Targeting and dynamics of gene repression during stem cell differentiation

Inauguraldissertation

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

von

Florian Lienert
aus Einsiedeln, SZ

Basel, 2011

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Susan Gasser, Prof. Dr. Primo Schär und Prof. Dr. Dirk Schübeler.

Basel, den 21. Juni 2011

Prof. Dr. Martin Spiess
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

MDR	methylation determining region
TF	Transcription factor
Pol II	RNA polymerase II
KAT	Histone acetyltransferase
HDAC	Histone deacetylase
KMT	Histone lysine methyltransferase
PRMT	Protein arginine methyltransferase
me	any methylation state of an arginine or lysine
me1	mono-methylation
me2	di-methylation
me3	tri-methylation
PcG	Polycomb group protein
TxG	Trithorax-group protein
ES cell	Embryonic stem cell
PRC	Polycomb repressive complex
E1	embryonic day 1
DNMT	DNA methyltransferase
ICR	imprinting control region
MBP	Methyl-CpG-binding protein
PGC	primordial germ cell
IAP	Intracisternal A particle (low copy retroviral-like element)
DMR	differentially methylated region

Protein names in capitals irrespective of mouse or human origin

Gene names italic

1. Summary

Summary

The identity and function of different cellular subtypes critically depend on their unique set of expressed genes. Gene expression programs and their changes during development are mainly controlled by sequence-specific DNA binding factors. It has recently become clear that chromatin modifications are important regulators of these processes. While there are several chromatin-based pathways that correlate with gene repression, their exact role in silencing remains elusive. Moreover, for many repressive chromatin modifications a complete picture of the genomic distribution and its dynamics during development is lacking. Finally, it is still unclear how these genomic patterns of repressive chromatin marks are established. We here set out to address these questions by studying the targeting of H3K9me2 and DNA methylation during cellular differentiation.

Our analysis revealed that H3K9me2 is highly abundant in embryonic stem cells and occurs in large domains that occupy more than half of the genome. H3K9me2 marks chromatin outside of transcribed, active or polycomb regulated sites, possibly keeping it in a repressed state. Importantly, abundance of H3K9me2 increases only slightly during neuronal differentiation, with a localized gain occurring at gene bodies of transcribed genes. By gene expression profiling we further show that the transcriptome complexity is very similar in stem cells and derived post-mitotic neurons. These data are in contrast to a previously suggested model which states that the pluripotent state of stem cells is accompanied by a global reduction in heterochromatin and a concomitant higher proportion of transcription. Together with results from other groups our data rather indicate that repressive chromatin is abundant in stem cells and upon differentiation gets redistributed only locally and not globally. It has been suggested that such a localized increase of repression at gene regulatory regions helps stabilizing lineage choices and differentiation processes.

In order to investigate how chromatin-based repression pathways are targeted to gene regulatory sites, we focused on DNA methylation, a modification whose catalysis and epigenetic propagation are well understood. By site-specific sequence integration experiments we show that 1 kb promoter elements are sufficient to recapitulate endogenous DNA methylation patterns in stem cells and their dynamic changes upon differentiation, in a process that is independent of transcription. In stem cells, promoters are protected from DNA methylation by small sequence elements that we termed methylation determining regions (MDRs). Protection from DNA methylation by MDRs depends on a combination of DNA binding motifs, which get recognized by transcription factors such as RFX2. It has been speculated before that establishment of an unmethylated promoter state is facilitated by proteins that recognize unmethylated CpGs. While not excluding a role in maintenance, our data suggest that CpG-richness alone is not sufficient for initiation of this chromatin state.

Remarkably, no additional sequence besides an MDR is needed to recapitulate differentiation-induced *de novo* methylation. Moreover, MDRs are able to protect neighboring sequences from DNA methylation in stem cells and from *de novo* methylation during differentiation. These results imply that one possible way of differentiation-induced *de novo* methylation could involve reduced binding of factors that protect from DNA methylation.

In summary, H3K9me2 and DNA methylation occupy per default most the genome, even in cells with a high developmental potential. Accordingly, cellular differentiation is accompanied by focal, rather than global changes in repressive chromatin modifications. In the case of DNA methylation, such local changes at gene regulatory sites are determined by the underlying sequence and likely involve binding of transcription factors that protect from DNA methylation.

2. Introduction

During mammalian development a single fertilized egg gives rise to hundreds of specialized cell types. During this process, cells go through a series of sequential lineage choices and gradually decrease their developmental potential. While the genetic information content of differentiating cells stays constant throughout development, their set of expressed genes is subject to major changes. Switches in transcriptional programs that govern embryogenesis are mainly determined by sequence specific DNA binding proteins. These transcription factors form gene regulatory networks that are reused during different steps throughout development (Davidson 2010). While DNA sequence recognition lies at the core of transcription networks, the packaging of DNA into chromatin allows for a second layer of regulation. Changes in the occupancy, structure and modifications of chromatin alter its accessibility and thereby influence binding of sequence specific factors. Furthermore, some chromatin modifications were demonstrated or proposed to be epigenetically inherited during cellular division (Margueron and Reinberg 2010). Given these properties of chromatin modifications, it has been speculated that they are major players in determining the plasticity and stability of lineage choices during development (Reik 2007).

The following paragraphs will summarize current knowledge on the structure and remodeling of chromatin, on chromatin modifications and the role of these processes in transcriptional regulation in relation to my PhD thesis project.

2.1. Gene regulation in bacteria and mammals

The genome of *E. coli*, the most studied bacterial organism, contains around 4'300 genes embedded in a circular genome of 4.6 x 10^6 basepairs (Blattner et al. 1997). In bacteria, regulation of gene expression is mostly needed to adjust the growth rate and metabolism in response to environmental changes. Jacob and Monod, who pioneered the study of bacterial gene regulation, suggested that promoters of structure genes are controlled by operator sites. Their operon model further stated that regulator genes act on the operator through action of a repressor, which itself can be modulated by small molecules (Jacob and Monod 1961). While this early model proposed that the repressor would be an RNA molecule and inhibit activity by base-pairing to the operator, it was later shown that transcriptional repressors represent proteins which bind to DNA in a sequence specific manner (Ptashne 1967). In bacteria, transcriptional repressors act by directly preventing access of the RNA polymerase to promoter regions. Transcriptional activators, on the other hand, facilitate binding of RNA polymerase and its associated factors and thereby induce transcription. While most transcription factors (TFs) show the same mode of activity (either repressive or active) on all their target genes, there are also examples where a TF acts as a repressor at some genes and as an activator at other target genes (Taniguchi and de Crombrugghe 1983). The activating and repressing transcription factors of *E. coli* form a highly interconnected network. More precisely, one TF regulates on average three genes, and one *E.coli* gene is directly controlled by two TFs (Thieffry et al. 1998). However, the connectivity of the TF network follows a power-law distribution and of the ~300 TFs that the *E. coli* genome codes for, only nine proteins control over half of all genes (Martinez-Antonio and Collado-Vides 2003). The TF network of *E. coli* is composed of smaller, repeatedly occurring network motifs, of which each has a specific function in determining gene expression (Shen-Orr et al. 2002). These functions include generation of temporal expression programs, control of responses to fluctuating external signals, adjustment of response time and generation of bi-stable gene expression (Alon 2007).

Most of the basic transcriptional mechanisms are conserved between bacteria and eukaryotes. However, it has been argued that the logic of gene regulation is fundamentally different in eukaryotes (Struhl 1999). Prokaryotic transcription is mainly determined by the quality of the promoter sequence, and the *in vivo* transcription rate of an isolated promoter is similar to that achieved *in vitro*. The ground state for prokaryotic transcription therefore seems to be non-restrictive. In eukaryotes, RNA polymerase II (Pol II) needs help of several general transcription factors to be able to initiate transcription *in vitro*. Furthermore, in

contrast to the situation in bacteria, a eukaryotic core promoter is not sufficient to induce transcription *in vivo*. The ground state for eukaryotic transcription is therefore thought to be restrictive (Struhl 1999). This is mainly a consequence of the fact that eukaryotic DNA is packaged into a dense chromatin structure. Notably, emergence of histones, the basic unit of this chromatin structure, accompanied the evolutionary transition from prokaryotes to eukaryotes. It has therefore been speculated that evolution of a restrictive chromatin structure has been instrumental in allowing to acquire more genes in the eukaryotic genome by reducing noisy transcription initiation (Bird 1995).

To overcome the restrictive state of chromatin, eukaryotic Pol II needs the action of transcriptional co-activators, which either directly interact with the Pol II complex or facilitate transcription by modifying the chromatin structure. Interaction of activators and the Pol II machinery further depends on a large protein complex known as the mediator complex [reviewed in (Malik and Roeder 2010)]. In addition, transcription initiation in higher eukaryotes is often regulated by enhancers; distal binding sites of transcriptional activators [reviewed in (Ong and Corces 2011)]. Together, these diverse transcriptional regulatory mechanisms enable eukaryotes to generate and fine-tune very complex temporal and spatial patterns of gene expression.

2.2. Chromatin

2.2.1. The nucleosome as the basic unit of chromatin

The term chromatin was first used by Walther Flemming to describe a structure in cell nuclei that strongly absorbed basophilic dyes (Flemming 1882). During the same time period, Miescher and Kossel investigated the chemical composition of the nucleus and identified nucleic acid and a protein portion that Kossel named 'histon' (Miescher 1871; Kossel 1884). Later, advances in fractionation methods revealed that the 'histon' protein part is composed of four different histones (Johns 1964). Experiments using endonuclease digestion further indicated that chromatin is composed of a sub-structure of around 200 bp in length (Hewish and Burgoyne 1973). Based on these results and cross-linking studies of histones it was proposed that the basic repeating unit of chromatin is formed by ~200 bp of DNA in complex with 4 histone pairs (Kornberg 1974; Kornberg and Thomas 1974). Electron microscopy revealed that this basic unit, which was later termed nucleosome, can be observed on isolated chromatin and that *in vitro* reconstituted nucleosomes can self-assemble (Oudet et al. 1975). Finally, by obtaining crystal structures it was shown that histones contain two different domains: a globular histone-fold domain, consisting of one long and two short hydrophobic alpha-helices and a long unstructured, hydrophilic N-terminal tail. Two histone H2A-H2B dimers and a histone H3-H4 tetramer build up the nucleosome core (Fig. 1). DNA is wrapped around this octamer of histones 1.65 times, corresponding to 147 bp of DNA (Luger et al. 1997). Repeats of nucleosome cores further assemble into higher-order chromatin structures which are stabilized by the linker histone H1, which is not part of the nucleosome core particle, but binds the nucleosome at the entry and exit sites of the DNA (Luger 2003).

The *in vivo* architecture of higher-order assemblies of nucleosomal arrays and its potential influence on transcription regulation are still unclear. Transcriptional regulation at the level of nucleosomes, on the other hand, has been intensely investigated. Initial *in vitro* studies revealed that nucleosomes bound to promoters inhibit initiation of transcription (Knezetic and Luse 1986; Lorch et al. 1987). Besides generally hindering access of the transcription machinery, nucleosome occupancy directly influences binding of transcription factors. Notably, the inherent ability to bind nucleosomal templates seems to vary among different TFs (Taylor et al. 1991). Nucleosomes are generally depleted from active regulatory sites in the yeast genome, partly trough action of polyA tract sequences that repel nucleosomes (Iyer and Struhl 1995; Lee et al. 2004; Kaplan et al. 2009; Zhang et al. 2009). At yeast promoters, an interplay of nucleosome-free regions and TF binding sites of varying

affinity is used to fine tune transcriptional responses (Lam et al. 2008). There is recent evidence that also in mammalian cells transcription factor binding sites might be predetermined by reduced nucleosomal occupancy (John et al. 2011).

2.2.2. Chromatin remodeling

Eukaryotes possess specific protein complexes which change nucleosome positioning and thereby potentially influence the accessibility of DNA. Genes that influence transcription by altering the chromatin structure were initially identified by screens for mutant yeast strains that affect mating-type switching and growth on sucrose (Neigeborn and Carlson 1984; Hirschhorn et al. 1992; Peterson and Herskowitz 1992). These so called chromatin remodeling complexes use ATP hydrolysis to disrupt DNA-nucleosome contacts, remove or exchange nucleosomes or to move nucleosomes along DNA. Chromatin remodelers can be grouped in to four different protein families: The SWI/SNF family, the INO80/SWR1 family, the ISWI family and the CHD family [reviewed in (Hargreaves and Crabtree 2011)]. To function *in vivo,* most chromatin remodeler ATPases form large complexes with additional factors. Reports on a neuron-specific SWI/SNF complex suggested that switches in the subunit composition of remodeler complexes are critically involved in cellular differentiation processes (Lessard et al. 2007). Since most functional *in vitro* studies on ATP dependent remodelers where performed without including all subunits of the complex, it is unclear how much such experiments reflect the *in vivo* function (Hargreaves and Crabtree 2011) . Despite this caveat, *in vitro* studies led to models, how remodeling complexes utilize ATP hydrolysis to move DNA around nucleosomes. In case of ACF, a member of the ISWI family, it has been proposed that the remodeler introduces a DNA loop at the nucleosome entry site that propagates and leads to repositioning of the nucleosome (Strohner et al. 2005). A way how such repositioning of nucleosomes can affect the expression of genes has been shown for the Isw2 complex in yeast. Here, the Isw2 complex is needed for positioning of nucleosomes over sequences at promoters that are thermodynamically unfavorable for nucleosomes, which leads to reduced accessibility and transcription factor binding (Whitehouse and Tsukiyama 2006). Moreover, this repositioning enforces the directionality of transcription initiation and prevents transcription from cryptic sites (Whitehouse et al. 2007). Conversely, remodelers also play a role in increasing accessibility of chromatin and can thereby enhance gene expression. The time point of the transcriptional cycle, during which the action of a particular remodeler is needed, varies from gene to gene (Clapier and Cairns 2009). The timing itself is likely regulated by specific transcription factors that recruit chromatin modelers. For instance, recruitment of SWI/SNF

by the transcription factor SWI5 represents the first step of transcription initiation at the yeast *HO* promoter (Cosma et al. 1999).

2.2.3. Histone variants

The first indications that histone proteins occur as different variants came from biochemical studies on calf thymus nuclei (Marzluff et al. 1972). It was later shown that these non-canonical histone-variants are found as single copies in the genome and are expressed throughout the cell cycle. This stands in contrast to canonical histones (H2A, H2B, H3 and H4), which occur as clustered arrays in the genome and are transcriptionally linked to DNA replication [for a recent review on histone variants see (Talbert and Henikoff 2010)]. Certain histone variants, including CENP-A, H3.3, H2A.Z and H2A.X, evolutionary date back to the earliest known diversifications of eukaryotic lineages, while others, such as Macro H2A, are only found in animals (Malik and Henikoff 2003; Talbert and Henikoff 2010). In general, these universal variants are thought to directly alter the nucleosome structure and thereby its stability.

CENP-A is an H3 like histone variant that is specifically found at centromeres and plays an essential role in assembly of the kinetochore (Palmer et al. 1991; Santaguida and Musacchio 2009). The histone variant H2AX, on the other hand, is involved in DNA repair processes. Phosphorylation of H2AX upon induction of double-strand breaks leads to recruitment of DNA repair proteins, histone modifying enzymes and chromatin remodeling complexes (van Attikum and Gasser 2009). Macro H2A occurs on the inactive X chromosome and was recently reported to bind to proteins modified by poly-ADP-ribosylation (Chadwick and Willard 2002; Timinszky et al. 2009). Further, it has been shown that *in vitro* assembled nucleosome containing macro H2A show an altered structure which prevents binding of a particular transcription factor (NK-kB) and impedes remodeling by the SWI/SNF complex (Angelov et al. 2003). In yeast, the histone variant H2A.Z occurs preferentially around nucleosome-free regions at promoters irrespective of their transcriptional status (Raisner et al. 2005). H2A.Z has seemingly contradictory influence on activation and repression of transcription, heterochromatin and DNA repair (Zlatanova and Thakar 2008). It has however been proposed that a common feature of these diverse roles might be the tendency of H2A.Z to form stably positioned nucleosomes (Talbert and Henikoff 2010). The histone variant H3.3 replaces nucleosomes that get evicted during the process of transcription (Wirbelauer et al. 2005). H3.3 is also enriched at DNA hypersensitive sites, at sites regulated by the trithorax and polycomb group proteins and at origins of replication (Mito et al. 2007). It has recently been reported that these different sites of H3.3 deposition

show differential turnover of nucleosomes, which might directly influence local accessibility and gene expression (Deal et al. 2010). However, transcriptional regulation does not seem to critically depend on the histone variant H3.3, since Drosophila mutants lacking both copies of H3.3 are viable (Hodl and Basler 2009).

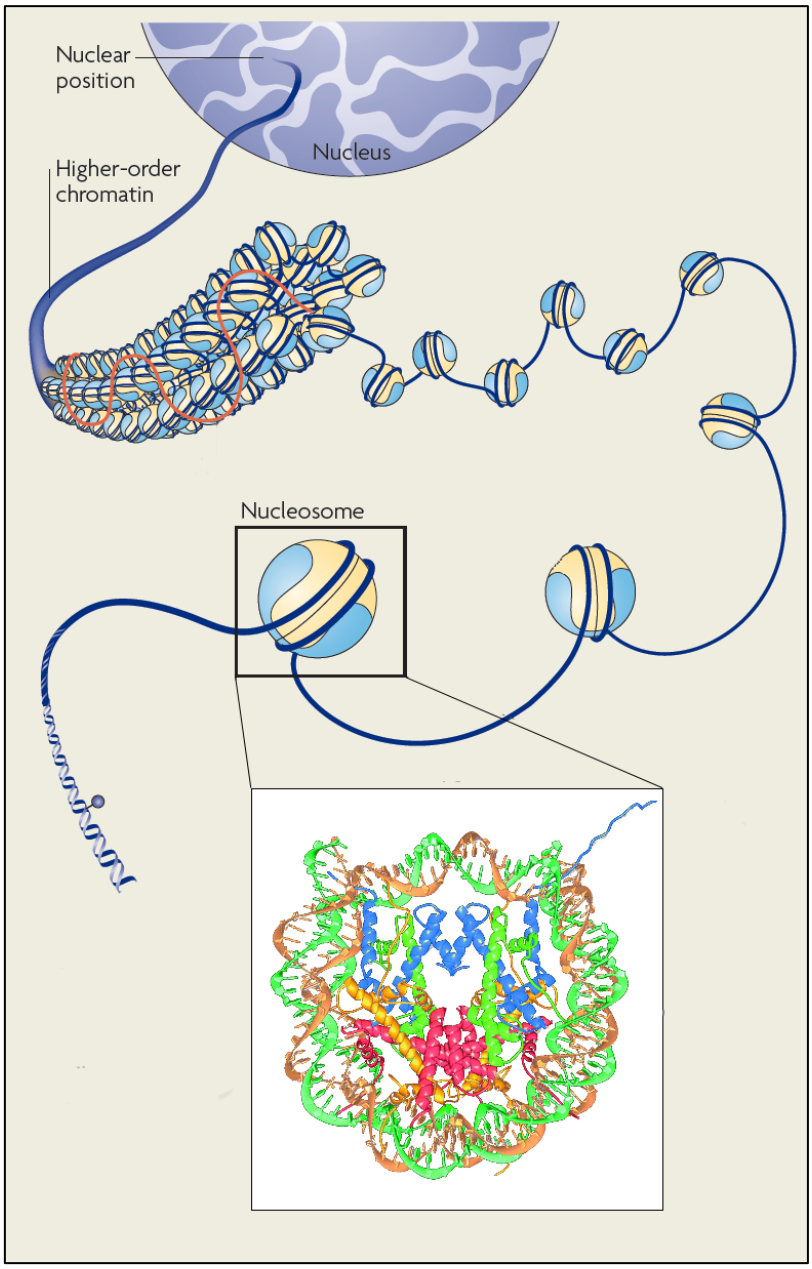


Figure 1. **Model of chromatin structure.**
The box shows the nucleosome core structure with the 146-bp DNA backbones (brown and turquoise) and the main chains of the core histones (blue: H3; green: H4; yellow: H2A; red: H2B. Adapted from (Probst et al. 2009) and (Luger et al. 1997).

2.3. Chromatin modifications

2.3.1. Histone modifications

The first indications that histones are posttranslationally modified came from an *in vitro* experiment using isolated calf thymus nuclei and labeled acetyl and methyl donors. This study further provided evidence that acetylated histones keep a high affinity for nucleosomes, yet lose some of their capacity to inhibit RNA synthesis (Allfrey et al. 1964). Nowadays, it is known that histones can be posttranslationally modified in many different ways (Fig. 2). These various histone modifications either influence chromatin structure over short or long distances and regulate binding of effector molecules [reviewed in (Bannister and Kouzarides 2011)]. Below I will discuss different histone modifications, concentrating on the ones that relate to my PhD thesis work.

2.3.1.1. Histone acetylation

In their pioneering study on histone modifications, Allfrey et al. suggested that facilitation of transcription by histone acetylation depends on the neutralization of positively charged lysines, which disrupts stable electrostatic interactions of DNA with histones (Allfrey et al. 1964). 30 years later, this notion was supported by the observation that sites of histone acetylation overlap with DNAse hypersensitive regions, which are thought to reflect regions of nucleosome displacement (Hebbes et al. 1994). A further link between histone acetylation and gene activation was provided by identification of a histone acetyltransferase with high homology to the yeast transcriptional activator GCN5 (Brownell et al. 1996). At the same time, a screen for histone deacetylases identified a homolog of the yeast transcriptional regulator RPD3, suggesting that also removal of histone acetylation might play a regulatory role in transcription (Taunton et al. 1996). Since then, a multitude of histone acetyltranserases (KAT) and histone decetylases (HDAC) have been identified, many of which with a role in transcription. Most members of the KAT and HDAC protein families occur as stable multiprotein complexes and show relatively relaxed substrate specificity; a single enzyme is often able to modify multiple lysine sites of histone proteins (Yang and Seto 2007). Genome-wide mapping of acetylation at all four core histones revealed that these marks often overlap and are highly correlative with transcription (Schubeler et al. 2004; Pokholok et al. 2005; Wang et al. 2008). Histone acetylation marks might therefore have an additive effect on chromatin accessibility and transcription. In support of this notion, combinatorial substitutions of lysine 5,8,12 of histone H4 in yeast lead to transcriptional changes that scale with the number of affected residues in a non-specific manner (Dion et al. 2005). However, this model does not exclude that some acetylated lysine residues have a

specific role. In fact, the study by Dion et al. further reported a unique expression pattern associated with the deletion of lysine K16 on histone H4 (H4K16) (Dion et al. 2005). A specific effect of this particular acetylated lysine was later supported by the finding that H4K16 acetylation alone can inhibit formation of a 30 nm higher-order chromatin structure *in vitro* (Shogren-Knaak et al. 2006). Besides disrupting nucleosome-DNA interaction and chromatin fiber formation, histone acetylation also influences DNA templated events through binding of mediator proteins. Recognition of acetylated lysines occurs almost exclusively through highly conserved bromodomains, which occur in most KAT proteins, in the general transcription factor TAF1 and in many chromatin remodeling complexes (Zeng and Zhou 2002; Taverna et al. 2007).

2.3.1.2. Histone methylation

Histone methylation occurs at lysines and arginines and, in contrast to acetylation, does not alter the amino acid charge. Histone lysine methyltransferases (KMTs) and protein arginine methyltransferases (PRMTs) catalyze addition of up to three methyl marks (me1, me2, me3). These enzymes show a high specificity for the residues they modify and for the number of added methyl groups (Bannister and Kouzarides 2011).

2.3.1.2.1. Arginine methylation

Shortly after the first report of a gene encoding a PRMT (Gary et al. 1996), Chen et al. identified CARM1 (now called PRMT4) that methylates histones and enhances transcription as a coactivator (Chen et al. 1999). Of the 11 PRMTs found in human, PRMT1 and 4 possess a role as transcriptional coactivators and methylate either TFs, other coactivators or histones [reviewed in (Bedford and Clarke 2009)]. In case of the transcription factor RUNX1, arginine methylation by PRMT1 leads to its dissociation from the corepressor SIN3A, thereby promoting transcriptional activation (Zhao et al. 2008b). The mechanism of transcriptional activation by arginine methylation of histones is less clear, since up to date no specific mediator which binds arginine methylation has been identified. PRMT5 and 6 are the only two PRMTs with a role in transcriptional repression, with both acting through methylation of histones (Bedford and Clarke 2009). Again, no proteins that recognize single methylated arginines and possess repressive function are known. However, it has been shown that PRMT6 mediated histone H3 arginine 2 (H3R2) methylation precludes binding of proteins that bind to methylated lysine 4 on the same histone (H3K4me), such as the H3K4 methyltransferase MLL1 (Hyllus et al. 2007; Iberg et al. 2008). This suggests that histone

arginine methylation might influence transcriptional activation through inhibition of H3K4 methylation, which is generally correlated with accessible chromatin.

2.3.1.2.2. H3K9 methylation

Only recently, the Jenuwein group identified the first histone lysine methyltransferase, SUV39H1, and showed that its SET domain catalyses H3K9 methylation (Rea et al. 2000). The same group later identified SUV39H2 and showed that deletion of both of these KMTs impairs pericentric hetereochromatin formation and genome stability (O'Carroll et al. 2000; Peters et al. 2001). This work led to a model how constitutive heterochromatin, which is mainly found at the repeat-rich centromeric and telomeric regions, is assembled and maintained. In this model, H3K9me3 recruits HP1, which itself recruits two different SUV420H enzymes that mediate H4K20me3 (Lachner et al. 2001; Schotta et al. 2004). The H4K20me3 mark might be directly involved in setting up a compact higher-order structure at heterochromatin, since *in vitro* reconstituted nucleosomal arrays carrying H4K20me3 show enhanced condensation (Lu et al. 2008). Intact constitutive heterchromatin further depends on the association of HP1 with the *de novo* DNA methyltransferase DNMT3B. In line with a connection to DNA methylation, *SUV39H* double-knockout cells show reduced DNA methylation and slight transcriptional up-regulation of major satellite repeats (Lehnertz et al. 2003). Recent evidence indicates that transcripts originating from major satellite repeats are the initial trigger for recruitment of HP1 at pericentric heterochromatin (Maison et al. 2011). Whether HP1 subsequently recruits SUV39H and thereby starts a self-reinforcing loop of heterochromatin formation remains to be determined.

SETDB1, G9a and GLP represent three additional KMTs with specificity towards H3K9. In contrast to the SUV39H homologs, these enzymes do not localize to constitutive heterochromatin and are rather involved in silencing repetitive DNA and retroviral repeats in euchromatic regions (Kouzarides 2007). It was recently reported that SETDB1, which catalyzes H3K9me3, acts together with the corepressor KAP1 in silencing of endogenous retroviruses during the period of early embryogenesis (Matsui et al. 2010). G9a and GLP form a heterodimeric complex and catalyze di-methylation of the H3K9 residue (Tachibana et al. 2005). G9a plays a role in targeting as well as in maintaining DNA methylation, which however occurs independently of its catalytic activity (Feldman et al. 2006; Dong et al. 2008). Similarly, while G9a seems to be involved in promoter silencing at several genes, such as MAGE-2, IF-β and P21, it is yet unclear whether the H3K9me2 mark itself is required for transcriptional regulation (Tachibana et al. 2002; Gyory et al. 2004; Nishio and Walsh 2004).

2.3.1.2.3. H3K27 methylation

Besides H3K9 methylation, H3K27me3 represents a second histone methyl mark that correlates with transcriptional repression. H3K27me3 is known as a mechanistic intermediate during transcriptional repression by Polycomb-group (PcG) proteins. PcG proteins and the antagonistically acting Trithorax-group (TrxG) proteins have initially been identified as regulators of HOX gene expression throughout drosophila embryonic development (Maeda and Karch 2006). While keeping a role in HOX gene regulation, PcG genes underwent a major expansion and diversification during vertebrate evolution (Whitcomb et al. 2007).

In mammals, H3K27me3 seems to be essential for embryonic development, as deletion of PcG proteins that set this mark is early embryonic lethal (Faust et al. 1995; O'Carroll et al. 2001; Pasini et al. 2004). In embryonic stem (ES) cells, H3K27me3 and PcG proteins occupy many inactive promoters of key developmental regulators and are therefore thought to maintain pluripotency and cellular identity in these cells (Bernstein et al. 2006; Boyer et al. 2006; Lee et al. 2006). Gene repression by PcG proteins has also been implicated in regulation of cellular differentiation during later steps of development (Mohn et al. 2008; Ezhkova et al. 2009).

Biochemical studies of PcG proteins revealed that they form at least two classes of complexes designated as polycomb repressive complexes 1 and 2 (PRC1 and PRC2), with each class including several complexes with distinct compositions (Kerppola 2009). PRC2 directly mediates H3K27me3, which in turn leads to recruitment of PRC1 through binding by chromodomains (Cao et al. 2002; Fischle et al. 2003). However, the dependency of PRC1 on PRC2 does not seem to be strict, as at certain polycomb targets recruitment of PRC1 seems to occur independently of H3K27me3 (Schoeftner et al. 2006). The PRC1 proteins RING1A and RING1B mediate monoubiquitination of histone H2A, a modification which was proposed to be directly responsible for PcG mediated gene repression (de Napoles et al. 2004; Wang et al. 2004). However, as a recent study revealed that PRC1 mediated chromatin compaction does not depend on the ubiquitin mark (Eskeland et al. 2010), the mechanism of gene repression by PcG proteins remains elusive.

It is also still not fully understood how PcG proteins are guided to their genomic targets. In drosophila, PRC2 gets recruited to defined sequences, termed polycomb response elements (PRE) [reviewed in (Muller and Kassis 2006)]. Binding of PREs is mediated by a complex pattern of different motifs recognized by various sequence-specific DNA-binding proteins, such as GAF, Zeste, DSP1, Pipsqueak, Grainyhead and SP1. However, as all of these factors also participate in processes other than PcG silencing, it

remains unclear which part of their range of activities is involved in targeting PREs (Beisel and Paro 2011). In mammals, the mechanism of polycomb recruitment is even less well understood. Recently reported mammalian PREs were of rather large size (2-3 kb) and their mode of specifying polycomb recruitment remained undefined (Sing et al. 2009; Woo et al. 2010). Moreover, based on integration of ectopic DNA in murine ES cells, Mendenhall et al. proposed that PRC2 might be recruited by default to any CpG-rich sequence that is depleted of activating motifs (Mendenhall et al. 2010). On the other hand, it has been suggested that long noncoding RNAs, such as XIST, HOTAIR and ANRIL, directly recruit PcG proteins (Rinn et al. 2007; Zhao et al. 2008a; Yap et al. 2010).

2.3.1.2.4. H3K4 methylation

The yeast protein SET1 was the first H3K4 methyltransferase to be identified (Briggs et al. 2001; Roguev et al. 2001). While these initial studies suspected H3K4 methylation to be involved in gene silencing, mapping of H3K4me in the yeast genome revealed that this mark is highly correlated with transcription. Remarkably, different methylation states of K4 show a distinct profile over expressed genes: K4me3 peaks at start sites, K4me1 towards the 3' end of genes and K4me2 in between (Pokholok et al. 2005). As H3K4 methylation can be bound by chromatin remodelling complexes and different histone acetyltransferases, it might be directly involved in facilitating transcription initiation (Santos-Rosa et al. 2003; Taverna et al. 2006). In yeast, SET1 gets recruited by the Ser5 phosphorylated C-terminal of PolII through a link of H2BK123 monoubiquitination by a yet elusive mechanism (Dehe and Geli 2006). In mammals, the recruitment mechanism of SET1 to transcribed genes seems to be largely conserved (Zhu et al. 2005). However, mammals have at least ten known or predicted H3K4 methyltransferases indicating a high degree of specialization and/or redundancy (Ruthenburg et al. 2007). The H3K4 KMT MLL1, for example, binds to only a subset of transcribed genes and mice with an inactive Mll1 show a defined homeotic developmental phenotype (Milne et al. 2005; Terranova et al. 2006). Furthermore, although H3K4me3 can be directly bound by the general transcription factor TFIID and thereby might facilitate transcription (Vermeulen et al. 2007), H3K4me is not exclusively located at transcribed regions in mammals. Genome-wide maps rather showed that H3K4 methylation marks all promoters with a high CpG density irrespective of their activity (Roh et al. 2006; Guenther et al. 2007; Mikkelsen et al. 2007; Weber et al. 2007). The location of H3K4 methylation at these regions might be linked to their general higher accessibility (Roh et al. 2005). As an additional layer of complexity different methylation states of H3K4 seem to mark regions with distinct functions. H3K4me2 and me3 were suggested to play unique roles during developmental switches in progenitor cells and H3K4me1 is part of a group of histone

modifications that uniquely marks enhancer regions (Heintzman et al. 2007; Orford et al. 2008). There is evidence that, similar to the situation at promoters, K4 methylation at enhancers may be involved in recruitment of chromatin modelers (Schnetz et al. 2009). It is not clear yet, how H3K4 HMTs are targeted to these regions, but it is likely that various mechanisms play a role, such as recruitment through TFs, RNAs or other histone modifications (Ruthenburg et al, 2007).

2.3.1.2.5. H3K36 and H3K79 methylation

In yeast, a second HMT, named SET2, is recruited by the elongating Pol II (Krogan et al. 2003). SET2 mediates H3K36 methylation, which in turn gets recognized by the histone deacetylase complex Rpd3 (Keogh et al. 2005). It has been suggested that H3K36 mediated deacetylation within transcribed genes is involved in preventing spurious transcription (Carrozza et al. 2005). While higher eukaryotes posses at least three additional H3K36 HMTs (Kouzarides 2007), the preferential location of H3K36 methylation at gene bodies of active genes is conserved (Barski et al. 2007; Bell et al. 2007). Experiments in drosophila revealed that two different HMTs mediate H3K36me2 and me3, with the latter mark showing preferred enrichment towards the 3' end of transcribed genes. Interestingly, while H3K36me3 in analogy to the situation in yeast signals reduction of H4K16ac, the dimethyl mark seems to have the opposing effect (Bell et al. 2007).

H3K79 methylation represents an additional mark that was shown to be enriched at transcribed genes (Schubeler et al. 2004). H3K79 methylation is catalyzed by DOT1, which is the only lysine HMT that does not contain a SET domain (Feng et al. 2002; van Leeuwen et al. 2002). DOT1 catalyses all three methylation variants of H3K79, which seem to colocalize in the genome (Frederiks et al. 2008). To date, no protein that specifically binds to H3K79me has been identified and the role of this modification in regulation of gene expression remains unclear.

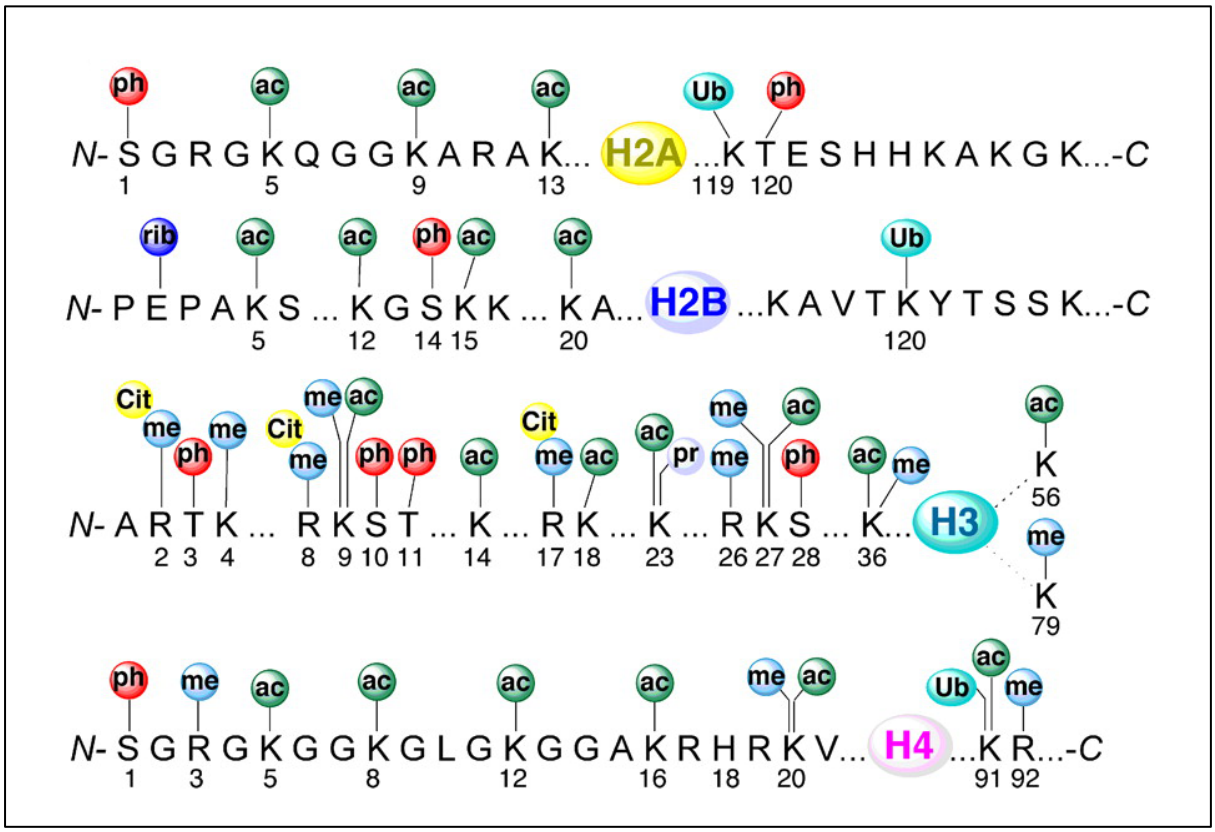


Figure 2. **Histone modifications.**
Overview of some of the known post-translational modifications on the N-terminal and C-terminal tails of canonical histones. Modifications groups are indicated as follows: *ac*, acetyl; *Cit*, citrullyl; *me*, methyl; *ph*, phosphoryl; *pr*, propionyl; *rib*, ADP-ribosyl; and *Ub*, ubiquityl. Adapted from (Chatterjee and Muir 2010).

2.3.2. DNA methylation

In prokaryotes, DNA methylation at cytosines and adenines is part of the so called restriction-modification system, a defense mechanism against invading foreign DNA [reviewed in (Kobayashi 2001)]. In addition, bacterial DNA methylation plays a role in cell cycle regulation, DNA repair and transcriptional regulation [reviewed in (Marinus and Casadesus 2009)].

In eukaryotes, DNA methylation occurs only at cytosines and is involved in maintaining a repressed chromatin state and stably silencing promoters (Bird and Wolffe 1999; Colot and Rossignol 1999). While in plants and fungi cytosines can be methylated in the context of CpG, CpNpG or even in cytosines followed by any other bases, DNA methylation in animals occurs almost exclusively in the context of CpG dinucleotides. Not only the sequence context but also the extent and global patterns of DNA methylation vary extensively among different eukaryote species.

2.3.2.1. DNA methylation in fungi and plants

While the yeast species *Saccharomyces cerevisiae* and *Saccharomyces pombe* are both entirely devoid of DNA methylation, *Neurospora crassa* shows moderate levels of DNA methylation. In this fungal species, DNA methylation exclusively localizes to relics of transposons that were subject to repeat-induced point mutation, a genome defense system that mutates duplicated sequences (Selker et al. 2003). It was proposed that recruitment of DNA methylation to these loci involves recognition of A:T rich repeated sequences, followed by H3K9 methylation which subsequently triggers binding of HP1 and the DNA methyltransferase DIM-2 (Lewis et al. 2010). A recent genome-wide survey of DNA methylation at base-pair resolution revealed that the DNA methylation pattern observed in *Neurospora crassa* seems to be conserved in many fungal species (Zemach et al. 2010).

The model plant *Arabidopsis thaliana* displays moderate levels of DNA methylation and a mosaic genomic pattern with the mark exclusively occurring at gene bodies, transposons and repetitive elements (Zhang et al. 2006). In contrast to fungi, *de novo* methylation of transposons and repeat elements in plants depends on a RNA-directed mechanism (Wassenegger et al. 1994). This process involves two plant-specific RNA polymerases and several proteins of the RNA interference machinery [reviewed in (Matzke et al. 2009)]. Notably, non-CpG methylation is abundant in plant transposons, with short elements particularly enriched for asymmetric methylation (Zemach et al. 2010). Methylation in gene bodies, on the other hand, occurs almost exclusively at CpG dinucleotides (Lister et al. 2008). Remarkably, the highest enrichment of gene body methylation is found in genes

with moderate expression levels and it has been speculated that this might prevent spurious initiation during transcription elongation (Zilberman et al. 2007). The validity of this model remains however unclear, as in mutants with low levels of DNA methylation the observed increase of anti-sense transcripts is relatively moderate and not correlated to gene-body methylation (Zhang et al. 2006).

From a comparative study of genome-wide methylomes of a variety of different species, it has been suggested that the common ancestor of plants, fungi and animals possessed a mosaic methylation pattern, with methylation at gene bodies and transposons, similarly to what is observed in *Arabidopsis thaliana* (Zemach et al. 2010). Loss of either or both targets of DNA methylation has occurred in many eukaryotic lineages, suggesting that methylation might come with a price in terms of evolutionary fitness.

2.3.2.2. DNA methylation in animals

While it has been reported that *Drosophila melanogaster* shows low levels of DNA methylation specifically during early embryonic development, subsequent studies could not reproduce these results and indicated that this insect species is entirely lacking DNA methylation (Phalke et al. 2009; Schaefer and Lyko 2010; Zemach et al. 2010). Moderate levels of DNA methylation have however been found in the following invertebrates; the honey bee *Apis melifera,* the silk moth *Bombyx mori*, the tunicate *Ciona intestinalis* and the anemone *Nematostella vectensis* (Wang et al. 2006; Zemach et al. 2010). These species show high CpG methylation in gene bodies of expressed genes, but no correlation between promoter methylation and transcription and no evidence of transposon methylation (Zemach et al. 2010).

In contrast to invertebrates, which posses a mosaic methylation pattern, the vast majority of the vertebrate genome is methylated. Methylation occurs in a bimodal distribution; most of the genome is highly methylated (80-100%) and a few regions are unmethyated (0-20%) (Eckhardt et al. 2006). While DNA methylation generally localizes exclusively to CpG dinucleotides, a recent whole genome bisulfite sequencing study showed that non-CpG methylation occurs in pluripotent stem cells and localizes mainly to gene bodies of transcribed genes (Lister et al. 2009). It remains to be determined whether these low levels of non-CpG methylation have any functional role.

How the evolutionary transition from mosaic to global DNA methylation in vertebrates was accomplished is unclear, yet is has been speculated that this change might have benefited the innate immune system (Suzuki and Bird 2008). Additionally, global DNA methylation might have been needed to more efficiently suppress noisy transcription in the

larger genomes of vertebrates (Bird 1995). Given the global methylation of vertebrate genomes, it is difficult to determine whether specific targeting of gene body methylation remained conserved from invertebrates to vertebrates. Evidence for a conservation of this relationship has been provided by studies on X chromosome inactivation. This phenomenon occurs in female mammals and is used to compensate the differing dosage of the X chromosome between females and males [reviewed in (Chow and Heard 2009)]. While on the inactivated X chromosome many promoter regions get hypermethylated, their associated gene bodies generally show less methylation, arguing that transcription in gene bodies and methylation might be linked (Hellman and Chess 2007). Interestingly, the connection between transcription elongation and DNA methylation might depend on direct recognition of H3K36 methylation by one of the mammalian DNA methyltransferases (Dhayalan et al. 2010).

2.3.2.3. The DNA methylation machinery

As described above, fungi, plants and animals show substantial differences in patterns and functional roles of DNA methylation. This variation is also reflected in the proteins involved in writing and reading DNA methylation. Below, I will concentrate on describing the mammalian version of the DNA methylation machinery. Comprehensive descriptions of these processes in plants and fungi can be found elsewhere (Law and Jacobsen 2010; Rountree and Selker 2010).

The first eukaryotic enzyme able to catalyze DNA methylation was purified and cloned by Bestor et al. and was later named DNA methyltransferase 1 (DNMT1) (Bestor et al. 1988). Subsequently, it was shown that purified DNMT1 preferentially acts on a hemimethylated DNA substrate (Yoder et al. 1997). While this *in vitro* study also reported reduced but significant activity on unmethylated DNA, recent structural work showed that DNMT1 contains a loop that prevents *de novo* methylation, confirming that DNMT1 solely acts on hemi-methylated DNA (Song et al. 2011b). Remarkably, the existence of a eukaryotic enzyme that methylates symmetric sites on a hemi-methylated substrate has been postulated more than ten years before the identification of DNMT1 (Holliday and Pugh 1975; Riggs 1975). Recognition and methylation of CpGs in hemi-methylated DNA by DNMT1 provides a simple model how DNA methylation is propagated and stably maintained during mitosis. In accordance with this model, DNMT1 interacts with PCNA, which localizes to sites of DNA replication during S-phase (Chuang et al. 1997). Recognition of hemi-methylated CpGs at replication forks is further aided by UHRF1 (also known as NP95) (Bostick et al. 2007; Sharif et al. 2007; Arita et al. 2008; Avvakumov et al. 2008). The

importance of an interaction between UHRF1 and DNMT1 is demonstrated by the high phenotypic resemblance of *UHRF1*-/- and *DNMT1*-/- mice, with both showing a developmental arrest shortly after gastrulation (Li et al. 1992; Sharif et al. 2007). The first indications that DNA methylation might have an effect on cellular differentiation resulted from studying the effect of a 5-azacytidine, an inhibitor of DNA methylation, on *in vitro* differentiation of myotubes (Jones and Taylor 1980). Remarkably, murine ES cells with a homozygous deletion of *DNTM1* are viable and can be maintained in culture over a long period of time even though their total DNA methylation is dramatically reduced (Li et al. 1992; Lei et al. 1996). However, consistent with the observed early lethality *in vivo*, *in vitro* differentiation of *DNMT1*-/- ES cells is not efficient (Li et al. 1992; Lei et al. 1996; Jackson et al. 2004). Furthermore, it has been reported that induced deletion of *DNMT1* in cultured fibroblasts results in cell death (Jackson-Grusby et al. 2001). The different sensitivity of ES cells and differentiated cell to loss of DNA methylation could be related to transposon control mechanisms. In ES cells and presumably in the early embryo, repression of transposable elements mostly relies on H3K9 methylation, through a mechanism that is independent of DNA methylation (Matsui et al. 2010). In contrast, loss of DNA methylation in somatic cells leads to highly elevated levels of transposable elements, which might have a deleterious effect (Walsh et al. 1998; Jackson-Grusby et al. 2001). Alternatively, the lethal phenotype of *DNMT1-/-* embryos might be attributed to the loss of imprinting or the misregulated inactivation of one of the X chromosomes (Li et al. 1993; Panning and Jaenisch 1996).

First evidence that early embryonic cells also encode an enzyme with *de novo* DNA methyltransferase activity was provided by experiments involving infection of *DNMT1-/-* ES cells with provirus DNA (Lei et al. 1996). This subsequently led to the identification of DNMT3A and DNMT3B, two highly homologous proteins which *in vitro* are able to *de novo* methylate fully unmethylated as well as hemi-methylated substrates (Okano et al. 1998). Only a combined deletion of both *DNTM3* genes leads to impaired *de novo* methylation of inserted proviral DNA, suggesting that DNMT3A and DNMT3B might have overlapping functions in ES cells and early embryos (Okano et al. 1999). However, their expression patterns and functional role seems to be partially distinct. During embryogenesis, *DNMT3B* is specifically expressed in preimplantation stages, while *DNMT3A* expression is detected in an ubiquitous manner from embryonic day 10.5 (E10.5) on (Watanabe et al. 2002). Mice deleted for both *DNMT3* homologues show largely reduced methylation and stop development shortly after gastrulation, similarly to the *DNMT1* mutant. Single deletions of *DNMT3A* and *DNMT3B* lead to less severe phenotypes, suggesting partially overlapping function during embryonic development. However, in accordance with their differential expression, single mutants of *DNMT3A* or *DNMT3B* show distinct phenotypes. Mice with a homozygous deletion of *DNMT3A* develop to term, but die 4 weeks after birth. In contrast,

embryos lacking *DNMT3B* die shortly after E9.5, showing multiple developmental effects, including growth impairment and neural tube defects (Okano et al. 1999). In line with these distinct deletion phenotypes, DNMT3A and DNMT3B vary in their target specificity. Promoter *de novo* methylation during early embryonic development mostly depends on DNMT3B (Borgel et al. 2010). During germ cell development, DNMT3A is required for methylation of most imprinting control regions (ICRs) and of SINEB1 elements, while only DNMT3B is involved in methylation of satellite repeats (Kato et al. 2007).

The DNMT3 protein family includes a third member, DNMT3L, which however does not possess a catalytic domain. DNMT3L is expressed in germ cells and in the early embryo and is needed for methylation of ICRs in both the male and female germ line, as well as for methylation of several repetitive elements in the male germ line (Bourc'his et al. 2001; Bourc'his and Bestor 2004; Kato et al. 2007). Crystallography and biochemical studies revealed that DNMT3L interacts with DNMT3A, with two DNMT3L/3A heterodimers forming a tetramer. This interaction was shown to stimulate *de novo* methylation activity and recruitment to chromatin (Jia et al. 2007; Ooi et al. 2007).

DNMT2 is a strongly conserved protein that is widely distributed among species, even occurring in many species without DNA methylation. Although structure and sequence comparison strongly suggest it to be an active DNA methyltransferase, DNMT2 does not show any catalytic activity on DNA (Goll and Bestor 2005). DNMT2 was however reported to act as a RNA methyltransferase, methylating a cytosine in the anticodon loop of the aspartic acid transfer RNA (Goll et al. 2006; Jurkowski et al. 2008).

2.3.2.4. The DNA methylation pattern in mammals

Initial chromatography-based experiments suggested that the majority of CpGs in mammalian genomes are methylated (Gruenbaum et al. 1981). Usage of methylation sensitive restriction enzymes subsequently led to the identification of a small genomic fraction (1-2%) that is unmethylated in all tissues (Bird et al. 1985). Later, sites of DNA methylation were mapped in a genome-wide manner using a variety of different methods [reviewed in (Zilberman and Henikoff 2007)]. These studies confirmed that DNA methylation levels occur in a bimodal distribution, with most of the genome being highly methylated and a minor proportion of genomic regions showing very low methylation (Weber et al. 2005; Weber et al. 2007; Meissner et al. 2008; Lister et al. 2009). Already the pioneering study by Bird et al. revealed that these unmethylated regions show an elevated CpG density (Bird et al. 1985). Due to this property, these regions were later termed CpG islands (Bird 1986). The occurrence of CpG islands is specific to genomes of vertebrates and a consequence of the

DNA methylation pattern found in these species. As C-to-T transitions occur in a higher frequency when a cytosine is methylated, the global DNA methylation in vertebrates led to an increased loss of CpGs during evolution (Illingworth and Bird 2009). Genomic regions that are unmethylated in the germ line, on the other hand, kept the expected CpG density and therefore appear as CpG islands (Weber et al. 2007). Different methods have been employed to predict CpG islands based on a limited set of sequence criteria such as GC content, ratio of the number of observed CpGs over expected, CpG clustering and region length (Gardiner-Garden and Frommer 1987; Takai and Jones 2002; Hackenberg et al. 2006). Depending on the combination and thresholds of these somewhat arbitrary definition criteria, it was estimated that mammalian genomes contain 30'000 to 200'000 CpG islands, which cover around 1 % of the genome. CpG islands represent a large fraction of gene-regulatory elements as around 70 % of all promoters contain CpG islands (Bajic et al. 2006). Additionally, there is evidence that CpG islands outside of promoters are involved in distal gene regulatory function (Tanay et al. 2007). While initiation at CpG poor promoters mostly relies on the TATA-box-binding protein and starts at a defined nucleotide position, initiation start sites in CpG islands are less strictly defined and fall in a broader sequence region of around 100 bp (Bajic et al. 2006; Carninci et al. 2006; Sandelin et al. 2007). Notably, CpG islands promoters are also more likely to initiate transcription in the antisense direction (Core et al. 2008; Seila et al. 2008). Further, it has been shown that activation of immediate early genes occurs more rapidly when their promoter contains a CpG island (Ramirez-Carrozzi et al. 2009). Together these findings indicate that CpG islands regions are relatively permissive for transcription initiation, a property that might be mediated by their unique chromatin state (Blackledge and Klose 2011). Histone acetylation, for example, is highly abundant at CpG islands irrespective of their association with promoters or transcriptional activity (Roh et al. 2005). Furthermore, unmethylated CpG islands are generally occupied by H3K4me2/me3, a mark that might facilitate binding of proteins that initiate transcription, such as TFIID, the nucleosome remodeler NURF or the HBO1 KAT complex (Wysocka et al. 2006; Mikkelsen et al. 2007; Vermeulen et al. 2007; Weber et al. 2007; Saksouk et al. 2009). A recent study further revealed that CpG islands are enriched for KDM2A, an H3K36me2 specific lysine demethylase (Blackledge et al. 2010). Depletion of H3K36me2 at CpG islands might prevent binding of HDAC complexes, thereby representing another mechanism to keep these regions in open chromatin state (Carrozza et al. 2005). In concordance with their open chromatin structure and less stringent transcriptional control, CpG-rich promoters often control housekeeping gene. However, this represents not a strict correlation, since also around 50% of tissue-specific genes are controlled by CpG-rich promoters. Interestingly, CpG-rich, tissue specific gene promoters are enriched among targets of polycomb group

proteins, indicating that this repression system might be needed to suppress intrinsic transcriptional noise specifically at this promoter class (Mohn and Schubeler 2009).

2.3.2.5. DNA methylation and transcription

The effect of DNA methylation on gene transcription seems to differ between CpG-poor and CpG-rich promoters. Single gene studies suggested that DNA methylation of CpG-poor promoters can preclude transcription (Boyes and Bird 1992; Schubeler et al. 2000). However, measurements of genome-wide DNA methylation revealed that the majority of CpG-poor promoters are methylated in a given cell type even when the associated gene is transcribed (Weber et al. 2007; Meissner et al. 2008; Ball et al. 2009). It was therefore concluded that methylation at this promoter class does not preclude transcription initiation. However, it has recently been shown that transcriptional oscillations at a number of oestrogen responsive, CpG-poor promoters are accompanied by a cyclic gain and loss of DNA methylation. This indicated that transient hypomethylation might be linked to transcription of CpG-poor promoters (Kangaspeska et al. 2008; Metivier et al. 2008). The inconsistency of this result with findings from genome-wide studies could be explained by the fact that only a minority of alleles are actively transcribed in a group of cells (Larson et al. 2009). Thus, by measuring DNA methylation of a cell population one could miss a possible, transient correlation of hypomethylation and transcription at CpG-poor promoters.

