-chip (5mC) and ChIP-chip (H3K27me3, H3K4me2, Pol II) between biological replicates. (A) Representative examples of pair-wise comparisons of log2 (IP/Input) measurements for all promoters in two independent neuronal differentiations. (B) Table showing the range of Pearson correlation coefficients (R) between individual replicates performed on independent neuronal differentiations. We performed 4 independent replicates for H3K27me3 in ES cells (*) and 3 independent replicates for DNA methylation (5mC, **) at all three cellular states. Other experiments were done in independent duplicates.

Figure S2

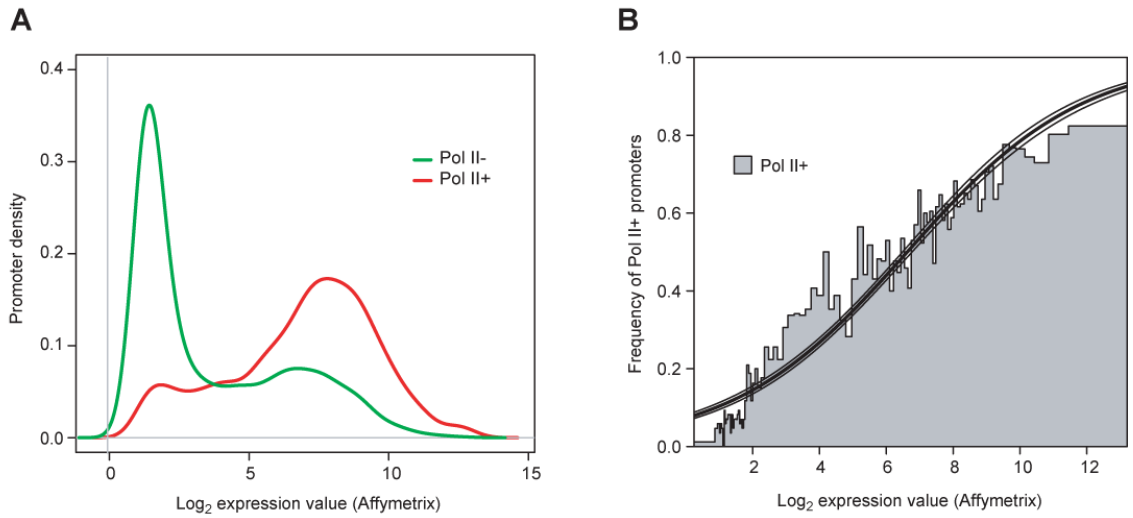


Figure S2. Correlation between polymerase occupancy at promoters and mRNA abundance. (A) Density plot comparing the distribution of mRNA levels measured with Affymetrix arrays (X-axis) for promoters scored as polymerase bound (Pol II+) and unbound (Pol II-). X-axis = relative expression value (log2). Note the low abundance of Pol II+ promoters with low mRNA levels, which validates the use of Pol II occupancy as a predictor of activity. Nonetheless, this small portion of promoters could be bound by stalled polymerase, ready for rapid induction. (B) Plotted is the mRNA expression value (X-axis) versus the probability of being scored Pol II+ (Y-axis). For plotting purposes only, genes were ranked according to their mRNA expression value and divided in 100 bins of equal gene numbers. Each column presents the percentage of Pol II+ genes in one bin (grey). The relationship between the probability of being Pol II+ at the promoter and mRNA levels was analyzed using logistic regression, which is a statistical method used if there are only two potential outcomes for one of the two variables. The resulting logistic regression curve (thick line) and 95% confidence intervals (outer lines) reveal a strong correlation between mRNA levels and percentage of Pol II+ promoters illustrating again the predictive power of Pol II for transcriptional activity at most genes.

Figure S3

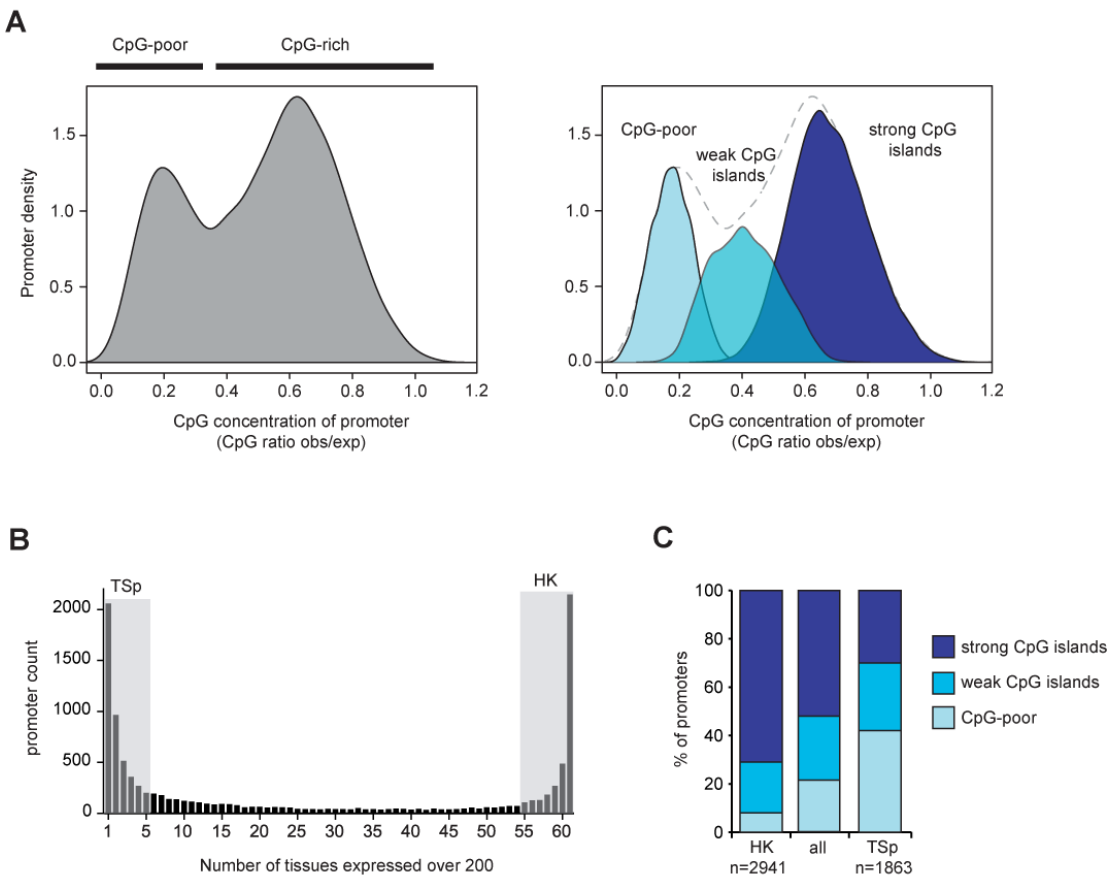


Figure S3. Promoter classification and tissue-specific expression of linked genes. (A). Density plots showing the CpG ratio observed/expected for all 15.100 high-confidence promoters. The total distribution (left) consisting of CpG-poor and CpG-rich promoters was split into 2 non-overlapping populations (right) of strong CpG islands (50% of total, dark blue) and CpG-poor promoters (23%, light blue). The remaining intermediate group was classified as weak CpG island promoters (27%, blue; see Experimental Procedures). (B) Histogram displaying the distribution of promoters according to the number of tissues where the linked genes show an expression value above 200 in the Symatlas database. Genes show either ubiquitous expression in most of the 61 probed tissues or very restricted expression in only few tissues (grey boxes). Promoters linked to genes expressed above 200 in 55 or more tissues were classified as ubiquitously expressed, i.e. “housekeeping” (HK), whereas genes expressed above 200 in at least one but less than 6 tissues were classified as tissue-specific (TSp). (C) The bar graph shows the distribution of HK and TSp among promoter classes. As expected HK are enriched in the strong CpG island promoter class (1.4x), slightly underrepresented in the weak CpG

island class (0.8x) and depleted form the CpG-poor promoters (0.3x) in regard to overall class frequencies. TSp show an opposite distribution as they are depleted in the strong CpG island promoter class (0.6x), close to expected in the weak CpG islands (1.1x) and overrepresented in the CpG-poor promoters (1.8x), which confirms previous reports (Saxonov et al., 2006; Schug et al., 2005; Vinogradov, 2005).

Figure S4

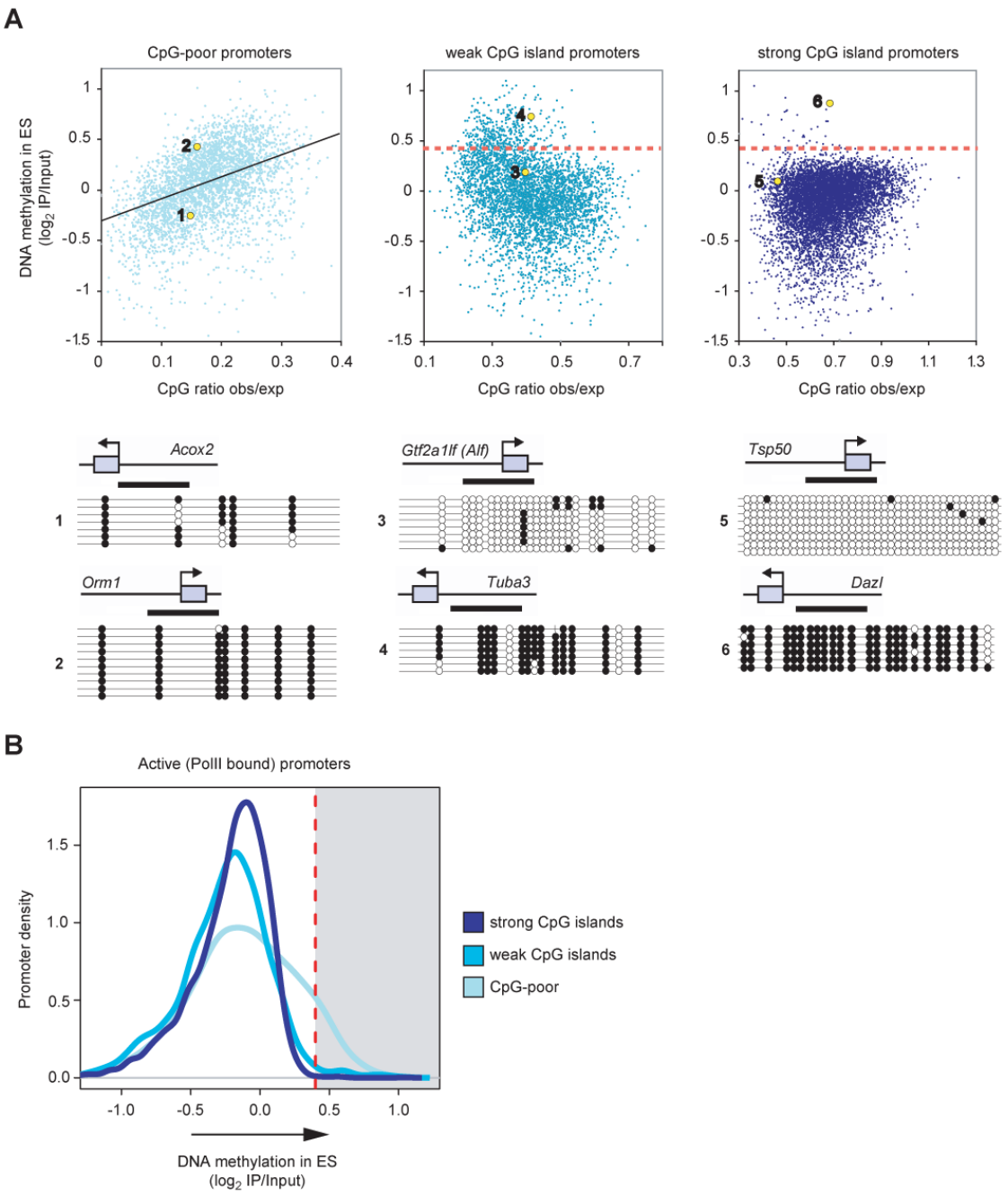


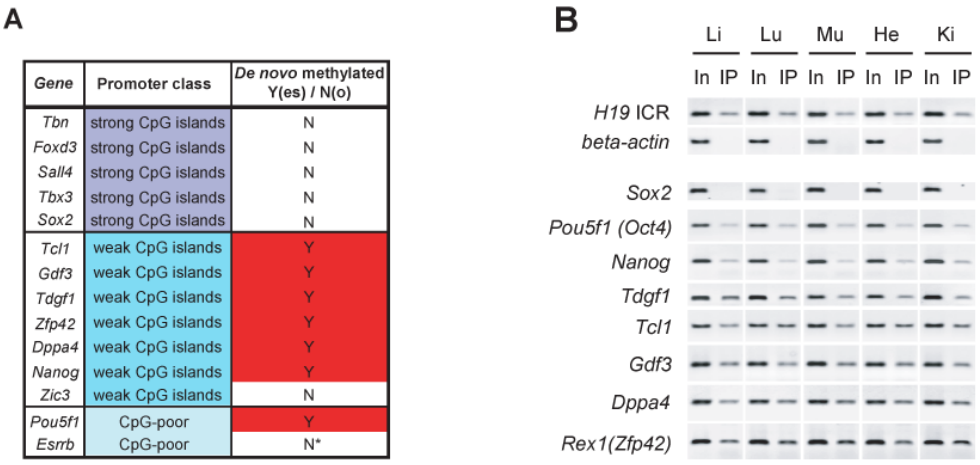
Figure S4. Promoter DNA methylation in ES cells. (A) Scatter-plot displaying DNA methylation values (log2 IP/Input) in relation to CpG content for all promoters in the three promoter classes. Note that for CpG-poor promoters DNA methylation increases linear with CpG content (black line indicates the linear regression curve, R = 0.44), indicating that the majority is DNA methylated by default (ref. 2). The red line indicates the cutoff for CpG island hypermethylation that is set at log2 (IP/Input) = 0.4.

Bisulfite sequencing of examples from the respective classes (yellow dots) are shown below the scatter plots. The black bar below the gene pictogram indicates the region analyzed. (B) Density plot for DNA methylation levels of active (PolII bound) promoters of the individual classes. The red line indicates the DNA methylation cutoff. Note that in the grey shaded area (DNA methylated promoters) almost no promoters of the strong and weak CpG island class are found whereas a significant amount of CpG-poor promoters are DNA methylated and expressed.

Figure S5

A

Gene	Promoter class	*De novo* methylated Y(es) / N(o)
Tbn	strong CpG islands	N
Foxd3	strong CpG islands	N
Sall4	strong CpG islands	N
Tbx3	strong CpG islands	N
Sox2	strong CpG islands	N
Tcl1	weak CpG islands	Y
Gdf3	weak CpG islands	Y
Tdgf1	weak CpG islands	Y
Zfp42	weak CpG islands	Y
Dppa4	weak CpG islands	Y
Nanog	weak CpG islands	Y
Zic3	weak CpG islands	N
Pou5f1	CpG-poor	Y
Esrrb	CpG-poor	N*



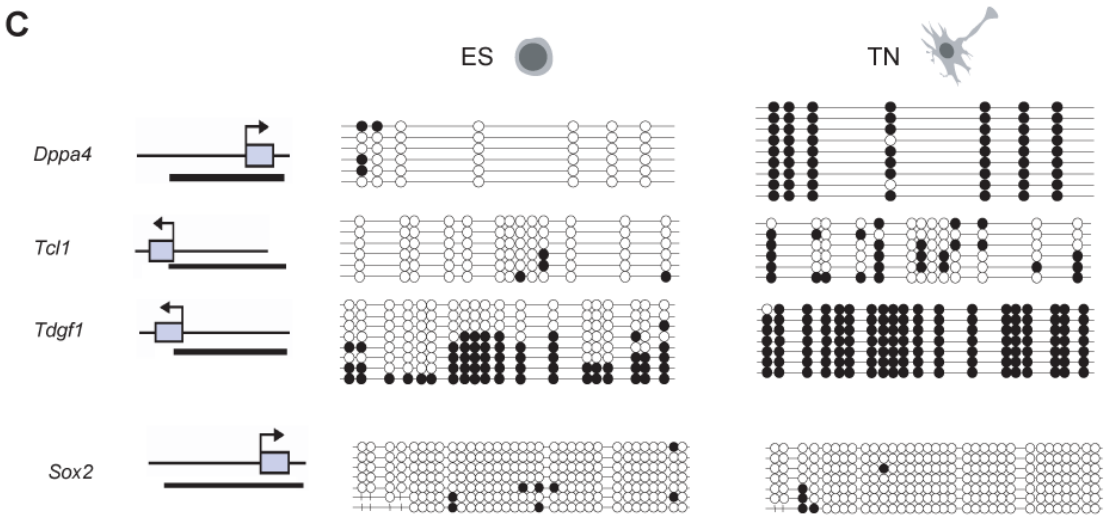


Figure S5. De novo methylation of pluripotency-associated genes. (A) Table indicating de novo methylation for all pluripotency genes. Note the uniform de novo methylation of weak CpG island promoters. (*) Esrrb is methylated at all stages. (B) Single-gene analysis of promoter DNA methylation of pluripotency-associated genes (see also Figure 2) in primary tissue samples (Li= liver, Lu= lung, Mu = muscle, He = heart, Ki = kidney). (C) Validation of de novo methylation for promoters of pluripotency genes by bisulfite sequencing. Open circles indicate unmethylated and full circles methylated CpGs, respectively. The black bar below the gene scheme shows the analyzed region respective to the

transcription start site (arrow). As a negative control Sox2 is included which does not undergo significant DNA methylation changes during differentiation (see also Figure 2).

Figure S6

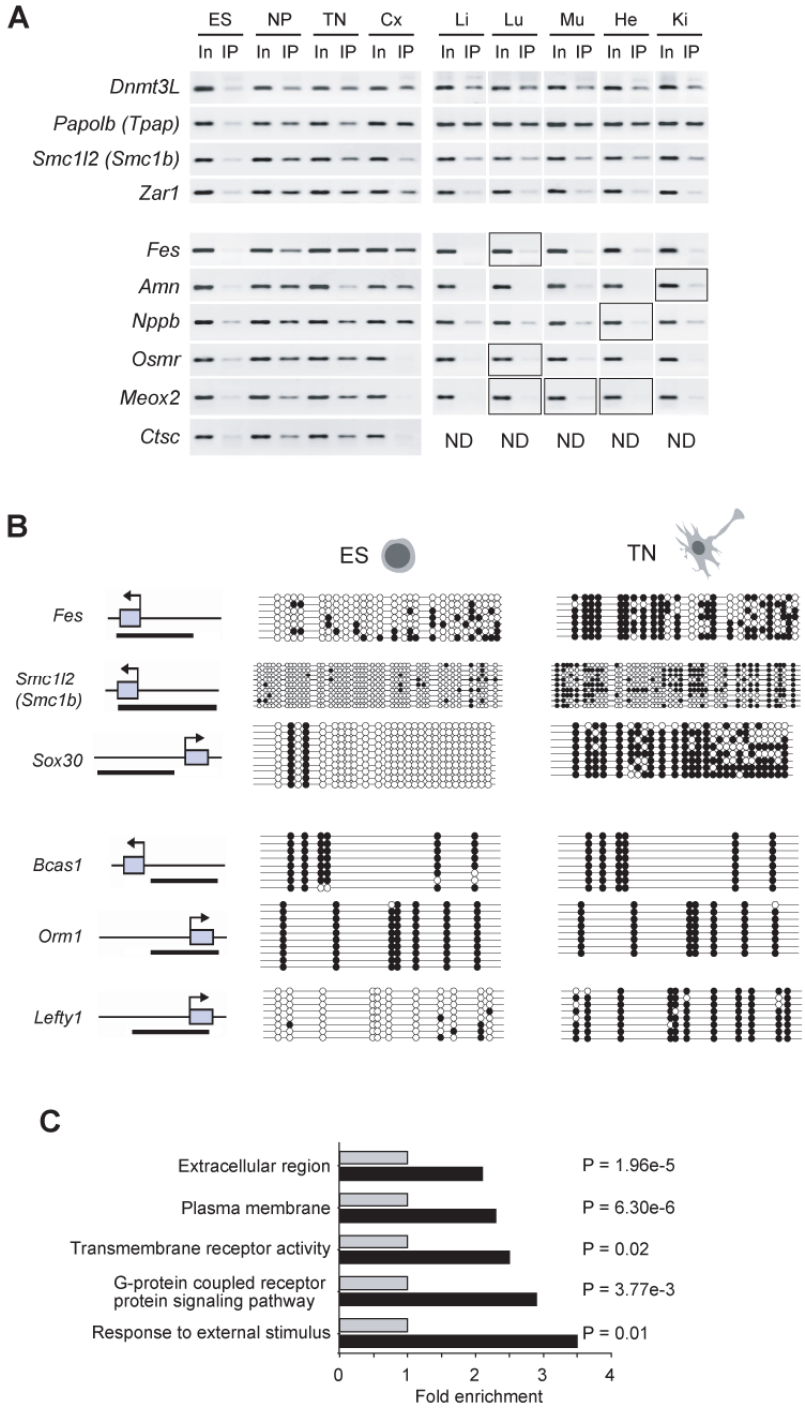


Figure S6. De novo methylation of germline-specific and other tissue-specifc genes. (A) Single-gene analysis of promoter DNA methylation of germline-specific genes (upper part) and somatic tissue-specific genes (lower part) in differentiating ES cells and primary tissue samples (Cx= cortex, Li= liver, Lu= lung, Mu = muscle, He = heart, Ki = kidney). Note that while most testis-specific genes

are methylated in all somatic tissues, other tissue-specific genes are mostly unmethylated in non-neuronal tissues. Black boxes mark those tissues in which the respective gene is expressed according to Symatlas. (B) Bisulfite sequencing validates de novo methlyation of some genes analysed in (A). Further, two examples of constitutively methylated CpG-poor promoters (Bcas1, Orm1; see also Figure S4) and one of the few CpG-poor promoters which gains DNA methylation (Lefty1) during differentiation are shown. (C) Gene onthology (GO) term analysis of de novo methylated promoters. The graph represents the enrichment of GO-terms among de novo methylated promoters (black bars) compared to all promoters (grey bars). Only significant GO-terms are shown ($P<0.05$).

Figure S7

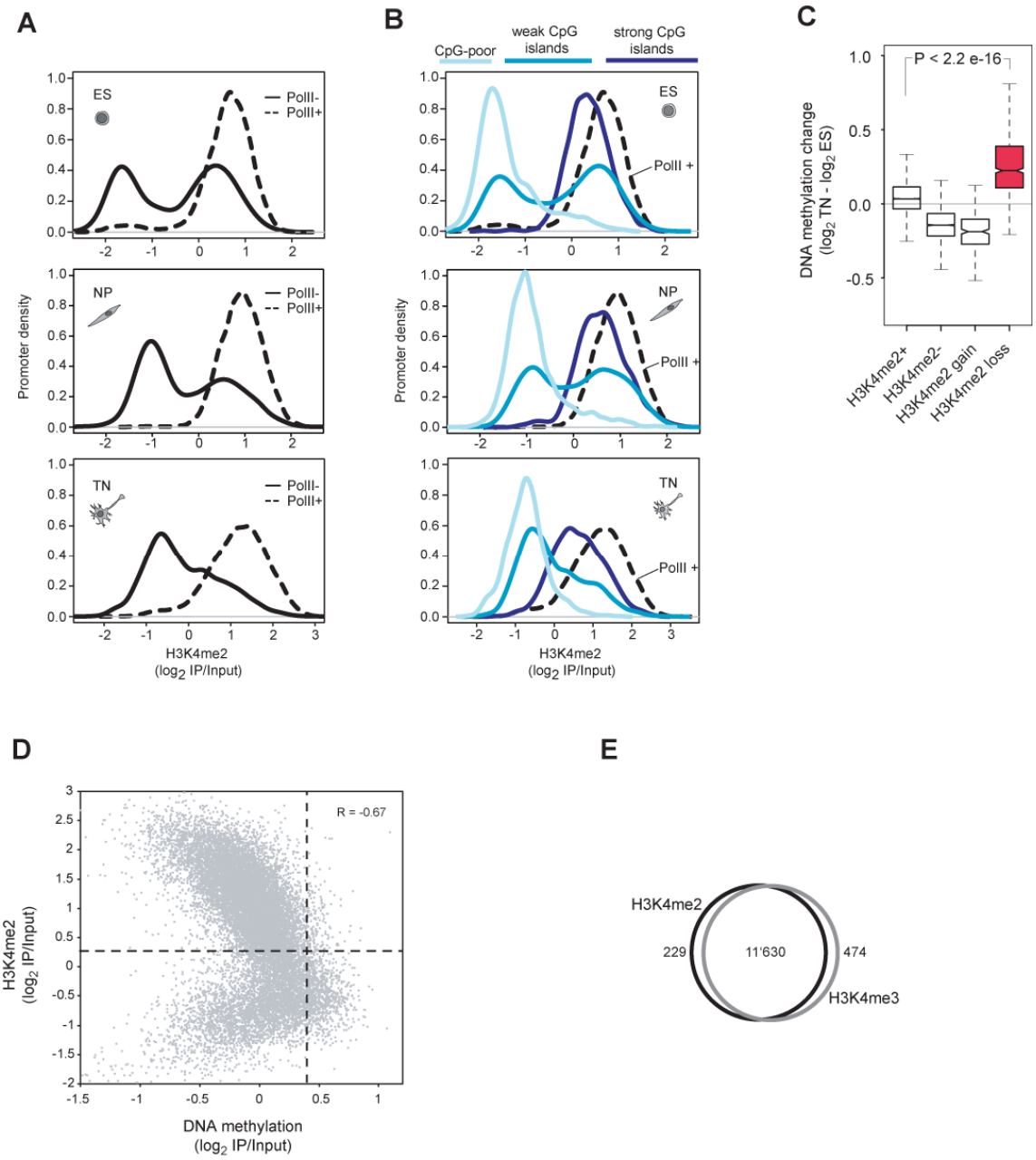


Figure S7. H3K4 methylation is present at inactive CpG-rich promoters if these are DNA unmethylated. (A) Density plot for H3K4 methylation for active (PolII+) and inactive (PolII-) promoters in stem cells (ES), neuronal progenitors (NP) and terminally differentiated neurons (TN). (B) Same plot as in (a) except inactive promoters are grouped and colored according to CpG class. (C) Boxplot showing the gain of DNA methylation at promoters which lose H3K4me2. (D) Scatter-plot comparing DNA methylation levels (X-axis) with H3K4 methylation (Y-axis) in neurons. The dotted lines illustrate

the cutoffs used for classifying DNA methylated and H3K4 methylated promoters. Note the strong negative correlation indicating that DNA methylation and H3K4me2 are mutually exclusive (R = -0.67). (E) Venn diagram showing the overlap between H3K4me2 and H3K4me3 positive promoters in stem cells indicating that H3K4 di-and trimethylation are highly correlated at the level of promoters (pairwise comparison: R = 0.85), which can also be deduced form previous global mapping data (Barski et al., 2007; Heintzman et al., 2007).

