Turnover and function of DNA methylation at transcription factor binding sites

INAUGURALDISSERTATION

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

ANGELIKA FELDMANN

aus Großheppach, Deutschland

Basel, 2014

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von Prof. Dr. Dirk Schübeler, Prof. Dr. Antoine Peters, Prof. Dr. Michael Rehli

Basel, den 25.03.2014

Prof. Dr. Jörg Schibler

(Dekan)

Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 3.0 Schweiz
(CC BY-NC-ND 3.0 CH)

Sie dürfen: **Teilen** — den Inhalt kopieren, verbreiten und zugänglich machen

Unter den folgenden Bedingungen:

Namensnennung — Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Keine kommerzielle Nutzung — Sie dürfen diesen Inhalt nicht für kommerzielle Zwecke nutzen.

Keine Bearbeitung erlaubt — Sie dürfen diesen Inhalt nicht bearbeiten, abwandeln oder in anderer Weise verändern.

Wobei gilt:

- **Verzichtserklärung —** Jede der vorgenannten Bedingungen kann **aufgehoben** werden, sofern Sie die ausdrückliche Einwilligung des Rechteinhabers dazu erhalten.
- **Public Domain (gemeinfreie oder nicht-schützbare Inhalte) —** Soweit das Werk, der Inhalt oder irgendein Teil davon zur Public Domain der jeweiligen Rechtsordnung gehört, wird dieser Status von der Lizenz in keiner Weise berührt.
- **Sonstige Rechte —** Die Lizenz hat keinerlei Einfluss auf die folgenden Rechte:
    - Die Rechte, die jedermann wegen der Schranken des Urheberrechts oder aufgrund gesetzlicher Erlaubnisse zustehen (in einigen Ländern als grundsätzliche Doktrin des **fair use** bekannt);
    - Die **Persönlichkeitsrechte** des Urhebers;
    - Rechte anderer Personen, entweder am Lizenzgegenstand selber oder bezüglich seiner Verwendung, zum Beispiel für **Werbung** oder Privatsphärenschutz.
- **Hinweis —** Bei jeder Nutzung oder Verbreitung müssen Sie anderen alle Lizenzbedingungen mitteilen, die für diesen Inhalt gelten. Am einfachsten ist es, an entsprechender Stelle einen Link auf diese Seite einzubinden.

Quelle: http://creativecommons.org/licenses/by-nc-nd/3.0/ch/ Datum: 12.11.2013

Für meine Eltern

Acknowledgments

First and foremost I would like to thank my PhD supervisor Dirk Schübeler for making my PhD experience as nice as it could possibly be. He did not only teach me how to approach a problem, but most importantly how to find the right question to ask. I am grateful for the inspiring discussions and the guidance I received during my thesis, but also for the trust to work independently on my project.

Furthermore I would like to thank all current and past members of the Schübeler lab for the truly scientific environment and critical discussions. Especially, I want to acknowledge Rabih Murr and Robert Ivanek for the fruitful collaboration on the manuscript and scientific advice, but also Arnaud Krebs, Altuna Akalin and Juliane Schmidt for technical help.

I would be far from finishing my PhD without our great facilities. I am deeply grateful for the help provided by Lukas Burger. He was open to any question I had and his invaluable bioinformatic support was as important to me as our countless discussions. Additionally, I would like to thank Sophie Dessus-Babus and Tim Roloff for the considerable amounts of sequencing data generated during my project.

The Marie Curie Initial Training Network “Nucleosome 4D” offered me the opportunity to get in contact with many bright people. I enjoyed our meetings and appreciate the financial and personal support I received from the network and its members.

I am thankful to Antoine Peters and Fred Zilbermann for the help with the conditional knockout system and to Frank Edenhofer from the University of Bonn for providing the Cre protein.

Finally, I would like to thank Michael Rehli and Antoine Peters as members of my thesis committee for their fruitful contributions during my thesis committee meetings and for evaluating my thesis.

Ich danke von ganzem Herzen meinen Eltern für ihren unerschütterlichen Glauben an mich und für ihr Interesse an meiner Arbeit. Ihr habt mich motiviert und aufgebaut, wann immer es nötig war. Meiner Schwester danke ich dafür, dass sie es immer wieder geschafft hat, mich auf den Boden der Tatsachen zu bringen und mir zu zeigen, dass ein Leben ausserhalb des Labors auch möglich ist. Ich danke meinen Freunden für die (nicht immer) wissenschaftlichen Diskussionen. Ein ganz besonderer Dank gilt meiner grossen Liebe Karsten für seine unendliche Geduld und Verständnis sowie für sein Vertrauen in mich.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ac	Acetylation
Afp	Alpha-fetoprotein
AID	Activation-induced deaminase
APOBEC	Apolipoprotein B mRNA editing enzyme catalytic polypeptide
BER	Base-excision repair
bp	Basepairs
CBP/p300	CREB-binding protein/E1A binding protein p300
CFP1	CXXC finger protein 1
CGI	CpG island
CHD	Chromodomain, helicase, DNA binding
ChIP-Seq	Chromatin immunoprecipitation followed by sequencing
CREB	cAMP response element-binding protein
CTCF	CCCTC-binding factor (zinc finger protein)
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
Egr-2	Early growth response 2
ES cell	Embryonic stem cell
FMR	Fully methylated region
FoxA1	Forkhead box A1
GATA	GATA binding protein 1
HDAC	Histone deacetylase
HP1	Heterochromatic protein 1
IAP	Intracisternal A particle
Igf2	Insulin-like growth factor 2
Igf-2r	Insulin-like growth factor 2 receptor
INO80	Inositol requiring 80
ISWI	Imitation switch
Isw2	Imitation switch protein 2
Jarid1b	Jumonji, AT rich interactive domain 1B

LMR	Low methylated region
LSD1	Lysine-specific demethylase 1
MBD	Methyl-CpG-binding domain
me1	Monomethylation
me2	Dimethylation
me3	Trimethylation
MeCP2	Methyl CpG binding protein 2
miR-17-92	microRNA-17-92
NDR	Nucleosome depleted region
NFR	Nucleosome-free region
NP	Neuronal progenitor
OCT4 (POU5F1)	POU Class 5 Homeobox 1
PCNA	Proliferating cell nuclear antigen
PHD	Plant homeo domain
PMD	Partially methylated domain
REST	RE1-silencing transcription factor
RNA	Ribonucleic acid
RSC	Remodel the structure of chromatin
SAGA	Spt-Ada-Gcn5-acetyltransferase
SAM	S-Adenosyl-Methionine
SBF	Swi4-Swi6 cell cycle box, CACGAAA
SET1	SET domain-containing 1
SETD1	SET domain-containing protein 1
SWI/SNF	Switch/sucrose nonfermentable
TDG	T DNA glycosylase
TET1/2/3	Ten-eleven translocation-1/2/3
Tex13	Testis-expressed gene 13
TF	Transcription factor
TSS	Transcription start site
UHRF1	Ubiquitin-like with PHD and ring finger domains 1
UMR	Unmethylated region
YY1	Yin and Yang 1 protein

1 Summary

Cell type identity is largely determined by regulatory networks consistent of various transcription factors. Transcription factor activity requires interaction with DNA and thus critically depends on the accessibility of binding motifs. Growing evidence suggests that interactions between transcription factors and DNA are modulated by distinct chromatin modifications which in turn are influenced by transcription factors. Thus, ultimately transcriptional output is a product of intimate interactions between DNA, transcription factors and chromatin modifications. While recent studies support a model in which DNA sequence in collaboration with transcription factors can autonomously determine chromatin states, exact relationship between all these components is not well understood.

Full genome single basepair resolution mammalian methylomes (Hodges *et al,* 2011; Stadler *et al,* 2011) demonstrated a correlation between transcription factor occupancy and hypomethylation at distal regulatory regions. Importantly, these low methylated states critically depend on the presence of transcription factors. Here we analyzed how DNA binding factors impact DNA methylation. Using chromatin immunoprecipitation followed by bisulfite sequencing, we show that CTCF bound molecules can vary in their methylation levels at such low methylated regions (LMRs). This observation suggests that no tight link exists between DNA binding of transcription factors and unmethylated state. While cytosines which are highly occupied by CTCF indeed are fully devoid of methylation, cytosines within sites of low occupancy display heterogeneous methylation levels. Moreover, at these sites CTCF occupancy correlates with the likelihood of being demethylated. 5-hydroxymethylcytosine (5hmC) is a putative intermediate of active demethylation. In support of a dynamic model of interaction between transcription factors and DNA methylation, we found that 5hmC is highly enriched at cell type specific and constitutive LMRs in embryonic stem cells and upon their neuronal differentiation. Furthermore, regions with hydroxymethylation changes between these cell types are enriched for cell type specific LMRs. This suggests a participation of transcription factor mediated oxidative demethylation in reprogramming of distal regulatory elements. Knockout of CTCF is lethal for embryonic stem cells. Therefore, in order to test the relationship between transcription factor binding and hydroxymethylation we chose an embryonic stem (ES) cell line with genetic deletion of REST, another factor previously shown to be involved in formation of low methylated states. Indeed, deletion of REST

decreased 5-hydroxymethylcytosine levels while concomitantly increasing methylation levels at its binding sites within the analyzed LMRs. These results indicate that transcription factor mediated turnover of DNA methylation acts in maintenance and reprogramming of distal regulatory regions.

To test whether the observed turnover is selective for active regulatory regions, we decided to delete the two *de novo* DNA methyltransferases DNMT3A and DNMT3B in embryonic stem cells. Surprisingly, using this approach we detected loss of methylation at both, low and fully methylated regions. In order to compare the turnover kinetics between different segment subtypes, we collected DNA from ES cells at various time points after DNMT3A/B deletion. This indeed revealed an accelerated turnover at low methylated regions. On average full demethylation was achieved after eight days, suggesting that binding of transcription factors can induce rapid changes in DNA methylation.

In summary, this study supports a model in which methylation at distal regulatory regions is maintained and reprogrammed by a transcription factor mediated turnover. We furthermore provide evidence that this turnover depends on TET proteins for demethylation and on DNMT3A/B for remethylation. Quantification suggests that while DNA methylation turnover is present throughout the genome it is accelerated at active distal regulatory elements.

2 Introduction

Mammalian development begins as a single fertilized oocyte followed by multiple cell divisions. During this period cells have to accomplish the complicated task of acquiring a new identity, ultimately differentiating into many distinct cell types that form an entire organism. While the genetic content in differentiated cells remains largely unchanged, this genetic information has to be correctly interpreted to execute cell type specific functions. As a result, precise regulation of gene expression in space and time is crucial for diversification and maintenance of cell fate (Davidson 2010). Considering that hundreds of cell types exist in the human body, this represents a challenging and highly complex task. To achieve precise orchestration of transcriptional programs, higher eukaryotes use several mechanistic layers (Struhl 1999). The first regulatory level is given by the patterns encoded in DNA which directly guide DNA binding factors to their site of action where these can initiate transcriptional programs. In a second layer, DNA is packaged into chromatin by being wrapped around proteins called histones whose presence impacts binding of transcription factors. Finally, chromatin can be chemically modified. These epigenetic marks can further influence chromatin structure and attract or repel additional proteins.

It is becoming increasingly appreciated that an intricate interplay exists between these layers and we are just beginning to understand how they affect each other and transcription as a whole. I have extended this knowledge by showing how factor binding influences the stability of an epigenetic modification.

In the following paragraphs I will give a more detailed introduction on gene regulation in the context of chromatin and epigenetic modifications.

2.1 Transcriptional regulation

Evolution has been accompanied by a burst of genome size disproportional to the increase in gene number. The human haploid genome with its 3.4 gigabases (Gregory 2014) exceeds the genome size of the bacteria *Echerischia coli* by a 1000 fold, but contains only about seven times as many protein-coding genes. On the other hand, the human genome is a 100 times smaller than that of the marbled lungfish *Protopterus aethiopicus*, the animal with the largest known genome. These examples nicely illustrate that the complexity of an organism is not dictated by the pure size of the genome.

In animals expansion in genome size is believed to be caused by an accumulation of transposons which make up almost 50% of the human genome (Kidwell 2002). Indeed, particularly organisms with obligatory sexual reproduction display an increased likelihood of transposon fixation. It has been demonstrated that a transposon has to cause more than 50% reduction in the fitness of such an organism in order to be deleted (Bestor 1999). As transposition represents a threat for genomic stability, parallel invention of mechanisms controlling their expression was critical for the survival of species with large genomes (Bird 1995; Bestor 1999). It is thus possible that genome expansion together with the necessity to repress transposons represent a major driving force for the evolution of complex gene regulation mechanisms.

The basic principle of gene regulation is the interaction of transcription factors with DNA sequence. Forced expression of only a few transcription factors is sufficient for reprogramming of an adult differentiated cell into a pluripotent stem cell (Takahashi and Yamanaka 2006) demonstrating the power of this phenomenon. However, this process is rather inefficient and stochastic, suggesting that additional barriers need to be overcome in order to ensure robust changes of gene expression. Indeed, while prokaryotes can regulate their genes through a combination of transcription (co-)factors and regulatory sequences, mammalian gene regulation employs more mechanisms (Joseph *et al,* 2010; Kaplan *et al,* 2011).

As a key difference to prokaryotes, eukaryotic DNA is compacted to chromatin by histone proteins. In addition to packaging DNA, chromatin creates a general physical barrier for transcription, as it renders the DNA less permissive for binding factors (Knezetic and Luse 1986; Struhl 1999; Levine and Tjian 2003). The level of compaction can further be modulated by chemical modifications of histone proteins which enable changes in chromatin accessibility. This can be achieved by altering the electric charge of DNA or creating binding sites for effector proteins (Bannister and Kouzarides 2011).

In addition, mammalian genomes are decorated by DNA methylation, a covalent modification of cytosines present only in large eukaryotic genomes (Bestor 1990). Similar to chromatin, it has been proposed to create a genome-wide restrictive state (Bird 1995). Recent studies extended the list of covalent DNA modifications to oxidation products of DNA methylation, namely hydroxy-, formyl- and carboxymethylation (He *et al,* 2011; Ito *et al,* 2011). However, their role in gene regulation is less understood (see 2.4.5). DNA and histone modifications are currently referred to as “epigenetic modifications”.

Given this restrictive environment, in order to gain access to their binding site, mammalian transcription factors need to overcome a number of physical barriers. As a result, mammalian gene regulation is a product of a close collaboration between DNA sequences, chromatin modifications and transcription factor binding. Importantly, while prokaryotic transcriptional regulation occurs in the absence of chromatin modifications, both transcription factors and epigenetic modifiers are essential for mammalian development (Nichols *et al,* 1998; Okano *et al,* 1999; Ringrose and Paro 2004). Thus, mammalian gene regulation cannot be considered separately from the chromatin context (Joseph *et al,* 2010; Kaplan *et al,* 2011). Exact crosstalk between all these components is not fully understood. Yet, several lines of evidence suggest an autonomous function of DNA sequence in determining its epigenetic and transcriptional state in a process which is largely transcription factor dependent (Lienert *et al,* 2011; Schubeler 2012; Arnold *et al,* 2013). For instance, promoters and distal regulatory elements can accurately reproduce spatial and temporal characteristics of their chromatin and DNA modification states when introduced at ectopic sites.

As this thesis investigates chromatin modifications at distal regulatory elements, I will first introduce the major *cis*-regulatory modules of the genome, namely promoters, enhancers and insulators.

2.1.1 *Cis*-regulatory elements

Promoters

Promoters of protein-coding genes function locally to initiate transcription from the transcriptional start site (TSS) by attracting the core transcriptional machinery consisting of general transcription factors (such as TFIID) and RNA polymerase II (Pribnow 1975; Schaller *et al,* 1975; Gannon *et al,* 1979; Corden *et al,* 1980; Grosschedl *et al,* 1981). Three major types have been described in metazoans: tissue-specific, constitutive and developmentally regulated promoters (reviewed in (Lenhard *et al,* 2012)). These

promoters differ in respect to their underlying sequence, their chromatin organization and modifications. For example, tissue-specific promoters are DNA methylated and contain a TATA-box for recruitment of the basal transcriptional machinery. In contrast, the other promoter subtypes display high level of CpG dinucleotides indicative of CpG islands and consequently remain mostly DNA unmethylated in any transcriptional state (see 2.4). Despite these differences, all active promoters possess similar histone modifications, as has been initially observed in yeast (Santos-Rosa *et al,* 2002; Pokholok *et al,* 2005) and later confirmed for the human genome (Heintzman *et al,* 2007). This suggests an intimate crosstalk between gene expression and their epigenetic state.

Enhancers

The activity of the core transcriptional machinery is further modulated by additional transcription factors (TFs). These bind to proximal and distal regulatory elements (enhancers or silencers) that can be located many megabases away (Banerji *et al,* 1981; Fromm and Berg 1983; Gillies *et al,* 1983; Scholer and Gruss 1984). Their interaction with the transcriptional machinery at promoters is therefore often regulated by chromatin looping (reviewed by (Chambeyron and Bickmore 2004; Fraser 2006)).

Enhancers usually contain clusters of short 6-12 basepair motifs presenting binding sites for different TFs (Arnosti and Kulkarni 2005; Boyer *et al,* 2005; Carroll *et al,* 2006; Spitz and Furlong 2012). In many cases enhancer activity directly depends on combinatorial binding of several transcription factors which can be modulated spatially by cell type specific (Mullen *et al,* 2011; Trompouki *et al,* 2011) or temporally by developmentally regulated sequential expression (Cirillo *et al,* 2002; Serandour *et al,* 2011).