While the effect of methylation at CpG-poor promoters requires further investigations, it is clear that DNA methylation of CpG island promoters is not compatible with transcription of the associated gene (Weber et al. 2007). This inhibitory effect of promoter DNA methylation happens at the step of transcription initiation and can be explained by two different models (Schubeler et al. 2000; Appanah et al. 2007). In the first model, DNA methylation precludes binding of methylation-sensitive transcription factors and thereby directly interferes with transcription initiation. Such a mode of action has been reported for several transcription factors including CREB, C-MYC and E2F (Iguchi-Ariga and Schaffner 1989; Prendergast and Ziff 1991; Campanero et al. 2000). In the second model, methylated promoters get recognized by proteins that specifically bind to methylated CpGs and recruit cofactors that in turn repress transcription. A variety of such methyl-CpG-binding proteins (MBPs) are known and get generally divided into two classes: those sharing a methyl-CpG-binding domain (MECP2, MBD1, MBD2 and MBD4) and those which bind via a zinc finger domain (KAISO, ZBTB4 and ZBTB38) (Hendrich and Tweedie 2003; Filion et al. 2006; Clouaire and Stancheva 2008; Dhasarathy and Wade 2008). For most of these proteins, it has been reported that they interact with factors that set up a repressive chromatin

environment. Among others, these include the SIN3A histone deacetylation co-repressor complex, H3K9 KMTs and the chromatin remodeling factors ATRX and BRM1 [reviewed in (Clouaire and Stancheva 2008)]. While such interactions provide a possible mechanism how DNA methylation leads to transcriptional repression, the precise function of MBPs at methylated promoter is still unclear. First of all, mice lacking one or several MBPs show mild and late-onset phenotypes which stands in contrast to the severe developmental phenotypes observed in mice lacking DNMT proteins (Li et al. 1992; Okano et al. 1999; Chen et al. 2001; Guy et al. 2001; Hendrich et al. 2001; Zhao et al. 2003; Martin Caballero et al. 2009). Secondly, the binding specificity of most MBPs has only been studied *in vitro* and genome-wide location studies of most MBPs are still lacking. Interestingly, the only study mapping a MBP revealed that MECP2 binds to most of the genome, thus questioning its role as a gene-specific transcriptional repressor (Skene et al. 2010).

2.3.2.6. Changes of DNA methylation during development

As discussed above, the mammalian genome is globally methylated with only a few regions, including CpG islands, being unmethylated. During mammalian development this methylation profile gets drastically modulated including gene-specific as well as genome-wide changes (Fig. 3). One of two major waves of reprogramming of DNA methylation occurs during germ cell development [reviewed in (Sasaki and Matsui 2008)]. In mouse embryos, primordial germ cells (PGCs) get specified around E7 and then migrate to the genital ridge where they settle by E11.5. As shown by immunofluorescence studies, migration and maturation of PGCs is accompanied by global changes in various chromatin modifications. This reprogramming involves a loss of H3K9me2 and a subsequent decrease in DNA methylation, which is followed by an increase in H3K27me3 (Seki et al. 2005; Seki et al. 2007). A recent bisulfite sequencing study revealed that this loss of DNA methylation is massive, with only around 10% of CpGs remaining methylated (Popp et al. 2010). This study also confirmed earlier findings that intracisternal A particle (IAP) transposons are escaping the global wave of demethylation in PGCs (Hajkova et al. 2008; Popp et al. 2010). It is likely that genome-wide demethylation in PGCs restricts the potential involvement of DNA methylation in transgenerational epigenetic inheritance. In line with that notion, the only well-studied example of transgenerational inheritance in mice involves differential methylation of an IAP element (Whitelaw and Whitelaw 2008). Besides a potential role in erasing epigenetic modifications from the previous generation, global demethylation in PGCs is needed to reset these cells to a more pluripotent state. It has further been shown that demethylation is crucial for reactivation of a set of germ-cell specific genes during development from the pre- to the postmigratory stage (Maatouk et al. 2006). In addition, methylation reprogramming is

crucial for resetting parental imprints, which serve as epigenetic marks that ensure parental-origin-specific monoallelic expression of around hundred imprinted genes in the next generation (Sasaki and Matsui 2008).

After cell-intrinsic and -extrinsic factor have initiated sex determination at E12.5, germ cells progressively reestablish genome-wide DNA methylation (Lin et al. 2008; Kota and Feil 2010). While the extent and timing of this global *de novo* methylation is unclear, it is crucial for meiosis, as male germ cells lacking DNMT3L show elevated expression levels of retrotransposons and meiotic failure (Bourc'his and Bestor 2004). *De novo* methylation is also crucial in female germ cells, since female mice with a deletion of the chromatin remodeler *LSH* show disrupted meiosis which is likely caused by reduced DNA methylation and elevated levels of retrotransposon transcription (De La Fuente et al. 2006). The wave of *de novo* methylation in germ cells is accompanied by reestablishment of parental imprints in a gender-specific manner. In the male germ line, imprints at three different well studied loci get reestablished between E14.5 and the newborn stage (Li et al. 2004). It was shown that methylation at these differentially methylated regions (DMRs) of *H19* and *DLK1/GTL2* depends on DNMT3A and DNMT3L, while the DMR at *RASGRF1* requires action of all three DNMT3 proteins. The fact that the latter DMR contains a sequence derived from an endogenous retrovirus might explain the requirement for DNMT3B, since this enzyme is also needed for methylation of IAP retrotransposon methylation (Kato et al. 2007; Watanabe et al. 2011). In the female germ line, various DMRs get methylated after birth during the phase of oocyte growth (Lucifero et al. 2004). While DNMT3A and DNMT3L are essential for maternal imprinting, DNMT3B seems to be dispensable (Bourc'his et al. 2001; Kaneda et al. 2004). The phenotype of *LSH* mutant mice indicates that besides its role in imprinting, DNA methylation is also crucially involved in retrotransposon silencing in oocytes (De La Fuente et al. 2006).

Shortly after fertilization a second wave of genome-wide demethylation occurs. While demethylation occurs rapidly and prior to DNA replication in the paternal pronucleus, the maternal genome gets demethylated in a more progressive manner (Mayer et al. 2000; Oswald et al. 2000; Santos et al. 2002). Since these results are mostly based on immunostainings, the extent and targets demethylation are still unclear (Reik 2007). It is thought that global demethylation after fertilization facilitates the reprogramming of mature gametes, which are highly specialized cells, to a pluripotent state. In line with this, several factors that regulate pluripotency in the preimplantation embryo were shown to be methylated in sperm cells (Imamura et al. 2006; Farthing et al. 2008). Importantly, there are also sequences that resist demethylation after fertilization, including IAPs and parental imprints at DMRs (Reik 2007). While, two different proteins, STELLA and ZFP57 were

shown to be involved in protection from demethylation, their mode of action still remains to be determined (Nakamura et al. 2007; Li et al. 2008).

Immunostaining experiments indicated that global DNA methylation levels get restored in the embryo at the blastocyst stage (Dean et al. 2001). This was recently confirmed by a microarray study that determined DNA methylation levels in 11 kb region around promoters during this stage of embryonic development (Borgel et al. 2010). In E3.5 blastocysts all examined regions were unmethylated, with the exception of genes carrying germline DMRs and a small set of gene promoters, which are enriched for genes expressed in the male germ line. In E6.5 epiblast, DNA methylation was restored at intergenic, intragenic and CpG-poor promoters, showing that the genome gets globally methylated during the implantation stage. While most CpG-rich promoters remain unmethylated in E6.5 epiblast cells and in E9.5 embryonic tissue, a subset of this promoter class was found to get *de novo* methylated in a DNMT3B dependent manner (Borgel et al. 2010). Borget at al. showed that these *de novo* methylated promoters are enriched for germline and pluripotency-specific genes, a finding that confirmed earlier studies using ES cells (Farthing et al. 2008; Mohn et al. 2008). Although being isolated from the inner cell mass of E3.5 blastocysts, *in vitro* cultured ES cells seem to recapitulate some of the *de novo* DNA methylation that happens *in vivo.* ES cells therefore show a DNA methylation pattern similar to that of epiblast cells, however without methylation of pluripotency genes (Borgel et al. 2010). When comparing ES cell with differentiated cells, it was found that DNA methylation seems to control a relatively small set of promoters (Farthing et al. 2008; Meissner et al. 2008; Mohn et al. 2008). Changes in differentiation-induced promoter DNA methylation are mainly unidirectional, with more *de novo* methylation than demethylation. Furthermore, genes of factors that are involved in maintaining the pluripotent state of ES cell are highly enriched among *de novo* methylated promoters. It has therefore been speculated that stable promoter repression by DNA methylation during embryonic development helps to restrict developmental potential and stabilizes cellular identity (Reik 2007; Guibert et al. 2009). In line with this model, DNA methylation of the *ELF5* promoter is critically preventing cells of the embryonic lineage to enter the trophoblast lineage (Ng et al. 2008). Moreover, DNA methylation seems to be a major barrier for reprogramming of differentiated cells to a pluripotent state (Mikkelsen et al. 2008). Similarly, it was found that *in vitro* differentiated ES cells that fail in methylating promoters of pluripotency factors can be more easily reverted to a pluripotent state (Epsztejn-Litman et al. 2008). DNA methylation of pluripotency factors might safeguard these genes from expression in somatic cells, as this could lead to dedifferentiation and cause a predisposition to cancer (Hochedlinger et al. 2005). Since the kinetics of repression has been studied for only a few promoters that get *de novo* methylated, it is currently unclear whether DNA methylation is directly triggering repression

of these genes or rather gets accumulated after silencing (Guibert et al. 2009). In case of the *OCT4* promoter, *de novo* methylation is one of the last steps in repression, occurring after loss of transcription factor binding, histone deacetylation and gain in H3K9 methylation (Feldman et al. 2006). On the other hand, loss of DNMT3B, which is the main enzyme responsible for promoter *de novo* methylation during implantation leads to improper activation of several normally methylated testis-specific genes in E9.5 embryos, showing that DNA methylation is directly involved in repression of these genes (Borgel et al. 2010).

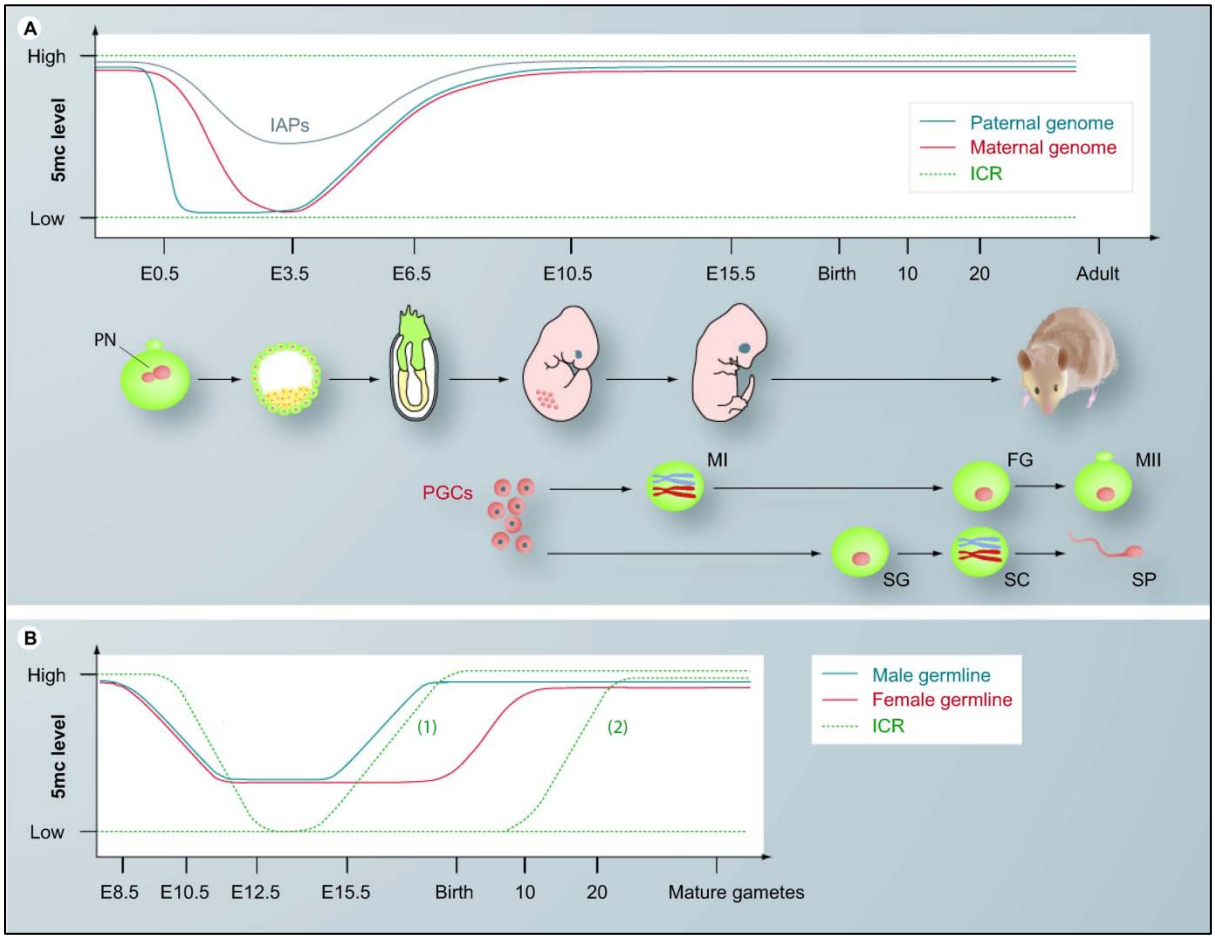


Figure 3. **Reprogramming of DNA methylation during mouse development.**
(**A**) Paternal (blue line) and maternal (red line) genomes are demethylated asynchronously during preimplantation development and reach a low point of methylation at the blastocyst stage (E3.5). IAP elements (gray line) partially resist this demethylation wave. After implantation, global DNA methylation patterns are restored. ICRs (green lines) escape global reprogramming during somatic development: the methylated allele resists global demethylation in preimplantation embryos, whereas the unmethylated allele resists *de novo* methylation in post-implantation embryos. (**B**) PGCs gradually loose global DNA methylation until E12.5. Methylation patterns are re-established at different developmental time points in male and female germline. In male germ cells (blue line), methylation patterns are regained during the maturation to spermatogonia between E14.5 and birth. Methylation of paternal ICRs (green line 1) occurs at the same time. In female embryos (red line), germ cells arrest in meiotic prophase around E13.0 and initiation of DNA methylation occurs after birth during oocyte growth. Methylation of maternal ICRs in the female germline (green line 2) occurs between 10 and 25 days after birth and is completed in fully grown oocytes.

PN: Pronuclei; FG: Fully grown oocyte; IAP: Intracisternal A-particle; ICR: Imprinting control region; MI: Meiosis I; MII: Meiosis II; PGC: Migrating primordial germ cell; SC: Spermatocyte; SG: Spermatogonia; SP: Spermatid.

Adapted from (Guibert et al. 2009).

2.3.2.7. The establishment of DNA methylation patterns

2.3.2.7.1. Keeping CpG islands unmethylated

Soon after the discovery of CpG islands in mammalian genomes, several studies addressed how these regions are kept unmethylated. Integration of multi-copy arrays of the THY-1 gene in fertilized eggs revealed that the unmethylated state of the *THY-1* CpG island promoter was recapitulated in all adult tissues, suggesting a *cis*-acting mechanism (Kolsto et al. 1986). Later it was shown that ES cells are able *de novo* methylate randomly integrated foreign DNA, which is in accordance with the wave of *de novo* methylation observed during early embryonic development. Based on genomic integration in ES cells it was found that a 214 bp fragment in the *THY-1* promoter region could protect from DNA methylation (Szyf et al. 1990). Subsequently, similar experiments revealed that binding sites of the transcription factor SP1 were crucial for the unmethylated state of the CpG island promoter of *APRT* (Brandeis et al. 1994; Macleod et al. 1994). However, it remained unclear whether elevated DNA methylation levels were directly linked to a loss of the SP1 binding sites or a consequence of reduced transcriptional activity. A direct involvement of SP1 in regulating DNA methylation was further questioned by the finding that *SP1* knock-out mice do not show elevated DNA methylation at CpG islands, including the *APRT* promoter (Marin et al. 1997). These contradictory results were recently resolved by showing that the SP1 binding site in the *APRT* promoter is bound by VEZF1, a zinc finger protein VEZF1 that acts as crucial mediator of the unmethylated state at the chicken β-globin enhancer (Dickson et al. 2010). In line with a role of transcription factors in defining hypomethylation of CpG islands, an algorithm, based on TF binding sites that are enriched in these genomic regions, could predict many hypomethylated regions (Straussman et al. 2009). However, a mechanistic model how TF binding could lead to protection from *de novo* DNA methylation is still lacking.

Based on experiments using the *E.coli* lac repressor/operator system it has been suggested protein binding itself might protect from DNA methylation in mammalian cells (Han et al. 2001). On the other hand, the unique chromatin features of CpG islands could play a role as regulators of DNA methylation levels (Blackledge and Klose 2011). Particularly H3K4 methylation, which strongly correlates with unmethylated CpG islands, has been mechanistically linked to DNA methylation. By crystallography it was shown that DNMT3L recognizes the histone H3 tail and that this binding is strongly inhibited by H3K4 methylation (Ooi et al. 2007). Recent studies indicated that the ADD domains of DNMT3A and 3B show the same binding specificity (Otani et al. 2009; Zhang et al. 2010). Furthermore, ectopic expression of DNMT3s in *S. cerevisiae* defined that the N-terminal part of histone H3 tail is needed for *de novo* methylation of chromatin. DNA methylation occurred in regions lacking H3K4 methylation and mutants without K4 methylation showed elevated DNA methylation

levels (Hu et al. 2009). While these experiments provide evidence that H3K4 methylation could mediate protection from DNA methylation, it remains unclear how this histone modification is targeted to CpG islands. One potential candidate involved in this process is CXXC finger protein 1 (CFP1), which was found to bind unmethylated CpGs *in vitro* and to be part of the SETD1 H3K4 methlytransferase complex (Voo et al. 2000; Lee and Skalnik 2005). Recent genome-wide mapping revealed that CFP1 exclusively binds to H3K4 methylated regions in the mouse brain. Moreover, knock-down of CFP1 led to reduced levels of H3K4me3 at CFP1 occupied regions (Thomson et al. 2010). Interestingly, the CXXC domain also occurs in a number of H3K4 KMTs, suggesting that these enzymes could directly bind to unmethylated CpG islands (Voo et al. 2000).

2.3.2.7.2. Targeting of *de novo* methylation

While most CpG-rich regions in the mammalian genome remain unmethylated in all cell types, a subset of these sequences gets *de novo* methylated during development (see previous chapter). There are different models how targets of *de novo* methylation are defined. Except for a tendency to methylate CpGs with an interval of eight to ten base pairs, DNMT3 enzymes themselves do not show specific sequence preferences (Jia et al. 2007). It has therefore been suggested that DNMT3 proteins might be directly recruited to their targets by binding to sequence specific TFs, such as MYC, PU.1 or GCNF (Brenner et al. 2005; Sato et al. 2006; Suzuki et al. 2006). However, for these factors, a direct induction of *de novo* methylation has not been shown genetically. Recent experiments, on the other hand, provided strong evidence that the transcription factor E2F6 is involved in maintenance of DNMT3B-mediated DNA methylation and in silencing of a group of germ-line-specific genes (Velasco et al. 2010). Of note, protein array experiments indicated that DNMT3 enzymes interact with many more TFs, suggesting that direct recruitment of *de novo* methyltransferase activity could be a wide spread mechanism (Hervouet et al. 2009).

Alternatively, it has been suggested that DNMT3 proteins could be targeted by noncoding RNAs. For instance, it was recently reported that a small RNA in the promoter of rRNAs can form a DNA:RNA triplex which is specifically recognized by DNMT3B (Schmitz et al. 2010). While binding of DNMT3B to this unusual nucleotide structure awaits further investigations, several recent reports demonstrated a role of small RNAs in DNA methylation mediated transposon silencing. This mechanism involves MILI and MIWI2, two proteins of the PIWI subfamily of Argonaute proteins. MILI and MIWI2 are expressed during gametogenesis and interact with PIWI-interacting RNAs, small RNAs that are derived from retrotransposons. Deletion of *MILI* and *MIWI2* leads to impaired *de novo* methylation of IAP and LINE-1 elements in male germ cells, causing to derepression of these repeat elements

(Carmell et al. 2007; Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008). Remarkably, in the male germ line the PIWI system is also needed for *de novo* methylation of the *RASGRF1* DMR, in a mechanism that seems to be unique to this imprinted locus (Watanabe et al. 2011).

A recent study suggested a link to transcription as an alternative mechanism of DNA methylation targeting. Chotalia et al. showed that methylation of the *GNAS* DMR in the maternal germ line depends on transcription of a protein-coding transcript through this locus. Notably, the authors reported that many more maternally methylated DMRs are positioned within transcription units, suggesting that this mechanism might not be limited to the *GNAS* locus (Chotalia et al. 2009). It will be interesting to see how this phenomenon relates to observations that DNA methylation is enriched in transcribed gene bodies and that intragenic CpG-rich promoters might be regulated by DNA methylation in somatic cells (Maunakea et al. 2010).

Finally, several different histone modifying enzymes, and the respective marks they set, might play a role in targeting of DNA methylation (Cedar and Bergman 2009). This includes H3K27me3, which in ES cells marks promoters that show an increased frequency of *de novo* DNA methylation during *in vitro* differentiation (Mohn et al. 2008). In accordance with this correlation, the polycomb group protein EZH2, which catalyses H3K27me3, was reported to directly interact with DNMT proteins (Vire et al. 2006). G9a (EHMT2), which catalyses H3K9me2, represents a second histone modifying enzyme with a role in recruiting DNA methylation. Overexpression of a catalytically inactive form of G9a in G9a knock-out cells revealed that G9a induces DNA methylation at repeat elements and at several promoters during ES cell differentiation independently of its KMT activity (Dong et al. 2008; Epsztejn-Litman et al. 2008). It was recently reported that symmetric dimethylation of H4R3 directly recruits DNMT3A at the γ-globin promoter (Zhao et al. 2009). However, the binding of the DNMT3A ADD domain to symmetric H4R3me2 could not be reproduced in an unbiased peptide binding screen (Zhang et al. 2010). Finally, H3K4 methylation represents a likely histone modification involved in regulating *de novo* DNA methylation, given its inhibitory influence on binding of DNMT3 proteins to chromatin (see above). Indeed, it has recently been shown that the H3K4 demethylase AOF1, which is almost exclusively expressed during oogenesis, is required for *de novo* methylation of some imprinted genes in oocytes (Ciccone et al. 2009). It is still open whether *de novo* methylation of promoters during somatic development might as well involve action of H3K4 demethylases. Furthermore, it remains to be determined how histone modifying enzymes that potentially induce *de novo* DNA methylation are recruited specific targets.

2.3.2.7.3. Demethylation

DNA demethylation can either happen through a passive mechanism, by progressive loss owing to an absence of maintenance of CpG methylation during cell replication, or through active demethylation. The global wave of demethylation in the female pronucleus is thought to occur in a passive manner, involving retention of DNTM1 in the cytosol (Cardoso and Leonhardt 1999). As global demethylation in the paternal pronucleus happens more rapid and prior to replication, the existence of an active demethylation process of yet unknown mechanism has been proposed (Mayer et al. 2000). Similarly, global demethylation during germ cell development might involve an active mechanism, given the slow cell cycle of PGCs and their rapid loss of DNA methylation (Gehring et al. 2009).

During somatic development, changes in DNA methylation are biased towards *de novo* DNA methylation and only few cases of demethylation have been reported. Interestingly, a group of genes that gets *de novo* methylated during early embryogenesis is active in distinct tissues in the adult mouse, implying that some demethylation must happen in a lineage-specific manner (Borgel et al. 2010). In line with this possibility, Klug et al. identified several genomic loci that get demethylated during differentiation of human monocytes into macrophages and dendritic cells. The fact that these cells are post-mitotic implies an active mechanism of demethylation (Klug et al. 2010).

Active demethylation has been proposed to operate via several different mechanisms (Ooi and Bestor 2008). While initial studies reported enzymes that directly remove the methyl-group from cytosines, their findings could not be reproduced and such a mechanism of demethylation therefore remains unlikely. An alternative model involves removal of entire nucleotide stretches by the nucleotide excision repair (NER) machinery. A screen for proteins that could reactivate a methylation-silenced reporter identified GADD45A as mediator of active demethylation through the NER pathway (Barreto et al. 2007). However, a similar study could not reproduce these results and a recent analysis found no DNA methylation changes in *GADD45A* mutant mice (Jin et al. 2008; Engel et al. 2009). A third model of active DNA demethylation proposes that proteins of the base excision repair (BER) machinery induce deamination of 5mC into thymine, followed by T-G mismatch repair. Candidate cytosine deaminases include AID and APOBEC1, which are expressed in PGCs and during early embryogenesis and are able to deaminate 5mC *in vitro* (Morgan et al. 2004). A recent study indeed revealed that PGCs show elevated DNA methylation levels in *AID* deleted mice (Popp et al. 2010). Furthermore, it was reported that AID is essential for cell-fusion induced reprogramming of a somatic genome to a pluripotent state, possibly by mediating demethylation of pluripotency genes (Bhutani et al. 2010). A recent study

supported a role of AID in active demethylation by showing that overexpression of AID, together with the thymine glycosylase MBD4, leads to reduced DNA methylation in zebrafish (Rai et al. 2008). Next to MBD4, TDG represents a second G:T mismatch-specific thymine glycosylase with a role in BER mediated DNA demethylation. Interestingly, TDG is the only DNA glycosylase whose deletion in mice results in embryonic lethality. Further, it was shown that lack of TDG has a minor effect on the repair of canonical base damage, but leads to increased DNA methylation at a set of promoters in mouse embryonic fibroblasts (Cortazar et al. 2011).

Several recent reports identified 5-hydroxmethylcytosine (5hmC) as a new potential factor involved in demethylation processes. 5hmC, albeit at low levels, was found to specifically occur in ES cells and cerebellar neurons and to result from hydroxylation of 5mC by the family of TET proteins (Kriaucionis and Heintz 2009; Tahiliani et al. 2009). First maps of 5hmC in the cerebellum revealed that this mark mainly occurs in gene bodies of active genes where its levels correlate with gene expression levels (Song et al. 2011a). Interestingly, recent mapping studies in ES cells showed that in this cell type 5hmC additionally occurs at a subset of CpG-rich promoters. These are enriched for polycomb targets and for promoters that gain 5mC during development, arguing for 5mC/5hmC turnover in differentiation (Pastor et al. 2011; Wu et al. 2011). It has been suggested that 5hmC might be involved in removal of 5mC, either by inhibiting maintenance of 5mC or by acting as substrate during BER mediated active demethylation (Guo et al. 2011). ChIPseq analysis revealed that TET1 occupies the majority of CpG islands, implicating an involvement in regulating the unmethylated state of these genomic regions. However, knock-down of TET1 resulted in a rather small increase of 5mC at a small set of promoters. Furthermore, TET1 seems to influence gene expression independently of its catalytically activity (Pastor et al. 2011). The role of TET1 and 5hmC in mediating 5mC methylation therefore remains elusive.

2.4. Scope of thesis

While sharing the same genetic information, different cell types of an individual organism show an enormous variety in shape and function. This variety results from transcription of the invariable genetic code into cell-specific gene expression programs, specified by a set of transcription factors that are unique to each cell type (Vaquerizas et al. 2009). Over the last 30 years it became clear that this transcriptional control by sequence specific TFs is strongly influenced by chromatin-based regulatory mechanisms.

When I started my PhD project in 2007, advances in sequencing technology led to the first whole genome maps of chromatin modifications in human cells (Barski et al. 2007). While such studies confirmed that a set of chromatin modifications correlates with gene repression, several aspects of their role in transcriptional silencing remained unclear. Are these chromatin modifications directly involved in transcriptional silencing? How dynamic is the genomic distribution of repressive marks and how are these patterns established? We here set out to address these questions by studying the targeting of H3K9me2 and DNA methylation during cellular differentiation.

While studies on G9a and GLP, which catalyze H3K9me2, suggested that this mark is involved in setting up a repressive chromatin state, the genomic pattern of H3K9me2 was unknown (Tachibana et al. 2005). We therefore determined the genome-wide distribution of H3K9me2 in ES cells by using chromatin-immunoprecipitation in combination with microarray technology (ChIP on chip). In order to relate H3K9me2 occupancy to an active and an additional repressive chromatin modification, we generated maps of H3K4me2 and H3K27me3 by ChIPseq technology. With the aim of studying the dynamics of repressive heterochromatin we further expanded our H3K9me2 ChIP on chip experiments to post-mitotic neurons, which were derived through *in vitro* differentiation of ES cells by employing a protocol that generates highly pure cell populations (Bibel et al. 2007). We then related the H3K9me2 occupancy in these two cell types to changes in gene expression as measured by RNAseq. Together these experiments allowed us to test a prevalent model in stem cell biology which suggests that loss of pluripotency entails a massive increase in heterochromatin, which in turn results in reduced transcription in differentiated cells (Meshorer and Misteli 2006).

In order to study how repressive chromatin modifications patterns are established, we focused our work on DNA methylation. By then it was known that DNA methylation can be epigenetically inherited and has an important role in silencing of repeats, genomic imprinting and X-chromosome inactivation (Walsh and Bestor 1999). Ongoing work in the lab further showed that promoter DNA methylation is dynamic during cellular differentiation and gets acquired *de novo* at a subset of genes (Mohn et al. 2008). As sequence analysis and

correlations to other chromatin modifications were not sufficient to predict the observed DNA methylation patterns, we chose to address this question using a different experimental approach. In order to determine whether promoter methylation is guided by the underlying DNA sequence, we chose a set of promoter regions with different DNA methylation patterns and inserted them at an ectopic site in the genome of ES cells. For that we made use of an insertion method that allows sequential insertions at the same genomic site without need of a transcribed marker gene (Seibler et al. 1998). By following DNA methylation levels at inserted promoter sequences during cellular differentiation we further aimed at studying the targeting mechanism of this repressive chromatin mark.

3. Results

3.1. Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells

Lienert F.*, Mohn F.*, Tiwari V.K., Baubec T., Roloff T.C., Gaidatzis D., Stadler M.B. and Schübeler D.

3.1.1. Summary

Cell-specific gene expression programs are defined by transcription factors and involve processes that depend on chromatin modifications. It has been proposed that pluripotent stem cells possess a unique chromatin structure that is characterized by low abundance of repressive heterochromatin and coinciding global transcription. In this model, loss of pluripotency entails a global increase in heterochromatin and coinciding shutdown of lineage unrelated genes (Gaspar-Maia et al. 2011).

We here tested this model by profiling the distribution of heterochromatic H3K9me2 in pluripotent ES cells and derived post-mitotic neurons, using a well-defined differentiation protocol (Bibel et al. 2004). Our analysis revealed that H3K9me2 is highly abundant in ES cells and occurs in large domains that occupy more than half of the genome. H3K9me2 marks all genomic regions that are devoid of transcription, H3K4me2 and polycomb mediated H3K27me3. This suggests that H3K9me2 is set by default and might be involved in establishment of a repressive chromatin state outside of regions that are marked by active chromatin modifications or regulated by polycomb. Importantly, abundance of H3K9me2 increases only slightly during differentiation of ES cell into neurons, with a localized gain occurring at gene bodies of transcribed genes. Moreover, gene expression profiling by RNAseq does not reveal evidence for a reduction of transcriptome complexity as stem cells become terminally differentiated neurons.

Based on these data and on results from many recent studies we suggest that pluripotent embryonic stem cells are not per se unique in regards to heterochromatin abundance and transcriptional plasticity as compared to somatic cells. Accordingly, differentiation entails focal, rather than global, changes in repressive chromatin, which might help to stabilize cellular states at any developmental stage (Farthing et al. 2008; Meissner et al. 2008; Mohn et al. 2008; Lister et al. 2009; Hawkins et al. 2010; Min et al. 2011).

OPEN ACCESS Freely available online

PLoS GENETICS

Genomic Prevalence of Heterochromatic H3K9me2 and Transcription Do Not Discriminate Pluripotent from Terminally Differentiated Cells

Florian Lienert[◔], Fabio Mohn[◔¤], Vijay K. Tiwari, Tuncay Baubec, Tim C. Roloff, Dimos Gaidatzis, Michael B. Stadler, Dirk Schübeler*

Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

Abstract

Cellular differentiation entails reprogramming of the transcriptome from a pluripotent to a unipotent fate. This process was suggested to coincide with a global increase of repressive heterochromatin, which results in a reduction of transcriptional plasticity and potential. Here we report the dynamics of the transcriptome and an abundant heterochromatic histone modification, dimethylation of histone H3 at lysine 9 (H3K9me2), during neuronal differentiation of embryonic stem cells. In contrast to the prevailing model, we find H3K9me2 to occupy over 50% of chromosomal regions already in stem cells. Marked are most genomic regions that are devoid of transcription and a subgroup of histone modifications. Importantly, no global increase occurs during differentiation, but discrete local changes of H3K9me2 particularly at genic regions can be detected. Mirroring the cell fate change, many genes show altered expression upon differentiation. Quantitative sequencing of transcripts demonstrates however that the total number of active genes is equal between stem cells and several tested differentiated cell types. Together, these findings reveal high prevalence of a heterochromatic mark in stem cells and challenge the model of low abundance of epigenetic repression and resulting global basal level transcription in stem cells. This suggests that cellular differentiation entails local rather than global changes in epigenetic repression and transcriptional activity.

Citation: Lienert F, Mohn F, Tiwari VK, Baubec T, Roloff TC, et al. (2011) Genomic Prevalence of Heterochromatic H3K9me2 and Transcription Do Not Discriminate Pluripotent from Terminally Differentiated Cells. PLoS Genet 7(6): e1002090. doi:10.1371/journal.pgen.1002090

Editor: Wolf Reik, The Babraham Institute, United Kingdom

Received January 21, 2011; **Accepted** March 17, 2011; **Published** June 2, 2011

Funding: FL and FM were supported by a PhD fellowship of the Boehringer Ingelheim Fonds. VKT is supported by Marie Curie International Incoming fellowship and an EMBO long-term postdoctoral fellowship. TB is supported by an EMBO longterm postdoctoral fellowship. Research in the laboratory of DS is supported by the Novartis Research Foundation, by the European Union (NoE "The Epigenome" LSHG-CT-2004-503433, LSHG-CT-2006-037415), the European Research Council (ERC-204264), SystemsX.ch, the Swiss initiative in Systems Biology, and the EMBO Young Investigator program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dirk@fmi.ch

¤ Current address: Institute of Molecular Biotechnology, Vienna, Austria

◔ These authors contributed equally to this work.

Introduction

Resetting of the transcriptional program is the key driver for cell type specification during organismal development [1,2]. While embryonic stem (ES) cells bear the fascinating ability to acquire very diverse fates, derived somatic stages are usually irreversible under physiological conditions. This unidirectionality has been suggested to depend in part on epigenetic repression of lineage unrelated genes [3,4]. Accordingly, ES cell plasticity was suggested to rely on a low prevalence of heterochromatin and coinciding promiscuous low-level expression of many genes in stem cells [5–11]. In line with this model, distinct changes in nuclear staining had previously been observed by electron microscopy during cellular differentiation [12,13]. Further, a subset of promoters was shown to become DNA methylated [14–16] and the repressive histone modifications H3K27me3 and H3K9me3 were reported to locally expand in differentiated cells [9].

Here, we set out to test the model of widespread heterochromatinization via monitoring of the differentiation-coupled dynamics of H3K9me2, a repressive epigenetic modification, which appears to be the most abundant heterochromatic modification and has recently been reported to cover large domains in differentiated cells [17]. Unexpectedly, we found that H3K9me2 is not only highly abundant in terminally differentiated cells, but already occupies large parts of the genome in pluripotent stem cells. In this cellular state, H3K9me2 occupies most genomic regions devoid of transcription and certain histone modifications. While our analysis revealed discrete local changes particularly at gene bodies, we observed little global increase in H3K9me2 during differentiation. This unexpected finding motivated us to revisit the model of promiscuous low-level gene expression in undifferentiated cells by quantitative RNA sequencing. Remarkably, we found the actual number of low-level expressed genes, postulated hallmarks of stem cells to be equal between both developmental states. Together, our findings challenge the model of promiscuous basal gene expression as a distinct property of pluripotency and a widespread increase of heterochromatin during cellular differentiation.

Author Summary

Epigenetic modifications of DNA and bound histones are major determinants of cell type–specific gene expression patterns. A prevalent model in stem cell biology suggests that the loss of pluripotency entails global increase in heterochromatin and coinciding shutdown of lineage unrelated genes. We performed analysis of both H3K9 dimethylation pattern and the global transcriptome in an advanced murine neuronal differentiation model. In this paradigm, we do not find evidence for a global increase in heterochromatic H3K9 dimethylation or reduction of transcriptome complexity as stem cells become terminally differentiated post-mitotic neurons. This suggests that pluripotent embryonic stem cells are not *per se* unique in regards to heterochromatin abundance and transcriptional plasticity as compared to somatic cells. Instead, focal changes in chromatin might help to stabilize cellular states at any developmental stage.

Results

H3K9me2 is nearly invariant and only displays distinct local changes between developmental stages

To asses differentiation associated dynamics of the repressive histone modification H3K9me2 we made use of a highly pure and robust murine *in vitro* neurogenesis model [18], which we previously used to profile histone and DNA methylation [14]. Here, we generated profiles for H3K9me2 in pluripotent embryonic stem cells and derived terminally differentiated pyramidal neurons. We made use of custom tiling arrays covering 10% of the mouse genome including all well-annotated promoters, several large multi-gene loci and the complete chromosome 19 (see Figure S1 and Text S1). The chromosomal profiles for H3K9me2 revealed domains of enrichments that upon visual inspection were highly comparable between stem cells and the neuronal state (Figure 1A), which is further supported by a high overall pair-wise correlation (Figure 1B). Despite this overall similarity we noticed confined regional differences (Figure 1A), a finding which is consistent with the fact that biological replicates of H3K9me2 are more similar than the patterns between cell states (Figure 1B). We also included in our comparison a recently published dataset for H3K9me2 in a distinct ES cell line [17], which shows high correlation to our ES cell datasets despite different experimental conditions (Figure S2). Of note, analysis of the H3K9me2 dataset from Wen et al. [17] revealed that chromosome 19 behaves similar to the other chromosomes (Figure S3), suggesting that our results can be extrapolated to the entire genome. Together this demonstrates that our H3K9me2 data are reproducible and of high resolution, yet overall patterns appear to be highly similar between a pluripotent and a terminally differentiated state. Visual inspection suggests that H3K9me2 covers large domains in both ES cells and neurons (Figure 1A). To quantitatively define the actual location and sizes of domains we applied a Hidden-Markov-Model (HMM) analysis to the microarray data. This unsupervised statistical method is a widely accepted approach for unbiased data segmentation in epigenome analysis [19,20]. The HMM analysis not only agreed with and statistically corroborated the visual impression of the raw data, but also yielded robust results under variable settings (Figure S4). It revealed that over 50% of chromosome 19 is covered by H3K9me2 in ES cells (Figure 1C). Using the same approach for the H3K9me2 data in the ES cell-derived terminally differentiated neurons, we detected a modest yet reproducible 5% increase of genomic regions covered by H3K9me2 (Figure 1C). We conclude that global coverage and size of H3K9me2 domains is nearly identical between ES cells and derived post-mitotic pyramidal neurons. In line with this finding we do not detect a significant change in global H3K9me2 levels by Western blot detection (Figure 1D). Moreover, H3K9me2 domain features of ES cells and neurons show similar size distribution and median length (Figure S4). Further analysis revealed that changes in H3K9me2 between the two examined cellular states are rare; 88% of H3K9me2 occupied regions in ES cells are also occupied in neurons (Figure 1F). Notably, regions that change in H3K9me2 state tend to be small and are below the average size of invariant domains (Figure S5). Consistent with the overall increase of 5% in H3K9me2 coverage during differentiation, regions which gain H3K9me2 are more frequent and of larger size than regions showing a loss of the mark (Figure S5 and Figure S6). Interestingly, most of the larger regions (>10 kb) that gain H3K9me2 are located within genes, starting downstream of the promoter region (Figure 2A and Figure S5). These global findings are fully reproducible in single gene controls (Figure 2B) and consistent with a focused comparison of only genic regions (Figure 2C). Importantly, this shows that our experimental and data analysis approach is indeed highly sensitive to detect differences if they do occur. Interestingly, many genes that acquire H3K9me2 show slightly reduced expression in many cases, while others increase expression upon gain of the modification (Figure 2D and Figure S5). This suggests that the gain of H3K9me2, while highly selective for gene bodies, cannot simply be explained by the silencing of gene activity.

H3K9me2 and H3K27me3 are mutually exclusive

Given the high prevalence of H3K9me2, we next asked how its presence relates to a distinct repressive chromatin modification, namely trimethylation of H3K27 (H3K27me3). This mark is set by the Polycomb pathway and often occurs in domains of several kilobases [9,21,22]. We find that both heterochromatic histone modifications occur mutually exclusive even when in direct neighborhood as illustrated by the sharp boundaries of the H3K9me2 signal next to H3K27me3 peaks (Figure 3A, 3B and 3C). This is consistent with a previous study in human embryonal carcinoma cells that was limited to promoters [23].

H3K9me2 is largely exclusive with active chromatin

We further related H3K9me2 occupancy to regions with transcriptional activity or presence of the active modification H3K4me2. Active regions are mutually exclusive with H3K9me2 in ES cells but surprisingly to a lesser extent in neurons (Figure 3D). The compatibility of H3K9me2 and gene expression in neurons is however limited to gene bodies and does not occur in the promoters of expressed genes, consistent with the former regions gaining H3K9me2 during differentiation (Figure 2A and Figure S7). We find the majority of H3K4me2 regions to be mutually exclusive with H3K9me2 in stem cells (Figure 3D). In neurons, a small number of regions become co-occupied, again most of these being within transcribed genes (Figure 3D). Importantly, an HMM independent analysis confirms that regions with high H3K9me2 enrichment do not overlap with transcribed genes in stem cells, yet a subset does in neurons (Figure 3E). We conclude that gain of H3K9me2 during differentiation has only a minor effect on the overall chromosomal coverage of the modification, yet it occurs highly localized and preferentially at genic regions.

Prevalent low-level transcription is not a stem cell–specific feature

Our finding of surprising conservation of heterochromatin patterns in a refined model of differentiation let us to revisit the

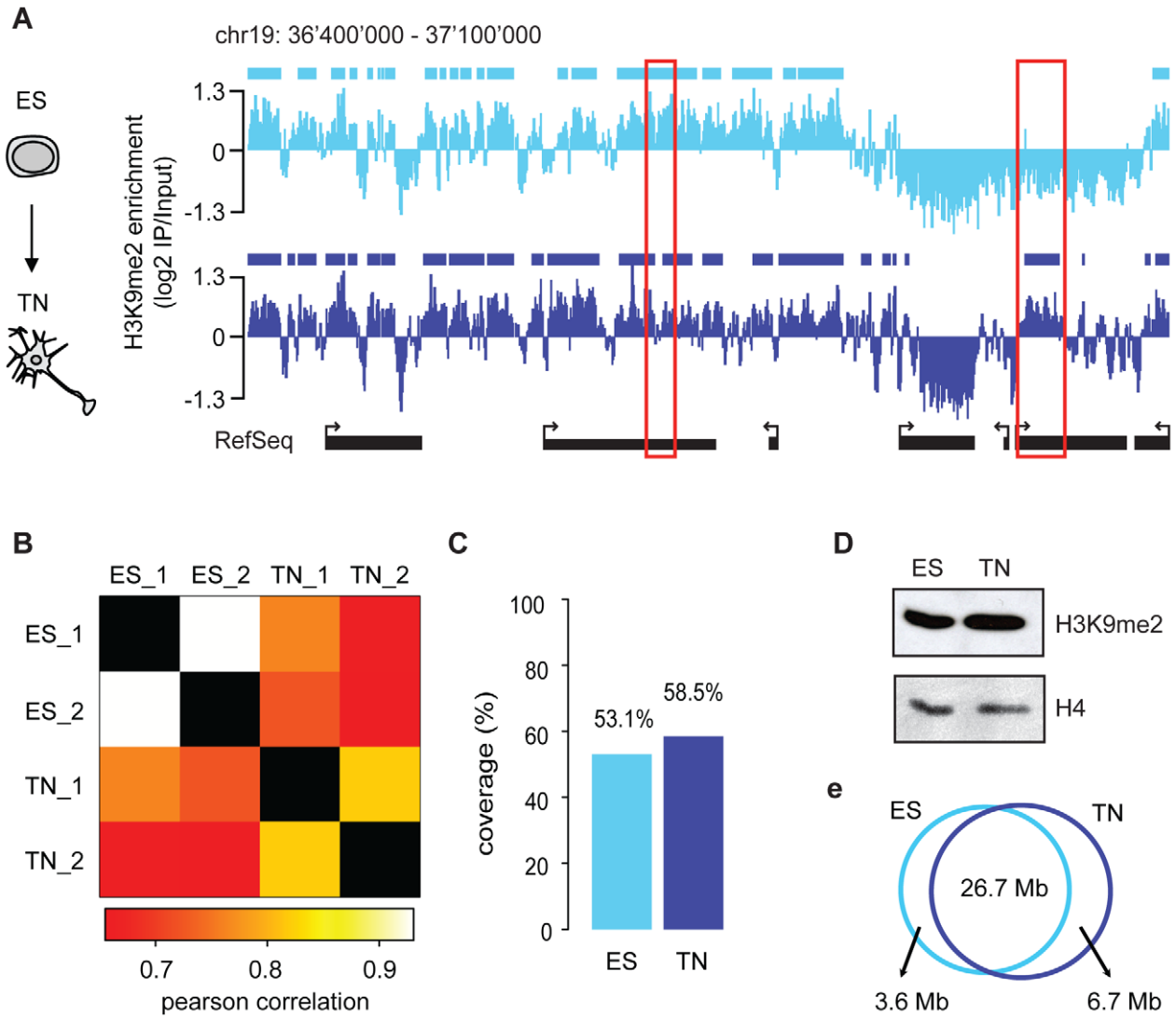


Figure 1. H3K9me2 covers large domains in pluripotent stem cells and derived neurons and is largely invariant between both states. (A) H3K9me2 localization in ES cells and neurons at a representative chromosomal region. Bars above each track indicate H3K9me2 domains determined by HMM. The red boxes highlight two regions that lose and gain H3K9me2, respectively. (B) Pair-wise correlation of H3K9me2 signal among different samples and experiments. Biological replicates are indicated as 1 and 2. Pearson correlations of IP/Input ratios were calculated for 500 bp windows on chromosome 19 (white/yellow corresponds to higher correlations). (C) Quantification of genomic coverage of H3K9me2 in ES cells and neurons applying a two-state HMM (i.e. "high" or "low"). Shown is the percentage of chromosome 19 that is in an H3K9me2 "high" state (i.e. enriched for H3K9me2) in both biological replicates. The value was normalized to the total coverage of the tiled region on the array. (D) Western blot detection of H3K9me2 levels in ES cells and neurons (TN). (E) Venn diagram showing the overlap of H3K9me2 enriched regions between ES cells and neurons (TN).
doi:10.1371/journal.pgen.1002090.g001

transcriptome in a quantitative manner using high throughput RNA sequencing (RNAseq). RNAseq in ES cells and derived neurons revealed the expected down regulation of stem cell specific genes and induction of neuron specific genes (Figure 4A). Further, when counting RNA molecules after gene mapping and normalization, both cell types displayed a characteristic bimodal distribution. This reflects a group of genes that is in a clear off state with no detectable RNA molecules and a second peak of expressed genes (Figure 4B). Separating these two groups of genes by a stringent cutoff revealed that out of 35′606 transcription units, 45% are expressed in ES cells. Interestingly, we identified a slightly higher number (50%) of expressed genes in terminally differentiated neurons, indicating that differentiation of stem cells is not coinciding with a reduced number of highly expressed genes. This agrees with a recent report that suggested that stem cells and somatic cells do mainly differ in the number of low-level expressed genes due to a global reduction of basal gene activity in the course of lineage-commitment and loss of pluripotency [10]. To test this in our *in vitro* differentiation system, we grouped the genes that could not clearly be assigned to the on or off state, into a separate class of genes expressed at low to background level (Figure 4B). This analysis reveals that in stem cells 16% of all transcription units show a basal expression level. Surprisingly however, the proportion of genes expressed at such low level (14%) is very similar in neurons. This unexpected finding prompted us to conduct an additional RNAseq experiment in a second fully differentiated somatic murine cell type; primary mouse embryonic fibroblasts (MEFs). Interestingly, also fibroblasts display a similar

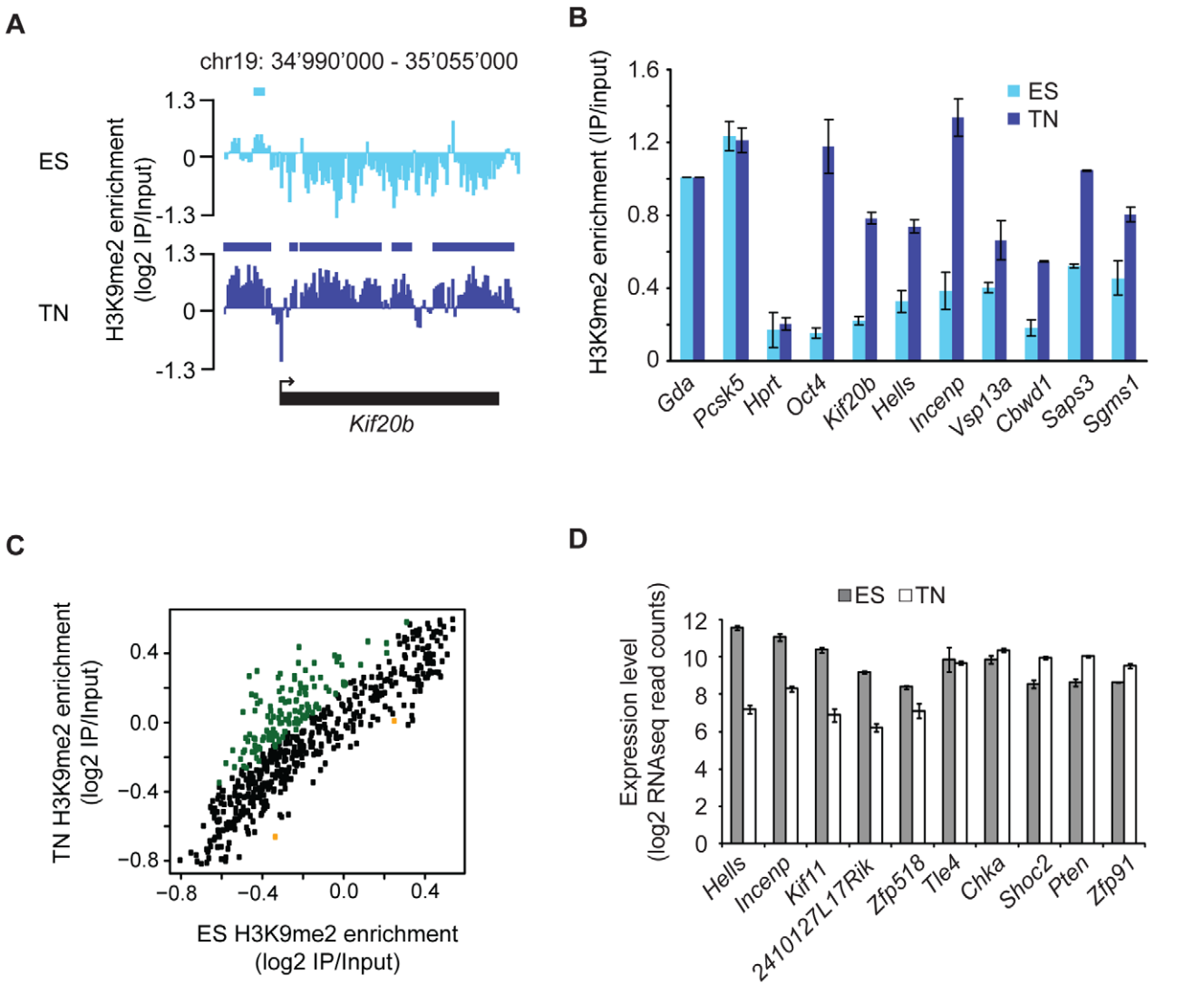


Figure 2. Changes in H3K9me2 during differentiation occur localized and at gene bodies. (A) H3K9me2 signal at the *Kif20b* locus as an example of a locus, which gains H3K9me2 in neurons. Bars above each track indicate HMM-defined H3K9me2 domains. (B) Validation of microarray results by single gene PCRs for a set of gene bodies that gain H3K9me2. Enrichments were normalized to *Gda*. Both *Gda* and *Pcsk5* have equally high H3K9me2 levels in both cellular states in the array data. Methylation levels of *Hprt* (negative control) and *Oct4* (*Pou5f1;* positive control [39]) were measured at the promoter. Shown are averages from 2 independent differentiation experiments. Error bars indicate standard deviation. (C) Comparison of H3K9me2 enrichments in gene bodies in ES cells and neurons. Genes, which gain and lose H3K9me2 significantly (adjusted P-Value <0.05) are depicted in green and orange, respectively. (D) RNAseq read counts of 10 genes that show the most significant gain of H3K9me2. Black and white bars indicate RNA in ES cells (ES) and neurons (TN), respectively.
doi:10.1371/journal.pgen.1002090.g002

transcriptional landscape as stem cells, with 46% of all transcription units being highly expressed and 13% being expressed at basal levels. Hence, this qualitative similarity of expression patterns is not specific to the neuronal subtype we generated *in vitro*, but appears to be a more general property of both undifferentiated and differentiated cells. Thus, while transcripts expressed at low levels show little overlap between stem cells and somatic cells (Figure S8), their numbers are remarkably similar. Stem cells do also not show an increased number of highly expressed genes. Based on additional analysis we can exclude that this similarity of the transcriptional landscape is a consequence of insufficient sampling (Figure S9). Moreover, it is not limited to genic regions as the abundance of transcripts generated from diverse classes of endogenous repeat is comparable between stem cells and neurons (Figure 4C).