Figure S8

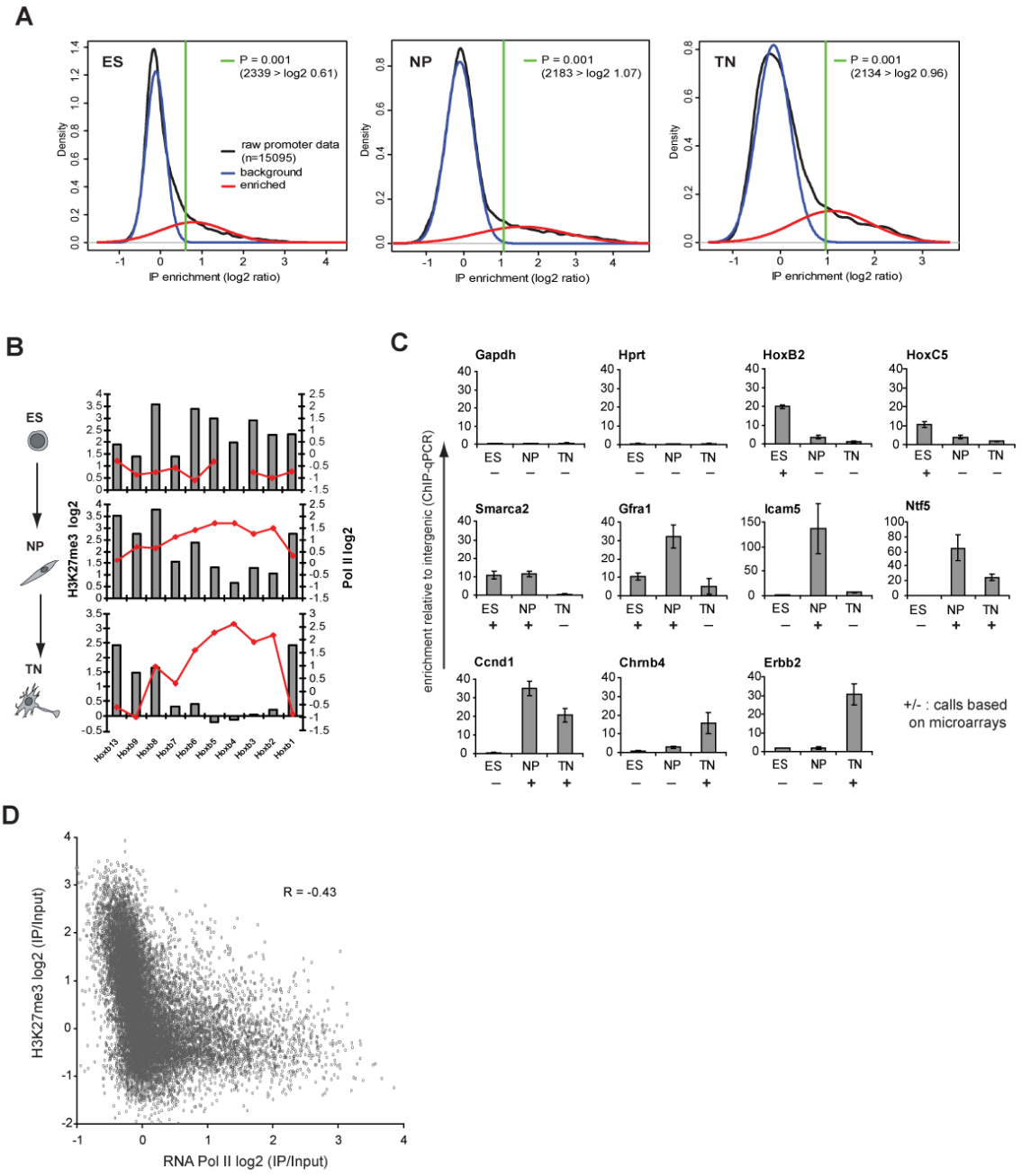


Figure S8. Illustration and validation of Polycomb targets and microarray cutoffs. (A) Illustration of cutoffs used for defining significantly enriched promoters in ChIP-chip experiments (see Supplemental Experimental Procedures). Briefly, two Gaussians were fitted into the distribution of log2 values (black) representing the background distribution (blue) and enriched promoters (red). P-values reflect significance of enrichment relative to the blue background curve. For H3K27me3 ChIP-chip a P-value cutoff of 0.001 was chosen based on validation by an independent method (see C). (B)

Examples of H3K27me3 levels at the HoxB cluster in ES, NP and TN (grey bars, left y-axis) ChIP-chip experiments. The red line indicates Pol II levels as a measure for gene activity (right y-axis). Shown are the log2 ratios (IP/Input) from the microarrays. (C) Validation by ChIP and quantitative real-time PCR (qPCR) of H3K27me3 positive and negative promoters at the three cellular stages (ES = stem cell, NP = neuronal progenitor, TN = terminal neuron) identified in ChIP-chip experiments. Plus and minus signs below the diagrams indicate the call for being H3K27me3 positive or negative based on the P-value cutoff. Error bars indicate SEM of at least two biological replicate experiments. (D) Scatter plot comparing H3K27me3 and RNA Pol II in ES cells. NP and TN are very similar (data not shown). Note the negative correlation and the low number of highly H3K27me3 and Pol II positive promoters (upper right quadrant), indicating that Pol II is in most cases excluded form H3K27me3 modified promoters.

Figure S9

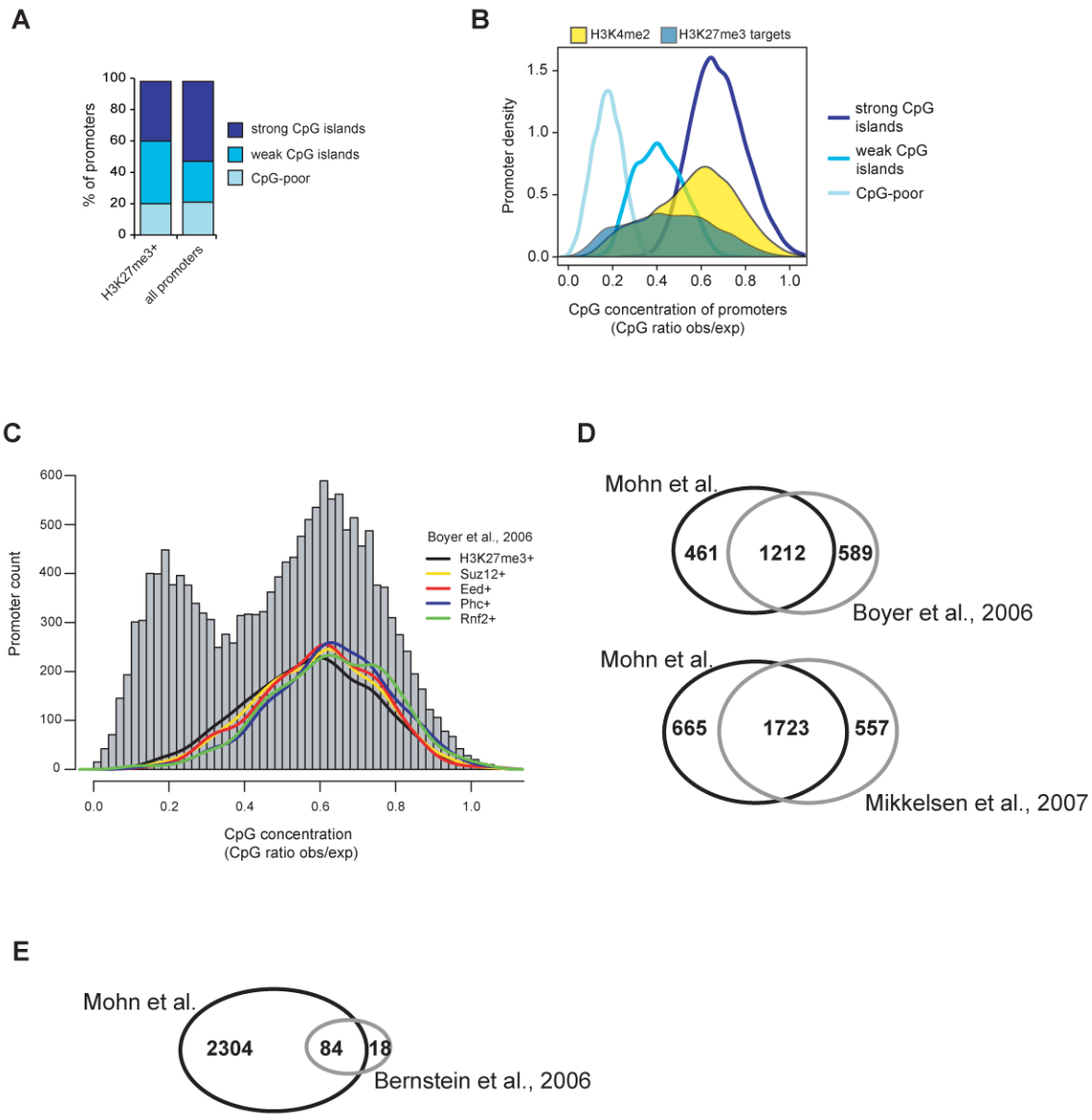


Figure S9. Targeting bias of Polycomb and comparison with published data on Polycomb targets in ES cells. (A) Distribution of promoter classes within the H3K27me3 targets (H3K27me3+) compared to all promoters. (B) Density plot for CpG ratio observed/expected for all promoter classes with superimposed density plots for targets H3K27 methylation (blue) and H3K4 methylation (yellow) at all three differentiation stages. (C) Histogram showing the distribution of all analyzed promoters according to the ratio of CpG observed / CpG expected. Superimposed are density plots for targets of H3K27me3 as well as Polycomb repressive complex 1 (Phc1, Rnf2) and 2 (Suz12, Eed) members identified by Boyer et al. (2006). Compare to Figure 5. Note the strong bias of H3K27me3 and

Polycomb targets towards CpG island promoters in our dataset aswell as in these published data. (D) Venn diagram illustrating the overlap of Polycomb targets in stem cells identified in our analysis and in previously published datasets by Boyer et al. (2006) and Mikkelsen et al. (2007). The comparisons include only sequences which were present in both studies. (E) Venn diagram showing the overlap of identified bivalent domains with data by Bernstein et al. (2006).

Table S1: Complete GO term lists of H3K27me3 targets that appear or dissolve between differentiation stages. Green P-values indicate over-representation in the analyzed gene list compared to the control set of all non-Polycomb targets present on the microarray, red P-values indicate under-representation. Black boxes specify clusters of GO terms that are derived from the same gene-list or from sublists which differ in less than 3 genes. Gene names of target genes which give rise to the GO term clusters are listed in the column "Genes in respective GO term cluster" for each cluster. "Targets" indicates number of genes in respective group of interest, "Total" indicates number of genes in the respective control set used for GO term analysis. P-values are corrected for the false positive discovery rate by Benjamini-Hochberg correction for multiple testing.

Table S2: Primer sequences.

PCR primers		
Gene	**Sense/Antisense**	**Sequence**
H19 ICR	s	GCATGGTCCTCAAATTCTGCA
	as	GCATCTGAACGCCCCAATTA
b-Actin	s	AGCCAACTTTACGCCTAGCGT
	as	TCTCAAGATGGACCTAATACGGC
Dppa4	s	GTTGAGGGTGGGACCAGAA
	as	TCCAGCAGTCTCCATCTTGA
Papolb (Tpap)	s	TGTCCTCTATCCCCAAGCAG
	as	GAGTATGGGCTGGCAGAAGA
Smc1l2 (Smc1b)	s	TGATGCAGGTGAACCTCTTG
	as	CGCGAGACTTCGAAAAGAAT
Zar1	s	AAAAAGCCCTACCACCTGCT
	as	TCCCCTACAAATCCTTGCTG
Amn	s	AGATGCTAAAGCCGCTGAGA
	as	GCTCTCACAGATGCGGACTT
Nppb	s	GAGCAGGAAGCAAGGACTGT
	as	GGGGTGGGGTTATCTCTGAT
Osmr	s	AATCAACTACGGGGCAAGTG
	as	ACCAGGAGCAAATTCCTGTG
Meox2	s	AGCAATGGGAACTGGAAATG
	as	AACTGCTTTCAGGGGGAAAT
Ctsc	s	ACCCTGTCCCTCCCTACATC
	as	GGAGCAGATGGCAGAGAAAG
Dnmt3L	s	ATCCACCCTCTGTCATTTGC
	as	TGGGGTAGGGGTAGGAGAAG
Gdf3	s	ATAAGGCTGCATGGTCGTCT
	as	TCTGAGGGGCTGAGAAGAGA
Rex1 (Zfp42)	s	TCTAGGCGGCTAGGAGTTCA
	as	ACCTTGAGCGCTTCTCATTG

Fes	s	GCCCGTGTTCCACATAAACT
	as	AAGCAGGAATCAGGAACTGG
Tdgf1	s	CAGCCAAGATCTCCGTGTAG
	as	CCACCTGAGGGTCCTACTCA
Pou5f1 (Oct4)	s	ACCTCCGTCTGGAAGACACA
	as	TCACCTAGGGACGGTTTCAC
Nanog	s	TGCAGGTGGGATTAACTGTG
	as	CAACGGCTCAAGGCGATAGA
Sox2	s	GCCAATATTCCGTAGCATGG
	as	GCTGGGGAACCTTTGTATCC
Tcl1	s	TCCTACACGGTGAGCATGAG
	as	TCCAGAAGTCCACGTTGTTG
Primers for bisulfite sequencing		
Sox30	s	AGGTGTTTTTATATTTGAGAATGATTAGAA
	as	ATTAAAACCCTTCCAAAACCTTAACTA
Tsp50	s	TAAAAATTGTTATTGAAGTTAAGTTTGG
	as	CTAAACCCTTTCTCTAAATCCCTATAC
Orm1	s	GTATAAAGTTGGTTTGAGGGAATATT
	as	AAAAATAACCTCCTACAATAAACAAATC
Bcas1	s	AGGTTTGTTGGTTGAATATATAGAGTAG
	as	ATAATAAAAATAACACCCAAAATCAAC
Acox2	s	GTTAGAGTATTGTATGAGGTTTTGTTGT
	as	TTTAACACTAATTCTTCACTAACACCAT
Tcl1	s	GTGTAGGTGTTTTTGTTTTGTGTGTT
	as	CAAACCCATATTTTCAAAAATAAATTTC
Tdgf1	s	AAAAGAAAGATATGTTAAATGAGAGAGGTT
	as	ATCCCTTAAAAACAAACAAAAAATCTC
Fes	s	TTAGGAATTAGATTGGAGGTTTTTAATAG
	as	AATCCATCCCCACTACATTTTAAAT
Sox2	s	GGGTTTTGTTTTATTTTGGTTTTAGTT
	as	AAACAACAAAATACTTTTCCCTTTTTAC
Lefty1	s	TTTATGTTGGGATTGAATTTAGGGT
	as	AAAAAATAACCATCCCTTCCACAT
Gdf3	s	TAGGGTAGGGTTAATAGAAGGAAGTTAA
	as	CCTTACATATCACAAACACACTAATATCTA
Dnmt3L	s	GTATTTTGGAGGGATTATTGGTTAT
	as	AAAAAAAACTAATTAAAAAAAACCTTACC
Smc1l2 (Smc1b)	s	GTTAATGATGTAGGTGAATTTTTTGAAA
	as	AAATAAAAATTTCCAAATCTAAACAACC
Dppa4	s	TTAGATTTATAGTTGTTAGGAGTAGGGG
	as	AAATAAAAAACCCTCATTTAAAAAACC
Dazl	s	GATTTTTGTTATTTTTTAGTTTTTTTAGGAT
	as	AAAATTCTCTCAACTAACCTAACTTATTTCT
Tuba3	s	ATTTTATTAATGATTGGATGTGGTTTAA
	as	AAATAAACAACTACTCACACAAACTTCC
Gtf2a1lf (Alf)	s	AGTGAGGTATGGTAAAAATAGGAATAATATT
	as	TACCCTAAAACCTAAATAACCTCAATTAA

3.2. DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity

Dong K.B., Maksakova I.A., Mohn F., Leung D., Appanah R., Lee S., Yang H.W., Lam L.L., Mager D.L., Schübeler D., Tachibana M., Shinkai Y., Lorincz M. C.

Summary

DNA methylation has been shown to be essential for silencing of endogenous retroviruses (ERV) such as intra cisternal A particles (IAP), which are severely upregulated in DNMT1 deficient mouse embryos (Walsh et al., 1998). Recent data indicated that these repetitive elements are further embedded in chromatin decorated with H3K9me2 and/or H3K9me3 (Martens et al., 2005; Mikkelsen et al., 2007; Peters et al., 2003). However, the exact interplay of H3K9me2/3 and DNA methylation for repressing these repetitive elements are not known. It was shown that the Suv39h1/2 enzymes, which are essential for pericentric heterochromatin maintenance, play no major role in H3K9 methylation and DNA methylation at ERVs (Lehnertz et al., 2003; Peters et al., 2003). Thus we set out to investigate whether G9a and its close homolog GLP/EuHMTas1 could be the enzymes involved in mediating H3K9me2 and DNA methylation at ERVs. Both proteins form a heteromeric complex localizing predominantly to euchromatin and mediate euchromatic H3K9 methylation (Tachibana et al., 2005). We find that L1 (LINE 1 retrotransposons), IAP and MLV (murine leukemia virus) elements show a strong decrease in DNA methylation in G9a$^{-/-}$ and also GLP $^{-/-}$ ES cells, but no transcriptional reactivation can be detected. This indicates that G9a and GLP are required for efficient DNA methylation at ERVs, but not for their transcriptional silencing. Notably, the loss of DNA methylation is slightly less severe than in DNMT1$^{-/-}$ ES cells, which show reactivation of IAPs. It might therefore be that DNA methylation levels are not below a critical threshold in order to allow reactivation of the repetitive elements. Alternatively, because H3K9me2 levels are only modestly reduced in G9a$^{-/-}$ ES cells while H3K9me3 and HP1a binding to the ERVs remains unchanged, a DNA methylation and G9a independent repressive pathway could backup the loss of DNA methylation. Setdb1, the only H3K9 KMT not yet tested for a function in ERV silencing, would be an obvious candidate.

Upon reintroduction of a catalytically inactive G9a transgene into G9a$^{-/-}$ ES cells, the DNA methylation defect is partially rescued. Moreover, not only DNA methylation at ERVs is affected in the G9a knock-out but also promoter regions of normally densely methylated genes show a reduction of DNA methylation which is partially rescued by a mutant G9a transgene similar to ERVs. Together our analysis suggest that DNA methylation, in contrast to plants and fungi (Freitag et al., 2004a; Jackson et al., 2002), does not depend on H3K9 methylation and is more likely to be directly recruited via G9a itself. This model is in agreement with data from Feldman and coworkers indicating that G9a can trigger DNA

methylation at the Oct4 promoter during cellular differentiation (Feldman et al., 2006). Given the effect observed on single copy genes, this pathway appears to affect global DNA methylation patterns which is supported by a report suggesting that direct interaction of G9a and DNMT1 might have a role in propagation of H3K9 methylation and DNA methylation during replication (Esteve et al., 2006).

The EMBO Journal (2008) 27, 2691–2701 | © 2008 European Molecular Biology Organization | All Rights Reserved 0261-4189/08
www.embojournal.org

THE EMBO JOURNAL

DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity

Kevin B Dong[1], Irina A Maksakova[1,2], Fabio Mohn[3], Danny Leung[1], Ruth Appanah[1], Sandra Lee[1], Hao W Yang[1], Lucia L Lam[1], Dixie L Mager[1,2], Dirk Schübeler[3], Makoto Tachibana[4,5], Yoichi Shinkai[4,5] and Matthew C Lorincz[1,*]

[1]Department of Medical Genetics, Life Sciences Institute, The University of British Columbia, Vancouver, British Columbia, Canada, [2]Terry Fox Laboratory, BC Cancer Agency, Vancouver, British Columbia, Canada, [3]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland, [4]Experimental Research Center for Infectious Diseases, Institute for Virus Research, Kyoto University, Kyoto, Japan and [5]Graduate School of Biostudies, Kyoto University, Sakyo-ku, Kyoto, Japan

Histone H3K9 methylation is required for DNA methylation and silencing of repetitive elements in plants and filamentous fungi. In mammalian cells however, deletion of the H3K9 histone methyltransferases (HMTases) Suv39h1 and Suv39h2 does not affect DNA methylation of the endogenous retrovirus murine leukaemia virus, indicating that H3K9 methylation is dispensable for DNA methylation of retrotransposons, or that a different HMTase is involved. We demonstrate that embryonic stem (ES) cells lacking the H3K9 HMTase G9a show a significant reduction in DNA methylation of retrotransposons, major satellite repeats and densely methylated CpG-rich promoters. Surprisingly, demethylated retrotransposons remain transcriptionally silent in *G9a*$^{-/-}$ cells, and show only a modest decrease in H3K9me2 and no decrease in H3K9me3 or HP1α binding, indicating that H3K9 methylation *per se* is not the relevant trigger for DNA methylation. Indeed, introduction of catalytically inactive G9a transgenes partially 'rescues' the DNA methylation defect observed in *G9a*$^{-/-}$ cells. Taken together, these observations reveal that H3K9me3 and HP1α recruitment to retrotransposons occurs independent of DNA methylation in ES cells and that G9a promotes DNA methylation independent of its HMTase activity.

The EMBO Journal (2008) **27,** 2691–2701. doi:10.1038/emboj.2008.193; Published online 25 September 2008
Subject Categories: chromatin & transcription
Keywords: chromatin; DNA methylation; ERV; H3K9 methylation; HP1α

*Corresponding author. Department of Medical Genetics, Life Sciences Institute Room 5-507, The University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia, Canada V6T 1Z3.
Tel.: +604 827 3965; Fax: +604 822 5348;
E-mail: mlorincz@interchange.ubc.ca

Received: 2 April 2008; accepted: 21 August 2008; published online: 25 September 2008

Introduction

Retrotransposons, including long terminal repeat (LTR) and non-LTR elements, are widely dispersed in the euchromatic compartment in higher mammals (Kuff and Lueders, 1988; Medstrand *et al*, 2002), constituting ~37% of the mouse genome (Mouse Genome Sequencing Consortium, 2002). A subset of these elements are transcriptionally competent, placing a significant mutational load on their hosts (Maksakova *et al*, 2006). To minimize the likelihood of retrotransposition, a number of pathways that function at the transcriptional or post-transcriptional stages of the replicative cycle have evolved to inhibit the expression of these parasitic elements. DNA methylation for example, has an important function in transcriptional silencing of retrotransposons in mammalian cells (Li *et al*, 1992; Yoder *et al*, 1997; Walsh *et al*, 1998), as illustrated by the high level of expression of the intracisternal A particle (IAP) endogenous retrovirus (ERV) in mouse embryos deficient in the DNA methyltransferase (DNMT), Dnmt1 (Walsh *et al*, 1998). DNA methylation also has a critical function in transcriptional silencing of repetitive elements and their relics in filamentous fungi and plants (Goyon *et al*, 1996; Lindroth *et al*, 2001; Zhou *et al*, 2001), substantiating the importance of this epigenetic mark in suppressing transposable elements in distantly related eukaryotes.

Repetitive elements in eukaryotes are also marked by specific covalent histone modifications (Bernstein *et al*, 2007). Methylation of lysine 9 of the histone H3 tail (H3K9) in particular, has an important function in silencing of these elements in yeast (Nakayama *et al*, 2001), filamentous fungi (Tamaru and Selker, 2001), plants (Jackson *et al*, 2002) and animals (Martens *et al*, 2005). Recent genome-wide studies reveal that ERVs are marked by H3K9 dimethylation (H3K9me2) and/or H3K9 trimethylation (H3K9me3) in murine cells (Peters *et al*, 2003; Martens *et al*, 2005; Mikkelsen *et al*, 2007); however, the specific histone methyltransferases (HMTases) responsible have not been identified.

Intriguingly, the H3K9 HMTase *DIM-5* is required for CpG methylation in *Neurospora* (Tamaru and Selker, 2001) and the H3K9 HMTase *KRYPTONITE* is required for CpNpG methylation in *Arabidopsis* (Jackson *et al*, 2002), suggesting the existence of an evolutionarily conserved silencing pathway in which H3K9 methylation promotes *de novo* DNA methylation of repetitive elements (Freitag and Selker, 2005; Stancheva, 2005). However, the role, if any, that H3K9 methylation has in DNA methylation of retrotransposons in mammalian cells has not been systematically addressed.

Five HMTases in the 'Suv39' subfamily of SET (Suv39, Enhancer of Zeste, Trithorax) domain-containing proteins with H3K9 catalytic activity, including Suv39h1 and the closely related Suv39h2, G9a and the closely related GLP/EuHMTase1 and SETDB1/Eset, have been characterized in mammalian cells. On the basis of its sequence similarity to SETDB1, the sixth Suv39 family member, SETDB2/CLLD8, is

also likely to have specificity for H3K9 (Mabuchi *et al*, 2001; Kouzarides, 2007). Although Suv39h1 and Suv39h2 double-negative (*Suv39h1/2*$^{-/-}$) embryonic stem (ES) cells show a dramatic reduction in H3K9me3 and DNA methylation at major satellite repeats, IAP elements show no reduction in H3K9 methylation (Peters *et al*, 2003; Martens *et al*, 2005; Mikkelsen *et al*, 2007) and murine leukaemia virus (MLV) ERVs show no reduction in DNA methylation (Lehnertz *et al*, 2003) in these cells. Taken together, these results indicate that Suv39h1 and Suv39h2 do not have a major function in H3K9 methylation or DNA methylation of LTR retrotransposons in mammalian cells.

In contrast to the Suv39h HMTases, G9a and GLP/Eu-HMTase1, which form a heteromeric complex *in vivo*, are widely dispersed in the euchromatic compartment and deletion of either leads to a dramatic decrease in H3K9me1 and H3K9me2 in ES cells (Tachibana *et al*, 2002, 2005). A recent analysis revealed that ~300–400 genes show altered expression in *G9a*$^{-/-}$ cells (Sampath *et al*, 2007) and several studies have shown that G9a regulates the expression and/or DNA methylation status of specific genes (Feldman *et al*, 2006; Ikegami *et al*, 2007). However, experiments aimed at determining whether G9a influences the expression and/or DNA methylation states of interspersed repetitive elements have not been reported.

Here, we investigated the function that G9a and GLP have in DNA methylation and silencing of potentially active ERVs and non-LTR retrotransposons. We show that DNA methylation of these elements, and a subset of non-repetitive sequences including CpG-rich promoters, is reduced in *G9a*$^{-/-}$ cells and that Dnmt3a recruitment to retrotransposons is decreased in these cells. However, H3K9me3 enrichment and HP1α binding are unaltered, demonstrating that an alternative H3K9 HMTase marks ERVs and that H3K9 methylation *per se* is not sufficient to promote DNA methylation of these elements. In support of this model, we show that the introduction of two different G9a transgenes that lack catalytic activity into *G9a*$^{-/-}$ cells partially 'rescues' the observed DNA methylation defect, indicating that G9a promotes DNA methylation of retrotransposons independent of its catalytic activity.

Results

G9a is required for DNA methylation of retrotransposons

To establish whether G9a is required for DNA methylation of ERVs, genomic DNA isolated from TT2 wild-type (wt) and *G9a*$^{-/-}$ ES cells (Tachibana *et al*, 2002) (Supplementary Figure S1) was analysed by Southern blotting using the methylation-sensitive restriction enzyme *Hpa*II and probes specific for IAP and MLV ERVs, of which there are ~1200 and ~60 copies in the mouse genome, respectively (Figure 1A and B). Genomic DNA samples isolated from *Dnmt1*$^{-/-}$ (Lei *et al*, 1996), *Suv39h1/2*$^{-/-}$ (Peters *et al*, 2001) and the wt parent ES cell lines from which they were derived were analysed in parallel. A dramatic reduction in DNA methylation of both ERVs was detected in the *G9a*$^{-/-}$ line relative to the wt control. At the resolution of Southern blot analysis, this reduction in methylation is not distinguishable from that observed for the *Dnmt1*$^{-/-}$ ES line, or TT2 genomic DNA digested with the methylation-insensitive isoschizomer *Msp*I.

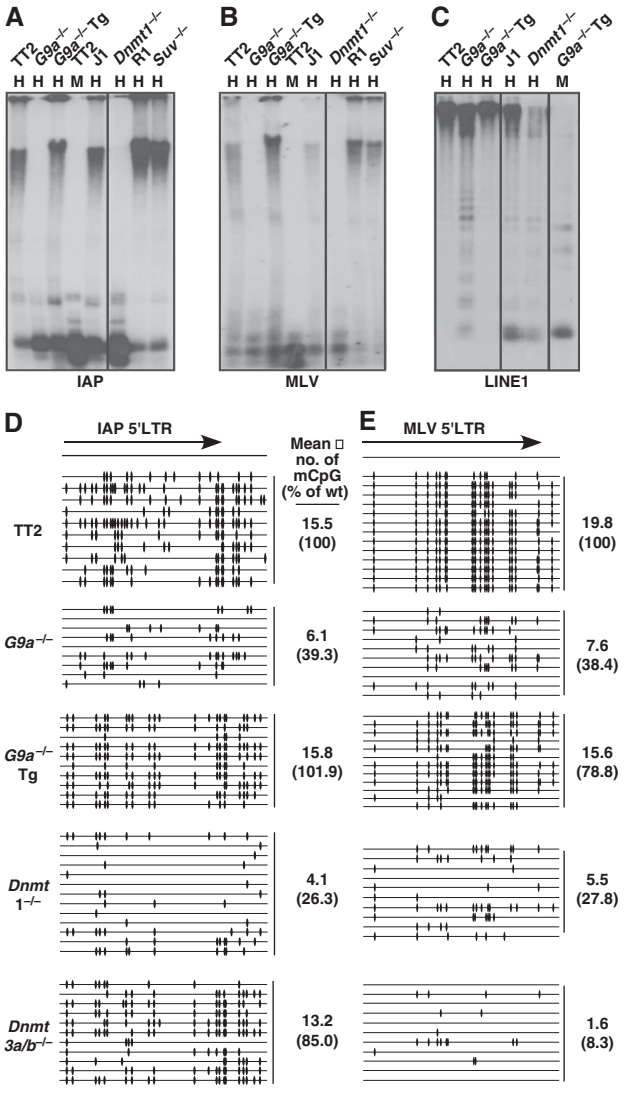


Figure 1 DNA methylation of MLV, IAP and LINE1 retrotransposons is reduced in *G9a*$^{-/-}$ cells. Genomic DNA isolated from *G9a*$^{-/-}$, *G9a*$^{-/-}$Tg, *Dnmt1*$^{-/-}$, *Suv39h1/2*$^{-/-}$ (*Suv*$^{-/-}$) and the wt parent lines TT2, J1 and R1, respectively, was digested with *Msp*I (M) or the methylation-sensitive restriction enzyme *Hpa*II (H) and subject to Southern blotting using probes specific for (**A**) IAP, (**B**) MLV or (**C**) LINE1 retrotransposons. The *G9a*$^{-/-}$ line shows a dramatic decrease in DNA methylation at each of these repetitive elements that is reversed in the *G9a*$^{-/-}$Tg line. In contrast, the *Suv39h1/2*$^{-/-}$ line shows no DNA methylation defect at IAP or MLV repeats. (**D**, **E**) Bisulphite analysis of the 5′LTR regions of IAP and MLV elements was conducted on TT2, *G9a*$^{-/-}$, *G9a*$^{-/-}$Tg (15-3), *Dnmt1*$^{-/-}$ and *Dnmt3a/b*$^{-/-}$ cells. For each molecule sequenced (horizontal bar), filled ovals represent the presence of an mCpG. The mean number of mCpGs per molecule sequenced is shown to the right of each set of sequenced samples. The mean % of mCpGs relative to the wild-type line is also shown (in parentheses).