Combinatorial binding is useful for many different reasons. In some cases, direct interactions can change TF affinity or specificity towards its binding site. For instance, cooperative binding can increase the motif affinity of binding partners (Johnson *et al,* 1979). Furthermore, interaction with a cofactor which does not bind to DNA itself can alter the DNA binding specificity of a TF (Siggers *et al,* 2011). Binding of one TF may also be necessary to recruit other complex-forming factors. In addition to direct protein-protein interactions, indirect cooperativity is possible by creation of an accessible binding site by nucleosome displacement during “assisted loading” (Voss *et al,* 2011) or “collaborative competition” of two transcription factors (Miller and Widom 2003). Another well-described phenomenon is chromatin remodeling by “pioneer factors” (Zaret and Carroll 2011). Indeed, chromatin accessibility of a motif increases the likelihood of binding even for a factor capable of occupying a closed side (John *et al,* 2011)

emphasizing the importance of chromatin remodeling for TF binding. Moreover, favorable changes in DNA conformation by a preceding factor can indirectly help recruitment of other factors ("bending", (Falvo *et al,* 1995)). As an additional mechanism, interaction with common enhancer-activating factors, such as CBP/p300, can enhance transcriptional activity (Merika *et al,* 1998).

Experimentally, location of transcription factors in the genome is determined by enrichment-based methods, such as chromatin immunoprecipitation followed by sequencing (Furey 2012). Such datasets of a quality sufficient to determine consensus motifs only exist for some transcription factors. Out of those, all TFs occupy only a small subset of their binding motifs present in the genome (Carr and Biggin 1999; Iyer *et al,* 2001; Joseph *et al,* 2010; Kaplan *et al,* 2011). How exactly TFs select between multiple options is not well understood, however this is likely to involve the different modes of cooperation described above.

Active and poised enhancers are furthermore characterized by specific chromatin modifications which enable identification of their genome-wide localization (Heintzman *et al,* 2009; Heintzman and Ren 2009). Although their function is not fully understood these modifications could regulate chromatin accessibility and thus enable transcription factor binding (Lupien *et al,* 2008). In addition, enhancers display local hypomethylation only when active in a manner that appears dependent on the binding of transcription factors (Stadler *et al,* 2011).

Insulators

Insulators exert their regulatory role over entire chromatin regions by separating two regulatory domains in the genome (Bell *et al,* 2001; Phillips-Cremins and Corces 2013). One classic insulator activity is enhancer blocking. The most prominent example is certainly the H19/Igf2 imprinted locus (Bell *et al,* 1999). In this locus CTCF binds at an insulator element exclusively in the maternal allele and regulates enhancer function to repress Igf2 in allele-specific manner. Importantly, allele-specific interaction with CTCF is guided by DNA methylation (Bell and Felsenfeld 2000; Hark *et al,* 2000; Kanduri *et al,* 2000).

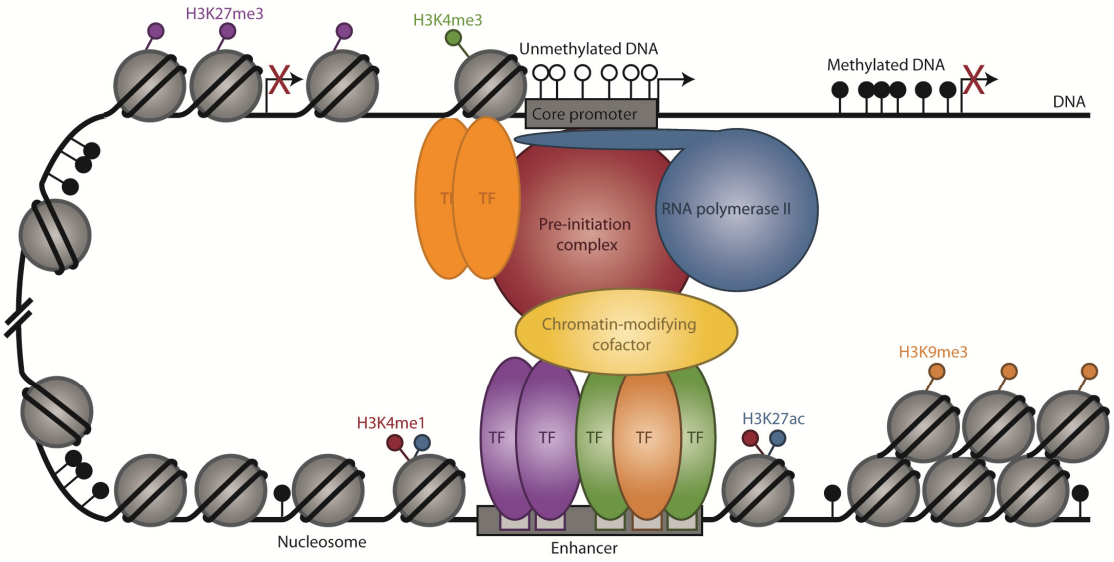

Figure 2-1 Model of transcriptional regulation. Adapted from (Bardet 2012).

In summary, concerted gene regulation in animals can be seen as a complex interplay between *cis*- and *trans*-acting factors, epigenetic modifications and higher-order chromatin structures, such as looping or locus position within the nucleus. Despite growing data accumulation, limited knowledge exists about exact causality. Do chromatin modifications influence transcriptional activity or are they just a footprint of preceding TF binding events? Regulatory function implies high stability and accurate inheritance of epigenetic marks. Yet, both phenomena are not well understood.

In the present work I examine the establishment and maintenance of DNA methylation patterns at the sites of transcription factor binding. Following paragraphs will introduce transcriptional regulatory principles relevant to my work with the main focus on DNA methylation.

2.2 Chromatin

In eukaryotes DNA is packaged to chromatin, a unifying term for DNA and all accessory proteins. Chromatin can be classified in two groups as defined in 1928 by Emil Heitz based on their staining characteristics (reviewed in (Zacharias 1995; Elgin 1996)). Euchromatin, the “proper” chromatin, decondensates during the interphase and thus behaves according to the original definition of chromatin made by Boveri in 1904 as a substance which forms chromosomes during mitosis (cited in (Zacharias 1995)). In contrast, following the original definition heterochromatic regions remain highly condensed and stained in the interphase. Today this definition has been refined and we know that repetitive sequences and transposable elements belong to heterochromatin even though for example telomeres do not show the characteristic interphase staining and condensation (Elgin 1996).

The two main chromatin states additionally differ in terms of their chromatin modifications, their gene density and their bound proteins. One such characteristic is heterochromatic protein HP1 which was originally identified as a protein abundantly present at pericentromeric heterochromatin (James *et al,* 1989). Furthermore, heterochromatin contains high levels of histone H3 lysine 9 and DNA methylation (Grewal and Rice 2004; Trojer and Reinberg 2007). As indicated by its condensed structure, this chromatin form must be rather inaccessible for transcription factors. Indeed, two key observations linked heterochromatin with gene silencing: the condensed structure of the inactivated X-chromosome (Barr body) and silencing of active genes in its vicinity, termed position effect variegation (reviewed in (Elgin 1996)). Based on the reversibility of the heterochromatic state heterochromatin is furthermore often subdivided into constitutive and facultative heterochromatin. Constitutive heterochromatin comprises repeat sequences and transposons which are obligatory silenced, whereas facultative heterochromatin contains genes (Trojer and Reinberg 2007). In contrast, euchromatic regions have higher accessibility and are enriched for active chromatin modifications (see 2.3). In the recent years, more detailed classifications have been proposed based on mapping of bound proteins and chromatin modifications (Filion *et al,* 2010; Kharchenko *et al,* 2011).

In the following I will introduce the components and organization of chromatin together with their role in gene regulation.

2.2.1 Nucleosomes

The nucleosome is the basic packaging unit of chromatin. This core subunit is formed when 146 basepairs of DNA are wrapped around an octamer of histone proteins in a 1.65 turn (Luger *et al,* 1997). Contacts between negatively charged DNA and the basic histone proteins are stabilized through a number of electrostatic interactions, predominantly at the phosphodiester backbone. Importantly, due to the helix turn such interactions can only occur approximately every 10 base pairs. Histone octamers contain pairs of each of the histone proteins H2A, H2B, H3 and H4 which are placed by histone chaperones. These assemble a tetramer of $(H3\text{-}H4)_2$ with two heterodimers of H2A-H2B (for review see (De Koning *et al,* 2007)). The canonical histone variants can alternatively also be replaced by relatively rare histone variants. For instance, histone H3.3 is deposited at active genes and was even proposed to transmit active epigenetic states (Ahmad and Henikoff 2002; Ng and Gurdon 2008b). Nucleosome core particles are connected by linker DNA (about 50-60bp in mammals) bound by a structural histone protein H1 that helps folding of nucleosome repeats to higher-order chromatin structures (Luger 2003).

Beyond packaging of DNA, nucleosomes generally render the chromatin less permissive, so that their depletion can result in a 10-20 fold increased accessibility of DNA binding factors (Liu *et al,* 2006). Notably, recruitment of the transcriptional machinery by TATA-box binding protein as well as the binding of the general transcription factor TFIIIC requires a nucleosome-free environment (Workman and Kingston 1998; Bartke *et al,* 2010). This directly illustrates the inhibitory impact of nucleosomes on transcription initiation. It is clear, however, that binding ability of nucleosomal templates differs between TFs and not all are repulsed by the presence of nucleosomes (Taylor *et al,* 1991). Availability of a partner TF can increase the potential to access nucleosomal DNA, as in electrophoretic mobility shift assays some factors can only cooperatively achieve efficient binding to nucleosomal arrays (Adams and Workman 1995). On the other hand, even TFs which can bind to inaccessible chromatin on their own seem to prefer preexisting accessible sites (John *et al,* 2011). TFs without an inherent nucleosome binding capability can gain access to their cognate motifs with the help of “pioneer factors”, such as GATA-4 and FoxA1. These factors access nucleosome occupied DNA and can induce chromatin remodeling, thereby opening the chromatin for other proteins (Cirillo *et al,* 2002; Zaret and Carroll 2011). Finally, a group of transcription factors, including the pluripotency factors Oct4 or Nanog, seems to bind to nucleosome-occupied regions without displacing the nucleosome (Teif *et al,* 2012).

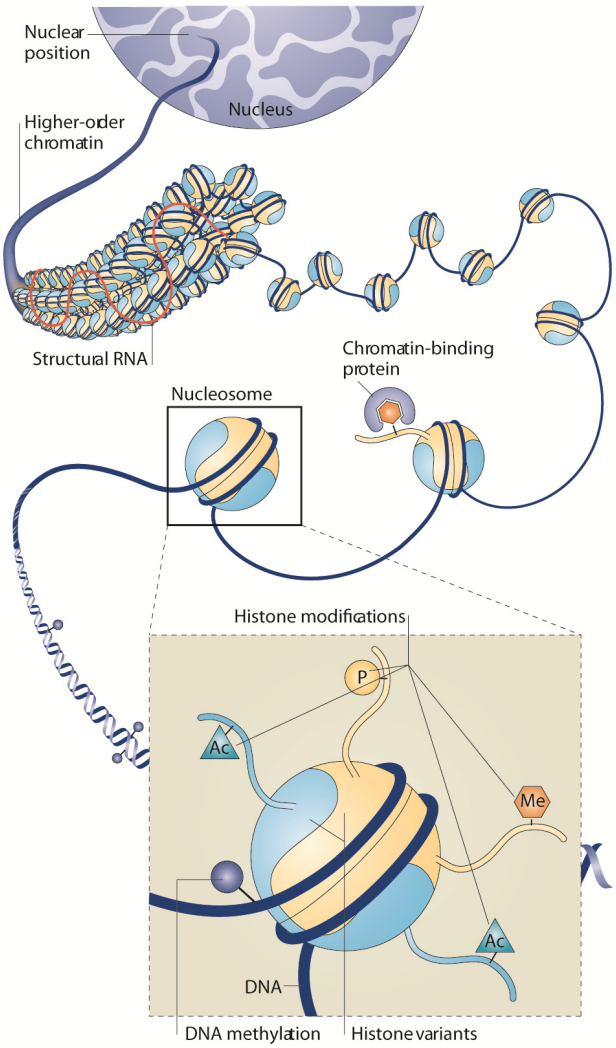

Figure 2-2 Chromatin organization. Adapted from (Probst *et al,* 2009)

2.2.2 Defining nucleosome positions

In line with their role in gene regulation, nucleosomes are not randomly distributed throughout the genome, but instead well positioned at active regulatory regions (Yuan *et al,* 2005; Schones *et al,* 2008). Active promoters are characterized by a nucleosome-free region (NFR) in the vicinity of their TSS in all analyzed eukaryotes. Its size and the extent of nucleosome depletion both correlate with transcriptional output (Teif *et al,* 2012). CpG rich promoters are exceptional, as they possess an NFR even when inactive (Teif *et al,* 2012). The first genome-wide nucleosome mapping was carried out in yeast (Yuan *et al,*

2005). In this model organism active promoter NFRs are typically flanked by strongly positioned nucleosomes at positions +1 and -1 and at least four additional well-positioned nucleosomes downstream of the active TSS. The authors report a high conservation of nucleosome-free regions, suggesting that nucleosome depletion is a general feature of regulatory elements. Indeed, measurements of genome accessibility by DNAseI digestion suggest a high overlap of open chromatin regions with regulatory sites (Song L. *et al,* 2011). Recent accessibility and nucleosome positioning studies demonstrate that open chromatin at a given enhancer can occur independently of the presence of individual factors, even master regulators of cell fate (Teif *et al,* 2012; McKay and Lieb 2013). Such observations open up the possibility that distal regulatory elements can be “recycled” for usage by different TFs.

Nucleosome positioning is nicely illustrated at insulator sequences bound by CTCF. A combinatorial profile shows that these are surrounded by 20 well-positioned nucleosomes (Fu *et al,* 2008). Importantly, in the absence of CTCF these sites tend to be occupied by a nucleosome, arguing that strong positioning is not encoded within the DNA sequence. Nucleosome phasing at these sites is furthermore recapitulated by a number of histone modifications (see 2.3.1) as well as by DNA methylation (Stadler *et al,* 2011; Kelly *et al,* 2012) (see 2.4).

Several lines of evidence suggest a role of DNA sequence in the positioning of nucleosomes (Struhl and Segal 2013). First, the repetitive nature of nucleosomes excludes contributions of highly specific sequences and thus a favoring sequence would simply be one that favors bending of DNA. From analysis of genomic DNA from chicken erythrocytes, it has been suggested that this is given in case of a ten-base periodic occurrence of AT, allowing helical bending around the histones (Satchwell *et al,* 1986). Indeed, the prevalence of such periodicity is increased at well-positioned nucleosomes in the yeast genome (Ioshikhes *et al,* 2006). Second, long stretches of dA:dT or dG:dC polymers, such as those present at many eukaryotic promoters, disfavor bending of DNA, thus providing a plausible explanation for nucleosome depletion at the TSS (Simpson and Shindo 1979). This does not seem a universal principle, as several yeast strains are capable of establishing promoter NFRs without having enrichments of monopolymer stretches (Tsankov *et al,* 2010). Third, *in vitro* reconstituted nucleosomes overall manage to recapitulate *in vivo* positioning of the yeast chromatin (Kaplan *et al,* 2009; Zhang *et al,* 2009).

While *in vitro* reconstitution experiments allow for a recapitulation of nucleosome-depleted sites, positioning of the nucleosomes at positions +1 and -1 can only be

achieved upon addition of ATP and a crude cell extract (Zhang *et al,* 2011). Thus, precise positioning of the TSS flanking nucleosomes cannot be solely guided by the DNA sequence. At the same time, exact positions of all nucleosomes are not reproduced in any of the *in vitro* assays (Zhang *et al,* 2011). In summary, these experiments argue that DNA sequence in cooperation with ATP dependent and independent factors present in the cell extract determines the positioning of nucleosomes around NFRs.

Nucleosomes are indeed reordered by specialized ATPases. (Clapier and Cairns 2009; Mueller-Planitz *et al,* 2013). Currently, four families of these nucleosome remodelers are known: SWI/SNF, ISWI, CHD and INO80 which are conserved between eukaryotes from yeast to humans. These families have different functions in assembly or disassembly of entire nucleosomes, whereas all of them seem to be capable of nucleosome dislocation. Mechanistically, this can be achieved by a localized destabilization of histone-DNA interactions as well as by destabilization of the DNA or of the histone octamer (reviewed in (Mueller-Planitz *et al,* 2013)). Furthermore, remodelers can display specialized functions in promoter activation and repression. This is exemplified in yeast, where the RSC remodeling complex is involved in nucleosome removal from promoter regions (Badis *et al,* 2008; Wippo *et al,* 2011). Conversely, the activity of Isw2 is needed to occlude nucleosome disfavoring promoter regions (Whitehouse and Tsukiyama 2006). Recent mapping of nucleosome remodelers in mouse cells suggests a highly overlapping synergistic as well as antagonistic function between different complexes (Morris *et al,* 2014). Targeting of remodelers is possible through specific recognition sequences (Badis *et al,* 2008), binding to nucleosomes and histone modifications or recruitment by specific TFs. As an example, at yeast HO promoter the SWI/SNF complex is recruited upon binding of the transcription factor Swi5p. This enables histone acetylation through the SAGA complex and ultimately binding of SBF (Cosma *et al,* 1999).