Discussion

Embryonic stem cells are characterized by their potential to differentiate into any cell type of the three germ layers in the developing embryo, while somatic cells lose this developmental plasticity upon lineage-commitment. Despite its relevance for our understanding of development and disease, the molecular determinants of pluripotency are still not fully understood and the factors responsible for this uniqueness of stem cells are actively debated [24]. Our study of gene expression and an abundant heterochromatin mark reveal surprising conservation of the transcriptome and epigenome landscape between pluripotent and unipotent cells. H3K9me2 is already highly prevalent in ES cells, arguing that the pathways that mediate H3K9me2 are highly active in stem cells, and serve similar functions as in somatic cells,

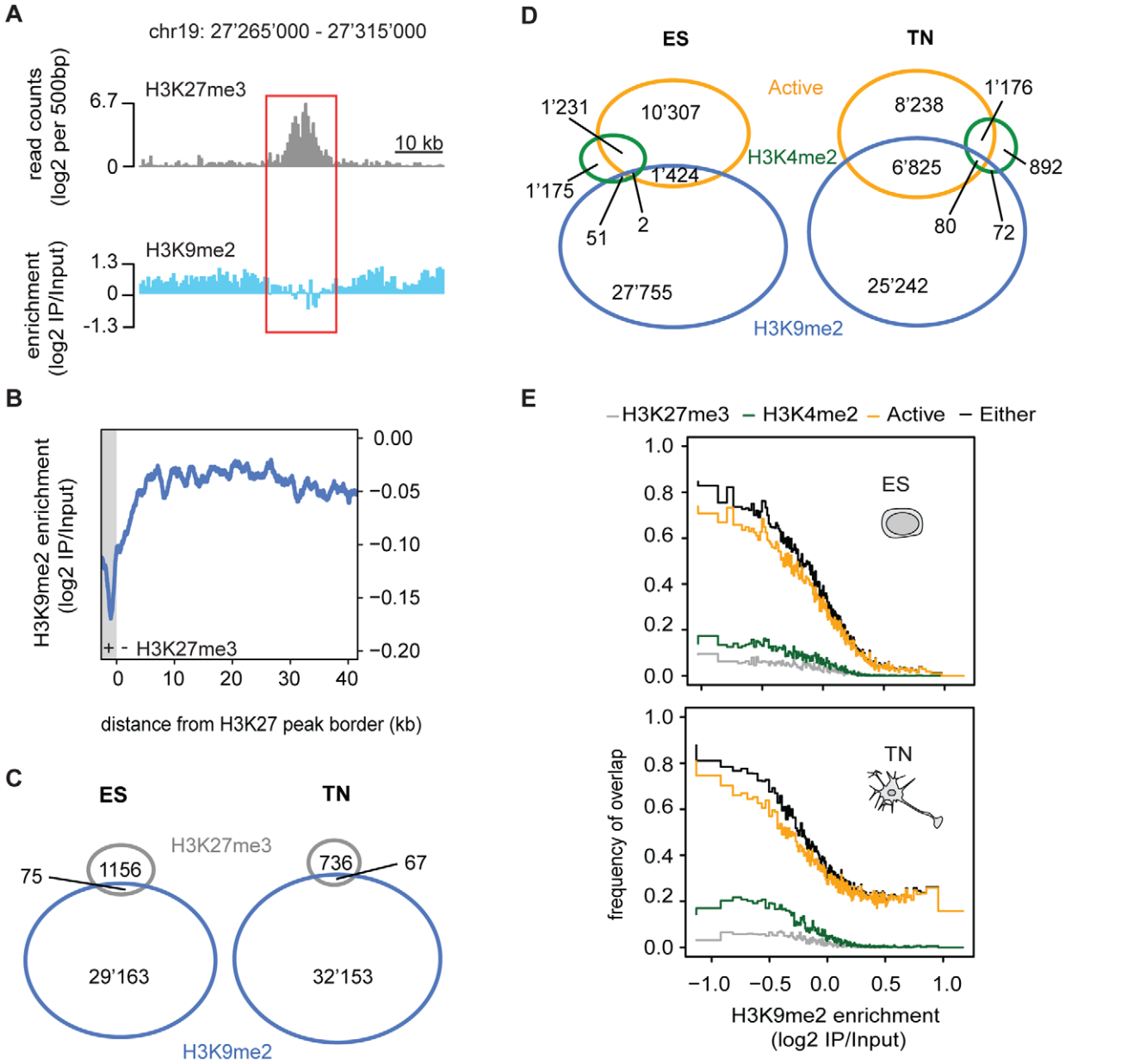


Figure 3. H3K9me2 occurs mutually exclusive with expression and distinct chromatin marks. (A) Enrichment for H3K27me3 and H3K9me2 in a representative region. (B) H3K9me2 enrichments relative to peaks of H3K27me3 illustrating the sharp boundaries between both marks. Plotted are the moving-window averages (median) of H3K9me2 enrichments with window sizes of 500 bp (blue line) around all H3K27me3 domains on chromosome 19 (shown in grey). (C) and (D) Venn diagram illustrating the actual overlap between H3K9me2 and H3K27me3 occupied regions (C) and between H3K9me2, H3K4me2 occupied and active regions (D). Numbers indicate base pair coverage on chromosome 19 (in kb). Note that areas are not drawn to scale. (E) The higher H3K9me2, the less likely a genomic region is active or contains H3K27me3 or H3K4me2. H3K9me2 enrichment values of 500 bp windows were binned into groups of equal size. Plotted is the frequency for H3K9me2 enrichment bins to overlap with active regions (orange), H3K27me3 occupied (gray), H3K4me2 occupied (green) and regions in either of these states (black).
doi:10.1371/journal.pgen.1002090.g003

which only show a slight increase of the mark (from 53 to 58%). Interestingly, the observed gain occurs very localized at gene bodies and does not necessarily coincide with lower transcription of the corresponding gene. The analysis of regions that acquire H3K9me2 during differentiation further revealed a differentiation specific coexistence of H3K9me2 and transcriptional activity, which is not detected in the pluripotent state and which could be involved in modulating expression in the differentiated cell. Nevertheless, despite the subtle increase we detected, an involvement of H3K9me2 in globally regulating cell-type specific gene repression appears unlikely. The limited dynamics as compared to massive transcriptome reprogramming and the limited correlation between expression and gain of H3K9me2 at target genes argue against H3K9me2 as being a major player in setting up gene expression programs.

These findings disagree with a recent report that suggested absence of large H3K9me2 domains in ES cells and found a striking increase in differentiated cells [17]. Notably, our ES cell profile for H3K9me2 is similar to the one generated previously (Figure 1B and Figure S2), making data analysis a likely

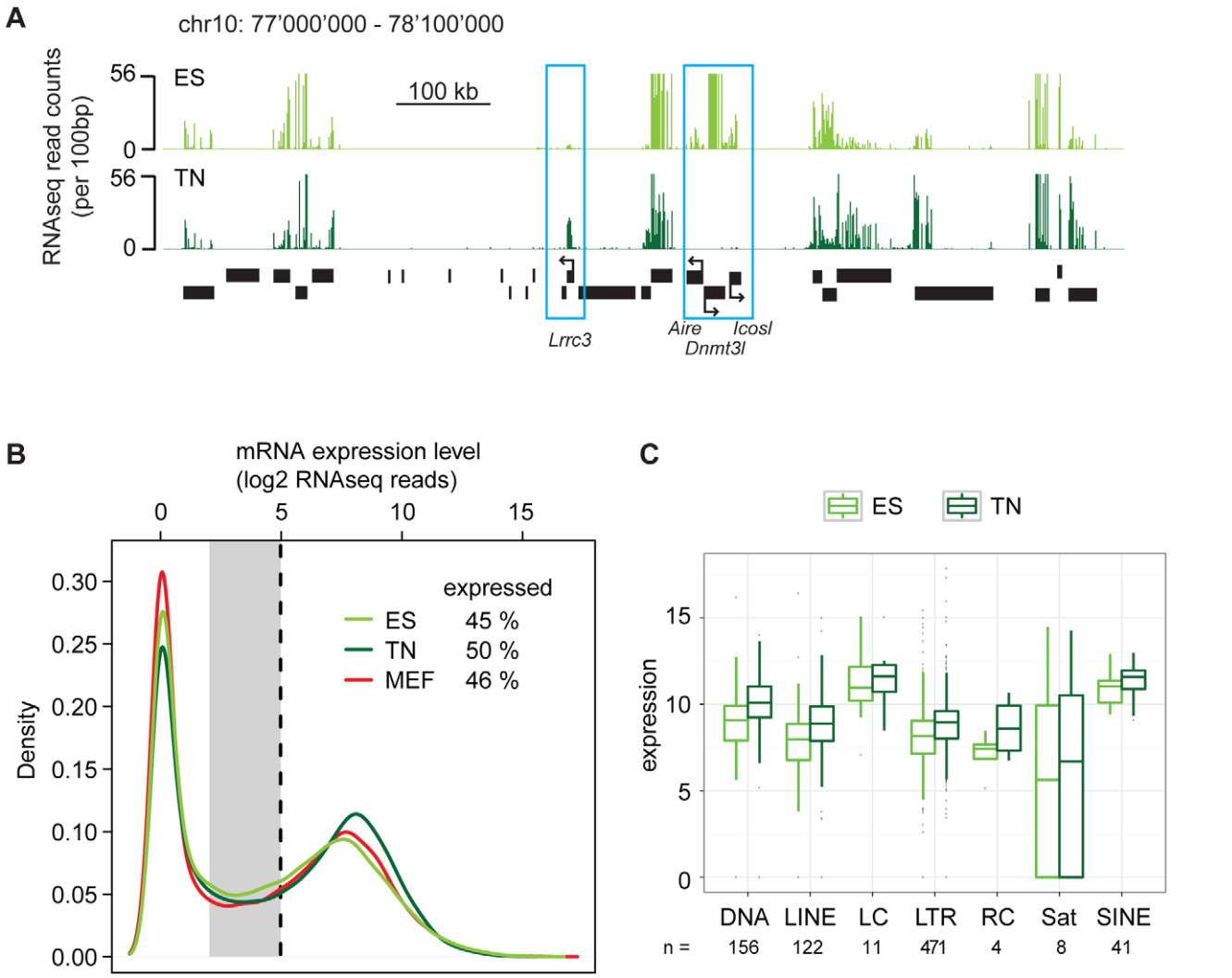


Figure 4. Promiscuous low-level expression is not a distinctive feature of the pluripotent state. (A) Representative region illustrating RNAseq data in ES cells and neurons (TN). Blue boxes indicate genes that either increase (*Lrrc3*, a neuron specific ion channel) or decrease (*Aire, Dnmt3l, Icosl*; three ES cell specific genes) expression during differentiation, while the other genes remain constant. (B) Density distribution of transcript abundance (measured as read counts) for all RefSeq transcripts in ES cells (green line), derived neurons (dark green line) and mouse embryonic fibroblasts (red line). Numbers indicate percentage of expressed genes in each cell type based on a threshold of log2(read count) = 5 (black dotted line). The gray area between 2 and 5 indicates the zone of low-level expression. (C) Boxplot representation of expression levels of prominent repeat classes in ES cells and neurons (TN). Numbers below indicate how many repeat elements are included in each repeat class.
doi:10.1371/journal.pgen.1002090.g004

explanation for the discrepancy. We applied an unbiased approach that is insensitive to variations between arrays and which we show to lead to similar results under various parameter settings (Figure 3 and Figure S10). As already discussed by Fillion and van Steensel [25], the previous study relied on defined thresholds, which can be prone to false estimation of differences particular in the absence of biological replicates [25].

Widespread low-level expression in stem cells has previously been reported and interpreted as a sign of pluripotency [5,6,10]. It has been speculated that this basal promiscuous activity would poise genes for rapid induction upon receipt of differentiation cues [5,6,10]. Using mRNA sequencing in our differentiation paradigm does not confirm this model. We do not find evidence of elevated transcription throughout the genome or on specific chromosome (Figure S8; [10]). A likely explanation for these discrepancies is that microarrays, which were used in the previous studies, overestimate low level signal due to cross-hybridization [26]. In the present study we utilized RNAseq, which permits an actual counting of RNA molecules and thus enables accurate discrimination between very low and no expression. Notably, RNA sequencing experiments have recently put other findings in question that relied on quantifying small transcriptome differences detected by microarrays. For example, recent RNAseq data challenged the presence of pervasive intergenic transcription [26] and the existence of transcriptional dosage compensation of the single male X chromosome [27]. In addition to the increased sensitivity of RNAseq that can explain the differences to previous studies, the presence of a small fraction of differentiated cells in suboptimal conditions of stem cell culture could similarly contribute to an overestimation of the number of genes that are actually expressed in stem cells. Notably, the ES cell differentiation protocol applied by us is optimized to reduce the number of differentiated cells in the culture [18].

In summary, our analysis suggests to revisit the model of massive heterochromatinization during cellular differentiation via a global increase in repressive histone marks [5–9,17] and coinciding repression of basal gene activity [5,6,10].

Our data together with previous reports on dynamics of DNA methylation, H3K9me3 and the Polycomb pathway between pluripotent and somatic cells [9,14,15,28–31] support a model whereby repressive chromatin is already highly active in stem cells and that epigenome reprogramming entails localized changes of repressive histone modifications and DNA methylation at regulatory regions that specify and stabilize lineage specification and terminal differentiation [32]. It will be interesting to determine if these local differences account for the observed changes in nuclear morphology [33]. Notably, epigenetic repression can be overcome by the local activity of transcription factors upon strong induction cues during normal differentiation or artificially during generation of induced pluripotent stem cells (iPS) [34] and might therefore safeguard rather than actively channel development via direct transcriptome regulation.

Methods

Cell culture

Wild-type embryonic stem cells (129Sv-C57Bl/6) were cultured and differentiated as previously described [14,35]. Fibroblasts were isolated from wild-type embryos (C57Bl/6).

Western blot and peptide dot blot analysis

Peptide sequences can be found in Table S2. Western blot analysis was performed with acid extracts using 1/1000 dilutions of either anti-H3K9me2 (Abcam no. 1220) or anti-H4 (Upstate, no. 07–108) antibodies. Blots were developed with ECL reagent (GE Healthcare).

Chromatin-IP (ChIP)

ChIP experiments were performed as described before [14], starting with 70 µg of chromatin and 5 µg of the following antibodies: anti-dimethyl-H3K9 (LP Bio, no. AR-0108), anti-dimethyl-H3K9 (Abcam no. 1220), anti-trimethyl-H3K27 (Upstate, no. 07–449), anti-dimethyl-H3K4 (Upstate, no. 07–030). H3K9me2 ChIP samples were amplified using the WGA2 kit (Sigma) and hybridized to a custom tiling microarray (NimbleGen Systems Inc., see below). H3K27me3 and H3K4me2 ChIP libraries for Illumina sequencing were prepared with the Illumina ChIP-Seq DNA Sample Prep Kit (Cat# IP-102-1001) according to Illumina's instructions and sequenced on the Genome Analyzer II following the manufacturer's protocols. ChIP-real time PCR was performed using SYBR Green chemistry (ABI) and 1/40 of ChIP or 20 ng of input chromatin per PCR reaction. Primers are listed in Table S1.

Microarray design

H3K9me2 ChIP samples were hybridized to custom designed microarrays representing all well-annotated promoters, several large multi-gene loci and the complete chromosome 19 with an average probe spacing of 100 bp and a total of 2.1 million features (HD2.1, NimbleGen Systems Inc).

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 log2 of the bound/input ratios for each individual oligo. Subsequently, for comparison all arrays were normalized to a median log2 = 0 and scale normalized to have the same median absolute deviation using the "LIMMA" R/Bioconductor package [36,37].

RNAseq data analysis

RNA from ES cells, neurons and fibroblasts of two independent biological replicates each was used for cDNA preparation using oligo dT primers followed by sequencing on an Illumina GA II analyzer. Reads were mapped to the *Mus musculus* transcriptome and normalized to transcript length and sequencing library size (for details see Text S1).

Bioinformatics

Unless otherwise stated, H3K9me2 enriched regions were identified by HMM and H3K4me2 and H3K27me3 peaks using MACS peak finder [38]. Active regions were defined as RefSeq transcription units with a normalized RNAseq log2 read count above 5 (for details see Text S1). Microarray design, hybridization and analysis, ChIPseq and RNAseq analysis and additional references are described in Text S1.

Datasets

Microarray and deep sequencing data were deposited at NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE27866 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE27866).

Supporting Information

Figure S1 Specificity of H3K9me2 antibodies used in this study. Indicated amounts of either modified or unmodified H3 and H4 peptides (H3 residues 1–20, 19–38 and 25–45, H4 residues 12–31) were spotted onto polyvinylidene difluoride membranes (GE) and probed with (A) anti-dimethyl-H3K9 (LP Bio, no. AR-0108) or (B) anti-dimethyl-H3K9 (Abcam no. 1220) at 1:1000 dilution each. Both antibodies are highly specific towards H3K9me2.
(TIF)

Figure S2 High similarity of ES cell H3K9me2 datasets. (A) Comparison of our H3K9me2 enrichment profiles in ES cells to a previously generated ES cell profile (Wen et al., 2009). (B) Venn diagram illustrating the overlap of H3K9me2-occupied regions between the two ES cell datasets from (A) (defined by an HMM approach). The overlap agrees with the high correlation for this previously generated ES cell dataset with our two ES cell replicates (Pearson correlation of 0.73 and 0.74, on 500 bp windows).
(TIF)

Figure S3 H3K9me2 enrichments on chromosome 19 are representative of the entire genome. (A) H3K9me2 enrichments in ES cells from a previously generated dataset (Wen et al., 2009). The boxplots show H3K9me2 enrichment values for all probes separated by chromosome, with chromosome 19 highlighted in blue. (B) H3K9me2 enrichments per 900 bp promoter windows (defined in Mohn et al., 2008) in ES cells. The boxplots represent average enrichments at promoters separated by chromosome with chromosome 19 highlighted in blue.
(TIF)

Figure S4 H3K9me2 enriched regions are robustly identified independent of HMM parameters and H3K9me2 domain size is highly similar in ES cells and neurons. (A) Shown is the comparison of H3K9me2 enrichments in ES cells and neurons using a 3 state HMM to define non-occupied (low), intermediately and highly occupied domains (top). Importantly, the state specific HMM distribution parameters (mean and variance of H3K9me2 enrichment) are very similar between ES cells and neurons

(bottom). Note that with a 3 state HMM the percentage of domains in the high state does not change between ES cells and neurons (TN). However, we detect an overall increase of the intermediate state of around 10%. (B) H3K9me2 domains were defined by HMM (see methods). While numbers of domains on chromosome 19 increase from 1618 to 1914 from ES cells to neurons, their domain size distribution is very similar. This was also seen using alternative domain definition methods (data not shown).
(TIF)

Figure S5 Regions that lose are smaller than regions that gain H3K9me2 during differentiation and gaining regions often overlap with gene bodies. (A and B) Histograms displaying the size distributions of regions that lose (A) and gain (B) H3K9me2 during neuronal differentiation. Adjacent 500 bp windows being in a low HMM state in both ES cell replicates and in a high HMM state in the neuron replicates were grouped into H3K9me2 losing and gaining regions, respectively. (C) Pie chart illustrating the percentage of large (>10kb) H3K9me2 gaining regions within (grey) and outside (white) of genes (left). This overlap is significantly higher than for all regions that have no H3K9me2 in ES cells, i.e. which are in a HMM low state in both replicates (right). (D) Boxplot showing expression levels of all genes (n = 138, adjusted P-value <0.05) that gain H3K9me2 in neurons. Note that the median expression does not change significantly between ES cells and neurons (TN), indicating that there is no global trend towards either up- or down-regulation upon accumulation of H3K9me2 in the gene bodies of these genes.
(TIF)

Figure S6 Regions losing H3K9me2 during differentiation are close to background. (A–C) Shown are representative genomic regions that significantly lose H3K9me2 as defined by HMM analysis. (D) ChIP-real time PCR validation of the regions shown in (A–C). Note that H3K9me2 enrichments are already low in ES and only slightly above background as exemplified by the active housekeeping gene *Hprt*, which does not carry H3K9me2. The loss of enrichment in neurons is small, though reproducible.
(TIF)

Figure S7 H3K9me2 in neurons rarely overlaps with promoter regions of active genes. (A and B) Pie chart illustrating the percentage of promoters (1 kb up- and downstream of TSS) (A) and gene bodies (B) of transcribed genes that overlap H3K9me2-occupied regions in neurons. Note that promoters are rarely H3K9 dimethylated when the gene is actively transcribed.
(TIF)

Figure S8 Overlap of low-level expressed genes and expression of transcripts per chromosome. (A) Venn diagram illustrating the overlap of transcripts expressed at a low level (log2 RNAseq reads between 2 and 5) in at least one of the examined cell types. Note that areas are not drawn to scale. (B) Boxplots showing expression levels of transcripts per chromosomes.
(TIF)

Figure S9 Absence of difference in transcriptome complexity is not a function of sample size in next generation sequencing. Shown is the number of detected Refseq transcripts (at least 32 reads) relative to the number of sampled reads. Number of detected transcripts is shown as the mean of 100 rounds of random subsampling with the same total read numbers. Irrespective of the sample size, a random sampling of RNAseq reads from ES cells (green line), neurons (dark green line) and MEFs (red line) revealed a smaller number of detected transcripts than an artificial sample with pooled reads from all three cell types (blue line) at any given sample size.
(TIF)

Figure S10 H3K9me2 domain definition by different statistical methods leads to similar results. (A) HMM-independent quantification of genomic coverage of H3K9me2 in ES cells and neurons. Shown is the percentage of chromosome 19 that lies within H3K9me2 peaks defined by a simple threshold method (t-peaks, see Text S1). (B) Diagram illustrating the actual overlap between regions defined to be occupied by H3K9me2 by two different methods; a simple threshold method (t-peaks; see Text S1) and by an HMM approach (HMM; see Text S1).
(TIF)

Table S1 List of real time PCR primers.
(TIF)

Table S2 Peptides used in dot blots in Figure S1.
(TIF)

Text S1 Supplementary methods.
(DOC)

Acknowledgments

We are grateful to Christian Beisel and Ina Nissen for processing deep sequencing samples within the ETH BSSE Deep Sequencing unit. We thank Antoine Peters and Frédéric Zilbermann for providing H3 and H4 peptides; Christiane Wirbelauer and Leslie Hoerner for technical assistance; and members of the Schübeler lab, Antoine Peters, and Marc Bühler for critical comments on the manuscript.

Author Contributions

Conceived and designed the experiments: FL FM DS. Performed the experiments: FL FM VKT. Analyzed the data: FL FM TB DS. Wrote the paper: FL FM DS. Designed the microarray: TCR. Designed tools to store and analyze deep sequencing data and contributed to data analysis and writing of the manuscript: MBS DG.

References

1. Boyer LA, Mathur D, Jaenisch R (2006) Molecular control of pluripotency. Curr Opin Genet Dev 16: 455–462.
2. Egli D, Birkhoff G, Eggan K (2008) Mediators of reprogramming: transcription factors and transitions through mitosis. Nat Rev Mol Cell Biol 9: 505–516.
3. Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. Science 293: 1089–1093.
4. Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. Cell 128: 669–681.
5. Niwa H (2007) Open conformation chromatin and pluripotency. Genes Dev 21: 2671–2676.
6. Meshorer E, Misteli T (2006) Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol 7: 540–546.
7. Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, et al. (2009) Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature 460: 863–868.
8. Persson J, Ekwall K (2010) Chd1 remodelers maintain open chromatin and regulate the epigenetics of differentiation. Exp Cell Res 316: 1316–1323.
9. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, et al. (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 6: 479–491.
10. Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, et al. (2008) Global transcription in pluripotent embryonic stem cells. Cell Stem Cell 2: 437–447.
11. Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M (2011) Open chromatin in pluripotency and reprogramming. Nat Rev Mol Cell Biol 12: 36–47.
12. Francastel C, Schubeler D, Martin DI, Groudine M (2000) Nuclear compartmentalization and gene activity. Nat Rev Mol Cell Biol 1: 137–143.
13. Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, et al. (2010) Global chromatin architecture reflects pluripotency and lineage commitment in

the early mouse embryo. PLoS ONE 5: e10531. doi:10.1371/journal.pone.0010531.
14. Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, et al. (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell 30: 755–766.
15. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766–770.
16. Farthing CR, Ficz G, Ng RK, Chan CF, Andrews S, et al. (2008) Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet 4: e1000116. doi:10.1371/journal.pgen.1000116.
17. Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP (2009) Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat Genet 41: 246–250.
18. Bibel M, Richter J, Schrenk K, Tucker KL, Staiger V, et al. (2004) Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nat Neurosci 7: 1003–1009.
19. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799–816.
20. Koch CM, Andrews RM, Flicek P, Dillon SC, Karaoz U, et al. (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res 17: 691–707.
21. Ringrose L, Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development 134: 223–232.
22. Schwartz YB, Pirrotta V (2007) Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet 8: 9–22.
23. O'Geen H, Squazzo SL, Iyengar S, Blahnik K, Rinn JL, et al. (2007) Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. PLoS Genet 3: e89. doi:10.1371/journal.pgen.0030089.
24. Hanna JH, Saha K, Jaenisch R (2010) Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell 143: 508–525.
25. Filion GJ, van Steensel B (2010) Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells. Nat Genet 42: 4; author reply 5–6.
26. van Bakel H, Nislow C, Blencowe BJ, Hughes TR (2010) Most "dark matter" transcripts are associated with known genes. PLoS Biol 8: e1000371. doi:10.1371/journal.pbio.1000371.
27. Xiong YY, Chen XS, Chen ZD, Wang XZ, Shi SH, et al. (2010) RNA sequencing shows no dosage compensation of the active X-chromosome. Nature Genetics 42: 1043-U1029.
28. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553–560.
29. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K (2007) The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol 27: 3769–3779.
30. Pan G, Tian S, Nie J, Yang C, Ruotti V, 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.
31. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322.
32. Mohn F, Schubeler D (2009) Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet 25: 129–136.
33. Meister P, Mango SE, Gasser SM (2011) Locking the genome: nuclear organization and cell fate. Curr Opin Genet Dev. In press.
34. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663–676.
35. Bibel M, Richter J, Lacroix E, Barde YA (2007) Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nat Protoc 2: 1034–1043.
36. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article3.
37. Smyth GK, Speed T (2003) Normalization of cDNA microarray data. Methods 31: 265–273.
38. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, et al. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9: R137.
39. Feldman N, Gerson A, Fang J, Li E, Zhang Y, et al. (2006) G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol 8: 188–194.

Supplementary Information

Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells

Florian Lienert, Fabio Mohn, Vijay K. Tiwari, Tuncay Baubec, Tim C. Roloff, Dimos Gaidatzis, Michael B. Stadler and Dirk Schübeler

Total supplementary items: 12

- Figure S1. Dot blots for assessing the specificity of H3K9me2 antibodies used in this study. Relates to Figure 1 and is discussed in the main text.

- Figure S2. Comparison of H3K9me2 profiles with previously published data. Relates to Figure 1.

- Figure S3. Comparison of H3K9me2 on chromosome 19 and other chromosome. Is discussed in the main text.

- Figure S4. Comparison of HMM analysis parameters. Relates to Figures 1 and 2 and is discussed in the main text.

- Figure S5. Analysis of regions that lose or gain H3K9me2 during differentiation. Relates to Figure 2.

- Figure S6. Examples of regions that lose H3K9me2 during differentiation. Relates to Figure 2.

- Figure S7. Analysis showing that in neurons H3K9me2 rarely overlaps with promoter regions of active genes. Relates to Figure 3.

- Figure S8. Overlap of low level expressed genes among cell types and comparison of transcript expression levels among chromosomes. Discussed in main text.

- Figure S9. Analysis to validate that the absence of difference in transcriptome complexity is not a function of sample size in next generation sequencing and does not reflect non-saturating sequencing. Relates to Figure 4 and is discussed in main text and Supplementary Methods.

- Figure S10. H3K9me2 domain definition by different statistical methods leads to similar results. Discussed in main text and Supplementary Methods.

- Table S1. List of real time PCR primers

- Table S2. List of peptides used in dot blots for Figure S1

- Supplementary Methods. Additional details of the bioinformatical analysis methods

- Supplementary References

Supplementary Methods

Segmentation of H3K9me2 profiles by Hidden Markov Models (HMMs). To define H3K9me2 enriched regions, we segmented ChIP-chip data using HMMs, as described (Birney et al., 2007). The basic premise of HMMs is that observed data are generated stochastically from a pre-determined number of hidden background probability distributions, or states. We used two states to distinguish enriched from not enriched regions (an analysis using three states led to similar results, see Figure S3). The parameters of the HMMs (emission probabilities, here modeled as normal distributions, and the transition probabilities between states) are estimated via unsupervised learning (Baum-Welch algorithm) from the H3K9me2 enrichment profile. For that purpose, log2 of bound/input ratios were calculated for 500 bp windows on the whole chromosome 19. For single windows not covered by an oligonucleotide on the array the average of the neighboring two windows was taken. Regions consisting of more than one 500 bp window not covered by array were omitted from the analysis. Enriched and not enriched states were assigned to genomic positions according to the most probable path through the trained model states given the observed data (Viterbi algorithm). The analysis was performed independently on each biological replicate and 500 bp windows being in a high state in both replicates were considered to be enriched for H3K9me2. Consecutive genomic positions with identical H3K9me2 enrichment states were merged. For the 3 state HMMs, we performed the analysis in the same way, except for using averaged H3K9me2 enrichment values from the two biological replicates. Analysis was done using the

"RHmm" R package (http://CRAN.R-project.org/package=RHmm (Taramasco, 2009)).

Analysis of H3K9me2 changes in gene bodies. We calculated the average H3K9me2 signal per transcript (RefSeq start to end, downloaded 06/24/2010) as the weighted sum of the log2(ChIP/input) enrichment of oligos within the transcript divided by the weighted number of oligos within the transcript, with oligo weights defined as the fraction of overlap with the particular transcript. Significant changes between ES cells and neurons were then determined using the "LIMMA" R package (Smyth, 2004; Smyth and Speed, 2003).

Definition of H3K9me2 enriched regions by a simple threshold method (t-peaks). To cross validate H3K9me2 enriched regions obtained from HMMs by an independent approach we made use of a simple threshold method. For that purpose, average log2 bound/input ratios from two replicates were calculated for 500 bp windows on the whole chromosome 19. For both samples (ES cells and neurons), we defined the respective median of log2 bound/input ratios as a threshold. We then merged adjacent regions with a log2 bound/input ratio above the threshold into domains. Domains separated by 1 kb or less were fused. These H3K9me2 enriched domains show high overlap with domains defined by the HMM approach (Figure S10).

Deep sequencing analysis. ChIPseq and RNAseq samples were sequenced on an Illumina GA II analyzer. Low quality reads (not passing Illumina chastity filter or containing more than two N's) and low complexity reads (based on entropy, typically less than 1%) were removed. We mapped reads to the Mus musculus genome (mm9) and transcriptome (RefSeq, downloaded on

07/17/2009) using bowtie (Langmead et al., 2009) allowing for up to two mismatches and retaining all best hits for reads with up to 100 alignments ignoring sequence qualities.

Identification of H3K27me3 and H3K4me2 peaks. H3K4me2 ChIP samples (from one differentiation) and H3K27me3 ChIP samples (from two independent biological replicates) and Input DNA from ES cells (as a background sample) were sequenced on an Illumina Genome Analyzer. Reads were filtered, mapped and counted as described above. We obtained the following number of reads mapping to genome (in Mio. reads); ES_H3K27me3_a1 4.96, ES_H3K27me3_a2 4.60, ES_H3K27me3_b 12.58, ES_H3K27me3_a1 4.59, ES_H3K27me3_a2 4.74, ES_H3K27me3_b 7.5, ES_H3K4me2 13.25, TN_H3K4me2 10.33 and Input 11.9 (with a and b being independent biological replicates and 1 and 2 being technical replicates). In order to identify peaks in the H3K27me3 and H3K4me2 datasets we used MACS peak finder (Zhang et al., 2008). MACS was used with standard settings (lambad-set: 2kb, 10kb, 20kb) and the Input sequencing sample as a background. Peaks were filtered in the following way. We counted the number of reads from the histone modification ChIP library and from the Input library in the peak regions identified by MACS. For each sample we then normalized the number of mapped reads per peak to the total number of reads which could be mapped to the genome. Next, we determined for each peak the enrichment ratio of normalized read numbers from ChIPseq to the normalized reads from Input-seq (after addition of eight pseudo-counts per peak to reduce sampling noise). For further analysis we only used peaks with a two-fold enrichment ratio and more than four mapped reads per 100 bp. We

crossvalidated peaks by comparison to peaks identified by two independent methods; BayesPeak (Spyrou et al., 2009) and a variation of the t-peaks method described above. The peaks identified by using these three different methods showed a high overlap (data not shown).

RNAseq data analysis. RNA from ES cells, neurons and MEFs from two independent biological replicates was used separately for cDNA preparation followed by sequencing on an Illumina GA II analyzer. Reads were filtered, mapped and counted as described above. We obtained the following number of reads mapping to the transcriptome (in Mio. reads); ES_RNA_a 13.3, ES_RNA_b 35.7, TN_RNA_a 10.8, TN_RNA_b 21.5, MEF_RNA_a 27.2 and MEF_RNA_b 27.5, (with a and b being independent biological replicates). Next, we calculated expression levels of the transcripts as the weighted sum of reads aligning to the transcript, with read weights defined as the observed count divided by the number of genomic hits of each read or the number of hits in a single transcript if greater than the number of genomic hits. This sum was length normalized by dividing through the transcript length and multiplying by the average length of all transcripts. In order to allow comparison among different cell types and biological replicates, transcript levels were normalized to the total number of reads mapping to transcripts for each sequence run. For Repeat analysis, genomic read alignments were compared to the Mus musculus Repeatmasker annotation (Smit, AFA & Green, P, www.repeatmasker.org) as described for transcriptome analysis. Read counts mapping to 3'927'115 repeat instances were collected in pools according to their corresponding repeat elements and were length-normalized to the summed length of all instances per element. For further analysis we

normalized read counts of repeat instances to the total number of reads mapping uniquely to the genome for each sequence run and merged read counts into their corresponding repeat classes.

Supplementary References

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E.*, et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

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.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

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.

Spyrou, C., Stark, R., Lynch, A.G., and Tavare, S. (2009). BayesPeak: Bayesian analysis of ChIP-seq data. BMC Bioinformatics *10*, 299.

Taramasco, O. (2009). RHmm: Hidden Markov Models simulations and estimations. R package version 1.3.1.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.

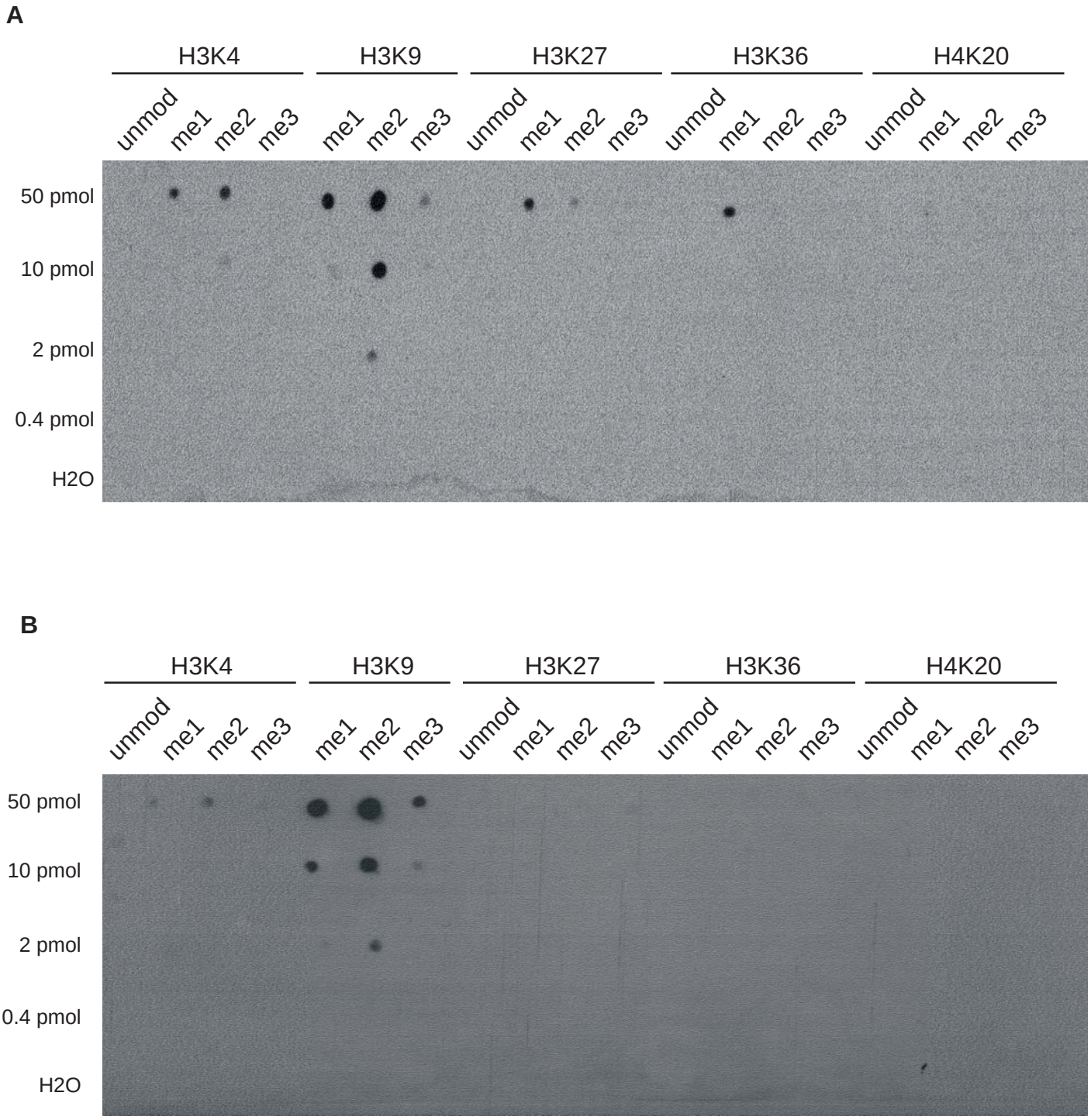


Figure S1. Specificity of H3K9me2 antibodies used in this study.
Indicated amounts of either modified or unmodified H3 and H4 peptides (H3 residues 1-20, 19-38 and 25-45, H4 residues 12-31) were spotted onto polyvinylidene difluoride membranes (GE) and probed with (**A**) anti-dimethyl-H3K9 (LP Bio, no. AR-0108) or (**B**) anti-dimethyl-H3K9 (Abcam no. 1220) at 1:1000 dilution each. Both antibodies are highly specific towards H3K9me2.

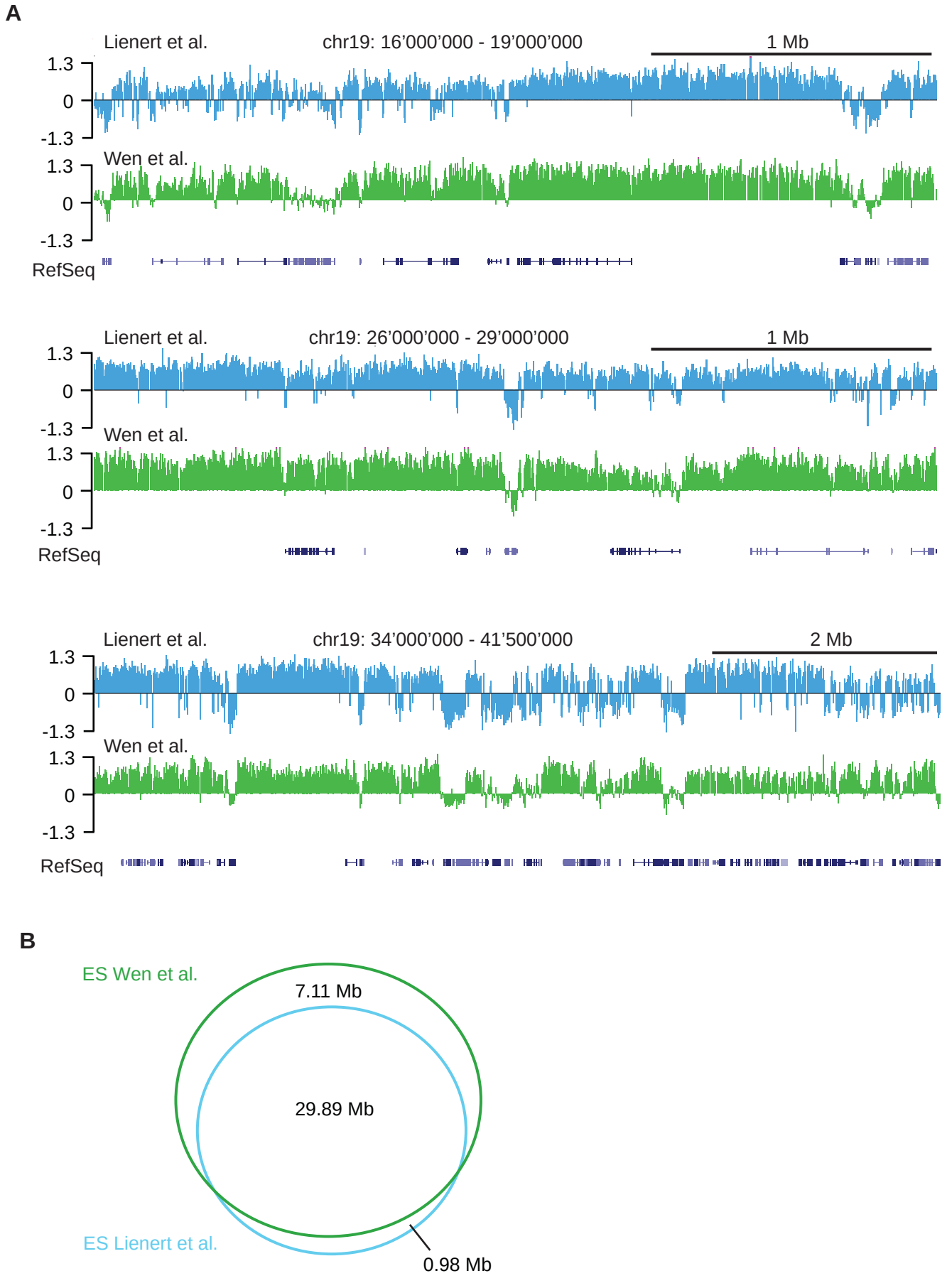


Figure S2. High similarity of ES cell H3K9me2 datasets.
(**A**) Comparison of our H3K9me2 enrichment profiles in ES cells to a previously generated ES cell profile (Wen et al.). (**B**) Venn diagram illustrating the overlap of H3K9me2-occupied regions between the two ES cell datasets from (A) (defined by an HMM approach). The overlap agrees with the high correlation we observe for this previously generated ES cell dataset with our two ES cell replicates (Pearson correlation of 0.73 and 0.74, on 500 bp windows).

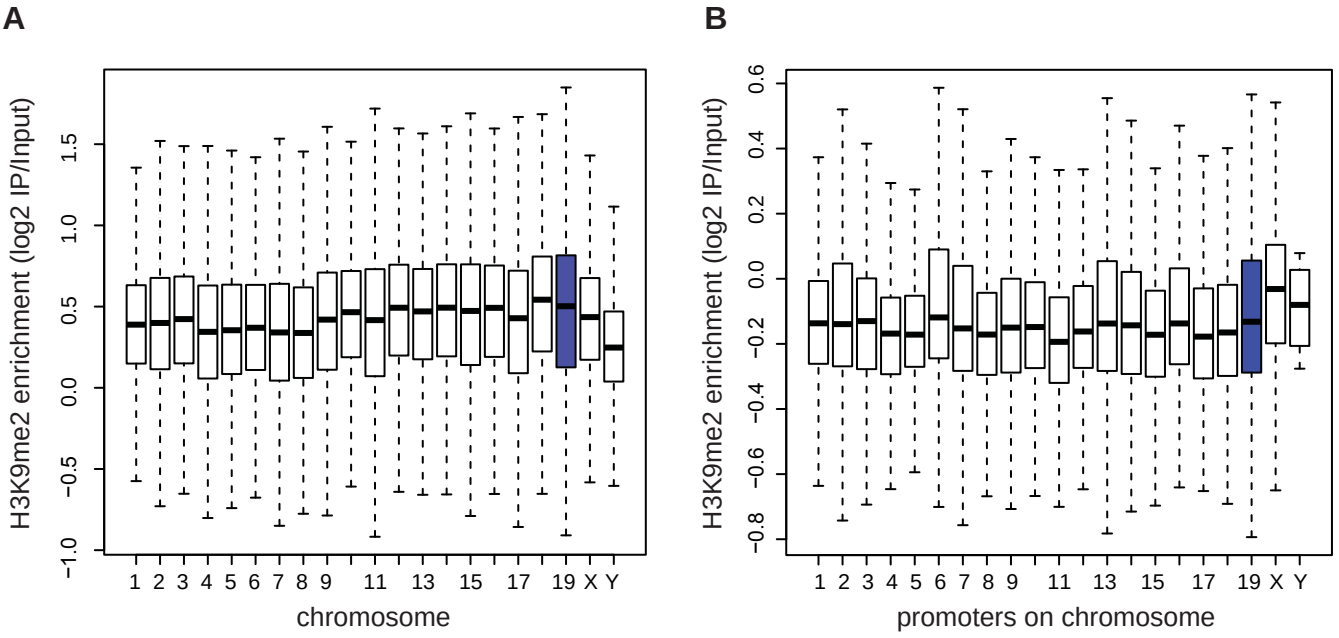
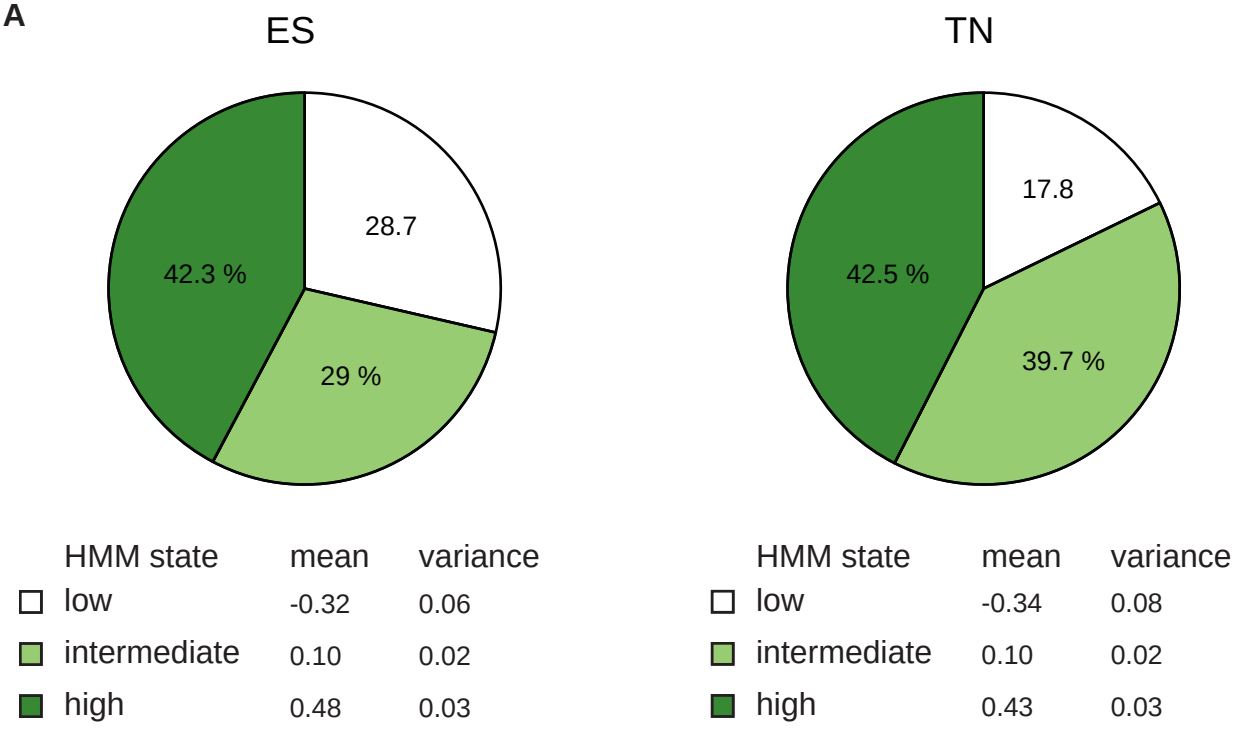


Figure S3. H3K9me2 enrichments on chromosome 19 are representative of the entire genome. (**A**) H3K9me2 enrichments in ES cells from a previously generated dataset (Wen et al.). The boxplots show enrichments on the probe level per chromosome with chromosome 19 highlighted in blue. (**B**) H3K9me2 enrichments per 900 bp promoter windows (defined in Mohn et al.) in ES cells. The boxplots represent average enrichments at promoters per chromosome with chromosome 19 highlighted in blue.

A



HMM state	mean	variance
low	-0.32	0.06
intermediate	0.10	0.02
high	0.48	0.03

HMM state	mean	variance
low	-0.34	0.08
intermediate	0.10	0.02
high	0.43	0.03

Figure S4. (A) H3K9me2 enriched regions are robustly identified independent of HMM parameters. Shown is the comparison of H3K9me2 enrichments in ES cells and neurons using a 3 state HMM to define non-occupied (low), intermediately and highly occupied domains (top). Importantly, the state specific HMM distribution parameters (mean and variance of H3K9me2 enrichment) are very similar between ES cells and neurons (bottom). Note that with a 3 state HMM the percentage of domains in the high state does not change between ES cells and neurons (TN). However, we detect an overall increase of the intermediate state of around 10%.

B

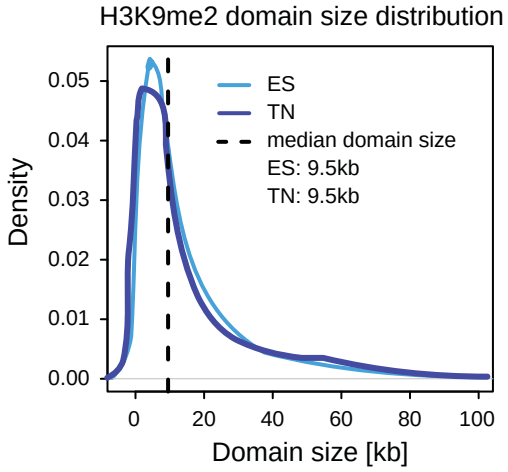


Figure S4. (B) H3K9me2 domain size is highly similar in ES cells and neurons.
H3K9me2 domains were defined by HMM (see methods). While numbers of domains on chromosome 19 increase from 1618 to 1914 from ES cells to neurons, their domain size distribution is very similar. This was also seen using alternative domain definition methdos (data not shown).

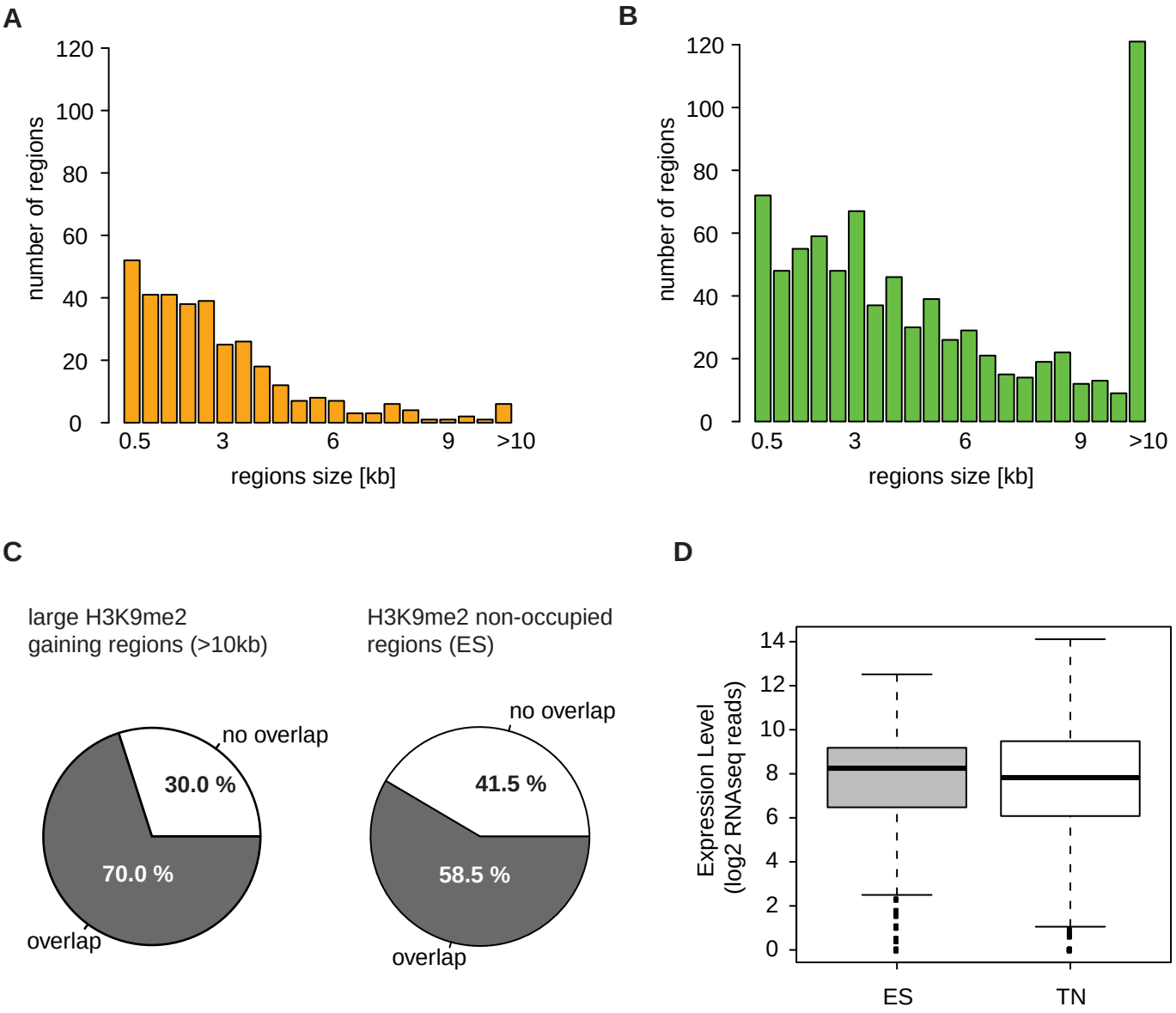


Figure S5. Regions that lose are smaller than regions that gain H3K9me2 during differentiation and gaining regions often overlap with gene bodies.
Histograms displaying the size distributions of regions that lose (**A**) and gain (**B**) H3K9me2 during neuronal differentiation, respectively. Adjacent 500 bp windows being in a low HMM state in both ES cell replicates and in a high HMM state in the neuron replicates were grouped into H3K9me2 losing regions and vice versa. (**C**) Pie chart illustrating the percentage of large (>10kb) H3K9me2 gaining regions within (grey) and outside (white) of genes (left). This overlap is significantly higher than for all regions which have no H3K9me2 in ES cells, defined as being in a HMM low state in both replicates (right). (**D**) Boxplot showing expression levels of all genes (n = 138, adjusted P-value < 0.05) that gain H3K9me2 in neurons. Note that the median expression does not change significantly between ES cells and neurons (TN), indicating that there is no global trend towards either up- or down-regulation upon accumulation of H3K9me2 in the gene bodies of these genes.

