

Identification of genetic elements that autonomously determine DNA methylation states

Florian Lienert^{1,2}, Christiane Wirbelauer¹, Indrani Som³, Ann Dean³, 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 established¹. In mammals, DNA methylation is essential² 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 islands³. 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 well¹³.

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 disease²² or environment-induced changes, such as those observed in the brain upon stimulation²³.

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 islands²⁴. 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}.

¹Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. ²Faculty of Science, University of Basel, Basel, Switzerland. ³Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, US National Institutes of Health, Bethesda, Maryland, USA. ⁴Present 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 erythropoiesis²⁶ 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 cells²⁷ (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 element³⁴ recapitulated the methylation state of the endogenous locus after insertion in the β -globin locus. That is, both the

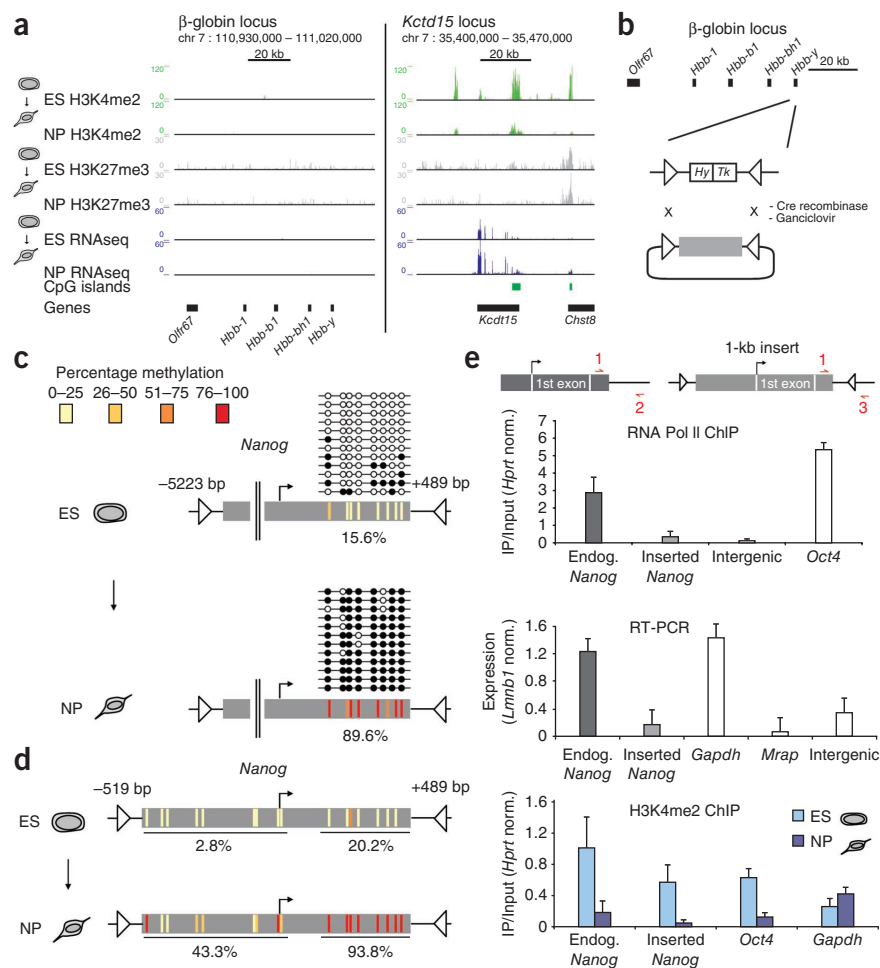
promoter and the enhancer region of the transgene remained unmethylated in stem cells (Fig. 1c and Supplementary Fig. 1). Upon *in vitro* differentiation, the enhancer region remained unmethylated, whereas the proximal promoter became *de novo* methylated to once again mirror the endogenous locus (Fig. 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 H3K27me3²⁷. 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 *Lmb1*. 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 survey¹⁵, 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.