It is not entirely clear, how positioning is regulated within the gene bodies. It has been speculated that elongating RNA polymerase II has a function in the positioning of nucleosome arrays downstream of the TSS (Struhl and Segal 2013). In support of this hypothesis, nucleosome remodelers bind to coding regions (Morris *et al,* 2014). It is furthermore conceivable that a histone passback mechanism during active transcription affects nucleosome organization (Radman-Livaja *et al,* 2011).

In the last years it became clear that nucleosomes themselves are not stable structures, but instead underlie a dynamic turnover (Ahmad and Henikoff 2002). Replication independent turnover was studied in G1 arrested yeast by competition between

constitutively expressed and inducible histones (Dion *et al,* 2007; Jamai *et al,* 2007; Rufiange *et al,* 2007). These experiments suggested a turnover of histone H3, H4 and H2B at promoters. High histone dynamics at sites containing epigenetic modifications was later indentified and quantified by a technique utilizing metabolic posttranslational histone labeling (Deal *et al,* 2010).

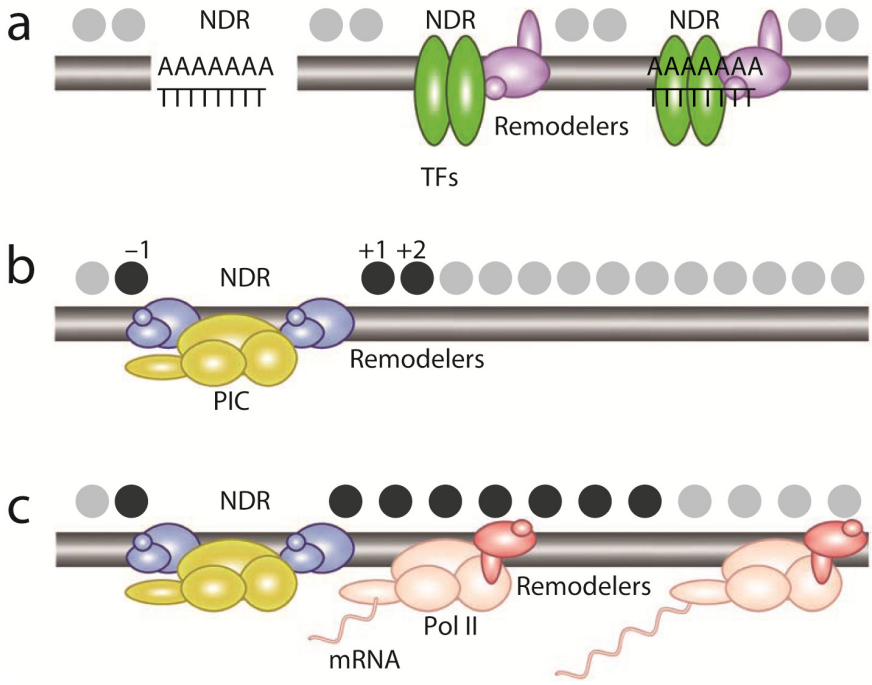

Figure 2-3 Model of nucleosome positioning. Adapted from (Struhl and Segal 2013)

Note that the here suggested model is based on yeast data. (A) Nucleosome depleted regions (NDR) are determined either by sequence polymer stretches and/or by transcription factors and recruited remodelers. Gray circles: nucleosomes (B) Preferred positions (black circles) of NDR-flanking nucleosomes are regulated by remodelers and pre-initiation complex (PIC). (C) Positioning of nucleosomes downstream of the TSS depends on elongation by RNA polymerase II (PolII) and elongation-associated remodeler complexes.

Nucleosome positioning and stability have been linked to chromatin modifications by several studies. It has been reported that histone acetylation (see 2.3) is a positive regulator of turnover (Rufiange *et al,* 2007; Verzijlbergen *et al,* 2011). Furthermore, DNA methylation has been implicated in stabilizing histone-DNA interactions (Collings *et al,* 2013) However, a simultaneous genome-wide mapping of nucleosome positioning and DNA methylation showed an anticorrelation of DNA methylation and nucleosome occupancy around CTCF sites (Kelly *et al,* 2012).

2.3 Chromatin modifications

Chromatin modifications are chemical groups covalently coupled to either DNA or histone proteins, the “epigenetic modifications”. In the original definition by Waddington 1942, the term “epigenetics” was applied to mechanisms which generate a phenotype from a genotype (republished in (Waddington 2012)). Later, this definition was extended to heritable changes in gene expression not involving alterations of the genomic sequence (Allis 2007).

Indeed, epigenetic modifications might fulfill these criteria as they possess gene-regulatory potential either directly by controlling the accessibility of chromatin through electric charges or indirectly by recruiting additional “readers” of the respective modification (Bannister and Kouzarides 2011). Mechanisms of inheritance and self-propagation have been proposed for DNA methylation, repressive and histone variant coupled active modifications (Okano *et al,* 1998; Ng and Gurdon 2008a; Probst *et al,* 2009). However, a mechanism for replicative transmission has not been described for all chromatin modifications. Transgenerational inheritance is even more questionable, since epigenetic marks can be removed during gametogenesis and development (Reik 2007).

These limitations are corrected in a third definition of epigenetics, made by Adrian Bird (Bird 2007). He proposed epigenetics to be “the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states”. Since this definition includes all chromatin-based processes it appears as the most unifying and contemporary compared to the definitions discussed above.

In the last two parts of the introduction I will discuss epigenetic modifications in light of their regulatory potential and stability.

2.3.1 Histone modifications

Histones can bear various posttranslational modifications (PTMs) either within their globular domains or at their N-terminal tails (Izzo and Schneider 2010). Structurally, histone tails are protruding from the octamer, suggesting that their modifications may have an increased potential to contribute to the overall stability of a nucleosome (Luger *et al,* 1997). Such modifications can be acetylation, methylation, phosphorylation, ubiquitinylation and ADP-ribosylation, with activating or repressive functions (Bannister and Kouzarides 2011). Combinatorial spatial or temporal activity of distinct modifications has been proposed to form a “histone code” (Strahl and Allis 2000).

The oldest example of how histone modifications can influence gene expression comes from a pioneering study demonstrating the posttranslational nature of acetylation and methylation of histone residues (Allfrey *et al,* 1964). Based on *in vitro* transcription in the presence of acetylated histones, Allfrey et al. show that the inhibitory effect of histones on RNA synthesis is decreased upon their acetylation despite preserved DNA binding capacity. They suggested that positively charged acetylated lysines partially disrupt histone-DNA interactions. In agreement with this hypothesis, lysine acetylation overlaps with active gene regulatory elements (Heintzman *et al,* 2007; Heintzman *et al,* 2009). Furthermore, histone acetyltransferases, such as the yeast protein Gcn5, have been linked to gene activation (Brownell and Allis 1996; Brownell *et al,* 1996). Consequently, inhibition of histone deacetylases enhances somatic cell reprogramming by a factor of 1000 (Huangfu *et al,* 2008). Histone acetylation can furthermore disrupt higher-order chromatin structure as has been demonstrated *in vitro* for nucleosomal arrays containing lysine 16 acetylation of histone H4 (Shogren-Knaak *et al,* 2006).

Contrary to acetylation, methylation (usually occurring at lysines or arginines) does not affect the charge of the histone proteins and thus can be activating or repressive dependent on the context (Bannister and Kouzarides 2011). A canonical example for a methylated residue associated with active state is lysine 4 of histone H3 (H3K4). Trimethylation typically occurs at the promoters of all active genes in a well-conserved manner among eukaryotes (Heintzman *et al,* 2007; Heintzman *et al,* 2009). Several chromatin remodeling complexes and histone acetyltransferases can read H3K4 methylation. Together with the evidence that the general transcription factor TFIID can bind H3K4me3 through its PHD domain this suggests a direct involvement of this mark in regulating transcriptional initiation at promoters (Santos-Rosa *et al,* 2003; Taverna *et al,* 2006; Vermeulen *et al,* 2007). There is furthermore experimental evidence that transcriptional activity directly affects the level of H3K4me3 at promoters in yeast, where the H3K4 methyltransferase SET1 is recruited by elongating RNA polymerase II (Krogan *et al,* 2003; Ng *et al,* 2003). In addition to active promoters, H3K4me3 decorates all CpG island promoters (see 2.4) independently of their activity but only in the absence of DNA methylation. This suggests a general recruitment to CpG-rich regions which might be achieved through binding of the zinc finger CXXC domain containing protein CFP1. CFP1 coexists in a complex with SETD1 H3K4 methyltransferase and was demonstrated to recruit H3K4me3 to an exogenous CpG island independently of transcription (Thomson *et al,* 2010).

In contrast to H3K4me3, polycomb mediated lysine 27 trimethylation of histone H3 (H3K27me3) has been widely associated with repression of developmental genes (Ringrose and Paro 2004; Mohn *et al,* 2008). Similarly to H3K4me3, H3K27me3 is enriched at CpG islands and its targeting to CpG-rich regions was proposed to depend on transcriptional inactivity (Mendenhall *et al,* 2010; Lynch *et al,* 2012). In *Drosophila*, polycomb targeting has been well described and is determined by sequences called polycomb response elements (PREs). In mammals several different mechanisms have been proposed, including recruitment by long noncoding RNAs or transcription factors (Ringrose and Paro 2004; Tsai *et al,* 2010; Arnold *et al,* 2013).

Genome-wide mapping of histone modifications suggests their highly characteristic distribution at *cis*-regulatory elements (Heintzman *et al,* 2007; Heintzman *et al,* 2009). In human cell lines active promoters are invariantly marked by H3K4me3 and histone acetylation and these marks are conserved across cell types. Enhancer landscape is more dynamic with high ratio of H3K4me1:H3K4me3 and H3K27 acetylation as the most characteristic signatures of activity. Importantly, these landscapes are so specific that they can be used for *de novo* prediction of enhancers (Heintzman *et al,* 2007).

Besides the previously mentioned direct influence on chromatin structure, histone modifications can affect binding of effector proteins (Bartke *et al,* 2010). These can be sequence-specific transcription factors or unspecific binders, such as chromatin remodeling complexes.

Influence of histone modifications on gene regulation at enhancers can be exemplified by the pioneer transcription factor FoxA1 (Lupien *et al,* 2008). Genome-wide mapping of FoxA1 binding sites in two human cancer cell lines shows a correlation with H3K4me1 and H3K4me2. Importantly, depletion of H3K4 dimethylation by overexpression of the histone methyltransferases LSD1 disrupts FoxA1 binding, suggesting a role of H3K4 methylation in FoxA1 recruitment. At the same time, H3K4 methylation does not seem to be the sole determinant of FoxA1 binding, as FoxA1 also occupies sites lacking H3K4me1/2 when overexpressed in a cancer cell line (Serandour *et al,* 2011).

Impact of TF binding on epigenetic landscapes can be illustrated in macrophage differentiation. In this system Egr-2 reduces H3K4me3 levels at the miR-17-92 promoter through recruitment of the demethylase Jarid1b (Pospisil *et al,* 2011).

Today it is evident that histone modifications regulate gene expression as part of highly coordinated events, where multiple effectors and pathways act in close collaboration to provide functional output (Cosma *et al,* 1999; Li *et al,* 2010).

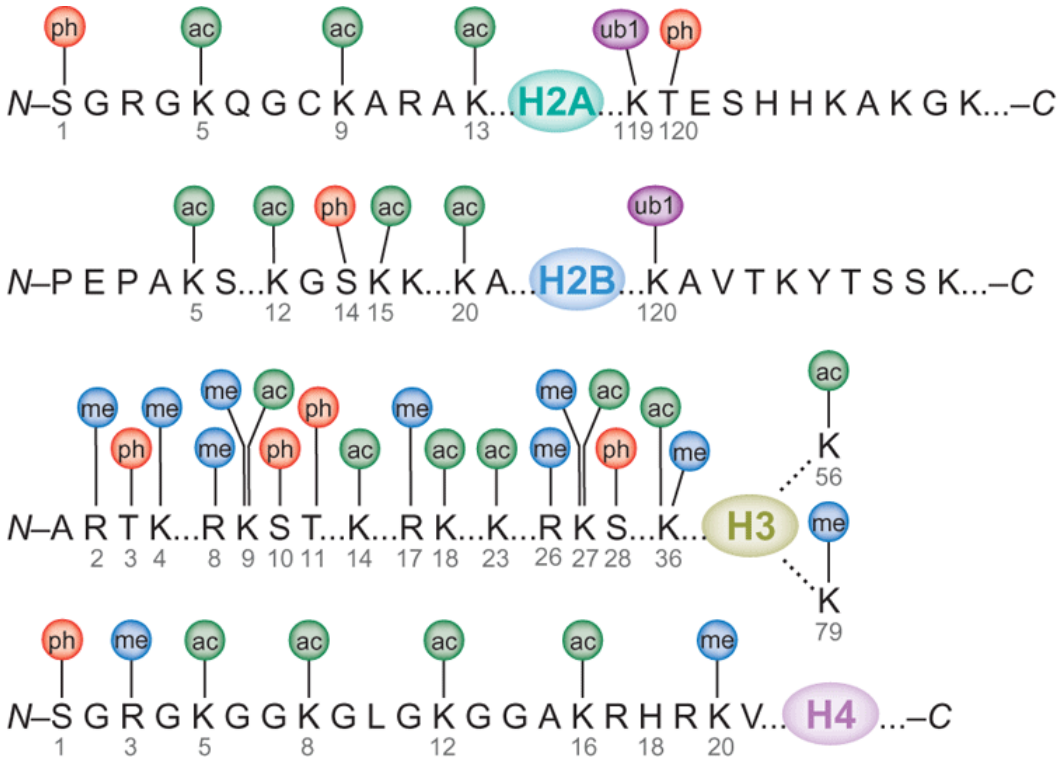

Figure 2-4 Posttranslational histone modifications. Adapted from (Bhaumik *et al,* 2007).

2.4 DNA methylation

Together with its derivatives DNA methylation is the only known covalent modification of DNA. Participation of DNA methylation in inheritance of epigenetic states as well as its role in repression of genes has been proposed as early as in 1975 (Holliday and Pugh 1975; Riggs 1975). Since then it became increasingly appreciated that DNA methylation correlates with gene repression, although a causal role appears to depend on the genomic context (Baubec and Schübeler, in press). More accepted is the function of DNA methylation in such fundamental processes like X-chromosome inactivation, imprinting and repeat silencing (Goll and Bestor 2005). It is furthermore essential for differentiation and development and consequently abnormal methylation patterns can be found in cancer and disease (Shirohzu *et al,* 2002; Plass *et al,* 2013).

In the following paragraphs, I will summarize current knowledge about this epigenetic mark with an emphasis on stability and regulatory function.

2.4.1 Evolution of DNA methylation

DNA methylation has been proposed to originate from the primitive prokaryotic immune system (Bestor 1990). Indeed, bacteria use methylation of adenine or cytosine in host defense for selective destruction of exogenous bacteriophage DNA (Goll and Bestor 2005). Such defense mechanism, however, has never been reported in eukaryotes.

While the genome of some lower eukaryotes contains methylated adenines (Gorovsky *et al,* 1973), in higher eukaryotes DNA methylation occurs exclusively at the fifth carbon of cytosines (5-methylcytosine, 5mC) (Wyatt 1951; Bird and Wolffe 1999; Goll and Bestor 2005). This modification is common to organisms with large genomes, suggesting that DNA methylation evolved as an additional regulatory layer to compensate for the increased genomic complexity (Bestor 1990). Thus, necessity to silence transposons combined with sexual reproduction has been proposed as the major driving force for the evolution of DNA methylation (Zemach and Zilberman 2010).

In the fungus *Neurospora crassa*, DNA methylation almost exclusively affects relics of transposons which were subject to repeat-induced point mutations as part of a genome defense mechanism (Selker *et al,* 2003). While this pattern is conserved among other fungal species with DNA methylation, the sequence context of methylated cytosines can vary (Zemach *et al,* 2010). Importantly, the most widely studied model fungi *Saccharomyces cerevisiae* and *Saccharomyces pombe* both lack DNA methylation.

In contrast to fungi, methylation in the plant *Arabidopsis thaliana* localizes not only to transposons but also to gene bodies and repetitive elements (Zhang *et al,* 2006). In this species cytosine methylation occurs in the context of CHH, CHG and CG sequences with an overall relatively small fraction (up to 25%) of methylated cytosines (Furner and Matzke 2011). Importantly, transposon and repeat upregulation in DNA methylation mutants suggest that their silencing might indeed be the key role of DNA methylation in plants (Zilberman *et al,* 2007; Lister *et al,* 2008; Tsukahara *et al,* 2009).