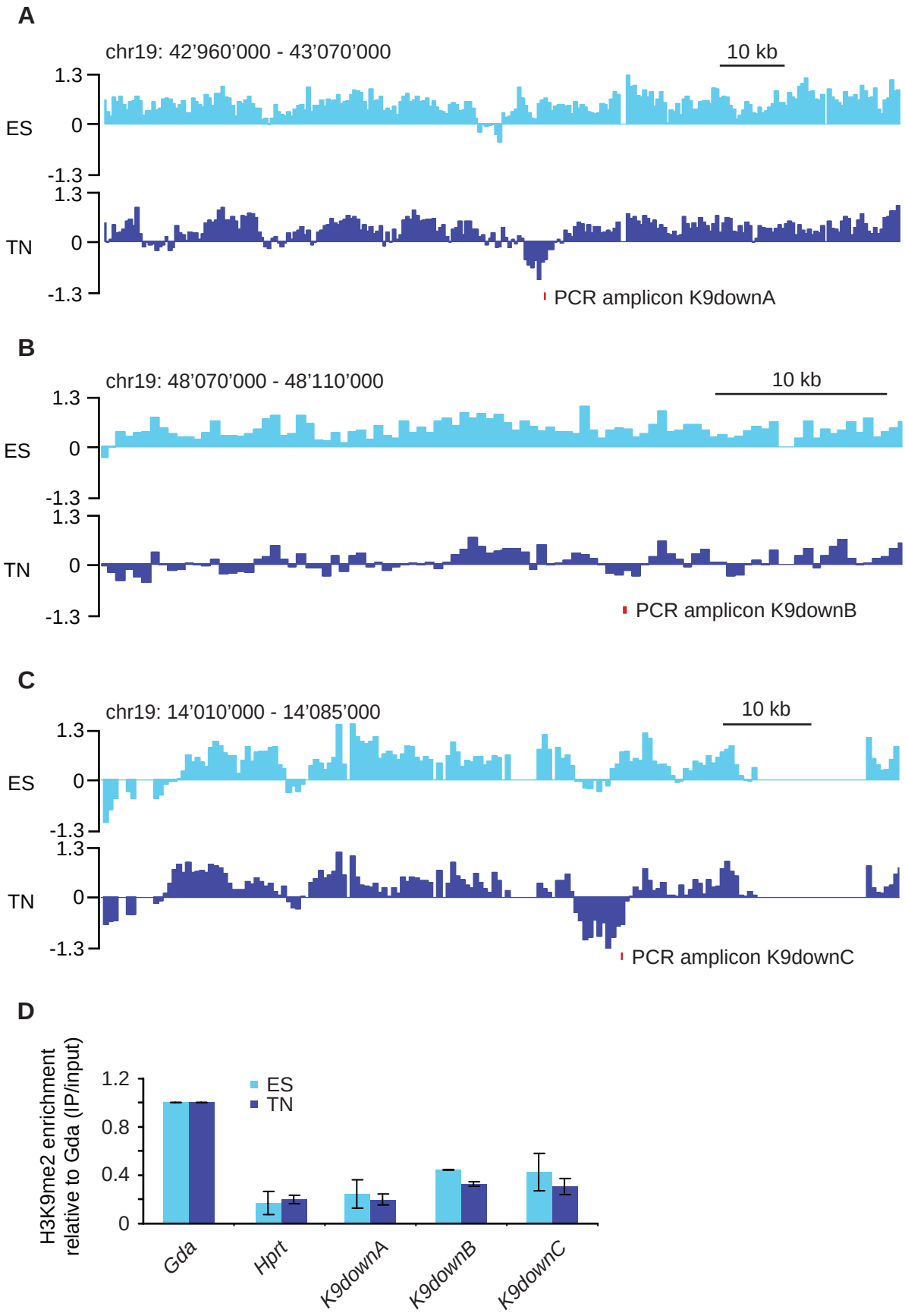


Figure S6. Regions losing H3K9me2 during differentiation are close to background. (**A-C**) Shown are representative genomic regions that significantly lose H3K9me2 as defined by HMM analysis. (D) ChIP-real time PCR validation of those regions shown in (A-C). Note that H3K9me2 enrichments are already low in ES and only slightly above background as exemplified by the active housekeeping gene Hprt, which does not carry H3K9me2. The loss of enrichment in neurons is small, though reproducible.

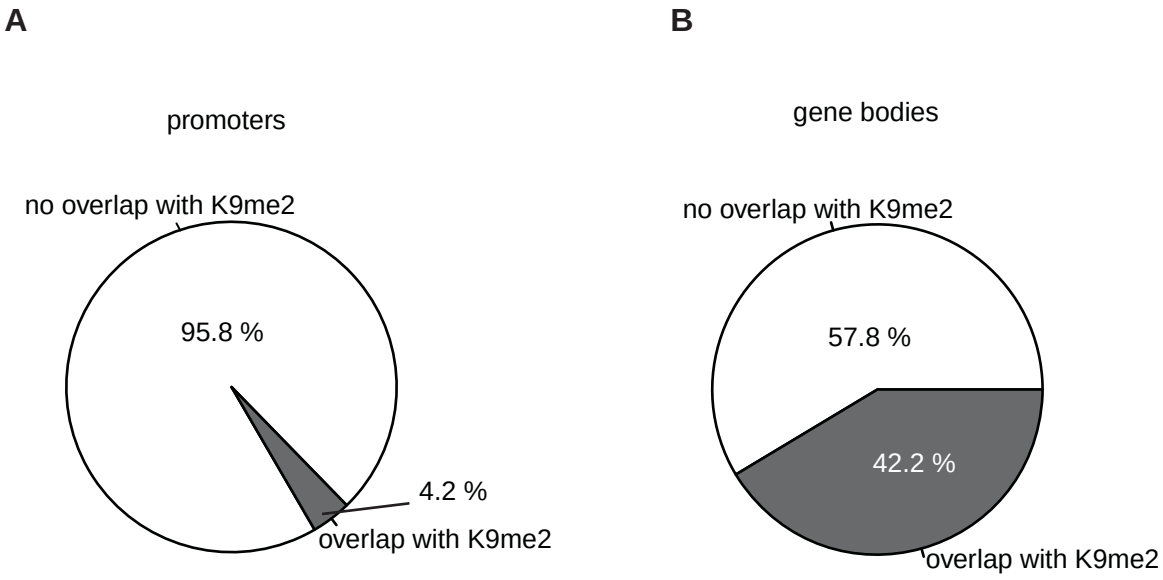


Figure S7. H3K9me2 in neurons rarely overlaps with promoter regions of active genes.
Pie chart illustrating the percentage of promoters (1kb up- and downstream of TSS) (**A**) and gene bodies (**B**) of transcribed genes that overlap H3K9me2-occupied regions in neurons. Note that promoters are rarely H3K9 dimethylated when the gene is actively transcribed.

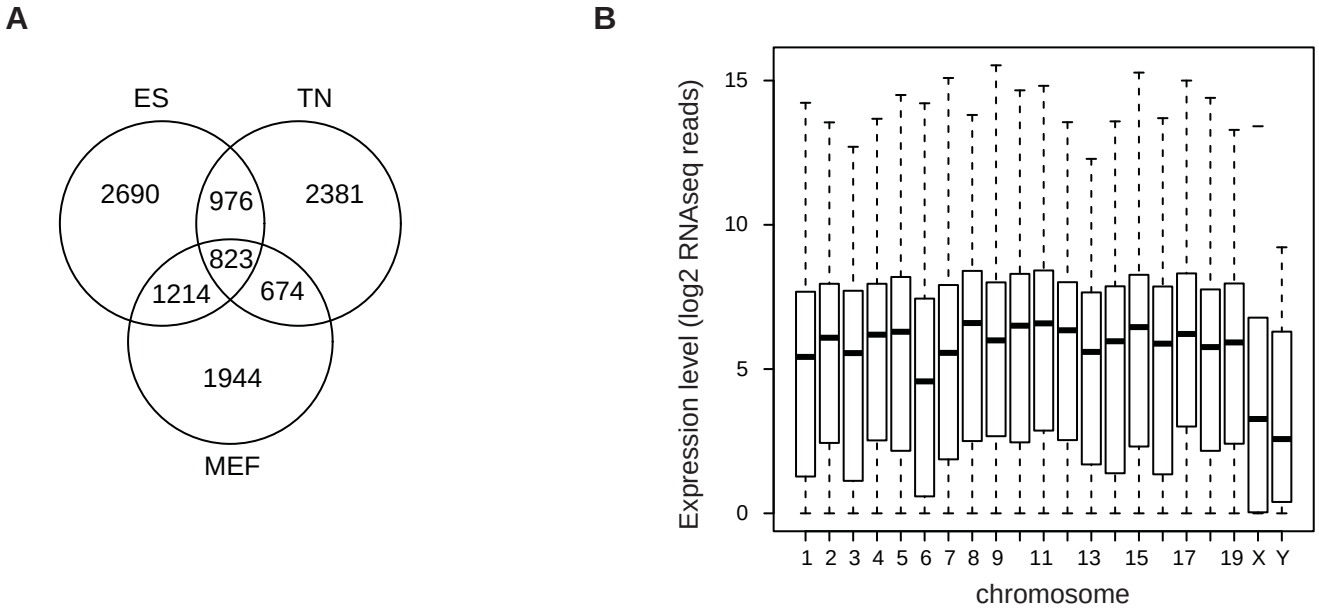


Figure S8.
(**A**) Venn diagram illustrating the overlap of transcripts expressed at a low level (log2 RNAseq reads between 2 and 5) in at least one of the examined cell types. Note that areas are not drawn to scale.
(**B**) Boxplots showing expression levels of transcripts per chromosomes.

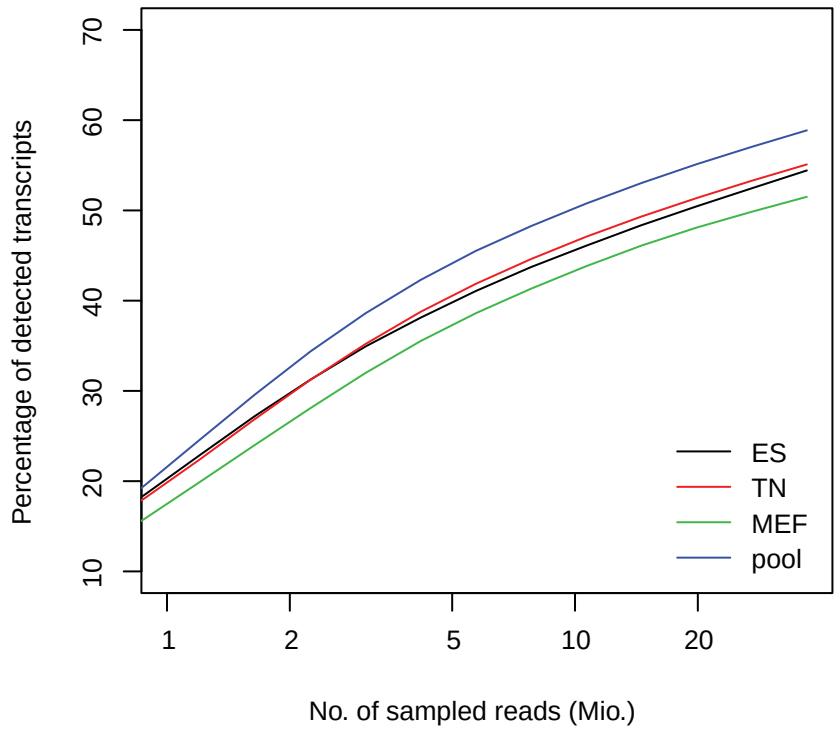


Figure S9. Absence of difference in transcriptome complexity is not a function of sample size in next generation sequencing.
Shown is the number of detected Refseq transcripts (at least 32 reads) relative to the the number of sampled reads. Number of detected transcripts is shown as the mean of a 100 different subsamples with the same read numbers. Irrespective of the sample size, a random sampling of RNAseq reads from ES cells (green line), neurons (dark green line) and MEFs (red line) revealed a smaller number of dected transcripts than an artificial sample with pooled reads from all three cell types (blue line).

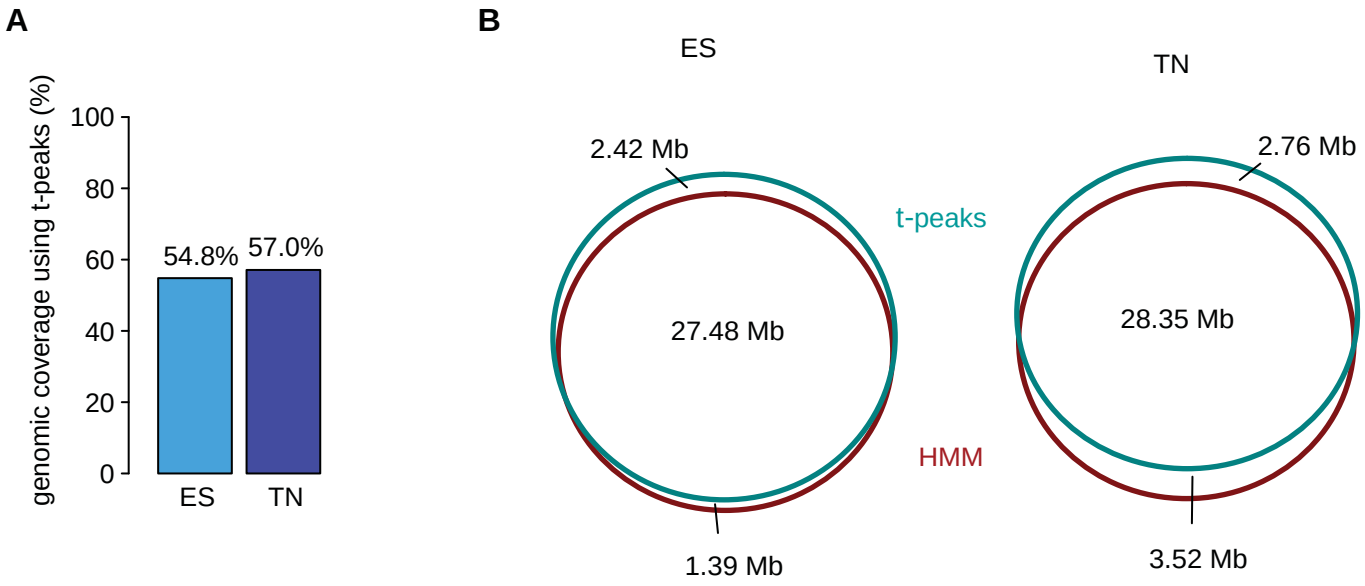


Figure S10. H3K9me2 domain definition by different statistical methods leads to similar results. (**A**) HMM-independent quantification of genomic coverage of H3K9me2 in ES cells and neurons. Shown is the percentage of chromosome 19 that lies within H3K9me2 peaks defined by a simple threshold method (see Methods). (**B**) Diagram illustrating the actual overlap between regions defined to be occupied by H3K9me2 by two different methods; a simple threshold method (t-peaks; see Methods) and by an HMM approach (HMM; see Methods).

Table S1. Real time PCR primers.

Genes	sense/antisense	Sequence
Gda	s	ACG ACG GCA CCA AAG AAT AC
	as	TTG CCA GAC AAG CAA TCA AC
Pcsk5	s	CAA TGA AGT GAT GCC CTC CT
	as	CAT GCT TGG CTA GTA GTG GAA
Hprt	s	CCA AGA CGA CCG CAT GAG AG
	as	CAA CGG AGT GAT TGC GCA TT
Oct_4	s	ACC TCC GTC TGG AAG ACA CA
	as	TCA CCT AGG GAC GGT TTC AC
Kif20b	s	GCC TCA AGG GGT TAA AGA GG
	as	TCT GCC CTA TAT GCA TGC TG
Hells	s	GAG AAT TGG GAC TGG AAG CA
	as	CCA GAG GAC CCA AGT TCA AA
Incenp	s	GCC TTG GTC TTG GAG AAA CA
	as	CTC ACA GGT CTG CCC ATT CT
Vsp13a	s	AGG CTT GTG TCT GAG CCA GT
	as	GCC TTT TTG ACG TCT TCA GC
Cbwd1	s	TTG GGA GCT GGT AAG ACA CA
	as	TCA CCA GCT CTT GCA AAA CA
Saps3	s	CTG GCC AAA TGT TAG CAC CT
	as	TCT AGG GAC CCT TGT GAT GG
Sgms1	s	CAG GGT TAC TGG ACC CCT TT
	as	AAC AAG CCC CTC ATC AAA TG
K9downA	s	AGC CCA GGG TGA GTC AGT T
	as	GCC ACC ATC ATA TCT CAC CA
K9downB	s	GGG GTA GAG GGA GAT GGA AA
	as	ATG CGC ATA CAG ATC ACG TC
K9downC	s	AAA GCA TCC TGG ACA TAA AAG AA
	as	GAA TCC TTT CGG CTT TAG GG

Table S2. Peptides used in dot blots in Figure S1.

protein	modification	sequence
H3 (1-20)	unmod	ARTKQTARKSTGGKAPRKQL-cys
H3 (1-20)	K4(Me)	ARTK(Me)QTARKSTGGKAPRKQL-cys
H3 (1-20)	K4(Me)2	ARTK(Me2)QTARKSTGGKAPRKQL-cys
H3 (1-20)	K4(Me)3	ARTK(Me3)QTARKSTGGKAPRKQL-cys
H3 (1-20)	K9(Me)	ARTKQTARK(Me)STGGKAPRKQL-cys
H3 (1-20)	K9(Me)2	ARTKQTARK(Me2)STGGKAPRKQL-cys
H3 (1-20)	K9(Me)3	ARTKQTARK(Me3)STGGKAPRKQL-cys
H3 (19-38)	unmod	QLATKAARKSAPATGGVKKP-cys
H3 (19-38)	K27(Me)	QLATKAARK(Me)SAPATGGVKKP-cys
H3 (19-38)	K27(Me)2	QLATKAARK(Me2)SAPATGGVKKP-cys
H3 (19-38)	K27(Me)3	QLATKAARK(Me3)SAPATGGVKKP-cys
H3 (25-45)	unmod	ARKSAPATGGVKKPHRYRPGT-cys
H3 (25-45)	K36(Me)	ARKSAPATGGVK(Me)KPHRYRPGT-cys
H3 (25-45)	K36(Me)2	ARKSAPATGGVK(Me2)KPHRYRPGT-cys
H3 (25-45)	K36(Me)3	ARKSAPATGGVK(Me3)KPHRYRPGT-cys
H4 (12-31)	unmod	KGGAKRHRKVLRDNIQGITK-cys
H4 (12-31)	K20(Me)	KGGAKRHRK(Me)VLRDNIQGITK-cys
H4 (12-31)	K20(Me)2	KGGAKRHRK(Me2)VLRDNIQGITK-cys
H4 (12-31)	K20(Me)3	KGGAKRHRK(Me3)VLRDNIQGITK-cys

3.2. Identification of genetic elements that autonomously determine DNA methylation states

Lienert F., Wirbelauer C., Som I., Dean A., Mohn F. and Schübeler D.

3.2.1. Summary

DNA methylation is an epigenetically inherited, repressive chromatin modification that has an important role in silencing of repeats, in genomic imprinting and during X-chromosome inactivation (Walsh and Bestor 1999). In mammals, genome-wide DNA methylation gets reprogrammed during germ cell development and early embryogenesis (Guibert et al. 2009). While most CpG-rich promoters remain unmethylated in all cell types, a subset of these regulatory regions gets *de novo* methylated during early development (Weber et al. 2007; Farthing et al. 2008; Illingworth et al. 2008; Meissner et al. 2008; Mohn et al. 2008; Borgel et al. 2010).

We here used a well-defined ES cell differentiation system (Bibel et al. 2004) to investigate how patterns of DNA methylation are established during loss of pluripotency and subsequent commitment to a neuronal lineage. In order to test whether promoter methylation is guided by the underlying DNA sequence, we chose a set of promoter regions with different DNA methylation levels and inserted them at an ectopic site in the genome. To do so, we made use of an insertion method that allows sequential insertions at the same genomic site without the need of a transcribed marker gene (Seibler et al. 1998). By following DNA methylation levels at ectopically inserted promoter elements during cellular differentiation we aimed to address the targeting principles of this repressive chromatin modification.

Our results show that 1 kb promoter elements are sufficient to recapitulate endogenous DNA methylation patterns in ES cells. This finding is in agreement with earlier reports on sequence mediated protection from DNA methylation at two different CpG island promoters in ES cells (Szyf et al. 1990; Brandeis et al. 1994; Macleod et al. 1994). Our study generalizes this finding and reveals that also differentiation-induced *de novo* methylation is determined by the underlying promoter sequence. We further define that DNA methylation patterns and their differentiation-induced changes are established independently of transcription. Moreover, insertion of numerous truncated promoter elements led to the identification of methylation determining regions (MDRs); sequence elements of varying size that are required for an unmethylated promoter state in ES cells. Mutations of predicted transcription factor binding sites in the MDR of Alf (*GTF2A1L*) suggest that protection from

DNA methylation at these sequence elements critically depends on TF binding. Specifically, we provide evidence that RFX2 represents a potential transcription factor involved in mediating an unmethylated state at the Alf MDR. In agreement with a general role of TFs in protecting from DNA methylation, it was recently reported that the zinc finger protein VEZF1 defines the unmethylated state of the chicken β-globin enhancer and the murine *APRT* CpG island (Dickson et al. 2010). Such TF mediated protection from DNA methylation might be facilitated by CXXC domain proteins that bind unmethylated CpGs and possess or recruit H3K4 methyltransferase activity (Thomson et al. 2010). However, CpG-richness alone does not seem to be sufficient for protection from DNA methylation, as the majority of ectopically inserted CpG-rich sequences from *E. coli* get methylated in ES cells. Remarkably, at promoters that get *de novo* methylated upon differentiation, MDRs encode both, protection from DNA methylation in ES cells as well as *de novo* methylation. We further show that MDRs can protect neighboring sequence from DNA methylation in ES cells and from *de novo* methylation during differentiation.

Based on these results we propose that TF factors are critically involved in locally protecting genomic regions from DNA methylation. We further speculate that differentiation-induced *de novo* methylation could be triggered by reduced binding of TFs that protect from DNA methylation.

nature genetics

Identification of genetic elements that autonomously determine DNA methylation states

Florian Lienert[1,2], Christiane Wirbelauer[1], Indrani Som[3], Ann Dean[3], Fabio Mohn[1,4] & Dirk Schübeler[1,2]

Cytosine methylation is a repressive, epigenetically propagated DNA modification. Although patterns of DNA methylation seem tightly regulated in mammals, it is unclear how these are specified and to what extent this process entails genetic or epigenetic regulation. To dissect the role of the underlying DNA sequence, we sequentially inserted over 50 different DNA elements into the same genomic locus in mouse stem cells. Promoter sequences of approximately 1,000 bp autonomously recapitulated correct DNA methylation in pluripotent cells. Moreover, they supported proper *de novo* methylation during differentiation. Truncation analysis revealed that this regulatory potential is contained within small methylation-determining regions (MDRs). MDRs can mediate both hypomethylation and *de novo* methylation in *cis*, and their activity depends on developmental state, motifs for DNA-binding factors and a critical CpG density. These results demonstrate that proximal sequence elements are both necessary and sufficient for regulating DNA methylation and reveal basic constraints of this regulation.

DNA methylation is an efficient repressor of transcriptional activity and represents a true epigenetic modification, as its mechanism of inheritance during the cell cycle is well established[1]. In mammals, DNA methylation is essential[2] and occurs almost exclusively at cytosines in the context of CpG dinucleotides. As a result of an increased mutation rate of methylated cytosines, most of the genome is depleted of CpGs except for small regions, which represent two-thirds of all mammalian promoters and are termed CpG islands[3]. Recent genome-wide analyses have corroborated that the majority of CpG islands are kept unmethylated at any given time during development, whereas most CpGs outside of CpG islands are methylated by default[4,5].

Several mechanisms have been proposed for how this global pattern of DNA methylation is established. Single-transgene studies have suggested that DNA-binding factors are involved in creating an unmethylated state[6–8]. In contrast, DNA methylation and certain active chromatin marks occur in a mutually exclusive manner[5,9,10]. Of note, in plants and fungi, a functional role for an interplay between DNA methylation and chromatin has been well established[11,12], making it possible that chromatin could be a determining factor in establishing global methylation patterns in mammals, as well[13].

Notably, DNA methylation patterns are not static. In mammals, a subset of promoters that are enriched for CpG islands of intermediate CpG density becomes *de novo* methylated in a cell lineage–dependent manner[9,14,15]. In this context, it has also been suggested that sequence-specific factors[16–19] and epigenetic modifications[20,21] may account for the observed selectivity of this *de novo* methylation process. However, these observations have not yet resulted in predictive models of dynamic DNA methylation, underscoring our limited understanding of the basic principles that govern this DNA modification. Such advanced knowledge of regulatory logic would further assist in the identification of potential mechanisms that underlie erroneous DNA methylation in disease-[22] or environment-induced changes, such as those observed in the brain upon stimulation[23].

To gain systematic insight into the constraints that define endogenous DNA methylation patterns, we integrated a multitude of different DNA elements into the same genomic locus in mouse embryonic stem (ES) cells. This experimental setup made it possible to control for chromosomal environment and also to dissect out potentially indirect effects, such as those mediated by ongoing transcription. This enabled us to measure the contribution of DNA sequence to the establishment of DNA methylation. Moreover, we were able to quantify the contribution of DNA sequence to methylation changes during cellular differentiation. In sum, we here demonstrate that proximal promoters show unexpected autonomy in defining their own DNA methylation states in pluripotent cells and in reprogramming these states during cellular differentiation. Furthermore, we show that this regulatory potential is genetically encoded by small sequence modules within these promoters.

RESULTS

Recapitulation of *Nanog* promoter methylation

To define sequence contribution to DNA methylation states, we used mouse embryonic stem cells as a cellular model. The methylation landscape of ES cells is comparable to epiblast cells, with established genome-wide methylation outside of CpG islands[24]. Furthermore, differentiating ES cells can recapitulate *in vivo* development and lineage-specific epigenetic reprogramming, as we have shown in an optimized system that generates highly purified neuronal progenitors[15,25].

[1]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. [2]Faculty of Science, University of Basel, Basel, Switzerland. [3]Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, US National Institutes of Health, Bethesda, Maryland, USA. [4]Present address: Institute of Molecular Biotechnology, Vienna, Austria. Correspondence should be addressed to D.S. (dirk@fmi.ch).

Received 8 April; accepted 25 August; published online 2 October 2011; doi:10.1038/ng.946

Here we have chosen the β-globin gene locus as a genomic site to test whether DNA methylation is DNA sequence dependent, and we uncouple this event from chromatin context and transcription. This well-studied locus is active only during erythropoiesis[26] and accordingly shows no transcriptional activity or presence of the active histone H3 lysine 4 dimethylation (H3K4me2) modification in either stem cells or derived neuronal progenitor cells[27] (**Fig. 1a**). In addition, this genomic region shows no occurrence of the repressive histone H3 lysine 27 trimethylation (H3K27me3) modification and harbors no CpG islands, suggesting that it represents an inert epigenetic environment in non-erythroid cells (**Fig. 1a**). We engineered a targeting site in this locus by replacing a 3.6-kb region around the *Hbb-y* gene with a selection cassette flanked by two inverted *loxP* sites via homologous recombination. This site enables efficient genomic insertions by recombinase-mediated cassette exchange (RMCE)[28–30]. Notably, any correct insertion will replace the selection marker during the recombination process and therefore circumvent the need of active transcription after targeting, which might otherwise override insert-driven epigenetic modifications (**Fig. 1b**).

We first inserted sequences from the well-characterized promoter region of *Nanog*; a gene encoding a transcription factor essential for maintaining the pluripotency of ES cells[31,32]. During differentiation, the *Nanog* promoter switches from a highly active, unmethylated state to a silent, methylated state[15,33]. A 5,712-bp *Nanog* sequence including the transcriptional start site (TSS), the proximal promoter and a strong enhancer element[34] recapitulated the methylation state of the endogenous locus after insertion in the β-globin locus. That is, both the promoter and the enhancer region of the transgene remained unmethylated in stem cells (**Fig. 1c** and **Supplementary Fig. 1**). Upon *in vitro* differentiation, the enhancer region remained unmethylated, whereas the proximal promoter became *de novo* methylated to once again mirror the endogenous locus (**Fig. 1c** and **Supplementary Fig. 1**). We conclude that a 5,712-bp sequence around the *Nanog* promoter is sufficient to recapitulate not only the unmethylated state in stem cells but also *de novo* methylation during differentiation.

To narrow down the functional DNA element within this 5,712-bp region, we next inserted a 1,013-bp fragment of the proximal promoter to test its behavior in the absence of the enhancer. Like the longer element, the 1,013-bp fragment was sufficient to recapitulate an unmethylated state in stem cells and a methylated state in neuronal progenitors (**Fig. 1d**). However, in contrast to the endogenous *Nanog* gene, the inserted sequence had no transcriptional activity and did not recruit RNA polymerase II in stem cells (**Fig. 1e**). This suggests that the generation of correct endogenous methylation patterns occurs independently of transcriptional activity. Also of note, we observed that the inserted fragment lost H3K4 dimethylation upon *de novo* DNA methylation in a similar fashion as the endogenous locus (**Fig. 1e**), in agreement with previous data[5,9]. This finding further suggests that sequences introduced ectopically into the β-globin locus can behave as their endogenous counterparts.

Autonomy of 1-kb promoter elements

To further investigate our observation of the epigenetic autonomy of the *Nanog* promoter, we similarly inserted nine additional promoter

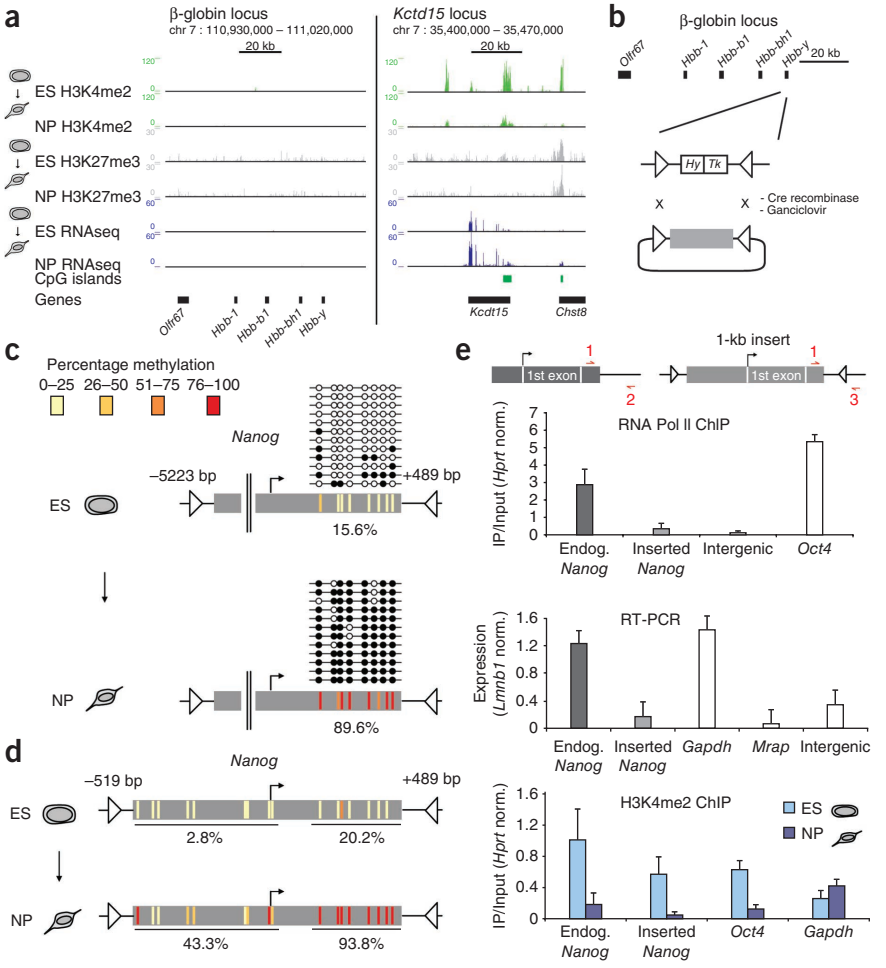


Figure 1 Ectopic *Nanog* promoter recapitulates the methylation state of the endogenous promoter. (**a**) RNAseq and chromatin immunoprecipitation sequencing (ChIPseq) tracks for H3K4me2 and H3K27me3[27]. Read counts (per 100 bp) are shown for the β-globin locus (left) and a randomly chosen region (kctd15; right) in ES cells and derived neuronal progenitors (NPs). The UCSC CpG island and gene tracks are shown below. (**b**) Sequence fragments (gray box) are integrated in the β-globin locus by Cre recombinase. The target site consists of two inverted *loxP* elements (triangles) flanking a fusion of a hygromycin-resistance (*Hy*) and a ganciclovir-sensitivity gene (*Tk*), which is replaced by the sequence of interest. (**c**) DNA methylation levels for single CpGs at the inserted 5.7-kb promoter fragment of *Nanog* are depicted as black (methylated) or white (unmethylated) circles. Every line corresponds to a sequenced bisulfite PCR amplicon. Colored vertical bars summarize these results, as defined by the color legend. (**d**) CpG methylation levels at the inserted 1-kb *Nanog* promoter fragment. (**e**) RNA Pol II occupancy and RNA levels (in ES cells) and H3K4me2 occupancy (in ES and NP cells) at the inserted 1-kb *Nanog* promoter fragment. ChIP enrichment for RNA Pol II and H3K4me2 was normalized to *Hprt*. RNA levels were determined using reverse transcription followed by real-time PCR and normalized to *Lmnb1*. Values at a methylated promoter and at an intergenic region are shown as a comparison. Error bars indicate s.d. from two independent biological replicates. The location of primers for the endogenous and inserted promoter is depicted above.

fragments ranging in size from ~700 to ~1,000 bp that had different CpG density to test the generality of this finding. The endogenous loci of these ten promoter sequences represent the different DNA methylation patterns that we have previously identified in our genome-wide survey[15], including maintaining high or low methylation levels in undifferentiated or differentiated cells or gaining methylation during cellular differentiation. Notably, nine out of ten inserted promoter fragments correctly recapitulated the endogenous methylation pattern in stem cells (**Fig. 2a** and **Supplementary Fig. 2**). The inserts recapitulated hypomethylated states, and promoters that are typically fully methylated at their endogenous loci were also correctly methylated when inserted. This rules out the possibility that the detected hypomethylation in stem cells is simply due to insertion at this particular ectopic site.

Of note, the sole fragment that did not recapitulate its endogenous epigenetic state was a 1,025-bp fragment encompassing the TSS of *Hes3*. In this case, the insert acquired methylation in stem cells, even though the endogenous locus is unmethylated. We therefore hypothesized that the fully functional promoter of this gene might be longer than 1,025 bp and tested this idea by inserting a 1,579-bp fragment that included a CpG-rich region downstream of the TSS (**Fig. 2b** and **Supplementary Fig. 2**). This longer fragment remained unmethylated (**Fig. 2b**), suggesting that the shorter fragment lacked a critical component.

Taken together, these data indicate that all ten promoter fragments tested, when of the correct length, direct their methylation levels in stem cells. Of note, the inserted fragments recapitulate the endogenous methylation with very high accuracy ($r = 0.93$, P value < 10^{-13}, **Fig. 2c**,**d** and **Supplementary Fig. 2**), underscoring their regulatory autonomy.

Genetic determination of dynamic DNA methylation

To ask whether the same regulatory logic determines DNA methylation states during cellular differentiation, we derived neuronal progenitors for each cell line containing one of the studied promoter inserts. Similar to the situation in stem cells, methylation levels of ectopically placed promoter fragments were in perfect accordance with their endogenous counterparts, showing either no change in DNA methylation or *de novo* methylation (**Fig. 2a** and **Supplementary Fig. 2**).

As for *Nanog*, all the inserted promoters that were tested recapitulated the endogenous H3K4 methylation pattern and its changes during differentiation (**Supplementary Fig. 3**).

These findings indicate that promoter DNA methylation is primarily regulated by *cis*-acting sequences and is thus genetically encoded. Our data also argue that methylation patterns and their changes can be established independently of transcription, as most studied promoters were inactive at their endogenous sites and when placed into the ectopic site (**Fig. 2a** and **Supplementary Fig. 4**). Further supporting the idea that maintenance of the unmethylated state is independent from transcriptional activity, the endogenously expressed *Nanog* and *Tcl1* recapitulated their dynamic DNA methylation patterns at the ectopic site without being transcriptionally active (**Fig. 1e** and **Supplementary Fig. 4**).

Methylation autonomy depends on critical regions

Given the ability of 1-kb elements to accurately recapitulate endogenous DNA methylation, we next asked if smaller fragments had similar capabilities. For this purpose, we inserted a total of 23 truncated variants derived from the initial set of unmethylated promoters. Unexpectedly, the truncated fragments behaved differently in many cases (**Fig. 3a** and **Supplementary Fig. 5**). Below a promoter size of approximately 700 bp, the otherwise unmethylated state became unstable, leading to increased DNA methylation in stem cells (**Fig. 3b**). This finding raises the possibility that the length of the insert is a critical determinant of its epigenetic state. Notably, our empirically identified size threshold is comparable to the average length of unmethylated regions around start sites reported in a recent genome-wide methylation survey[4]. We suggest that these regions in most cases harbor the *cis*-regulatory information sufficient for directing hypomethylation.

Our truncation experiments further revealed several smaller elements that contain all information necessary to recreate the correct endogenous methylation pattern. In the *Gtf2a1l* (also known as *Alf*) and *Hes3* promoters, these elements consist of 176 and 301 bp, respectively (**Fig. 3c** and **Supplementary Figs. 5** and **6**). For *Syt1*, *Nanog* and *Trf*, the critical region ranges in size from 515 to 720 bp (**Fig. 3c** and

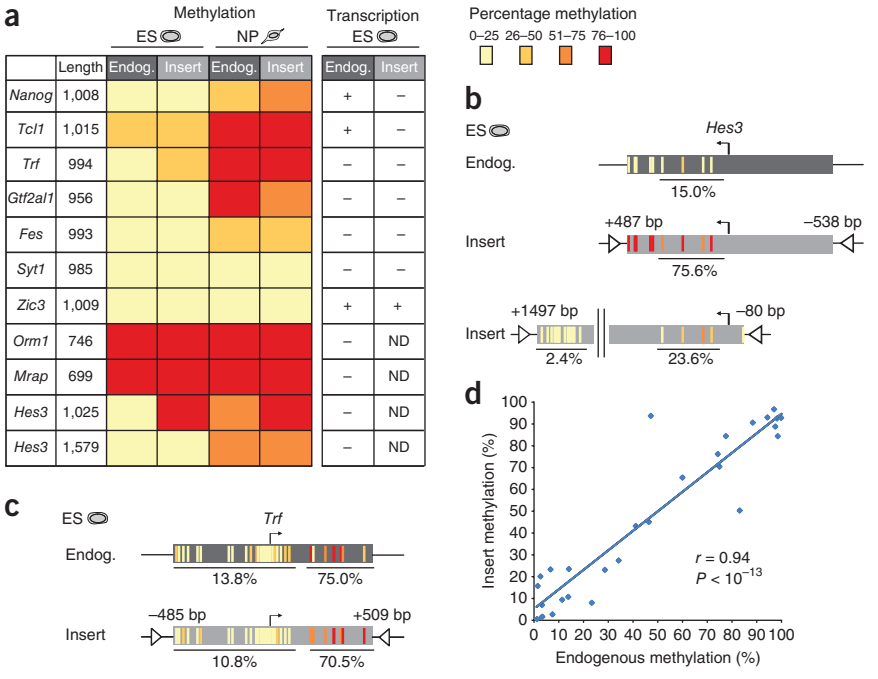


Figure 2 One-kb elements autonomously set DNA methylation state. (**a**) Summary table of inserted fragments. Length of fragments is given in bp. The heat map summarizes DNA methylation levels of endogenous and inserted promoters in ES cells and in NPs, as defined in the color legend. Transcriptional activity of endogenous and inserted promoters is indicated for ES cells. (Endog., endogenous promoter; Insert, inserted promoter element; ND, not determined). (**b**) CpG methylation levels in ES cells at the endogenous *Hes3* promoter and the inserted 1-kb and 1.6-kb *Hes3* fragments. Methylation levels of the regions indicated by a horizontal bar are shown below each fragment. (**c**) CpG methylation levels in ES cells at the endogenous and the ectopically inserted *Trf* promoter. (**d**) Comparison of methylation levels between inserted fragment and endogenous promoter, as determined by bisulfite PCR, illustrating the quantitative similarity in DNA methylation levels (see also **Supplementary Figs. 1** and **2**).

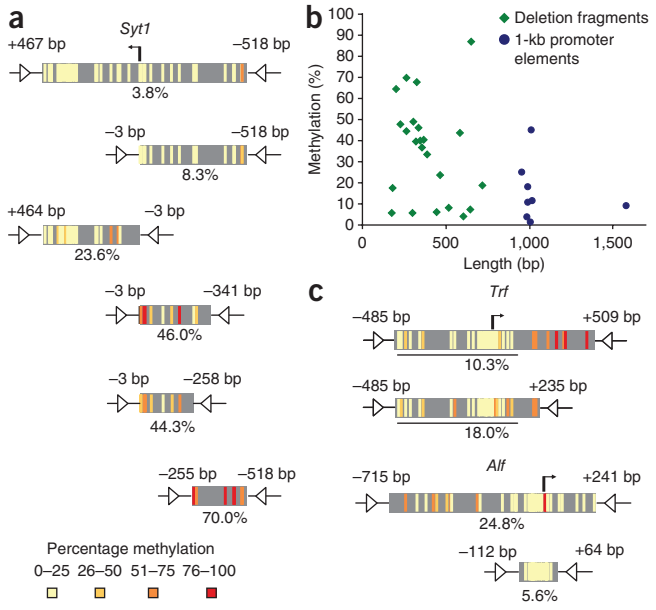


Figure 3 Truncation experiments identify methylation-determining regions. (**a**) CpG methylation levels in ES cells at the inserted 1-kb *Syt1* promoter and at five truncated versions. Deletion fragments are aligned to the largest insert. (**b**) Comparison of methylation levels in ES cells with length of inserted promoter elements and deletion fragments. (**c**) CpG methylation levels in stem cells at two inserted 1-kb promoters and at their respective MDRs.

Supplementary Fig. 5). Of note, in the cases of *Hes3* and *Syt1*, the respective sequences do not include the transcriptional start site, further supporting the notion that actual transcription is not needed for protection from DNA methylation. We conclude that the regulation of promoter hypomethylation depends on smaller embedded regions, which we refer to as methylation-determining regions (MDRs).

CpG density alone does not account for MDR function

To gain further insight into the regulatory logic of MDRs, we compared the sequence features of all promoter fragments and their DNA methylation in stem cells. From this analysis, we noticed a significant anti-correlation ($r = -0.49$, $P < 0.05$) between hypermethylation and the number of CpGs (**Fig. 4a**). We hypothesized that CpG density, as a distinguishing feature of CpG islands, might be critical to establish a hypomethylated state. Regardless of their CpG content, promoters contain many transcription factor–binding sites whose occupancy might be involved in regulating DNA methylation. Thus, we reasoned that a proper test for CpG dependency would require DNA sequence that contains no such binding sites. To maximize evolutionary distance and thus to minimize the likelihood of functional binding sites for mammalian transcription factors, we inserted ten random genomic regions from the prokaryote *Escherichia coli*. The tested *E. coli* sequences had an average length of 780 bp and varied in CpG density from 4.4 to 6.8 CpGs per 100 bp (**Supplementary Fig. 7**). This corresponds to a CpG density where almost all (>98%) endogenous mouse promoters are unmethylated[15] (M. Stadler, R. Murr, L. Burger and D.S., unpublished data). Unexpectedly, upon insertion seven out of ten *E. coli* fragments became fully methylated in stem cells, whereas the remaining three showed a level of methylation below 50% (**Fig. 4b** and **Supplementary Fig. 7**). These three fragments also showed enrichment for the H3K4me2 modification (**Supplementary Fig. 8**). Consistent with earlier findings that CpG density is highly correlated with hypomethylation[5,35], the inserted bacterial DNA shows a trend toward hypomethylation with increasing CpG content. However, despite the fact that all ten *E. coli* fragments have a CpG content equal to or greater than most unmethylated mouse promoters, the majority were still methylated. Together, these findings suggest that for most mouse promoters, CpG content alone cannot account for their unmethylated state *in vivo*.

Mutation of DNA-binding motifs impairs MDR function

We next tested whether motifs for DNA-binding factors, potentially in conjunction with CpG density, contribute to the unmethylated state in stem cells. To this end, we generated several 4-bp

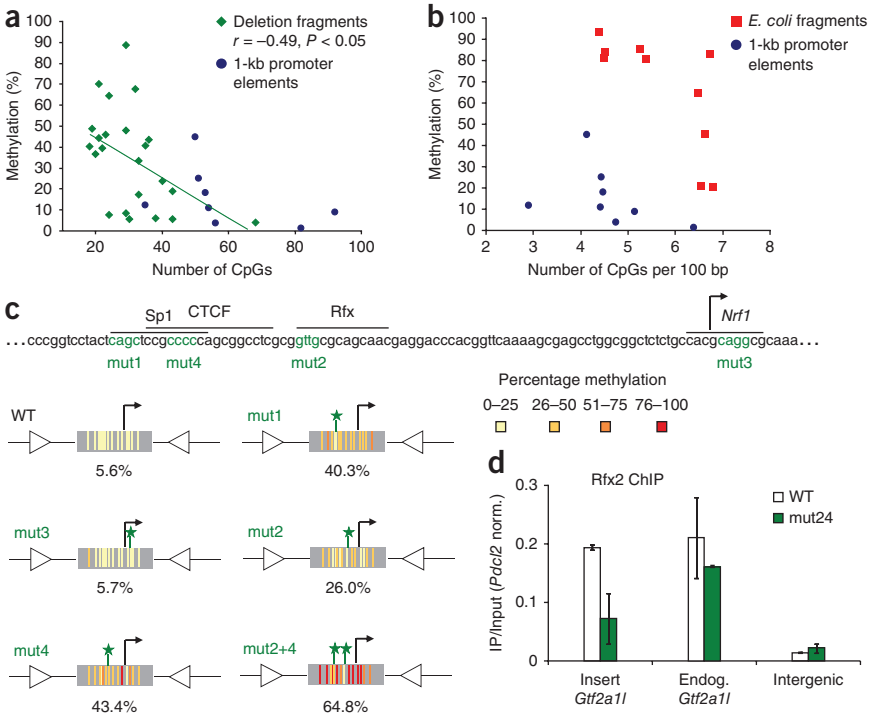


Figure 4 MDR function depends on CpG density and DNA-binding motifs. (**a**) Methylation level in ES cells plotted against number of CpGs at inserted promoter elements and deletion fragments. (**b**) Methylation level in ES cells plotted against CpG density at inserted promoter elements and *E. coli* sequence fragments. (**c**) Part of the inserted *Gtf2a1l* promoter sequence (104 of 176 bp) with predicted DNA-binding motifs[36,49] and mutation deletions (in green) indicated. CpG methylation levels in ES cells are shown for each mutated MDR fragment. (**d**) Rfx2 occupancy of the inserted *Gtf2a1l* MDR. ChIP enrichments were normalized to *Pdcl2*, which also contains a Rfx-binding site[38]. Error bars indicate s.d. from two independent biological replicates. WT, wild type.

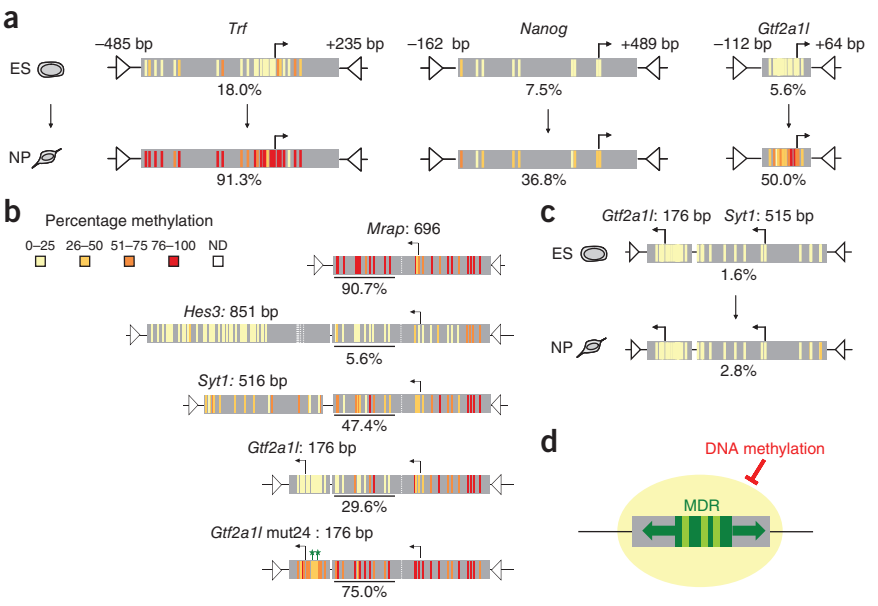


Figure 5 MDRs control *de novo* methylation and function *in cis* on heterologous DNA. (**a**) CpG methylation levels in ES cells and NPs at MDRs of promoters that get *de novo* methylated, showing that MDRs are properly reprogrammed. (**b**) CpG methylation levels in ES cells at a hybrid of the *Mrap* promoter fragment with MDRs of the *Syt1*, *Hes3* and *Gtf2a1l* promoters. In each case the MDR confers partial hypomethylation *in cis*. This effect is diminished in a hybrid with a mutated *Gtf2a1l* MDR. (**c**) CpG methylation levels at a hybrid of the *Gtf2a1l* and *Syt1* MDRs revealing inhibition of *de novo* methylation. (**d**) Model of MDR function in mediating hypomethylation. Sequence dependence of DNA methylation relies on MDR regions (dark green) that reside within hypomethylated promoters (gray box). MDR activity is dependent on DNA-binding motifs (light green bars) and mediates proximal hypomethylation *in cis*, probably by protecting against DNA methylation.

deletions in predicted DNA-binding motifs in the short MDR of the well-characterized *Gtf2a1l* promoter[36]. Notably, the deletions did not affect CpGs, allowing us to study the dependency of DNA methylation on DNA-binding motifs uncoupled from CpG density (**Fig. 4c**). We inserted these mutated fragments into the target site and determined their DNA methylation in ES cells (**Fig. 4c** and **Supplementary Fig. 9**). Three of four tested mutations in predicted DNA-binding motifs resulted in increased methylation at the promoter element compared to the original sequence. These mutations affected DNA sequences that were predicted to bind Sp1, CTCF or members of the Rfx winged-helix transcription factor family[36]. A combination of mutations affecting the SP1, CTCF and Rfx binding sites led to an even higher methylation level, suggesting an additive effect of factor binding in MDRs. As Sp1 and CTCF have been previously implicated in the regulation of DNA methylation[6,7,37], we tested whether Rfx factors might have a similar role. We measured binding by Rfx2, which has previously been reported to occupy the *Gtf2a1l* promoter in spermatocytes[38]. These experiments revealed that Rfx2 binds the *Gtf2a1l* MDR in ES cells and that this binding is strongly diminished when its binding site is mutated (**Fig. 4d**). We conclude that the unmethylated state of the *Gtf2a1l* MDR requires the combinatorial presence of DNA-binding motifs and presumably their corresponding factors. Of note, this effect is independent of transcription *per se*, because the inserted *Gtf2a1l* MDR is inactive in stem cells (**Supplementary Fig. 6**). Thus, DNA binding motifs in MDRs contribute to hypomethylation even in the absence of productive transcription.

MDRs regulate *de novo* methylation

Having identified MDRs and sequence features that mediate hypomethylation in stem cells, we next wanted to identify promoter regions that are crucial for inducing *de novo* methylation during cellular differentiation. We first tested the MDRs of the *Gtf2a1l*, *Trf* and *Nanog* promoters during differentiation from stem cells into neuronal progenitors (**Fig. 5a** and **Supplementary Fig. 10**). All three fragments were *de novo* methylated in progenitors similarly to their endogenous counterparts. We therefore conclude that MDRs require no additional sequence information for differentiation-induced *de novo* methylation.