To obtain a more accurate measure of the methylation status of these elements, high-resolution bisulphite sequencing analysis was conducted using primers specific for the CpG-rich 5′LTR regions of IAP and MLV elements. A >2.5-fold decrease in the mean number of mCpGs per molecule sequenced was detected in the *G9a*$^{-/-}$ line relative to the wt parent line (Figure 1D and E), with a subset of sequenced molecules in the *G9a*$^{-/-}$ line showing almost complete loss of methylation in these regions. A decrease in DNA methylation

across the LTR and downstream regions of the potentially active class II ERV MusD (Mager and Freeman, 2000), of which there are ~90 full-length copies in the mouse genome, was also detected in the *G9a*$^{-/-}$ line (Supplementary Figure S2). All three of these LTR retrotransposons show an even more severe DNA methylation defect in *Dnmt1*$^{-/-}$ cells (Figure 1D and E; Supplementary Figure S2). Consistent with the observations of Chen *et al* (2003), early passage *Dnmt3a/b*$^{-/-}$ cells (Okano *et al*, 1999) show a significantly more severe DNA methylation defect for MLV elements than IAP elements. Interestingly, whereas IAP elements show a less severe defect in *Dnmt3a/b*$^{-/-}$ cells than in *G9a*$^{-/-}$ cells, the reverse is true of MLV repeats. Thus, although it is clear that G9a is required for DNA methylation of distantly related ERVs in murine ES cells, the degree of demethylation is distinct from that observed for *Dnmt1*$^{-/-}$ or *Dnmt3a/b*$^{-/-}$ cells. In contrast, no reduction in DNA methylation of MLV or IAP elements was detected in *Suv39h1/2*$^{-/-}$ cells (Figure 1A and B), consistent with a previous report showing that Suv39h1 and Suv39h2 are not required for DNA methylation of MLV (Lehnertz *et al*, 2003).

DNA methylation of LINE1 (L1) elements, non-LTR retrotransposons that comprise ~20% of the mouse genome (Mouse Genome Sequencing Consortium, 2002), also depends on the presence of both Dnmt1 and Dnmt3a and/or Dnmt3b in ES cells (Liang *et al*, 2002). To determine whether G9a has a function in DNA methylation of this class of interspersed repeats, Southern blot analysis was conducted with a probe that spans the promoter region of the L1Md-A2 subfamily of L1 elements. A significant decrease in DNA methylation of L1 elements is also apparent in *G9a*$^{-/-}$ cells, although this defect is not as severe as that detected in the DNMT mutant lines (Figure 1C). Taken together, these observations indicate that G9a influences DNA methylation of both LTR and non-LTR retrotransposons in ES cells.

To determine whether G9a is also required for DNA methylation of tandem repeats in ES cells, Southern blot analysis using a probe specific for major satellite repeats (present at approximately 700 000 copies per cell) was conducted using the methylation-sensitive restriction enzyme HpyCH4IV (Supplementary Figure S3). Consistent with a previous report showing that Suv39h1 and Suv39h2 are required for methylation of major satellite repeats (Lehnertz *et al*, 2003), a dramatic reduction in DNA methylation of major satellite repeats was detected in *Suv39h1/2*$^{-/-}$ cells. Unexpectedly, a dramatic reduction in DNA methylation of this class of repeats was also detected in the *G9a*$^{-/-}$ line, revealing that G9a is required for DNA methylation of pericentromeric heterochromatin as well.

Introduction of a G9a transgene rescues the DNA methylation defect observed in G9a$^{-/-}$ *cells*

Reintroduction of Dnmt3a, Dnmt3a2 (the predominant isoform of Dnmt3a in ES cells; Chen *et al* (2002)) or Dnmt3b1 into *Dnmt3a/b*$^{-/-}$ ES cells restores DNA methylation of MLV and IAP elements (Chen *et al*, 2003), indicating that the *de novo* DNMTs are capable of reestablishing DNA methylation patterns in these cells. To determine whether reintroduction of G9a is capable of reversing the DNA methylation defect observed in *G9a*$^{-/-}$ cells, the methylation status of these elements was also analysed in a *G9a*$^{-/-}$ line stably expressing a wt G9a transgene (*G9a*$^{-/-}$*Tg*) (Tachibana *et al*, 2002) at a level similar to that of the endogenous protein (Supplementary Figure S1). Strikingly, the DNA methylation state of MLV, IAP, L1 (Figure 1) and MusD (Supplementary Figure S2) retrotransposons and major satellite repeats (Supplementary Figure S3) in the *G9a*$^{-/-}$*Tg* resembles that of the original wt parent line (TT2) rather than the *G9a*$^{-/-}$ line from which they were directly derived. These observations indicate that loss of DNA methylation in *G9a*$^{-/-}$ ES cells is not an irreversible process and that reintroduction of G9a is sufficient for the reestablishment of DNA methylation in G9a-deficient ES cells.

GLP is required for DNA methylation of retrotransposons

As G9a forms a complex with the closely related HMTase GLP, and both are required for the deposition of the H3K9me2 mark (Tachibana *et al*, 2005), we next determined whether GLP is also required for DNA methylation of retrotransposons. Genomic DNA isolated from wt TT2 and *GLP*$^{-/-}$ ES cells (Tachibana *et al*, 2005) was analysed by Southern blotting as above, using probes specific for IAP, MLV (Supplementary Figure S4) and L1 elements (data not shown). A significant DNA methylation defect is apparent for all three elements in the *GLP*$^{-/-}$ line as well, with IAP elements showing the most dramatic decrease. Furthermore, introduction of a wt GLP transgene into the *GLP*$^{-/-}$ line (generating the *GLP*$^{-/-}$*Tg* line (see Supplementary Figure S1; Tachibana *et al*, 2005) rescues the IAP DNA methylation defect and partially rescues the MLV methylation defect, revealing that DNA methylation can be reestablished on reintroduction of this HMTase as well. Consistent with these results, bisulphite sequencing analysis of polytrophic MLV elements reveals an ~40% reduction in DNA methylation across the 5′LTR in the *GLP*$^{-/-}$ line relative to the wt control, and a partial rescue of this methylation defect in the *GLP*$^{-/-}$*Tg* line (Supplementary Figure S4). Thus, both G9a and GLP have a function in DNA methylation of retrotransposons in ES cells.

DNA methylation at non-repetitive genomic regions is reduced in G9a$^{-/-}$ *cells*

To determine whether this DNA methylation defect extends to non-repetitive elements in the genome, we carried out MeDIP (Weber *et al*, 2005) on genomic DNA isolated from TT2, *G9a*$^{-/-}$ and *G9a*$^{-/-}$Tg ES cells and analysed the methylation status of 11 single-copy genomic regions, including 9 CpG-rich promoters shown previously to be methylated in ES cells (Mohn *et al*, 2008) (Figure 2A). Strikingly, all of the regions that are highly methylated in the TT2 line show a significant decrease in the *G9a*$^{-/-}$ line, including the germline-specific gene *Mage-a2*, shown previously to be aberrantly expressed in *G9a*$^{-/-}$ cells (Tachibana *et al*, 2002). As for the repetitive elements, DNA methylation is increased at most of these regions in the *G9a*$^{-/-}$Tg line. The DNA methylation defect was confirmed through bisulphite sequencing of the Dazl and Tuba3 promoter regions, both of which show an ~40% reduction in DNA methylation in the *G9a*$^{-/-}$ line (Figure 2B). The degree of demethylation across the Dazl promoter is similar to that observed in *Dnmt1*$^{-/-}$ and *Dnmt3a/b*$^{-/-}$ ES cells. In contrast, the degree of demethylation across the Tuba3 promoter in *G9a*$^{-/-}$ cells more closely resembles that observed in the *Dnmt3a/b*$^{-/-}$ line. Thus, although DNA

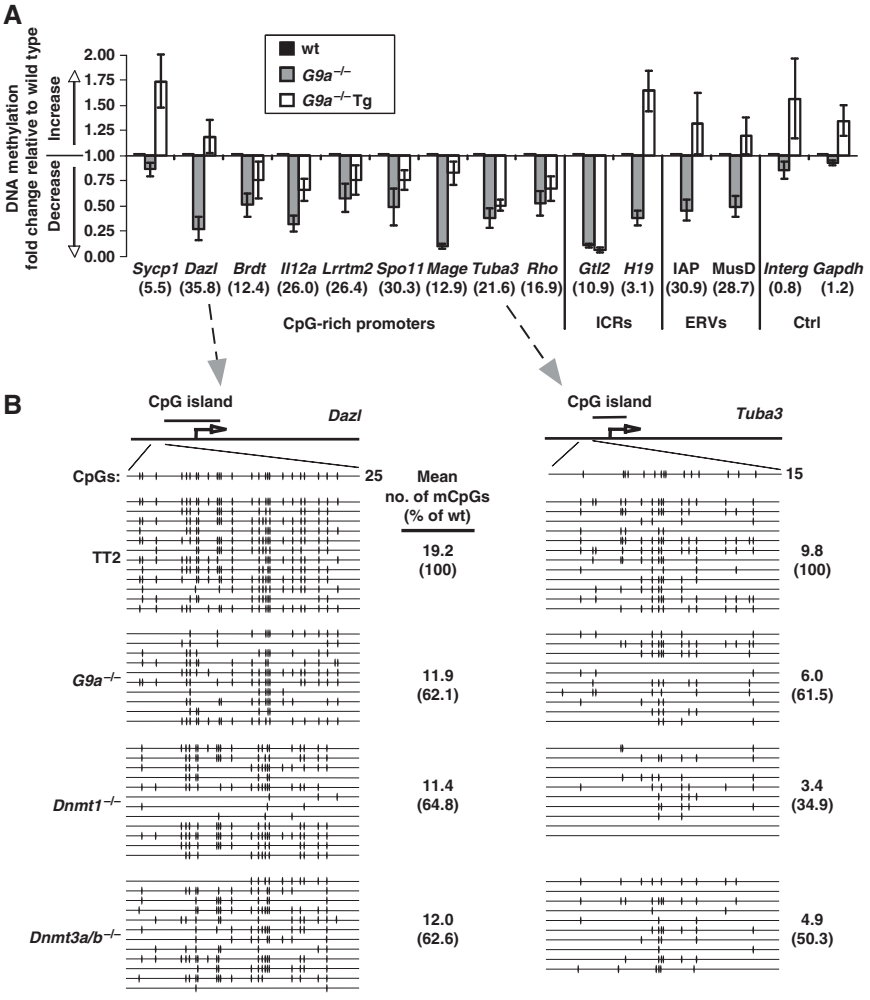


Figure 2 DNA methylation of promoter regions is reduced in *G9a*$^{-/-}$ cells. (**A**) MeDIP followed by quantitative PCR of nine CpG-rich promoter regions and two imprinting control loci (ICR) shown previously to be methylated in ES cells (Mohn *et al*, 2008) was conducted on wt, *G9a*$^{-/-}$ and *G9a*$^{-/-}$Tg lines. IAP and MusD amplicons were included as positive controls. An active housekeeping gene (Gapdh) and a CpG-poor intergenic region (Interg) were included as negative controls. A bar graph illustrating DNA methylation changes in *G9a*$^{-/-}$ and *G9a*$^{-/-}$Tg ES cells relative to wt ES cells (set to 1) is shown. The fold change is normalized to an unmethylated control gene (Hprt). Numbers in parentheses indicate the enrichment in MeDIP relative to Hprt. Error bars indicate the s.e.m. of at least three independent experiments. A lower level of methylation was detected in the *G9a*$^{-/-}$ line than the wt or rescued lines for all of the genes that show a high level of methylation in the TT2 parent line. (**B**) DNA methylation status of the germline-specific Dazl and Tuba3 genes in wt, *G9a*$^{-/-}$, *Dnmt1*$^{-/-}$ and *Dnmt3a/b*$^{-/-}$ cells was confirmed by bisulphite sequencing. The mean number of mCpGs per molecule sequenced is shown, along with the mean % of mCpGs relative to the wild-type line (in parentheses). Both promoters show an ~40% reduction in DNA methylation density in the *G9a*$^{-/-}$ line.

methylation of CpG-rich promoter regions is also dependent on G9a, the degree to which G9a influences DNA methylation state depends on the genomic context.

DNMT expression is not dramatically altered in G9a$^{-/-}$ *cells*

The observed DNA methylation defect prompted us to address whether Dnmt1, Dnmt3a, Dnmt3b or DNMT-like (Dnmt3L) are downregulated in *G9a*$^{-/-}$ ES cells. Quantitative RT–PCR analysis did not reveal a significant difference in mRNA levels of any of the DNMT family members in these lines (Figure 3A). Similarly, quantitative western blot analyses revealed a <2-fold difference in Dnmt1 or Dnmt3a2 expression levels and an ~2-fold higher level of Dnmt3b expression in the *G9a*$^{-/-}$ line than the wt parent line (Figure 3B). These data indicate that the DNA methylation defect observed in *G9a*$^{-/-}$ cells is unlikely to be a consequence of decreased DNMT expression.

Dnmt3a recruitment is reduced in G9a$^{-/-}$ *ES cells*

Introduction of Dnmt3a, Dnmt3a2 or Dnmt3b1 is sufficient to restore DNA methylation of retrotransposons in highly demethylated *Dnmt3a/b*$^{-/-}$ ES cells (Chen *et al*, 2003), indicating that *de novo* DNMT activity is required to maintain retrotransposons in a densely methylated state. To determine whether DNMT recruitment to such elements is perturbed in *G9a*$^{-/-}$ cells, chromatin immunoprecipitation (ChIP) was conducted using Dnmt1-, Dnmt3a/Dnmt3a2- or Dnmt3b-specific antibodies and unmodified histone H3 as an internal control. For Dnmt1 and Dnmt3b, ChIP experiments with two different antibodies specific for each did not yield conclusive results (data not shown). In contrast, a significant reduction in enrichment of Dnmt3a was detected in the LTR regions of MLV, IAP and MusD retrotransposons in the *G9a*$^{-/-}$ line (Figure 4A). Thus, the decrease in DNA methylation observed in *G9a*$^{-/-}$ cells can be attributed at

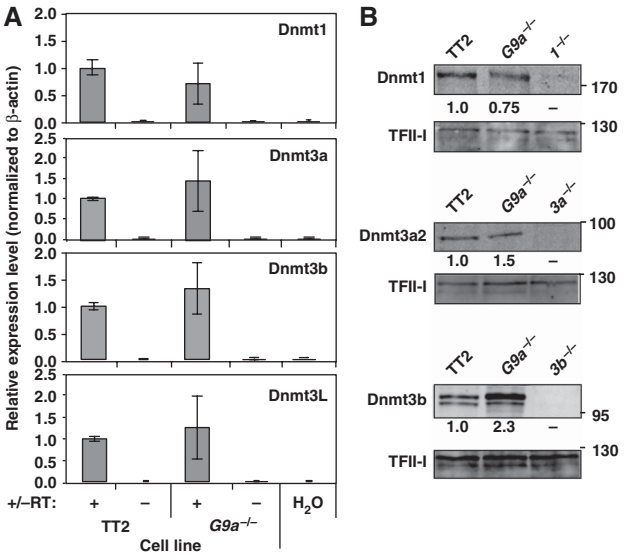


Figure 3 Expression of DNMTs in *G9a*$^{-/-}$ cells. (**A**) RNA was isolated from TT2 wt and *G9a*$^{-/-}$ cells and expression levels of Dnmt1, Dnmt3a, Dnmt3b and Dnmt3L were determined by real-time quantitative RT–PCR, normalized to β-actin (RT–reverse transcriptase). Values represent the mean (±s.d.) expression level relative to the wild-type line, from three independent experiments. (**B**) Western blot analyses using quantitative two-colour fluorescence imaging was performed on nuclear extracts isolated from TT2 and *G9a*$^{-/-}$ cells, using antibodies specific for Dnmt1, Dnmt3a and DNMT3b and TFII-I as an internal control. Extract isolated from DNMT-deficient cells was used as a control for antibody specificity. Relative protein expression levels, normalized to TFII-I, are shown beneath each blot.

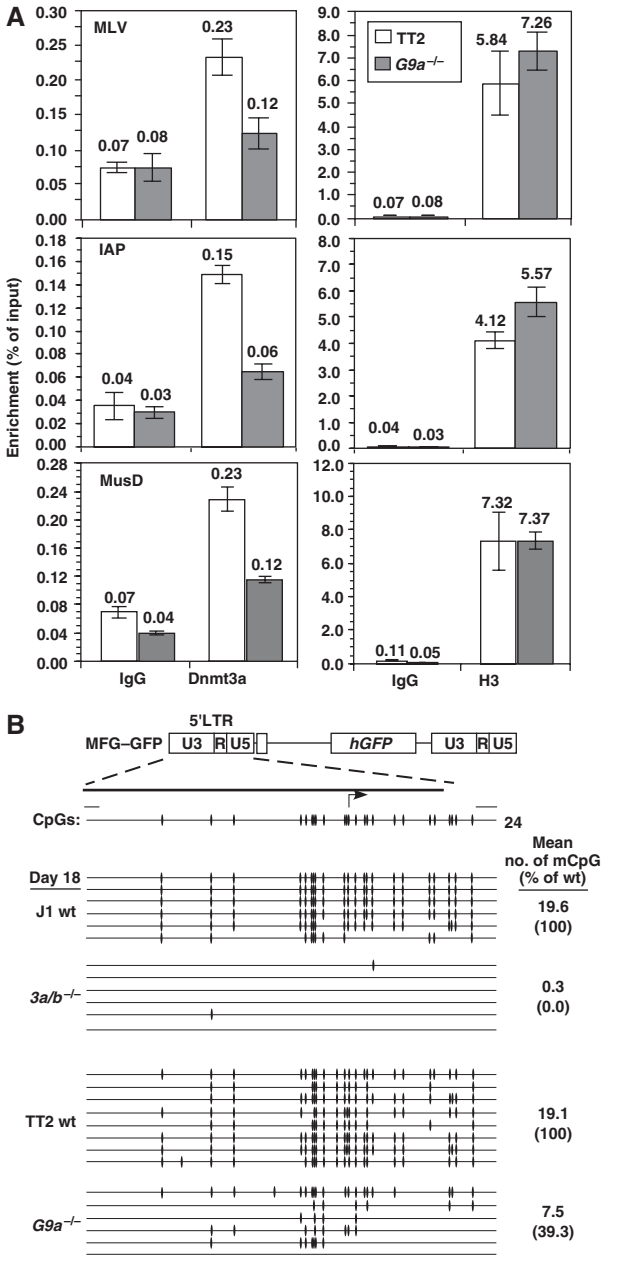


Figure 4 *G9a*$^{-/-}$ ES cells show defects in recruitment of Dnmt3a to ERVs and *de novo* methylation of introduced retroviruses. (**A**) Formaldehyde-fixed chromatin was isolated from TT2 and *G9a*$^{-/-}$ lines and ChIP was conducted using nonspecific IgG or antisera raised against Dnmt3a or unmodified histone H3. Real-time PCR using primers specific for the LTR regions of MLV, IAP and MusD ERVs was carried out and values are presented as percentage of input precipitated (±s.d.) relative to the input in the representative experiment shown. A significant reduction in Dnmt3a enrichment in the *G9a*$^{-/-}$ line relative to the wt control is clearly apparent. (**B**) TT2 wt, *G9a*$^{-/-}$, J1 wt and *Dnmt3a/b*$^{-/-}$ (*3a/b*$^{-/-}$) lines were infected with the retroviral vector MFG–GFP and passaged in the absence of selection. Genomic DNA was isolated on day 18 post-infection and analysed by bisulphite genomic sequencing.

least in part to a decrease in the efficiency of recruitment of *de novo* DNMT activity.

To independently assess whether deletion of G9a not only impairs maintenance of DNA methylation on already methylated loci but also influences the efficiency of *de novo* DNA methylation on previously unmethylated DNA, TT2 wt, *G9a*$^{-/-}$, J1 wt and *Dnmt3a/b*$^{-/-}$ ES cells were infected with the MLV-based retroviral vector MFG–GFP and the DNA methylation status of the proviral LTR was analysed at day 18 post-infection (Figure 4B). As expected, infected *Dnmt3a/b*$^{-/-}$ cells show virtually no DNA methylation at this time point. Strikingly, infected *G9a*$^{-/-}$ cells also show a significantly lower level of DNA methylation (>2.5-fold) than the parent line from which they were derived. Although not as dramatic as the methylation defect observed in the *Dnmt3a/b*$^{-/-}$-deficient cell line, this observation indicates that G9a is required for efficient *de novo* methylation in ES cells.

G9a is not required for transcriptional silencing of retrotransposons

As Dnmt1 was previously shown to be required for silencing of IAP elements in embryos (Walsh *et al*, 1998), we next determined whether the defect in DNA methylation of retrotransposons in *G9a*$^{-/-}$ cells is associated with aberrant expression of these potentially active endogenous elements. Expression of MLV and LINE1 elements was not detected above background levels in wt, *G9a*$^{-/-}$, *Dnmt1*$^{-/-}$ *or* *Dnmt3a/b*$^{-/-}$ lines by northern blotting (Figure 5A and data not shown). In contrast, IAP elements of each subtype (I, IΔ1 and II) (Kuff and Lueders, 1988) (Figure 5B) and MusD elements (Supplementary Figure S5) are expressed at a significantly higher level in the *Dnmt1*$^{-/-}$ line than the *G9a*$^{-/-}$ line, relative to the parent lines from which they were derived. Although not as dramatic as that observed in the *Dnmt1*$^{-/-}$ line, aberrant IAP expression was also observed in the *Dnmt3a/b*$^{-/-}$ line by RT–PCR (Figure 5C). As the G9a and

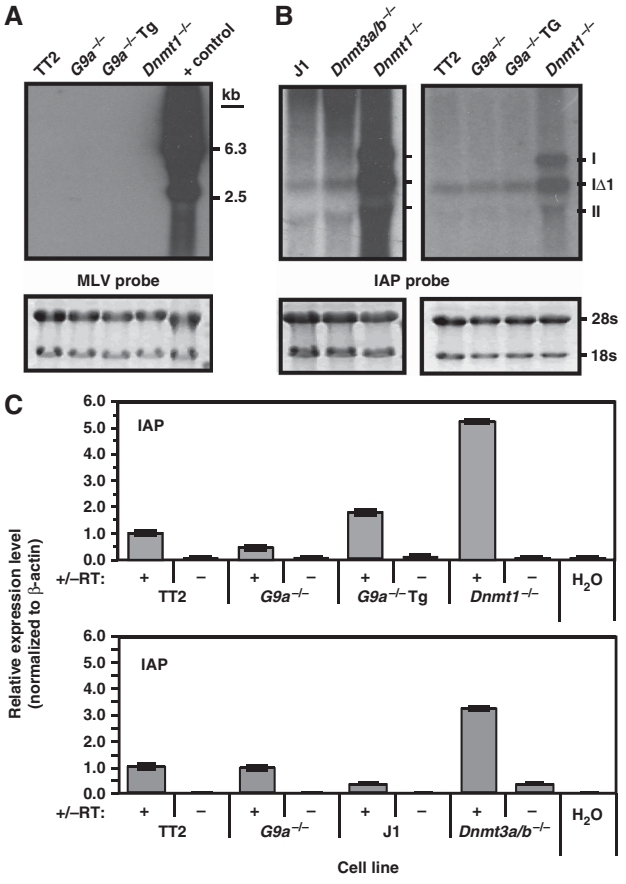


Figure 5 ERVs are not aberrantly expressed in $G9a^{-/-}$ cells. RNA was isolated from TT2 wt, $G9a^{-/-}$, $G9a^{-/-}$Tg, J1 wt, $Dnmt3a/b^{-/-}$ and $Dnmt1^{-/-}$ cells and analysed by northern blotting. RNA isolated from a cell line harbouring an active MLV-based retroviral vector was used as a positive control. 18 and 28s RNA loading controls are shown for each blot. (**A**) No expression of MLV was detected in any of the lines tested, using a probe specific for the MLV LTR region. (**B**) A high level of aberrant expression of the three subtypes (I, IΔ1 and II) of IAP elements was detected in the $Dnmt1^{-/-}$ line, but not the $G9a^{-/-}$ line, using a probe specific for the IAP LTR region. (**C**) Quantitative RT–PCR (+/−RT) using primers specific for the *Pol* region of full-length IAP elements revealed no increase in expression in the parent or $G9a^{-/-}$ lines, but a significant increase in expression in the $Dnmt1^{-/-}$ and $Dnmt3a/b^{-/-}$ lines.

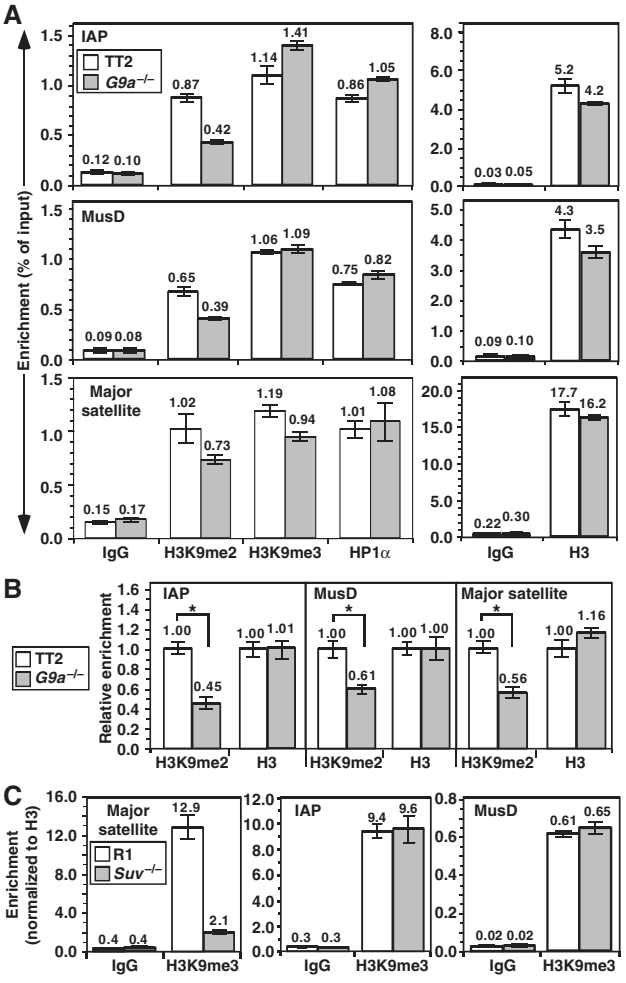


Figure 6 ERVs show a reduction in H3K9 dimethylation in $G9a^{-/-}$ cells, but no reduction in H3K9 trimethylation or HP1α binding. TT2 wt and $G9a^{-/-}$ ES cells were analysed using ChIP with specific for H3K9me2, H3K9me3, HP1α, unmodified H3 and nonspecific IgG (IgG) as a control. Quantitative real-time PCR was conducted using primers specific for IAP or MusD retrotransposons or major satellite repeats. (**A**) Mean enrichment values are presented as percentage of input precipitated (±s.d.), relative to the input in the representative experiment shown. (**B**) Plotting the mean relative enrichment (±s.d.) of H3K9me2 and H3 from three independent experiments reveals an ~2-fold decrease in H3K9me2 in the $G9a^{-/-}$ line relative to the parent line (*$P<0.05$, by Student's *t*-test) but no difference in H3 occupancy at these elements. (**C**) Relative to the R1 wt parent line, $Suv39h1/2^{-/-}$ ($Suv^{-/-}$) ES cells show a dramatic decrease in H3K9me3 only at major satellite repeats.