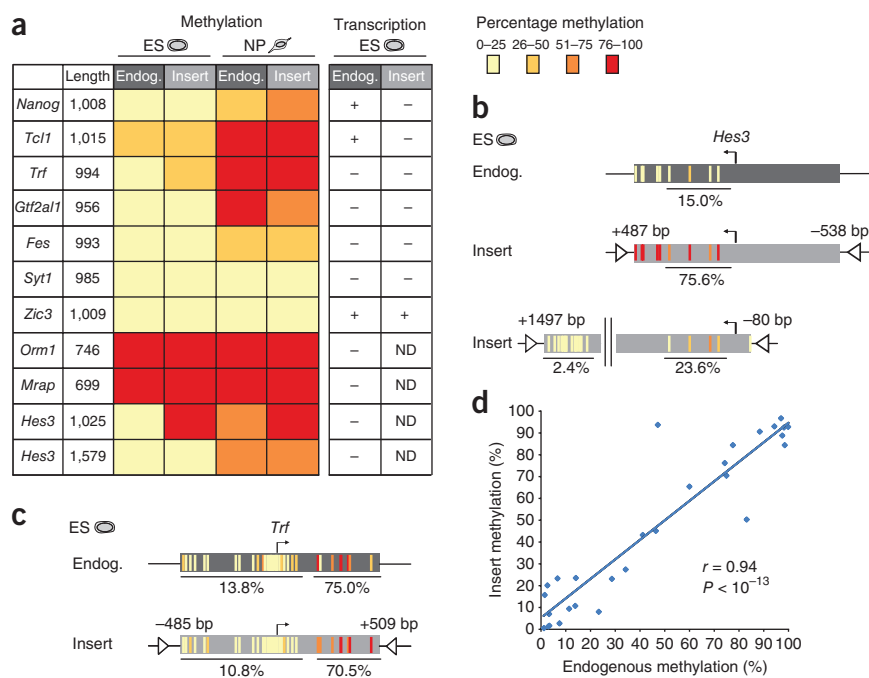
Of note, the sole fragment that did not recapitulate its endogenous epigenetic state was a 1,025-bp fragment encompassing the TSS of *Hes3*. In this case, the insert acquired methylation in stem cells, even though the endogenous locus is unmethylated. We therefore hypothesized that the fully functional 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).

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).



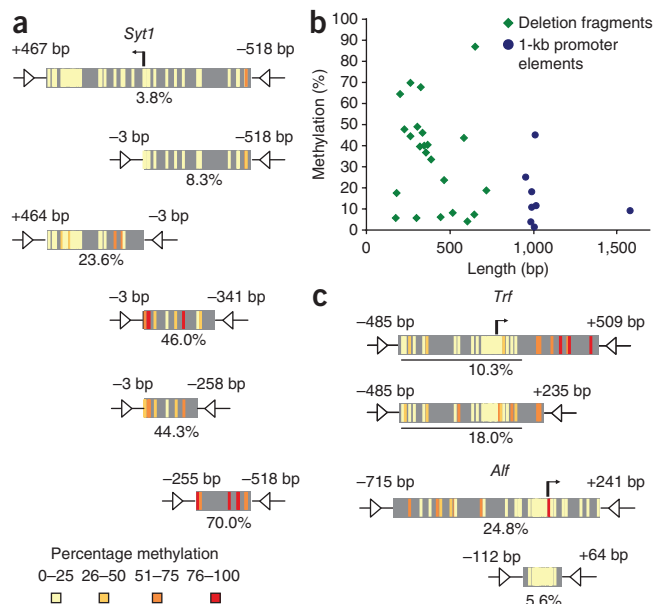
As for *Nanog*, all the inserted promoters that were tested recapitulated the endogenous H3K4 methylation pattern and its changes during differentiation (Supplementary Fig. 3).

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 *Tcf1* 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 survey⁴. 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 *Gtf2a1* (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



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

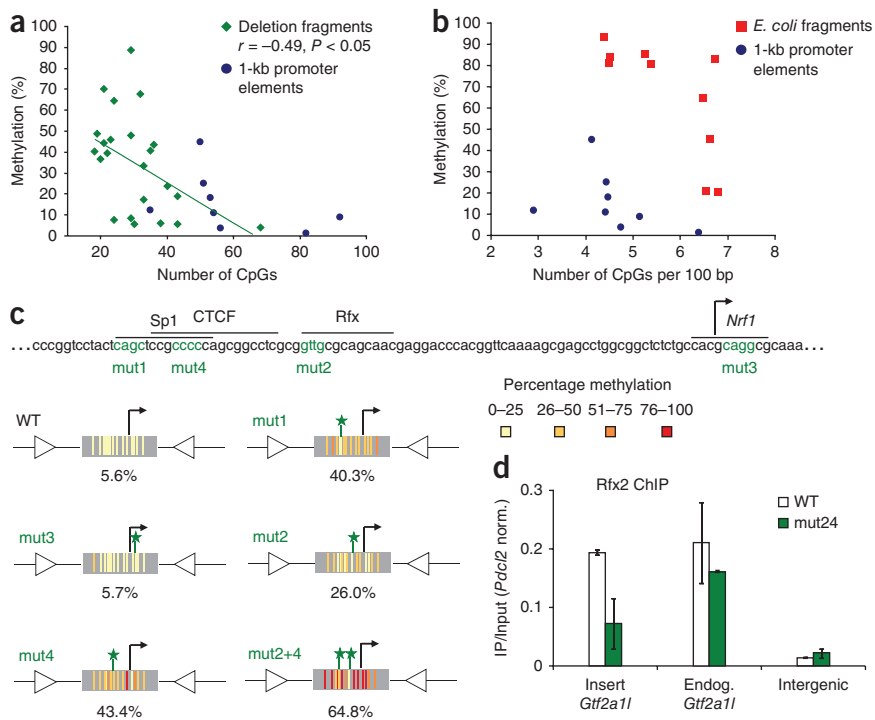


Figure 4 MDR function depends on CpG density and DNA-binding motifs. (a) Methylation level in ES cells plotted against number of CpGs at inserted promoter elements and deletion fragments. (b) Methylation level in ES cells plotted against CpG density at inserted promoter elements and *E. coli* sequence fragments. (c) Part of the inserted *Gtf2a1* 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 *Gtf2a1* MDR. ChIP enrichments were normalized to *Pdcl2*, which also contains a Rfx-binding site³⁸. Error bars indicate s.d. from two independent biological replicates. WT, wild type.