MDRs act *in cis* on heterologous regions

Because MDRs are critical for the correct methylation of adjacent promoter sequences, we next asked whether they are able to confer DNA methylation states *in cis* on heterologous fragments. For these studies, we turned to the *Hes3* promoter, whose hypomethylation depends on an MDR downstream of the transcriptional start site (**Fig. 2b**). We fused the MDR of *Hes3* to the promoters of *Mrap* and *Orm1*, which when inserted individually are fully methylated in stem cells (**Fig. 2a**). When fused to the *Hes3* MDR, both the *Mrap* and *Orm1* promoter had much less DNA methylation (**Fig. 5b** and **Supplementary Fig. 11**), showing that an MDR associated with hypomethylation can independently confer this epigenetic state on adjacent heterologous sequences. In all cases examined, the induced hypomethylation occurred primarily at sequences directly adjacent to the fused MDR. This was not limited to the *Hes3* MDR, as the MDRs of *Syt1* and *Gtf2a1l* were equally capable of inducing hypomethylation when fused to *Mrap* and *Orm1* (**Fig. 5b** and **Supplementary Fig. 11**). We therefore conclude that the ability to confer hypomethylation on adjacent sequences seems to be a general feature of MDRs that induce an unmethylated state in stem cells. To test whether this ability to induce the hypomethylation of heterologous sequences in *cis* similarly relies on DNA-binding motifs, we fused the mutated *Gtf2a1l* MDR to *Mrap* and *Orm1*. In both cases, we saw a detectably weaker spread of the unmethylated state, suggesting that transcription factor–binding sites determine both the methylation state of an MDR and its ability to function *in cis* (**Fig. 5b** and **Supplementary Fig. 11**). Whereas the *Gtf2a1l* and *Syt1* MDRs are both capable of conferring hypomethylation on heterologous sequences in stem cells, they act in opposing ways during differentiation. *Gtf2a1l* induces *de novo* methylation and *Syt1* maintains hypomethylation (**Fig. 2a**). Thus, we wondered how these seemingly opposing activities would influence each other. We fused the two MDRs and inserted them into the β-globin locus. In stem cells, the hybrid of the two MDRs caused a hypomethylated state, as one would expect from the single insertion of each MDR (**Figs. 3a**,**c** and **5c** and **Supplementary Fig. 12**). The hybrid stayed unmethylated, even after differentiation into neuronal progenitors (**Fig. 5c** and **Supplementary Fig. 12**). This finding suggests that the *Syt1* MDR overrides the *de novo* methylation signal in the *Gtf2a1l*

MDR, further supporting the notion that MDRs are regulatory modules that enact hypomethylated states in *cis*.

DISCUSSION

The epigenetic nature of inheritance of symmetric DNA methylation is well understood in mammals, where Dnmt1 and Uhrf1 were shown to function in a *bona fide* copying mechanism[39]. However, the rules that govern the establishment of DNA methylation patterns remain undefined, despite emerging genome-wide data sets of DNA methylation patterns, their developmentally and environmentally driven variability and their correlations with histone modifications[4,5,9,14,15,35,40]. Using a highly controlled system in which we repeatedly target the same chromosomal locus, we show that promoter sequences of ~1,000 bp are generally sufficient to precisely recapitulate DNA methylation patterns in stem cells and to replicate the changes that occur during differentiation. These results argue that DNA methylation levels are genetically defined *in cis*. This is compatible with several previous single-transgene studies, which mostly did not separate effects arising from the genomic integration site from those caused by transcription[6–8]. Indeed, transcription has been suggested as a likely mediator of a hypomethylated state, as it provokes high local levels of activating histone modifications, such as H3K4 methylation[41]. Unexpectedly, we observed correct methylation of *cis*-regulatory regions that are transcriptionally inactive, suggesting that transcription does not play an essential part in the establishment of DNA methylation states.

We further show that DNA sequence–driven patterns rely on even smaller methylation determining regions (MDRs) that reside within promoter elements. These elements are necessary and sufficient for correct DNA methylation. On the basis of truncation and mutagenesis analyses, we show that DNA-binding motifs are critical for generating a hypomethylated state. This finding suggests that DNA-binding factors mediate hypomethylation without necessarily leading to active transcription. These results further argue against a locus-specific, distal regulation of DNA methylation by transcription, regulatory regions or chromatin states.

Given the overall absence of DNA methylation at CpG islands and the existence of proteins capable of recognizing unmethylated CpGs[42], it would be conceivable that CpG density alone is sufficient to mediate a hypomethylated state. Our observation that nonmammalian CpG-rich sequences are frequently methylated after insertion argues against such a scenario. We favor a model in which CpG density, in combination with DNA-binding motifs, generates functional MDRs capable of establishing a hypomethylated state (**Fig. 5d**). This idea is consistent with a recent evolutionary analysis of primate CpG islands that suggests that there is no particular selective constraint on CpGs in these regions[43].

We show that MDRs confer hypomethylation on larger promoter fragments, which is compatible with a model that they function via a protective mechanism that extends to neighboring sequences (**Fig. 5d**). In line with this model, we have shown that MDRs are capable of conferring a hypomethylated state on heterologous DNA in ES cells and that they can further protect against *de novo* methylation during differentiation. As this protective activity of MDRs seems to depend on DNA-binding motifs, it is conceivable that DNA methylation states are influenced by tissue-specific expression of *trans*-acting factors. In line with this notion, differentiation-induced *de novo* methylation of the *Gtf2a1l* MDR coincides with transcriptional downregulation of Rfx2 (data not shown), which binds this MDR in ES cells (**Fig. 4d**). The expressed repertoire of *trans*-acting factors might therefore explain the observed tissue specificity of promoter DNA methylation[4,5,9,14,15,44], as well as the variability of DNA methylation in regions adjacent to CpG islands, referred to as CpG island shores[45–47]. Similarly, loss of MDR protective activity, for example, through loss of DNA-binding factors in disease, might define aberrant targets of *de novo* methylation, as has been suggested in cancer[48].

These findings do not exclude the possibility that chromatin structure is crucial in mediating local DNA methylation; however, our results show that the local DNA sequence is the primary determinant of target specification for DNA methylation in mammals. This genetic determination further predicts that sequence variation between individuals can contribute to differential DNA methylation patterns, which needs to be taken into account by any study linking DNA methylation differences to phenotypes.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are grateful to M. Pietrzak for sequencing. We thank M. Lorincz of the University of British Columbia–Vancouver for providing plasmids for RMCE and S. Fiering for advice. We would also like to thank members of the Schübeler group and S. Gasser for critical comments on the manuscript. F.L. is supported by a PhD fellowship of the Boehringer Ingelheim Fonds. Research in the laboratory of A.D. is supported by the Intramural Program of National Institute of Diabetes and Digestive and Kidney Diseases, US National Institutes of Health. Research in the laboratory of D.S. is supported by the Novartis Research Foundation, by the European Union (NoE "EpiGeneSys" FP7-HEALTH-2010-257082, LSHG-CT-2006-037415), the European Research Council (ERC-204264) and by the RTD "Cellplasticity" of the Swiss initiative in Systems Biology (SystemsX.ch).

AUTHOR CONTRIBUTIONS

F.L. and C.W. performed experiments. I.S. and A.D. generated the target ES cell line. F.L., F.M. and D.S. designed the study, analyzed data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Law, J.A. & Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
2. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
3. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
4. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
5. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).
6. Brandeis, M. *et al.* Sp1 elements protect a CpG island from *de novo* methylation. *Nature* **371**, 435–438 (1994).
7. Macleod, D., Charlton, J., Mullins, J. & Bird, A.P. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* **8**, 2282–2292 (1994).
8. Dickson, J. *et al.* VEZF1 elements mediate protection from DNA methylation. *PLoS Genet.* **6**, e1000804 (2010).
9. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
10. Hawkins, R.D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
11. Tamaru, H. & Selker, E.U. A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature* **414**, 277–283 (2001).
12. Jackson, J.P., Lindroth, A.M., Cao, X. & Jacobsen, S.E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**, 556–560 (2002).

13. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10**, 295–304 (2009).
14. Farthing, C.R. *et al.* Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* **4**, e1000116 (2008).
15. Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).
16. Brenner, C. *et al.* Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J.* **24**, 336–346 (2005).
17. Suzuki, M. *et al.* Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. *Oncogene* **25**, 2477–2488 (2006).
18. Sato, N., Kondo, M. & Arai, K. The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing. *Biochem. Biophys. Res. Commun.* **344**, 845–851 (2006).
19. Velasco, G. *et al.* Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl. Acad. Sci. USA* **107**, 9281–9286 (2010).
20. Zhao, Q. *et al.* PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat. Struct. Mol. Biol.* **16**, 304–311 (2009).
21. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2006).
22. Robertson, K.D. DNA methylation and human disease. *Nat. Rev. Genet.* **6**, 597–610 (2005).
23. Dulac, C. Brain function and chromatin plasticity. *Nature* **465**, 728–735 (2010).
24. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* **42**, 1-93–1100 (2010).
25. Bibel, M. *et al.* Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat. Neurosci.* **7**, 1003–1009 (2004).
26. Fromm, G. & Bulger, M. A spectrum of gene regulatory phenomena at mammalian beta-globin gene loci. *Biochem. Cell Biol.* **87**, 781–790 (2009).
27. Lienert, F. *et al.* Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS Genet.* **7**, e1002090 (2011).
28. Feng, Y.Q. *et al.* Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. *J. Mol. Biol.* **292**, 779–785 (1999).
29. Schübeler, D. *et al.* Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation. *Mol. Cell. Biol.* **20**, 9103–9112 (2000).
30. Lorincz, M.C., Schubeler, D., Hutchinson, S.R., Dickerson, D.R. & Groudine, M. DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent *de novo* methylation. *Mol. Cell. Biol.* **22**, 7572–7580 (2002).
31. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
32. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
33. Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M. & Walker, P.R. Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. *J. Biol. Chem.* **280**, 6257–6260 (2005).
34. Levasseur, D.N., Wang, J., Dorschner, M.O., Stamatoyannopoulos, J.A. & Orkin, S.H. Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev.* **22**, 575–580 (2008).
35. Illingworth, R. *et al.* A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* **6**, e22 (2008).
36. Kim, M. *et al.* Regulatory factor interactions and somatic silencing of the germ cell–specific *ALF* gene. *J. Biol. Chem.* **281**, 34288–34298 (2006).
37. Pant, V. *et al.* The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev.* **17**, 586–590 (2003).
38. Horvath, G.C., Kistler, M.K. & Kistler, W.S. RFX2 is a candidate downstream amplifier of A-MYB regulation in mouse spermatogenesis. *BMC Dev. Biol.* **9**, 63 (2009).
39. Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908–912 (2007).
40. Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D. & Pfeifer, G.P. A human B cell methylome at 100–base pair resolution. *Proc. Natl. Acad. Sci. USA* **106**, 671–678 (2009).
41. Deaton, A.M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
42. Thomson, J.P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082–1086 (2010).
43. Cohen, N.M., Kenigsberg, E. & Tanay, A. Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell* **145**, 773–786 (2011).
44. Schilling, E. & Rehli, M. Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics* **90**, 314–323 (2007).
45. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* **41**, 1350–1353 (2009).
46. Irizarry, R.A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41**, 178–186 (2009).
47. Ji, H. *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338–342 (2010).
48. Gebhard, C. *et al.* General transcription factor binding at CpG islands in normal cells correlates with resistance to *de novo* DNA methylation in cancer cells. *Cancer Res.* **70**, 1398–1407 (2010).
49. Pachkov, M., Erb, I., Molina, N. & van Nimwegen, E. SwissRegulon: a database of genome-wide annotations of regulatory sites. *Nucleic Acids Res.* **35**, D127–D131 (2007).

ONLINE METHODS

Cell culture. TC-1 ES cells (background 129S6/SvEvTac) were cultured and differentiated as previously described[15,50].

Homologous recombination. The pZRMCE plasmid used for targeting mouse TC-1 ES cells (background 129S6/SvEvTac) in the β-globin locus was constructed in the pZERO multiple cloning site (MCS) and included a 2.4-kb NotI–XhoI fragment designated upstream arm (from positions –3,700 to –1,300 relative to the *Hbb-y* ATG start) and a 3.0-kb KpnI–NotI downstream arm (positions +2,332 to +5,432) cloned 5′ and 3′, respectively, to the selection cassette that was flanked by inverted *loxP* sites (L1-Hyg^{R}-TK-1L[28]). Further 5′ to the upstream arm at the FspI site of the vector, a SalI–ClaI fragment containing the gene encoding diptheria toxin A (DTA) was inserted.

TC-1 ES cells were electroporated with 100 µg of pZRMCE plasmid using a Bio-Rad Gene Pulser (at 500 µF and 250 V/cm). Cells were selected with 150 µg/ml hygromycin for 7–10 d after transfection. Clones were tested for successful recombination events by Southern blot analysis.

Recombinase-mediated cassette exchange. For targeted insertion, DNA fragments were cloned into a plasmid containing a multiple cloning site flanked by two inverted L1 *Lox* sites (a kind gift from M. Lorincz). Promoter regions were amplified from TC-1 ES cell genomic DNA. Start and end sites relative to the TSS of promoter fragments are depicted in **Supplementary Figures 1, 2, 5, 9, 10, 11** and **12**, with the TSS corresponding to the following genomic coordinates (mm9): *Nanog* (chr6:122′657′610), *Tcl1* (chr12:106′460′914), *Trf* (chr9:103′132′448), *Gtf2a1l* (chr17:89′068′017), *Fes* (chr7:87′532′781), *Syt1* (chr10:108′448′031), *Zic3* (chrX:55′283′805), *Orm1* (chr4:63′005′600), *Mrap* (chr16:90′738′569) and *Hes3* (chr4:151′665′771).

RMCE was performed as described[28] with slight modifications. TC-1 ES cells were selected under hygromycin (250 µg/ml, Roche) for 10 d. Next, 4×10^6 cells were electroporated (Amaxa nucleofection, Amaxa) with 25 µg of L1-promoter-1L plasmid and 15 µg of pIC-Cre (a kind gift from R. Terranova). Selection with 3 µM Ganciclovir (Roche) was started 2 d after transfection and continued for 7–10 d. Clones were tested for successful insertion events by PCR and Southern blot analysis.

Chromatin immunoprecipitation. ChIP experiments were performed as described[15], starting with 70 µg of chromatin and 5 µg of antibodies to the following: RNA Pol II (Santa Cruz Biotechnology, no. SC899), dimethyl-H3K4 (Upstate, no. 07-030) and Rfx2 (Santa Cruz Biotechnology, no. SC10557). Real-time PCR was performed using SYBR Green chemistry (Applied Biosystems) and 1/80 of the ChIP reaction or 20 ng of input chromatin per PCR reaction. Primer sequences are listed in **Supplementary Table 1**.

RT-PCR. RNA was isolated using TRIzol (Invitrogen) and subsequently DNase digested (RQ1 RNase-free DNase, Promega). First-strand cDNA was generated using random hexamers (Superscript III, Invitrogen) and analyzed by real-time PCR using SYBR Green chemistry. Expression levels of primary transcripts were calculated as $1 / C_t$, followed by subtraction of the control lacking reverse transcriptase and normalization for amplification efficiency. Primer sequences are listed in **Supplementary Table 1**.

Bisulfite sequencing. Genomic DNA (2 µg) was bisulfite converted with the EpiTec Bisulfite Kit (QIAGEN). Regions of interest were amplified by PCR and cloned by TOPOTA cloning (Invitrogen). Sequences were analyzed using BiQ Analyzer[51]. Primer sequences for PCR are listed in **Supplementary Table 1**.

50. Bibel, M., Richter, J., Lacroix, E. & Barde, Y.A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protoc.* **2**, 1034–1043 (2007).
51. Bock, C. *et al.* BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* **21**, 4067–4068 (2005).

doi:10.1038/ng.946

SUPPLEMENTARY INFORMATION

Identification of genetic elements that autonomously determine DNA methylation states

Florian Lienert[1,2], Christiane Wirbelauer[1], Indrani Som[3], Ann Dean[3], Fabio Mohn[1§] and Dirk Schübeler[1,2]

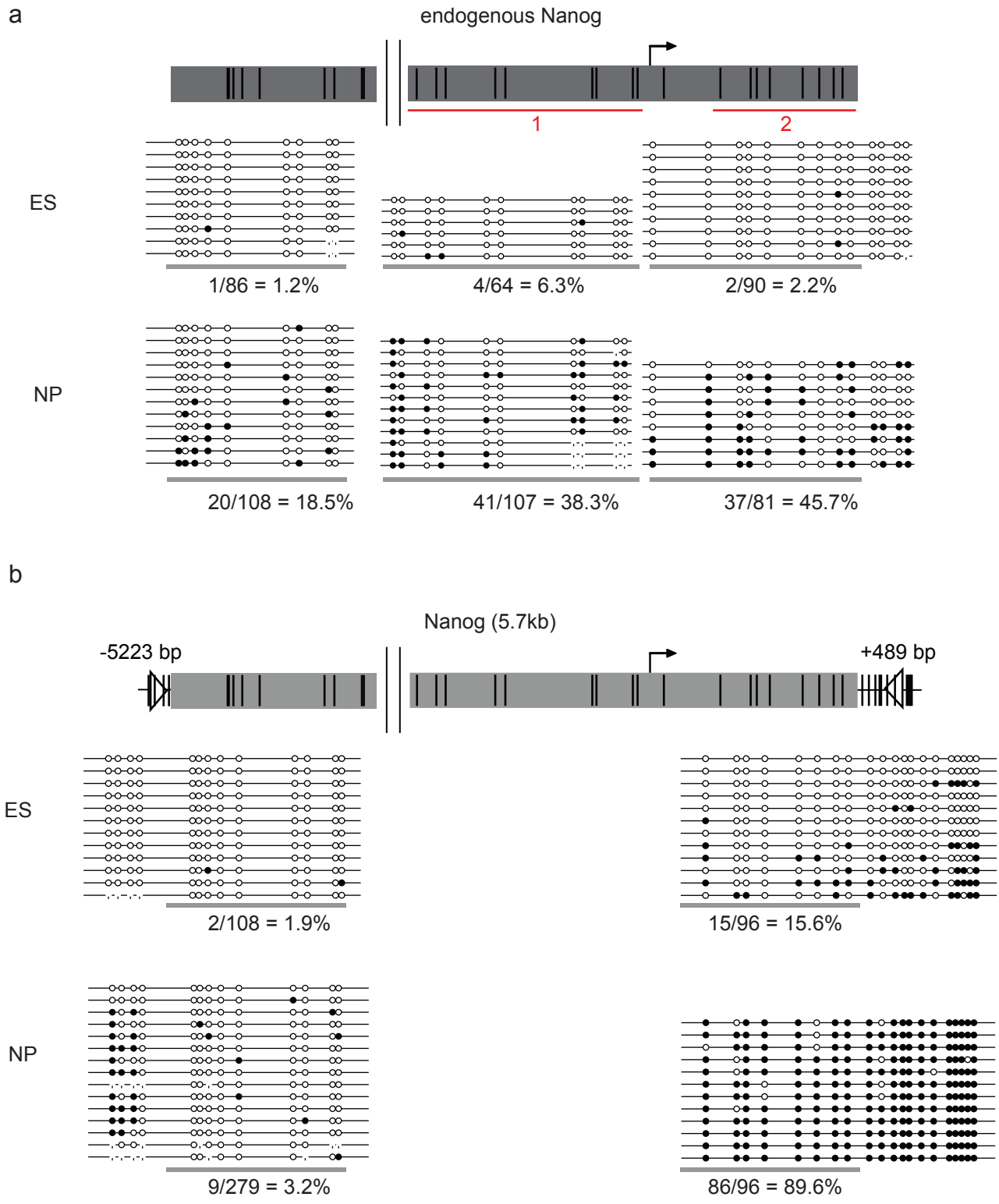
[1] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH 4058 Basel, Switzerland
[2] Faculty of Science, University of Basel, Klingelbergstrasse 50, CH 4056 Basel, Switzerland
[3] National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Laboratory of Cellular and Developmental Biology, 50 South Dr., Bethesda, MD, 20892, USA

[§] Present address: Institute of Molecular Biotechnology GmbH, Dr. Bohr-Gasse 3,

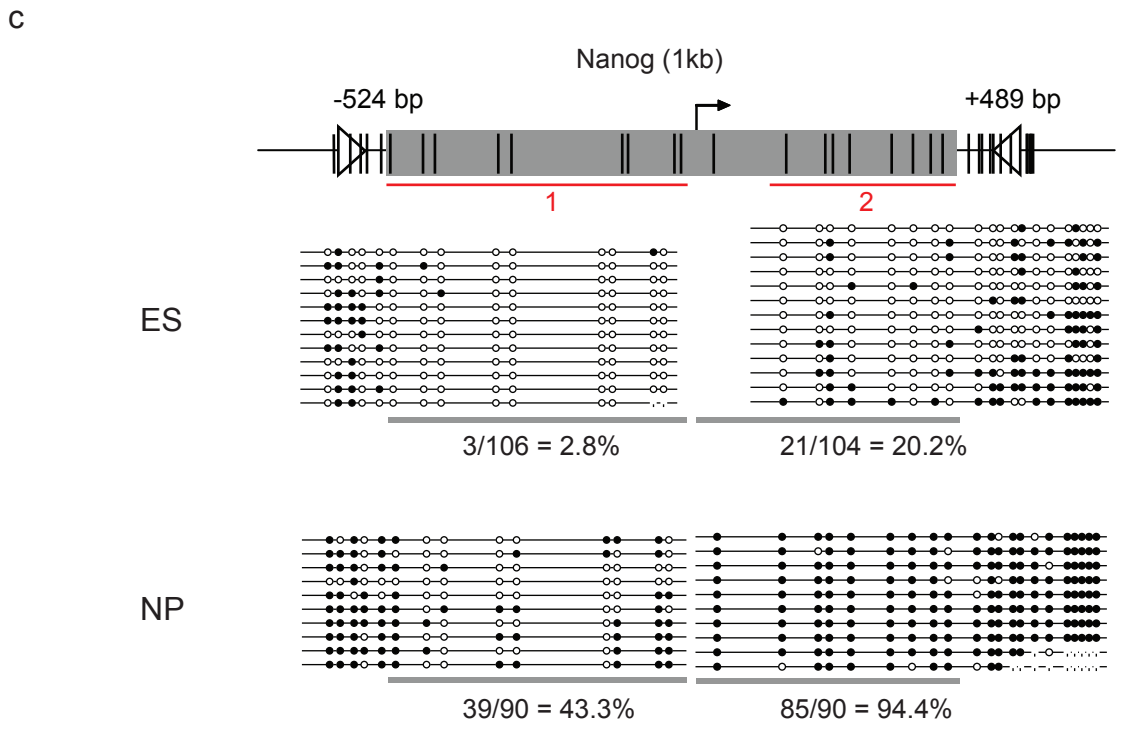
1030 Vienna, Austria

Correspondence should be addressed to D.S. (dirk@fmi.ch).

Supplementary Figure 1

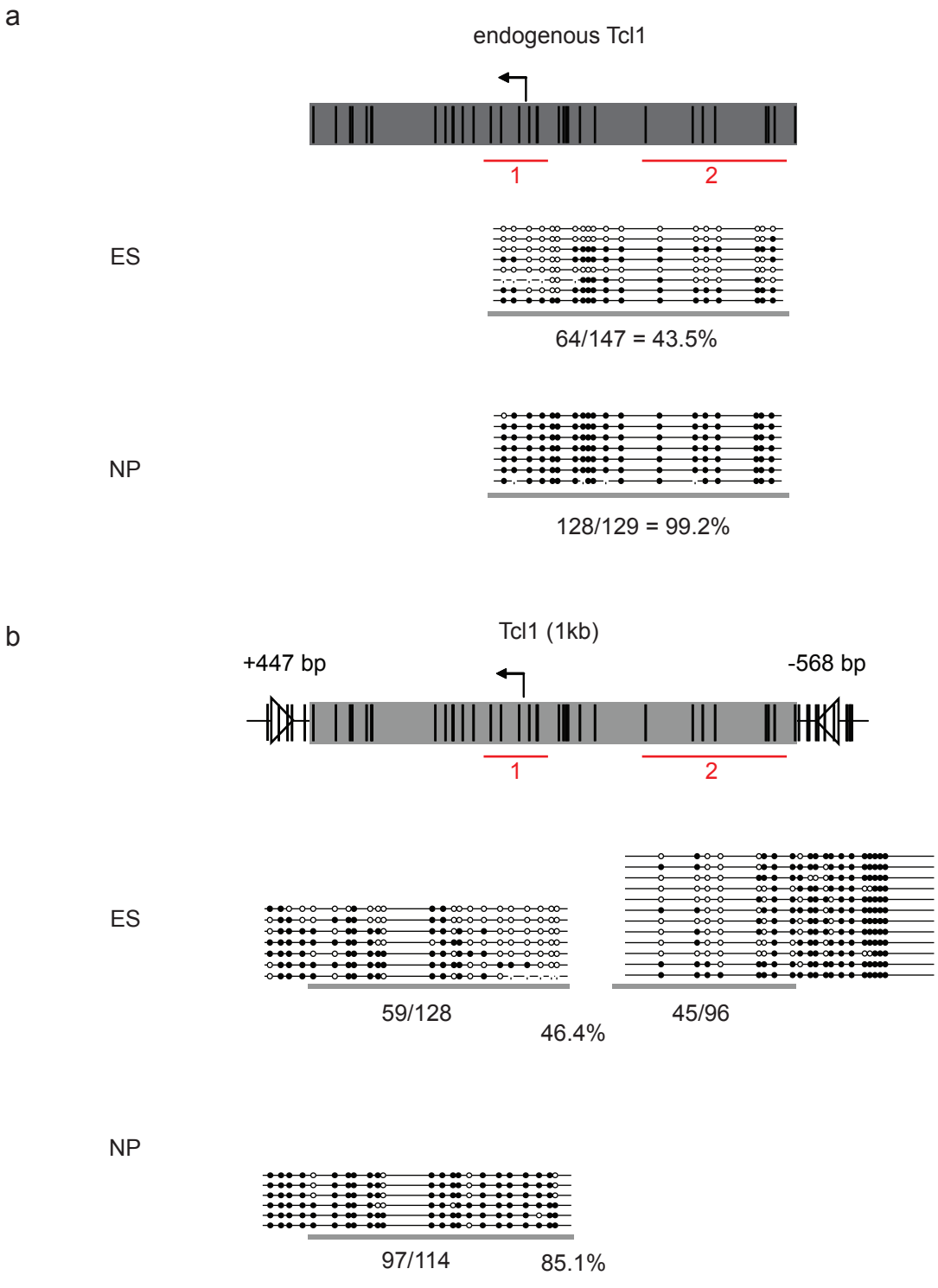


Supplementary Figure 1

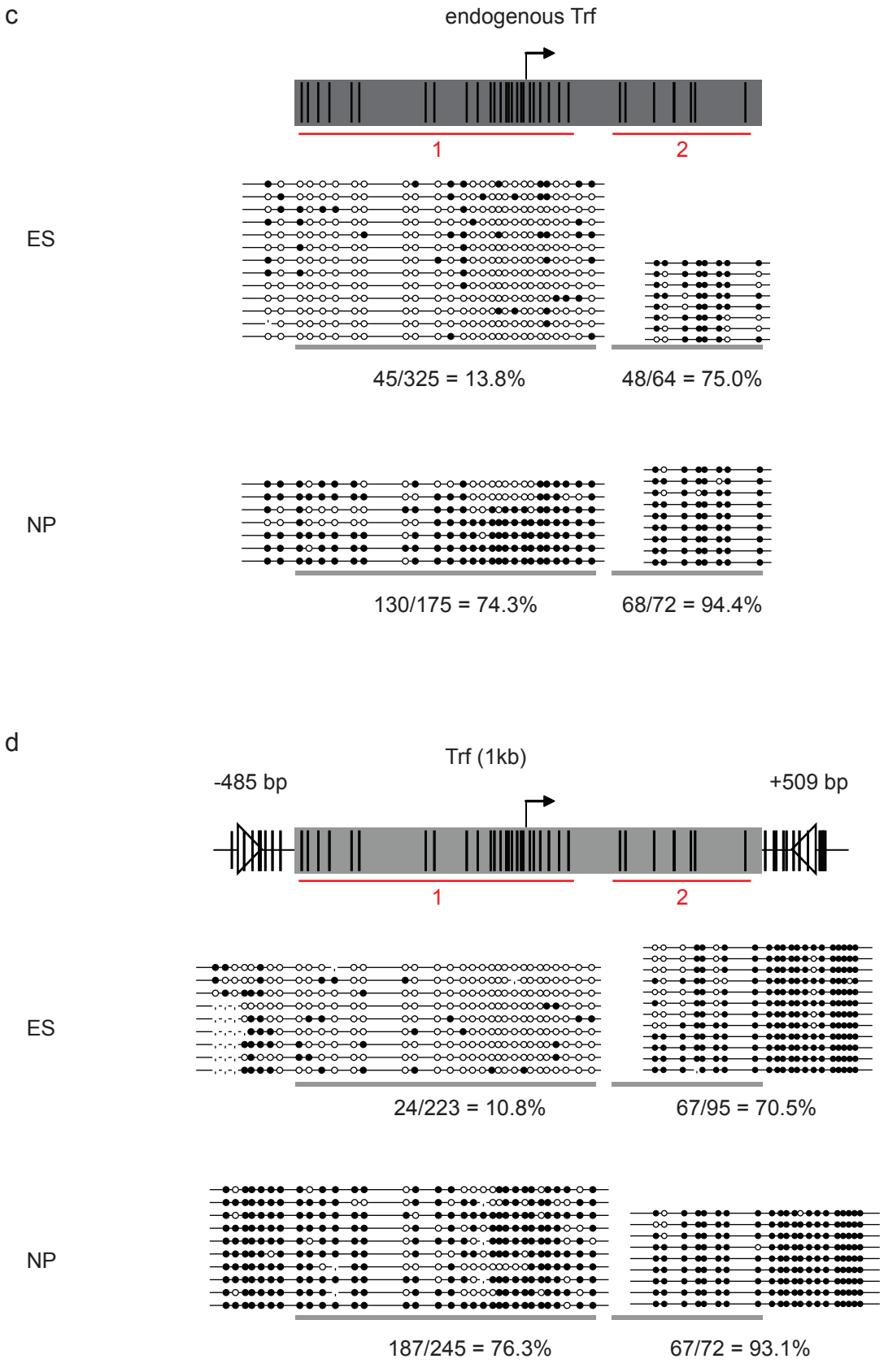


Supplementary Figure 1 Recapitulation of DNA methylation at ectopic Nanog promoter. (**a-c)** DNA methylation levels are shown for the endogenous Nanog locus (a) and the inserted 5.7 kb (b) and 1 kb (c) promoter fragments in stem cells (ES) and neuronal progenitors (NP). CpGs are depicted as vertical bars and the two loxP sites by triangles. The gray box indicates sequence regions corresponding to the endogenous promoter with start and end of the region relative to the transcriptional start site indicated above. DNA methylation levels for single CpGs are depicted as black (methylated) or white (unmethylated) circles. Every line corresponds to a sequenced bisulfite PCR amplicon. The numbers indicate percentage of methylated CpGs in the promoter regions highlighted by a gray bar. Red bars indicate regions whose methylation levels are compared in Supplementary Fig. 2t.