DNMT deletions were generated in ES cells of differing genetic backgrounds, it is not possible to attribute the differences in ERV expression exclusively to the genes deleted. Nevertheless, taken together with the DNA methylation data, these results indicate either that the level of residual DNA methylation is sufficient to maintain potentially active retroelements in a silent state in TT2 $G9a^{-/-}$ cells, or that an alternative repressive pathway exerts an effect on these elements independent of DNA methylation.

IAP and MusD ERVs show reduced H3K9 dimethylation in $G9a^{-/-}$ cells, whereas H3K9 trimethylation and HP1α binding are unaffected

Several groups have reported that ERVs and other repetitive sequences are marked by H3K9me2 and/or H3K9me3 in murine ES cells (Peters *et al*, 2003; Martens *et al*, 2005; Mikkelsen *et al*, 2007). As G9a is responsible for the majority of H3K9me2 in euchromatin (Tachibana *et al*, 2002), we next wished to determine whether ERVs show a decrease in either of these marks in $G9a^{-/-}$ cells. In addition, as methylation of H3K9 creates a binding site for the HP1 family of transcriptional repressor proteins (Lachner *et al*, 2001; Smallwood *et al*, 2007), we also wished to determine whether recruitment of HP1α is disrupted in the absence of G9a.

ChIP experiments using chromatin isolated from TT2 and $G9a^{-/-}$ cells and antisera specific for H3K9me2, H3K9me3a, HP1α and unmodified H3 revealed an ~2-fold reduction in H3K9me2 in the $G9a^{-/-}$ line in the LTR regions of IAP and MusD elements, in the absence of a significant change in H3 occupancy (Figure 6A and B). Consistent with a previous report indicating that H3K9me2 is decreased in the

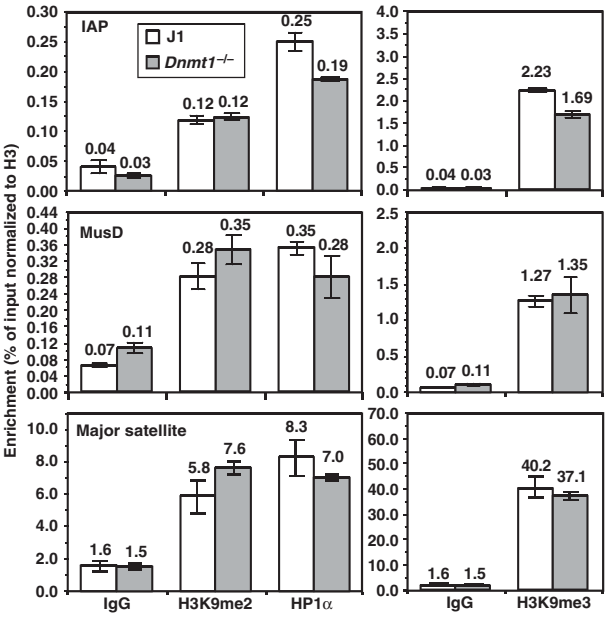


Figure 7 H3K9 methylation and HP1α binding at MusD, IAP and major satellite repeats in *Dnmt1*$^{-/-}$ cells. ChIP was conducted on J1 wt and *Dnmt1*$^{-/-}$ ES cells using antibodies specific for H3K9me2, H3K9me3, HP1α and unmodified H3. Nonspecific IgG was used as a control. Real-time PCR of reverse-crosslinked material using primers specific for IAP, MusD or major satellite repeats was conducted in triplicate and enrichment (±s.d.) is presented as the mean percentage of input material immunoprecipitated, normalized to unmodified H3. IAP elements show a modest reduction in H3K9me3 and HP1α binding in the *Dnmt1*$^{-/-}$ line, but no change in H3K9me2 enrichment. No significant difference in any of these features was detected at MusD or major satellite repeats.

pericentromeric compartment in *G9a*$^{-/-}$ ES cells (Rice *et al*, 2003), an ~2-fold decrease in H3K9me2 was also observed at major satellite repeats, indicating that G9a activity is not confined to the euchromatic compartment and perhaps explaining why major satellite repeats show a DNA methylation defect in *G9a*$^{-/-}$ cells. As we are simultaneously surveying the modification status of histones associated with multiple copies of each repeat, we cannot discriminate between complete loss of H3K9me2 at approximately half of the repeats, or a uniform ~2-fold decrease in this mark across all repeats. Nevertheless, these observations indicate that a significant number of IAP and MusD retrotransposons are direct targets of G9a.

Strikingly, no decrease in H3K9me3 or HP1α binding was detected at IAP or MusD retrotransposons in the *G9a*$^{-/-}$ line (Figure 6A), indicating that an alternative H3K9 HMTase is responsible for the deposition of the H3K9me3 mark at these elements. Although the HMTases Suv39h1 and Suv39h2 are required for H3K9me3 of major satellite repeats (Peters *et al*, 2003), they do not have a major function in the deposition of the H3K9me3 mark at IAP or MusD retrotransposons, as H3K9me3 enrichment is not reduced at these elements in *Suv39h1/2*$^{-/-}$ ES cells either (Figure 6C). In contrast, the promoter region of the G9a-regulated *Mage-a2* gene, which is significantly enriched for H3K9me2 in wt but not *G9a*$^{-/-}$ cells (Tachibana *et al*, 2002; Supplementary Figure S6), showed only very low levels of H3K9me3 enrichment in either line. Taken together, these results indicate that an HMTase with specificity for H3K9 other than G9a, Suv39h1 or Suv39h2 is responsible for H3K9 trimethylation of ERVs and that the deposition of this mark occurs independent of the DNA methylation state of these elements.

In contrast to *G9a*$^{-/-}$ cells, *Dnmt1*$^{-/-}$ cells show no decrease in H3K9me2 across the LTR regions of IAP and MusD elements when normalized to unmodified histone H3 (Figure 7), indicating that Dnmt1 is not required for the deposition of this mark by G9a. Surprisingly, HP1α binding and H3K9me3 across the LTR region of IAP and MusD elements also show only a modest or no reduction, respectively, despite the fact that these elements are hypomethylated and aberrantly expressed in *Dnmt1*$^{-/-}$ cells. The simplest explanation for this observation is that only a small number of proviruses of each class are actually transcribed in *Dnmt1*$^{-/-}$ cells and in turn depleted of H3K9me3 and HP1α. Alternatively, severely hypomethylated IAP and MusD elements may be transcriptionally active despite the presence of these repressive marks. In either case, these observations clearly demonstrate that H3K9me3 and HP1α binding at these elements are not dependent on the presence of dense DNA methylation.

Introduction of a catalytically inactive G9a transgene partially rescues the DNA methylation defect observed in G9a$^{-/-}$ *cells*

The observation of a decrease in DNA methylation of ERVs in *G9a*$^{-/-}$ cells, despite the presence of a high level of residual H3K9 methylation, is surprising, given the known function of K9 methylation in controlling DNA methylation in plants and filamentous fungi (Jackson *et al*, 2002; Freitag and Selker, 2005). To directly address whether DNA methylation of repetitive elements in mammalian cells is dependent on the catalytic activity of G9a, we took advantage of the previously described observation that the DNA methylation defect observed in *G9a*$^{-/-}$ cells is rescued by the introduction of a wt G9a transgene (see Figure 1).

Constructs encoding two G9a mutants (*G9a*$^{-/-}$Tg(C1168A) and *G9a*$^{-/-}$Tg(Y1120V;Y1207F)), each of which harbour amino-acid substitutions in the SET domain that reduce catalytic activity to <1% of that of wt G9a, but do not affect the ability of the encoded protein to form a complex with GLP (see Tachibana *et al*, this issue), were stably introduced into the *G9a*$^{-/-}$ line. Western blot analyses of cell lines stably expressing each of these transgenes revealed that the exogenous wt and mutant proteins are produced at the expected molecular weight (Figure 8A). Furthermore, quantitative western blot analysis reveals that GLP is expressed at similar levels in the wt, *G9a*$^{-/-}$ *and G9a*$^{-/-}$ Tg lines, confirming that the expression of G9a does not significantly influence the stability of its binding partner GLP (Tachibana *et al*, 2005) (Figure 8B). As expected, Southern blot analysis with an IAP-specific probe reveals that DNA methylation is reduced in the *G9a*$^{-/-}$ parent line and restored to wt levels in the *G9a*$^{-/-}$Tg(wt) line (Figure 8C). Strikingly, DNA methylation of IAP elements is also increased in cells expressing either of the catalytic mutants, albeit not to the same level as observed for the wt transgene (Figure 8C–E). DNA methylation levels at MusD (Supplementary Figure S7) and L1 repeats (data not shown) were also significantly increased in the *G9a*$^{-/-}$Tg(C1168A) line relative to the parent *G9a*$^{-/-}$ line.

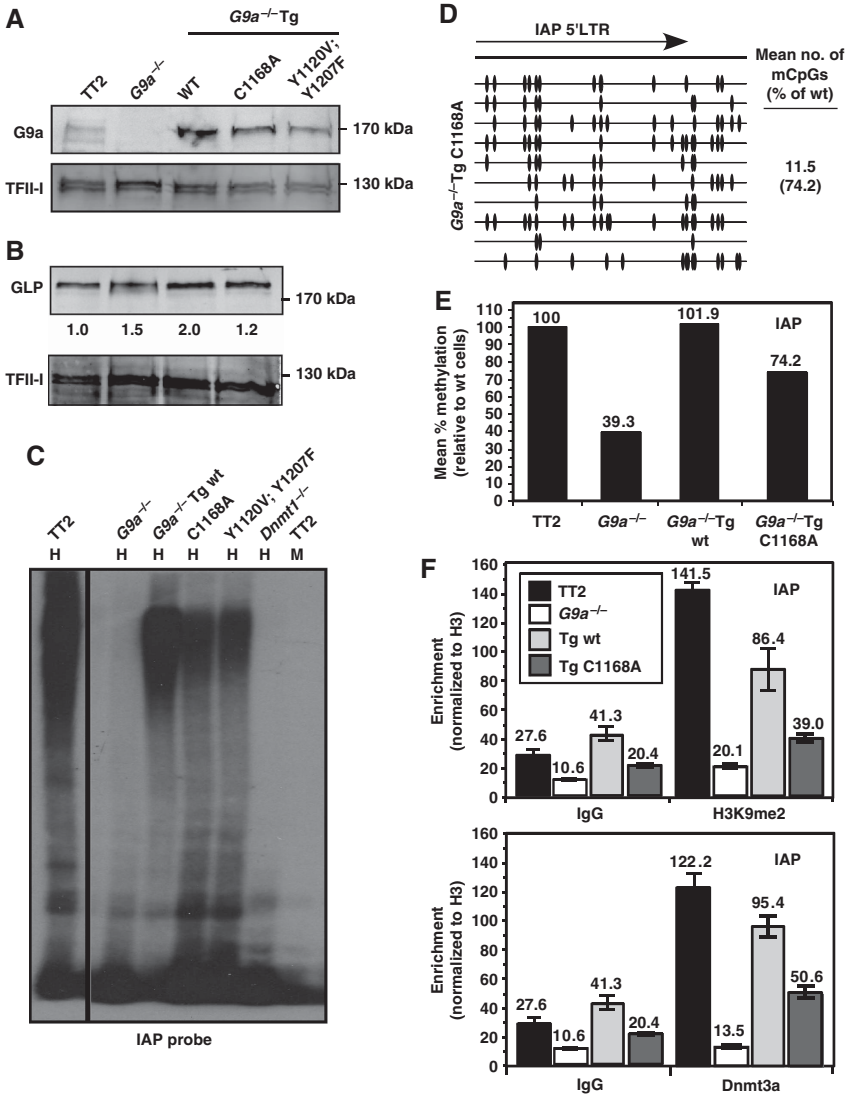


Figure 8 Stable expression of catalytically inactive G9a transgenes is sufficient to rescue the DNA methylation defect observed in $G9a^{-/-}$ ES cells. The $G9a^{-/-}$ line 2-3 was stably transfected with constructs encoding a wt G9a transgene $G9a^{-/-}$Tg(wt), or the mutant G9a transgenes $G9a^{-/-}$Tg(C1168A) and $G9a^{-/-}$Tg(Y1120V;Y1207F). Each of the latter transgenes harbour amino-acid substitutions in the SET domain that reduce the catalytic activity of the encoded protein to <1% of that of the wt protein. (**A**) Western blot analysis of cell lines expressing each of these transgenes reveals that the mutant proteins are expressed at the expected molecular weight. An antibody specific for TFII-I was used as a loading control. (**B**) Quantitative western blot analysis of GLP expression in these lines was determined by normalizing to the signal obtained for endogenous TFII-I on the same blot. (**C**) Genomic DNA isolated from wt (TT2), $G9a^{-/-}$ (2-3), $G9a^{-/-}$Tg(wt), $G9a^{-/-}$Tg(C1168A), $G9a^{-/-}$Tg(Y1120V;Y1207F) and $Dnmt1^{-/-}$ ES cells was digested with *Hpa*II (H) and subject to Southern blotting using an IAP-specific probe. *Msp*I (M) was used as a control. (**D**) Bisulphite analysis using primers specific for the IAP 5′LTR confirms that expression of catalytically inactive G9a partially rescues the DNA methylation defect. (**E**) A bar graph showing the mean no. of mCpGs per molecule sequenced is shown for the bisulphite data presented in (D) and Figure 1D. (**F**) ChIP was conducted on the wt (TT2), $G9a^{-/-}$ (2-3), $G9a^{-/-}$Tg(wt) and $G9a^{-/-}$Tg(C1168A) lines using antisera raised against H3K9me2, Dnmt3a or unmodified H3. Nonspecific IgG was included as a control. Real-time PCR of reverse-crosslinked material using IAP-specific primers was carried out in triplicate and enrichment (±s.d.) is presented as the mean percentage of input material immunoprecipitated, normalized to unmodified H3.

Importantly, the level of H3K9me2 enrichment in the $G9a^{-/-}$Tg(C1168A) line remains significantly below that observed for the TT2 and $G9a^{-/-}$Tg(wt) lines (Figure 8F). Furthermore, rescue of the DNA methylation defect in the $G9a^{-/-}$Tg(C1168A) line is accompanied by recruitment of Dnmt3a to the IAP LTR, although at a level lower than that observed in the wt parent line or the $G9a^{-/-}$ line rescued with the wt G9a transgene (Figure 8F). Taken together, these results reveal that independent of its catalytic activity, G9a promotes *de novo* DNA methylation through enhancing recruitment of Dnmt3a.

Discussion

We demonstrate that G9a is required for DNA methylation of representative LTR and non-LTR retrotransposons and a number of CpG-rich promoters in murine ES cells. However, unlike the H3K9 HMTases in plants and filamentous fungi (Jackson *et al*, 2002; Freitag and Selker, 2005), G9a promotes DNA methylation of repetitive elements independent of its catalytic activity. How might G9a influence DNA methylation, given that H3K9 methylation *per se* does not seem to be the predominant trigger? Although it is possible

that deletion of G9a influences DNA methylation through an indirect mechanism, given that H3K9me2 is decreased at each of the genomic regions analysed, we favour the possibility that G9a–GLP exert an effect *in cis* to promote DNA methylation. Indeed, two groups recently reported that in somatic cells, G9a interacts directly with Dnmt1 in a complex that includes PCNA (Esteve *et al*, 2006; Sharif *et al*, 2007), indicating that Dnmt1 and G9a coordinate H3K9 methylation and maintenance DNA methylation at the replication fork.

Our survey of the DNA methylation status of a number of repetitive and single-copy genomic sequences reveals that the extent of the defect in *G9a*$^{-/-}$ ES cells is generally not as severe as that in *Dnmt1*$^{-/-}$ ES cells. Although we did detect an interaction between G9a and GLP through co-immunoprecipitation, as described earlier (Tachibana *et al*, 2005), we were unable to detect an interaction between G9a and any of the DNMTs in ES cells (MCL and SL, unpublished data). On the other hand, we did find that Dnmt3a recruitment to the promoter regions of LTR retrotransposons was reduced in the *G9a*$^{-/-}$ line relative to the wt control.

Thus, although we cannot confirm whether G9a influences Dnmt1 activity in ES cells, we propose that G9a regulates DNA methylation in these cells at least in part by promoting *de novo* DNMT activity in *cis*. In support of this model, we find that introduced MLV-based retroviral vectors, which are unmethylated at the time of integration, are not efficiently *de novo* methylated in *G9a*$^{-/-}$ cells and show a defect in silencing similar to that observed in *Dnmt3a/b*$^{-/-}$ ES cells infected with the same virus (KBD and MCL, in preparation). As maintenance methylation by Dnmt1 is reported to be an inefficient process in ES cells (Liang *et al*, 2002), continual *de novo* methylation may be required to preserve DNA methylation homoeostasis. Alternatively, active demethylation by an as yet unidentified DNA demethylase may necessitate ongoing *de novo* methylation by Dnmt3a and/or Dnmt3b to maintain steady-state levels of this epigenetic mark.

Intriguingly, despite the fact that the potentially active IAP and MusD retrotransposons show a dramatic reduction in DNA methylation in *G9a*$^{-/-}$ cells, these interspersed repetitive elements remain transcriptionally inactive. As *G9a*$^{-/-}$ cells show a somewhat higher level of residual DNA methylation than *Dnmt1*$^{-/-}$ ES cells, in which these elements are aberrantly expressed, it is possible that potentially active ERVs are expressed only when DNA methylation density drops below a critical threshold. Consistent with this model, unlike Dnmt1-null mice, which show high levels of IAP expression (Walsh *et al*, 1998), compound heterozygous mice carrying a hypomorphic Dnmt1 allele over a null allele show genome-wide hypomethylation but no detectable IAP expression (Howard *et al*, 2008).

Alternatively, as H3K9me3 enrichment and HP1α binding at IAP and MusD elements are not dramatically reduced in *G9a*$^{-/-}$ or *Dnmt1*$^{-/-}$ cells, it is possible that an alternative repressive pathway maintains the vast majority of ERVs in a silent state independent of DNA methylation. Our observations clearly show that an HMTase with specificity for H3K9 other than G9a, GLP, Suv39h1 or Suv39h2 marks LTR retrotransposons in ES cells, leaving SETDB1, which shows di- and tri-methyl HMTase activity towards H3K9 *in vitro* and *in vivo* (Wang *et al*, 2003), or the closely related SETDB2, as the remaining candidates in the Suv39 family of HMTases for this activity.

Given the well-documented deleterious effects of retrotransposition on genomic integrity, the existence of a DNA methylation-independent silencing pathway may serve to minimize proviral expression during those stages in embryonic development when DNA methylation levels are relatively low, such as following fertilization or in the developing germ line. Intriguingly, deletion of the Piwi protein Mili leads to derepression of L1 and IAP ERVs and the loss of DNA methylation at L1 elements (Aravin *et al*, 2007), revealing that Piwi-interacting RNAs (piRNAs) generated by transposable elements in the germ line are required to maintain these elements in a silent state. Such piRNAs may have a function in the targeting of H3K9 HMTase activity to homologous repetitive elements prior to *de novo* DNA methylation of these elements.

We show that G9a is required for DNA methylation in ES cells not only of repetitive elements but also of the CpG-rich promoter regions of a number of genes that are normally densely methylated in ES cells. These results are consistent with those reported by Tachibana and colleagues (see accompanying paper by Tachibana *et al*), and indicate that in addition to the deposition of the H3K9me2 mark, the G9a–GLP complex may have a genome-wide influence on DNA methylation homoeostasis in ES cells. As GLP expression is downregulated in primordial germ cells, coincident with genome-wide DNA demethylation in these cells (Seki *et al*, 2005, 2007), it is possible that the G9a–GLP heteromeric complex has a function in the programmed changes in DNA methylation that occur not only following fertilization but also in the developing germ line.

Materials and methods

Cell lines

J1 wt (129S4/SvJae), *Dnmt1*$^{c/c}$ (Dnmt1$^{-/-}$) (Lei *et al*, 1996), Dnmt3a and Dnmt3b double-negative (*Dnmt3a/b*$^{-/-}$) (Okano *et al*, 1999), TT2 wt (c57BL/6xCBA), *G9a*$^{-/-}$ (clones 2-3 and 22-10), *G9a*$^{-/-}$*Tg* (clone 15-3) (Tachibana *et al*, 2002), *GLP*$^{-/-}$, *GLP*$^{-/-}$*Tg* (Tachibana *et al*, 2005), *G9a*$^{-/-}$Tg(wt), *G9a*$^{-/-}$Tg(C1168A) (clone G4), *G9a*$^{-/-}$Tg(Y1120V;Y1207F) (clone G7), R1 wt (129X1/SvJ × 129S1) and Suv39h1 and Suv39h2 double-negative (*Suv39h1/2*$^{-/-}$) (Peters *et al*, 2001) ES cells were passaged every 48–72 h in DMEM supplemented with 15% FBS (HyClone), 20 mM HEPES, 0.1 mM non-essential amino acids, 0.1 mM 2-mercaptoethanol, 100 U/ml penicillin, 0.05 mM streptomycin, leukaemia-inhibitory factor and 2 mM glutamine on gelatinized plates.

Bisulphite sequencing and MeDIP analyses

Genomic DNA was subject to sodium bisulphite conversion using the EZ DNA Methylation-Gold kit (Zymo Research) as described earlier (Appanah *et al*, 2007). MeDIP was conducted as described earlier (Weber *et al*, 2007). Detailed protocols are provided in the Supplementary data.

Northern and Southern blot analyses

Southern blot analyses, restriction digests, membrane transfers and preparation of the DNA probe were performed by standard methods. A detailed protocol is provided in the Supplementary data.

Quantification of proviral mRNA levels

RNA was isolated using Tri reagent (Sigma) according to the manufacturer's protocol. DnaseI-treated RNA was subject to first-strand cDNA synthesis using RevertAid H Minus kit (Fermentas) in the presence or absence of reverse transcriptase. Quantitative RT–PCR using MLV-, IAP- and MusD-specific primers, or β-actin-specific primers as an internal control (all primer sequences are listed in Supplementary Table 1), was conducted with EvaGreen dye (Biotium) on an Opticon 2 thermal cycler (Bio-Rad). Relative expression levels were determined by normalizing to the *β-actin* gene.

Antibodies and ChIP experiments

ChIP for histones (Appanah *et al*, 2007) and non-histone proteins (O'Geen *et al*, 2007) was conducted as described. Details are provided in the Supplementary data.

Western blot analysis

Nuclear extractions were conducted as described (Tachibana *et al*, 2002). Western blot analyses were conducted using the Odyssey Infrared Imaging System (LI-COR Biosciences), as described in the Supplementary data.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Dr En Li for the *Dnmt1*$^{-/-}$ and *Dnmt3a/b*$^{-/-}$ cell lines, Dr Thomas Jenuwein for the *Suv39h1/2*$^{-/-}$ ES cell line and Dr Stephen Smale for HP1-specific antibodies and the UBC flow cytometry and Nucleic Acid Protein Service facilities. We thank Mark Groudine, James Ellis, Carolyn Brown and Jacob Hodgson for comments on the article. This study was supported by CIHR grant 77805 to MCL, 10825 to DLM and a Grant-in Aid from the Ministry of Education, Science, Technology and Culture of Japan to MT and YS. MCL is a Scholar of the Michael Smith Foundation for Health Research. Research in the laboratory of DS is supported by the Novartis Research Foundation and the European Union (LSHG-CT-2004-503433 and LSHG-CT-2006-037415).

References

Appanah R, Dickerson DR, Goyal P, Groudine M, Lorincz MC (2007) An unmethylated 3′ promoter-proximal region is required for efficient transcription initiation. *PLoS Genet* **3:** e27

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316:** 744–747

Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. *Cell* **128:** 669–681

Chen T, Ueda Y, Dodge JE, Wang Z, Li E (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* **23:** 5594–5605

Chen T, Ueda Y, Xie S, Li E (2002) A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active *de novo* methylation. *J Biol Chem* **277:** 38746–38754

Esteve PO, Chin HG, Smallwood A, Feehery GR, Gangisetty O, Karpf AR, Carey MF, Pradhan S (2006) Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev* **20:** 3089–3103

Feldman N, Gerson A, Fang J, Li E, Zhang Y, Shinkai Y, Cedar H, Bergman Y (2006) G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* **8:** 188–194

Freitag M, Selker EU (2005) Controlling DNA methylation: many roads to one modification. *Curr Opin Genet Dev* **15:** 191–199

Goyon C, Rossignol JL, Faugeron G (1996) Native DNA repeats and methylation in *Ascobolus*. *Nucleic Acids Res* **24:** 3348–3356

Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A (2008) Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* **27:** 404–408

Ikegami K, Iwatani M, Suzuki M, Tachibana M, Shinkai Y, Tanaka S, Greally JM, Yagi S, Hattori N, Shiota K (2007) Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. *Genes Cells* **12:** 1–11

Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416:** 556–560

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128:** 693–705

Kuff EL, Lueders KK (1988) The intracisternal A-particle gene family: structure and functional aspects. *Adv Cancer Res* **51:** 183–276

Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410:** 116–120

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* **13:** 1192–1200

Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, Li E (1996) *De novo* DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122:** 3195–3205

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69:** 915–926

Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA (2002) Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol Cell Biol* **22:** 480–491

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292:** 2077–2080

Mabuchi H, Fujii H, Calin G, Alder H, Negrini M, Rassenti L, Kipps TJ, Bullrich F, Croce CM (2001) Cloning and characterization of CLLD6, CLLD7, and CLLD8, novel candidate genes for leukemogenesis at chromosome 13q14, a region commonly deleted in B-cell chronic lymphocytic leukemia. *Cancer Res* **61:** 2870–2877

Mager DL, Freeman JD (2000) Novel mouse type D endogenous proviruses and ETn elements share long terminal repeat and internal sequences. *J Virol* **74:** 7221–7229

Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL (2006) Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet* **2:** e2

Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* **24:** 800–812

Medstrand P, van de Lagemaat LN, Mager DL (2002) Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res* **12:** 1483–1495

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520–562

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–560

Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schubeler D (2008) Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30:** 755–766

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292:** 110–113

O'Geen H, Squazzo SL, Iyengar S, Blahnik K, Rinn JL, Chang HY, Green R, Farnham PJ (2007) Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet* **3:** e89

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99:** 247–257

Peters AH, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AA, Perez-Burgos L, Kohlmaier A, Opravil S, Tachibana M, Shinkai Y, Martens JH, Jenuwein T (2003) Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell* **12:** 1577–1589

Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, Jenuwein T (2001) Loss of the

Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107:** 323–337

Rice JC, Briggs SD, Ueberheide B, Barber CM, Shabanowitz J, Hunt DF, Shinkai Y, Allis CD (2003) Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol Cell* **12:** 1591–1598

Sampath SC, Marazzi I, Yap KL, Sampath SC, Krutchinsky AN, Mecklenbrauker I, Viale A, Rudensky E, Zhou MM, Chait BT, Tarakhovsky A (2007) Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Mol Cell* **27:** 596–608

Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y (2005) Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278:** 440–458

Seki Y, Yamaji M, Yabuta Y, Sano M, Shigeta M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M (2007) Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134:** 2627–2638

Sharif J, Muto M, Takebayashi SI, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, Tajima S, Mitsuya K, Okano M, Koseki H (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450:** 908–912

Smallwood A, Esteve PO, Pradhan S, Carey M (2007) Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev* **21:** 1169–1178

Stancheva I (2005) Caught in conspiracy: cooperation between DNA methylation and histone H3K9 methylation in the establishment and maintenance of heterochromatin. *Biochem Cell Biol* **83:** 385–395

Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H, Shinkai Y (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* **16:** 1779–1791

Tachibana M, Ueda J, Fukuda M, Takeda N, Ohta T, Iwanari H, Sakihama T, Kodama T, Hamakubo T, Shinkai Y (2005) Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev* **19:** 815–826

Tamaru H, Selker EU (2001) A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414:** 277–283

Walsh CP, Chaillet JR, Bestor TH (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20:** 116–117

Wang H, An W, Cao R, Xia L, Erdjument-Bromage H, Chatton B, Tempst P, Roeder RG, Zhang Y (2003) mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol Cell* **12:** 475–487

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* **37:** 853–862

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39:** 457–466

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13:** 335–340

Zhou Y, Cambareri EB, Kinsey JA (2001) DNA methylation inhibits expression and transposition of the *Neurospora* Tad retrotransposon. *Mol Genet Genomics* **265:** 748–754

Supplementary Methods

Bisulphite sequencing analyses. Nested or semi-nested PCR was conducted on converted DNA using primers specific for each of the elements analyzed (primer sequences are listed in Supplementary Information, Table 1). Amplification products were cloned via T/A cloning using the pGEM-T easy kit (Promega) and individual clones were sequenced using BigDye v3.1 chemistry. Sequences were analyzed using Sequencher software (Gene Codes). The mean number of methylated CpGs (mCpGs)/molecule sequenced is presented for each set of samples. As the number of CpGs in the 5'LTR regions of ERVs is heterogeneous, a CpG density "key" is not shown for these elements.