While DNA methylation is common to all vertebrate genomes, its prevalence is limited among non-vertebrate animals. The genomes of *Drosophila melanogaster* and *Caenorhabditis elegans* both are devoid of DNA methylation. Non-vertebrate animals with DNA methylation – such as the honeybee *Apis mellifera* – show incomplete mosaic or “fractional” methylation with moderate methylation levels, mostly within gene bodies (Zemach *et al,* 2010; Deaton and Bird 2011). The transition to vertebrate lineages was accompanied by the acquisition of genome-wide DNA methylation (Tweedie *et al,* 1997; Deaton and Bird 2011). Here, methyl groups predominantly occur in the context of CpG dinucleotides, although rare cases of cytosine methylation in a non-CpG context have been reported in stem cells and brain tissue (Ramsahoye *et al,* 2000; Lister *et al,* 2009; Lister *et al,* 2013). Genome-wide DNA methylation brought up the hypothesis that vertebrate genomes are methylated by default and targeted demethylation is the key regulatory mechanism (Bird and Wolffe 1999). In summary, methylation differs between eukaryotic lineages in terms of preferences for sequence contexts and spatial distribution.

In their comparative study of methylomes from a variety of species Zemach et al. suggested that the last common ancestor of plants, fungi and vertebrates possessed all tools of the DNA methylation machinery (Zemach *et al,* 2010). If this hypothesis is true, then DNA methylation has been lost in several lineages, such as in *D. melanogaster* and *C. elegans* (Dean *et al,* 2001; Suzuki and Bird 2008). This could be enabled partially due to the lack of selective pressure and partially because other compensatory mechanisms have evolved (Brennecke *et al,* 2007). Loss of DNA methylation in some lineages argues that it might have opposing effects on the fitness of an organism (Hollister and Gaut 2009; Zemach *et al,* 2010). Genome-wide DNA methylation in vertebrates might indeed have evolved as a silencing mechanism for transposons and repeats. Following this, their occasional insertion in introns could have spread the methylation over the coding gene regions (Jahner and Jaenisch 1985). In this case, the benefits of preserving the

genome's stability by transposon inactivation must have outperformed the potential disadvantages of aberrant gene silencing.

2.4.2 DNA methylation patterns in vertebrates

Genome-wide methylation patterns in vertebrates are mainly established during gametogenesis and postimplantation development following global demethylation (Morgan *et al,* 2005; Borgel *et al,* 2010). During gametogenesis parental imprint methylation is reestablished which later resists global demethylation upon fertilization of the oocyte (Tucker *et al,* 1996; Dean *et al,* 2001). Further programmed methylation takes place during development and differentiation and affects somatic imprints and gene promoters (Mohn *et al,* 2008; Borgel *et al,* 2010).

In general, the majority of cytosines in the context of CpG dinucleotides (about 80%) are methylated, thus allowing DNA methylation to extend its function beyond silencing of transposons (Lister *et al,* 2009; Zemach *et al,* 2010; Deaton and Bird 2011; Long *et al,* 2013b). Since methylated cytosines are predisposed to deamination, germline mutations render mammalian genomes globally depleted in CpGs (Coulondre *et al,* 1978; Bird 1980; Schorderet and Gartler 1992; Freitag *et al,* 2002; Zemach *et al,* 2010). A notable exception are CpG islands (CGI) with a locally high concentration of CpG dinucleotides (ration observed/expected >0.5 dependent on the algorithm) (Bird *et al,* 1985). About 70% of genes contain a CGI in their promoter and only about 3% of those become methylated in adult tissues (Deaton and Bird 2011; Long *et al,* 2013b). Thus, while methylation of CpG island promoters usually coincides with gene silencing, most of them remain unmethylated but inactive (Stein *et al,* 1982; Schilling and Rehli 2007; Shen *et al,* 2007; Weber *et al,* 2007; Mohn *et al,* 2008; Payer and Lee 2008). Promoters are furthermore frequently silenced by other epigenetic marks before acquiring DNA methylation (Feldman *et al,* 2006). Taken together these observations led to the speculation that DNA methylation has a role in "locking-in" the repressive state of genes and thus is required wherever stable silencing is needed.

A variety of methods have been developed for the analysis of unmethylated or methylated DNA. These include endonuclease digestion by enzymes with different sensitivity towards DNA methylation (Bird and Southern 1978), affinity purification of methylated or unmethylated DNA (Cross *et al,* 1994; Weber *et al,* 2005; Blackledge *et al,* 2012) and conversion of unmethylated cytosines to uracil by bisulfite treatment (Wang *et al,* 1980; Frommer *et al,* 1992). The latter method allows for a single base resolution

global methylation analysis when coupled with genome-wide sequencing technologies and was recently applied by several groups to mammalian genomes (Lister *et al,* 2009; Hodges *et al,* 2011; Stadler *et al,* 2011; Xie *et al,* 2013; Ziller *et al,* 2013). These studies revealed that the genome-wide methylation landscape is characterized by segment-specific methylation signatures with more or less uniformly methylated blocks. In agreement with previous observations, the majority of cytosines indeed fall into fully methylated or unmethylated regions (FMRs and UMRs, respectively). In this classification UMRs largely correspond to unmethylated CpG islands (Lister *et al,* 2009; Stadler *et al,* 2011).

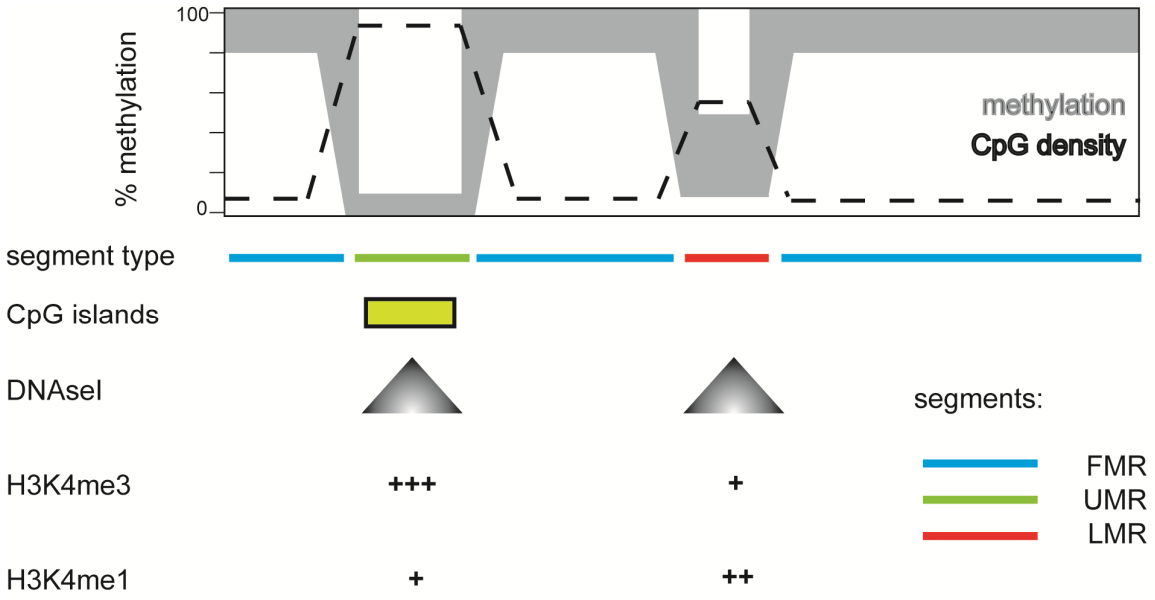

Figure 2-5 Schematic representation of DNA methylation landscape in vertebrates.

Three major segment types are shown as defined in Stadler et al. (Stadler *et al,* 2011): Fully methylated regions (FMR, blue) with mostly 80-100% methylated CpGs, unmethylated regions (UMR, green) with methylation ranging from 0 to 10% and low methylated regions (LMR, red) with about 10-50% methylation. Grey shadows represent the variability in methylation between individual cytosines. CpG density (black, dashed line) is elevated at UMRs which mostly correspond to CpG islands and to a lower extent at LMRs. DNAseI cuts in UMRs and LMRs, indicative of an open chromatin state and transcription factor presence within these regions. LMRs are marked by low H3K4me3 and high H3K4me1, characteristic for enhancers.

However, in addition to these previously anticipated domains, a new class of low CpG-density segments has been identified which is hypomethylated but not fully unmethylated like CpG islands and therefore termed low methylated regions or LMRs (Stadler *et al,* 2011). Average methylation within these regions comprises 30%, although the methylation state of individual cytosines can vary throughout the domain. Importantly, LMRs possess all features of active distal regulatory elements with enrichments of cell

type specific DNA-binding factors and enhancer-characteristic histone marks. Furthermore, insertion of unmethylated or *in vitro* methylated DNA fragments at an ectopic locus in mouse embryonic stem cells showed that binding of the factor CTCF is necessary and sufficient for creating a hypomethylated state. This observation argues against an instructive role of DNA methylation for transcription factor recruitment to these regions. Importantly, methylation at LMRs changes dynamically during neuronal differentiation correlating with changes in the expression of cell type specific TFs and with active enhancer signatures. The presence of low methylated regions has been described in many cell types, confirming these initial findings in embryonic stem cells (Hodges *et al,* 2011; Stadler *et al,* 2011; Burger *et al,* 2013; Hon *et al,* 2013; Xie *et al,* 2013; Jeong *et al,* 2014).

More recent global studies on DNA methylation in human ES cell differentiation and hematopoietic stem cells reported presence of methylation depleted regions exceeding the size of LMRs or UMRs (median in ES cells 324bp and 581bp, respectively) by several fold (median size: >5kb). These regions were termed methylation “valleys” or “canyons” (Xie *et al,* 2013; Jeong *et al,* 2014). Besides the difference in size they largely resemble CpG islands, cover conserved motif-rich sequences and are conserved between cell lineages. Interestingly, loss of the DNA methyltransferases DNMT3A mediated DNA methylation increases the size of canyons, suggesting a role of DNA methylation turnover in the maintenance of border methylation at these regions (Jeong *et al,* 2014). Regulatory potential of unmethylated and hypomethylated sequences is further demonstrated by a recent study showing evolutionary conservation of hypomethylated regions between vertebrate species (Long *et al,* 2013a).

Beyond these segments with well-defined methylation states, large partially methylated domains (PMD) with apparently deregulated variable methylation are detectable in some mammalian cell types (Lister *et al,* 2009; Gaidatzis *et al,* 2014).

2.4.3 DNA methylation machinery

Methyl groups are added to cytosines in an S-Adenosyl-Methionine (SAM) dependent reaction by a conserved group of enzymes called DNA methyltransferases (DNMTs). In their speculative pioneer work Holliday and Pugh predicted the presence of two enzyme activities to maintain a cell’s methylation levels: a *de novo* methyltransferase and a maintenance methyltransferase (Holliday and Pugh 1975). The maintenance enzyme was predicted to share the capacity of bacterial methyltransferases to act only on

hemimethylated DNA and thereby copy methylation to the palindromic CpG sequence of the daughter cell. This traditional separation into maintenance and *de novo* methylating enzymes remains in use today.

As originally postulated, the maintenance DNA methyltransferase DNMT1 prefers hemimethylated DNA as substrate (Stein *et al,* 1982; Bestor *et al,* 1988; Okano *et al,* 1998). Initial *in vitro* methylation assays (Okano *et al,* 1998) found confirmation in the structural analysis showing autoinhibition of DNMT1 catalytic center upon binding of a fully unmethylated substrate DNA (Song J. *et al,* 2011; Song *et al,* 2012). DNMT1 is recruited to the replication forks by UHRF1, a protein which interacts with PCNA during DNA synthesis (Sharif *et al,* 2007). Such maintenance provides a potential mechanism for epigenetic memory and inheritance (Holliday and Pugh 1975; Riggs 1975). Indeed, methylation at CGIs is accurately transmitted, as has been demonstrated by integration of premethylated DNA fragments (Wigler *et al,* 1981; Schubeler *et al,* 2000). Certain sequences, however, can autonomously determine their methylation state independently of preceding manipulation (Lienert *et al,* 2011; Stadler *et al,* 2011).

Lei et al. reported the first experimental evidence that mammals possess separate enzymes for the catalysis of *de novo* DNA methylation in 1996, when they noticed that proviral DNA can be *de novo* methylated in the absence of DNMT1 (Lei *et al,* 1996). The *de novo* DNA methyltransferases DNMT3A and DNMT3B were cloned and characterized a few years later, confirming that they indeed are able to methylate unmodified cytosines (Okano *et al,* 1998; Okano *et al,* 1999).

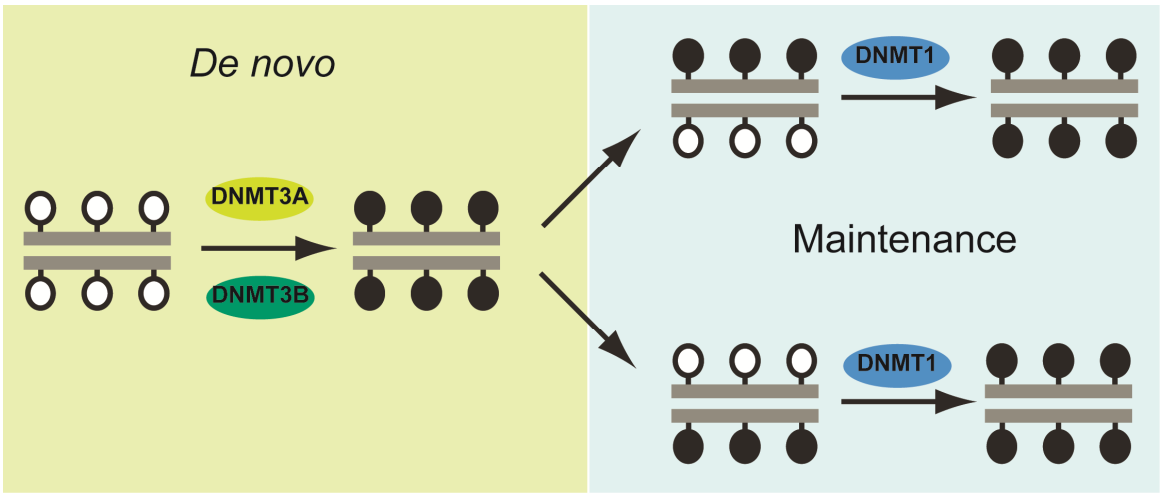

Figure 2-6 Schematic representation of *de novo* and maintenance DNA methylation
Methylated CpG: black lollipops, unmethylated CpG: white lollipops

Although common consent accepts this role distribution, it is probably not very accurate. For example, the *in vitro de novo* methylation activity of DNMT1 is five times higher than that of DNMT3A/B (Okano *et al,* 1998). Similarly, DNMT3A/B possess maintenance activity *in vitro* (Okano *et al,* 1998) and *in vivo*, as deletion of DNMT1 in embryonic stem cells does not result in complete loss of methylation (Lei *et al,* 1996; Jackson *et al,* 2004). There is furthermore evidence that *de novo* DNA methyltransferases participate in local or global methylation maintenance *in vivo* (Chen *et al,* 2003; Jackson *et al,* 2004; Arand *et al,* 2012; Jeong *et al,* 2014). Initially, nearest-neighbor analysis suggested a progressive loss of methylation in DNMT3A/B double-knockout embryonic stem cells, resulting in a global methylation decrease by 50% at passage 20 upon knockout (Jackson *et al,* 2004). Arand et al. found varying dependency of genomic regions on DNMT3A/B for methylation maintenance when they analyzed methylation of DNMT mutant mouse ES cell lines by hairpin-bisulfite PCR (Arand *et al,* 2012). For instance, methylation at Tex13, Afp, IAPs and mSat can be maintained by either enzyme, while Igf2, Snrpn, B1 and L1 repeats require cooperativity from both enzymes. Surprisingly, these characteristics differ from one region to another and no common rule distinguishes between single genes and repetitive sequences.

Analysis of DNMT3A and DNMT3B knockout mice and embryonic stem cell lines revealed target specificity of *de novo* DNA methyltransferases (Okano *et al,* 1999). For instance, centromeric minor, but not major satellite repeats are hypomethylated in both DNMT3B knockout and hypomorphic mutant mice (Okano *et al,* 1999; Velasco *et al,* 2010). How exactly DNMT3A/B are targeted to specific loci remains to be determined. Contribution of flanking sequences to targeting specificity of DNMTs has been reported (Lin *et al,* 2002; Handa and Jeltsch 2005; Gowher *et al,* 2006) as well as recruitment by transcription factors, such as E2F6 (Velasco *et al,* 2010). DNMT3A/B have been shown to anchor to methylated nucleosomes in cancer cell lines (Jeong *et al,* 2009).

DNMTs are differentially expressed throughout development and cell differentiation (Okano *et al,* 1998; Okano *et al,* 1999; La Salle *et al,* 2004; Watanabe *et al,* 2006). DNMT3B expression is widespread in the early embryo and is restricted mostly to the developing brain upon embryonic day E9.5, the time-point when DNMT3A expression becomes ubiquitous (Okano *et al,* 1999). DNMT3A functions predominantly in germ cells as the main methyltransferase in the establishment of parental imprints and is moreover upregulated later in development (Watanabe *et al,* 2006; Nguyen *et al,* 2007; Challen *et al,* 2012). Different temporal and spatial expression patterns are mirrored in the phenotypes of knockout mice (Okano *et al,* 1999). In general, knockout of any enzymatic

DNMT is lethal. However, while DNMT3B and DNMT1 knockout mice die during embryogenesis, DNMT3A knockout animals survive until 4 weeks after birth. A role of DNMT3A in postnatal cells has been described for hematopoietic stem cell differentiation upon conditional knockout (Challen *et al,* 2012). The authors used bone-marrow transplanted conditional knockout hematopoietic stem cells to track their *in vivo* differentiation potential. They found that the differentiation was compromised and the cells were biased towards the stem cell state. Unexpectedly, reduced-representation bisulfite sequencing revealed equal amounts of both demethylated and hypermethylated regions upon DNMT3A knockout and little correlation to gene expression changes.