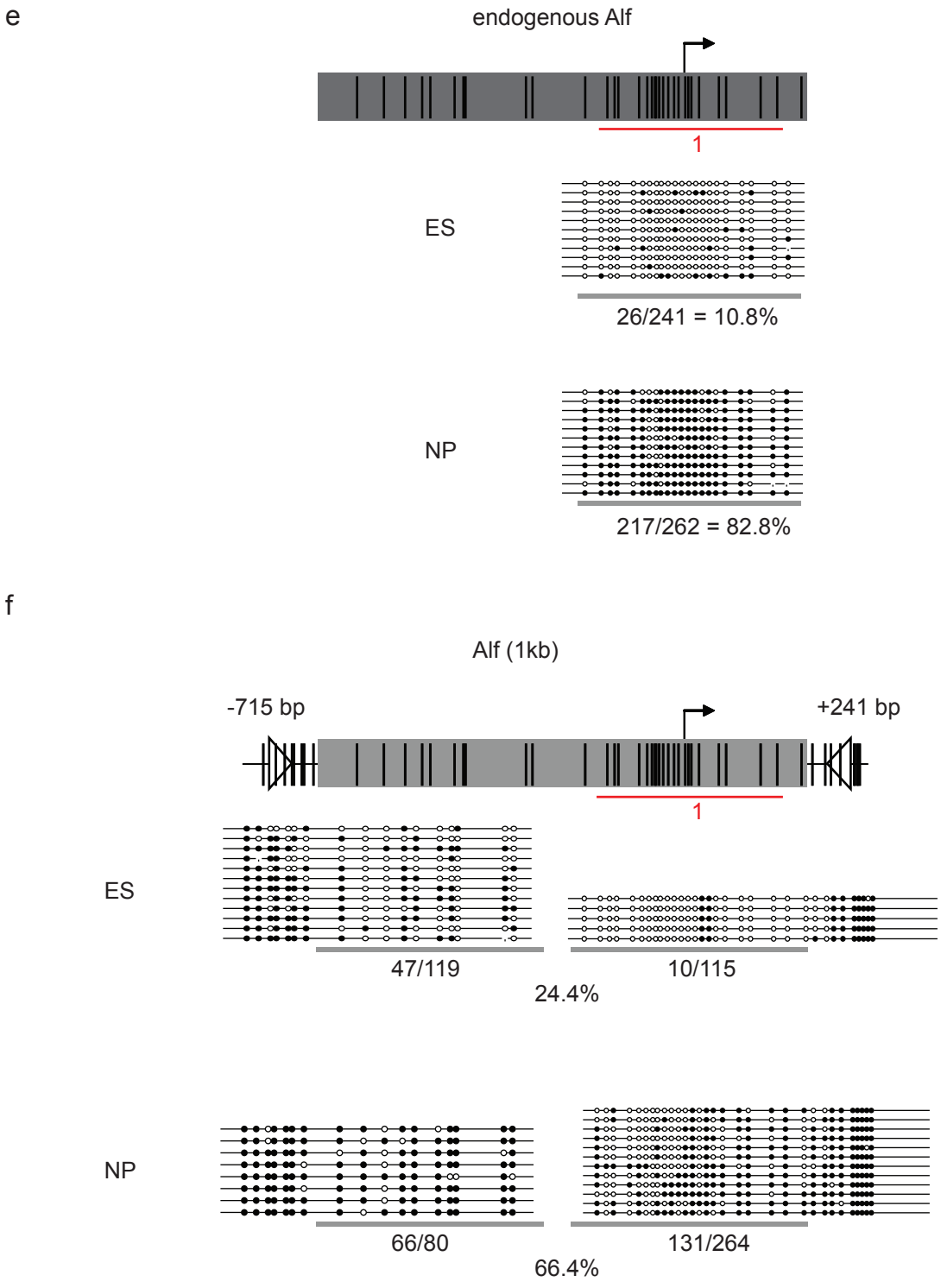
Supplementary Figure 2



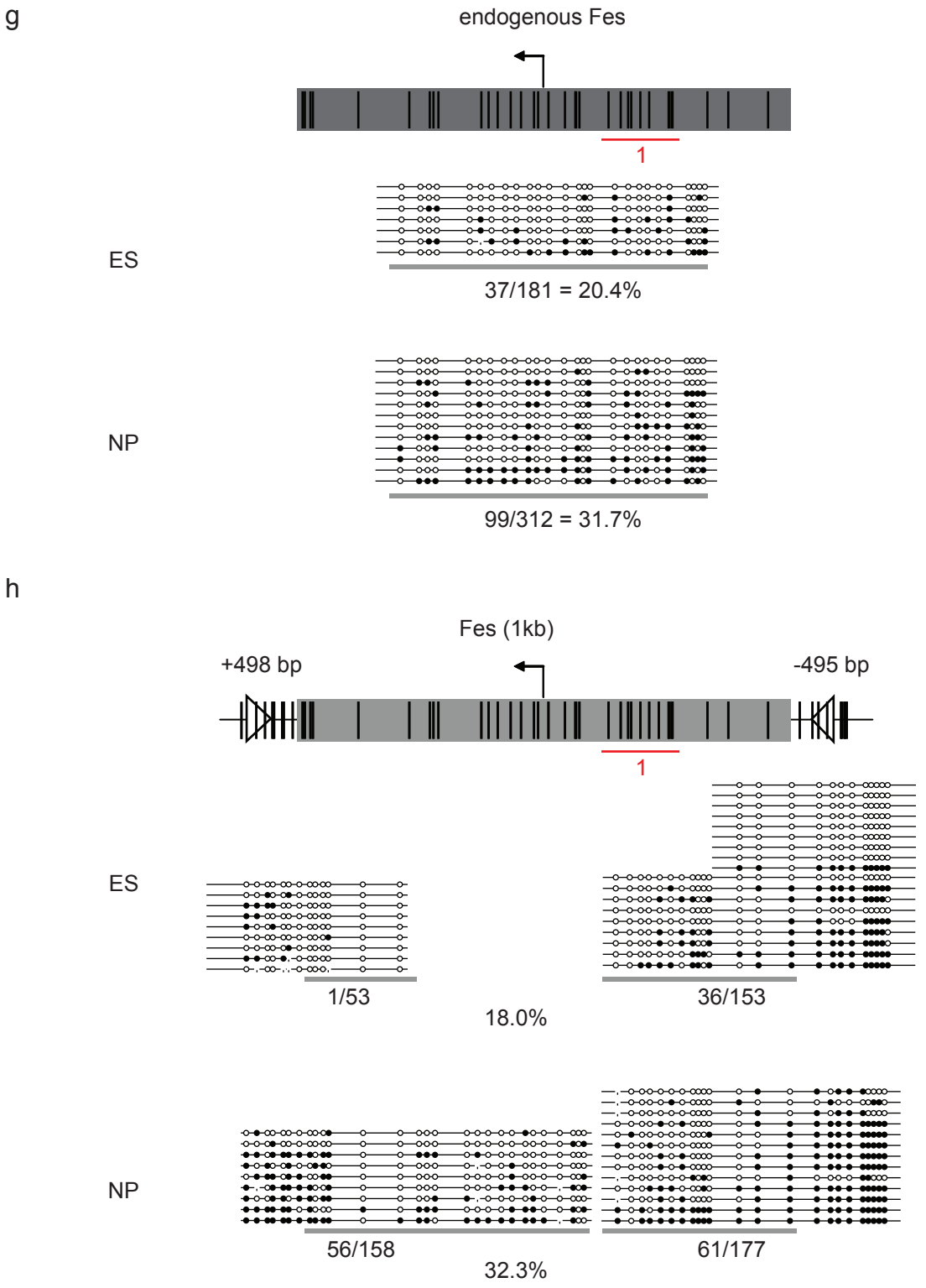
Supplementary Figure 2



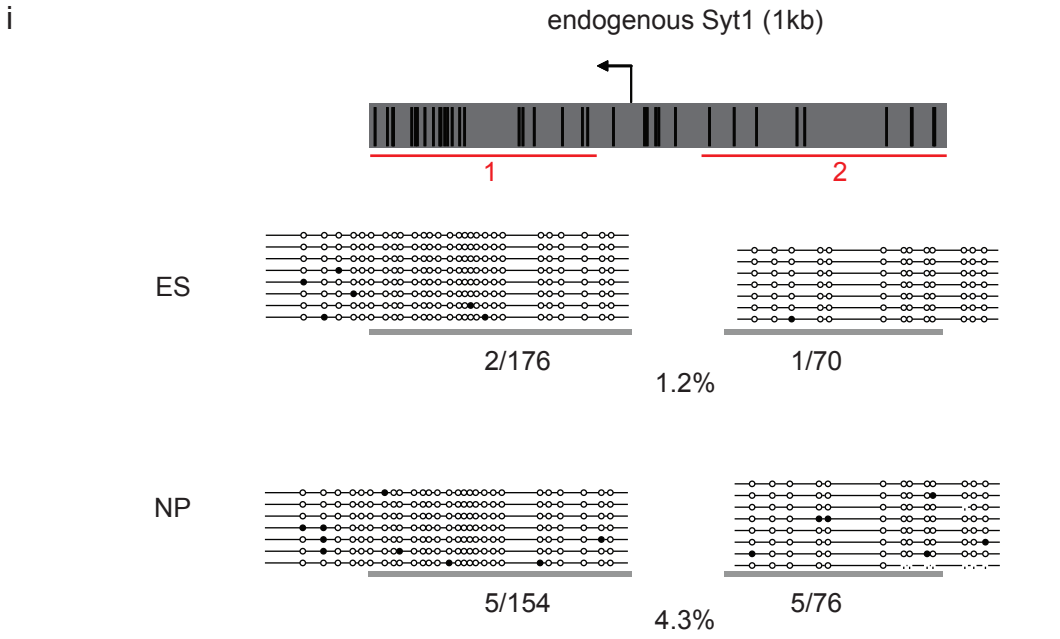
Supplementary Figure 2

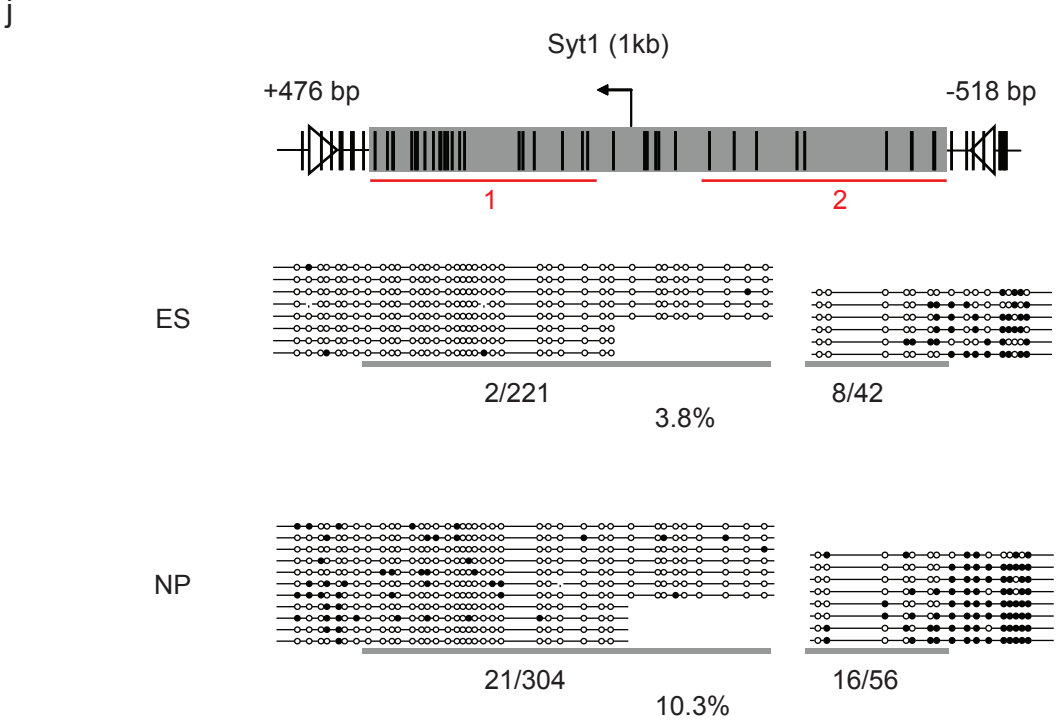


Supplemetary Figure 2



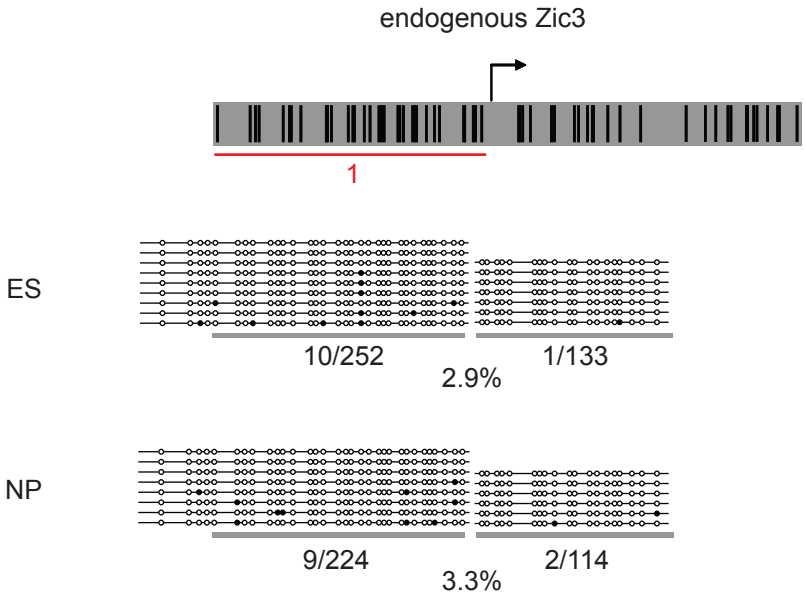
Supplementary Figure 2



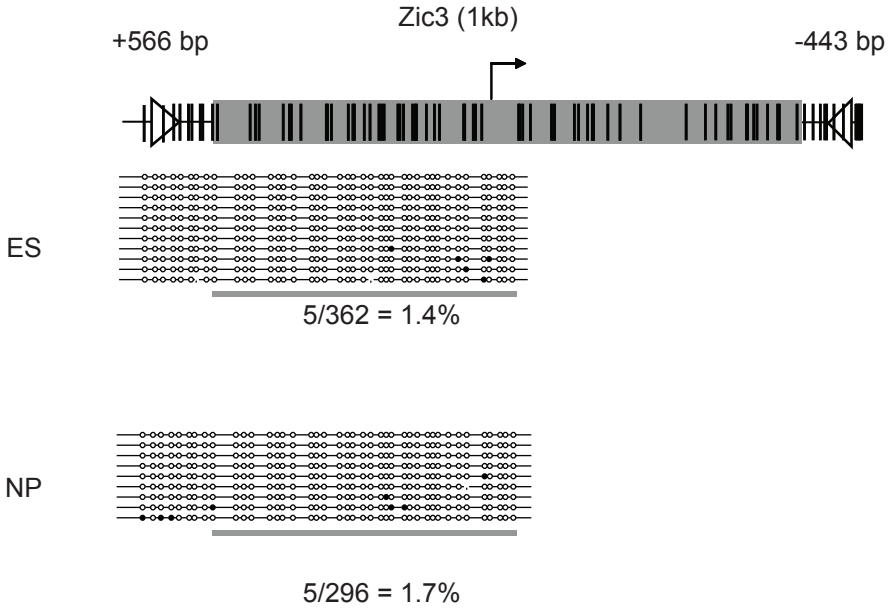


Supplementary Figure 2

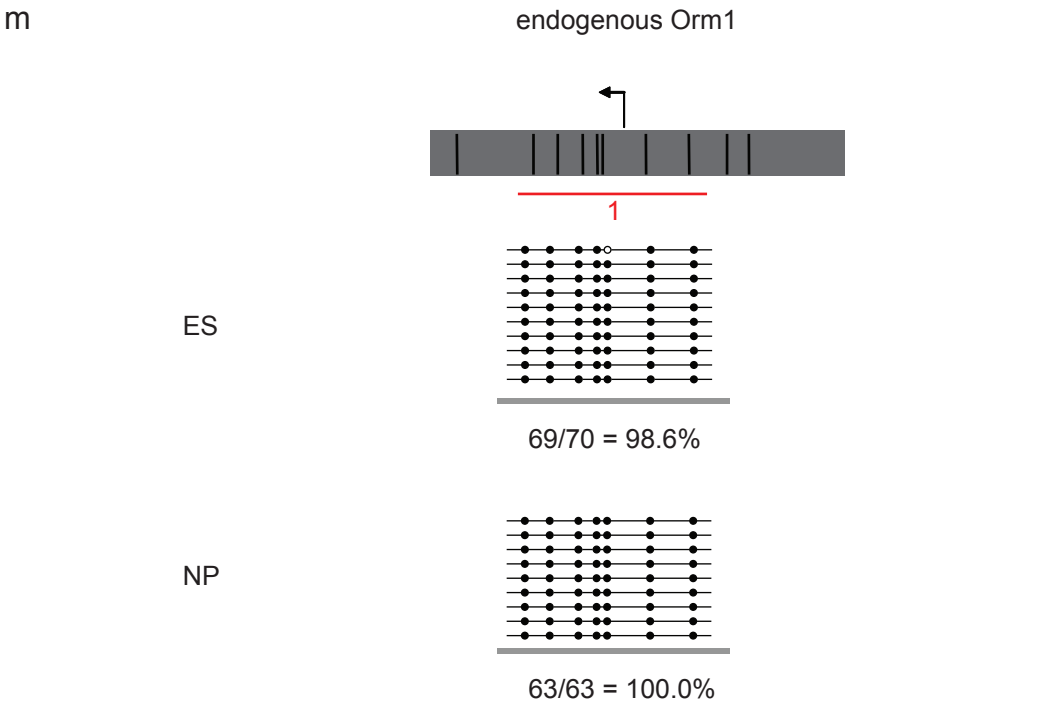
k

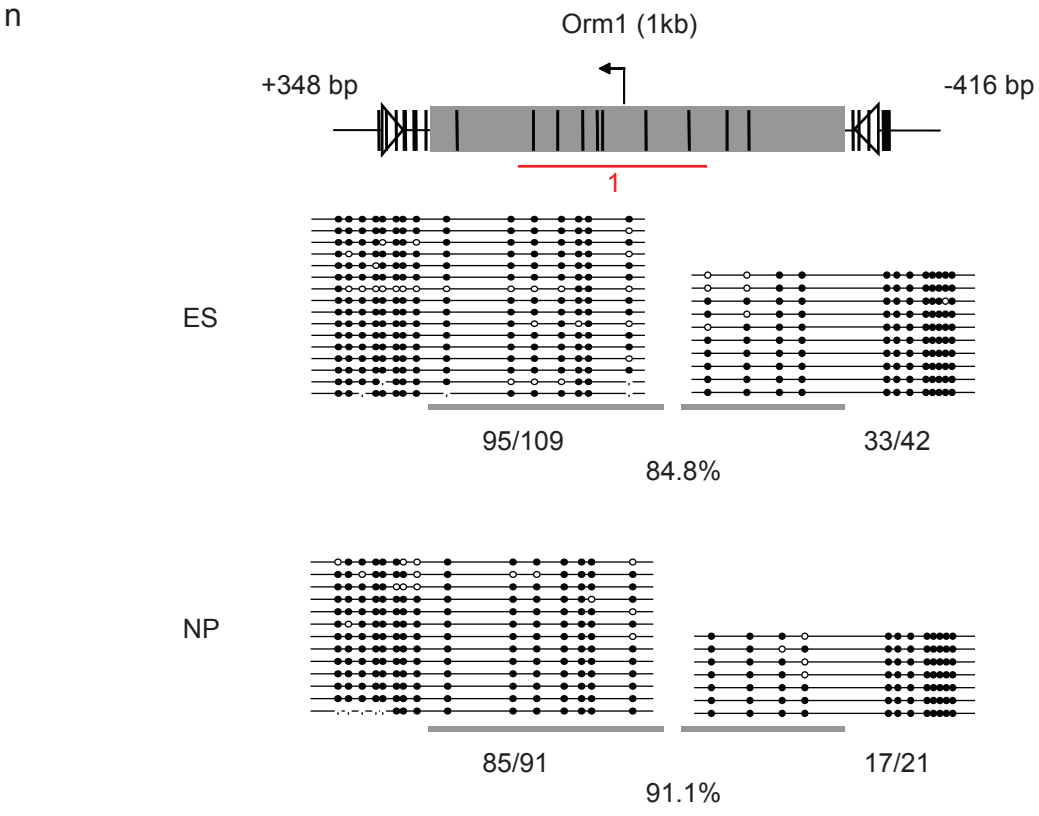


l

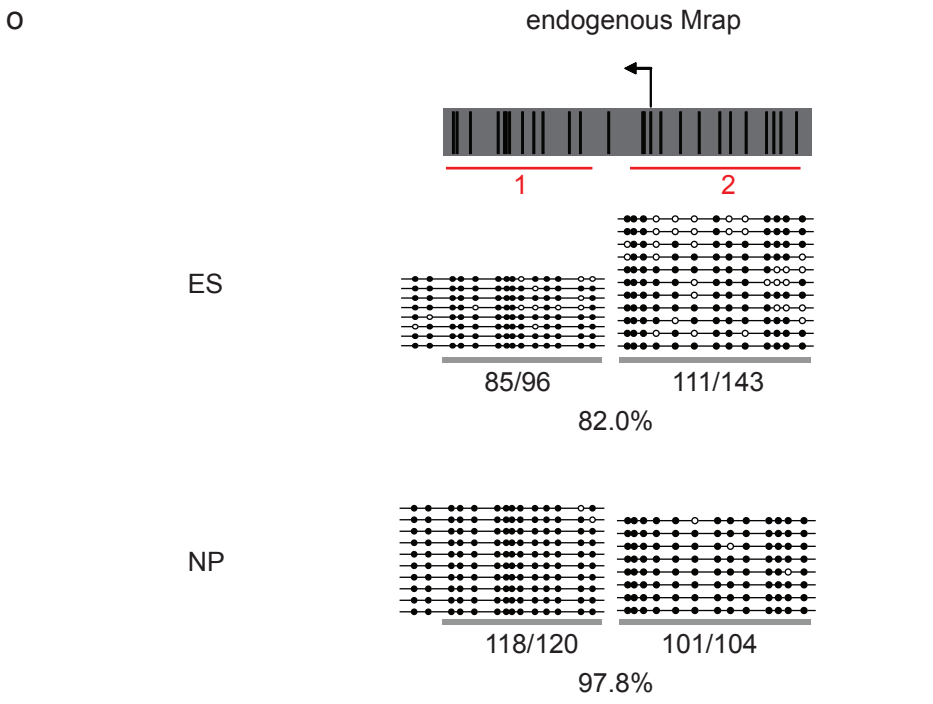


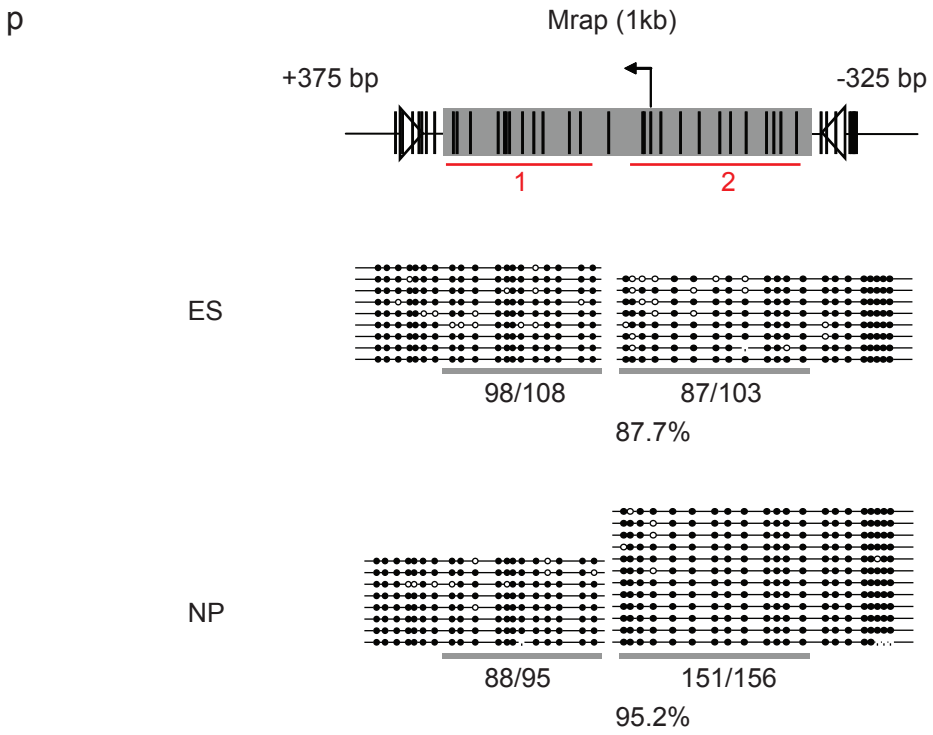
Supplementary Figure 2



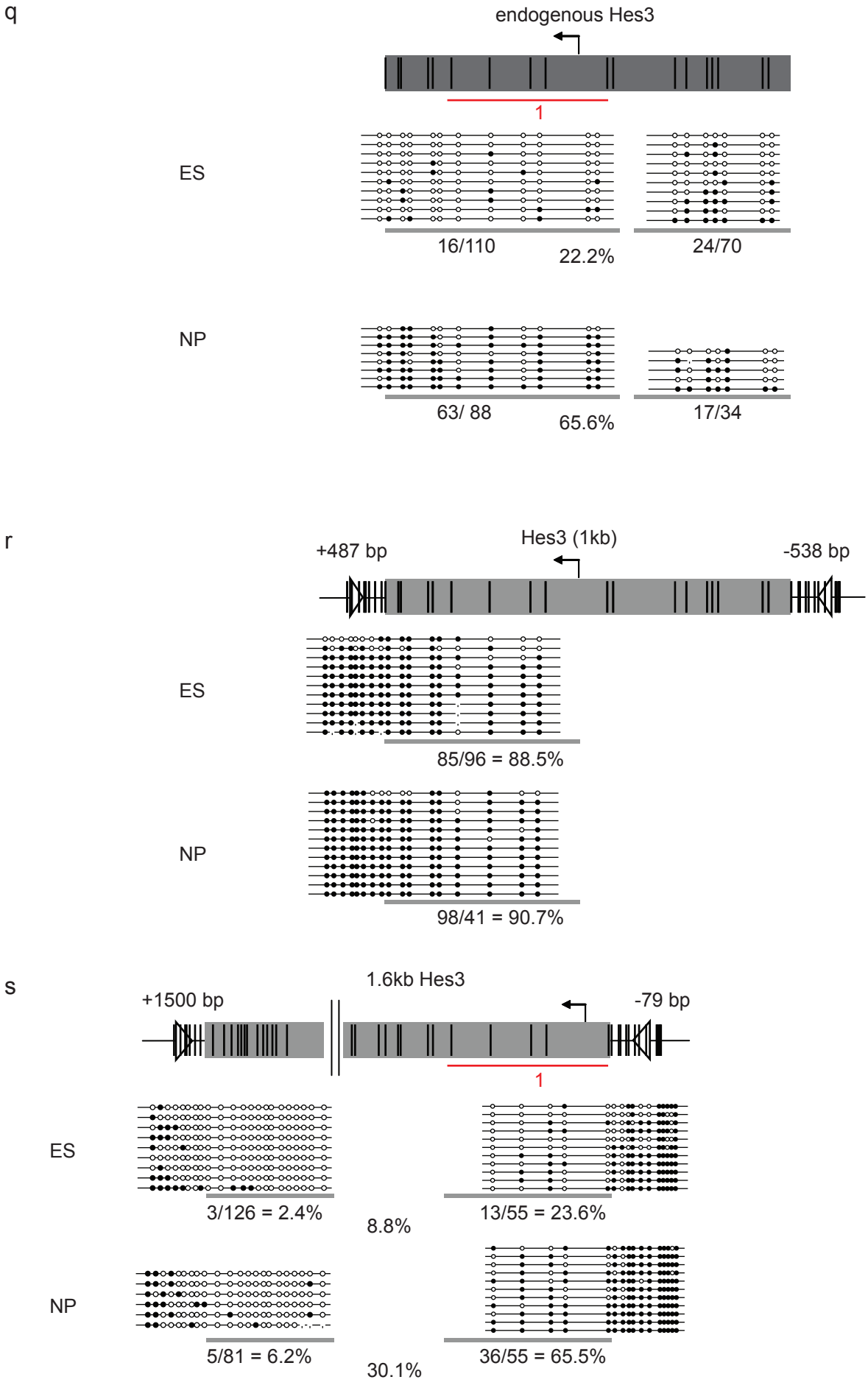


Supplementary Figure 2



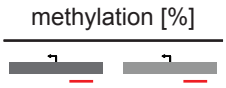


Supplementary Figure 2

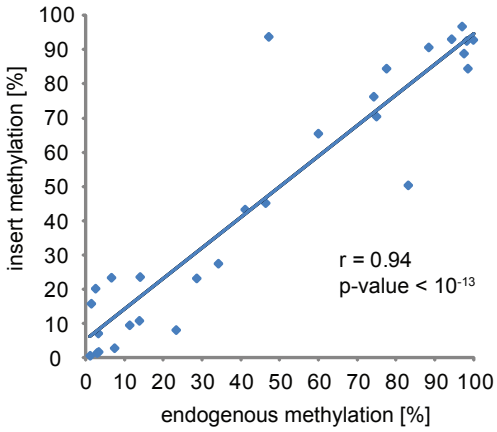


Supplementary Figure 2

t

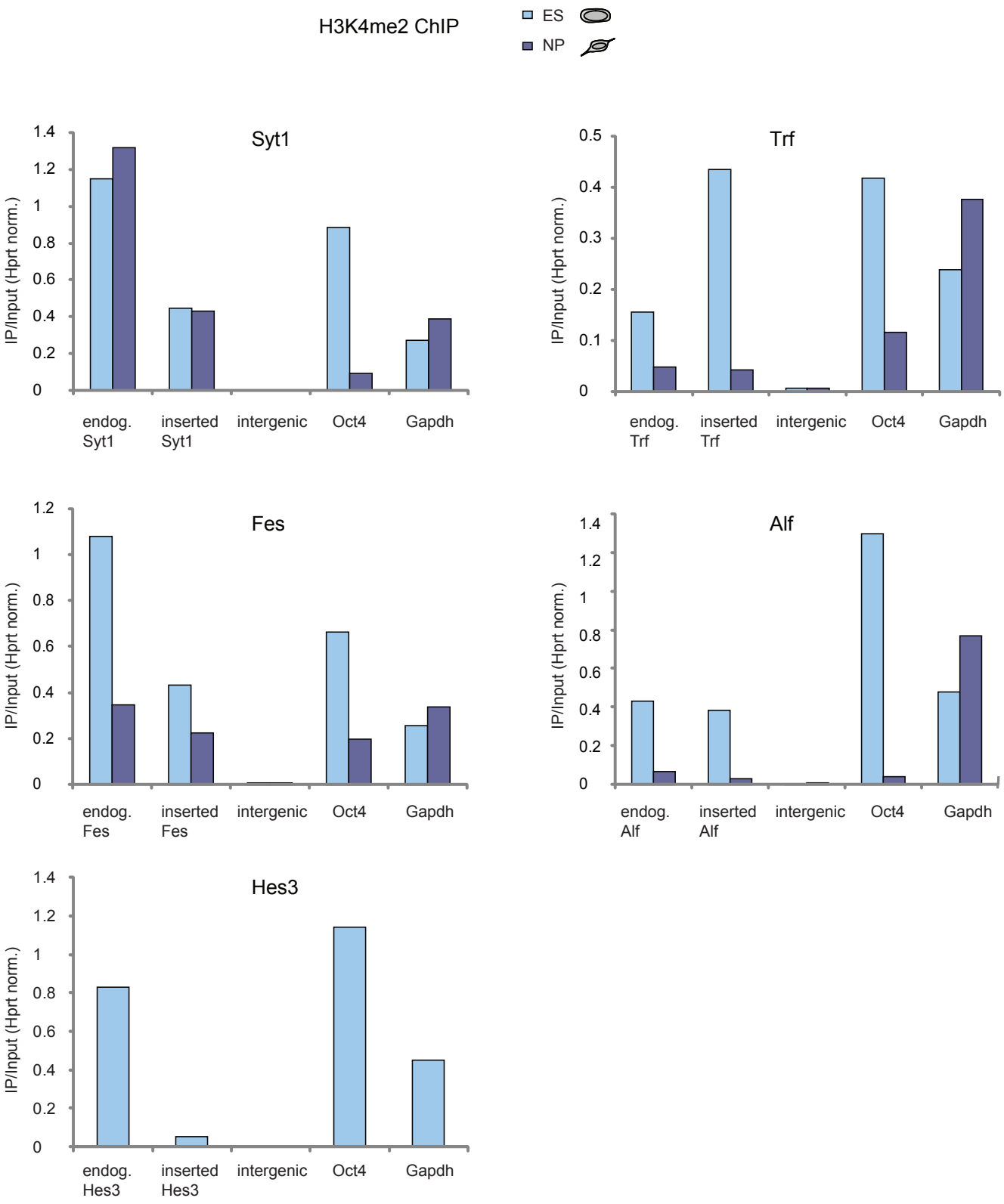


promoter	region	state	endog.	insert
Nanog	1	ES	7.4	2.8
Nanog	1	NP	41.1	43.3
Nanog	2	ES	2.5	20.2
Nanog	2	NP	47.2	93.8
Tcl1	1	ES	23.3	8.1
Tcl1	1	NP	97.6	88.9
Tcl1	2	ES	46.4	45.2
Trf	1	ES	13.8	10.8
Trf	2	ES	75.0	70.5
Trf	1	NP	74.3	76.3
Trf	2	NP	94.4	93.1
Alf	1	ES	11.3	9.5
Alf	1	NP	83.2	50.4
Fes	1	ES	28.6	23.2
Fes	1	NP	34.2	27.5
Syt1	1	ES	1.1	0.6
Syt1	1	NP	3.2	7.1
Syt1	2	ES	1.4	15.8
Syt1	2	NP	6.6	23.4
Zic3	1	ES	2.9	1.4
Zic3	1	NP	3.3	1.7
Orm1	1	ES	98.6	84.5
Orm1	1	NP	100.0	92.9
Mrap	1	ES	88.5	90.7
Mrap	1	NP	98.3	92.6
Mrap	2	ES	77.6	84.5
Mrap	2	NP	97.1	96.8
Hes3 (1.6kb)	1	ES	14.0	23.6
Hes3 (1.6kb)	1	NP	60.0	65.5



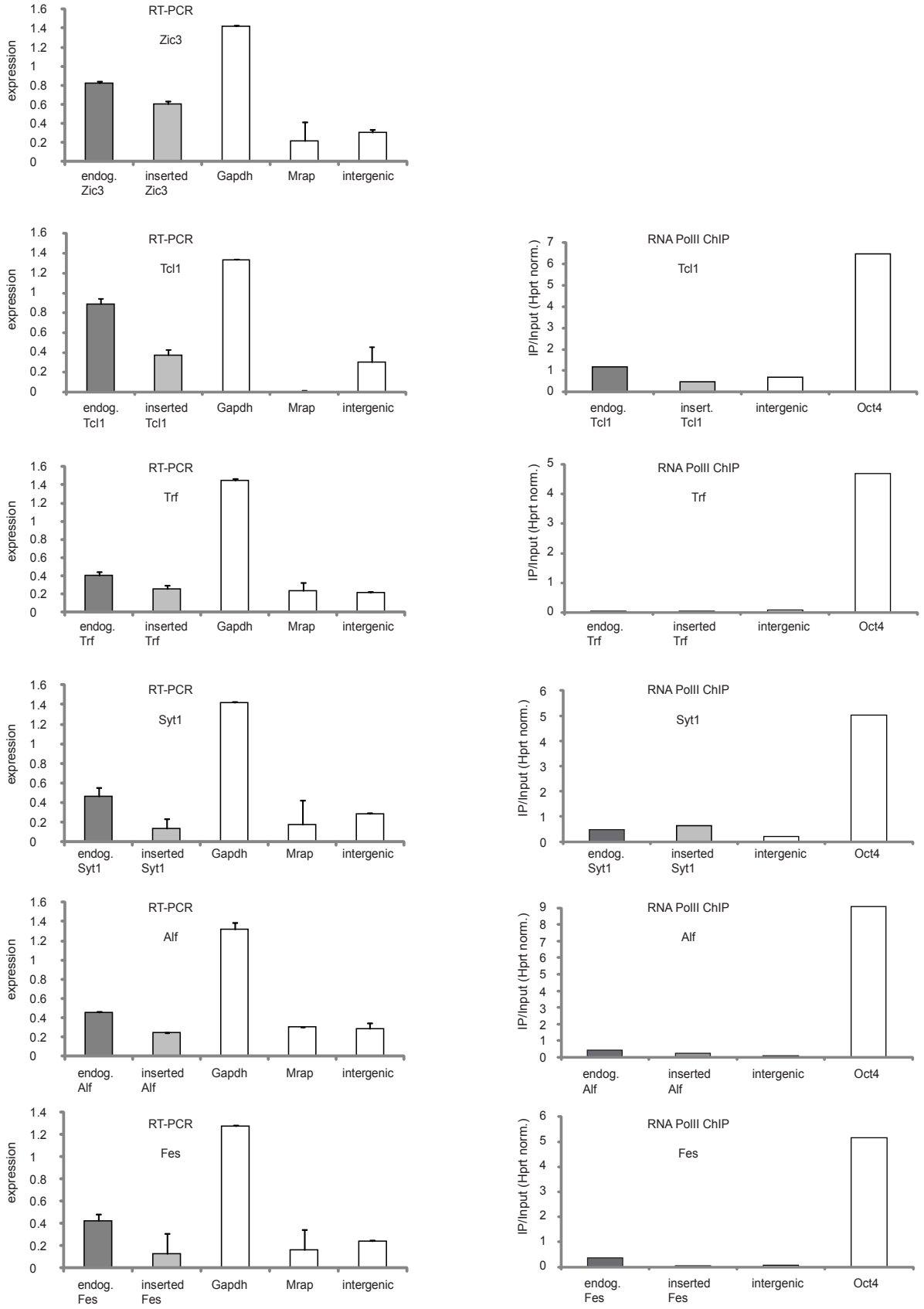
Supplementary Figure 2 Recapitulation of DNA methylation at ectopic promoter fragments. (**a-s**) DNA methylation levels are shown as in Supplementary Fig. 1. For most inserts methylation is also shown for the endogenous promoter as a comparison. Some of these has been published before (Mohn et al.,Mol Cell, 2008). (**t**) Comparison of methylation levels between inserted fragment and endogenous promoter as determined by bisulfite PCR illustrating the quantitative similarity in DNA methylation levels. Analyzed regions are indicated by a red bar and number in Supplementary Figs. 1 and 2 a-s.

Supplementary Figure 3



Supplementary Figure 3 H3K4me2 occupancy (in ES and NP cells) at inserted 1 kb promoter fragments. ChIP enrichments were normalized against Hprt.

Supplementary Figure 4

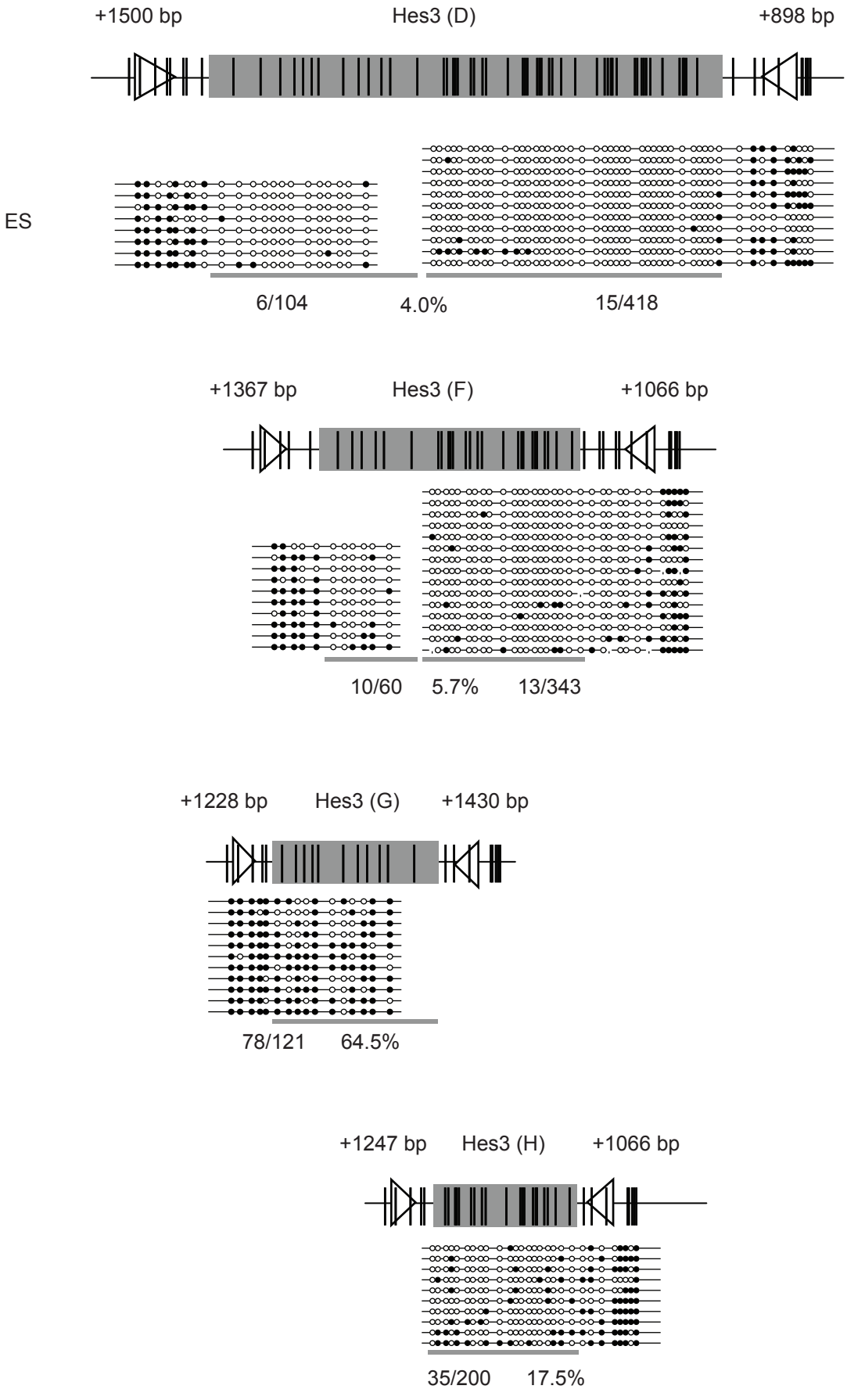


Supplementary Figure 4 RNA levels and RNA PolII occupancy in ES cells at inserted 1 kb promoter fragments. RNA levels were determined using reverse transcription followed by real-time PCR and normalized to LaminB. Values at a methylated promoter and at an intergenic region are shown as a comparison. Error bars indicate standard deviation from two independent biological replicates. ChIP enrichments for RNA PolII were normalized against Hprt.

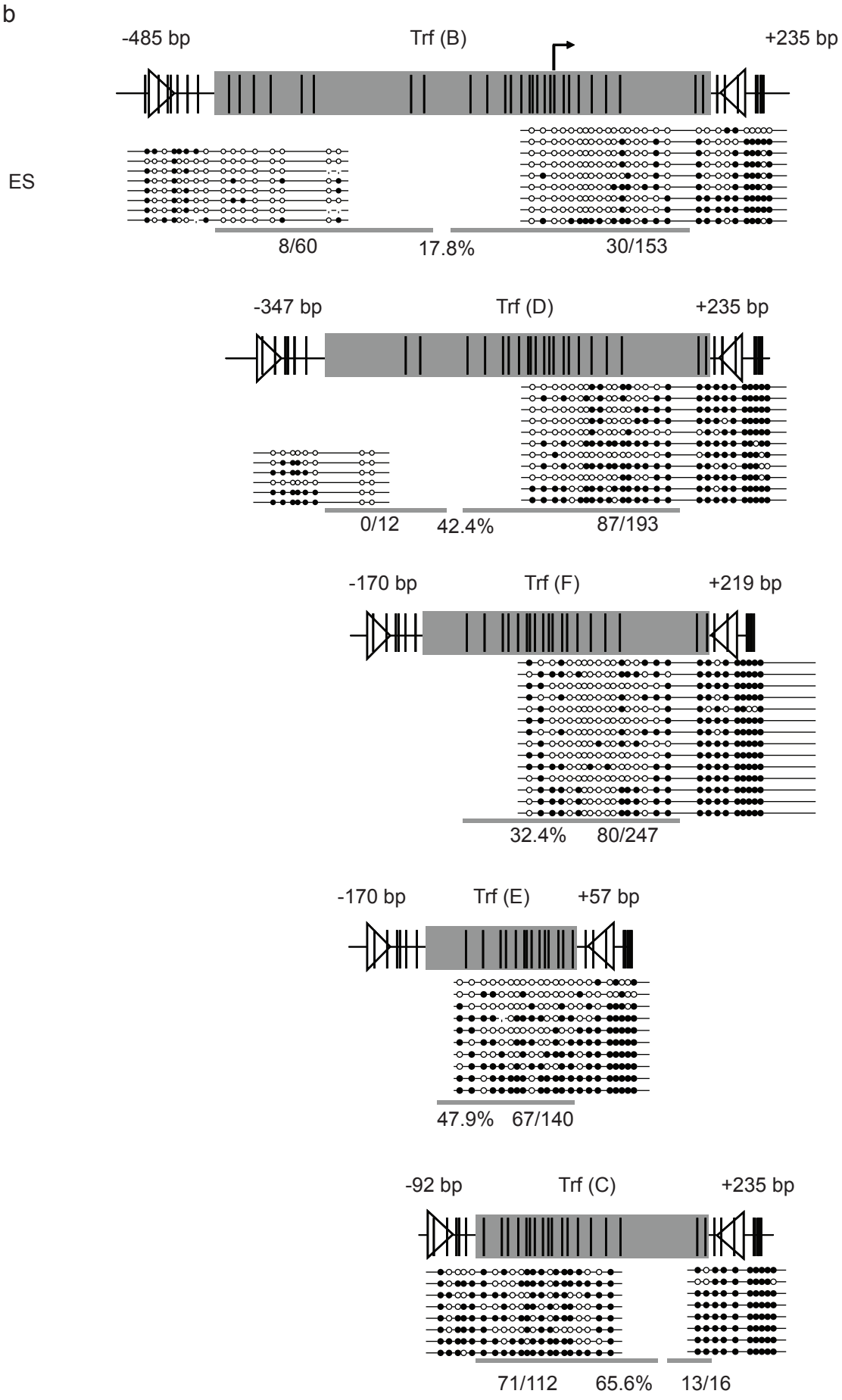
Nature Genetics: doi:10.1038/ng.946

Supplementary Figure 5

a

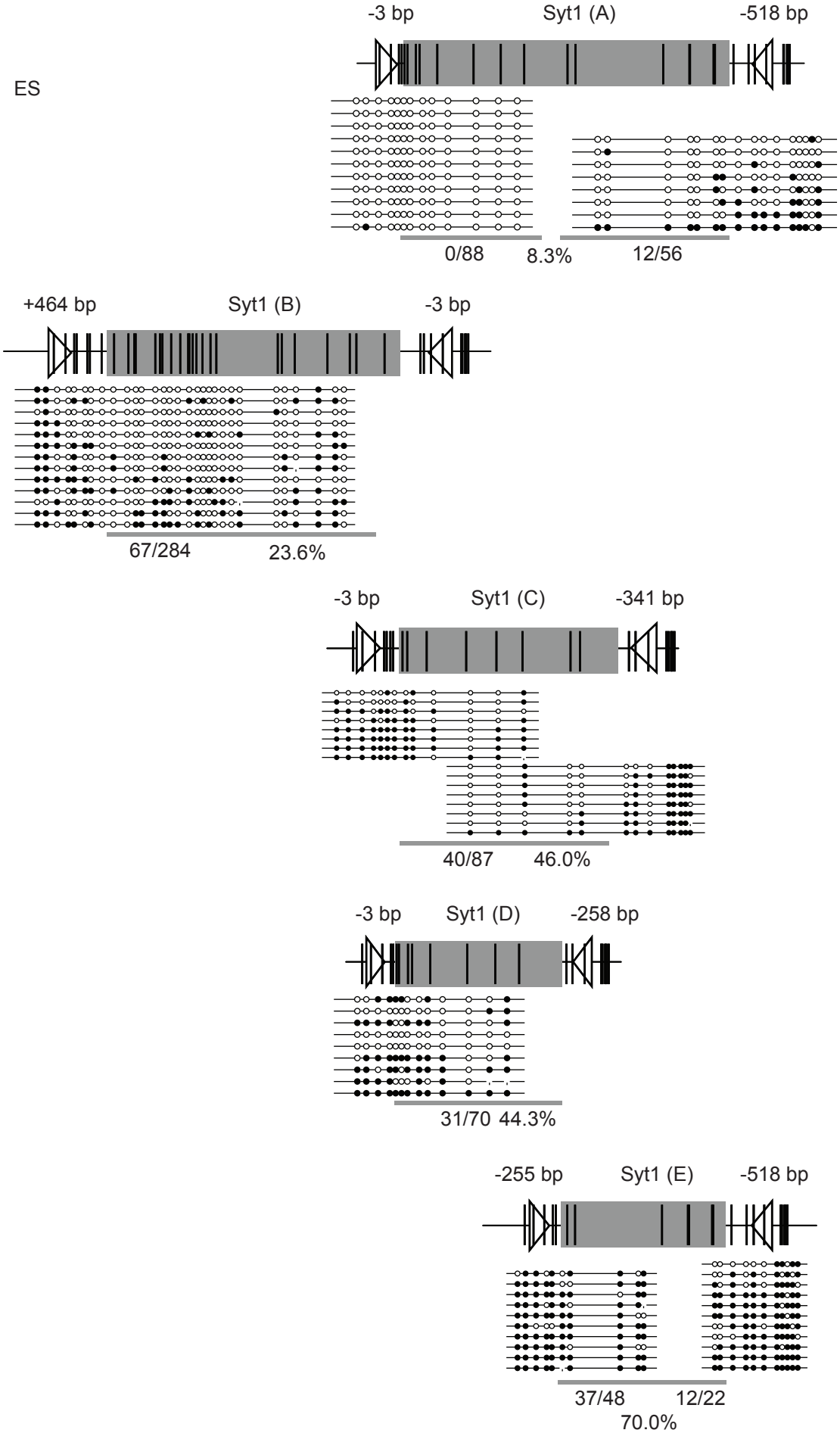


Supplementary Figure 5



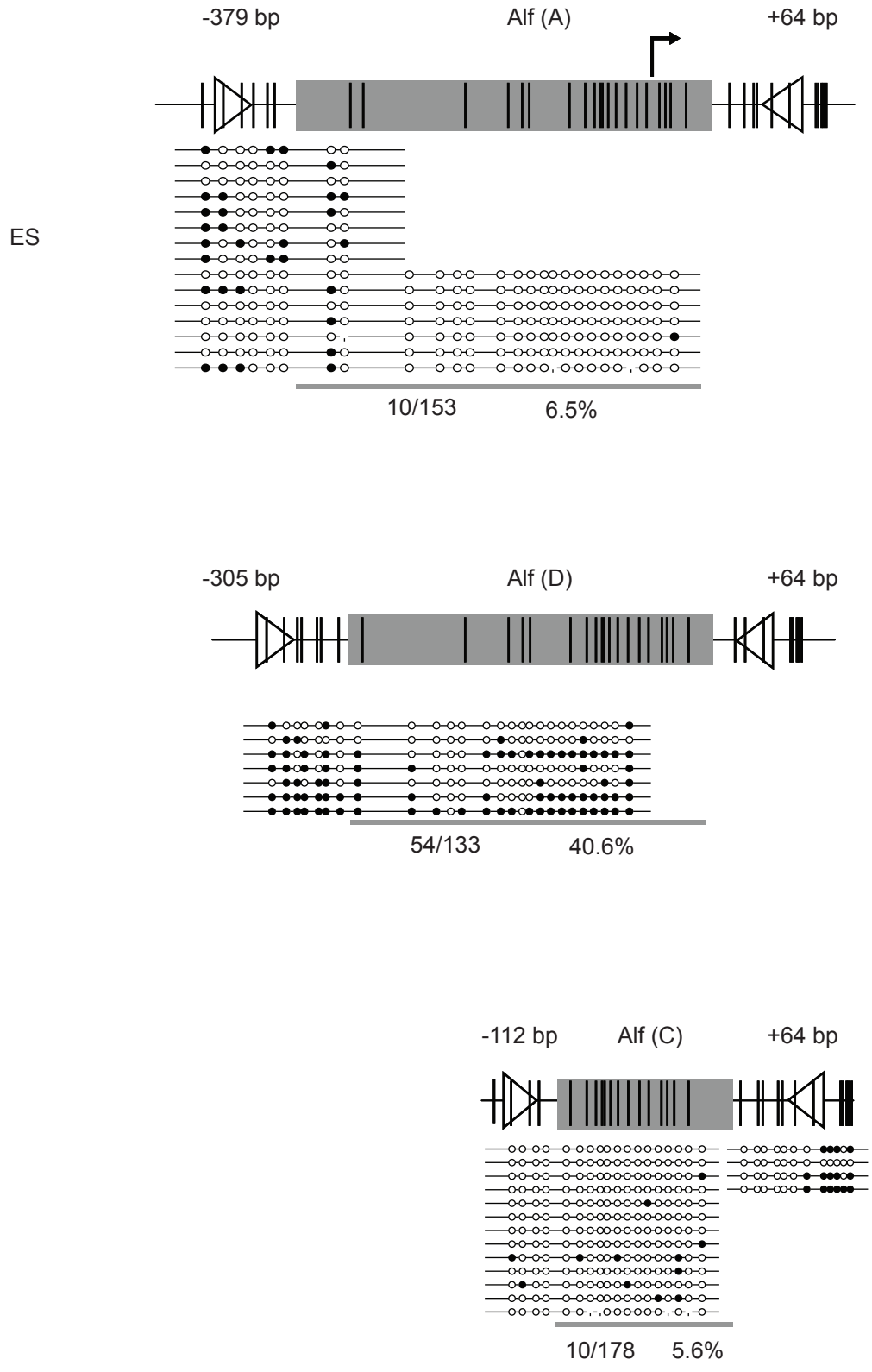
Supplementary Figure 5

c



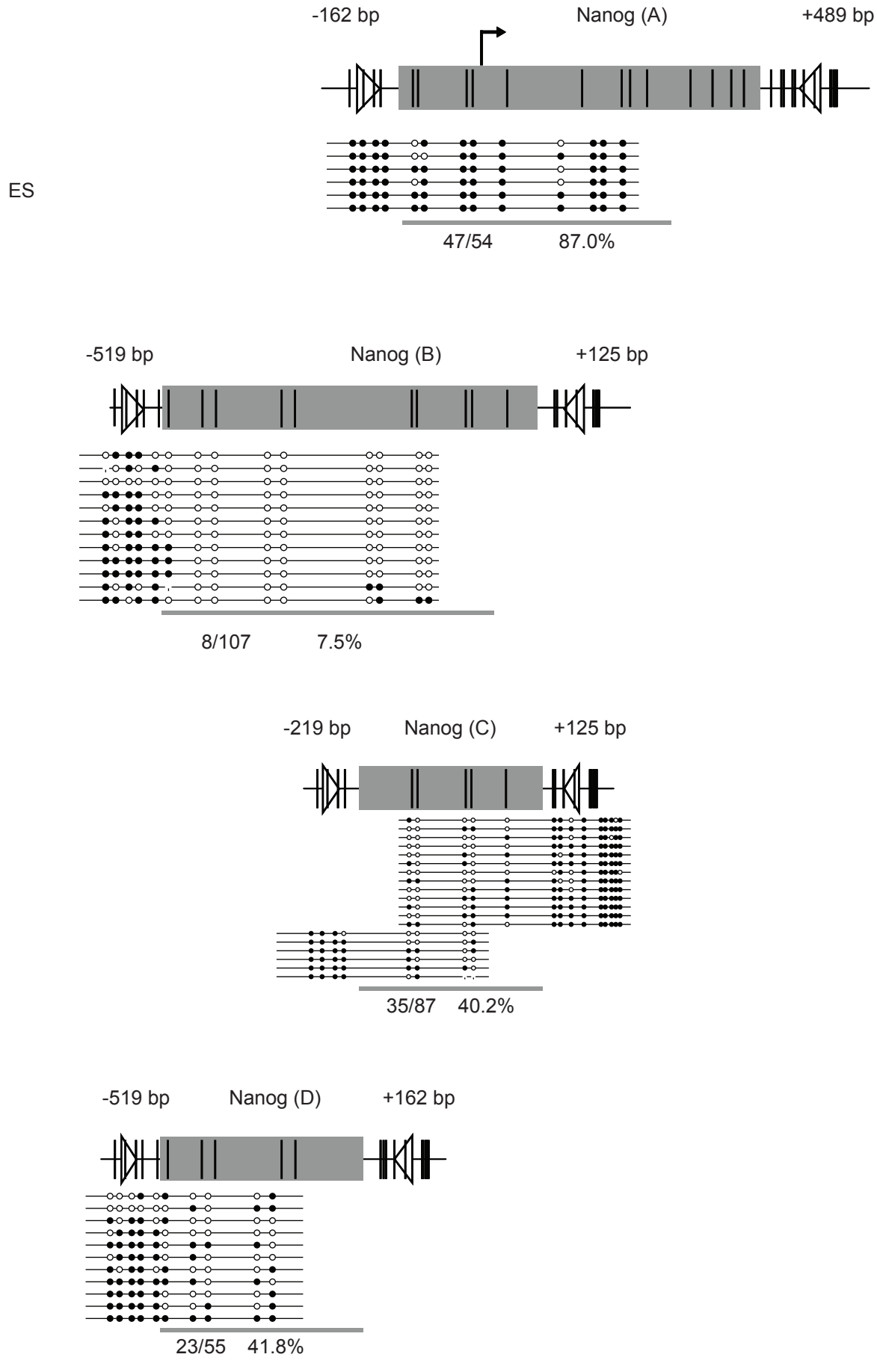
Supplementary Figure 5

d



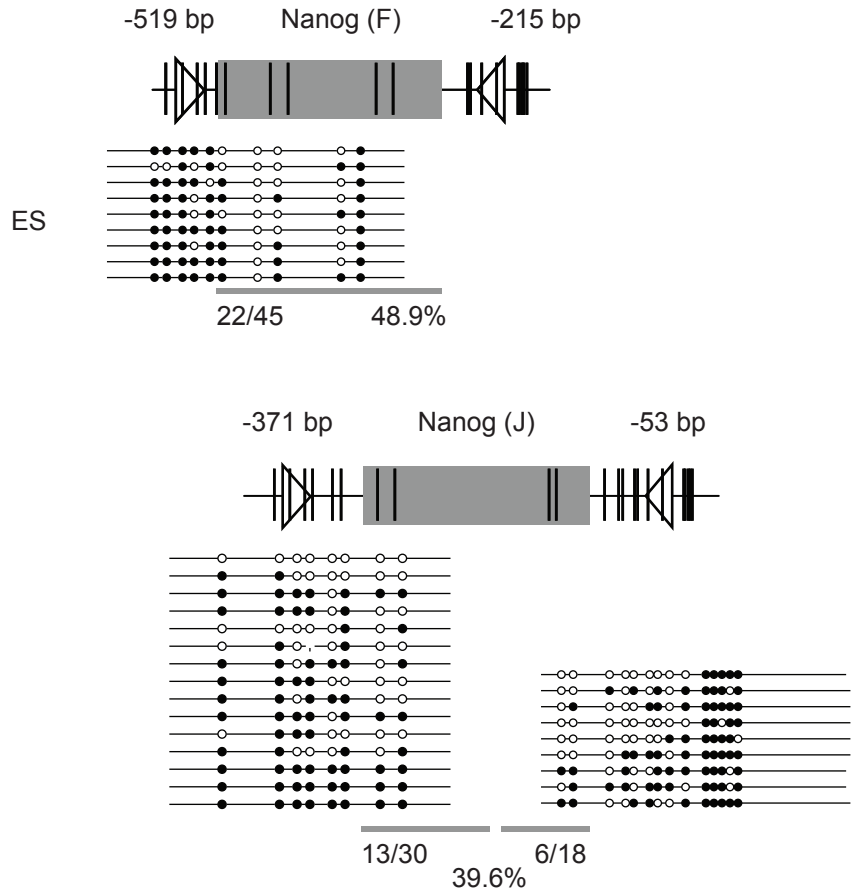
Supplementary Figure 5

e



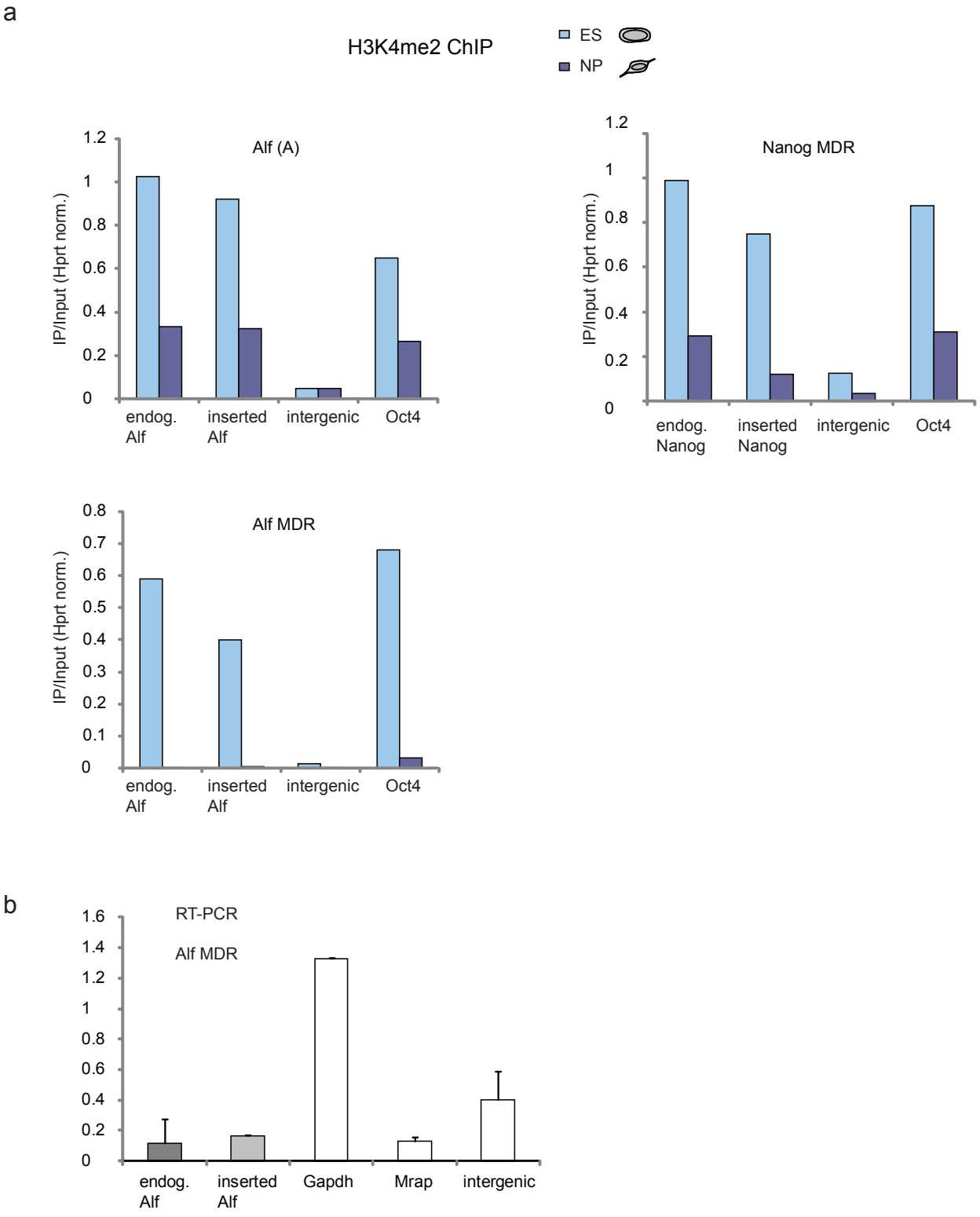
Supplementary Figure 5

e



Supplementary Figure 5 DNA methylation at inserted truncated promoter fragments. (**a-e**) DNA methylation levels are shown as in Supplementary Fig. 1.

Supplementary Figure 6



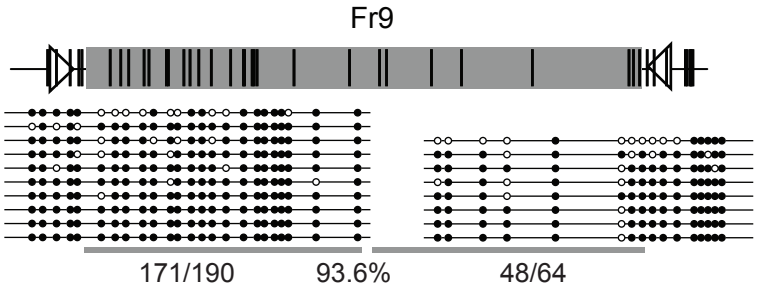
Supplementary Figure 6 H3K4 dimethylation and RNA levels at inserted truncated promoter fragments. (**a**) H3K4me2 occupancy (in ES and NP cells) at three inserted truncated promoter fragments. ChIP enrichments for H3K4me2 were normalized against Hprt. (**b**) RNA levels in ES cells at inserted Alf MDR. RNA levels were determined using reverse transcription followed by real-time PCR and normalized to LaminB. Values at a methylated promoter and at an intergenic region are shown as a comparison. Error bars indicate standard deviation from two independent biological replicates.

Supplementary Figure 7

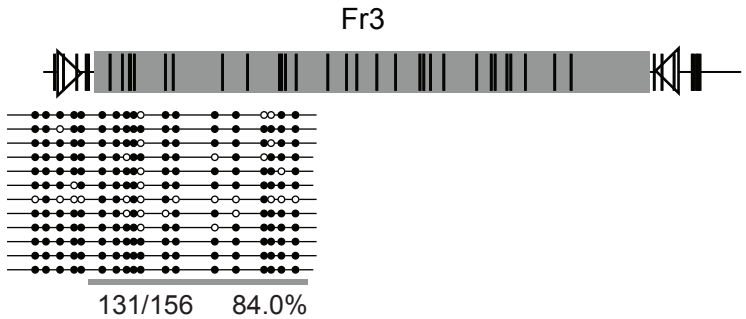
a

	Length [bp]	CpG frequency [%]	Primer s	Primer as
Fr9	780	4.4	CCT TTG ATG ACA GCC CAG TT	TCG TTT ACT CGT ATG GCG TTT
Fr3	798	4.5	AAG CCC TTC ATC AGC AAA GA	GGC CTA CAG TCC TGG TAC AAA
Fr6	762	4.5	AAC TGG GGA TTG TTC AGG TG	AAT GTA TCA GTT CGC TTG CTC A
Fr5	782	5.3	AAA AGG TCG GGG AAC TGA TT	TAT TGA AAC AGC GAC GAT GC
Fr2	780	5.4	AAC GTG CCT ACC GTA AGC TG	CCT TTG CGA CCA ACT CAT C
Fr1	832	6.5	GAT GGG CGA GGT TGT CTA AC	GCA GGG TTA ACG CTT TCT GA
Fr4	781	6.5	CTA CGG TGC TGA AGC AAC AA	GAT GTC CGA CCA GGA TTA GC
Fr12	761	6.6	GGT AAA CCA GTG AGC CGA AA	CTG CCA GAC CTT CAG GGA TA
Fr10	761	6.7	TAC CGG TGA ACT GGT GGG TA	TTC TGG TGT GAT CCC GTT CT
Fr11	796	6.8	GCC AGC TGG TAT CTT CGA CT	AGA GGG AAA CAA CCC AGA CC

b

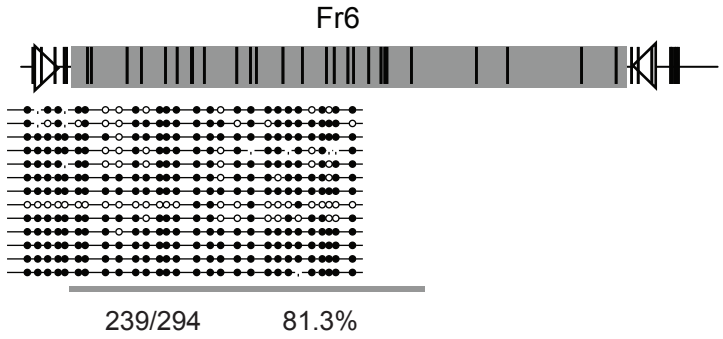


c

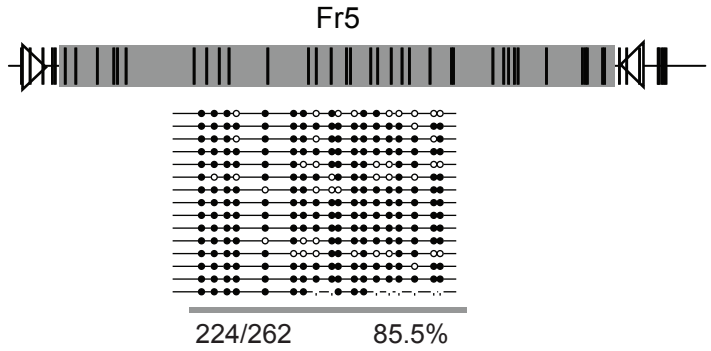


Supplementary Figure 7

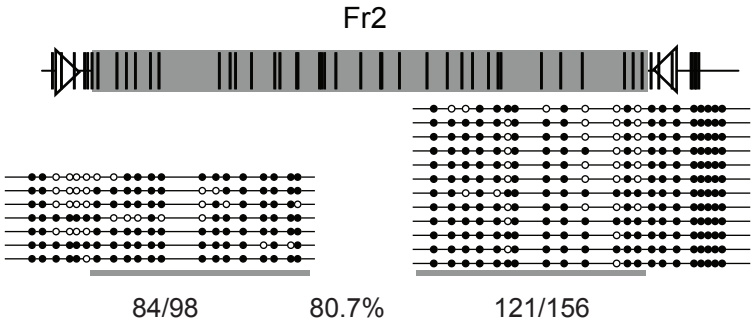
d



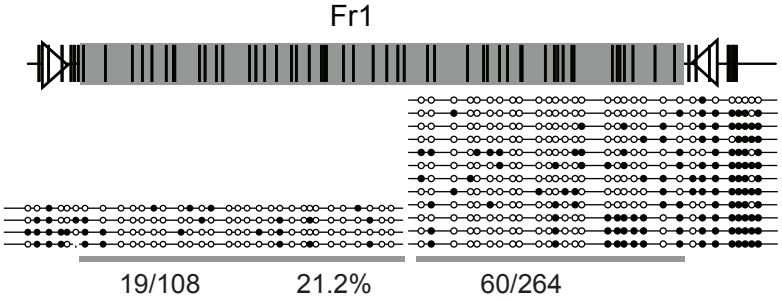
e



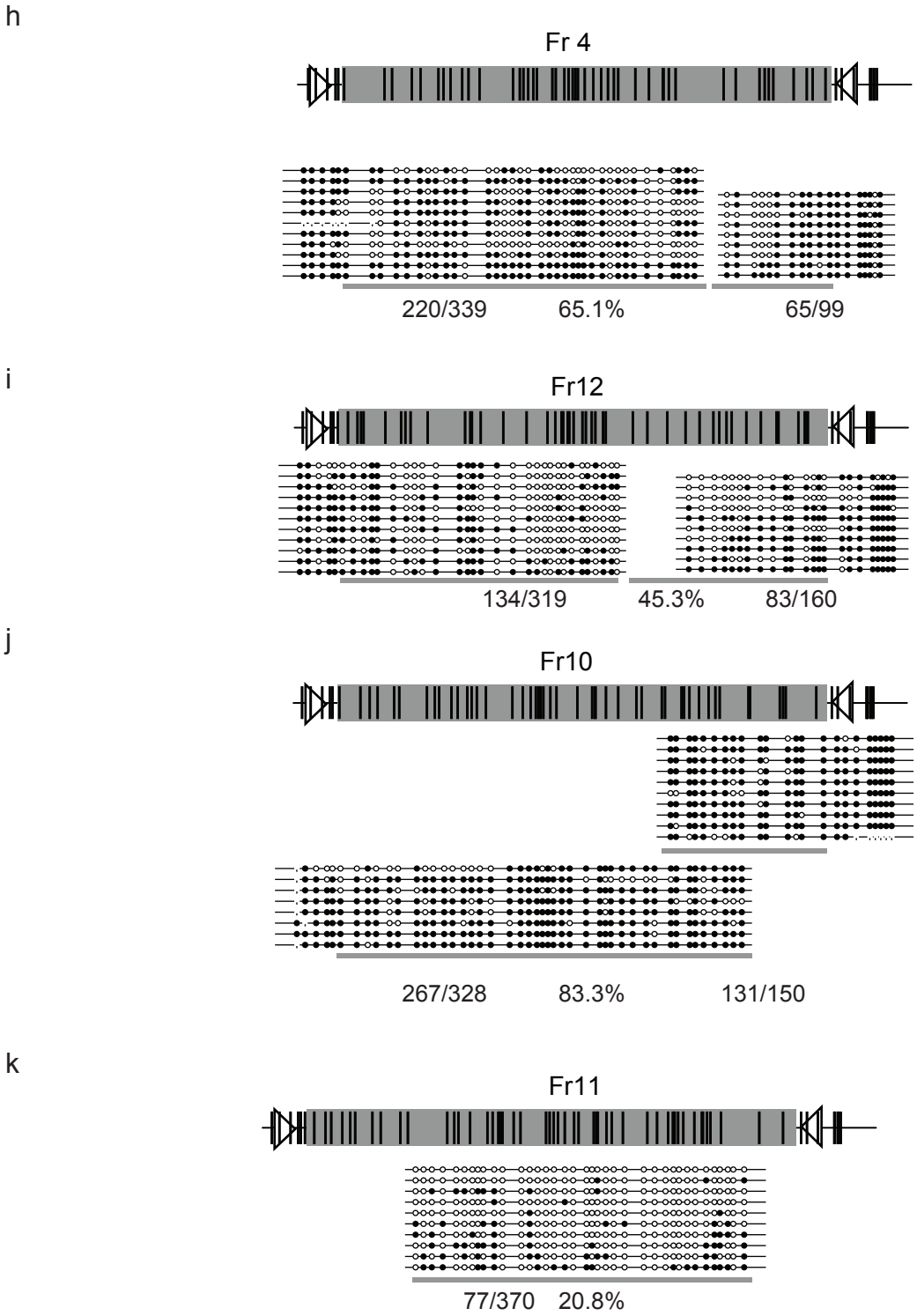
f



g

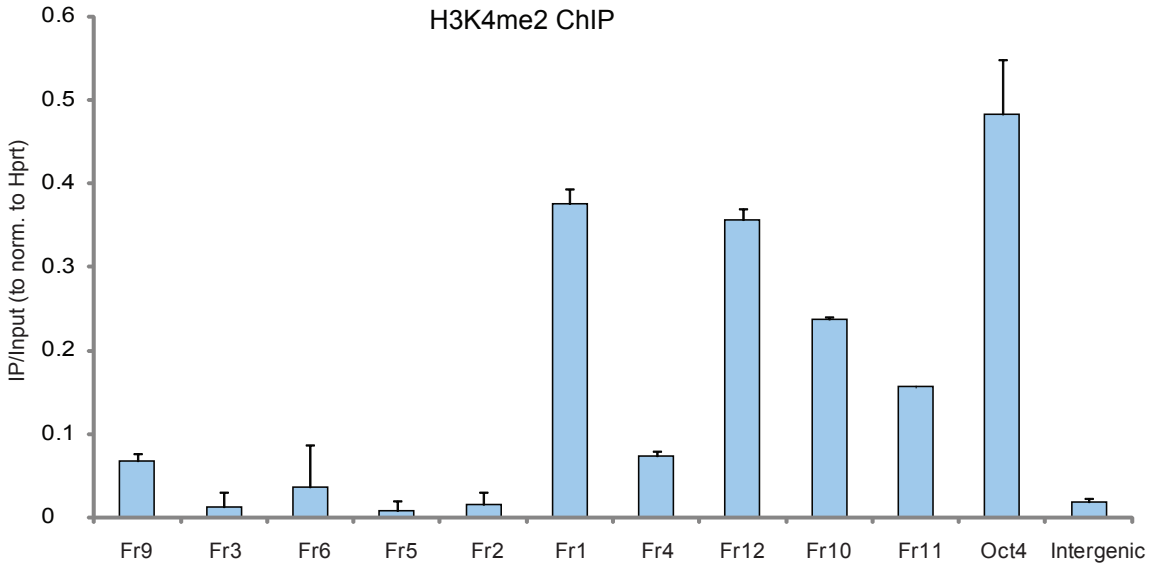


Supplementary Figure 7



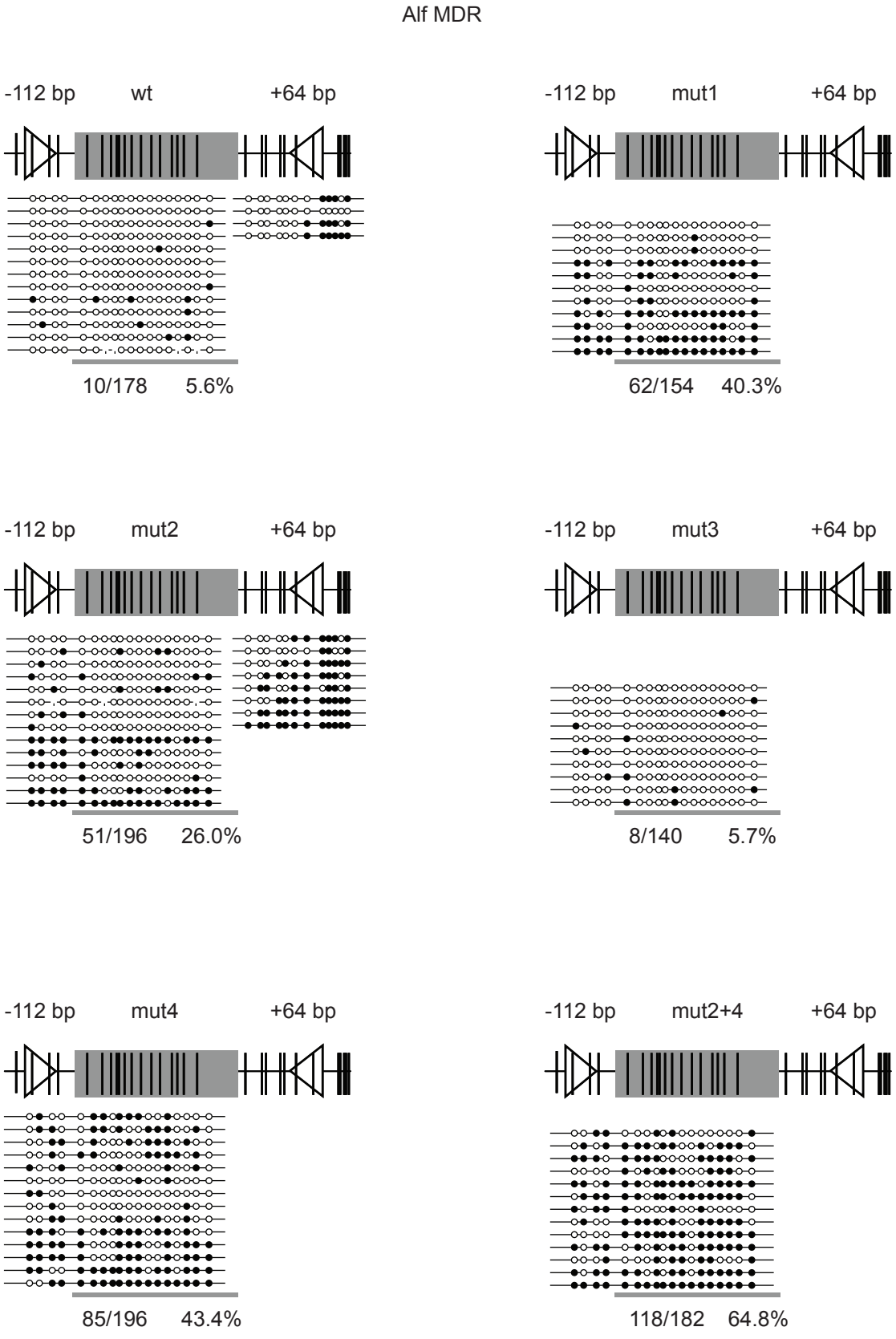
Supplementary Figure 7 DNA methylation at inserted sequences from E.coli. (**a**) Table showing length and CpG density of each fragment. The last two columns indicate primer sequences that were used for amplification of fragments from E.coli genomic DNA (TOP10, Invitrogen). (**b-k**) DNA methylation levels of inserted E.coli sequences in stem cells are shown as in Supplementary Fig. 1.