Northern and Southern analysis. Genomic DNA was extracted using DNAzol (Invitrogen) according to the manufacturer's instructions. For each digested sample, 2-3 mg of genomic DNA was loaded per lane. LTR-specific probes were used for analysis of IAP and MLV elements. Probes were labeled with [α-^{32}P]dCTP using the Random Primers DNA Labeling System (Invitrogen). Membranes were pre-hybridized in ExpressHyb (BD Biosciences) at 60°C for 3-5 hours, hybridized overnight at the same temperature in fresh ExpressHyb, washed according to the manufacturer's protocol and exposed to film.

For Northern analyses, total RNA was extracted using the RNeasy RNA isolation kit (Qiagen) according to the manufacturer's instructions. For each lane, 5-8 mg of RNA was denatured, subject to electrophoresis in a 0.8% agarose, 1.9% formaldehyde gel in

1xMOPS buffer and transferred overnight to a Zeta-probe nylon membrane (Bio-Rad). MusD, IAP and MLV specific probes were labeled with [α-^{32}P]dCTP as above. Membranes were pre-hybridized at 68°C for 3-5 hours, hybridized overnight at the same temperature in fresh ExpressHyb, washed according to the manufacturer's protocol and exposed to film. The probes for the MusD *gag* region and the IAP *pol* gene were synthesized by PCR from C57BL6 genomic DNA.

ChIP experiments. ChIP for histones was conducted as described (Appanah et al, 2007). Briefly, 2.4-4x10^7 exponentially growing ES cells were incubated in the presence of 1% (v/v) formaldehyde for 10 minutes at 37°C and chromatin was sonicated to fragment sizes of 300-1000bp using the Bioruptor sonicator (Diagenode). Purified, reverse-crosslinked material was resuspended in 50 ml elution buffer (Qiagen) and quantified by real-time quantitative PCR using 2 µl of template with EvaGreen dye and hot-start Taq polymerase (Fermentas). ChIP for non-histone proteins was conducted as described (O'Geen et al, 2007). Purified rabbit IgG (10µg, Sigma) was used as a control. Antibodies specific for H3 (2.5µg; Abcam, ab1791), H3K9me2 (5µg; Upstate, 07-441), H3K9me3 (5µg; Upstate, 07-442 or 2.5-5µg; Abcam, ab8898), HP-1α (kind gift of Stephen Smale), and Dnmt3a (5µg; Imgenex, IMG-268A) were used. Of note, while Dnmt3a2 is the predominant protein recognized by the Dnmt3a "specific" antibody, this reagent also shows weak cross-reactivity with Dnmt3b1, as determined by Western blotting (Chen et al, 2002).

Reactions were carried out in triplicate (conditions available upon request). For each amplicon, the amount of input and immunoprecipitated DNA was calculated using the standard curve method. The "percent of input" was subsequently calculated by taking the ratio of these values. For the major satellite amplicon, DNA was diluted 1:500 prior to PCR.

MeDIP and real-time PCR. Genomic DNA was isolated from 2-3 independent passages of wildtype, G9a$^{-/-}$ and G9a$^{-/-}$Tg ES cells and prepared for MeDIP as previously described (Weber et al, 2007). In brief, 2µg sonicated (300-1000bp) genomic DNA was precipitated with 2µg antibody against 5-methylcytidine (Eurogentec, BI-MECY-1000) and bound DNA was recovered with 8µl M-280 Sheep anti-mouse IgG Dynabeads (Invitrogen). For real-time PCR, 30 ng sonicated genomic input DNA and 1/40 of a MeDIP reaction were used as template and quantified with SYBR Green PCR mastermix (Applied Biosystems) using standard cycling conditions on an ABI Prism 7000 detection system. Relative enrichments were determined by normalizing against an active gene (Hprt). Reactions were performed in duplicates, averaged for each PCR and standard errors were calculated between the averaged duplicate reactions of biological replicate experiments. Primers for DNA methylated loci in ES cells were chosen based on a recent genome-wide study (Mohn et al, 2008) and are listed in Supplementary Table 1.

Western analysis. For Western blotting, 60-100 µg of nuclear extract was added per lane and quantitative western analyses were conducted using the Odyssey Infrared Imaging System (LI-COR Biosciences), according to the manufacturer's protocol. Antibodies

used include: G9a (1:2000; PPMX, A8620A), GLP (1:1000; PPMX, B0422), Suz12 (Kind gift from Yi Zhang), Bmi1 (1:6000; Upstate, 05-637), TFII-I (1:2000; kind gift of Ivan Sadowski), Dnmt1 (1:300; Imgenex, IMG-261A), Dnmt3A (1:300; IMG-268A) and Dnmt3b (1:1000; Imgenex, IMG-184A).

Supplementary References

Appanah R, Dickerson DR, Goyal P, Groudine M, Lorincz MC (2007) An unmethylated 3' promoter-proximal region is required for efficient transcription initiation. *PLoS Genet* **3:** e27

Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schubeler D (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30:** 755-766

O'Geen H, Squazzo SL, Iyengar S, Blahnik K, Rinn JL, Chang HY, Green R, Farnham PJ (2007) Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet* **3:** e89

Tachibana M, Ueda J, Fukuda M, Takeda N, Ohta T, Iwanari H, Sakihama T, Kodama T, Hamakubo T, Shinkai Y (2005) Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev* **19:** 815-826

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39:** 457-466

Supplementary Figure Legends

Supplementary Figure S1. Western analyses of G9a and GLP expression in the TT2, *G9a*-/-, *G9a*-/-*Tg*, *GLP*-/- and *GLP* -/-*Tg* lines. (A) Nuclear extract was isolated from TT2, *G9a*-/- and *G9a*-/-*Tg* lines and analyzed by Western blotting using antibodies specific for G9a and Suz12 as a loading control. While no G9a expression was detected in the *G9a*-/- line, a significant level of G9a expression was detected in the *G9a*-/-*Tg* line. (B) Similarly, while no expression of GLP was detected in the *GLP*-/- line, a significant level of GLP was detected in the *GLP* -/-*Tg* line. These data confirm that G9a and GLP transgenes are stably expressed in the *G9a*-/-*Tg* and *G9a*-/-*Tg* lines, respectively.

Supplementary Figure S2. MusD ERVs show a dramatic reduction in DNA methylation density in *G9a*-/- cells. Bisulphite analysis of TT2, *G9a*-/-, *Dnmt1*-/- and *G9a*-/-Tg was conducted using primers specific for the 5' LTR and downstream regions of MusD. In the *G9a*-/- and *Dnmt1*-/- lines, these class II ERVs show a significantly lower level of methylation across the LTR and downstream region than in the TT2 line. Introduction of a G9a transgene (*G9a*-/-Tg) rescues the observed DNA methylation defect. The mean number of CpGs/molecule sequenced is shown to the right of each set of sequenced samples, along with the mean % of mCpGs relative to the wildtype line (in parentheses).

Supplementary Figure S3. Major satellite repeats show a DNA methylation defect in G9a-/- ES cells. Genomic DNA isolated from R1 and *Suv39h1/2*$^{-/-}$ lines as well as TT2,

G9a-/-, *G9a-/-* (2-3), *G9a*$^{-/-}$Tg (15-3), J1, *Dnmt1*$^{-/-}$ and *Dnmt3a/b*$^{-/-}$ lines was digested with the methylation-sensitive restriction enzyme HpyCH4IV and subject to Southern analysis using a probe specific for major satellite repeats. The *G9a-/-* line shows a DNA methylation defect similar to that observed in the *Suv39h1/2*$^{-/-}$, *Dnmt1*$^{-/-}$ and *Dnmt3a/b*$^{-/-}$ lines, but not in the parent lines from which these mutants were derived. This defect is reversed in the *G9a*$^{-/-}$Tg line.

Supplementary Figure S4. DNA methylation of MLV and IAP ERVs is reduced in *GLP*$^{-/-}$ cells. (A) Genomic DNA isolated from TT2 *wt*, *GLP*$^{-/-}$ and *GLP*$^{-/-}$Tg lines was digested with HpaII (H) and subject to Southern blotting using probes specific for MLV or IAP elements. While the *GLP*$^{-/-}$ line shows a dramatic reduction in DNA methylation relative to the TT2 parent line, the *GLP* $^{-/-}$Tg line shows a significantly higher level of DNA methylation of both ERVs. (B) Bisulphite analysis confirms that a significantly lower level of methylation is present across the 5' LTR of endogenous MLV elements in the *GLP*$^{-/-}$ line. The mean number of mCpGs/molecule sequenced (combined with the data shown in Figure 1d for the TT2 line) is shown to the right of each set of sequenced samples, along with the mean % of mCpGs relative to the wildtype line (in parentheses).

Supplementary Figure S5. MusD ERVs are transcriptionally silent in *G9a-/-* cells. (A) Northern analysis or RNA isolated from J1, *Dnmt3a/b-/-*, *Dnmt1-/-*, TT2, *G9a-/-* and *G9a-/-*Tg lines reveals aberrant expression of MusD elements (~7.5 kb) in the *Dnmt1-/-* line, but not the *Dnmt3a/b-/-* or *G9a-/-*, or *G9a-/-*Tg lines. Longer exposure of the blot revealed a low level of expression in the *wt* and *G9a-/-* lines (data not shown). (B)

Quantitative RT-PCR (+/-RT) using primers specific for MusD elements revealed a significantly higher level of expression in the *Dnmt1-/-* line than the *G9a-/-* line, consistent with the results obtained by Northern blotting.

Supplementary Figure S6. The promoter region of the Mage-a2 gene is marked by H3K9me2 but not H3K9me3 in wildtype ES cells. ChIP was conducted on *wt* and *G9a-/-* ES cells using antisera specific for H3K9me2, H3K9me3, unmodified H3 and non-specific IgG as a control. Real-time PCR was carried out using primers specific for the promoter region of the Mage-a2 gene. Enrichment (+/-SD) is presented as the percentage of input material immunoprecipitated. A significantly lower level of enrichment of H3K9me2 was detected in the *G9a-/-* line than the *wt* parent line at the promoter region of the Mage-a2 gene. In contrast, both lines showed only very low levels of enrichment of H3K9me3 in this region. The lower level of H3 occupancy observed in the *G9a-/-* line likely reflects transcription-coupled depletion of nucleosomes in the promoter region of the Mage-a2 gene, which is aberrantly expressed in this line (Tachibana et al, 2002).

Supplementary Figure S7. Stable expression of a catalytically inactive G9a transgene in G9a-/- ES cells rescues the DNA methylation defect at MusD elements. Bisulphite analysis of the *G9a-/-* line 2-3 stably transfected with constructs encoding a *wt* G9a transgene *G9a-/-*Tg(*wt*), or the mutant G9a transgenes *G9a-/-*Tg(C1168A) was conducted using primers specific for the 5' LTR and downstream regions of MusD. The mean number of mCpGs/molecule sequenced is shown, along with the mean % of mCpGs relative to the wildtype line (in parentheses). Taken together with the data shown

in figure S2 (bar graph), these data confirm that expression of catalytically inactive G9a rescues the DNA methylation defect observed in the parent *G9a-/-* line. Interestingly, the clone expressing the wildtype transgene, which is expressed at significantly higher levels than the endogenous protein (see Figure 7A) shows a higher level of methylation than the TT2 parent line, suggesting that expression of endogenous G9a is limiting with respect to its influence on DNA methylation in ES cells.

Table 1. Primers used in this study

RT-PCR		
β-actin	+	TCATGAAGTGTGACGTTGACATCCGT
	-	CCTAGAAGCACTTGCGGTGCACGATGGAG
MusD	+	GTGGTATCTCAGGA(G/A)GAGTGCC
		GGGCAGCTCCTCTATCTGAGTG
IAP	+	AAGCAGCAATCACCCACTTTGG
	_	CAATCATTAGATG(T/C)GGCTGCCAAG
Dnmt1	+	TCGGCTGAACAACCCCGGCACCAC
	_	CTTCAGCACCATGGAGCGTCTGTAGG
Dnmt3a	+	GTCCGCAGCGTCACACAGAAGC
	_	TCTTTGGCGTCAATCATCACGG
Dnmt3b	+	AGGTTTATATGAGGGCACAGGAAGGC
	_	CATGTTGGACACGTCCGTGTAGTGAGC
Dnmt3L	+	ACTGAGGATGACCAAGAGACAAC
	_	CTCTTCAGCCCTGGAATGTTGCTC
Bisulfite analysis		Y= C or T, R= A or G
MLV 1st round:	+	TATTTTGTAAGGTATGAAAAAGTATTAGAGT
(RLTR4_Mm-int)	-	AAATCRATAATCCCTAAACAAAAATCTCCA
2nd round:	+	TAAATTTGTGTGTTTGTTAATGTTTTGATT
	-	AAATCRATAATCCCTAAACAAAAATCTCCA
IAP 5'LTR 1st round:	+	GGYGTTGATAGTTGTGTTTTAAGTGGTAAAT
	-	ATTCTAATTCTAAAATAAAAAATCTTCCTTA
2nd round:	+	GATAGTTGTGTTTTAAGTGGTAAATAAATA
	-	ATTCTAATTCTAAAATAAAAAATCTTCCTTA
MusD 5'LTR 1st round:	+	AAATTTGAGTTTTGATTAGTATGAAATTGT
	-	AATCTAATATTTCTTCTTCCTTAAACCATA
2nd round:	+	AAATTTGAGTTTTGATTAGTATGAAATTGT
	-	AACTTTAAACCCTTTCTTCTTCCACCTAAA
Dazl 1st round:	+	GGTTYGAGTTTTATTGATAGATAGATGGAT
	-	AACACCCTACAACTCAACTCTACTATAA
2nd round:	+	GATTTTTGTTATTTTTTAGTTTTTTTAGGAT
	-	AAAATTCTCTCAACTAACCTAACTTATTTCT
Tuba3 1st round:	+	TTAGGGGYGGTTTTAGGTTTTATATTTTAT
	-	ATTACCACCAAACATAACACACATCTATAA
2nd round:	+	ATTTTATTAATGATTGGATGTGGTTTAA
	-	AAATAAACAACTACTCACACAAACTTCC
ChIP analysis		
Major satellite	+	GACGACTTGAAAAATGACGAAATC
	-	CATATTCCAGGTCCTTCAGTGTGC
***IAP* (LTR)**	+	CTCCATGTGCTCTGCCTTCC
	-	CCCCGTCCCTTTTTTAGGAGA
MusD (LTR)	+	CCCTTCCTTCATAACTGGTGTCGCA
	-	TAGCATCTCTCTGCCATTCTTCAGG
Mage-a2	+	TTGGTGGACAGGGAAGCTAGGGGA
	-	CGCTCCAGAACAAAATGGCGCAGA
Primers used for generating Probes		

MusD (gag)	+	GAGTTGTTTCAGGCCAGAGGAGTAAGG
	-	GGGCAGCTCCTCTATCTGAGTG
IAP (LTR) Southern	+	CAGAAGATTCTGGTCTGTGGTGTT
	-	GAATTCATACAGTTGAATCCTTCT
IAP (pol) Northern	+	AAGCAGCAATCACCCACTTTGG
	-	CAATCATTAGATG(T/C)GGCTGCCAAG
MLV (LTR)	+	CATGTGAAAGACCCCACCTGTAG
	-	AGTCGGATGCAACTGCAAGAGGG
Major satellite	+	GACGACTTGAAAAATGACGAAATC
	-	CATATTCCAGGTCCTTCAGTGTGC
LINE1 (L1Md-A2)	+	TCCCAACATAGAGTCCTGAG
	-	TCAGTGGGCAGAGTATTCTC
Primers used for meDIP		
MusD f		CCCTTCCTTCATAACTGGTGTCGCA
MusD r		TAGCATCTCTCTGCCATTCTTCAGG
IAP f		CTCCATGTGCTCTGCCTTCC
IAP r		CCCCGTCCCTTTTTTAGGAGA
Mage f		TTGGTGGACAGGGAAGCTAGGGGA
Mage r		CGCTCCAGAACAAAATGGCGCAGA
Gapdh f		CTCTGCTCCTCCCTGTTCC
Gapdh r		TCCCTAGACCCGTACAGTGC
Sycp1 f		TGGACCAACCGTTAAATTGAG
Sycp1 r		GCGCTCCTTTATGAAGACGA
Brdt f		GCGGGTGAGTCCCATAAAG
Brdt r		CGATCACCCTTTCAGTTTGC
Interg3 f		ATGCCCCTCAGCTATCACAC
Interg3 r		GGACAGACATCTGCCAAGGT
Rho f		CTGGAGCCATGTGGAGAAGT
Rho r		GGTGGAGGCCCTTAGGTAAA
Il12a f		CAGTGTCCACGATGGAGAGA
Il12a r		ACCACACTCAGAGCGAAAGC
Lrrtm2 f		CTTCCCGCAGCTGTTAGTTC
Lrrtm2 r		TCGCAGCACATAAGCAAATC
H19_ICR f		GCATGGTCCTCAAATTCTGCA
H19_ICR r		GCATCTGAACGCCCCAATTA
Hprt f		AGCGTTTCTGAGCCATTGCT
Hprt r		GCAAAAAGCGGTCTGAGGAG
Dazl f		AATGCCCGCAGAAATAGAAA
Dazl r		TTCGGGCATTTATTTGAAGG
Spo11 f		CCAAACCAGGCAGAAATGTT
Spo11 r		ATCTCTGGGGTCGAGGTTTT
Gtl_ICR f		CTTTTGTTGACCACAACCCTTG
Gtl_ICR r		AATCCCACCACAGCTTCTTAGC
Tuba3 f		GCGCAGATAACATACGCAGA
Tuba3 r		ATGTGGCTCAAGTTGCAGTG

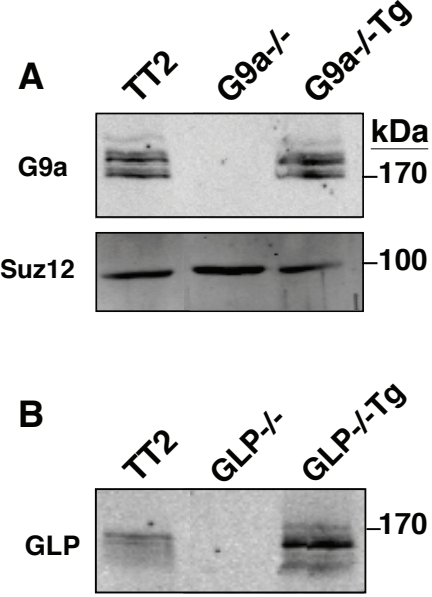


Figure S1

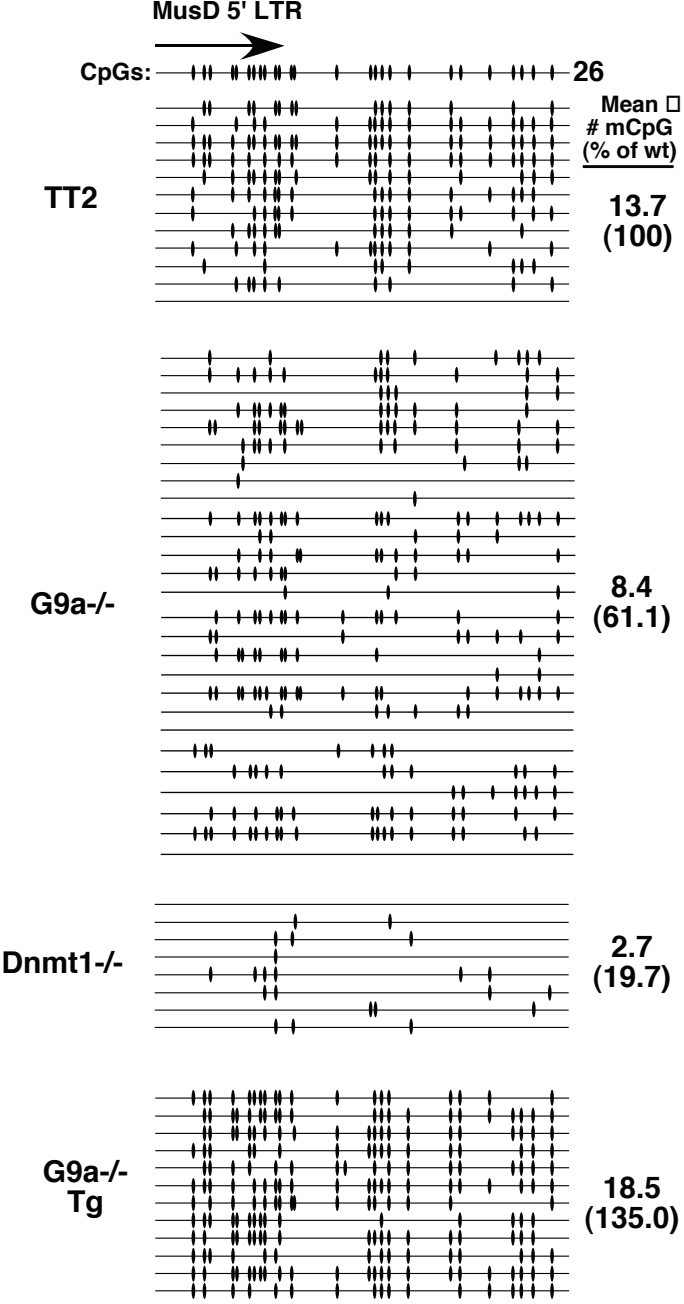
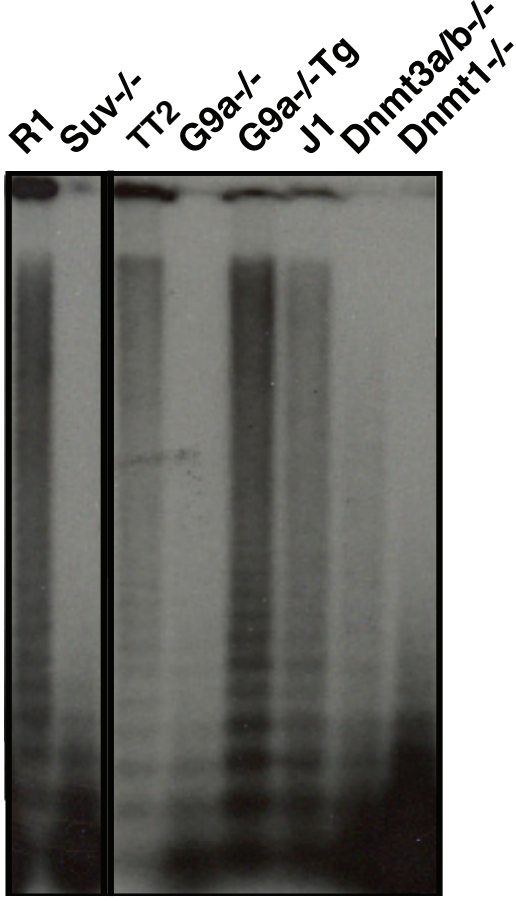
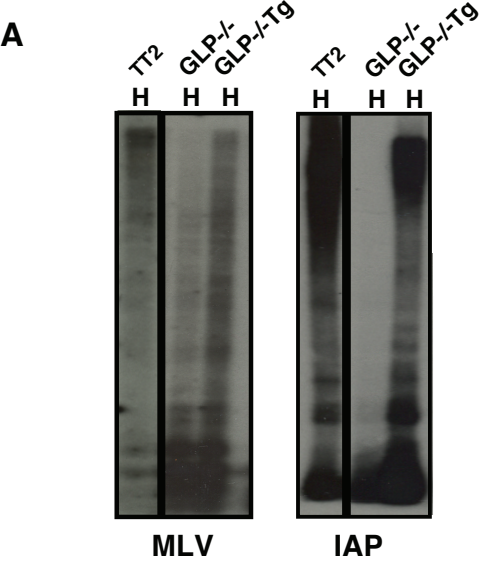

MusD 5' LTR
CpGs:
26
Mean # mCpG (% of wt)
TT2
13.7
(100)
G9a-/-
8.4
(61.1)
Dnmt1-/-
2.7
(19.7)
G9a-/-
Tg
18.5
(135.0)

Figure S2



Major Satellite

Figure S3


A
TT2
GLP-/-
GLP-/-Tg
H H H
TT2
GLP-/-
GLP-/-Tg
H H H
MLV
IAP

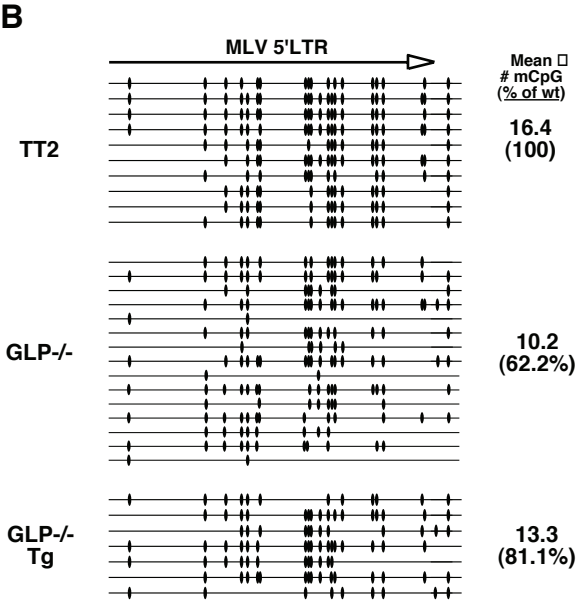

B
MLV 5'LTR
Mean # mCpG (% of wt)
TT2
16.4 (100)
GLP-/-
10.2 (62.2%)
GLP-/- Tg
13.3 (81.1%)

Figure S4

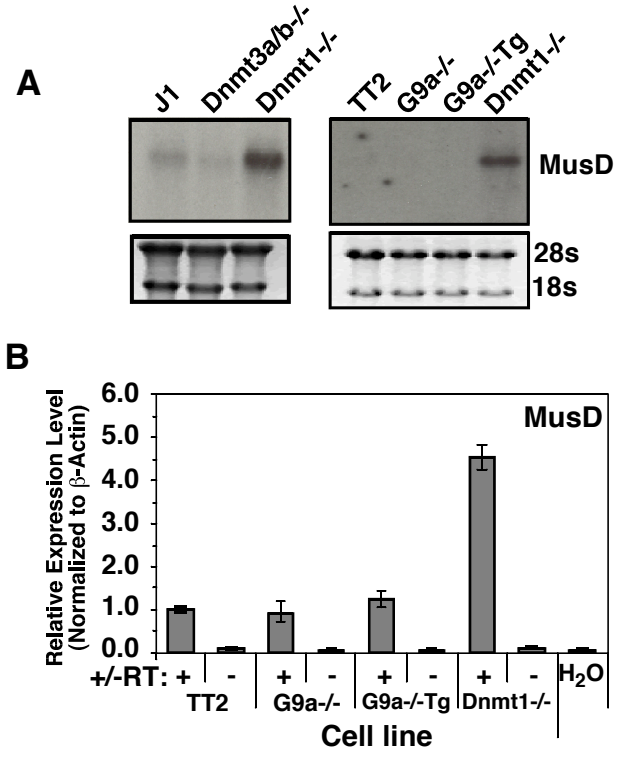

A
J1
Dnmt3a/b-/-
Dnmt1-/-
TT2
G9a-/-
G9a-/-Tg
Dnmt1-/-
MusD
28s
18s
B
Relative Expression Level
(Normalized to β-Actin)
6.0
5.0
4.0
3.0
2.0
1.0
0.0
MusD
+/-RT: + - + - + - + - H2O
TT2
G9a-/-
G9a-/-Tg
Dnmt1-/-
Cell line

Figure S5

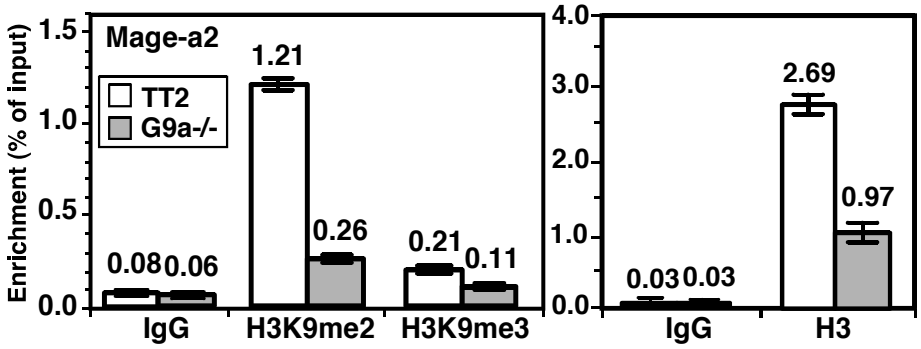

Mage-a2
TT2
G9a-/-
Enrichment (% of input)
1.5
1.0
0.5
0.0
0.08
0.06
1.21
0.26
0.21
0.11
IgG
H3K9me2
H3K9me3
4.0
3.0
2.0
1.0
0.0
0.03
0.03
2.69
0.97
IgG
H3

Figure S6

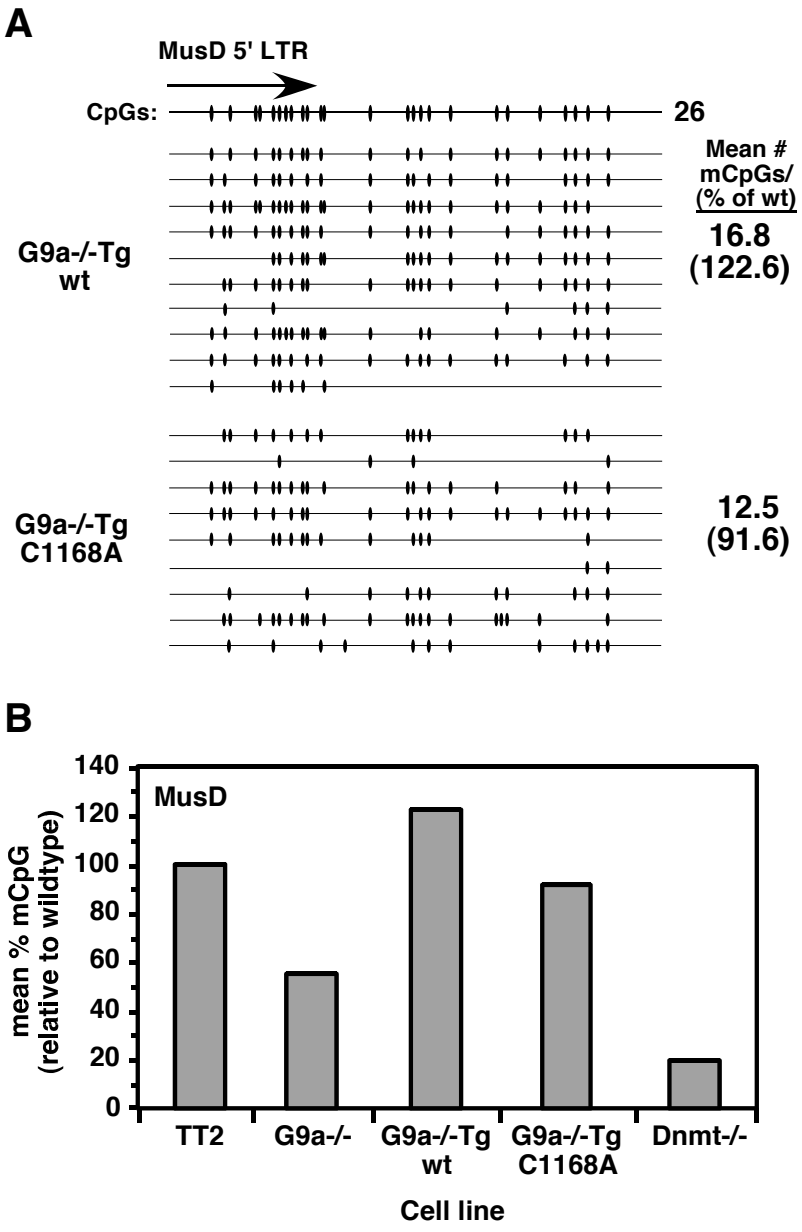

A
MusD 5' LTR
CpGs:
26
Mean # mCpGs/ (% of wt)
G9a-/-Tg wt
16.8 (122.6)
G9a-/-Tg C1168A
12.5 (91.6)
B
MusD
mean % mCpG (relative to wildtype)
140
120
100
80
60
40
20
0
TT2
G9a-/-
G9a-/-Tg wt
G9a-/-Tg C1168A
Dnmt-/-
Cell line

Figure S7

3.3. Genetics and epigenetics: stability and plasticity during cellular differentiation

Mohn F. and Schübeler D.