It seems that the presence of DNMTs is overall dispensable for embryonic stem cells. ES cells lacking all three DNMTs are viable and even retain the differentiation capability to embryonic lineages as long as DNMT1 is present (Jackson *et al,* 2004; Tsumura *et al,* 2006). Survival of extraembryonic lineages, on the other hand, appears independent of DNA methylation (Sakaue *et al,* 2010).

Even though in general DNMTs are downregulated upon development and differentiation, notable amounts can still be detected in adult postmitotic cells (Goto *et al,* 1994). However, the molecular function of DNMTs in adult tissues is not well understood. For example, Nguyen et al. demonstrated neuromuscular defects in a conditional knockout of DNMT3A in mouse neurons (Nguyen *et al,* 2007). Only very limited demethylation was detected at one of the investigated gene promoters in the adult mouse brain. This subtle effect does not allow establishing a link between observed defects and the methyltransferase activity of DNMT3A. Other studies suggest a role of DNA methylation in neuronal activity, although the link between DNA methylation, gene expression and reported defects remains unexplained (Feng *et al,* 2010; LaPlant *et al,* 2010; Guo *et al,* 2011b).

Together with conditional knockout experiments, presence of the DNA methylation machinery in adult cells argues that maintaining plasticity of DNA methylation might be important beyond development and differentiation (Feng *et al,* 2010; Guo *et al,* 2011b). The task of the DNA methyltransferases in differentiated tissues could involve correction of DNA damage, maintenance of transcriptional silencing capacity or participation in turnover.

2.4.4 DNA demethylation

2.4.4.1 Passive and active demethylation

Because of its heritability, its covalent coupling to DNA and the stability of the C-C bond between the fifth carbon of the cytosine and the methyl group, DNA methylation has been considered a stable modification whose main function is to ensure long-term silencing (Wu and Zhang 2010). Insights from developmental studies, however, suggest that it is more dynamic than anticipated. In mammalian development DNA methylation is removed in two main waves: during gametogenesis and preimplantation (Mayer *et al,* 2000; Oswald *et al,* 2000; Hajkova *et al,* 2002; Lee *et al,* 2002; Santos *et al,* 2002; Yamazaki *et al,* 2003; Wu and Zhang 2010; Seisenberger *et al,* 2013). Interestingly, during zygotic demethylation maternal and paternal pronuclei seem to rely on different demethylation mechanisms. Paternal pronucleus undergoes rapid demethylation, whereas the maternal pronucleus gradually loses its methyl mark. This difference in kinetics probably exemplifies active and passive demethylation.

Passive demethylation occurs when replication dependent maintenance is compromised. This can be mediated by exclusion of DNMT1 from the nucleus, as reported for preimplantation development (Monk *et al,* 1987; Howlett and Reik 1991; Carlson *et al,* 1992), or protection of DNA from the maintenance machinery (Hsieh 1999).

On the other hand, active demethylation requires a mechanism that catalyzes removal of the methyl group (Wu and Zhang 2010). In case of zygotic development, demethylation of the paternal pronucleus coincides with oxidation of 5-methylcytosine to 5-hydroxymethylcytosine (see 2.4.4.4). During these events the maternal pronucleus is most likely protected from active demethylation by Stella/DPPA3 (Gu *et al,* 2011; Wossidlo *et al,* 2011). The possibility of active demethylation causes a lot of excitement as it would allow for increased plasticity in gene regulation (Bird 2002). Nevertheless, while the presence of active demethylation has been widely accepted, the field suffers from unclarity in mechanisms.

The main challenge in the field is distinguishing between passive and active demethylation. This can be achieved by excluding replication mediated effects as is the case for demethylation of paternal pronucleus. Furthermore, active demethylation implies selectivity, so that it should be site-specific. Multiple studies report occurrence of local demethylation during differentiation of dividing (Mohn *et al,* 2008; Stadler *et al,* 2011; Serandour *et al,* 2012) or post-mitotic (Klug *et al,* 2010) cells. Active demethylation has also been reported in postmitotic neurons upon induction of neuronal activity *in vitro* at

the Bdnf promoter (Martinowich *et al,* 2003) and *in vivo* using genome-wide mapping of methylation-sensitive restriction sites (Guo *et al,* 2011b). However, overall demethylation in these systems was low and limited to individual cytosines within a few regions.

Other reports suggest a connection between demethylation and transcriptional activity at promoters (Kangaspeska *et al,* 2008; Metivier *et al,* 2008) or enhancers (Stadler *et al,* 2011; Wiench *et al,* 2011; Shen *et al,* 2013; Song *et al,* 2013). This hypothesis points directly to the question of whether DNA methylation is instructive for gene activity at all.

Mechanisms suggested to regulate active demethylation can generally be subdivided into three different categories: direct demethylation, targeted DNA repair and oxidation mediated demethylation (Franchini *et al,* 2012). In many cases a combination of different mechanisms is possible.

2.4.4.2 Direct removal of the methyl group

Direct removal of 5-methylcytosine is considered as chemically challenging, since it would require a direct cleavage of the C-C bond between the cytosine and the methyl group (Wu and Zhang 2010; Franchini *et al,* 2012). Direct demethylation has never been reported *in vivo*. However, in an *in vitro* assay all three mammalian DNMTs have been proposed to act as direct demethylases in the presence of Ca^{2+} and absence of the methyl group donor S-Adenosyl-Methionine (Chen *et al,* 2013a).

2.4.4.3 Targeted DNA repair

DNA repair mediated active demethylation pathways involve glycosylation and deamination of 5mC followed by base-excision repair (BER).

Glycosylation mediated demethylation is indeed known from plants, where it is executed by the Demeter family of DNA glycosylases (Gehring *et al,* 2009). A mammalian glycosylase suggested to participate in active demethylation is the T DNA glycosylase TDG (Zhu *et al,* 2000). Its homologue has initially been purified from chicken embryo extracts as demethylating enzyme which prefers hemimethylated over bi-stranded methylated or unmethylated DNA as substrate (Jost 1993). Quantification of its catalytic activity revealed that it is considerably higher at mismatched G/T than at 5mC (Zhu *et al,* 2000). TDG has been recently implicated in oxidation mediated demethylation pathways (He *et al,* 2011) (see 2.4.4.4).

Deamination followed by BER appears to be a mechanism more suitable for the activity of TDG, as deamination of 5mC creates its preferred substrate, the T/G mismatch. The

main deaminases implicated in active demethylation are the activation-induced deaminase AID and the family of apolipoprotein B mRNA editing enzyme catalytic polypeptide proteins (APOBEC) (Morgan *et al,* 2004). Mice deficient for AID have mildly increased methylation in primordial germ cells (Popp *et al,* 2010). However, the difference to wild type animals is so small that AID is most likely not the major demethylase in the germline.

Both mechanisms require targeted lesion of DNA in order to act at specific sites and it has been reported that AID catalyzes site-specific demethylation (Bhutani *et al,* 2010). It is unclear, however, how such targeting might be achieved in the absence of a DNA binding domain.

2.4.4.4 Oxidation mediated demethylation

TET proteins

In the last years, most mechanisms postulated to play a role in active demethylation involve oxidation of 5-methylcytosine to 5-hydroxymethylcytosine (5hmC) by the TET family proteins (Wyatt and Cohen 1953; Kriaucionis and Heintz 2009; Tahiliani *et al,* 2009). The discovery of TET proteins was inspired by a search for mammalian homologues of J-binding proteins JBP. JBP contribute to the generation of the “base J” by oxidation of the methyl group of a modified thymine in *Trypanosoma brucei* (Tahiliani *et al,* 2009). At the moment three proteins of this family of Fe^{2+}- and 2-oxoglutarate dependent dioxygenases are known: TET1, TET2 and TET3 (Ito *et al,* 2010). All of them share a homologous C-terminal domain with a capacity to convert 5mC in 5hmC. In contrast to TET2, TET1 and TET3 both contain a CXXC domain which guides their binding to CpG islands (Xu *et al,* 2011; Xu *et al,* 2012). Thus, it is likely that different family members have distinct substrate specificities, as suggested in a recent profiling of hydroxymethylation in TET1 and TET2 knockdown embryonic stem cells (Huang *et al,* 2014).

Evidence that TET proteins are involved in active DNA demethylation first came from *in vitro* and over-expression studies (Guo *et al,* 2011a; He *et al,* 2011; Ito *et al,* 2011). Recent tethering of TET1 to a specific locus showed that its catalytic activity only leads to a minor and locally defined loss of methylation (Maeder *et al,* 2013). Simultaneous knockout of TET1 and TET2 proteins, however, increases global cytosine methylation from approximately 5.3% to 5.75% in a mass spectrometry-based quantification (Dawlaty *et al,* 2013). This finding extends to various tissues with a surprisingly low correlation

between the abundance of 5hmC and the extent of methylation gain in TET1/2 knockout. For TET3 a role in demethylation has only been reported in preimplantation and brain development, although quantitative data are still missing (Gu *et al,* 2011; Xu *et al,* 2012).

In addition to their demethylating activity, TET proteins have also been implicated in gene silencing. The first hint came from an early genome-wide map of TET1 and 5hmC in murine ES cells (Williams *et al,* 2011). The authors observed oxidation-independent gene upregulation in the absence of TET1 and subsequently linked this effect to a physical interaction with the SIN3A co-repressor complex. Another study linked hydroxymethylation with gene silencing by reporting a localization of the methylated DNA binding protein MBD3 to hydroxymethylated regions (Yildirim *et al,* 2011). However, this is unlikely a consequence of 5hmC mediated recruitment, as MBD3 localization remains unaltered in ES cells lacking all three DNMTs (Baubec *et al,* 2013).

5-Hydroxymethylcytosine

5hmC is present in many mammalian cell types with particular abundance in the brain tissues (Kriaucionis and Heintz 2009; Tahiliani *et al,* 2009; Globisch *et al,* 2010; Dawlaty *et al,* 2013). Mass spectrometry based quantifications of 5hmC revealed that its overall amount is rather low and comprises about 1-20% of 5mC, dependent on the analyzed tissue or cell types (Globisch *et al,* 2010; Ito *et al,* 2011; Dawlaty *et al,* 2013). For instance, embryonic stem cells contain 3% methylated and 0.13% hydroxymethylated cytosines which make up approximately 4.25% of methylated cytosines (Ito *et al,* 2011). On the other hand, brain cortex contains 3.1% methylated and 0.67% hydroxymethylated cytosines which comprise more than 20% of 5mC. It is possible that increased persistence of 5hmC in the brain reflects its accumulation in the absence of replication.

5hmC has been reported to be enriched in gene bodies and distal regulatory regions, however absent from CpG islands, the major sites of TET1 protein binding (Pastor *et al,* 2011; Song C.X. *et al,* 2011; Williams *et al,* 2011). Its prevalence led to the hypothetic function as a signaling module on itself, particularly in neural tissues (Mellen *et al,* 2012). This seems unlikely, as only few 5hmC-specific readers could be identified in recent proteomic analysis of hydroxymethylated baits (Iurlaro *et al,* 2013; Spruijt *et al,* 2013).

Further processing of 5hmC involves different mechanisms, often co-occurring with other demethylation pathways (Figures 2-7 and 2-8). Presence of 5hmC could facilitate passive demethylation by exclusion of DNMT1 from hemi-hydroxymethylated DNA (Valinluck and Sowers 2007). This mechanism has been reported to act in the

demethylation of paternal pronucleus (Inoue and Zhang 2011). However, UHRF1 binding at hemihydroxymethylated DNA has been suggested to be reduced (Hashimoto *et al,* 2012) but present (Frauer *et al,* 2011). Thus, DNMT1 recruitment to hemi-hydroxymethylated DNA cannot formally be excluded, suggesting possible maintenance methylation to occur in the presence of 5hmC. Further supporting this hypothesis, hydroxymethylated episomal plasmids become fully methylated upon transfection into 293T cells (Kubosaki *et al,* 2012).

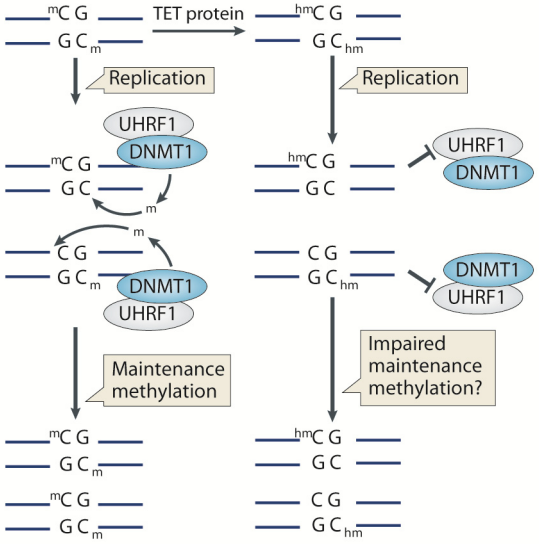

Figure 2-7 5hmC-induced passive demethylation.

Hydroxymethylation can facilitate passive demethylation through exclusion of DNMT1-UHRF1 binding from hemi-hydroxymethylated DNA. Adapted from (Pastor *et al,* 2013).

In addition, active demethylation has been proposed to be involved in removal of 5hmC. An intriguing hypothesis has been raised by an *in vitro* study, suggesting that DNMT3A and DNMT3B function as dehydroxymethylases (Chen *et al,* 2012). However, the relevance and necessity of such reaction is unclear, considering that DNMT3A/B should also be capable of direct conversion of 5mC to C according to the same research group (Chen *et al,* 2013a).

In an alternative scenario 5hmC is deaminated to 5-hydroxymethyluracil by the AID/APOBEC family of deaminases followed by DNA repair mediated excision as described above. Initial evidence for such reaction came from a study in which TET1 was overexpressed in HEK293 cells together with different enzymes of the AID/APOBEC family (Guo *et al,* 2011a). However, Nabel et al. detected only negligible amounts of

deamination products upon overexpression of AID/APOBEC along with TET2 in the same cell line (Nabel *et al,* 2012). Deamination of 5hmC by AID has furthermore been challenged by two publications where AID and APOBEC activities on differentially modified substrate DNA have been investigated *in vitro* (Nabel *et al,* 2012; Rangam *et al,* 2012). Both studies report a failure to deaminate hydroxymethylcytosine. Furthermore they show that the deaminase activity is anticorrelated to the electron cloud size of the cytosine modification. This size selectivity might exclude 5hmC from the catalytic pocket.

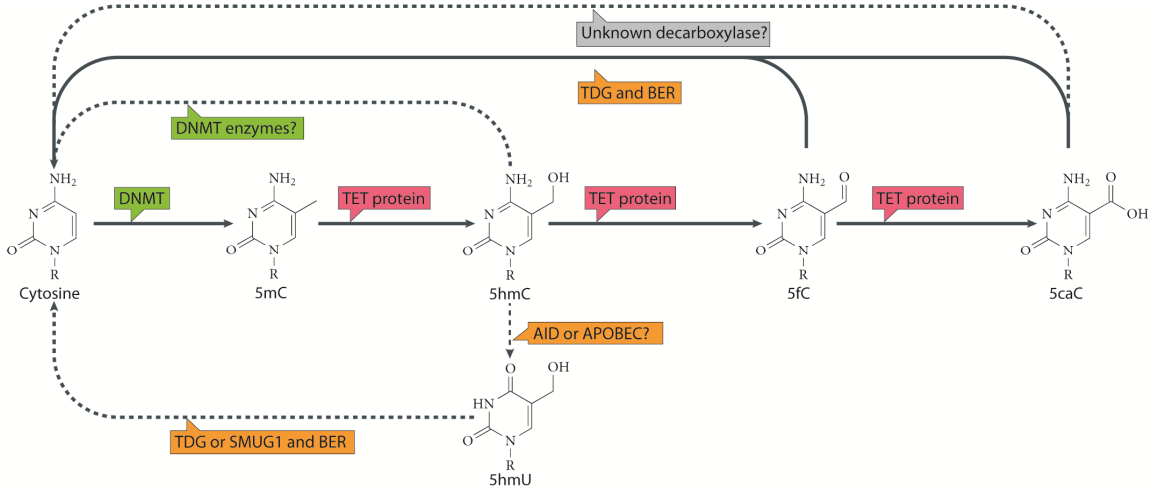

Figure 2-8 5hmC mediated active demethylation.

5hmU: 5-hydroxymethyluracil; green: DNMT, pink: TET proteins, orange: DNA repair pathways, grey: unidentified enzymatic activity. Adapted from (Pastor *et al,* 2013).

Following oxidation of 5mC, TET proteins can sequentially oxidize 5hmC to two additional products: 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Globisch *et al,* 2010; He *et al,* 2011; Ito *et al,* 2011). Enzymatic activity of TET proteins declines for later oxidation steps (Ito *et al,* 2011) which might explain the overall low abundance of 5fC and 5caC (about 1.5% and 0.2% of 5hmC in ES cells, respectively). This observation makes it less likely that oxidation is the major demethylation pathway (Globisch *et al,* 2010). Nevertheless, it has been proposed that these products are removed by DNA repair machinery with TDG as the main candidate (He *et al,* 2011; Ito *et al,* 2011). In support of this hypothesis, TDG knockout indeed displays an increase of both modifications in embryonic stem cells (Shen *et al,* 2013; Song *et al,* 2013). Alternatively, 5fC and 5caC could be directly removed by a yet unidentified enzyme present in embryonic stem cell extracts (Schiesser *et al,* 2013).