Supplementary Figure 8



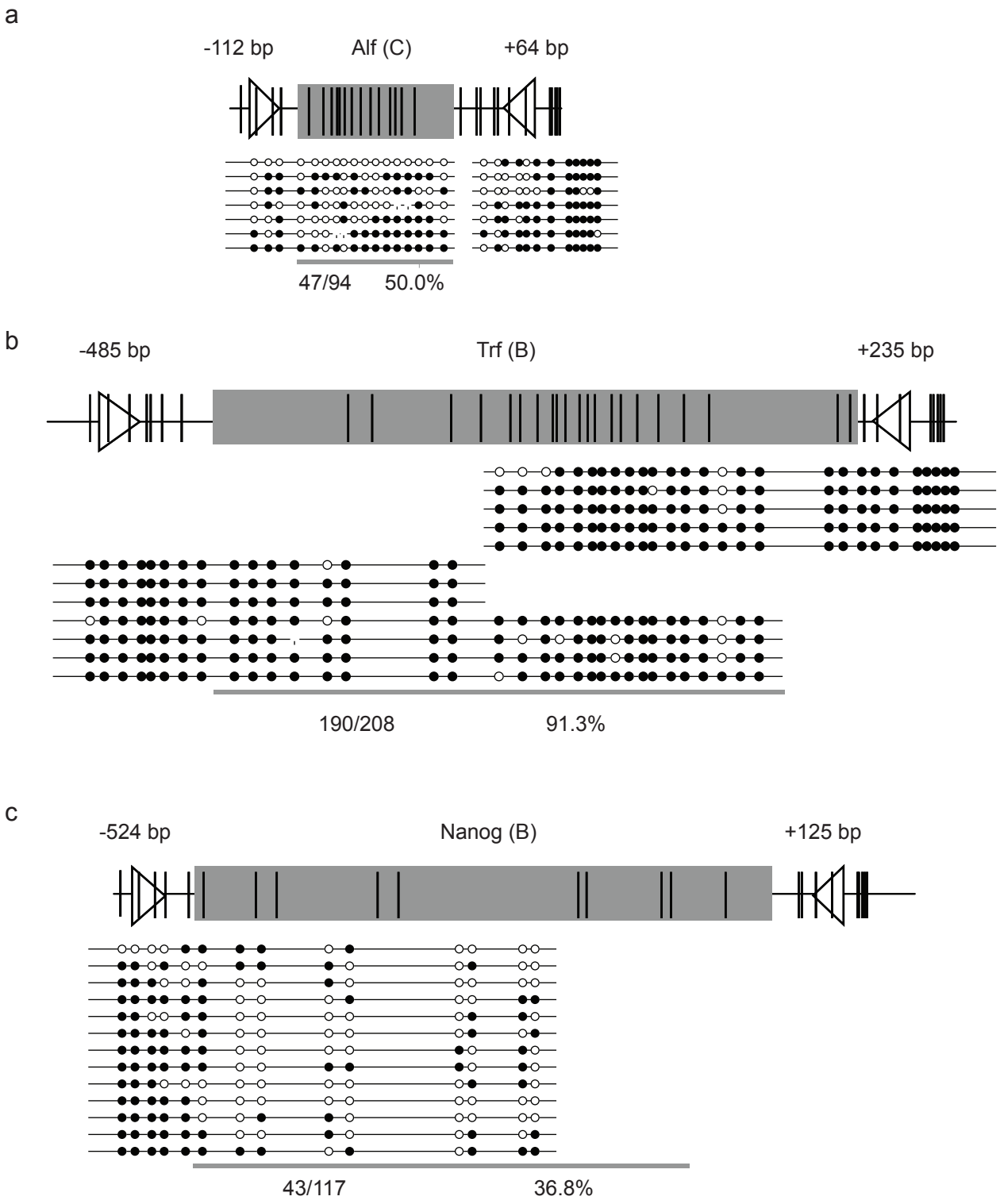
Supplementary Figure 8 H3K4me2 occupancy in ES cells at inserted E.coli sequences. ChIP enrichments for H3K4me2 were normalized against Hprt. The three fragments that possess low DNA methylation show enrichment for H3K4me2. Interestingly, Fr10 is also occupied by H3K4me2 although being methylated to 83.3%. At this fragment, H3K4me2 might be specifically enriched at unmethylated alleles in analogy to what has been shown for a CpG-rich transgene (Thomson J.P., Nature,2010).

Supplementary Figure 9



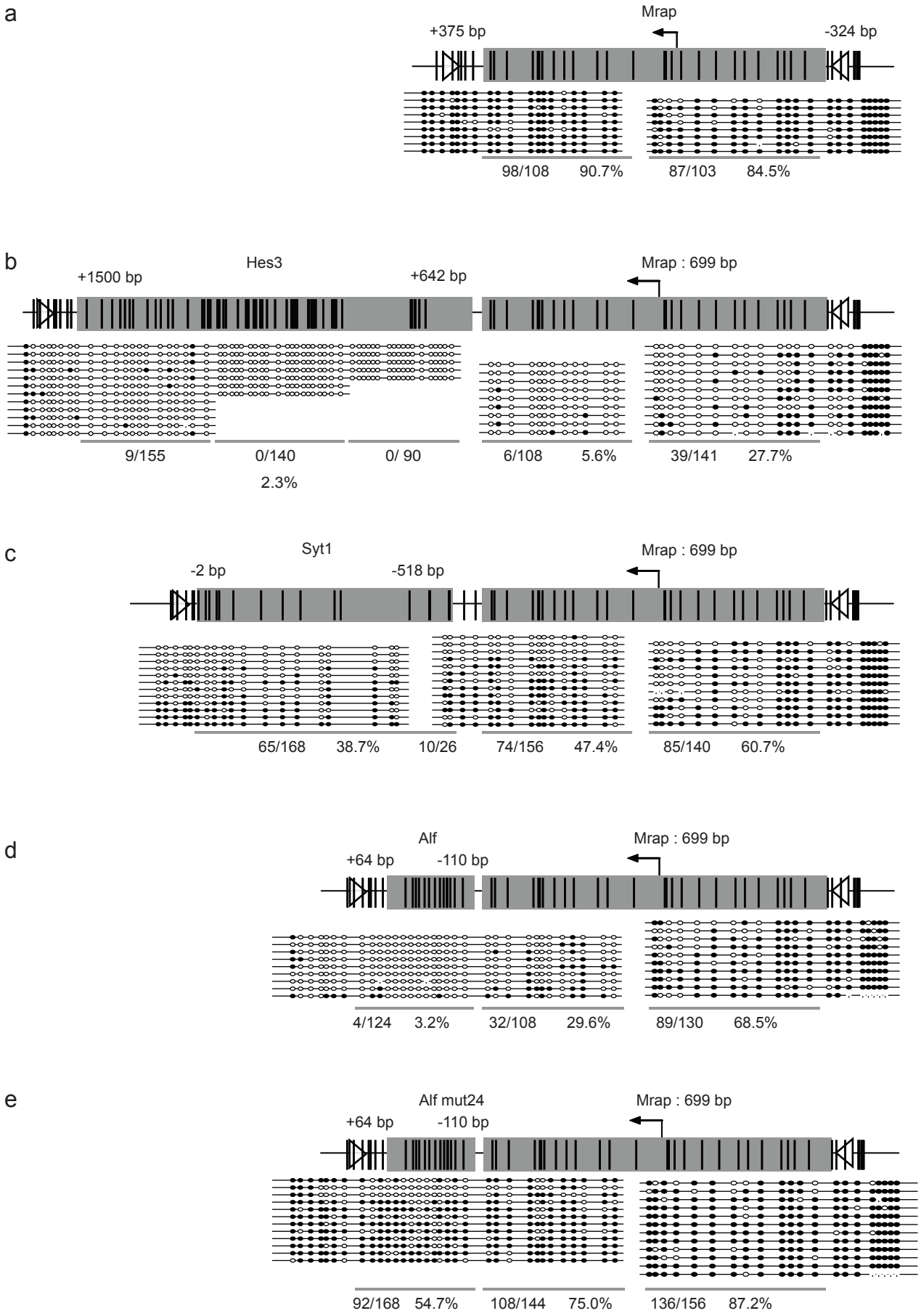
Supplementary Figure 9 DNA methylation levels at inserted Alf MDR containing different mutations.
DNA methylation levels in stem cells are shown as in Supplementary Fig. 1. Sequence mutations are shown in Fig. 4.

Supplementary Figure 10

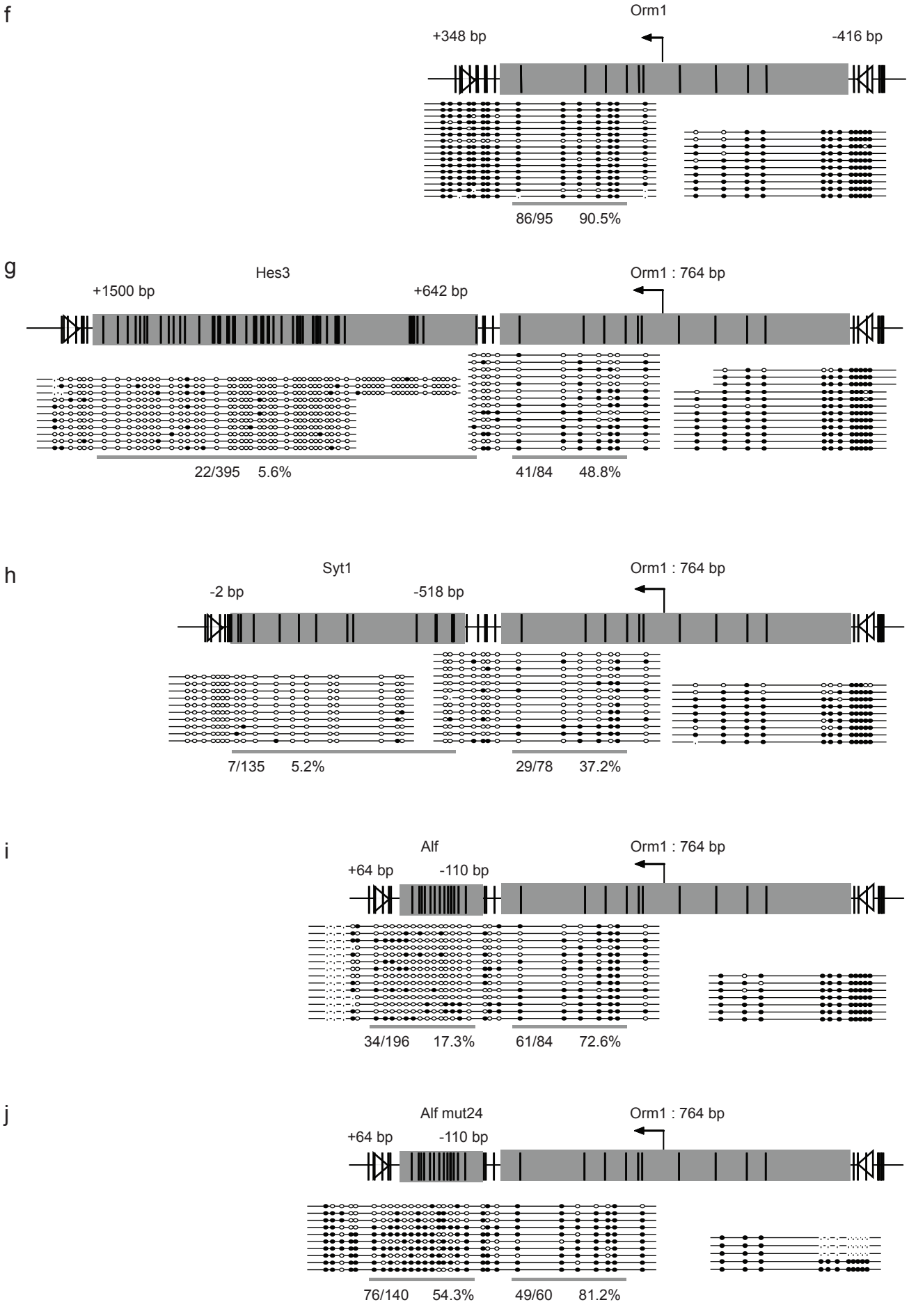


Supplementary Figure 10 DNA methylation levels at inserted MDRs in neuronal progenitors. DNA methylation levels in neuronal progenitors are shown as in Supplementary Fig. 1.

Supplementary Figure 11

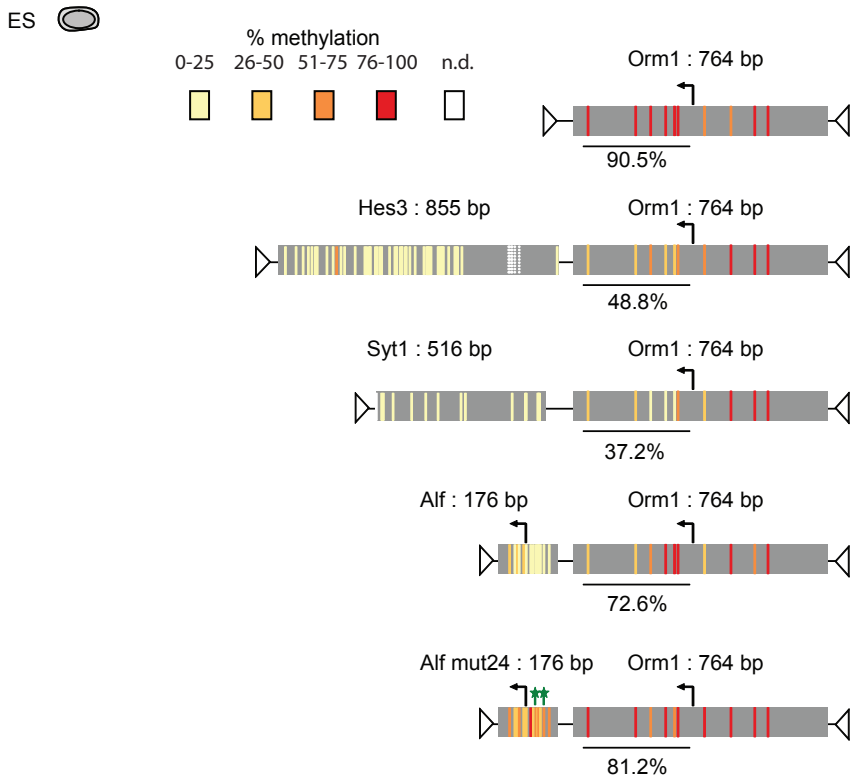


Supplementary Figure 11



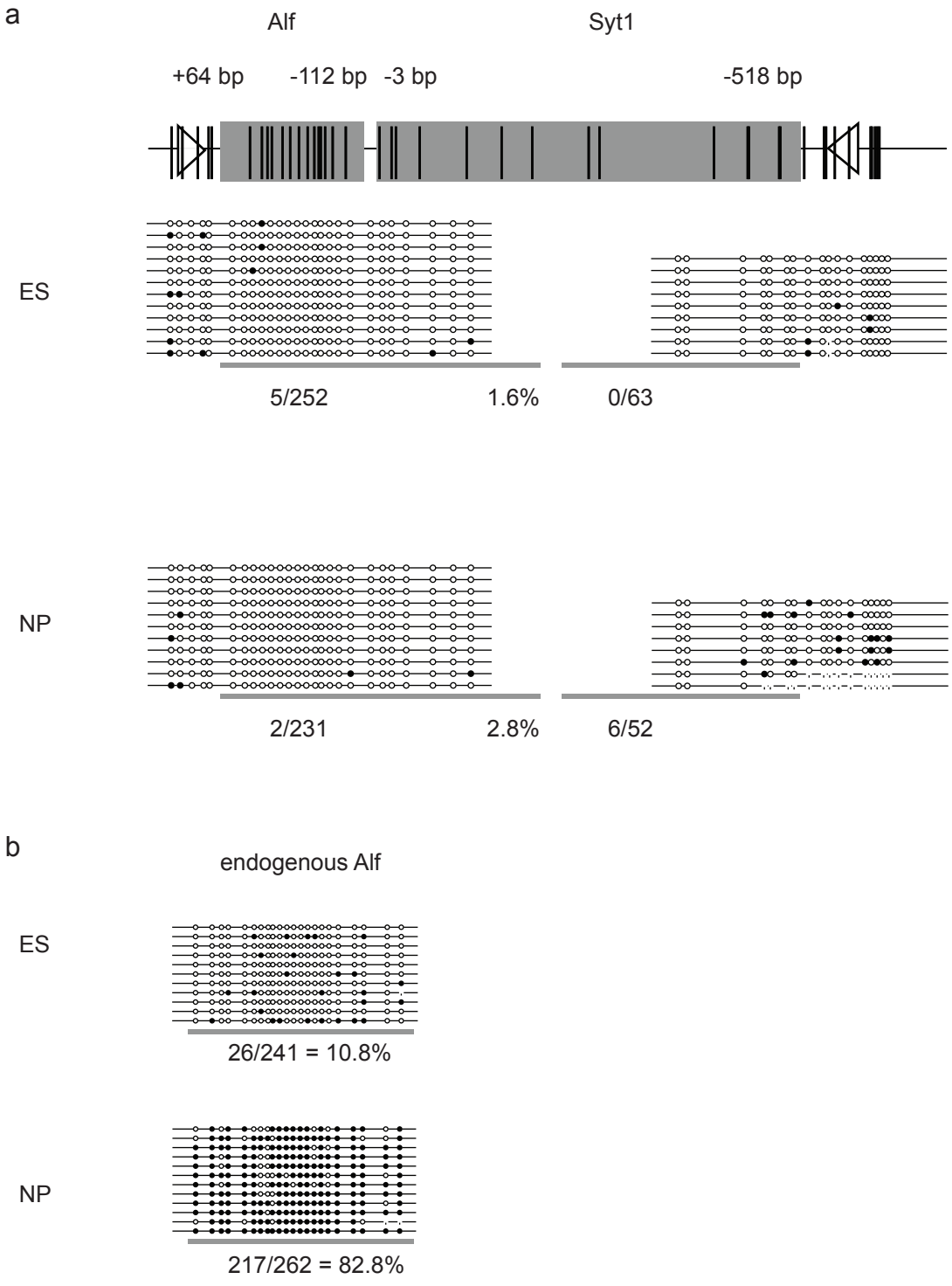
Supplementary Figure 11

k



Supplementary Figure 11 DNA methylation levels at inserted hybrid fragments.
(**a-j**) DNA methylation levels in stem cells are shown as in Supplementary Fig. 1. (**k**) DNA methylation levels of Orm1 hybrids are shown as in Fig. 5b. (n.d. = not determined)

Supplementary Figure 12



Supplementary Figure 12 Syt1 MDR overrides *de novo* methylation of Alf MDR. (**a**) DNA methylation levels at inserted hybrid of the Alf and Syt1 MDR in embryonic stem cells (ES) and neuronal progenitors (NP) are shown as in Supplementary Fig. 1. (**b**) DNA methylation at the endogenous Alf promoter in the same clone is shown as a control.

Supplementary Table 1

Primer sequences

The first three subtables indicate the combinations of primers used for reverse transcription RT-PCR, ChIP RT-PCR and bisulfite PCR.
The last subtable indicates the primer sequences.

Reverse Transcription RT-PCR

promoter	location	fragment	s-primer #	as-primer #
Nanog	endog.		Nanog outs.1	Nanog RT1
Nanog	insert.	1kb	44	Nanog RT1
Zic3	endog.		520	521
Zic3	insert.	1kb	44	6
Tcl1	endog.		101	85
Tcl1	insert.	1kb	30	85
Trf	endog.		51	65
Trf	insert.	1kb	44	51
Syt1	endog.		73	45
Syt1	insert.	1kb	30	45
Alf	endog.		10	56
Alf	insert.	1kb	44	10
Alf	insert.	MDR	44	386
Fes	endog.		49	67
Fes	insert.	1kb	30	49
LaminB	endog.		518	519
Gapdh	endog.		516	517
Mrap	endog.		522	523
intergenic	endog.		interg3 s	interg3 as

ChIP RT-PCR

promoter	location	fragment	s-primer #	as-primer #
Nanog	endog.		Nanog outs.1	Nanog RT1
Nanog	insert.	1kb	30	Nanog RT1
Nanog	insert.	MDR	30	fm60
Tcl1	endog.		101	85
Tcl1	insert.	1kb	30	85
Trf	endog.		51	65
Trf	insert.	1kb	44	51
Syt1	endog.		73	45
Syt1	insert.	1kb	44	46
Alf	endog.		56	10
Alf	endog.		mGTFA1L as2	mGTFA1L s2
Alf	insert.	1kb	44	10
Alf	insert.	MDR	44	386
Alf	insert.	Alf (A)	44	171
Fes	endog.		49	67
Fes	insert.	1kb	44	50
Hes3	endog.		53	27
Hes3	insert.	1kb	30	27
Pdcl2	endog.		502	503
Oct4	endog.		Oct s	Oct as
E.coli Fr9	insert.		30	354
E.coli Fr3	insert.		44	389
E.coli Fr6	insert.		30	351
E.coli Fr5	insert.		44	391
E.coli Fr2	insert.		30	347
E.coli Fr1	insert.		44	387
E.coli Fr4	insert.		30	349
E.coli Fr12	insert.		30	415
E.coli Fr10	insert.		30	411
E.coli Fr11	insert.		30	413

Bisulfite PCR

promoter	location	fragment	s-primer #	as-primer #
Nanog	endog.		327	317
Nanog	endog.		255	Nanog bis. 1
Nanog	endog.		437	302
Nanog	insert.	5.7 kb	Mb 1	317
Nanog	insert.	5.7 kb	277	303
Nanog	insert.	1 kb	Mb 1	Nanog bis. 1
Nanog	insert.	1 kb	277	302
Nanog	insert.	1 kb	277	303
Tcl1	insert.	1kb	Mb 1	Tcl1 bis. A
Tcl1	insert.	1kb	277	137
Trf	endog.		107	42
Trf	endog.		108	43
Trf	insert.	1kb	Mb 1	42

Trf	insert.	1kb	277	43
Alf	endog.		39	bGTF2A1LF as1
Alf	insert.	1kb	Mb 1	38
Alf	insert.	1kb	277	123
Fes	insert.	1kb	Mb 1	326
Fes	insert.	1kb	Mb 1	36
Fes	insert.	1kb	277	37
Syt1	endog.		321	105
Syt1	insert.	1kb	Mb 1	105
Syt1	insert.	1kb	Mb 1	35
Syt1	insert.	1kb	277	34
Zic3	insert.	1kb	Mb 1	62
Orm1	insert.	1kb	Mb 1	279
Orm1	insert.	1kb	277	299
Mrap	insert.	1kb	Mb 1	289
Mrap	insert.	1kb	277	301
Hes3	endog.		59	32
Hes3	endog.		33	50
Hes3	insert.	1kb	Mb 1	404
Hes3	insert.	1.6kb	Mb 1	315
Hes3	insert.	1.6kb	277	159
Hes3	insert.	Hes3 (D)	Mb 1	315
Hes3	insert.	Hes3 (D)	277	325
Hes3	insert.	Hes3 (F)	Mb 1	315
Hes3	insert.	Hes3 (F)	277	325
Hes3	insert.	Hes3 (G)	Mb 1	315
Hes3	insert.	Hes3 (H)	277	431
Trf	insert.	Trf (B)	Mb 1	135
Trf	insert.	Trf (B)	277	133
Trf	insert.	Trf (D)	Mb 1	135
Trf	insert.	Trf (D)	277	133
Trf	insert.	Trf (F)	277	133
Trf	insert.	Trf (E)	277	133
Trf	insert.	Trf (C)	Mb 1	42
Trf	insert.	Trf (C)	277	43
Syt1	insert.	Syt1 (A)	Mb 1	35
Syt1	insert.	Syt1 (A)	277	34
Syt1	insert.	Syt1 (B)	Mb 1	105
Syt1	insert.	Syt1 (C)	Mb 1	35
Syt1	insert.	Syt1 (C)	277	134
Syt1	insert.	Syt1 (D)	Mb 1	35
Syt1	insert.	Syt1 (E)	Mb 1	292
Syt1	insert.	Syt1 (E)	277	293
Alf	insert.	Alf (A)	Mb 1	38
Alf	insert.	Alf (A)	Mb 1	128
Alf	insert.	Alf (D)	Mb 1	128
Alf	insert.	Alf (C)	Mb 1	128
Alf	insert.	Alf (C)	277	129
Nanog	insert.	Nanog (A)	Mb 1	122
Nanog	insert.	Nanog (B)	Mb 1	Nanog bis. 1

Nanog	insert.	Nanog (C)	Mb 1	Nanog bis. 1
Nanog	insert.	Nanog (C)	277	144
Nanog	insert.	Nanog (D)	Mb 1	143
Nanog	insert.	Nanog (F)	Mb 1	143
Nanog	insert.	Nanog (J)	Mb 1	143
Nanog	insert.	Nanog (J)	277	144
E.coli Fr9	insert.		Mb 1	381
E.coli Fr9	insert.		277	382
E.coli Fr3	insert.		Mb 1	370
E.coli Fr6	insert.		Mb 1	375
E.coli Fr5	insert.		373	374
E.coli Fr2	insert.		Mb 1	368
E.coli Fr2	insert.		277	357
E.coli Fr1	insert.		Mb 1	356
E.coli Fr1	insert.		277	355
E.coli Fr4	insert.		Mb 1	371
E.coli Fr4	insert.		277	372
E.coli Fr12	insert.		Mb 1	424
E.coli Fr12	insert.		277	423
E.coli Fr10	insert.		Mb 1	418
E.coli Fr10	insert.		277	417
E.coli Fr11	insert.		420	434
Hes3 - Mrap	insert.		Mb 1	315
Hes3 - Mrap	insert.		Mb 1	314
Hes3 - Mrap	insert.		Mb 1	141
Hes3 - Mrap	insert.		278	289
Hes3 - Mrap	insert.		277	301
Syt1 - Mrap	insert.		Mb 1	292
Syt1 - Mrap	insert.		293	289
Syt1 - Mrap	insert.		277	301
Alf - Mrap	insert.		Mb 1	289
Alf - Mrap	insert.		277	301
Hes3 - Orm1	insert.		Mb 1	314
Hes3 - Orm1	insert.		Mb 1	141
Hes3 - Orm1	insert.		435	279
Hes3 - Orm1	insert.		277	299
Syt1 - Orm1	insert.		Mb 1	292
Syt1 - Orm1	insert.		293	279
Syt1 - Orm1	insert.		277	299
Alf - Orm1	insert.		Mb 1	279
Alf - Orm1	insert.		277	280
Alf - Syt1	insert.		Mb 1	35
Alf - Syt1	insert.		277	34

primer #	sequence
6	AAG TTG CAG CTC CGG GTA G
10	TTT TCA CCA TGG GGT ACA GG
27	CGG CAG GCC TGT TAA ATA GA
30	TCT TGG AAG AGA AAC TCT TAG GG
32	CTT TCC AAA TAC CTT AAA AAC CCT A
33	GGA GGA GGG AAA TTT AGG TTT TT
34	GGG GTG AAT TAA GAG GAG ATT AGA T
35	AAA TCT AAT CTC CTC TTA ATT CAC C
36	AAC TAT ACT CTC TCT TCC CTA TAA TAA ACC
37	GGT TTA TTA TAG GGA AGA GAG AGT ATA GTT T
38	CAT ACC TCA CTA AAT ATC ACT CAC AAC TA
39	TTT TAT TTT TTA AAG ATG AGG TTG TTT TAT
42	AAA TTA CAT ACA TAA CCA ACC AAT CC
43	GAT TGG TTG GTT ATG TAT GTA ATT TTT
44	TGT ATA CAG ATC TAC CAA CAT TAC GA
45	AGG TAA GGC AGC CCC ACT
46	GCG AGA GAG GCA GAA GTT TG
49	CAG AAC CCC ACC CAT GTC
50	AGG CGC CAG CCA CTA TTT
50	AGG CGC CAG CCA CTA TTT
51	GCT GGT GGC TTA GAC TAG GG
53	CAC ACG TGG ATG GCT AAA TG
56	AAA CAC AGA GGA AAT CAT CCT GA
59	TTT GGT TGG AGG AAG TGA GG
62	CTC TTC ACC TCC AAA ATA ATA TCC
65	GGT GAG GAG TCA GCA GGA TG
67	CTT CCT GCA TCT GCT GCA CT
73	ATG CAC TTG TGG TCT TGC AT
85	ATC ATT CGG GGT GAA ACA AG
101	GGA AGG TGC CTC TAG AAG AGA T
105	CAA CCC CCA ATA TCT TTC CT
107	ATA GAG GGA GGG AGA GTT GTT TT
108	AAT TTA AAC CCC AAA ATT CTA ACT TC
122	TCT ATA CAA AAC ATC TCA ATA ACA AAC C
123	TTT AGT GAG GTA TGG TAA AAA TAG GAA
128	CTT CCT CCT TAC CAC CAA ATT A
129	TTA ATT TGG TGG TAA GGA GGA A
133	GGT GAT TAT TAG GGT AAG AAA GGA
134	TTT GGG TTT TAG AAG AAA AAT AAA T
135	CCC TAA TAA TCA CCT CAT TTC CT
137	GGA ATA GGG GTA AGG TTT AGG TT
141	CAA AAC AAA TTA CCC TAC TAC TAA CTT AAC
143	AAA CAT CCT CTA ATC TAA AAA CAT CC
144	GAT TAA TTG TGA ATT TAT AGG GTT GG
159	TTT AAA GTT TGT TTT TGT AGT TGA GG
171	GCA AAC GGT TAG GCA CAG AG
255	GTA TGG TGG TAG ATA AGT TTG GTT TAT
277	ATA TAA AAT AAT AAC AAT ATA TAC AAA TCT ACC AAC

278	GTT TTT GGA TAG GGG TTT GTT
279	TAC TAT CCC TAA CTT CAA TCC CAT AC
280	TTT ATT AGT ATT TGG GAA ATG TTT TG
289	CCC CTA ACC TTC CAA CCA TA
292	TAT CAA ATA ACC AAA AAC AAA CTT CT
293	TAG AAG TTT GTT TTT GGT TAT TTG AT
299	ATT TAG AGT TAT TAG GTT TGG GAA AA
301	GTA GGG TTT TTG GTT TTT GTT TTT
302	TTG ATT TGG TTG GTG TTT TG
303	AGA TAT TGA GTT TTT TGG TTG TTG
314	CCA AAC ACA ATA CCC TTA ACA CT
315	ACT TCA ATC ACA TCA CCA ACC
317	AAC CCC CAT TCC TAT CCT AC
321	ATA GAG GAG AGA ATG GAA GAA GTA AG
325	GTT GGT GAT GTG ATT GAA GTT TA
326	AAC ACT ATT AAA AAC CCT AAA C
327	TTT TAA TTT TAG TTT TTT AGA GTT AGA GGT AG
347	CTT CGC CAT CTC CAT CAT CT
349	CTA CCT GAC CAC CTG TGT CG
351	TAT ACC AGA GGC GGT TCC TG
354	TGC GTA AGT CGT ACC TGC AC
355	GGG GTG TTG AAG GTT TTA TAG ATA G
356	TAT AAA ACC TTC AAC ACC CCA CTA C
357	GGT TAT TAT TAT GAA AGT ATT AGT TTT AGG TAT G
368	AAC TCT CAA CCT TAA ACA AAA TCA A
370	CTT CAT TTC CAA ACA ACA TTA AAA C
371	ACC TAT CCT ACA CAT CAA AAC TCA A
372	ATT GAG TTT TGA TGT GTA GGA TAG G
373	TGT GAT GTA TTT TGG AAA TTG AA
374	CAA AAT CTT ACT AAA AAT TCT TAC ACT TAC T
375	CAA TTA TCA TTA CAT CAT TCC CTT T
381	CCT AAA AAT ATA CCT TTC CCT TCA
382	TGA AGG GAA AGG TAT ATT TTT AGG
386	CAA CCT GGT GGT AAG GAG GA
387	AAC GCT TGC CAT TGC TTA CT
389	CAG TGT GAA ATT GAC ATA AAT GTG G
391	TGC AGG TAG CCC AGT TCA A
404	CAA AAT AAA CAA CTA TCT CTC CTT ACC
411	AGT CCC CAC ACG CAT ATC TC
413	GCA AAA TGG TGC CGT AAC TT
415	GAA GTT GGG GAT GAC ATT GC
417	TGG TAG GGT TTA TAT TGT TGT GTT
418	CCA TCA AAT CCC CAC ATA AA
420	TTT GAT GTT ATT TGA TGT TTA GAG G
423	TTG TGT TGG GTT ATG GAA TG
424	TCC ATT AAT AAA ATA AAC AAC ATA ATC C
431	GAT TTT GTG ATT GAA GTT TAG AGG
434	ACC CAA ACA ACC AAA ATA TTA ACT T
435	TTA AGT TAG TAG TAG GGT AAT TTG TTT TG
437	CCC TAC TAC TAA AAA CAC CAC TCA

502	ATC AGG GCC TCA AAC TTC CT
503	AGA TGC CCC AGA ACT TGA TG
516	TAT GCC CGA GGA CAA TAA GG
517	ACC GCC GTT ATG AAA TCT TG
518	CGT GGC GAT CAT TTC TAG GT
519	AAC CCA CCA GAT GCT TTG TC
520	CTC CCC AAC TCA GTC AGC TC
521	CAA AAT AGG GGG AGG GAG AG
522	CAA GCG TAA GTC GGA ACA CA
523	GTT TGA GTC CCA TGC CTG AT
bGTF2A1LF as1	TAC CCT AAA ACC TAA ATA ACC TCA ATT AA
Oct as	TCA CCT AGG GAC GGT TTC AC
fm60	GCG TTT CTT GCT TGC TCT TC
Oct s	ACC TCC GTC TGG AAG ACA CA
interg3 as	GGA CAG ACA TCT GCC AAG GT
interg3 s	ATG CCC CTC AGC TAT CAC AC
Mb 1	ATT AAA TAA AAT GAA AGT TTT GGA AGA G
mGTFA1L as2	AGT AGG ACC GGG CAA GAA AG
mGTFA1L s2	TGA GGC ATG GTA AAA ACA GG
Nanog bis. 1	AAA ACA CCA ACC AAA TCA ACC
Nanog outs.1	TCC TGT GTC CCG GAT CTC TA
Nanog RT1	AAA CGG GCT GAA GGG TTA TT
Tcl1 bis. A	CAA ACC CAC ACC TAA TAA TAA TAA ATA A

3.2.3. Additional results

Demethylation of *in vitro* methylated promoters

Our site-specific insertion experiments revealed that the underlying sequence is sufficient to recapitulate an unmethylated promoter state in ES cells. While the promoter fragments used in these experiments were cloned in *E.coli* and therefore were unmethylated at the time point of insertion, previous work indicated that ES cells are able to demethylate DNA that is *in vitro* methylated before genomic integration (Frank et al. 1991). In order to test this possibility, we *in vitro* methylated sequence fragments of the Alf and Syt1 promoter and inserted them in the beta globin locus. Bisulfite PCR analysis revealed that both *in vitro* methylated promoters were unmethylated after genomic insertion in ES cells (Fig. 5). It cannot be excluded that a minor fraction of promoter plasmids resisted *in vitro* methylation and preferentially integrated in the genome. However, it is more likely that the *in vitro* methylated promoter fragments got demethylated after genomic insertion. As the cells went through several rounds of replication between the time of insertion and extraction of DNA for bisulfite PCR, the mechanism of demethylation could be active or passive. These results further implicate that the binding of transcription factors that are potentially protecting these promoters from DNA methylation is methylation-insensitive.

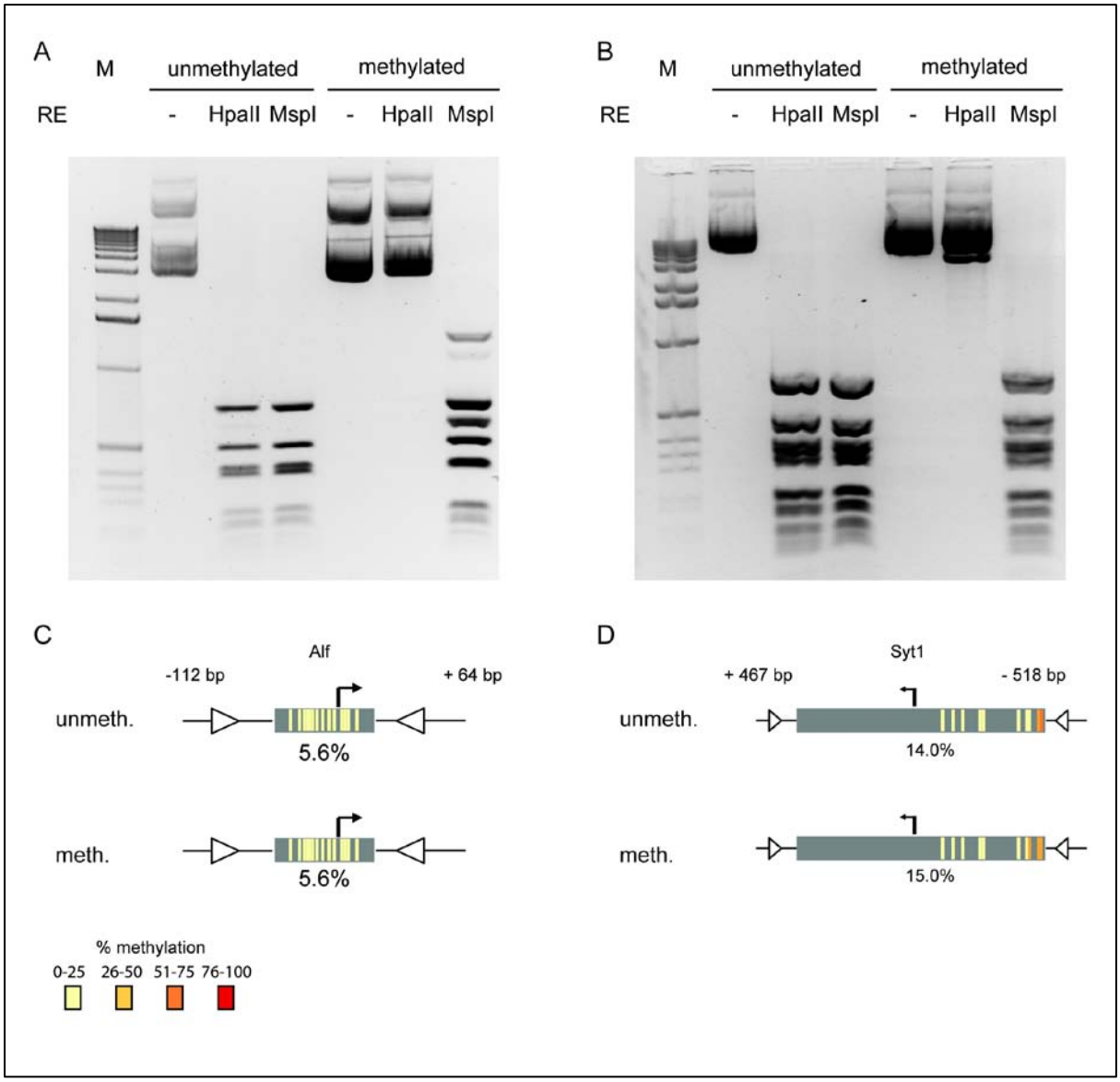


Figure 5. ***In vitro* methylated promoters get demethylated upon insertion in ES cells.**
(**A** and **B**) Restriction digest analysis of plasmids that were *in vitro* methylated at CpGs using M.SssI as described previously (Schubeler et al. 2001). The restriction enzyme (RE) MspI cuts both plasmids irrespective of methylation. *In vitro* methylated plasmids are not cut by the methylation-sensitive RE HpaII, confirming the efficiency of the M.SssI reaction. M = marker (**C** and **D**) Representation of methylation levels at Alf and Syt1 promoter fragments after *in vitro* methylation and insertion in ES cells. Methylation levels at fragments that were inserted without *in vitro* methylation are shown as a comparison.

4. General conclusions

While there are several chromatin marks that correlate with gene repression, the dynamics and establishment of their genomic distribution is still poorly understood. During my PhD studies I have investigated the genomic targeting of two repressive chromatin modifications: H3K9me2 and DNA methylation. Genome-wide mapping of H3K9me2 during neuronal differentiation revealed that this heterochromatic mark is almost as highly abundant in pluripotent ES cells as in post-mitotic neurons. Together with results from other groups our data suggest that cellular differentiation entails local, rather than global changes in repressive chromatin modifications. Our study on the establishment of local patterns of DNA methylation revealed that this process is mediated by the underlying sequence and likely involves binding of transcription factors that protect from DNA methylation.

Below I will discuss the main findings of my PhD project and how they relate to what is currently known about the targeting and function of H3K9me2 and DNA methylation.

4.1. Genome-wide distribution of H3K9me2

By genome-wide mapping we show that the H3K9me2 mark is highly abundant and occurs in large domains that occupy more than half of the genome. In ES cells, H3K9me2 is only depleted from regions that are transcribed, enriched for H3K4me2 or targeted by polycomb mediated H3K27me3. Remarkably, together these chromatin features include around 90% of all sites that are devoid of H3K9me2. These findings therefore suggest that H3K9me2 marks per default all euchromatic regions that are inactive and not repressed by polycomb. Importantly, this is in accordance with current knowledge about the protein machinery that sets the H3K9me2 mark. While G9a and GLP, the two enzymes which catalyze dimethylation of H3K9, occur as a heteromeric complex, experiments using methyltransferase-defective mutants provided evidence that most of the *in vivo* KMT activity depends on G9a (Tachibana et al. 2005; Tachibana et al. 2008). Interestingly, *in vitro* purified G9a is not able to modify H3 peptides that are K9 or K4 acetylated, or are pre-modified by K4 methylation (Chin et al. 2005). As K27me3 was not included in this study, it remains unclear whether this mark might have a similar influence on the activity of G9a. Nevertheless, this *in vitro* specificity might explain why H3K9me2 is excluded from regions with H3K4 methylation or active transcription. Additionally, structural analysis revealed that the ankyrin repeats of G9a and GLP can themselves bind to H3K9me1 and H3K9me2 (Collins et al. 2008) and it has therefore been speculated that establishment of H3K9me2 involves a spreading mechanism (Collins and Cheng 2010). Such spreading could potentially help to set up and maintain the large domains of H3K9me2 that we and others have observed (Wen et al. 2009).

Given its occurrence at most euchromatic sites that are inactive, it is tempting to speculate that H3K9me2 helps to keep these regions in an inaccessible chromatin state. In line with that, G9a seems to be involved in promoter silencing at several genes, such as MAGE-2, IF-β and P21 (Tachibana et al. 2002; Gyory et al. 2004; Nishio and Walsh 2004). However, the interaction of G9a with co-repressor machineries, such as the CtBP complex, suggests that G9a can induce silencing independently of its catalytic activity (Ueda et al. 2006). It is therefore unclear whether H3K9me2 is directly involved in transcriptional repression and how such a process could be mediated.

HP1 represents one candidate protein with the potential to directly bind to H3K9me2 and impact chromatin accessibility. While best studied for its interplay with H3K9me3 at pericentric heterochromatin, HP1 also shows *in vitro* binding activity towards H3K9me2 (Hughes et al. 2007). Furthermore, G9a and GLP seem to be required for binding of HP1 to euchromatin, as knockout of either enzyme diminishes euchromatic HP1 staining (Tachibana et al. 2005). It is therefore possible that HP1 keeps the large domains of H3K9me2 in a repressed chromatin state, in analogy to the situation at pericentric heterochromatin (Shinkai and Tachibana 2011).

Interestingly, H3K9me2 is enriched within domains of nuclear lamina association and within regions of late replication timing, which both are generally associated with repressed genes (Hiratani et al. 2008; Wen et al. 2009; Peric-Hupkes et al. 2010). Late-replication timing is however not functionally linked to H3K9me2, as an induced knock-out of G9a in ES cells does not alter replication timing (Yokochi et al. 2009). This study further revealed that deletion of G9a leads to upregulation of around 170 genes, which mostly locate to the nuclear periphery (6 out of 6 tested genes). However, loss of H3K9me2 and accompanying transcriptional upregulation does not affect location to the nuclear periphery (Yokochi et al. 2009). It therefore remains unclear how the interaction with nuclear lamina or replication timing might affect H3K9me2 occupied regions.

Alternatively, domains of H3K9me2 could be functionally linked to DNA methylation, since G9a knock-out ES cells show reduced DNA methylation at repeat elements and at promoters that get *de novo* methylated upon differentiation. However, these defects are rescued upon over-expression of a catalytic mutant of G9a, making it unlikely that the H3K9me2 mark is directly involved in targeting of DNA methylation (Dong et al. 2008; Epsztejn-Litman et al. 2008). On the other hand, H3K9me2 was suggested to facilitate recruitment of DNMT1, since UHRF1, a major mediator of this process, binds to H3K9me2 peptides *in vitro* (Hashimoto et al. 2009). However, a recent structural study questioned this result by showing that UHRF1 exclusively binds H3K9me3 and it therefore remains unclear whether H3K9me2 plays a role in ensuring proper inheritance of DNA methylation during the cell cycle (Rottach et al. 2010).

In conclusion, the large domains of H3K9me2 might be established by binding of G9a to nucleosomes that are unmodified at the most N-terminal part of H3. This does not exclude transcription factor mediated targeting of G9a to specific loci, such as to promoters during the process of transcriptional silencing. Establishment and maintenance of H3K9me2 domains might be further facilitated by a self-enforcing loop through binding of G9a to the modification it mediates. The role of H3K9me2 in these large domains remains elusive, but mostly likely involves an interaction with HP1 that could establish a repressive chromatin state.

4.2. Chromatin modifications during cellular differentiation

Cell lineage specification involves a gradual decrease in the developmental potential of differentiating cells. This process is most famously represented in Conrad Waddington's concept of the epigenetic landscape (Waddington 1957). During recent years it has become clear that chromatin modifications are involved in shaping the trajectories of this lineage specification landscape (Hemberger et al. 2009). Accordingly, it has been suggested that loss of pluripotency, the earliest step of differentiation, is accompanied by a major increase in repressive histone modifications (Gaspar-Maia et al. 2011).

In contrast to a previous study (Wen et al. 2009), our results show that the genomic occupancy of H3K9me2 is similar between pluripotent ES cells and differentiated neurons. This finding is in accordance with several additional studies on repressive chromatin modifications. It has, for example, been shown that only a small subset of promoters is subject to *de novo* DNA methylation during ES cell differentiation (Farthing et al. 2008; Meissner et al. 2008; Mohn et al. 2008). Further, while H3K27me3 and H3K9me3 occupancy was reported to be slightly higher in differentiated cells than in ES cells, the observed gain reflected a localized expansion of preexisting domains, rather than a global increase (Hawkins et al. 2010). Additionally, it was shown that the number of polycomb regulated promoters stays constant during ES differentiation (Mohn et al. 2008). Taken together, these genome-wide mapping studies suggest that repressive chromatin modifications are not generally depleted in pluripotent cells.

Even without a global depletion of repressive chromatin modifications, certain properties of the chromatin structure seem to differ between ES cells and differentiated cells (Meister et al. 2011). For instance, the histones H2A, H3 and H1, as well as HP1, show a higher turnover rate in ES cells than in committed neuronal precursor cells (Meshorer et al. 2006). Secondly, electron spectroscopic imaging revealed that ES cells possess highly dispersed chromatin, in contrast to differentiated cells that contain compact chromatin

domains (Ahmed et al. 2010). Remarkably, this apparent more open chromatin arrangement in ES cells does not affect the general mobility of genomic loci, which rather seems to be lower in ES cells when compared to differentiated cells (Masui et al. 2011). Furthermore, while imaging studies suggested a differentiation-induced increase of compact chromatin at the nuclear periphery, genome-wide mapping revealed few changes in nuclear lamina association during ES cell differentiation (Peric-Hupkes et al. 2010).

Importantly, our transcriptome analysis establishes that potential ES specific chromatin properties are not reflected by a net transcriptional output that is unique to a pluripotent cell type. This result is supported by experiments using genome-wide nuclear run-on (GRO-seq), which did not reveal any evidence for increased intergenic transcription in ES cells (Min et al. 2011).

In conclusion, our results with data from other studies show that rearrangements in the global distribution of repressive chromatin modifications are not involved in shaping Waddington's landscape of lineage specification, and suggest that this process rather entails changes of repressive histone modifications and DNA methylation that are localized to regulatory regions.

4.3. Mechanisms of setting up DNA methylation patterns

Our study on the establishment of DNA methylation patterns revealed small sequence elements (MDRs) that protect their immediate sequence environment from DNA methylation in ES cells and during differentiation. Mutations in the Alf (*GTF2A1L*) MDR showed that these elements depend on a combination of DNA binding motifs and led to the identification of RFX2 as a potential transcription factor involved in this process. RFX2 might act in similarity to the zinc finger protein VEZF1, which was recently reported to mediate the unmethylated state at the chicken β-globin enhancer and the *APRT* CpG island in ES cells (Dickson et al. 2010). In line with a role of TFs in mediating an unmethylated DNA state, it was reported that a combination of different TFs is needed to protect CpG islands from aberrant DNA methylation in tumor cells (Gebhard et al. 2010).