Summary

From an evolutionary perspective, epigenetic repressive pathways such as DNA methylation and Polycomb co-evolved in parallel to the dramatic increase in genome-size observed in vertebrates. Based on this it was previously proposed that evolution could only proceed owing to these novel inventions which might serve to lower the load of non-specific transcription in genomes that mostly consist of repetitive and transposable elements (Bestor, 1990; Bird, 1995). This model predicts partitioning of the genome into accessible genic and regulatory regions and inaccessible repetitive and non-coding regions mediated in part by epigenetic pathways. In this review we propose that epigenetic mechanisms not only act on such a global scale but also locally help to damped spurious transcription at CpG-rich promoters which mostly reside in transcriptionally permissive chromatin even when inactive. We further argue that DNA sequence composition has a strong impact on the epigenetic landscape and the target specification of repressive pathways such as Polycomb, DNA methylation and H3K9 methylation.

Vertebrate genomes are globally depleted of CpG dinucleotides, which is at large due to the mutagenic effect of DNA methylation (Weber et al., 2007). CpG islands are protected against DNA methylation in the germline and therefore kept their original CpG content (see Introduction). While most genes contain a CpG island in their promoter about 20-30% of mammalian promoters are CpG-poor (Ioshikhes and Zhang, 2000; Saxonov et al., 2006). Beyond their differential sequence composition, these two promoter classes are regulated in fundamentally distinct ways in terms of transcription initiation and also regarding epigenetic mechanisms. CpG-poor promoters essentially behave like invertebrate genes and in general appear to not require additional repression when activating signals are missing. CpG-rich promoters on the other hand always reside in transcriptionally permissive chromatin and therefore might require active repression to safe-guard this state and to prevent illegitimate transcription initiation in absence of specific signals. Remarkably, the Polycomb pathway almost exclusively targets CpG-rich sequences (Bernstein et al., 2006; Mikkelsen et al., 2007; Mohn et al., 2008; Tanay et al., 2007) for repression which is in line with our hypothesis. But considering that CpG island promoters mostly control ubiquitously expressed genes, this is not very intuitive. Nevertheless, we find that over 1000 tissue-specific genes are controlled by CpG-rich promoters. Remarkably CpG-rich promoters appear evolutionary older than CpG-poor promoters, which suggests that Polycomb targets evolutionary ancient

genes which might have been more ubiquitously expressed in invertebrates but then acquired a more specific regulation to allow further evolution of organismal complexity. This would also explain the absence of additional repressive modifications at tissue-specific CpG-poor promoters because many of these genes are specific to the vertebrate and/or mammalian lineage and evolved coinciding with or after the genome-wide depletion of CpGs.

Taken together, sequence features and epigenetic pathways might have co-evolved in order to partition the genome. The resulting accessible regulatory regions and inaccessible repetitive and non-coding regions facilitate appropriate gene regulation and prevent spurious transcriptional initiation outside of promoters on a global scale and at tissue-specific CpG-rich promoters in a locus-specific fashion.

Review

Genetics and epigenetics: stability and plasticity during cellular differentiation

Fabio Mohn and Dirk Schübeler

Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

Stem cells and multipotent progenitor cells face the challenge of balancing the stability and plasticity of their developmental states. Their self-renewal requires the maintenance of a defined gene-expression program, which must be stably adjusted towards a new fate upon differentiation. Recent data imply that epigenetic mechanisms can confer robustness to steady state gene expression but can also direct the terminal fate of lineage-restricted multipotent progenitor cells. Here, we review the latest models for how changes in chromatin and DNA methylation are regulated during cellular differentiation. We further propose that targets of epigenetic repression share common features in the sequences of their regulatory regions, thereby suggesting a co-evolution of epigenetic pathways and classes of *cis*-acting elements.

Epigenetic mechanisms and chromatin modulate gene expression

The tight control of gene expression programs at a given developmental stage is crucial to govern cell function and identity. The balance of stability versus plasticity in transcriptional programs presents an inherent regulatory challenge for developing organisms [1]. This difficulty is most obvious in mammalian embryonic stem (ES) cells, which have the potential to develop into every cell type of the adult organism. At the same time, these cells can be readily maintained in a pluripotent state *ex vivo* under defined culture conditions but can also be induced rapidly to differentiate.

The most important mediators for turning on or off expression of particular genes are DNA-sequence-specific transcription factors. Over the past few years, a large body of evidence indicated that chromatin-based regulatory mechanisms, in addition to transcription factors, could have important roles in establishing and maintaining transcriptional programs. This layer of control comprises post-translational modifications of DNA-bound histones, DNA methylation and chromatin remodeling [2,3]. All these pathways are currently referred to as being epigenetic, which, by stringent definition, involves a sequence-independent inheritance pattern during cell division in the absence of the initial trigger [4]. Currently, however, the mode of propagation is only known for DNA methylation [5], whereas several models have been proposed for post-translational histone modifications, including the involvement of positive-feedback loops [2,6,7]. Such models are compatible with the phenotypes observed in relevant knockout models and predict a self-perpetuation of modifications after the deposition of new nucleosomes, which is supported by protein interaction data [2,6,8]. However, it is inherently difficult to clearly distinguish between sequence-independent self-propagation of chromatin states and a re-establishment after cell division mediated by sequence-dependent recruitment of proteins or RNA, which in turn modify chromatin. To circumvent a discussion of the use of the term epigenetics, we will hereafter refer to a recent definition, which states that lasting chromatin changes can be termed epigenetic irrespective of proof of inheritance [9] (this nevertheless emphasizes the need for further investigation into if and how chromatin modifications can be inherited).

In eukaryotes, the packaging of DNA into nucleosomes provides a basic layer of repression because it reduces DNA access [10–12]. In a simplified view, any additional change in chromatin structure could thus further restrict access for DNA-binding factors or relieve repression and, therefore, potentially reside upstream of sequence-based regulation. These epigenetic modifications are thought to modulate DNA accessibility for transcription factors and the transcription machinery itself. Although the exact molecular mechanisms and their interplay with epigenetic modifications in controlling accessibility remain incompletely understood, it is hypothesized that they operate by directly blocking transcription-factor binding and/or by establishing higher-order chromatin structures, which would be either permissive or restrictive for transcription. The best studied example for the latter is histone H4 Lys16 (H4K16) acetylation, which can prevent higher-order structure formation, thus, indicating that H4K16 acetylation is directly involved in mediating an accessible chromatin state [13]. Conversely, it seems that Polycomb target genes localize to structures termed Polycomb bodies and adopt a higher-order confirmation, which excludes active chromatin and is thought to enhance repression [14,15].

Here, we review recent insight into epigenome changes during cellular differentiation and their potential impact on gene regulation and further developmental potential. In particular, we discuss evidence that CpG-rich and CpG-poor promoters are differentially regulated by epigenetic pathways, which is compatible with a model of genome partitioning through chromatin modifications in vertebrates.

Genome partitioning by chromatin

Over evolutionary time, genome sizes increased along with organismal complexity. For example, the

Corresponding authors: Mohn, F. (fabio.mohn@fmi.ch); Schübeler, D. (dirk@fmi.ch)

0168-9525/$ – see front matter doi:10.1016/j.tig.2008.12.005 Available online 29 January 2009

roundworm *Caenorhabditis elegans* genome ($\sim 10^8$ base pairs [bp]) is 30 times smaller than the human genome ($\sim 3 \times 10^9$ bp). However, both contain approximately the same number of genes (~20 000), showing that increased organismal complexity and genome size is not paralleled by a rise in gene number (the so-called C-value enigma [16]). In humans, and vertebrates in general, genome expansion results largely from the accumulation of repetitive and transposable elements, which is suggested to be a consequence of obligatory sexual reproduction [17]. As a result, only a small portion of vertebrate genomes encode proteins or regulatory RNA [18]. The transcriptional machinery therefore faces the challenge of locating *cis*-regulatory regions in a 'sea' of seemingly non-functional DNA sequence. One might expect that the complexity of sequence motifs that are recognized by transcription factors has similarly increased to enable specific binding to defined genomic sites. Surprisingly, however, eukaryotic transcription-factor-recognition motifs tend to be as short (6–8 bp) as those in prokaryotes and, in many cases, their binding sites are degenerate [19]. To illustrate the problem, consider that any 6-mer recognition motif occurs by chance every 4096 bp. If we assume a random sequence distribution, this would predict >781 200 binding sites in the human genome, which, for a given transcription factor, needs to be multiplied by the number of degenerate motifs it can recognize. *In vivo*, only a subset of these millions of sites is occupied, raising the question of how specificity is generated. One prominent way is cooperative binding of multiple factors [20]. In addition, the large genomes of higher eukaryotes require further structuring to direct transcription factors to appropriate targets and to reduce random binding, which in turn would dilute the pool of available factors and potentially lead to inappropriate gene regulation [21]. Epigenetic pathways that modify chromatin and DNA also coevolved along with increasing genome size. They are thus *bona fide* candidates to function in a potential partitioning of the genome into 'accessible' genic and regulatory compartments and 'inaccessible' repeat-containing regions. Indeed, previous work indicated that two major evolutionary steps, the origin of eukaryotes and the origin of vertebrates, were only possible owing to the parallel evolution of new mechanisms to control 'transcriptional noise' as an otherwise unavoidable by-product of increasing genome complexity [22]. The prokaryote-to-eukaryote transition was paralleled by the appearance of nucleosomes, which, compared with naked DNA, reduce the chance of aberrant transcription initiation [10,11]. The invertebrate-to-vertebrate step was accompanied by the advent of genome-wide DNA methylation, a modification that enables efficient transcriptional repression [23,24]. A second major repressive epigenetic pathway that coevolved in multicellular organisms along with increasing genome size and organismal complexity is mediated by Polycomb group (PcG) proteins. PcG proteins underwent marked expansion over evolutionary time [25], which is in line with the concept that increasing genome size requires additional repressive mechanisms to enhance specificity of transcription initiation and suppression of 'transcriptional noise'.

DNA methylation shapes mammalian promoter structure and stabilizes pluripotency shut-down during differentiation

DNA methylation is an efficient epigenetic repression pathway, which, in vertebrates, occurs only at cytosines in the context of CpG dinucleotides. It is catalyzed by three DNA methyltransferases (Dnmts), which are all essential [5]. *Dnmt1*- or *Dnmt3b*-deficient mouse embryos die by embryonic day 10.5 and *Dnmt3a*-deficient mice are born occasionally but suffer serious malformations and die within weeks [5]. Species that undergo widespread DNA methylation in their genome have lost CpG dinucleotides over evolutionary time. This is a direct consequence of DNA methylation because it results from increased C-to-T transitions that occur after deamination of methylated cytosines [26–28]. This loss, however, is non-uniform because certain regions are 'CpG-rich' and display the expected frequency of CpGs. They are referred to as CpG islands [29,30] and represent a large fraction of *cis*-regulatory sequence because ~60% of all mammalian gene promoters are CpG-rich [26,31,32]. In addition, several studies indicate that many non-promoter CpG islands probably serve an important regulatory function as distal regulators such as insulators and enhancers [33]. The localized depletion of CpGs results in a characteristic bimodal distribution of CpGs across vertebrate genomes (Figure 1). In invertebrates, DNA methylation is only present in some species, in which it occurs in mosaic patterns that are not genome-wide; consequently, no depletion or resulting bimodal CpG distribution is observed [34,35] (Figure 1b).

Recent genome-wide surveys revealed that DNA methylation at CpG-rich sequences is very low in stem cells [36–38]. During cellular differentiation, hypermethylation can occur at CpG island promoters and at CpG-rich sequences outside of promoter regions [37,38]. Remarkably, almost no demethylation is detected, indicating that DNA-methylation-mediated epigenetic repression increases during lineage-specification. Direct comparisons of differentiated cell types support this conclusion [28,39–41]. Notably, sequence-based detection methods such as microarrays and high-throughput sequencing, which have been employed for these studies, cannot comprehensively measure repetitive DNA, leaving the dynamics of DNA methylation at non-unique sequences an open question.

Many of the identified targets of differentiation-coupled *de novo* DNA methylation are promoters of stem-cell- and germline-specific genes [28,38,39]. One interpretation of this selectivity is that DNA methylation might stably repress the pluripotency program and prevent its aberrant reactivation and de-differentiation under physiological conditions. Experimental support for this model comes from a recent report showing that reprogramming of somatic cells into 'induced pluripotent' stem (iPS) cells is greatly enhanced upon treatment with the DNA methyltransferase inhibitor 5-aza-cytidine [42]. Collectively, the genetic and molecular data are compatible with a role for DNA methylation in the shut-down of pluripotency and, eventually, cellular specification. Nevertheless, formal proof for this model is still missing because the requirement of DNA methylation for development could reflect its

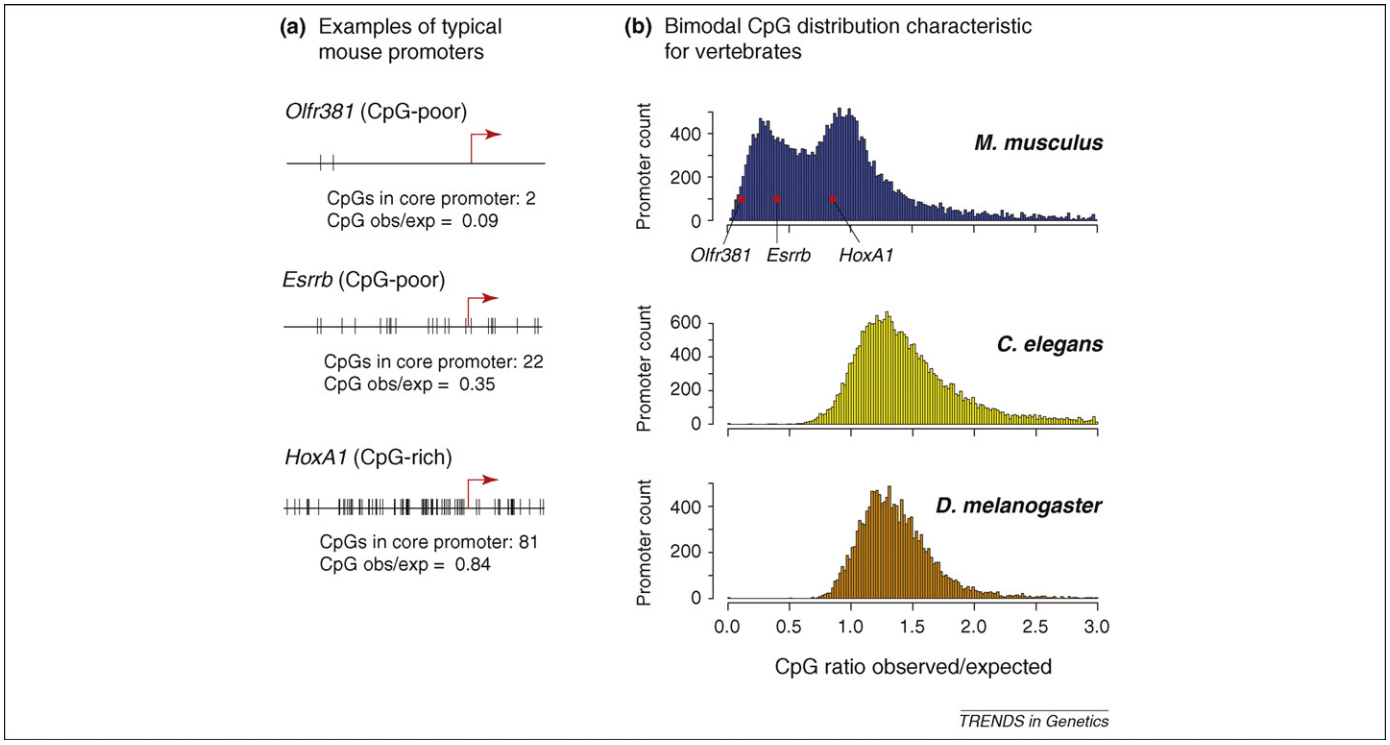


Figure 1. CpG distribution is markedly different in the genomes of vertebrates and invertebrates. Invertebrate genomes show homogenous distribution and no depletion of CpG dinucleotides. Vertebrate genomes, by contrast, are globally depleted for CpG dinucleotides, except at CpG-islands, which often mark regulatory elements such as enhancers and promoters. **(a)** Representative examples of mouse promoters with up to 40-fold different abundance of CpG dinucleotides. CpG-poor: olfactory receptor 381 (*Olfr381*) and estrogen-related receptor β (*Esrrb*); CpG-rich: *HoxA1*. In each case, a 1.3-kb window around the transcription start site (red arrow) is shown. Each vertical line represents an individual CpG dinucleotide. **(b)** CpG dinucleotide distribution at promoter sequences in a vertebrate (*Mus musculus*) and two invertebrate (*C. elegans* and *D. melanogaster*) genomes. For each genome, we calculated the CpG content of annotated transcription start sites and calculated the relative abundance as the ratio of observed CpGs versus the number of CpGs expected based on sequence composition assuming equal abundance of all bases in a 1.3-kb window. The resulting ratio is plotted as a histogram for all promoters. Vertebrates (here, mouse) show a bimodal distribution of promoter CpGs with two major peaks corresponding to CpG-poor (left peak) and CpG-rich (right peak) promoters. Invertebrates, illustrated by *D. melanogaster* and *C. elegans*, do not show such a distribution; instead, the major peak of CpG content is close to 1, indicating that invertebrates contain the expected number of CpGs in their promoters.

function in repeat inactivation [24] and maintenance of differentiated states rather than their establishment.

Profound differences in the regulation of CpG-rich versus CpG-poor promoters

In vertebrates, CpG-rich and CpG-poor promoters differ not only in sequence and DNA methylation but also in the spatial precision of transcriptional initiation and in their chromatin states. Massively parallel Cap-analysis gene expression (CAGE)-tag sequencing revealed that transcription initiates at defined nucleotide positions in CpG-poor promoters, which mostly rely on the initiation factor TATA-box-binding protein (TBP). At CpG-rich promoters, transcriptional start sites (TSS) are loosely defined and initiation can occur in a region spanning 10–60 bp [32,43]. This finding seems to contrast with the general precision of initiation reported for *in vitro* systems, however, these biochemical assays were performed almost exclusively with TBP-dependent CpG-poor promoters. The underlying difference(s) in TSS definition and regulation remains elusive, but a hint comes from genome-wide profiles of histone modifications (see later).

H3K4 methylation at CpG islands

Several recent studies show a differential distribution of active chromatin marks, such as methylation of Lys4 of histone H3 (H3K4me), at CpG-rich versus CpG-poor promoters. H3K4me creates a chromatin signal that is recognized by a large number of multiprotein complexes and which has been implicated in many gene activation pathways [44,45]. CpG-poor promoters harbor di- or tri-methylated H3K4 only when the gene is actively transcribed. This is reminiscent of the situation in invertebrates in which promoter H3K4 methylation reflects the active state [46,47]. By contrast, CpG-rich promoters behave differently: they are methylated at H3K4 constitutively and independently of the transcriptional activity of the corresponding gene [28,38,48,49]. Furthermore, a study in T cells showed that CpG-rich promoters are hyperacetylated at H3 in a transcription-independent manner [50]. Interestingly, low levels of RNA polymerase (Pol) II levels can be detected at many of these inactive CpG islands [49], raising the question of whether active histone modifications are a cause or consequence of this low and inefficient Pol II recruitment. It is conceivable that these promoters, albeit inactive, reside in an 'open', transcriptionally permissive environment, which leads to occasional Pol II binding but does not enable productive elongation [51]. This state, however, requires the absence of DNA methylation [28,36–38], implying that unmethylated CpG-rich elements are recognized by *trans*-acting factors that mediate the unique chromatin state of CpG islands (Figure 2). Such factors could include chromatin modifying enzymes (e.g. mixed-lineage leukemia [MLL] H3K4 methyltransfer-

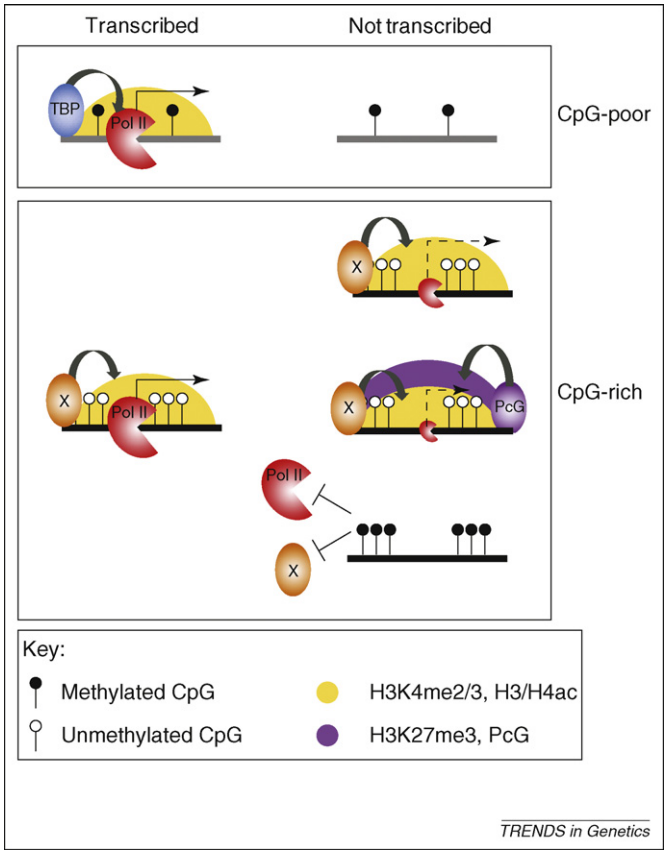


Figure 2. CpG-poor and CpG-rich promoters are differentially regulated, which is reflected in their chromatin configuration in the transcriptionally inactive state. CpG-poor promoters, which are often regulated via TBP (blue)-dependent pathways, only carry H3K4me and H3 and H4 acetylation (H3/H4ac) marks (yellow) when actively transcribed by Pol II (red). CpG-rich promoters are mostly DNA unmethylated (white lollipops), decorated by H3K4me and H3/H4ac and display low levels of Pol II even when inactive. This state could be mediated by proteins that recognize unmethylated CpG motifs, such as CXXC domain proteins (X, orange). When repressed by PcG proteins (purple), the active environment persists, indicating transient repression. Upon DNA methylation (black lollipops) of CpG-rich promoters, active histone modifications and Pol II can no longer be detected, indicating more stable silencing.

ases) containing CXXC domains, which bind preferentially to unmethylated CpGs [52]. Moreover, this 'active' chromatin environment at CpG-rich sequences could protect regulatory elements against DNA methylation and a resulting loss of 'accessibility'. Direct evidence for H3K4 methylation protection against DNA methylation comes from a structural analysis of Dnmt3L, a germline-specific co-factor essential for *de novo* methylation of imprinting control regions [53,54]. Dnmt3L in complex with the *de novo* methyltransferase Dnmt3a can only bind nucleosomes that are unmodified at H3K4, whereas H3K4 methylation blocks this interaction and prevents DNA methylation. Hence, it is tempting to speculate that an H3K4-dependent pathway also operates during stem-cell-differentiation-coupled *de novo* methylation. By definition, *de novo* methylation can only occur at unmethylated sequences, which, at large, are the CpG-rich sequences in the genome [21,28,36–38,40]. The fact that H3K4 methylation and DNA methylation are mutually exclusive [28] predicts that *de novo* methylation of CpG-rich sequences coincides with a loss of H3K4 methylation, which is indeed the case during stem-cell differentiation [37,38].

Only a small fraction of CpG-rich sequences is *de novo* DNA methylated during cellular differentiation, hence, the majority of CpG-rich regulatory regions in mammalian genomes are accessible at any developmental stage. By contrast, the non-regulatory CpG-poor part of the genome would reside in a less accessible chromatin environment, which in turn might reduce the binding of transcription factors to randomly occurring sites.

In support of the concept that DNA sequence features are sufficient to establish an accessible chromatin state, a recent study in mosaic mice carrying human DNA indicated that homologous transcription factors can establish tissue-specific transcription and H3K4 methylation in closely related species [55].