2.4.5 Regulatory potential of DNA methylation

The exact transcriptional role of DNA methylation has been discussed for many years (Holliday and Pugh 1975; Riggs 1975). A first experimental evidence for a role in gene repression came from methylation comparison within the chicken ß-globin locus between an expressing and a non-expressing cell line (McGhee and Ginder 1979). One year later Jones and Taylor showed the importance of DNA methylation for cellular differentiation by incorporation of the cytosine analogue 5-azacytidine into DNA (Jones and Taylor 1980). The first direct link between DNA methylation and gene silencing was established when retroviral genomes introduced during mouse embryogenesis were *de novo* methylated and silenced (Jahner *et al,* 1982). However, it is not entirely clear, whether methylation always has a regulatory role or whether its presence is a consequence of a lack of activation of individual genes.

The role of DNA methylation in silencing is well established for three phenomena: transposon silencing, imprinting and inactivation of genes at the X-chromosome in females. DNMT1 knockout mice display increased intracisternal A-type particle (IAP) expression and retrotransposition (Walsh *et al,* 1998; Gaudet *et al,* 2004). At the same time, monoallelic expression of the Igf-2, Igf-2r and H19 genes is misregulated in these embryos (Li *et al,* 1993). Stable silencing of one X-chromosome in female animals was demonstrated to rely on DNA methylation in embryonic, but not in extra-embryonic tissues using a DNMT1 knockout (Sado *et al,* 2000). While DNA methylation is recruited to most silenced genes on the inactivated X-chromosome this appears after initial silencing by histone modifications (Okamoto and Heard 2009). Thus, it seems likely that methylation is involved in the maintenance of X-inactivation.

2.4.5.1 Gene-specific regulation

Genome-wide DNA methylation in vertebrate genomes theoretically allows for gene-specific regulation. Consequently, it has been suggested that methylation negatively affects transcription by impeding initiation (Klose and Bird 2006). However, microarray analysis of an embryonic stem cell line in which all three DNMTs have been deleted argues for a relatively small impact of methylation on gene expression, as only about 50 genes are significantly affected (Sakaue *et al,* 2010). Furthermore, relatively few promoters significantly change their methylation status during development and differentiation (Mohn *et al,* 2008; Borgel *et al,* 2010; Challen *et al,* 2012). At least for some of them it has been reported that *de novo* methylation occurs after initial silencing by histone modifications (Feldman *et al,* 2006). Nevertheless, several studies

demonstrated that germline-specific genes rely on DNA methylation for silencing in thymus, MEFs, primordial germ cells or embryos (Borgel *et al,* 2010; Velasco *et al,* 2010; Hackett *et al,* 2012).

Regions with a high correlation between silencing and DNA methylation have an elevated CpG density, suggesting that the density of DNA methylation is more decisive for gene repression than the pure presence of the mark (Boyes and Bird 1992; Weber *et al,* 2007). This association has been proposed by an early study in which the effect of different fractions of *in vitro* methylated CpGs within an episomal plasmid on gene expression has been analyzed (Hsieh 1994). In this report a methylation of only 7% of the CpG dinucleotides decreases transcription of a luciferase reporter gene to 10%. Recently, DNA methylation density has been reported to correlate with the enrichment of methyl-CpG binding proteins which affect transcription by changing the chromatin environment (Nan *et al,* 1998; Baubec *et al,* 2013). Nevertheless, a negative correlation between activity and DNA methylation can also occur at CpG poor regions, as has been demonstrated for distal regulatory regions (Stadler *et al,* 2011).

2.4.5.2 Mechanisms of methylation mediated gene repression

An interesting hypothesis suggests that global DNA methylation generally decreases the genome accessibility and thus increases the barrier for active transcription and reduces transcriptional noise (Bird 1995). Possible silencing mechanisms involve exclusion of activators and attraction of repressors (Klose and Bird 2006). A direct inhibition can occur when a transcription factor is sensitive to CpG methylation within its binding site. However, only a few TFs have been reported to be directly repelled by DNA methylation in their motif including E2F, CREB and YY1 (Iguchi-Ariga and Schaffner 1989; Campanero *et al,* 2000; Kim *et al,* 2003; Elliott *et al,* 2010). Other factors, for example Sp1, appear insensitive (Harrington *et al,* 1988; Tate and Bird 1993). CTCF, the canonical protein for selective binding of unmethylated DNA regions, can occupy methylated CpG poor regions (Bell and Felsenfeld 2000; Stadler *et al,* 2011).

Indirect inhibition can be mediated by proteins recognizing methylated CpG dinucleotides. Two families have been described. Methyl-CpG-binding domain (MBD) proteins directly recognize methylated CpGs and occupy DNA in a methyl-CpG density dependent fashion (Tate and Bird 1993; Baubec *et al,* 2013). The second group contains proteins which recognize methylated CpGs in a sequence-specific context by their zinc-fingers (including KAISO, ZBTB4, ZBTB38 and ZFP57, (Filion *et al,* 2006)). Both could sterically counteract activator binding or recruit repressive histone modifiers. In their

pioneering work the research groups of Bird and Wolffe reported physical interaction between MeCP2 and Sin3A which in turn forms a complex with HDAC1/2. Upon binding to methylated regions MeCP2 thus triggers their deacetylation and ultimately silencing. Importantly, silencing depends on the deacetylase activity, as its inhibition by Trichostatin A derepresses reporter genes (Jones *et al,* 1998; Nan *et al,* 1998). In this case, while facilitated by DNA methylation, silencing is mediated through other factors.

2.4.5.3 Protection from DNA methylation

It is possible that gene activity and active chromatin counteract DNA methylation. For example, *in vitro* assays demonstrated that trimethylation of lysine 4 at histone 3 sterically excludes DNMT3L from CpG islands and thus protects them from recruitment of DNMT3A/B (Ooi *et al,* 2007). CpG islands can furthermore attract CXXC domain proteins which might contribute to their active chromatin state (Long *et al,* 2013b). Such a protein, TET1, which is strongly enriched at CpG islands (2.4.4.4) could actively keep them unmethylated and accessible for other factors (Pastor *et al,* 2013).

In addition, binding of certain transcription factors might protect from DNA methylation (Brandeis *et al,* 1994; Macleod *et al,* 1994; Lienert *et al,* 2011). Importantly, TF occupancy is not solely of a protective nature, but can induce demethylation *in vivo* via a yet unidentified mechanism (Thomassin *et al,* 2001; Xu *et al,* 2009; Stadler *et al,* 2011). Recruitment of the *E. coli* Lac repressor LacI to *in vitro* premethylated Lac operator sequence induces demethylation in a human cell line. Interestingly, titration of LacI by supplementation with IPTG suggests that the extent of demethylation directly correlates with occupancy of the target site but not with transcriptional activity (Lin *et al,* 2000). Similar conclusions were obtained from a study in fertilized *Xenopus* eggs for different transactivator domains (Matsuo *et al,* 1998). This study carries the functional analysis further and shows that demethylation by binding of TFs requires ongoing replication, suggesting a passive demethylation mechanism.

2.4.5.4 Role of 5mC derivatives

In theory, oxidation products of 5mC could also participate in gene regulation, although here the described relationship is even more complicated (Kriaucionis and Heintz 2009; Tahiliani *et al,* 2009; He *et al,* 2011; Ito *et al,* 2011). As intermediates of active demethylation and if DNA methylation is repressive, one would expect them to have an activating role. However, high 5-hydroxymethylcytosine enrichments at the TSS negatively correlate with gene expression (Pastor *et al,* 2011). Considering that most

promoters are CpG islands and bisulfite sequencing does not distinguish between 5mC and 5hmC (Huang *et al,* 2010), this anticorrelation is hardly surprising. Conversely, high gene-body enrichments have been associated with increased gene expression in brain tissue and in embryonic stem cells (Song C.X. *et al,* 2011; Wu *et al,* 2011; Mellen *et al,* 2012; Colquitt *et al,* 2013). This pattern resembles the distribution of 5mC so much that the presence of 5hmC within these sites is likely to be a consequence of substrate enrichment. Gel shift assays furthermore suggest that MeCP2 is a reader of 5hmC (Mellen *et al,* 2012), a finding confirmed by a mass spectrometry approach (Spruijt *et al,* 2013). However, these results are challenged by an older study in which the binding affinity of MeCP2 was measured *in vitro* for oxidative damaged 5mC DNA (Valinluck *et al,* 2004).

On the other hand, further oxidation of 5hmC appears to be detrimental for transcription, as *in vitro* elongation efficiency of the RNA polymerase II decreases at 5fC and 5caC containing targets (Kellinger *et al,* 2012).

Despite these correlative findings, changes in 5hmC levels could not be linked to concomitant changes in gene expression in a knockdown of TET1 (Williams *et al,* 2011). In summary, these findings do not allow to clearly discriminate the transcriptional effects of 5mC and its derivatives.

2.5 Scope of this thesis

Cell type specific gene expression programs require regulatory mechanisms which are capable to correctly interpret genetic information for each cell. Gene regulation ultimately relies on the interaction of transcription factors with their cognate sequences. Positive or negative modulation of such interactions is thus the function of any gene regulatory mechanism. In eukaryotes, this is achieved through the concerted action of nucleosomes and chromatin modifications.

When I started my PhD thesis, growing evidence suggested a role of DNA sequences in determination of epigenetic states, such as DNA methylation and trimethylation of lysine 27 of histone H3 (Lienert *et al,* 2011; Stadler *et al,* 2011; Arnold *et al,* 2013). In these studies establishment of chromatin state critically depended on the presence of transcription factors.

While DNA methylation has long been considered a stably repressive modification which is reprogrammed in germ cells and early embryogenesis (Wu and Zhang 2010), it became increasingly clear that this mark is reversible upon cellular differentiation (Mohn *et al,* 2008; Klug *et al,* 2010). Importantly, this process appears to depend on binding of transcription factors (Stadler *et al,* 2011). Advances in sequencing technologies allowed for a direct genome-wide measurement of DNA methylation. This revealed that its dynamic changes are especially pronounced at low methylated regions which are on average 30% methylated and often coincide with distal regulatory elements (Stadler *et al,* 2011). How exactly transcription factors induce hypomethylation and how stable is the achieved hypomethylated state remained undetermined.

We used two approaches to further investigate the relationship between transcription factors and DNA methylation at low methylated regions. First, we performed genome-wide chromatin immunoprecipitation of CTCF followed by bisulfite conversion and sequencing of the enriched DNA (Brinkman *et al,* 2012; Statham *et al,* 2012) to assess whether methylation is directly coupled to TF occupancy at the level of single molecules. By the time of performed experiments it has been suggested that DNA demethylation is achieved through oxidation of 5-methylcytosines to 5-hydroxymethylcytosine by the TET family of proteins (Kriaucionis and Heintz 2009; Tahiliani *et al,* 2009). Therefore, in a second approach we tested this hypothesis by using hydroxymethylated DNA immunoprecipitation followed by sequencing (Weber *et al,* 2005; Wu *et al,* 2011) during neuronal differentiation of embryonic stem cells. To further characterize the relationship

between transcription factor binding and DNA hydroxymethylation, we used a cell line bearing a genetic deletion of the factor REST. We then assessed the stability of DNA methylation in embryonic stem cells with a stable knockout or conditional inactivation of the two *de novo* DNA methyltransferases DNMT3A and DNMT3B. In order to quantify a potential turnover we measured methylation throughout a time-course of DNMT3A/B deletion by sequencing bisulfite PCR amplicons derived from different genomic regions.

3 Results

3.1 Transcription factor occupancy can mediate active turnover of DNA methylation at distal regulatory regions

Feldmann A*, Ivanek R*, Murr R*, Gaidatzis D, Burger L and Schübeler D

3.1.1 Summary

Cellular differentiation and development are largely regulated by distal regulatory elements, including enhancers and insulators. Binding of transcription factors to these elements is critical for their activity and coincides with locally reduced methylation (Stadler *et al,* 2011). Importantly, transcription factor binding is necessary and sufficient for formation of such low methylated regions (LMRs), even if they were previously methylated. How exactly these hypomethylated states are created and how transcription factor occupancy translates into DNA methylation is poorly understood.

Here we chose the DNA binding factor CTCF which has been previously shown to create LMRs as test case to investigate the relationship between DNA methylation and transcription factor binding. Using chromatin immunoprecipitation followed by bisulfite sequencing we show that in contrast to imprinted loci where only the unmethylated allele is occupied, CTCF can bind any methylation state at LMRs. We find an inverse relationship between occupancy and DNA methylation. Cytosines within sites of high CTCF enrichments have a high probability to be unmethylated as opposed to low occupancy sites, for which DNA methylation is heterogeneous. Our data suggest that CTCF binding is not statically linked to an unmethylated state but instead argue for a dynamic model of interaction in which methylated or unmethylated cytosines can be bound. In line with this model, we observe high enrichments of 5-hydroxymethylcytosine at cell type specific LMRs in embryonic stem cells and upon their neuronal differentiation, providing a mechanism for transcription factor dependent local demethylation. Moreover, cell type specific LMRs are enriched among sites of dynamic hydroxymethylation. As deletion of CTCF is lethal for embryonic stem cells, the link between transcription factor occupancy and hydroxymethylation was tested by profiling hydroxymethylation in a cell

line in which REST - another transcription factor implicated in LMR formation - was genetically deleted. This revealed an increase in DNA methylation and concomitant decrease of hydroxymethylation at REST bound LMRs.

Our results argue that transcription factors have the potential to induce TET protein dependent turnover of DNA methylation for maintenance and reprogramming of distal regulatory regions.

3.1.2 Published Manuscript

OPEN ACCESS Freely available online

PLOS | GENETICS

Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions

Angelika Feldmann[1◕], Robert Ivanek[1,2◕¤], Rabih Murr[1◕], Dimos Gaidatzis[1,2], Lukas Burger[1,2], Dirk Schübeler[1,3]*

1 Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland, **2** Swiss Institute of Bioinformatics, Basel, Switzerland, **3** Faculty of Science, University of Basel, Basel, Switzerland

Abstract

Distal regulatory elements, including enhancers, play a critical role in regulating gene activity. Transcription factor binding to these elements correlates with Low Methylated Regions (LMRs) in a process that is poorly understood. Here we ask whether and how actual occupancy of DNA-binding factors is linked to DNA methylation at the level of individual molecules. Using CTCF as an example, we observe that frequency of binding correlates with the likelihood of a demethylated state and sites of low occupancy display heterogeneous DNA methylation within the CTCF motif. In line with a dynamic model of binding and DNA methylation turnover, we find that 5-hydroxymethylcytosine (5hmC), formed as an intermediate state of active demethylation, is enriched at LMRs in stem and somatic cells. Moreover, a significant fraction of changes in 5hmC during differentiation occurs at these regions, suggesting that transcription factor activity could be a key driver for active demethylation. Since deletion of CTCF is lethal for embryonic stem cells, we used genetic deletion of REST as another DNA-binding factor implicated in LMR formation to test this hypothesis. The absence of REST leads to a decrease of hydroxymethylation and a concomitant increase of DNA methylation at its binding sites. These data support a model where DNA-binding factors can mediate turnover of DNA methylation as an integral part of maintenance and reprogramming of regulatory regions.

Citation: Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L, et al. (2013) Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions. PLoS Genet 9(12): e1003994. doi:10.1371/journal.pgen.1003994

Editor: Gregory S. Barsh, Stanford University School of Medicine, United States of America

Received April 23, 2013; **Accepted** October 14, 2013; **Published** December 19, 2013

Funding: AF is supported by a predoctoral fellowship from the Marie Curie training network "Nuc4D", RM is supported by an EMBO long-term postdoctoral fellowship. Research in the laboratory of DS is supported by the Novartis Research Foundation, the European Union (NoE "EpiGeneSys" FP7-HEALTH-2010-257082 and the "Blueprint" consortium FP7-282510), the European Research Council (EpiGePlas), the SNF Sinergia program and the Swiss initiative in Systems Biology (RTD Cell Plasticity). 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: Department of Biomedicine, University of Basel, Basel, Switzerland.

◕ These authors contributed equally to this work.

Introduction

Correct spatial and temporal regulation of genes depends on distal regulatory elements. Reprogramming the activity of these elements is thus central for successful cellular specialization [1,2]. Active distal regulatory elements are characterized by an open chromatin structure, corresponding to DNaseI hypersensitive sites, specific histone variants and histone modifications [3,4]. These modifications are thought to regulate the accessibility of the regulatory sequence and thus facilitate transcription factor (TF) binding [5].

Distal regulatory regions that reside outside of CpG islands are further unique, as they show reduced levels of DNA methylation when active [6–8]. Importantly, this feature is consistent between cell types so that it can be implemented to identify cell-type specific active regulatory elements as Low Methylated Regions (LMR) [6,7,9–11]. Although reduced, DNA methylation at LMRs is maintained at a residual level. This reflects heterogeneity within the population of sequenced DNA molecules, given that DNA methylation is binary for any particular cytosine. Functional experiments suggested that reduced methylation at LMRs critically depends on binding of transcription factors [7], but their role in creating methylation heterogeneity and whether this occurs via a passive and/or an active demethylation remains to be identified.