There are different scenarios how TF binding might induce an unmethylated state. Firstly, it has been shown that DNA binding of a bacterial protein in mammalian cells can protect from DNA methylation, opening the possibility that TFs could exclude the DNA methylation machinery by steric hindrance (Han et al. 2001). Alternatively, a combinatorial binding of different TFs could induce a nucleosome free region, which might represent an unfavorable substrate for *de novo* methyltransferases. However, since *in vitro* experiments suggest that DNMT3s favor naked DNA over chromatinized sequences, it is unclear whether

such a mechanism could occur *in vivo* (Zhang et al. 2010). Finally, transcription factors could directly recruit KMTs that catalyze H3K4 methylation, a histone modification that inhibits binding of DNMT3 proteins (Ooi et al. 2007; Otani et al. 2009; Zhang et al. 2010). At CpG islands the establishment of H3K4 methylation could be further facilitated through action of CXXC domain containing proteins, such as *CFP1*, which bind unmethylated CpGs and possess or recruit H3K4 KMT activity (Thomson et al. 2010). Of note, recruitment and local spreading of H3K4 methylation could explain our observation that MDRs can confer protection from DNA methylation to neighboring sequence regions. In fact, it is also plausible that TF binding at DMRs protects from DNA methylation by both, steric hindrance and recruitment of H3K4 methylation.

Our study further revealed that besides MDRs no additional promoter sequences are needed for recapitulation of differentiation-induced *de novo* methylation. Moreover, a MDR that remains unmethylated during differentiation can protect a neighboring MDR from *de novo* methylation. Together these findings suggest that *de novo* methylation could reflect a loss of protection, rather than active recruitment of DNMT3s. In such a model, regions that are constantly unmethylated would be protected from DNA methylation by a combination of TFs that bind throughout development. Conversely, at promoters that get *de novo* methylated, binding of such TFs would get lost during early embryogenesis. At CpG-rich promoters, protection from DNA methylation could be further facilitated through recognition of unmethylated CpGs by CXXC domain proteins that induce H3K4 methylation, which in turn inhibits binding of DNMT3s. This self-reinforcing loop might be weaker or missing at promoters that get *de novo* methylated during early development. In line with this notion, it was shown that *de novo* methylation occurs at a higher frequency at promoters that are enriched for an intermediate CpG density (Mohn et al. 2008; Borgel et al. 2010). Importantly, the here proposed model (Fig. 6) does not exclude the involvement of TFs that recruit DNMT3s to *de novo* methylated promoters, as reported for E2F6 (Velasco et al. 2010). In addition, it is also possible that TFs can initiate *de novo* methylation by targeting of H3K4 demethylases.

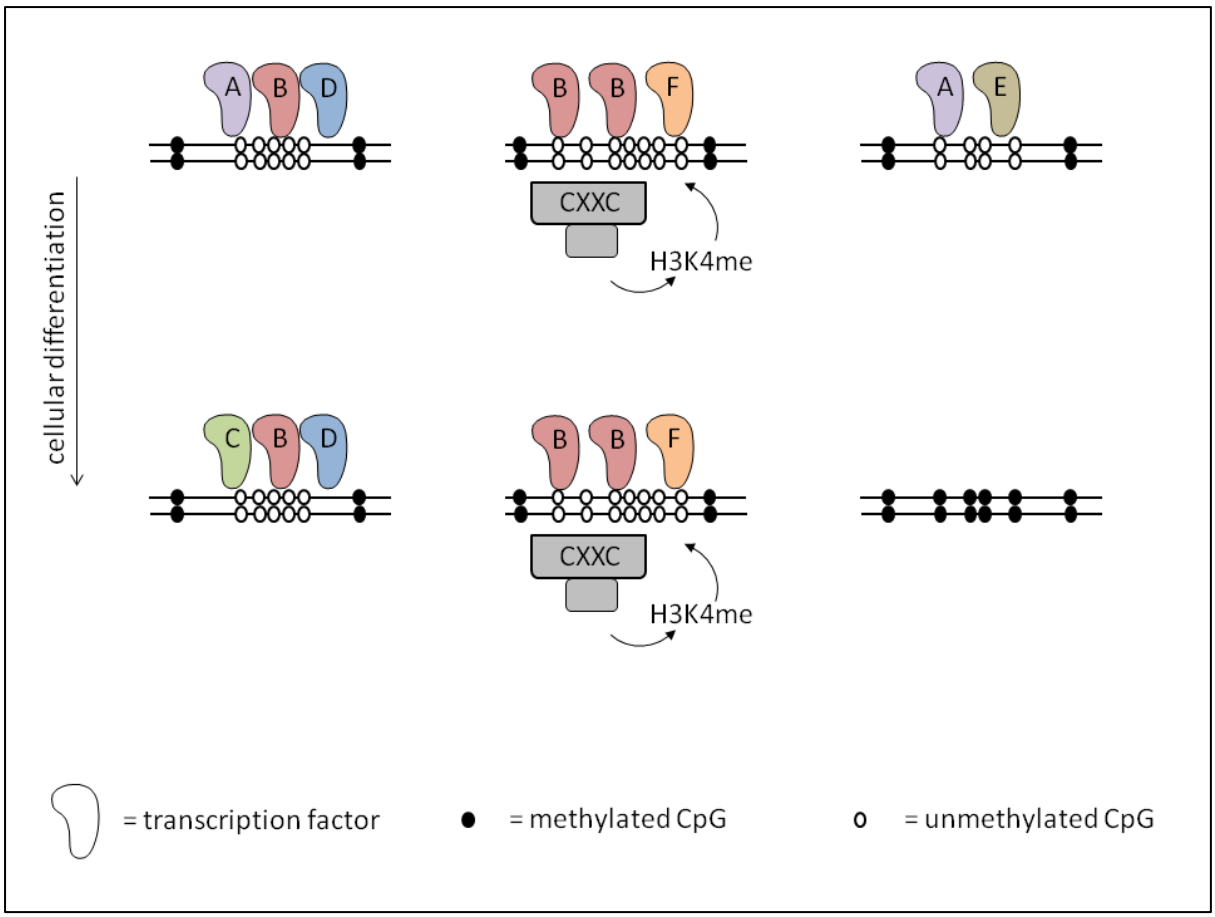


Figure 6. **Model of establishment of DNA methylation patterns.**
Combinatorial binding of transcription factors (A-F) protects from DNA methylation, possibly by steric hindrance or by recruitment of H3K4 KMTs. Regions that lose binding of these factors during cellular differentiation get *de novo* methylated (right panel). At CpG-rich sequences, protection from DNA methylation could be further facilitated through recognition of unmethylated CpGs by CXXC domain proteins that induce H3K4 methylation, which in turn inhibits binding of DNMT3s.

5. Acknowledgements

First of all I would like to thank my PhD supervisor Dirk Schübeler for giving me the opportunity to work under his guidance. He was always helpful and trained me to fight difficulties and challenges. While providing support whenever needed, he gave me the freedom to develop my own scientific ideas.

I further thank all current and past members of the Schübeler group. They always helped when necessary and made the time in the lab and outside the FMI very enjoyable. In particular I would like to point out Fabio and Christiane who provided invaluable support during my time at the FMI.

I thank Claudia, my family and my friends for being so patient and supportive.

Finally, I would like to acknowledge Susan Gasser 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

Ahmed, K., H. Dehghani, et al. (2010). "Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo." PLoS One **5**(5): e10531.

Allfrey, V. G., R. Faulkner, et al. (1964). "Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis." Proc Natl Acad Sci U S A **51**: 786-94.

Alon, U. (2007). An Introduction to Systems Biology: Design Principles of Biological Circuits. Boca Raton, FL, Chapman and Hall/CRC Press.

Angelov, D., A. Molla, et al. (2003). "The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling." Mol Cell **11**(4): 1033-41.

Appanah, R., D. R. Dickerson, et al. (2007). "An unmethylated 3' promoter-proximal region is required for efficient transcription initiation." PLoS Genet **3**(2): e27.

Aravin, A. A., R. Sachidanandam, et al. (2008). "A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice." Mol Cell **31**(6): 785-99.

Arita, K., M. Ariyoshi, et al. (2008). "Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism." Nature **455**(7214): 818-21.

Avvakumov, G. V., J. R. Walker, et al. (2008). "Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1." Nature **455**(7214): 822-5.

Bajic, V. B., S. L. Tan, et al. (2006). "Mice and men: their promoter properties." PLoS Genet **2**(4): e54.

Ball, M. P., J. B. Li, et al. (2009). "Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells." Nat Biotechnol **27**(4): 361-8.

Bannister, A. J. and T. Kouzarides (2011). "Regulation of chromatin by histone modifications." Cell Res **21**(3): 381-95.

Barreto, G., A. Schafer, et al. (2007). "Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation." Nature **445**(7128): 671-5.

Barski, A., S. Cuddapah, et al. (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-37.

Bedford, M. T. and S. G. Clarke (2009). "Protein arginine methylation in mammals: who, what, and why." Mol Cell **33**(1): 1-13.

Beisel, C. and R. Paro (2011). "Silencing chromatin: comparing modes and mechanisms." Nat Rev Genet **12**(2): 123-35.

Bell, O., C. Wirbelauer, et al. (2007). "Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila." EMBO J **26**(24): 4974-84.

Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-26.

Bestor, T., A. Laudano, et al. (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**(4): 971-83.

Bhutani, N., J. J. Brady, et al. (2010). "Reprogramming towards pluripotency requires AID-dependent DNA demethylation." Nature **463**(7284): 1042-7.

Bibel, M., J. Richter, et al. (2007). "Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells." Nat Protoc **2**(5): 1034-43.

Bibel, M., J. Richter, et al. (2004). "Differentiation of mouse embryonic stem cells into a defined neuronal lineage." Nat Neurosci **7**(9): 1003-9.

Bird, A., M. Taggart, et al. (1985). "A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA." Cell **40**(1): 91-9.

Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." Nature **321**(6067): 209-13.

Bird, A. P. (1995). "Gene number, noise reduction and biological complexity." Trends Genet **11**(3): 94-100.

Bird, A. P. and A. P. Wolffe (1999). "Methylation-induced repression--belts, braces, and chromatin." Cell **99**(5): 451-4.

Blackledge, N. P. and R. Klose (2011). "CpG island chromatin: a platform for gene regulation." Epigenetics **6**(2): 147-52.

Blackledge, N. P., J. C. Zhou, et al. (2010). "CpG islands recruit a histone H3 lysine 36 demethylase." Mol Cell **38**(2): 179-90.

Blattner, F. R., G. Plunkett, 3rd, et al. (1997). "The complete genome sequence of Escherichia coli K-12." Science **277**(5331): 1453-62.

Borgel, J., S. Guibert, et al. (2010). "Targets and dynamics of promoter DNA methylation during early mouse development." Nat Genet **42**(12): 1093-100.

Bostick, M., J. K. Kim, et al. (2007). "UHRF1 plays a role in maintaining DNA methylation in mammalian cells." Science **317**(5845): 1760-4.

Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." Nature **431**(7004): 96-9.

Bourc'his, D., G. L. Xu, et al. (2001). "Dnmt3L and the establishment of maternal genomic imprints." Science **294**(5551): 2536-9.

Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-53.

Boyes, J. and A. Bird (1992). "Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein." EMBO J **11**(1): 327-33.

Brandeis, M., D. Frank, et al. (1994). "Sp1 elements protect a CpG island from de novo methylation." Nature **371**(6496): 435-8.

Brenner, C., R. Deplus, et al. (2005). "Myc represses transcription through recruitment of DNA methyltransferase corepressor." Embo J **24**(2): 336-46.

Briggs, S. D., M. Bryk, et al. (2001). "Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae." Genes Dev **15**(24): 3286-95.

Brownell, J. E., J. Zhou, et al. (1996). "Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation." Cell **84**(6): 843-51.

Campanero, M. R., M. I. Armstrong, et al. (2000). "CpG methylation as a mechanism for the regulation of E2F activity." Proc Natl Acad Sci U S A **97**(12): 6481-6.

Cao, R., L. Wang, et al. (2002). "Role of histone H3 lysine 27 methylation in Polycomb-group silencing." Science **298**(5595): 1039-43.

Cardoso, M. C. and H. Leonhardt (1999). "DNA methyltransferase is actively retained in the cytoplasm during early development." J Cell Biol **147**(1): 25-32.

Carmell, M. A., A. Girard, et al. (2007). "MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline." Dev Cell **12**(4): 503-14.

Carninci, P., A. Sandelin, et al. (2006). "Genome-wide analysis of mammalian promoter architecture and evolution." Nat Genet **38**(6): 626-35.

Carrozza, M. J., B. Li, et al. (2005). "Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription." Cell **123**(4): 581-92.

Cedar, H. and Y. Bergman (2009). "Linking DNA methylation and histone modification: patterns and paradigms." Nat Rev Genet **10**(5): 295-304.

Chadwick, B. P. and H. F. Willard (2002). "Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome." J Cell Biol **157**(7): 1113-23.

Chatterjee, C. and T. W. Muir (2010). "Chemical approaches for studying histone modifications." J Biol Chem **285**(15): 11045-50.

Chen, D., H. Ma, et al. (1999). "Regulation of transcription by a protein methyltransferase." Science **284**(5423): 2174-7.

Chen, R. Z., S. Akbarian, et al. (2001). "Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice." Nat Genet **27**(3): 327-31.

Chin, H. G., M. Pradhan, et al. (2005). "Sequence specificity and role of proximal amino acids of the histone H3 tail on catalysis of murine G9A lysine 9 histone H3 methyltransferase." Biochemistry **44**(39): 12998-3006.

Chotalia, M., S. A. Smallwood, et al. (2009). "Transcription is required for establishment of germline methylation marks at imprinted genes." Genes Dev **23**(1): 105-17.

Chow, J. and E. Heard (2009). "X inactivation and the complexities of silencing a sex chromosome." Curr Opin Cell Biol **21**(3): 359-66.

Chuang, L. S., H. I. Ian, et al. (1997). "Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1." Science **277**(5334): 1996-2000.

Ciccone, D. N., H. Su, et al. (2009). "KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints." Nature **461**(7262): 415-8.

Clapier, C. R. and B. R. Cairns (2009). "The biology of chromatin remodeling complexes." Annu Rev Biochem **78**: 273-304.

Clouaire, T. and I. Stancheva (2008). "Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin?" Cell Mol Life Sci **65**(10): 1509-22.

Collins, R. and X. Cheng (2010). "A case study in cross-talk: the histone lysine methyltransferases G9a and GLP." Nucleic Acids Res **38**(11): 3503-11.

Collins, R. E., J. P. Northrop, et al. (2008). "The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules." Nat Struct Mol Biol **15**(3): 245-50.

Colot, V. and J. L. Rossignol (1999). "Eukaryotic DNA methylation as an evolutionary device." Bioessays **21**(5): 402-11.

Core, L. J., J. J. Waterfall, et al. (2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters." Science **322**(5909): 1845-8.

Cortazar, D., C. Kunz, et al. (2011). "Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability." Nature **470**(7334): 419-23.

Cosma, M. P., T. Tanaka, et al. (1999). "Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter." Cell **97**(3): 299-311.

Davidson, E. H. (2010). "Emerging properties of animal gene regulatory networks." Nature **468**(7326): 911-20.

De La Fuente, R., C. Baumann, et al. (2006). "Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells." Nat Cell Biol **8**(12): 1448-54.

de Napoles, M., J. E. Mermoud, et al. (2004). "Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation." Dev Cell **7**(5): 663-76.

Deal, R. B., J. G. Henikoff, et al. (2010). "Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones." Science **328**(5982): 1161-4.

Dean, W., F. Santos, et al. (2001). "Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos." Proc Natl Acad Sci U S A **98**(24): 13734-8.

Dehe, P. M. and V. Geli (2006). "The multiple faces of Set1." Biochem Cell Biol **84**(4): 536-48.

Dhasarathy, A. and P. A. Wade (2008). "The MBD protein family-reading an epigenetic mark?" Mutat Res **647**(1-2): 39-43.

Dhayalan, A., A. Rajavelu, et al. (2010). "The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation." J Biol Chem **285**(34): 26114-20.

Dickson, J., H. Gowher, et al. (2010). "VEZF1 elements mediate protection from DNA methylation." PLoS Genet **6**(1): e1000804.

Dion, M. F., S. J. Altschuler, et al. (2005). "Genomic characterization reveals a simple histone H4 acetylation code." Proc Natl Acad Sci U S A **102**(15): 5501-6.

Dong, K. B., I. A. Maksakova, et al. (2008). "DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity." EMBO J **27**(20): 2691-701.

Eckhardt, F., J. Lewin, et al. (2006). "DNA methylation profiling of human chromosomes 6, 20 and 22." Nat Genet **38**(12): 1378-85.

Engel, N., J. S. Tront, et al. (2009). "Conserved DNA methylation in Gadd45a(-/-) mice." Epigenetics **4**(2): 98-9.

Epsztejn-Litman, S., N. Feldman, et al. (2008). "De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes." Nat Struct Mol Biol **15**(11): 1176-83.

Eskeland, R., M. Leeb, et al. (2010). "Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination." Mol Cell **38**(3): 452-64.

Ezhkova, E., H. A. Pasolli, et al. (2009). "Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells." Cell **136**(6): 1122-35.

Farthing, C. R., G. Ficz, et al. (2008). "Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes." PLoS Genet **4**(6): e1000116.

Faust, C., A. Schumacher, et al. (1995). "The eed mutation disrupts anterior mesoderm production in mice." Development **121**(2): 273-85.

Feldman, N., A. Gerson, et al. (2006). "G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis." Nat Cell Biol **8**(2): 188-94.

Feng, Q., H. Wang, et al. (2002). "Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain." Curr Biol **12**(12): 1052-8.

Filion, G. J., S. Zhenilo, et al. (2006). "A family of human zinc finger proteins that bind methylated DNA and repress transcription." Mol Cell Biol **26**(1): 169-81.

Fischle, W., Y. Wang, et al. (2003). "Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains." Genes Dev **17**(15): 1870-81.

Flemming, W. (1882). Zellsubstanz, Kern und Zellteilung. Leipzig, F.C.W Vogel.

Frank, D., I. Keshet, et al. (1991). "Demethylation of CpG islands in embryonic cells." Nature **351**(6323): 239-41.

Frederiks, F., M. Tzouros, et al. (2008). "Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states." Nat Struct Mol Biol **15**(6): 550-7.

Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-82.

Gary, J. D., W. J. Lin, et al. (1996). "The predominant protein-arginine methyltransferase from Saccharomyces cerevisiae." J Biol Chem **271**(21): 12585-94.

Gaspar-Maia, A., A. Alajem, et al. (2011). "Open chromatin in pluripotency and reprogramming." Nat Rev Mol Cell Biol **12**(1): 36-47.

Gebhard, C., C. Benner, et al. (2010). "General transcription factor binding at CpG islands in normal cells correlates with resistance to de novo DNA methylation in cancer cells." Cancer Res **70**(4): 1398-407.

Gehring, M., W. Reik, et al. (2009). "DNA demethylation by DNA repair." Trends Genet **25**(2): 82-90.

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

Goll, M. G., F. Kirpekar, et al. (2006). "Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2." Science **311**(5759): 395-8.

Guenther, M. G., S. S. Levine, et al. (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**(1): 77-88.

Guibert, S., T. Forne, et al. (2009). "Dynamic regulation of DNA methylation during mammalian development." Epigenomics **1**(1): 81-98.

Guo, J. U., Y. Su, et al. (2011). "Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain." Cell **145**(3): 423-34.

Guy, J., B. Hendrich, et al. (2001). "A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome." Nat Genet **27**(3): 322-6.

Gyory, I., J. Wu, et al. (2004). "PRDI-BF1 recruits the histone H3 methyltransferase G9a in transcriptional silencing." Nat Immunol **5**(3): 299-308.

Hackenberg, M., C. Previti, et al. (2006). "CpGcluster: a distance-based algorithm for CpG-island detection." BMC Bioinformatics **7**: 446.

Hajkova, P., K. Ancelin, et al. (2008). "Chromatin dynamics during epigenetic reprogramming in the mouse germ line." Nature **452**(7189): 877-81.

Han, L., I. G. Lin, et al. (2001). "Protein binding protects sites on stable episomes and in the chromosome from de novo methylation." Mol Cell Biol **21**(10): 3416-24.

Hargreaves, D. C. and G. R. Crabtree (2011). "ATP-dependent chromatin remodeling: genetics, genomics and mechanisms." Cell Res **21**(3): 396-420.

Hashimoto, H., J. R. Horton, et al. (2009). "UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications." Epigenetics **4**(1): 8-14.

Hawkins, R. D., G. C. Hon, et al. (2010). "Distinct epigenomic landscapes of pluripotent and lineage-committed human cells." Cell Stem Cell **6**(5): 479-91.

Hebbes, T. R., A. L. Clayton, et al. (1994). "Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain." EMBO J **13**(8): 1823-30.

Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet **39**(3): 311-8.

Hellman, A. and A. Chess (2007). "Gene body-specific methylation on the active X chromosome." Science **315**(5815): 1141-3.

Hemberger, M., W. Dean, et al. (2009). "Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal." Nat Rev Mol Cell Biol **10**(8): 526-37.

Hendrich, B., J. Guy, et al. (2001). "Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development." Genes Dev **15**(6): 710-23.

Hendrich, B. and S. Tweedie (2003). "The methyl-CpG binding domain and the evolving role of DNA methylation in animals." Trends Genet **19**(5): 269-77.

Hervouet, E., F. M. Vallette, et al. (2009). "Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation." Epigenetics **4**(7): 487-99.

Hewish, D. R. and L. A. Burgoyne (1973). "Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease." Biochem Biophys Res Commun **52**(2): 504-10.

Hiratani, I., T. Ryba, et al. (2008). "Global reorganization of replication domains during embryonic stem cell differentiation." PLoS Biol **6**(10): e245.

Hirschhorn, J. N., S. A. Brown, et al. (1992). "Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure." Genes Dev **6**(12A): 2288-98.

Hochedlinger, K., Y. Yamada, et al. (2005). "Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues." Cell **121**(3): 465-77.

Hodl, M. and K. Basler (2009). "Transcription in the absence of histone H3.3." Curr Biol **19**(14): 1221-6.

Holliday, R. and J. E. Pugh (1975). "DNA modification mechanisms and gene activity during development." Science **187**(4173): 226-32.

Hu, J. L., B. O. Zhou, et al. (2009). "The N-terminus of histone H3 is required for de novo DNA methylation in chromatin." Proc Natl Acad Sci U S A **106**(52): 22187-92.

Hughes, R. M., K. R. Wiggins, et al. (2007). "Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect." Proc Natl Acad Sci U S A **104**(27): 11184-8.

Hyllus, D., C. Stein, et al. (2007). "PRMT6-mediated methylation of R2 in histone H3 antagonizes H3 K4 trimethylation." Genes Dev **21**(24): 3369-80.

Iberg, A. N., A. Espejo, et al. (2008). "Arginine methylation of the histone H3 tail impedes effector binding." J Biol Chem **283**(6): 3006-10.

Iguchi-Ariga, S. M. and W. Schaffner (1989). "CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation." Genes Dev **3**(5): 612-9.

Illingworth, R., A. Kerr, et al. (2008). "A novel CpG island set identifies tissue-specific methylation at developmental gene loci." PLoS Biol **6**(1): e22.

Illingworth, R. S. and A. P. Bird (2009). "CpG islands--'a rough guide'." FEBS Lett **583**(11): 1713-20.

Imamura, M., K. Miura, et al. (2006). "Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells." BMC Dev Biol **6**: 34.

Iyer, V. and K. Struhl (1995). "Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure." EMBO J **14**(11): 2570-9.

Jackson-Grusby, L., C. Beard, et al. (2001). "Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation." Nat Genet **27**(1): 31-9.

Jackson, M., A. Krassowska, et al. (2004). "Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells." Mol Cell Biol **24**(20): 8862-71.

Jacob, F. and J. Monod (1961). "Genetic regulatory mechanisms in the synthesis of proteins." J Mol Biol **3**: 318-56.

Jia, D., R. Z. Jurkowska, et al. (2007). "Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation." Nature **449**(7159): 248-51.

Jin, S. G., C. Guo, et al. (2008). "GADD45A does not promote DNA demethylation." PLoS Genet **4**(3): e1000013.

John, S., P. J. Sabo, et al. (2011). "Chromatin accessibility pre-determines glucocorticoid receptor binding patterns." Nat Genet **43**(3): 264-8.

Johns, E. W. (1964). "Studies on histones. 7. Preparative methods for histone fractions from calf thymus." Biochem J **92**(1): 55-9.

Jones, P. A. and S. M. Taylor (1980). "Cellular differentiation, cytidine analogs and DNA methylation." Cell **20**(1): 85-93.

Jurkowski, T. P., M. Meusburger, et al. (2008). "Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism." RNA **14**(8): 1663-70.

Kaneda, M., M. Okano, et al. (2004). "Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting." Nature **429**(6994): 900-3.

Kangaspeska, S., B. Stride, et al. (2008). "Transient cyclical methylation of promoter DNA." Nature **452**(7183): 112-5.

Kaplan, N., I. K. Moore, et al. (2009). "The DNA-encoded nucleosome organization of a eukaryotic genome." Nature **458**(7236): 362-6.

Kato, Y., M. Kaneda, et al. (2007). "Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse." Hum Mol Genet **16**(19): 2272-80.

Keogh, M. C., S. K. Kurdistani, et al. (2005). "Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex." Cell **123**(4): 593-605.

Kerppola, T. K. (2009). "Polycomb group complexes--many combinations, many functions." Trends Cell Biol **19**(12): 692-704.

Klug, M., S. Heinz, et al. (2010). "Active DNA demethylation in human postmitotic cells correlates with activating histone modifications, but not transcription levels." Genome Biol **11**(6): R63.

Knezetic, J. A. and D. S. Luse (1986). "The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro." Cell **45**(1): 95-104.

Kobayashi, I. (2001). "Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution." Nucleic Acids Res **29**(18): 3742-56.

Kolsto, A. B., G. Kollias, et al. (1986). "The maintenance of methylation-free islands in transgenic mice." Nucleic Acids Res **14**(24): 9667-78.

Kornberg, R. D. (1974). "Chromatin structure: a repeating unit of histones and DNA." Science **184**(139): 868-71.

Kornberg, R. D. and J. O. Thomas (1974). "Chromatin structure; oligomers of the histones." Science **184**(139): 865-8.

Kossel, A. (1884). "Ueber einen peptonartigen Bestandtheil des Zellkerns." Zeitschr. phys. Chemi **8**: 511-515.

Kota, S. K. and R. Feil (2010). "Epigenetic transitions in germ cell development and meiosis." Dev Cell **19**(5): 675-86.

Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.

Kriaucionis, S. and N. Heintz (2009). "The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain." Science **324**(5929): 929-30.

Krogan, N. J., M. Kim, et al. (2003). "Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II." Mol Cell Biol **23**(12): 4207-18.

Kuramochi-Miyagawa, S., T. Watanabe, et al. (2008). "DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes." Genes Dev **22**(7): 908-17.

Lachner, M., D. O'Carroll, et al. (2001). "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins." Nature **410**(6824): 116-20.

Lam, F. H., D. J. Steger, et al. (2008). "Chromatin decouples promoter threshold from dynamic range." Nature **453**(7192): 246-50.

Larson, D. R., R. H. Singer, et al. (2009). "A single molecule view of gene expression." Trends Cell Biol **19**(11): 630-7.

Law, J. A. and S. E. Jacobsen (2010). "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." Nat Rev Genet **11**(3): 204-220.

Lee, C. K., Y. Shibata, et al. (2004). "Evidence for nucleosome depletion at active regulatory regions genome-wide." Nat Genet **36**(8): 900-5.

Lee, J. H. and D. G. Skalnik (2005). "CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex." J Biol Chem **280**(50): 41725-31.

Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-13.

Lehnertz, B., Y. Ueda, et al. (2003). "Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin." Curr Biol **13**(14): 1192-200.

Lei, H., S. P. Oh, et al. (1996). "De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells." Development **122**(10): 3195-205.

Lessard, J., J. I. Wu, et al. (2007). "An essential switch in subunit composition of a chromatin remodeling complex during neural development." Neuron **55**(2): 201-15.

Lewis, Z. A., K. K. Adhvaryu, et al. (2010). "DNA methylation and normal chromosome behavior in Neurospora depend on five components of a histone methyltransferase complex, DCDC." PLoS Genet **6**(11): e1001196.

Li, E., C. Beard, et al. (1993). "Role for DNA methylation in genomic imprinting." Nature **366**(6453): 362-5.

Li, E., T. H. Bestor, et al. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell **69**(6): 915-26.

Li, J. Y., D. J. Lees-Murdock, et al. (2004). "Timing of establishment of paternal methylation imprints in the mouse." Genomics **84**(6): 952-60.

Li, X., M. Ito, et al. (2008). "A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints." Dev Cell **15**(4): 547-57.

Lin, Y., M. E. Gill, et al. (2008). "Germ cell-intrinsic and -extrinsic factors govern meiotic initiation in mouse embryos." Science **322**(5908): 1685-7.

Lister, R., R. C. O'Malley, et al. (2008). "Highly integrated single-base resolution maps of the epigenome in Arabidopsis." Cell **133**(3): 523-36.

Lister, R., M. Pelizzola, et al. (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." Nature **462**(7271): 315-22.

Lorch, Y., J. W. LaPointe, et al. (1987). "Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones." Cell **49**(2): 203-10.

Lu, X., M. D. Simon, et al. (2008). "The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure." Nat Struct Mol Biol **15**(10): 1122-4.

Lucifero, D., M. R. Mann, et al. (2004). "Gene-specific timing and epigenetic memory in oocyte imprinting." Hum Mol Genet **13**(8): 839-49.

Luger, K. (2003). "Structure and dynamic behavior of nucleosomes." Curr Opin Genet Dev **13**(2): 127-35.

Luger, K., A. W. Mader, et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-60.

Maatouk, D. M., L. D. Kellam, et al. (2006). "DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages." Development **133**(17): 3411-8.

Macleod, D., J. Charlton, et al. (1994). "Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island." Genes Dev **8**(19): 2282-92.

Maeda, R. K. and F. Karch (2006). "The ABC of the BX-C: the bithorax complex explained." Development **133**(8): 1413-22.

Maison, C., D. Bailly, et al. (2011). "SUMOylation promotes de novo targeting of HP1alpha to pericentric heterochromatin." Nat Genet **43**(3): 220-7.

Malik, H. S. and S. Henikoff (2003). "Phylogenomics of the nucleosome." Nat Struct Biol **10**(11): 882-91.

Malik, S. and R. G. Roeder (2010). "The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation." Nat Rev Genet **11**(11): 761-72.

Margueron, R. and D. Reinberg (2010). "Chromatin structure and the inheritance of epigenetic information." Nat Rev Genet **11**(4): 285-96.

Marin, M., A. Karis, et al. (1997). "Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation." Cell **89**(4): 619-28.

Marinus, M. G. and J. Casadesus (2009). "Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more." FEMS Microbiol Rev **33**(3): 488-503.

Martin Caballero, I., J. Hansen, et al. (2009). "The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation." PLoS One **4**(1): e4315.

Martinez-Antonio, A. and J. Collado-Vides (2003). "Identifying global regulators in transcriptional regulatory networks in bacteria." Curr Opin Microbiol **6**(5): 482-9.

Marzluff, W. F., Jr., L. A. Sanders, et al. (1972). "Two chemically and metabolically distinct forms of calf thymus histone F3." J Biol Chem **247**(7): 2026-33.

Masui, O., I. Bonnet, et al. (2011). "Live-Cell Chromosome Dynamics and Outcome of X Chromosome Pairing Events during ES Cell Differentiation." Cell **145**(3): 447-58.

Matsui, T., D. Leung, et al. (2010). "Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET." Nature **464**(7290): 927-31.

Matzke, M., T. Kanno, et al. (2009). "RNA-mediated chromatin-based silencing in plants." Curr Opin Cell Biol **21**(3): 367-76.

Maunakea, A. K., R. P. Nagarajan, et al. (2010). "Conserved role of intragenic DNA methylation in regulating alternative promoters." Nature **466**(7303): 253-7.

Mayer, W., A. Niveleau, et al. (2000). "Demethylation of the zygotic paternal genome." Nature **403**(6769): 501-2.

Meissner, A., T. S. Mikkelsen, et al. (2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells." Nature **454**(7205): 766-70.

Meister, P., S. E. Mango, et al. (2011). "Locking the genome: nuclear organization and cell fate." Curr Opin Genet Dev **21**(2): 167-74.

Mendenhall, E. M., R. P. Koche, et al. (2010). "GC-rich sequence elements recruit PRC2 in mammalian ES cells." PLoS Genet **6**(12): e1001244.

Meshorer, E. and T. Misteli (2006). "Chromatin in pluripotent embryonic stem cells and differentiation." Nat Rev Mol Cell Biol **7**(7): 540-6.

Meshorer, E., D. Yellajoshula, et al. (2006). "Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells." Dev Cell **10**(1): 105-16.

Metivier, R., R. Gallais, et al. (2008). "Cyclical DNA methylation of a transcriptionally active promoter." Nature **452**(7183): 45-50.

Miescher, F. (1871). "Ueber die chemische Zusammensetzung der Eiterzellen." med. chem. Unters.(4): 441-460.

Mikkelsen, T. S., J. Hanna, et al. (2008). "Dissecting direct reprogramming through integrative genomic analysis." Nature **454**(7200): 49-55.

Mikkelsen, T. S., M. Ku, et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature **448**(7153): 553-60.

Milne, T. A., Y. Dou, et al. (2005). "MLL associates specifically with a subset of transcriptionally active target genes." Proc Natl Acad Sci U S A **102**(41): 14765-70.

Min, I. M., J. J. Waterfall, et al. (2011). "Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells." Genes Dev **25**(7): 742-54.

Mito, Y., J. G. Henikoff, et al. (2007). "Histone replacement marks the boundaries of cis-regulatory domains." Science **315**(5817): 1408-11.

Mohn, F. and D. Schubeler (2009). "Genetics and epigenetics: stability and plasticity during cellular differentiation." Trends Genet **25**(3): 129-36.

Mohn, F., M. Weber, et al. (2008). "Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors." Mol Cell **30**(6): 755-66.

Morgan, H. D., W. Dean, et al. (2004). "Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming." J Biol Chem **279**(50): 52353-60.

Muller, J. and J. A. Kassis (2006). "Polycomb response elements and targeting of Polycomb group proteins in Drosophila." Curr Opin Genet Dev **16**(5): 476-84.

Nakamura, T., Y. Arai, et al. (2007). "PGC7/Stella protects against DNA demethylation in early embryogenesis." Nat Cell Biol **9**(1): 64-71.

Neigeborn, L. and M. Carlson (1984). "Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae." Genetics **108**(4): 845-58.

Ng, R. K., W. Dean, et al. (2008). "Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5." Nat Cell Biol **10**(11): 1280-90.

Nishio, H. and M. J. Walsh (2004). "CCAAT displacement protein/cut homolog recruits G9a histone lysine methyltransferase to repress transcription." Proc Natl Acad Sci U S A **101**(31): 11257-62.

O'Carroll, D., S. Erhardt, et al. (2001). "The polycomb-group gene Ezh2 is required for early mouse development." Mol Cell Biol **21**(13): 4330-6.

O'Carroll, D., H. Scherthan, et al. (2000). "Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression." Mol Cell Biol **20**(24): 9423-33.

Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-57.

Okano, M., S. Xie, et al. (1998). "Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases." Nat Genet **19**(3): 219-20.

Ong, C. T. and V. G. Corces (2011). "Enhancer function: new insights into the regulation of tissue-specific gene expression." Nat Rev Genet **12**(4): 283-93.

Ooi, S. K. and T. H. Bestor (2008). "The colorful history of active DNA demethylation." Cell **133**(7): 1145-8.

Ooi, S. K., C. Qiu, et al. (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." Nature **448**(7154): 714-7.

Orford, K., P. Kharchenko, et al. (2008). "Differential H3K4 methylation identifies developmentally poised hematopoietic genes." Dev Cell **14**(5): 798-809.

Oswald, J., S. Engemann, et al. (2000). "Active demethylation of the paternal genome in the mouse zygote." Curr Biol **10**(8): 475-8.

Otani, J., T. Nankumo, et al. (2009). "Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain." EMBO Rep **10**(11): 1235-41.

Oudet, P., M. Gross-Bellard, et al. (1975). "Electron microscopic and biochemical evidence that chromatin structure is a repeating unit." Cell **4**(4): 281-300.

Palmer, D. K., K. O'Day, et al. (1991). "Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone." Proc Natl Acad Sci U S A **88**(9): 3734-8.

Panning, B. and R. Jaenisch (1996). "DNA hypomethylation can activate Xist expression and silence X-linked genes." Genes Dev **10**(16): 1991-2002.

Pasini, D., A. P. Bracken, et al. (2004). "Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity." EMBO J **23**(20): 4061-71.

Pastor, W. A., U. J. Pape, et al. (2011). "Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells." Nature **473**(7347): 394-7.

Peric-Hupkes, D., W. Meuleman, et al. (2010). "Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation." Mol Cell **38**(4): 603-13.

Peters, A. H., D. O'Carroll, et al. (2001). "Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability." Cell **107**(3): 323-37.

Peterson, C. L. and I. Herskowitz (1992). "Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription." Cell **68**(3): 573-83.

Phalke, S., O. Nickel, et al. (2009). "Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2." Nat Genet **41**(6): 696-702.

Pokholok, D. K., C. T. Harbison, et al. (2005). "Genome-wide map of nucleosome acetylation and methylation in yeast." Cell **122**(4): 517-27.

Popp, C., W. Dean, et al. (2010). "Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency." Nature **463**(7284): 1101-5.

Prendergast, G. C. and E. B. Ziff (1991). "Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region." Science **251**(4990): 186-9.

Probst, A. V., E. Dunleavy, et al. (2009). "Epigenetic inheritance during the cell cycle." Nat Rev Mol Cell Biol **10**(3): 192-206.

Ptashne, M. (1967). "Specific binding of the lambda phage repressor to lambda DNA." Nature **214**(5085): 232-4.

Rai, K., I. J. Huggins, et al. (2008). "DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45." Cell **135**(7): 1201-12.

Raisner, R. M., P. D. Hartley, et al. (2005). "Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin." Cell **123**(2): 233-48.

Ramirez-Carrozzi, V. R., D. Braas, et al. (2009). "A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling." Cell **138**(1): 114-28.

Rea, S., F. Eisenhaber, et al. (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases." Nature **406**(6796): 593-9.

Reik, W. (2007). "Stability and flexibility of epigenetic gene regulation in mammalian development." Nature **447**(7143): 425-32.

Riggs, A. D. (1975). "X inactivation, differentiation, and DNA methylation." Cytogenet Cell Genet **14**(1): 9-25.

Rinn, J. L., M. Kertesz, et al. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell **129**(7): 1311-23.

Roguev, A., D. Schaft, et al. (2001). "The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4." EMBO J **20**(24): 7137-48.

Roh, T. Y., S. Cuddapah, et al. (2006). "The genomic landscape of histone modifications in human T cells." Proc Natl Acad Sci U S A **103**(43): 15782-7.

Roh, T. Y., S. Cuddapah, et al. (2005). "Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping." Genes Dev **19**(5): 542-52.

Rottach, A., C. Frauer, et al. (2010). "The multi-domain protein Np95 connects DNA methylation and histone modification." Nucleic Acids Res **38**(6): 1796-804.

Rountree, M. R. and E. U. Selker (2010). "DNA methylation and the formation of heterochromatin in Neurospora crassa." Heredity **105**(1): 38-44.

Ruthenburg, A. J., C. D. Allis, et al. (2007). "Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark." Mol Cell **25**(1): 15-30.

Saksouk, N., N. Avvakumov, et al. (2009). "HBO1 HAT complexes target chromatin throughout gene coding regions via multiple PHD finger interactions with histone H3 tail." Mol Cell **33**(2): 257-65.

Sandelin, A., P. Carninci, et al. (2007). "Mammalian RNA polymerase II core promoters: insights from genome-wide studies." Nat Rev Genet **8**(6): 424-36.

Santaguida, S. and A. Musacchio (2009). "The life and miracles of kinetochores." EMBO J **28**(17): 2511-31.

Santos-Rosa, H., R. Schneider, et al. (2003). "Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin." Mol Cell **12**(5): 1325-32.

Santos, F., B. Hendrich, et al. (2002). "Dynamic reprogramming of DNA methylation in the early mouse embryo." Dev Biol **241**(1): 172-82.

Sasaki, H. and Y. Matsui (2008). "Epigenetic events in mammalian germ-cell development: reprogramming and beyond." Nat Rev Genet **9**(2): 129-40.

Sato, N., M. Kondo, et al. (2006). "The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing." Biochem Biophys Res Commun **344**(3): 845-51.

Schaefer, M. and F. Lyko (2010). "Lack of evidence for DNA methylation of Invader4 retroelements in Drosophila and implications for Dnmt2-mediated epigenetic regulation." Nat Genet **42**(11): 920-1; author reply 921.

Schmitz, K. M., C. Mayer, et al. (2010). "Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes." Genes Dev **24**(20): 2264-9.

Schnetz, M. P., C. F. Bartels, et al. (2009). "Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns." Genome Res **19**(4): 590-601.

Schoeftner, S., A. K. Sengupta, et al. (2006). "Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing." EMBO J **25**(13): 3110-22.

Schotta, G., M. Lachner, et al. (2004). "A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin." Genes Dev **18**(11): 1251-62.

Schubeler, D., M. C. Lorincz, et al. (2000). "Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation." Mol Cell Biol **20**(24): 9103-12.

Schubeler, D., M. C. Lorincz, et al. (2001). "Targeting silence: the use of site-specific recombination to introduce in vitro methylated DNA into the genome." Sci STKE **2001**(83): pl1.

Schubeler, D., D. M. MacAlpine, et al. (2004). "The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote." Genes Dev **18**(11): 1263-71.

Seibler, J., D. Schubeler, et al. (1998). "DNA cassette exchange in ES cells mediated by Flp recombinase: an efficient strategy for repeated modification of tagged loci by marker-free constructs." Biochemistry **37**(18): 6229-34.

Seila, A. C., J. M. Calabrese, et al. (2008). "Divergent transcription from active promoters." Science **322**(5909): 1849-51.

Seki, Y., K. Hayashi, et al. (2005). "Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice." Dev Biol **278**(2): 440-58.

Seki, Y., M. Yamaji, et al. (2007). "Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice." Development **134**(14): 2627-38.

Selker, E. U., N. A. Tountas, et al. (2003). "The methylated component of the Neurospora crassa genome." Nature **422**(6934): 893-7.

Sharif, J., M. Muto, et al. (2007). "The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA." Nature **450**(7171): 908-12.

Shen-Orr, S. S., R. Milo, et al. (2002). "Network motifs in the transcriptional regulation network of Escherichia coli." Nat Genet **31**(1): 64-8.

Shinkai, Y. and M. Tachibana (2011). "H3K9 methyltransferase G9a and the related molecule GLP." Genes Dev **25**(8): 781-8.

Shogren-Knaak, M., H. Ishii, et al. (2006). "Histone H4-K16 acetylation controls chromatin structure and protein interactions." Science **311**(5762): 844-7.

Sing, A., D. Pannell, et al. (2009). "A vertebrate Polycomb response element governs segmentation of the posterior hindbrain." Cell **138**(5): 885-97.

Skene, P. J., R. S. Illingworth, et al. (2010). "Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state." Mol Cell **37**(4): 457-68.

Song, C. X., K. E. Szulwach, et al. (2011a). "Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine." Nat Biotechnol **29**(1): 68-72.

Song, J., O. Rechkoblit, et al. (2011b). "Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation." Science **331**(6020): 1036-40.

Straussman, R., D. Nejman, et al. (2009). "Developmental programming of CpG island methylation profiles in the human genome." Nat Struct Mol Biol **16**(5): 564-71.

Strohner, R., M. Wachsmuth, et al. (2005). "A 'loop recapture' mechanism for ACF-dependent nucleosome remodeling." Nat Struct Mol Biol **12**(8): 683-90.

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

Suzuki, M., T. Yamada, et al. (2006). "Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b." Oncogene **25**(17): 2477-88.

Suzuki, M. M. and A. Bird (2008). "DNA methylation landscapes: provocative insights from epigenomics." Nat Rev Genet **9**(6): 465-76.

Szyf, M., G. Tanigawa, et al. (1990). "A DNA signal from the Thy-1 gene defines de novo methylation patterns in embryonic stem cells." Mol Cell Biol **10**(8): 4396-400.

Tachibana, M., Y. Matsumura, et al. (2008). "G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription." EMBO J **27**(20): 2681-90.

Tachibana, M., K. Sugimoto, et al. (2002). "G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis." Genes Dev **16**(14): 1779-91.

Tachibana, M., J. Ueda, et al. (2005). "Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9." Genes Dev **19**(7): 815-26.

Tahiliani, M., K. P. Koh, et al. (2009). "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1." Science **324**(5929): 930-5.

Takai, D. and P. A. Jones (2002). "Comprehensive analysis of CpG islands in human chromosomes 21 and 22." Proc Natl Acad Sci U S A **99**(6): 3740-5.

Talbert, P. B. and S. Henikoff (2010). "Histone variants--ancient wrap artists of the epigenome." Nat Rev Mol Cell Biol **11**(4): 264-75.

Tanay, A., A. H. O'Donnell, et al. (2007). "Hyperconserved CpG domains underlie Polycomb-binding sites." Proc Natl Acad Sci U S A **104**(13): 5521-6.

Taniguchi, T. and B. de Crombrugghe (1983). "Interactions of RNA polymerase and the cyclic AMP receptor protein on DNA of the E. coli galactose operon." Nucleic Acids Res **11**(15): 5165-80.

Taunton, J., C. A. Hassig, et al. (1996). "A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p." Science **272**(5260): 408-11.

Taverna, S. D., S. Ilin, et al. (2006). "Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs." Mol Cell **24**(5): 785-96.

Taverna, S. D., H. Li, et al. (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nat Struct Mol Biol **14**(11): 1025-40.

Taylor, I. C., J. L. Workman, et al. (1991). "Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains." Genes Dev **5**(7): 1285-98.

Terranova, R., H. Agherbi, et al. (2006). "Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll." Proc Natl Acad Sci U S A **103**(17): 6629-34.

Thieffry, D., A. M. Huerta, et al. (1998). "From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli." Bioessays **20**(5): 433-40.

Thomson, J. P., P. J. Skene, et al. (2010). "CpG islands influence chromatin structure via the CpG-binding protein Cfp1." Nature **464**(7291): 1082-6.

Timinszky, G., S. Till, et al. (2009). "A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation." Nat Struct Mol Biol **16**(9): 923-9.

Ueda, J., M. Tachibana, et al. (2006). "Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP." J Biol Chem **281**(29): 20120-8.

van Attikum, H. and S. M. Gasser (2009). "Crosstalk between histone modifications during the DNA damage response." Trends Cell Biol **19**(5): 207-17.

van Leeuwen, F., P. R. Gafken, et al. (2002). "Dot1p modulates silencing in yeast by methylation of the nucleosome core." Cell **109**(6): 745-56.

Vaquerizas, J. M., S. K. Kummerfeld, et al. (2009). "A census of human transcription factors: function, expression and evolution." Nat Rev Genet **10**(4): 252-63.

Velasco, G., F. Hube, et al. (2010). "Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues." Proc Natl Acad Sci U S A **107**(20): 9281-6.

Vermeulen, M., K. W. Mulder, et al. (2007). "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4." Cell **131**(1): 58-69.

Vire, E., C. Brenner, et al. (2006). "The Polycomb group protein EZH2 directly controls DNA methylation." Nature **439**(7078): 871-4.

Voo, K. S., D. L. Carlone, 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**(6): 2108-21.

Waddington, C. H. (1957). The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology. London, Allen & Unwin.

Walsh, C. P. and T. H. Bestor (1999). "Cytosine methylation and mammalian development." Genes Dev **13**(1): 26-34.

Walsh, C. P., J. R. Chaillet, et al. (1998). "Transcription of IAP endogenous retroviruses is constrained by cytosine methylation." Nat Genet **20**(2): 116-7.

Wang, H., L. Wang, et al. (2004). "Role of histone H2A ubiquitination in Polycomb silencing." Nature **431**(7010): 873-8.

Wang, Y., M. Jorda, et al. (2006). "Functional CpG methylation system in a social insect." Science **314**(5799): 645-7.

Wang, Z., C. Zang, et al. (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." Nat Genet **40**(7): 897-903.

Wassenegger, M., S. Heimes, et al. (1994). "RNA-directed de novo methylation of genomic sequences in plants." Cell **76**(3): 567-76.

Watanabe, D., I. Suetake, et al. (2002). "Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis." Mech Dev **118**(1-2): 187-90.

Watanabe, T., S. Tomizawa, et al. (2011). "Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus." Science **332**(6031): 848-52.

Weber, M., J. J. Davies, et al. (2005). "Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells." Nat Genet **37**(8): 853-62.

Weber, M., I. Hellmann, et al. (2007). "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome." Nat Genet **39**(4): 457-66.

Wen, B., H. Wu, et al. (2009). "Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells." Nat Genet **41**(2): 246-50.

Whitcomb, S. J., A. Basu, et al. (2007). "Polycomb Group proteins: an evolutionary perspective." Trends Genet **23**(10): 494-502.

Whitehouse, I., O. J. Rando, et al. (2007). "Chromatin remodelling at promoters suppresses antisense transcription." Nature **450**(7172): 1031-5.

Whitehouse, I. and T. Tsukiyama (2006). "Antagonistic forces that position nucleosomes in vivo." Nat Struct Mol Biol **13**(7): 633-40.

Whitelaw, N. C. and E. Whitelaw (2008). "Transgenerational epigenetic inheritance in health and disease." Curr Opin Genet Dev **18**(3): 273-9.

Wirbelauer, C., O. Bell, et al. (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**(15): 1761-6.

Woo, C. J., P. V. Kharchenko, et al. (2010). "A region of the human HOXD cluster that confers polycomb-group responsiveness." Cell **140**(1): 99-110.

Wu, H., A. C. D'Alessio, et al. (2011). "Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells." Nature **473**(7347): 389-93.

Wysocka, J., T. Swigut, et al. (2006). "A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling." Nature **442**(7098): 86-90.

Yang, X. J. and E. Seto (2007). "HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention." Oncogene **26**(37): 5310-8.

Yap, K. L., S. Li, et al. (2010). "Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a." Mol Cell **38**(5): 662-74.

Yoder, J. A., N. S. Soman, et al. (1997). "DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe." J Mol Biol **270**(3): 385-95.

Yokochi, T., K. Poduch, et al. (2009). "G9a selectively represses a class of late-replicating genes at the nuclear periphery." Proc Natl Acad Sci U S A **106**(46): 19363-8.

Zemach, A., I. E. McDaniel, et al. (2010). "Genome-wide evolutionary analysis of eukaryotic DNA methylation." Science **328**(5980): 916-9.

Zeng, L. and M. M. Zhou (2002). "Bromodomain: an acetyl-lysine binding domain." FEBS Lett **513**(1): 124-8.

Zhang, X., J. Yazaki, et al. (2006). "Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis." Cell **126**(6): 1189-201.

Zhang, Y., R. Jurkowska, et al. (2010). "Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail." Nucleic Acids Res **38**(13): 4246-53.

Zhang, Y., Z. Moqtaderi, et al. (2009). "Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo." Nat Struct Mol Biol **16**(8): 847-52.

Zhao, J., B. K. Sun, et al. (2008a). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." Science **322**(5902): 750-6.

Zhao, Q., G. Rank, et al. (2009). "PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing." Nat Struct Mol Biol **16**(3): 304-11.

Zhao, X., V. Jankovic, et al. (2008b). "Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity." Genes Dev **22**(5): 640-53.

Zhao, X., T. Ueba, 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**(11): 6777-82.

Zhu, B., S. S. Mandal, et al. (2005). "The human PAF complex coordinates transcription with events downstream of RNA synthesis." Genes Dev **19**(14): 1668-73.

Zilberman, D., M. Gehring, et al. (2007). "Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription." Nat Genet **39**(1): 61-9.

Zilberman, D. and S. Henikoff (2007). "Genome-wide analysis of DNA methylation patterns." Development **134**(22): 3959-65.

Zlatanova, J. and A. Thakar (2008). "H2A.Z: view from the top." Structure **16**(2): 166-79.

Florian Lienert

Leonhardsstr. 29, 4051 Basel, Switzerland florian.lienert@fmi.ch +41 (61) 697 8270

EDUCATION

08/2011 – present **Postdoctoral training in Biology**

Friedrich Miescher Institute, Basel, Switzerland
Supervisor: Prof. Dr. Dirk Schübeler
Project synopsis:

- Sequence determinants of DNA methylation at imprinting control regions

04/2007 – 08/2011 **PhD studies in Biology**

Friedrich Miescher Institute, Basel, Switzerland
Title: Targeting and dynamics of gene repression during stem cell differentiation.
Supervisor: Prof. Dr. Dirk Schübeler
Project synopsis:

- Systematically tested the sequence dependency of promoter DNA methylation
- Mapped histone modifications and analyzed transcriptional landscape during differentiation of embryonic stem cells

11/2006 **Master of Science ETH**

ETH Zürich, Zürich, Switzerland

07/2005 – 05/2006 **Diploma thesis**

Brain Research Institute, ETH Zürich, Zürich, Switzerland
Title: Cadherins as molecular markers for different motor and sensory neuron subtypes in the chicken embryo.
Supervisors: Prof. Dr. Martin E. Schwab and Dr. Matthias Gesemann

10/2004 – 11/2006 **Advanced studies in neuroscience, cell biology and immunology**

ETH Zürich, Zürich, Switzerland

10/2002 – 10/2004 **Basic studies in biology**

ETH Zürich, Zürich, Switzerland

RESEARCH EXPERIENCE AND TECHNICAL SKILLS

Basic molecular biology techniques
Cell culture
Epigenome and transcriptome analysis
Bioinformatics analysis of ChIP on chip, ChIPseq and RNAseq data

POSTERS AND PRESENTATIONS

F. Lienert, C. Wirbelauer, I. Som, A. Dean, F. Mohn and D. Schübeler (December 2010) DNA methylation states of *cis*-regulatory elements are defined genetically in a modular fashion. NoE EpiGeneSys meeting, Paris, 2010. (Poster)

F. Lienert, F. Mohn, A. Dean and D. Schübeler (March 2010) Proximal promoter elements determine DNA methylation during somatic differentiation. Cold Spring Harbor Laboratory meeting on Systems Biology: Global Regulation of Gene Expression, CSHL, 2010. (Presentation)

ACHIEVEMENTS

- Selected to give a talk at the Cold Spring Harbor Laboratory meeting on Systems Biology: Global Regulation of Gene Expression, CSHL, 2010.

- Received PhD fellowship of the Boehringer Ingelheim Fonds.