Lessons from Polycomb

PcG-mediated gene repression regulates gene transcription during development. Originally discovered in *Drosophila melanogaster* as a system that controls *Hox* gene expression for correct body patterning [56,57], Polycomb has a broad regulatory potential in mouse and human ES cells because it targets many developmental transcription factors [58,59]. PcG-mediated repression entails histone H3 Lys27 trimethylation (H3K27me3), which is set by the Polycomb repressive complex 2 (PRC2) [56,57]. Polycomb repression can be overcome upon gene activation by specific stimuli once pluripotent cells are induced to differentiate, whereas non-induced Polycomb targets maintain H3K27me3 and PRC2 occupancy [58]. These data led to the model that Polycomb targets are specified early in development and continue to be repressed unless activated. However, this view has been challenged by a series of recent studies that reveal additional, non-stem-cell-specified Polycomb targets in various primary and transformed mammalian cell types [38,60–66]. For example, in neuronal progenitors, genes required for further developmental fates comprise differentiation-specific PcG targets [38]. Together, these findings support a model in which Polycomb repression could act not only in pluripotent stem cells to ensure proper lineage choice, but also in progenitor cells to guide their further developmental potential by ensuring proper regulation of subtype-specific genes. These findings further exemplify that stem cells are not necessarily unique with regard to the epigenetic mechanisms they employ but, rather, are unique in the genomic targets of these pathways. Cell-type-specific recruitment is probably mediated by transcription factors. For example, many targets of Oct4, a key transcription factor in the control of stem-cell pluripotency, carry the repressive H3K27me3 mark, pointing to a mechanism by which PcG proteins can be recruited to cell-type-specific targets [58]. Moreover, in cancer cells, PcG proteins are recruited by the transcription factor Snail1 to silence E-cadherin expression, which often correlates with increasing invasiveness and malignancy of the cancer owing to detachment from the extracellular matrix upon E-cadherin loss [67].

Several recent reports showed that H3K27me3 in mammalian cells is largely confined to CpG-rich sequences [33,38,62,68]. This is unexpected because the majority of tissue-specific genes, the *bona fide* targets of Polycomb

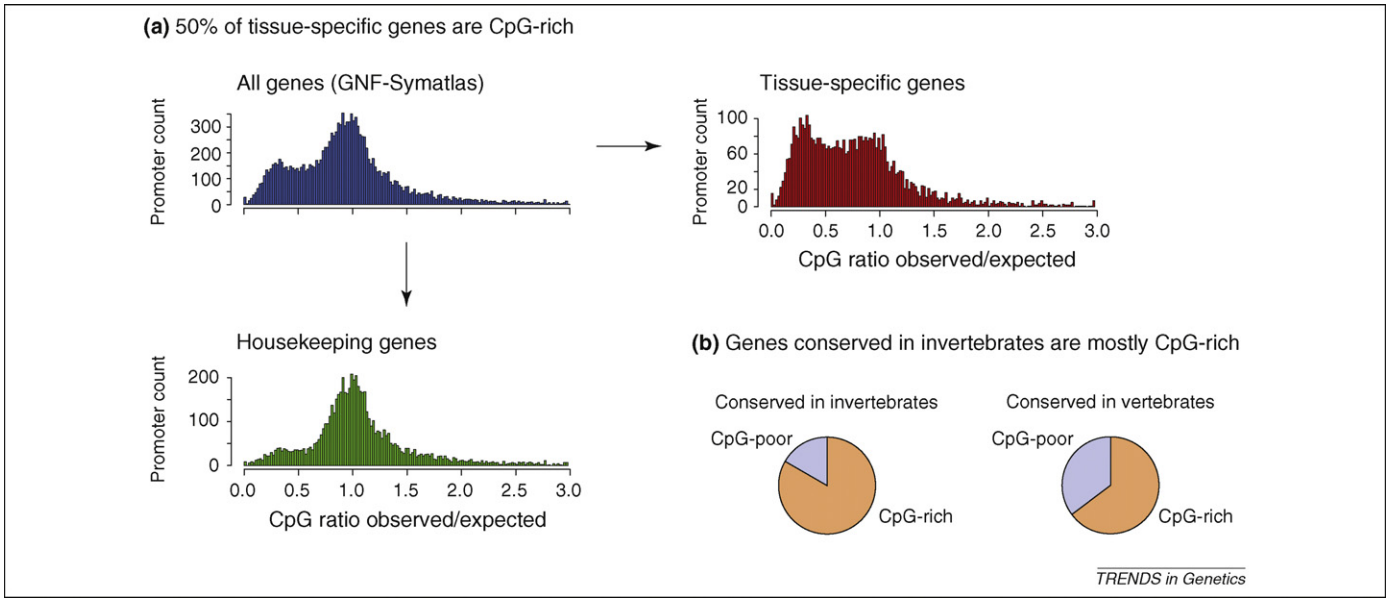


Figure 3. Genes can be grouped as housekeeping or tissue-specific based on a broad or restricted activity in different tissues. Most housekeeping genes are under the control of CpG island promoters. Nevertheless, many CpG island promoters control tissue-specific genes, which tend to regulate ancestral genes. **(a)** Gene expression data for 61 mouse tissues were retrieved from Symatlas (http://symatlas.gnf.org) [87] and CpG content was determined for all promoters in the dataset (n = 13 729) as described in Figure 1. A histogram is shown of the CpG ratios for all genes, including both tissue-specific and broadly expressed housekeeping genes. All genes display a similar bimodal distribution (top left) as shown in Figure 1 [69]. As expected, ubiquitously expressed 'housekeeping' genes (expression value >200 in >50 tissues out of 61; n = 5481) are mostly CpG-rich. However, tissue-specific genes (expression value >200 in <11 tissues out of 61; n = 5294) show almost equal numbers of CpG-rich and -poor promoters. These CpG-rich tissue-specific genes are preferably regulated by DNA methylation and Polycomb, whereas CpG-poor promoters seem not to require active repression. **(b)** Interestingly, 'old' genes (i.e. conserved from invertebrates) are mostly CpG-rich, whereas 'newer' genes, which arose in the vertebrate lineage, are more often controlled by CpG-poor promoters.

regulation, are under the control of CpG-poor promoters [69] (Figure 3). This surprising preference towards CpG-rich elements indicates a sequence contribution to the yet-to-be-determined targeting mechanism of the Polycomb machinery (see later). It furthermore provides a DNA-sequence-based explanation for the existence of so-called 'bivalent' chromatin domains, which harbor both H3K4 methylation and H3K27 methylation. As outlined earlier, CpG-rich promoters are ubiquitously marked by H3K4me. Thus, any CpG-rich sequence targeted by Polycomb will consequently form a bivalent domain carrying both the 'activating' H3K4 and the 'repressive' H3K27 methylation mark [68]. This model is compatible with the finding that bivalent domains are not a unique feature of pluripotent cells but, instead, are also present in differentiated cell types and can even form *de novo* during cellular differentiation [38,61–64]. In agreement with a CpG island explanation for bivalent chromatin, both marks are mutually exclusive in *D. melanogaster* [70], the genome of which does not contain global DNA methylation or CpG islands.

It remains to be tested if some aspects of the regulatory functions of Polycomb are unique to mammalian (and probably vertebrate) promoters, which display bivalency. One possibility is that, in addition to the function in invertebrates, Polycomb repression could be used to reduce partially intrinsic transcriptional noise at selected CpG island promoters because these reside in an 'accessible' chromatin state with detectable levels of Pol II even when inactive (Figure 2 and Box 1).

Control of CpG-poor promoters

In vertebrates, CpG-poor promoters are largely DNA methylated independently of activity state. When they are not active, they are neither H3K4 methylated nor bound by Pol II [38,49,62]. Thus, they are reminiscent of inactive promoters in lower eukaryotes such as yeast, in which nucleosomal packaging and the resulting decreased sequence accessibility seem sufficient to mediate a stable off-state [10]. Similarly to the general situation in yeast, and unlike CpG-rich promoters, CpG-poor promoters contain very precise start sites that are set by sequence-specific activators (see earlier). Interestingly, Orford and colleagues reported a set of CpG-poor promoters in hematopoietic progenitor cells (erythroid-myeloid-progenitors [EML]), which show H3K4 dimethylation (H3K4me2) at lineage-specific targets in the absence of transcription [71]. These EML-specific H3K4me2-marked promoters control several hematopoietic genes, which are expressed in either erythroid or myeloid lineages. Upon stimulation of EML cells towards erythroid differentiation, H3K4me2 positive erythroid genes are activated and retain H3K4me2, whereas myeloid-specific genes remain silent but lose the H3K4me2 mark. Hence, in this system, H3K4 methylation seems to mark lineage-specific genes in progenitor cells for later activation. In the absence of transcription, H3K4 methylation might lead to a more accessible chromatin environment at these CpG-poor promoters and it remains to be tested whether they are also occupied by Pol II or Polycomb, similarly to the situation at CpG-rich promoters.

Despite this example, 'open' chromatin is usually absent at inactive CpG-poor promoters. This finding might explain why they tend to rely less frequently on additional epigenetic repression, illustrated by the marked bias of Polycomb binding for CpG-rich promoters [33,68]. At the same time, H3K9 methylation and heterochromatin

Box 1. Is lineage choice driven by transcriptional oscillations from CpG-rich promoters?

Most evidence in mammalian systems and derived models indicates that tight control of gene 'on' and 'off' states defines cell identity. In many prokaryotic systems, however, stochastic gene expression and oscillations of transcription levels seem to function as triggers or switches in decision making [81]. Such events also occur in mammalian cells, for example, in the stochastic expression of odorant receptors in the mouse sensory neurons [82]. A recent report showed that clonal populations of mouse hematopoietic progenitor cells sorted for either high or low expression of the stem-cell marker Sca-1 can reconstitute the full parental spectrum of Sca-1 expression within a few days after isolation [83]. Notably, not only is Sca-1 expression different, but the entire expression program is markedly altered between the sorted populations and reverts back to the average of the unsorted pool of hematopoietic progenitors within a few days. Cells expressing low Sca-1 levels are prone to erythroid differentiation, whereas cells expressing high Sca-1 levels are more prone to differentiate into a myeloid lineage, indicating that these apparently random fluctuations have functional implications [83]. Interestingly, recent data showed that ES cells are also more heterogeneous than previously thought and seem to oscillate between different states. For example, Nanog, a transcription factor important for maintenance of pluripotency, and Stella, a protein specifically expressed in pre-implantation embryos and the germ line, show fluctuating expression in ES cells, which has an impact on the differentiation potential of the cells [84,85]. On a related note, caudal type homeobox 2 (Cdx2) and Eomes, two key transcription factors of the trophoblast lineage, show stochastic expression in the early blastocyst, which eventually is sufficient to promote expression of the Elf5 transcription factor and induction of trophoblast differentiation [86]. However, in the inner cell mass (ICM) and in ES cells, the *Elf5* promoter is DNA methylated and the sporadic activity of Cdx2 and Eomes is not sufficient to activate *Elf5*. Consequently, the ICM remains pluripotent and will give rise to all cells of the embryo but will not contribute to trophectoderm [86]. These examples illustrate how stochastic gene expression, in combination with epigenetic pathways, can contribute to cell-fate decision switches. Moreover, it is tempting to speculate that these random fluctuations are facilitated by promoter structure and, thus, mostly occur at CpG-rich promoters, which seem to be less stringently controlled than CpG-poor promoters. Along the same line, it will be interesting to determine if other genes show similar behavior and how this relates to decision making *in vivo*.

protein 1 (HP1), which binds H3K9 methylated nucleosomes, seem to confer repression to some CpG-poor promoters [62,72]. These H3K9 targets are often members of large gene families such as olfactory receptors and Krüppel-associated box (KRAB)-zinc finger genes, which are clustered in the genome. Their repetitive nature might explain the recruitment of H3K9 and HP1, which have a crucial role in silencing centromeric and intergenic repeats at constitutive heterochromatin [8]. Remarkably, H3K9 methylation does not occur at CpG-rich promoters [62], indicating that H3K9 and H3K27 methylation function at separate targets that differ in sequence composition.

Promoter structure and epigenetic regulation

Transient and more stable forms of epigenetic repression are targeted to particular promoter classes based upon CpG content. The observed bias of tissue-specific DNA methylation and Polycomb repression toward CpG islands seems surprising because CpG islands are thought to mostly regulate housekeeping genes [43,69]. This, however, is an oversimplification. Several thousand CpG-rich promoters control genes that are expressed in a tissue-specific manner (Figure 3). These promoters reside in accessible chromatin, as indicated by H3K4 methylation, and low, but detectable, Pol II levels (Figure 2). Their chromatin state might reflect a form of genome partitioning that maintains regulatory regions accessible and free of DNA methylation. As a consequence, this open structure might require active repression via chromatin and DNA methylation to stabilize gene 'off' states. The transcriptionally permissive environment at CpG-rich promoters could furthermore result in stochastic transcription from inactive or oscillatory expression at active promoters, which has been implicated in contributing to cell fate decisions (Box 1). Moreover, recent studies found divergent short sense and antisense transcripts at the transcription start sites of active genes, which might reflect less stringent control of transcription initiation owing to the accessible environment at CpG-rich promoters [73–76].

How did these remarkably different promoter classes evolve? To maintain high concentrations of CpG over evolutionary time, CpG islands must be unmethylated in the germline. At first glance, tissue-specific genes controlled by CpG islands tend to have arisen in invertebrates before the advent of global DNA methylation. For example, nearly all *Hox* genes are under the control of CpG island promoters. Conversely, vertebrate-specific classes of genes, such as those encoding immunoglobulins or olfactory receptors, tend to be under the control of CpG-poor promoters (Figure 3). It thus seems plausible that promoter classes primarily reflect the evolutionary history of the genes they control rather than their precise biological function. Further work is required to better understand the evolution of *cis*-regulatory regions; however, CpG-rich and CpG-poor promoters seem to be regulated differently, not only at the level of DNA methylation but also in their selective use of the Polycomb system.

Concluding remarks and future perspectives

Stem-cell-based differentiation models have proven informative for studying the cellular changes that accompany the loss of pluripotency and terminal differentiation. Because these systems can be genetically modified and generate pure populations of primary, non-transformed cells with defined developmental potential and function, they will continue to provide important information. Indeed, in this context, the application of modern genomics and proteomics tools can be used to identify regulatory principles that can be tested *in vivo*. Such approaches will also guide our understanding of epigenetic gene regulation; however, limitations of these systems should not be ignored. Rapidly dividing stem cells are often compared with post-mitotic differentiated cells bearing the risk that 'stem-cell-specific' characteristics are actually general features of mitotically dividing cells. In addition, current *in vitro* stem-cell systems do not enable the study of asymmetric cell division, a key feature of stem-cell maintenance *in vivo*, nor the testing of potential asymmetric segregation of epigenetic information on chromatin.

Equally importantly, most current assays that map sites of epigenetic modifications, such as chromatin immunoprecipitation (ChIP) coupled to microarrays or massive parallel sequencing, cannot account for heterogeneity

within pools of cells. New developments that enable ChIP to be performed on few cells [77] or high-throughput sequencing of bisulfite-converted DNA [37,78] will increase resolution and enable the identification of differences in populations of cells as in rare primary cells, for example, in the early embryo. Another important question for the future concerns the role of non-coding RNAs in differentiation processes; indeed, these RNAs can target sites of chromatin changes in several organisms, including vertebrates, and help to regulate DNA methylation in plants [79].

In this review, we have highlighted the interplay between promoter sequence features and chromatin and DNA modifications. A logical next step will be to identify the exact contribution of a given DNA sequence motif to chromatin-regulatory processes. Such information might provide a key to understanding targeting pathways and the propagation of epigenetic states and potentially link them to extracellular signaling pathways that are crucial for stem-cell function *in vivo* [80].

Acknowledgements

We thank Marc Bühler, Antoine Peters, Susan Gasser and members of the Schübeler laboratory for critical reading of the manuscript and helpful comments. Furthermore, we apologize to colleagues whose work could not be cited owing to space limitations. Research in the laboratory of D.S. is supported by the Novartis Research Foundation and the European Union (LSHG-CT-2004–503433 and LSHG-CT-2006–037415). F.M. is supported by a pre-doctoral fellowship from the Boehringer Ingelheim Fonds.

References

1 Reik, W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447, 425–432
2 Bernstein, B.E. *et al.* (2007) The mammalian epigenome. *Cell* 128, 669–681
3 Cairns, B.R. (2005) Chromatin remodeling complexes: strength in diversity, precision through specialization. *Curr. Opin. Genet. Dev.* 15, 185–190
4 Ptashne, M. (2007) On the use of the word 'epigenetic'. *Curr. Biol.* 17, R233–R236
5 Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514
6 Groth, A. *et al.* (2007) Chromatin challenges during DNA replication and repair. *Cell* 128, 721–733
7 Hansen, K.H. *et al.* (2008) A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* 10, 1291–1300
8 Peters, A.H. and Schubeler, D. (2005) Methylation of histones: playing memory with DNA. *Curr. Opin. Cell Biol.* 17, 230–238
9 Bird, A. (2007) Perceptions of epigenetics. *Nature* 447, 396–398
10 Struhl, K. (1999) Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98, 1–4
11 Workman, J.L. and Kingston, R.E. (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* 67, 545–579
12 Lam, F.H. *et al.* (2008) Chromatin decouples promoter threshold from dynamic range. *Nature* 453, 246–250
13 Shogren-Knaak, M. *et al.* (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844–847
14 Lanzuolo, C. *et al.* (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* 9, 1167–1174
15 Terranova, R. *et al.* (2008) Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell* 15, 668–679
16 Gregory, T.R. (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev. Camb. Philos. Soc.* 76, 65–101
17 Bestor, T.H. (2003) Cytosine methylation mediates sexual conflict. *Trends Genet.* 19, 185–190
18 Waterston, R.H. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562
19 Wray, G.A. *et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419
20 Bilu, Y. and Barkai, N. (2005) The design of transcription-factor binding sites is affected by combinatorial regulation. *Genome Biol.* 6, R103
21 Rollins, R.A. *et al.* (2006) Large-scale structure of genomic methylation patterns. *Genome Res.* 16, 157–163
22 Bird, A.P. (1995) Gene number, noise reduction and biological complexity. *Trends Genet.* 11, 94–100
23 Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21
24 Walsh, C.P. *et al.* (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* 20, 116–117
25 Whitcomb, S.J. *et al.* (2007) Polycomb Group proteins: an evolutionary perspective. *Trends Genet.* 23, 494–502
26 Antequera, F. and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. U. S. A.* 90, 11995–11999
27 Bestor, T.H. and Coxon, A. (1993) Cytosine methylation: the pros and cons of DNA methylation. *Curr. Biol.* 3, 384–386
28 Weber, M. *et al.* (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* 39, 457–466
29 Gardiner-Garden, M. and Frommer, M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.* 196, 261–282
30 Takai, D. and Jones, P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3740–3745
31 Ioshikhes, I.P. and Zhang, M.Q. (2000) Large-scale human promoter mapping using CpG islands. *Nat. Genet.* 26, 61–63
32 Bajic, V.B. *et al.* (2006) Mice and men: their promoter properties. *PLoS Genet.* 2, e54
33 Tanay, A. *et al.* (2007) Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5521–5526
34 Glass, J.L. *et al.* (2007) CG dinucleotide clustering is a species-specific property of the genome. *Nucleic Acids Res.* 35, 6798–6807
35 Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Natl. Rev.* 9, 465–476
36 Fouse, S.D. *et al.* (2008) Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* 2, 160–169
37 Meissner, A. *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770
38 Mohn, F. *et al.* (2008) Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* 30, 755–766
39 Farthing, C.R. *et al.* (2008) Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* 4, e1000116
40 Illingworth, R. *et al.* (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* 6, e22
41 Eckhardt, F. *et al.* (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 38, 1378–1385
42 Mikkelsen, T.S. *et al.* (2008) Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454, 49–55
43 Carninci, P. *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38, 626–635
44 Sims, R.J., 3rd and Reinberg, D. (2008) Is there a code embedded in proteins that is based on post-translational modifications? *Nat. Rev. Mol. Cell Biol.* 9, 815–820
45 Taverna, S.D. *et al.* (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* 14, 1025–1040
46 Pokholok, D.K. *et al.* (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122, 517–527
47 Schubeler, D. *et al.* (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* 18, 1263–1271

48 Barrera, L.O. *et al.* (2008) Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res.* 18, 46–59
49 Guenther, M.G. *et al.* (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77–88
50 Roh, T.Y. *et al.* (2005) Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* 19, 542–552
51 Lorincz, M.C. and Schubeler, D. (2007) RNA polymerase II: just stopping by. *Cell* 130, 16–18
52 Voo, K.S. *et al.* (2000) Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.* 20, 2108–2121
53 Jia, D. *et al.* (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. *Nature* 449, 248–251
54 Ooi, S.K. *et al.* (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* 448, 714–717
55 Wilson, M.D. *et al.* (2008) Species-specific transcription in mice carrying human chromosome 21. *Science* 322, 434–438
56 Schwartz, Y.B. and Pirrotta, V. (2007) Polycomb silencing mechanisms and the management of genomic programmes. *Natl. Rev.* 8, 9–22
57 Schuettengruber, B. *et al.* (2007) Genome regulation by polycomb and trithorax proteins. *Cell* 128, 735–745
58 Boyer, L.A. *et al.* (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353
59 Lee, T.I. *et al.* (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301–313
60 Barski, A. *et al.* (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837
61 Bracken, A.P. *et al.* (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* 20, 1123–1136
62 Mikkelsen, T.S. *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560
63 Pan, G. *et al.* (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1, 299–312
64 Pasini, D. *et al.* (2007) The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* 27, 3769–3779
65 Roh, T.Y. *et al.* (2006) The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15782–15787
66 Squazzo, S.L. *et al.* (2006) Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome Res.* 16, 890–900
67 Herranz, N. *et al.* (2008) Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. *Mol. Cell. Biol.* 28, 4772–4781
68 Bernstein, B.E. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326
69 Schug, J. *et al.* (2005) Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* 6, R33
70 Tolhuis, B. *et al.* (2006) Genome-wide profiling of PRC1 and PRC2 polycomb chromatin binding in *Drosophila melanogaster*. *Nat. Genet.* 38, 694–699
71 Orford, K. *et al.* (2008) Differential H3K4 methylation identifies developmentally poised hematopoietic genes. *Dev. Cell* 14, 798–809
72 Vogel, M.J. *et al.* (2006) Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res.* 16, 1493–1504
73 Core, L.J. *et al.* (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848
74 He, Y. *et al.* (2008) The antisense transcriptomes of human cells. *Science* 322, 1855–1857
75 Preker, P. *et al.* (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851–1854
76 Seila, A.C. *et al.* (2008) Divergent transcription from active promoters. *Science* 322, 1849–1851
77 O'Neill, L.P. *et al.* (2006) Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nat. Genet.* 38, 835–841
78 Cokus, S.J. *et al.* (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219
79 Zaratiegui, M. *et al.* (2007) Noncoding RNAs and gene silencing. *Cell* 128, 763–776
80 Moore, K.A. and Lemischka, I.R. (2006) Stem cells and their niches. *Science* 311, 1880–1885
81 Raj, A. and van Oudenaarden, A. (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216–226
82 Vassar, R. *et al.* (1993) Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell* 74, 309–318
83 Chang, H.H. *et al.* (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453, 544–547
84 Chambers, I. *et al.* (2007) Nanog safeguards pluripotency and mediates germline development. *Nature* 450, 1230–1234
85 Hayashi, K. *et al.* (2008) Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* 3, 391–401
86 Ng, R.K. *et al.* (2008) Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat. Cell Biol.* 10, 1280–1290
87 Su, A.I. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6062–6067

4. General discussion and outlook

Several pathways exist that are involved in chromatin-based gene repression. During my PhD studies I have investigated the genomic distribution and developmental plasticity of DNA methylation and Polycomb-mediated repression as two prominent modes of gene silencing. My results together with recent data from other groups suggest a stable mode of repression mediated by DNA methylation and/or H3K9 methylation and HP1, which is responsible for silencing repetitive elements and a number of gene promoters. This repression is set up early in development and appears incompatible with active chromatin and high levels of transcription. Polycomb mediated repression on the other hand appears to be more dynamic, cell state specific and can be overcome by specific activation signals. This would suggest a more transient mode of repression, which is dedicated to safe-guard "noisy" lineage specific CpG island promoters residing in transcriptionally permissive chromatin.

Below I will discuss the main findings of my PhD project and the impact of the underlying DNA sequence on observed epigenetic targeting and reprogramming.

4.1. Targeting of DNA methylation and its relationship to H3K4 methylation

The question how CpG islands remain unmethylated is fundamental for our understanding of the preservation of CpG islands over evolutionary time. Various evidence suggest that H3K4 methylation protects CpG islands form DNA methylation and thus from loss of CpGs due to DNA methylation dependent mutations. Hence the mechanism by which CpG-rich sequences are recognized and protected is likely to provide insight into the targeting of DNA methylation. Further it might also shed light on the shaping of genome sequence via interplay of antagonistic genetic and epigenetic pathways over evolutionary time.

We and others have shown that CpG-rich sequences are generally marked by H3K4me2/3 independent of cellular state and transcription (Guenther et al., 2007; Mikkelsen et al., 2007; Mohn et al., 2008; Weber et al., 2007). Most of these CpG-rich sequences are not DNA methylated in somatic cells, representing the only regions in vertebrate genomes which are not DNA methylated by default. Further, in order to maintain their CpG content, these regions have to be unmethylated in the germline (Antequera and Bird, 1993). Notably those CpG islands which are DNA methylated in somatic cells are devoid of H3K4 methylation (Meissner et al., 2008; Mohn et al., 2008; Weber et al., 2007) and our results demonstrate that H3K4 methylation is lost coinciding with *de novo* DNA methylation during cellular differentiation. This mutual exclusivity of DNA methylation and H3K4 methylation is compatible with a model that H3K4 methylation could protect CpG islands from becoming DNA methylated. If true, the H3K4 methylation mark would have to be removed by sequence-specific recruitment of H3K4 KDMs. Consequently, DNA methylation would not

involve sequence-specific recruitment of DNMTs but rather occur due to loss of protection from DNA methylation, which appears to be the default state for the rest of the genome. Alternatively, H3K4 KDMs and DNMTs could be recruited together as part of a "*de novo* DNA methylation complex". In a third scenario, DNMTs could be recruited in a sequence-specific fashion and confer *de novo* methylation of an H3K4 methylated locus. The subsequent loss of H3K4 methylation could be passive via cell division, if the DNA methylation interferes with further activity of H3K4 KMTs. In our analysis we find that H3K4 methylation levels are dramatically reduced at the neuronal progenitor state when *de novo* DNA methylation has occurred. Interestingly, K4 levels at *de novo* methylated promoters are even further reduced during terminal differentiation (Mohn et al, 2008, Fig. 3A), which would support the passive dilution model upon targeted *de novo* DNA methylation. Such local removal of active histone modifications and putative compaction of chromatin around methylated CpG islands is further supported by biochemical data suggesting association of MBDs with co-repressor complexes containing HDACs and KMTs [reviewed in (Clouaire and Stancheva, 2008)]. In line with a protective function for H3K4me, recent reports showed that the DNMT3L/DNMT3A complex only recognizes nucleosomes with unmethylated H3K4 (Jia et al., 2007; Ooi et al., 2007). This would support a model entailing active removal of H3K4 methylation prior to *de novo* DNA methylation. However a similar DNMT3L-mediated targeting process of DNMTs during somatic differentiation appears unlikely because *DNMT3L* expression is down-regulated early in embryonic development and its promoter gets DNA methylated upon induction of differentiation. Moreover, *DNMT3L* mutant mice develop normally and do not recapitulate the dramatic phenotypes of DNMT mutants (Bourc'his et al., 2001).

Of note, ES cells lacking both *de novo* methyltransferases *DNMT3A* and *DNMT3B* differentiate into post-mitotic neurons indistinguishable form corresponding wild-type cells, while *DNMT1* knock-out cells fail to complete differentiation (Mohn and Schübeler, unpublished data), which is in agreement with previous findings using less defined differentiation models (Jackson et al., 2004; Okano et al., 1999). While the DNA methylation status of the *DNMT3* mutant cells need to be investigated it can be hypothesized that *de novo* methylation of CpG-rich sequences during cellular differentiation is not essential for the differentiation process itself. Consequently the lethality caused by *DNMT* mutants could reflect the essential function of DNA methylation in repeat silencing, regulation of imprinting and genome stability.

4.2. Pol II and H3K4 methylation: Who is first?

Inactive CpG islands are always H3K4 methylated as reported by us and others (Guenther et al., 2007; Mikkelsen et al., 2007; Mohn et al., 2008; Weber et al., 2007). Many of them bear low but detectable levels of Pol II without detectable full-length mRNA

production (Guenther et al., 2007). This raises the question whether H3K4me is deposited with or without the involvement of Pol II. Either H3K4 methylation is deposited independently of Pol II, maybe via CxxC-domain containing H3K4 KMTs such as MLL or transcription factors which specifically recognize unmethylated CpG-rich sequences, and create a "transcriptionally permissive" chromatin environment. This in turn could attract transcription initiation factors and Pol II resulting in short non-productive visits of Pol II at such promoters, which lack the necessary activation signals to promote successful transcription initiation. Alternatively, occasional transcription initiation might occur at CpG-rich sequences due to unspecific binding of transcription factors. Due to missing elongation factors and improper activation this would again lead leading to non-productive recruitment of Pol II and by that conferring H3K4 methylation to these regions in a Pol II dependent pathway.