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.

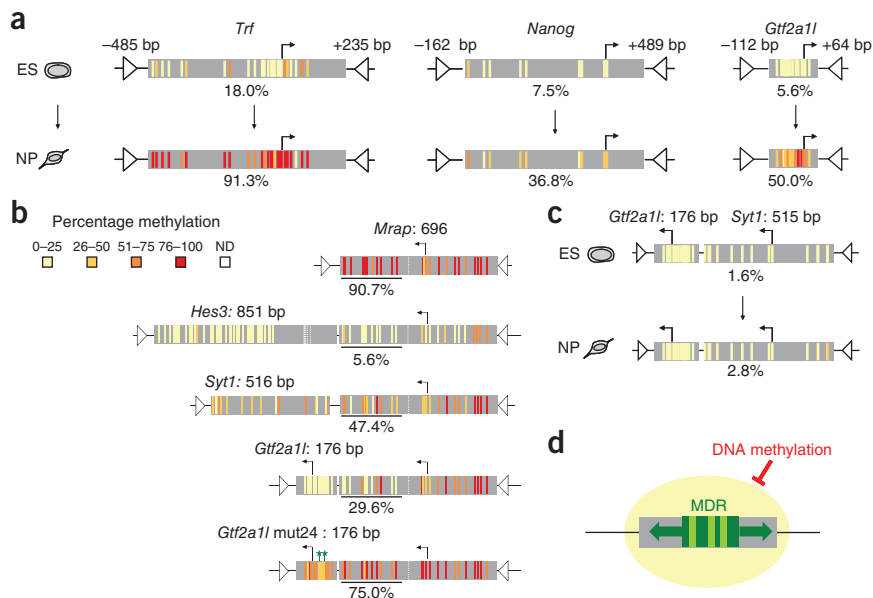
would require DNA sequence that contains no such binding sites. To maximize evolutionary distance and thus to minimize the likelihood of functional binding sites for mammalian transcription factors, we inserted ten random genomic regions from the prokaryote *Escherichia coli*. The tested *E. coli* sequences 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 unmethylated¹⁵ (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 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 *Gtf2a11* promoters. In each case the MDR confers partial hypomethylation *in cis*. This effect is diminished in a hybrid with a mutated *Gtf2a11* MDR. (c) CpG methylation levels at a hybrid of the *Gtf2a11* 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 *Gtf2a11* promoter³⁶. 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 family³⁶. 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 *Gtf2a11* promoter in spermatocytes³⁸. These experiments revealed that Rfx2 binds the *Gtf2a11* 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 *Gtf2a11* 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 *Gtf2a11* 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 *Gtf2a11*, *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 *Gtf2a11* 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 *Gtf2a11* 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 *Gtf2a11* and *Syt1* MDRs are both capable of conferring hypomethylation on heterologous sequences in stem cells, they act in opposing ways during differentiation. *Gtf2a11* 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 *Gtf2a11*

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 mechanism³⁹. 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 methylation⁴¹. 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 CpGs⁴², 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 regions⁴³.

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 *Gtf2a11* 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 cancer⁴⁸.

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>.

- Law, J.A. & Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
- Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
- Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
- Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
- Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).
- Brandeis, M. *et al.* Sp1 elements protect a CpG island from *de novo* methylation. *Nature* **371**, 435–438 (1994).
- 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).
- Dickson, J. *et al.* VEZF1 elements mediate protection from DNA methylation. *PLoS Genet.* **6**, e1000804 (2010).
- Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
- Hawkins, R.D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
- Tamaru, H. & Selker, E.U. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414**, 277–283 (2011).
- 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-1L²⁸). Further 5' to the upstream arm at the FspI site of the vector, a Sall–ClaI fragment containing the gene encoding diphtheria toxin A (DTA) was inserted.

TC-1 ES cells were electroporated with 100 μ g of pZRMCE plasmid using a Bio-Rad Gene Pulser (at 500 μ F and 250 V/cm). Cells were selected with 150 μ g/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 described²⁸ with slight modifications. TC-1 ES cells were selected under hygromycin (250 μ g/ml, Roche) for 10 d. Next, 4×10^6 cells were electroporated (Amaxa nucleofection, Amaxa) with 25 μ g of

L1-promoter-1L plasmid and 15 μ g of pIC-Cre (a kind gift from R. Terranova). Selection with 3 μ M 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 described¹⁵, starting with 70 μ g of chromatin and 5 μ g 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 $1 / C_p$, 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 μ g) was bisulfite converted with the EpiTect Bisulfite Kit (QIAGEN). Regions of interest were amplified by PCR and cloned by TOPOTA cloning (Invitrogen). Sequences were analyzed using BiQ Analyzer⁵¹. 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).