Several lines of evidence further link DNA demethylation to enhancer activity. Demethylation occurs at glucocorticoid receptor binding sites [8] and 5-hydroxymethylcytosine (5hmC), an intermediate of active demethylation via oxidation of 5-methylcytosines (5mC) by TET proteins [12–16], is present at active enhancers in embryonic stem (ES) cells as well as during neuronal and adipocyte differentiation [7,17–21]. Importantly, 5hmC can readily be detected in various cell types and thus utilized to locate regions of active DNA demethylation [22,23].

Here we addressed, whether heterogeneous methylation at LMRs reflects differential occupancy by transcription factors at individual molecules, using the DNA binding factor CTCF as an example. We show that CTCF-bound molecules display similar methylation levels as those observed in the entire cell population at CTCF binding sites. Moreover, for cytosines located within the CTCF motif, we find that binding affinity correlates with the likelihood of being unmethylated, so that CTCF is able to bind

Author Summary

Cell identity is determined by differential gene expression, which in turn is controlled by the combined activity of proximal and distal regulatory elements such as enhancers. DNA within active enhancer elements is marked by a hypomethylated state as a result of transcription factor (TF) binding. Here, using CTCF as an example for a DNA-binding factor, we explore the relationship between binding and DNA methylation at the level of single molecules by enriching for CTCF occupied DNA. To our surprise, methylation at molecules which are bound by CTCF does not differ from the average methylation levels at the binding sites defined by whole-genome bisulfite sequencing. We find that binding strength inversely correlates with DNA methylation within the CTCF motif with heterogenic methylation levels at low occupancy sites, suggesting that CTCF can bind to molecules with different methylation states. Moreover, we observed enrichment of 5-hydroxymethylcytosines at constitutive and cell-type specific TF binding sites indicative of an active demethylation process. To test the requirement of TF binding for the observed hydroxymethylation, and as CTCF deletion is incompatible with the survival of embryonic stem cells, we made use of cells in which REST – a factor which was previously shown to be involved in LMR formation - was genetically deleted. This deletion leads to loss of hydroxymethylation at its binding sites, suggesting that binding is necessary for turnover. Our data support a model in which TF occupancy mediates a continuous turnover of DNA methylation during maintenance and formation of active regulatory regions.

any methylation state within low occupancy sites. On the other hand, we find that high levels of hydroxymethylation coincide with the observed low methylation at LMRs, in a process that accounts for up to 20% of the genome-wide dynamics of 5hmC during neuronal differentiation of ES cells. Moreover, the presence of hydroxymethylation depends, at least partially, on TF binding, since genetic deletion of RE1-silencing transcription factor (REST) results in reduced hydroxymethylation at bound LMRs. Our results support a model where TF binding can occur at methylated regions and induce methylation turnover within active regulatory elements.

Results

Relation between CTCF occupancy and methylation states at CpG poor regions

Apart from CpG islands, mammalian genomes are mostly methylated. Notable exceptions are LMRs, CpG poor regions that display an average methylation level of 30% as measured by bisulfite sequencing (BisSeq). This reduced methylation marks active distal regulatory regions as it coincides with DNaseI hypersensitivity and enhancer-characteristic histone modifications [7]. We previously showed that, in the case of REST and CTCF, binding of trans-acting factors to DNA is required for LMR formation, yet it remains unclear whether and how this binding is related to the observed variation of DNA methylation between sequenced molecules [7]. Assuming a static model, unmethylated DNA would be limited to those molecules that are occupied by a TF, which in turn predicts that methylated molecules are not occupied, as has been established for imprinted CpG islands (Figure 1A, left) [24,25]. Alternatively, TFs could occupy all variations of methylation levels within LMRs (Figure 1A, right).

To test the first scenario, we performed Chromatin-IP (ChIP) in ES cells against the DNA binding factor CTCF and conducted bisulfite sequencing of the immunoprecipitated CTCF-bound DNA (ChIP-BisSeq) (Figure 1B) [26,27]. Importantly, the CTCF-ChIP enrichments recovered in our ChIP-BisSeq samples highly correlate with published ChIP enrichments [7] ($r = 0.91$ and 0.90 for replicate1 and replicate2, respectively) as well as between the replicate experiments ($r = 0.91$) (Figure S1A). Equally important, methylation for single cytosines correlates between the two replicates ($r = 0.8$, Figure S1B).

Only those CpGs, which show intermediate levels of methylation in BisSeq, can be informative to address our hypothesis. Therefore we first focused on CTCF sites located within LMRs. In this context it should be mentioned that the mean methylation of 30% observed at an LMR represents an average of individual cytosines within this LMR that can vary widely in their methylation percentage ([7] and data not shown). To ask if this heterogeneity is reduced at the occupied molecules, we compared methylation levels between the CTCF-bound fraction and the total population of cells. We first analyzed CpGs residing in sites of known allelic variation in CTCF binding, corresponding to DMRs, where we indeed only recover the unmethylated alleles in the ChIP-BisSeq assay (Figure 1B–C). This agrees with a recent report and confirms that our ChIP-BisSeq provides correct methylation status of bound molecules [25,26].

Next we asked if methylation patterns at CTCF-bound LMRs differ between exclusively bound molecules and the total population of DNA molecules. We analyzed average methylation levels for 200 bp regions centered at a CTCF motif only for those motifs which (1) overlap with LMRs, (2) are bound by CTCF as determined by ChIP enrichments and (3) for which all considered cytosines are covered at least 10 times in both ChIP-BisSeq and whole-genome (WG-) BisSeq. It is important to mention here, that while our ChIP does not allow for calling high resolution peaks such as those determined by other methods like ChIP-exo [28], our analysis pipeline is able to correctly identify high confidence bound sites as it requires CTCF motif in the center of the analyzed region in addition to high ChIP enrichment. This revealed a positive correlation with an equal spread of the data over the entire range ($r = 0.67$), arguing that LMRs do not display global differences in methylation levels at CTCF binding sites between the fraction of molecules bound by CTCF and those representing the total population of molecules in cells (Figure 1B). This finding is illustrated at individual loci (Figure 1C) and extends to CTCF binding sites outside of LMRs (Figure S1C).

We notice however that while entire LMRs do not display reduced methylation in the actually bound fraction of molecules, some individual cytosines in the vicinity of the CTCF motif do so (for example LMR1 in Figure 1C). To determine whether reduced methylation in CpGs close to the CTCF motif is a global phenomenon, we correlated changes in methylation between ChIP-BisSeq and WG-BisSeq with the distance to the nearest CTCF motif for individual cytosines with a minimal coverage of 10 fold in WG-BisSeq and ChIP-BisSeq in CTCF-bound Low Methylated Regions. This analysis revealed no correlation, suggesting that CTCF binding does not affect the methylation of proximal cytosines more than it does for the distal ones (Figure S1D). Therefore, the heterogeneity of occupancy by CTCF cannot explain the observed heterogeneity of methylation within LMRs, even though these can form upon CTCF binding and thus, at least in part, are CTCF dependent [7].

To further test the relationship between occupancy and methylation state, we next focused our analysis exclusively on CpGs that reside within a CTCF motif. We and others have

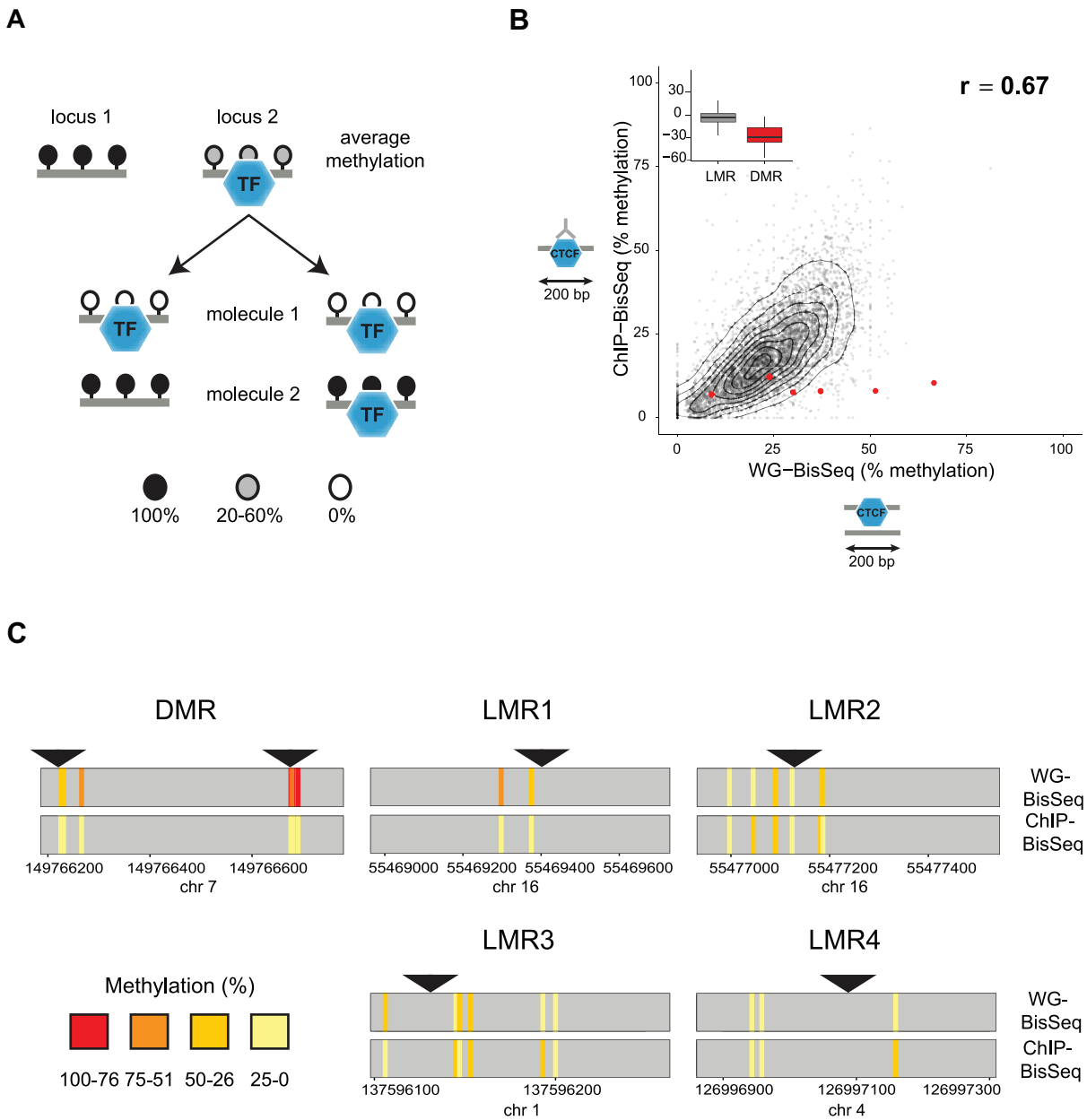

Figure 1. Relation between CTCF occupancy and methylation states in CpG poor regions. (A) LMRs are bound by transcription factors (TF) and have intermediate average methylation levels. There are two possible scenarios how TF binding and DNA methylation at CpG poor regions could be linked. In a static situation (left), TF binding would be linked to the unmethylated state of the bound molecule, whereas unbound molecules are fully methylated as previously shown for imprinted CpG islands. In an unlinked model (right), TF binding is independent of the DNA methylation state, therefore bound molecules display the same variation of methylation levels as the entire population. (B) To distinguish these scenarios we enrich for bound molecules by ChIP and determine their methylation by bisulfite sequencing (ChIP-BisSeq). This results in a high correlation of methylation levels between ChIP-BisSeq (y-axis) and whole genome bisulfite sequencing (WG-BisSeq, x-axis). Each point represents average methylation over a 200 bp region. Shown are only regions centered at a bound CTCF motif which overlaps with an LMR and for which all considered cytosines have a minimal coverage of 10× in both, WG-BisSeq and ChIP-BisSeq. Red points represent average for 200 bp windows centered on CTCF motifs located within DMRs. Boxplots show mean deviation of methylation levels in ChIP-BisSeq from those in WG-BisSeq at LMRs and DMRs in percent methylation. (C) Examples of single cytosine methylation levels in WG-BisSeq (top bars) and ChIP-BisSeq (bottom bars). For LMRs a whole segment is shown. Each bar represents a cytosine. Methylation is shown in a color code (red: high, yellow: low). Position of CTCF motifs is indicated by black triangles. Only cytosines with at least 10× coverage in both, WG-BisSeq and ChIP-BisSeq, are shown.
doi:10.1371/journal.pgen.1003994.g001

previously shown that methylation around occupied CTCF sites is the lowest at the actual binding motif and increases outwards [7,21]. Notably, 57% of all occupied sites by CTCF do not contain a CpG within the binding motif, yet display the same methylation pattern around the site (Figure 1B, Figure S1 and data not shown). Out of all predicted CTCF binding sites, 24.5% contain at least one CpG (Figure 2A–B). For all these sites, we related the strength of binding by CTCF as measured by ChIP enrichment to the methylation state of single CpGs within the motif. This reveals that strongly and weakly bound sites indeed differ in their methylation

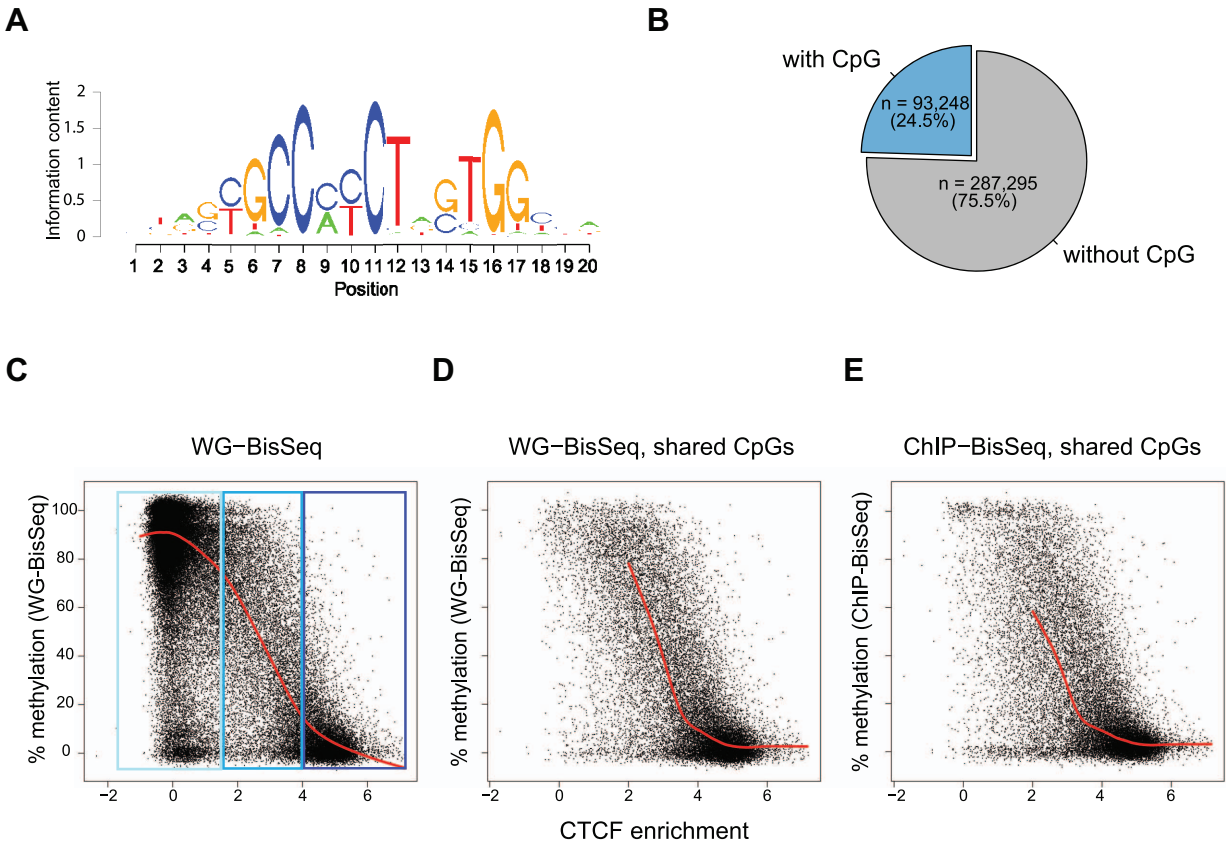

Figure 2. Relationship between binding strength and DNA methylation within the CTCF motif. (A) CTCF consensus motif used in this study [7]. (B) Percent of predicted CTCF sites containing a CpG within the motif. Exclusively these CpGs are shown in the plots (C–E). (C–E) Each point represents one individual CpG within a CTCF motif. (C) Correlation of methylation and CTCF enrichment identifies three classes of CTCF sites: unbound (light-blue), strongly bound and unmethylated (dark-blue), weakly bound with intermediate levels of methylation (blue). The red line represents a running mean measurement of methylation. (D) Same as C, but only showing cytosines covered in both WG-BisSeq and CTCF ChIP-BisSeq. (E) Same as D but only showing methylation levels derived from CTCF ChIP-BisSeq. In each case bound molecules show the same pattern as the entire population. Only cytosines residing within the CTCF binding motif and with a minimal coverage of 10× are shown. In order to prevent over-plotting the points were jittered with a standard deviation of 2%.
doi:10.1371/journal.pgen.1003994.g002

(Figure 2C). CpGs within highly occupied sites tend to be completely unmethylated, while methylation shifts towards intermediate levels with decreasing binding affinity. This links frequency of occupancy to methylation levels within the CTCF motif.