4.3. How does Polycomb mediate repression?

In contrast to DNA methylation we find that Polycomb-mediated H3K27me3 is very dynamic during cellular differentiation. The numbers of novel and resolved targets are roughly equal during both lineage commitment, when stem cells lose pluripotency, and during terminal neuronal differentiation, when progenitor cells mature into functional post-mitotic neurons. Together with recent findings reporting novel Polycomb targets in somatic cells and cancer cell lines (Azuara et al., 2006; Bracken et al., 2006; Mikkelsen et al., 2007; Pan et al., 2007; Squazzo et al., 2006) this suggests that Polycomb not only has a role in early development, but also during cellular differentiation and in terminally differentiated cells. These findings however do not address the mode of action and the events taking place at the molecular level. Immunofluorescence and chromosome looping data from flies indicates that Polycomb repressed genes would coordinately localize to so called Polycomb bodies which are devoid of Pol II and active histone modifications (Lanzuolo et al., 2007). In mammals such accumulations of Polycomb targets have not been reported. Nonetheless, recent evidence from mouse indicates that Polycomb could create a local repressive environment excluding Pol II and leading to locus contraction or higher order chromatin structure changes (Terranova et al., 2008; Tiwari et al., 2008). These structural changes could contribute to repression via rendering a locus less accessible to the transcriptional machinery. Along the same line recent electron microscopy data and *in vitro* experiments indicate that a PRC2 complex containing Ezh1 could be responsible for locus contraction at sites of H3K27me3 modification (Margueron et al., 2008). These data would in principle support the long standing model of epigenetic restriction. It predicts that repressive modifications such as H3K27me3, H3K9me and DNA methylation would lead to a decrease of DNA accessibility and thereby reduce the likelihood of transcription factor binding and transcriptional activation in the absence of strong and specific signals. The mutual exclusivity of DNA methylation and

H3K9 methylation with active marks such as H3 and H4 acetylation and H3K4 methylation further supports this. Hhowever in case of Polycomb-mediated repression in mammals, this does not hold true (see below).

4.4. Polycomb is strongly biased towards CpG-rich sequences

We and others have shown that Polycomb targeting is strongly biased towards CpG-rich sequences. These are by default in a "transcriptionally permissive" environment characterized by H3K4 methylation even when inactive (Azuara et al., 2006; Bernstein et al., 2006; Bracken et al., 2006; Mikkelsen et al., 2007; Mohn et al., 2008; Pan et al., 2007; Squazzo et al., 2006). Therefore most Polycomb targets contain both the repressive H3K27 and the active H3K4 methylation marks; a so-called bivalent chromatin state (Bernstein et al., 2006). This is incompatible with the previously mentioned accessibility model unless repressive marks would be dominant over active modifications.

Interestingly, and somewhat in contrast to the locus contraction and repression idea, it was proposed that bivalent chromatin would "poise" genes for later activation and facilitate their transcriptional induction (Bernstein et al., 2006; Spivakov and Fisher, 2007). However, there is no evidence supporting such a gene "poising" via chromatin modifications despite the fact that many ES cell specific Polycomb targets are activated during differentiation. Moreover given that virtually all CpG islands are H3K4 methylated by default, a bivalent chromatin state might rather reflect transient repression of an otherwise accessible region in order to ensure correct expression of such critical genes.

Consistent with this idea is the fact that in Drosophila, which does not display a bimodal CpG distribution at promoters (Fig. 1), H3K4 and H3K27 methylation are generally exclusive and the frequency of bivalent domains is minute compared to mammals (Schuettengruber et al., 2009).

4.5. Impact of epigenetic repression on DNA accessibility

Naively, if the difference in strength of repression between distinct epigenetic pathways is real, it should be detectable in terms of DNA accessibility and chromatin structure. Accessibility would thus gradually decrease ranging from "high" at promoters of actively transcribed genes, to intermediate at CpG-rich promoters of inactive genes without additional epigenetic repression, to low at inactive non-CpG island and CpG-rich promoters with additional Polycomb repression, and ultimately to very low at DNA methylated CpG islands and H3K9me3/HP1 bound regions. Although a popular model concerning the impact of epigenetic modifications on chromatin structure, experimental data demonstrating such a mode of action is still lacking.

Equally important, the field currently ignores upstream signaling pathways which integrate extrinsic and intrinsic signals and probably induce the changes we observe on the chromatin level.

5. Acknowledgements

First I would like to thank Dirk Schübeler for giving me the opportunity to conduct my PhD studies in his lab. He always had an open door for critical and fruitful discussions. I highly appreciate his supervision that combined lots of individual freedom for my own ideas regarding experiments and data analysis while providing support whenever needed.

Further I would like to express my thanks to all the current and past members of the Schübeler laboratory who contributed a lot in creating a stimulating working atmosphere and who were also good friends outside of FMI. In particular I would like to point out Christiane Wirbelauer who keeps the lab up and running and provides invaluable technical assistance.

Ultimately, I would like to acknowledge Susan Gasser, Fred Meins Jr., Irina Stancheva and Primo Schär as members of my thesis committee for taking their time, contributing ideas and suggestions during the thesis committee meetings and for evaluating the thesis.

6. Bibliography

Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A *90*, 11995-11999.

Antequera, F., Tamame, M., Villanueva, J.R., and Santos, T. (1984). DNA methylation in the fungi. J Biol Chem *259*, 8033-8036.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Molecular cell *31*, 785-799.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science *316*, 744-747.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature *455*, 818-821.

Armstrong, J.A., Papoulas, O., Daubresse, G., Sperling, A.S., Lis, J.T., Scott, M.P., and Tamkun, J.W. (2002). The Drosophila BRM complex facilitates global transcription by RNA polymerase II. Embo J *21*, 5245-5254.

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature *455*, 822-825.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M.*, et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Bajic, V.B., Tan, S.L., Christoffels, A., Schonbach, C., Lipovich, L., Yang, L., Hofmann, O., Kruger, A., Hide, W., Kai, C.*, et al.* (2006). Mice and men: their promoter properties. PLoS Genet *2*, e54.

Bantignies, F., and Cavalli, G. (2006). Cellular memory and dynamic regulation of polycomb group proteins. Curr Opin Cell Biol *18*, 275-283.

Bantignies, F., Grimaud, C., Lavrov, S., Gabut, M., and Cavalli, G. (2003). Inheritance of Polycomb-dependent chromosomal interactions in Drosophila. Genes Dev *17*, 2406-2420.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F.*, et al.* (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671-675.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A., and Schubeler, D. (2008). Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster. Mol Cell Biol *28*, 3401-3409.

Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N., Schwaiger, M., MacAlpine, D.M., Zilbermann, F., van Leeuwen, F., Bell, S.P., Imhof, A.*, et al.* (2007). Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. Embo J *26*, 4974-4984.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol *203*, 971-983.

Bestor, T.H. (1990). DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. Philosophical transactions of the Royal Society of London *326*, 179-187.

Bestor, T.H. (2003). Cytosine methylation mediates sexual conflict. Trends Genet *19*, 185-190.

Bestor, T.H., and Coxon, A. (1993). Cytosine methylation: the pros and cons of DNA methylation. Curr Biol *3*, 384-386.

Bhattacharya, S.K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999). A mammalian protein with specific demethylase activity for mCpG DNA. Nature *397*, 579-583.

Bibel, M., Richter, J., Schrenk, K., Tucker, K.L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nat Neurosci *7*, 1003-1009.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev *16*, 6-21.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396-398.

Bird, A.P. (1995). Gene number, noise reduction and biological complexity. Trends Genet *11*, 94-100.

Black, B.E., and Bassett, E.A. (2008). The histone variant CENP-A and centromere specification. Curr Opin Cell Biol *20*, 91-100.

Blais, A., van Oevelen, C.J., Margueron, R., Acosta-Alvear, D., and Dynlacht, B.D. (2007). Retinoblastoma tumor suppressor protein-dependent methylation of histone H3 lysine 27 is associated with irreversible cell cycle exit. The Journal of cell biology *179*, 1399-1412.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536-2539.

Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). The SANT domain: a unique histone-tail-binding module? Nat Rev Mol Cell Biol *5*, 158-163.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev *20*, 1123-1136.

Brehm, A., Tufteland, K.R., Aasland, R., and Becker, P.B. (2004). The many colours of chromodomains. Bioessays *26*, 133-140.

Brenner, C., Deplus, R., Didelot, C., Loriot, A., Vire, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T.*, et al.* (2005). Myc represses transcription through recruitment of DNA methyltransferase corepressor. Embo J *24*, 336-346.

Brero, A., Easwaran, H.P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H., and Cardoso, M.C. (2005). Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. The Journal of cell biology *169*, 733-743.

Britten, R.J. (1997). Mobile elements inserted in the distant past have taken on important functions. Gene *205*, 177-182.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.

Caretti, G., Di Padova, M., Micales, B., Lyons, G.E., and Sartorelli, V. (2004). The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes Dev *18*, 2627-2638.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Developmental cell *12*, 503-514.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C.*, et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P.*, et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nature reviews *10*, 295-304.

Chahrour, M., and Zoghbi, H.Y. (2007). The story of Rett syndrome: from clinic to neurobiology. Neuron *56*, 422-437.

Chamberlain, S.J., Yee, D., and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem cells (Dayton, Ohio) *26*, 1496-1505.

Chen, R.Z., Pettersson, U., Beard, C., Jackson-Grusby, L., and Jaenisch, R. (1998). DNA hypomethylation leads to elevated mutation rates. Nature *395*, 89-93.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.

Clapier, C.R., and Cairns, B.R. (2009). The Biology of Chromatin Remodeling Complexes. Annual review of biochemistry.

Clouaire, T., and Stancheva, I. (2008). Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? Cell Mol Life Sci *65*, 1509-1522.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215-219.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185-196.

Dame, R.T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. Molecular microbiology *56*, 858-870.

Das, C., Lucia, M.S., Hansen, K.C., and Tyler, J.K. (2009). CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature *459*, 113-117.

Deragon, J.M., and Capy, P. (2000). Impact of transposable elements on the human genome. Annals of medicine *32*, 264-273.

Dhasarathy, A., and Wade, P.A. (2008). The MBD protein family-reading an epigenetic mark? Mutation research *647*, 39-43.

Dion, M.F., Altschuler, S.J., Wu, L.F., and Rando, O.J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. Proc Natl Acad Sci U S A *102*, 5501-5506.

Dong, A., Yoder, J.A., Zhang, X., Zhou, L., Bestor, T.H., and Cheng, X. (2001). Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. Nucleic Acids Res *29*, 439-448.

Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature *284*, 601-603.

Downs, J.A., Kosmidou, E., Morgan, A., and Jackson, S.P. (2003). Suppression of homologous recombination by the Saccharomyces cerevisiae linker histone. Molecular cell *11*, 1685-1692.

Draker, R., and Cheung, P. (2009). Transcriptional and epigenetic functions of histone variant H2A.Z. Biochemistry and cell biology = Biochimie et biologie cellulaire *87*, 19-25.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A.*, et al.* (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet *38*, 1378-1385.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H.*, et al.* (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nature structural & molecular biology *15*, 1176-1183.

Erhardt, S., Su, I.H., Schneider, R., Barton, S., Bannister, A.J., Perez-Burgos, L., Jenuwein, T., Kouzarides, T., Tarakhovsky, A., and Surani, M.A. (2003). Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. Development *130*, 4235-4248.

Esteve, P.O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev *20*, 3089-3103.

Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.H., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell *136*, 1122-1135.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell *123*, 1199-1212.

Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet *4*, e1000116.

Feinberg, A.P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. Nature reviews *7*, 21-33.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol *8*, 188-194.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG Methylation Contributes to ES Cell Gene Regulation in Parallel with Oct4/Nanog, PcG Complex, and Histone H3 K4/K27 Trimethylation. Cell Stem Cell *2*, 160-169.

Francastel, C., Schubeler, D., Martin, D.I., and Groudine, M. (2000). Nuclear compartmentalization and gene activity. Nat Rev Mol Cell Biol *1*, 137-143.

Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. Cell *137*, 110-122.

Freitag, M., Hickey, P.C., Khlafallah, T.K., Read, N.D., and Selker, E.U. (2004a). HP1 is essential for DNA methylation in neurospora. Molecular cell *13*, 427-434.

Freitag, M., Lee, D.W., Kothe, G.O., Pratt, R.J., Aramayo, R., and Selker, E.U. (2004b). DNA methylation is independent of RNA interference in Neurospora. Science *304*, 1939.

Fujita, N., Shimotake, N., Ohki, I., Chiba, T., Saya, H., Shirakawa, M., and Nakao, M. (2000). Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. Mol Cell Biol *20*, 5107-5118.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res *31*, 2305-2312.

Futcher, B., Reid, E., and Hickey, D.A. (1988). Maintenance of the 2 micron circle plasmid of Saccharomyces cerevisiae by sexual transmission: an example of a selfish DNA. Genetics *118*, 411-415.

Futscher, B.W., Oshiro, M.M., Wozniak, R.J., Holtan, N., Hanigan, C.L., Duan, H., and Domann, F.E. (2002). Role for DNA methylation in the control of cell type specific maspin expression. Nat Genet *31*, 175-179.

Fyodorov, D.V., Blower, M.D., Karpen, G.H., and Kadonaga, J.T. (2004). Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. Genes Dev *18*, 170-183.

Galagan, J.E., and Selker, E.U. (2004). RIP: the evolutionary cost of genome defense. Trends Genet *20*, 417-423.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J Mol Biol *196*, 261-282.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell *124*, 495-506.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395-398.

Gonzalo, S., Jaco, I., Fraga, M.F., Chen, T., Li, E., Esteller, M., and Blasco, M.A. (2006). DNA methyltransferases control telomere length and telomere recombination in mammalian cells. Nat Cell Biol *8*, 416-424.

Gregory, T.R. (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biological reviews of the Cambridge Philosophical Society *76*, 65-101.

Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006). RNAi components are required for nuclear clustering of Polycomb group response elements. Cell *124*, 957-971.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. Science *315*, 1143-1147.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol *10*, 1291-1300.

Heard, E., Clerc, P., and Avner, P. (1997). X-chromosome inactivation in mammals. Annu Rev Genet *31*, 571-610.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol *18*, 6538-6547.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev *15*, 710-723.

Hendrich, B., and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet *19*, 269-277.

Henikoff, S., and Comai, L. (1998). A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis. Genetics *149*, 307-318.

Herranz, N., Pasini, D., Diaz, V.M., Franci, C., Gutierrez, A., Dave, N., Escriva, M., Hernandez-Munoz, I., Di Croce, L., Helin, K.*, et al.* (2008). Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. Mol Cell Biol *28*, 4772-4781.

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., and Sasaki, H. (2008). Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev *22*, 1607-1616.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J.*, et al.* (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS biology *6*, e22.

Ioshikhes, I.P., and Zhang, M.Q. (2000). Large-scale human promoter mapping using CpG islands. Nat Genet *26*, 61-63.

Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddeloh, J.A., Wen, B., and Feinberg, A.P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). Genome Res *18*, 780-790.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E.*, et al.* (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet *27*, 31-39.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556-560.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol Cell Biol *24*, 8862-8871.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248-251.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nature reviews *10*, 161-172.

Jin, S.G., Guo, C., and Pfeifer, G.P. (2008). GADD45A does not promote DNA demethylation. PLoS Genet *4*, e1000013.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683-692.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet *19*, 187-191.

Jost, J.P. (1993). Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. Proc Natl Acad Sci U S A *90*, 4684-4688.

Jurkowski, T.P., Meusburger, M., Phalke, S., Helm, M., Nellen, W., Reuter, G., and Jeltsch, A. (2008). Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. RNA (New York, NY *14*, 1663-1670.

Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? Genes Dev *19*, 295-310.

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. Nature *452*, 112-115.

Kato, M., Miura, A., Bender, J., Jacobsen, S.E., and Kakutani, T. (2003). Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Curr Biol *13*, 421-426.

Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P., and Akhtar, A. (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. Cell *133*, 813-828.

Kireeva, N., Lakonishok, M., Kireev, I., Hirano, T., and Belmont, A.S. (2004). Visualization of early chromosome condensation: a hierarchical folding, axial glue model of chromosome structure. The Journal of cell biology *166*, 775-785.

Klebes, A., Sustar, A., Kechris, K., Li, H., Schubiger, G., and Kornberg, T.B. (2005). Regulation of cellular plasticity in Drosophila imaginal disc cells by the Polycomb group, trithorax group and lama genes. Development *132*, 3753-3765.

Klenov, M.S., Lavrov, S.A., Stolyarenko, A.D., Ryazansky, S.S., Aravin, A.A., Tuschl, T., and Gvozdev, V.A. (2007). Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. Nucleic Acids Res *35*, 5430-5438.

Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. Nature reviews *7*, 715-727.

Kobor, M.S., Venkatasubrahmanyam, S., Meneghini, M.D., Gin, J.W., Jennings, J.L., Link, A.J., Madhani, H.D., and Rine, J. (2004). A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. PLoS biology *2*, E131.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M.*, et al.* (2003a). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Molecular cell *11*, 721-729.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C.*, et al.* (2003b). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol *23*, 4207-4218.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S.*, et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet *4*, e1000242.

Kuhlmann, M., Borisova, B.E., Kaller, M., Larsson, P., Stach, D., Na, J., Eichinger, L., Lyko, F., Ambros, V., Soderbom, F.*, et al.* (2005). Silencing of retrotransposons in Dictyostelium by DNA methylation and RNAi. Nucleic Acids Res *33*, 6405-6417.

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. Development *130*, 5083-5090.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W.*, et al.* (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev *22*, 908-917.

Kurdistani, S.K., Tavazoie, S., and Grunstein, M. (2004). Mapping global histone acetylation patterns to gene expression. Cell *117*, 721-733.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lam, F.H., Steger, D.J., and O'Shea, E.K. (2008). Chromatin decouples promoter threshold from dynamic range. Nature *453*, 246-250.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol *9*, 1167-1174.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73-84.

Lee, N., Maurange, C., Ringrose, L., and Paro, R. (2005). Suppression of Polycomb group proteins by JNK signalling induces transdetermination in Drosophila imaginal discs. Nature *438*, 234-237.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol *13*, 1192-1200.

Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195-3205.

Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H., Aebersold, R., Graef, I.A., and Crabtree, G.R. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neuron *55*, 201-215.

Levine, S.S., King, I.F., and Kingston, R.E. (2004). Division of labor in polycomb group repression. Trends Biochem Sci *29*, 478-485.

Lewis, Z.A., Honda, S., Khlafallah, T.K., Jeffress, J.K., Freitag, M., Mohn, F., Schubeler, D., and Selker, E.U. (2009). Relics of repeat-induced point mutation direct heterochromatin formation in Neurospora crassa. Genome Res *19*, 427-437.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362-365.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523-536.

Lu, X., Wontakal, S.N., Emelyanov, A.V., Morcillo, P., Konev, A.Y., Fyodorov, D.V., and Skoultchi, A.I. (2009). Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. Genes Dev *23*, 452-465.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nature structural biology *10*, 882-891.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Molecular cell *32*, 503-518.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. Embo J *24*, 800-812.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838-849.

Martinez, A.M., Colomb, S., Dejardin, J., Bantignies, F., and Cavalli, G. (2006). Polycomb group-dependent Cyclin A repression in Drosophila. Genes Dev *20*, 501-513.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y.E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science *302*, 890-893.

Matzke, M., Kanno, T., Huettel, B., Daxinger, L., and Matzke, A.J. (2007). Targets of RNA-directed DNA methylation. Current opinion in plant biology *10*, 512-519.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol *7*, 540-546.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Developmental cell *10*, 105-116.

Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G.*, et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature *452*, 45-50.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-55.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science *297*, 403-405.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet *37*, 1090-1097.

Mizuguchi, G., Tsukiyama, T., Wisniewski, J., and Wu, C. (1997). Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. Molecular cell *1*, 141-150.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Molecular cell *30*, 755-766.

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197-208.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M.*, et al.* (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol *9*, 64-71.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. Mol Cell Biol *21*, 4330-4336.

Ohm, J., McGarvey, K., Yu, X., Cheng, L., Schuebel, K., Cope, L., Mohammad, H., Chen, W., Daniel, V., Yu, W.*, et al.* (2007). A stem cell -like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Okano, M., Xie, S., and Li, E. (1998a). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet *19*, 219-220.

Okano, M., Xie, S., and Li, E. (1998b). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res *26*, 2536-2540.

Ooi, S.K., and Bestor, T.H. (2008). The colorful history of active DNA demethylation. Cell *133*, 1145-1148.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and A. Thomson, J.A. (2007). Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. Cell Stem Cell *1*, 299-312.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev *10*, 1991-2002.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol *27*, 3769-3779.

Penterman, J., Uzawa, R., and Fischer, R.L. (2007a). Genetic interactions between DNA demethylation and methylation in Arabidopsis. Plant physiology *145*, 1549-1557.

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007b). DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A *104*, 6752-6757.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y.*, et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Molecular cell *12*, 1577-1589.

Peters, A.H., and Schubeler, D. (2005). Methylation of histones: playing memory with DNA. Curr Opin Cell Biol *17*, 230-238.

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M.C., and Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. Nat Genet.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E.*, et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517-527.

Polach, K.J., and Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. J Mol Biol *258*, 800-812.

Ptashne, M. (2007). On the use of the word 'epigenetic'. Curr Biol *17*, R233-236.

Rabinowicz, P.D., Palmer, L.E., May, B.P., Hemann, M.T., Lowe, S.W., McCombie, W.R., and Martienssen, R.A. (2003). Genes and transposons are differentially methylated in plants, but not in mammals. Genome Res *13*, 2658-2664.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D.*, et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Reisenauer, A., Kahng, L.S., McCollum, S., and Shapiro, L. (1999). Bacterial DNA methylation: a cell cycle regulator? Journal of bacteriology *181*, 5135-5139.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223-232.

Robinson, P.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D. (2008). 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. J Mol Biol *381*, 816-825.

Robinson, P.J., and Rhodes, D. (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Current opinion in structural biology *16*, 336-343.

Roh, T.Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. Proc Natl Acad Sci U S A *103*, 15782-15787.

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., and Bestor, T.H. (2006). Large-scale structure of genomic methylation patterns. Genome Res *16*, 157-163.

Roloff, T.C., Ropers, H.H., and Nuber, U.A. (2003). Comparative study of methyl-CpG-binding domain proteins. BMC genomics *4*, 1.

Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. Nat Rev Mol Cell Biol *6*, 139-149.

Sarraf, S.A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Molecular cell *15*, 595-605.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol *7*, 557-567.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A *103*, 1412-1417.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E.*, et al.* (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet *39*, 232-236.

Schubeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J.*, et al.* (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev *18*, 1263-1271.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS biology *7*, e13.

Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert, C.J., Jr. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. Genome biology *6*, R33.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nature reviews *8*, 9-22.

Selker, E.U., Tountas, N.A., Cross, S.H., Margolin, B.S., Murphy, J.G., Bird, A.P., and Freitag, M. (2003). The methylated component of the Neurospora crassa genome. Nature *422*, 893-897.

Sharif, J., Muto, M., Takebayashi, S.I., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.

Simpson, V.J., Johnson, T.E., and Hammen, R.F. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res *14*, 6711-6719.

Song, F., Smith, J.F., Kimura, M.T., Morrow, A.D., Matsuyama, T., Nagase, H., and Held, W.A. (2005). Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A *102*, 3336-3341.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer *6*, 846-856.

Spivakov, M., and Fisher, A.G. (2007). Epigenetic signatures of stem-cell identity. Nature reviews *8*, 263-271.

Squazzo, S.L., O'Geen, H., Komashko, V.M., Krig, S.R., Jin, V.X., Jang, S.W., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2006). Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res *16*, 890-900.

Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A., and Cedar, H. (1982). Clonal inheritance of the pattern of DNA methylation in mouse cells. Proc Natl Acad Sci U S A *79*, 61-65.

Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell *98*, 1-4.

Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., Hara, E., Tenen, D.G., Hozumi, N., and Oikawa, T. (2006). Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. Oncogene *25*, 2477-2488.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nature reviews *9*, 465-476.

Suzuki, M.M., Kerr, A.R., De Sousa, D., and Bird, A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res *17*, 625-631.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev *19*, 815-826.

Tanay, A., O'Donnell, A.H., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proc Natl Acad Sci U S A *104*, 5521-5526.

Tang, L.Y., Reddy, M.N., Rasheva, V., Lee, T.L., Lin, M.J., Hung, M.S., and Shen, C.K. (2003). The eukaryotic DNMT2 genes encode a new class of cytosine-5 DNA methyltransferases. J Biol Chem *278*, 33613-33616.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408-411.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nature structural & molecular biology *14*, 1025-1040.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Developmental cell *15*, 668-679.

Tiwari, V.K., McGarvey, K.M., Licchesi, J.D., Ohm, J.E., Herman, J.G., Schubeler, D., and Baylin, S.B. (2008). PcG proteins, DNA methylation, and gene repression by chromatin looping. PLoS biology *6*, 2911-2927.

Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. Cell *128*, 651-654.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. Embo J *13*, 3822-3831.

Ushinsky, S.C., Bussey, H., Ahmed, A.A., Wang, Y., Friesen, J., Williams, B.A., and Storms, R.K. (1997). Histone H1 in Saccharomyces cerevisiae. Yeast (Chichester, England) *13*, 151-161.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell *109*, 745-756.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nature reviews *10*, 252-263.

Vaughn, M.W., Tanurd Ic, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P.D., Dedhia, N., McCombie, W.R., Agier, N., Bulski, A.*, et al.* (2007). Epigenetic Natural Variation in Arabidopsis thaliana. PLoS biology *5*, e174.

Vettese-Dadey, M., Grant, P.A., Hebbes, T.R., Crane- Robinson, C., Allis, C.D., and Workman, J.L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. Embo J *15*, 2508-2518.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M.*, et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

Vogel, M.J., Guelen, L., de Wit, E., Peric-Hupkes, D., Loden, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.K., and van Steensel, B. (2006). Human heterochromatin proteins form large domains containing KRAB-ZNF genes. Genome Res *16*, 1493-1504.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol *20*, 2108-2121.

Waddington, C.H. (1942). The epigenotype. Endeavour *1*, 18-20.

Waddington, C.H. (1957). The Strategy of the Genes (Allen & Unwin).

Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F., and Wolffe, A.P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet *23*, 62-66.

Walsh, C.P., and Bestor, T.H. (1999). Cytosine methylation and mammalian development. Genes Dev *13*, 26-34.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.

Wang, Y., Jorda, M., Jones, P.L., Maleszka, R., Ling, X., Robertson, H.M., Mizzen, C.A., Peinado, M.A., and Robinson, G.E. (2006). Functional CpG methylation system in a social insect. Science *314*, 645-647.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897-903.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet *37*, 853-862.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Whitcomb, S.J., Basu, A., Allis, C.D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. Trends Genet *23*, 494-502.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I.*, et al.* (2007). Epigenetic stem cell signature in cancer. Nat Genet *39*, 157-158.

Wirbelauer, C., Bell, O., and Schubeler, D. (2005). Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. Genes Dev *19*, 1761-1766.

Wolffe, A.P., and Hayes, J.J. (1999). Chromatin disruption and modification. Nucleic Acids Res *27*, 711-720.

Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A.M., Lipkin, M.*, et al.* (2002). Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. Proc Natl Acad Sci U S A *99*, 14937-14942.

Workman, J.L., and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annual review of biochemistry *67*, 545-579.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P.*, et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature *402*, 187-191.

Yasui, D.H., Peddada, S., Bieda, M.C., Vallero, R.O., Hogart, A., Nagarajan, R.P., Thatcher, K.N., Farnham, P.J., and Lasalle, J.M. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc Natl Acad Sci U S A *104*, 19416-19421.

Yoder, J.A., and Bestor, T.H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet *7*, 279-284.

Yoon, H.G., Chan, D.W., Reynolds, A.B., Qin, J., and Wong, J. (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Molecular cell *12*, 723-734.

Zeyl, C., Bell, G., and Green, D.M. (1996). Sex and the spread of retrotransposon Ty3 in experimental populations of Saccharomyces cerevisiae. Genetics *143*, 1567-1577.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E.*, et al.* (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189-1201.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev *13*, 1924-1935.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev *15*, 2343-2360.

Zhao, X., Ueba, T., Christie, B.R., Barkho, B., McConnell, M.J., Nakashima, K., Lein, E.S., Eadie, B.D., Willhoite, A.R., Muotri, A.R.*, et al.* (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. Proc Natl Acad Sci U S A *100*, 6777-6782.

Zhu, J., Kapoor, A., Sridhar, V.V., Agius, F., and Zhu, J.K. (2007). The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis. Curr Biol *17*, 54-59.

Zilberman, D. (2008). The evolving functions of DNA methylation. Current opinion in plant biology *11*, 554-559.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet *39*, 61-69.