Again we can ask if heterogeneous methylation at weakly bound sites reflects actual occupancy at the level of individual molecules by analyzing their methylation in the bound fraction that was enriched by CTCF-ChIP. Also at these selected CpGs the methylation of exclusively occupied molecules is similar to the methylation of the total population (Figure 2D–E). Importantly, this relationship between the methylation state and CTCF binding is not dependent on the position of the analyzed CpG, as illustrated by the analysis of CpGs positioned exclusively at position 5–6 of the consensus motif (Figure S2).

Together, our data suggest that actual factor occupancy at the level of single molecules does not explain the observed DNA methylation heterogeneity adjacent to CTCF sites within LMRs or at the motif itself throughout the genome. This argues against a scenario of static methylation at CpG poor regions (Figure 1A, left), where DNA in a fraction of cells is bound by a TF and unmethylated, while other molecules are never occupied and remain methylated. Alternative scenarios could involve binding of a TF independently of methylation states, which in turn could trigger active demethylation (Figure 1A, right).

Hydroxymethylation marks LMRs in a cell-type specific and transcription factor binding dependent fashion

To ask if LMRs are indeed sites of active DNA methylation turnover, we determined the presence of 5hmC, the intermediate of TET mediated oxidation. Notably, bisulfite does not convert 5hmC and thus a fraction of the residual unconverted cytosines at LMRs could represent hydroxymethylcytosines [29,30]. We enriched for this modification by performing hydroxymethylcytosine DNA-immunoprecipitation (hMeDIP) followed by high throughput sequencing (hMeDIP-seq) in stem cells [31,32]. Analysis of the 5hmC profiles revealed its enrichment at LMRs of ES cells in line with other reports that suggested its presence at stem cell enhancers (Figure 3A) [19,21]. Analysis of an existing map of genomic binding sites further reveals that also TET1, an enzyme that mediates oxidation to 5hmC, is strongly enriched at LMRs in ES cells (Figure 3A) [33].

To address, whether the presence of 5hmC at LMRs is limited to stem cells or conserved in committed cells, we performed hMeDIP-Seq in neuronal progenitors (NP), derived through controlled differentiation of ES cells [34]. We previously showed

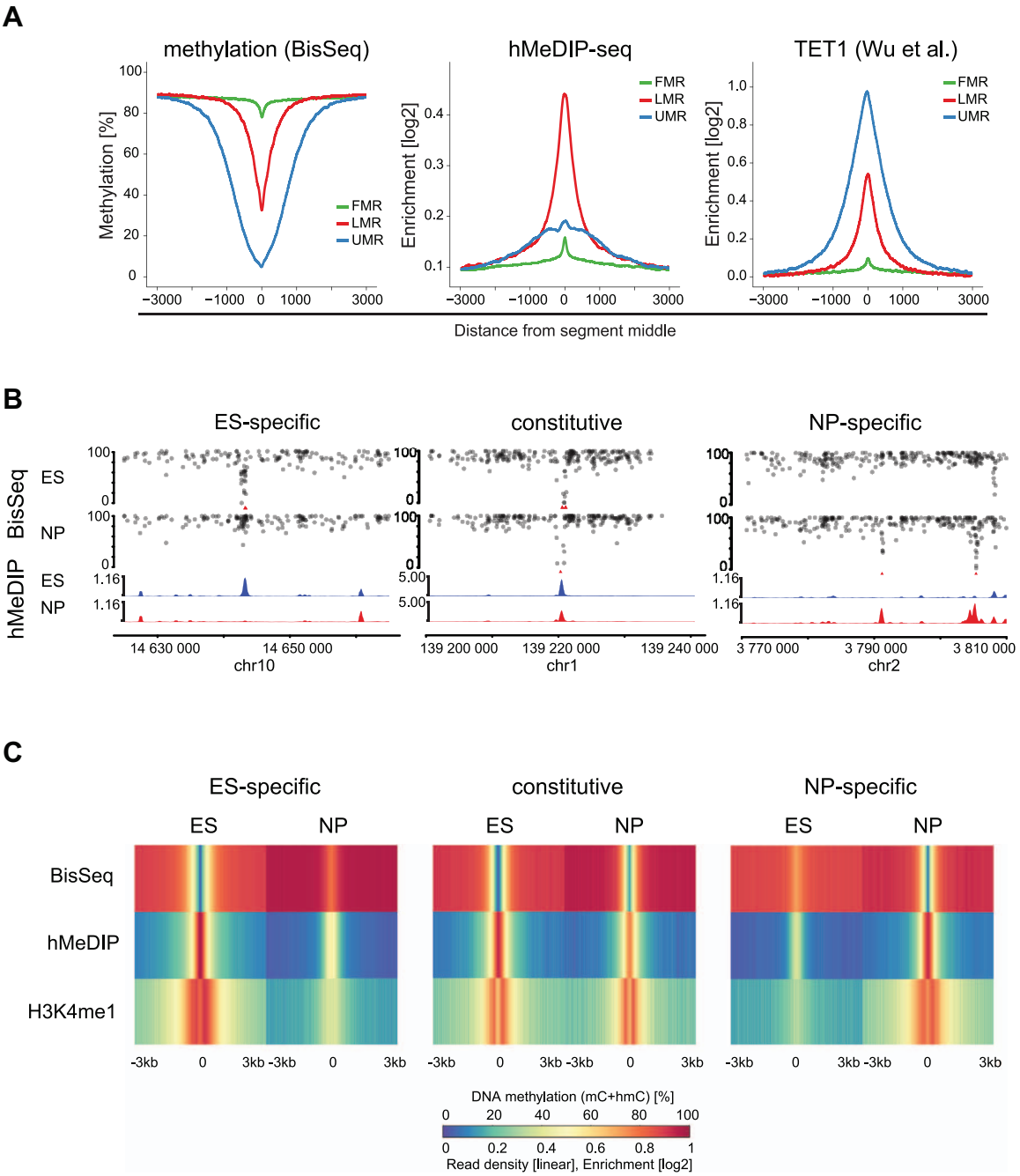

Figure 3. 5hmC marks LMRs in a cell-type specific fashion. (A) Average profiles for methylation (WG-BisSeq), 5hmC (hMeDIP-seq) and TET1 occupancy at Fully Methylated, Unmethylated and Low Methylated Regions (FMRs, UMRs and LMRs, respectively) in ES cells. (B) DNA methylation (upper tracks) and enrichment of 5hmC (lower tracks) in ES cells and NP of representative ES-specific, constitutive and NP-specific LMRs. (C) Average profiles representing methylation (WG-BisSeq), hMeDIP-seq and H3K4me1 ChIP-Seq in ES cells and NP ±3 kb around the segment middle.
doi:10.1371/journal.pgen.1003994.g003

in the same differentiation system that a large set of LMRs is cell-type specific, reflecting the extensive reprogramming of distal regulatory regions during somatic differentiation [7]. The resulting genomic 5hmC profiles reveal its enrichment at LMRs also in NP (Figure 3B–C). LMRs that are constitutive in both cell types show constitutive hydroxymethylation, suggesting that oxidation of 5-methylcytosine at LMRs also occurs in somatic cells (Figure 3B–C). ES-specific LMRs gain methylation and concomitantly lose hydroxymethylation in NP, suggesting that the state of reduced methylation and the presence of 5hmC coincide at active regulatory elements (Figure 3B–C, Figure S2). Similarly, NP-specific LMRs show a decrease in methylation and gain of hydroxymethylation along differentiation (Figure 3B–C, Figure S3). Notably, these NP-specific LMRs are enriched for neuron-specific TF binding sites, further confirming the link between TF binding at CpG poor regions and the presence of 5hmC [7]. The observed reciprocal behavior between loss of 5mC and gain of 5hmC is a general feature, as a genome-wide anti-correlation

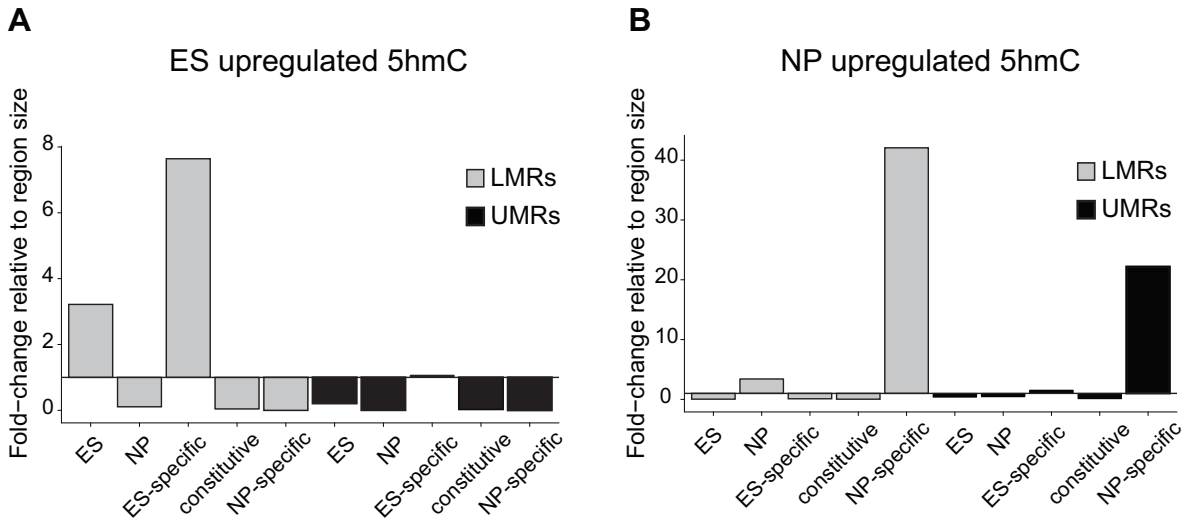

Figure 4. 5hmC dynamics during differentiation occurs preferentially at LMRs. (A–B) Shown is the relative frequency of changes in 5hmC at LMRs and UMRs normalized for genome coverage at the ES (A) and NP state (B). The y-axis shows observed linear fold enrichment relative to expected enrichments (see Materials and Methods). Note that 5hmC is changing preferentially at cell-type specific LMRs.
doi:10.1371/journal.pgen.1003994.g004

between changes in hMeDIP-Seq and WG-BisSeq ($r = -0.58$) as well as between changes in hMeDIP-Seq and MeDIP-Seq ($r = -0.30$, Figure S3) exists at LMRs.

To determine, if the observed turnover is selective for LMRs, we quantified 5hmC enrichments by hMeDIP-Seq throughout the genome and calculated the differences between ES cells and NP in order to identify genomic regions that show changes in the level of 5hmC. This revealed that cell-type specific enrichments for 5hmC show a large overlap with cell-type specific LMRs. This selectivity is further evident when calculating the occurrence in relation to genomic coverage (Figure 4). In this analysis, ES-specific LMRs are eightfold overrepresented in genomic regions that show enrichment for 5hmC in ES cells and the selectivity is even higher in NP, where NP-specific LMRs are more than 40-fold overrepresented.

This strong correlation suggests that transcription factors are required to induce hydroxymethylation. Indeed, 5hmC is more enriched at bound than at unbound CTCF motifs (Figure S4). To directly test whether increased 5hmC enrichment is a consequence of TF binding, we wanted to use a loss of function approach. Absence of CTCF, notably in ES cells, is cellular lethal [35–38], which precludes monitoring changes in methylation in cells that lack CTCF but otherwise are phenotypically normal. Effective depletion of CTCF would however be required in order to directly test its requirement in trans, since conserved binding sites remain occupied upon knockdown of CTCF [39]. As CTCF deletion is incompatible with survival of ES cells, we made use of a phenotypically normal ES cell line in which the *Rest* gene, coding for a different TF that is enriched within LMRs, had been genetically deleted. More specifically, we determined the level of hydroxymethylation at REST-bound LMRs. These regions become fully methylated in the absence of REST as measured by bisulfite sequencing, which is not discriminating between 5mC and 5hmC (Figure 5A–B). When measuring hydroxymethylation specifically by hMeDIP (see Table S1 for primers) we find that 5hmC levels are significantly reduced at these binding sites in REST knockout ES cells (Figure 5C). This indicates that factor activity in trans is required for increased hydroxymethylation at LMRs within a given cell type.

These observations are compatible with a scenario in which reduced DNA methylation at regulatory regions entails the presence of active DNA methylation turnover in both stem and differentiated cells.

Discussion

Using CTCF as example, this study provides further evidence that maintenance and reprogramming of correct DNA methylation levels at distal regulatory regions can entail active turnover as a function of transcription factor binding. We show that the loss of methylation at these regions during cellular differentiation involves a reciprocal gain of 5hmC and vice versa. This process occurs preferentially at LMRs and we demonstrate that it accounts for up to 20% of all observed changes in 5hmC during differentiation. These findings are compatible with previous reports of dynamic hydroxymethylation [18,40]. Importantly, this association is not limited to stem cells, even though these have been suggested to display higher global levels of 5hmC than differentiated cells [16]. We also show that this phenomenon can go beyond correlation, since genetic deletion of the TF REST results in reduced hydroxymethylation at its binding sites already in stem cells. Our results obtained from CTCF and REST mechanistically link binding of TF at regulatory regions with active demethylation. However, in light of the estimated 1400 different TFs encoded in mammalian genomes, it would be premature to generalize these findings.

The fact that CTCF can occupy different methylation states in CpG poor regions together with the presence of both 5hmC and TET1 at these sites is compatible with a scenario, where TF binding triggers an active demethylation process. In case of CTCF it is evident that the binding strength determined by ChIP relates directly to the level of demethylation within the binding motif. The frequency of binding correlates with the likelihood of a demethylated state for a cytosine within the binding site. Assuming that this relation extends to factors other than CTCF adds yet another dimension to whole-genome basepair methylomes by providing not only information about the activity of regulatory regions, but also about the strength of binding of trans-acting factors. It is important to note however that both CTCF and REST are rather

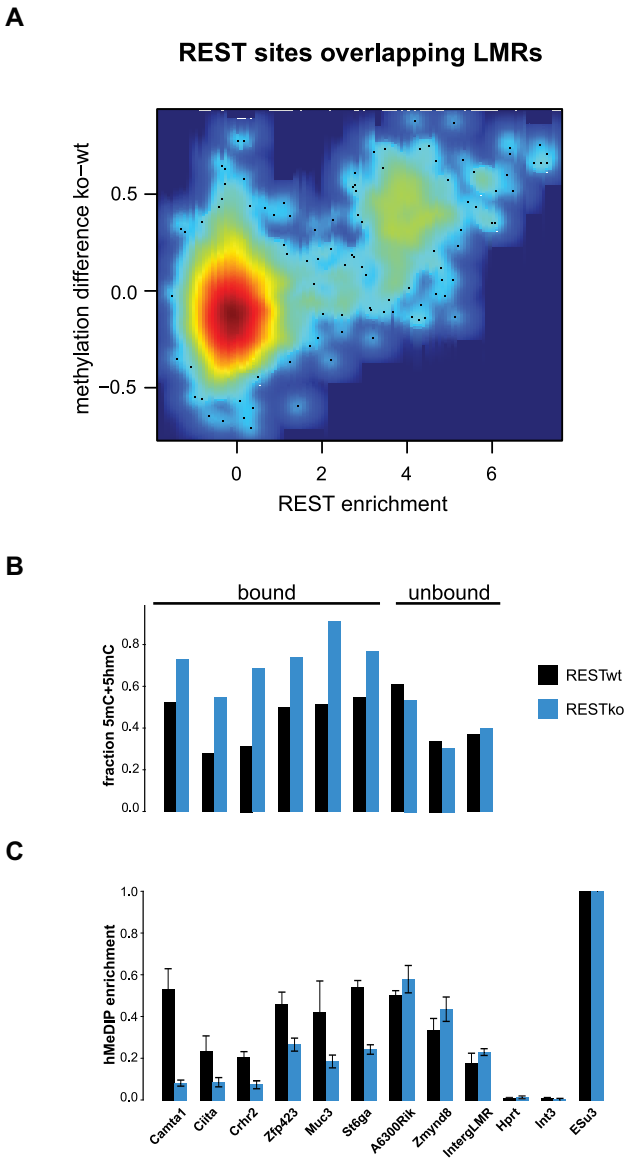

Figure 5. 5hmC enrichment at REST-bound LMRs is partially dependent on the presence of REST. (A) Relative methylation changes between REST wildtype and REST knockout ES cells are correlated to REST ChIP enrichment. Methylation was determined 200 bp around the REST motif at all REST sites overlapping with LMRs. The point density is colour-coded (red: high, blue: low point density). Methylation determined by BisSeq (B) and hMeDIP qPCR enrichments (C) at REST motif containing LMRs bound and not bound by REST in wildtype (wt, dark blue) and REST knockout (ko, blue) ES cells. Error bars in (C) represent standard deviation in three replicate experiments normalized to a positive control.
doi:10.1371/journal.pgen.1003994.g005

special in regards to the large size of their sequence motifs (20 and 21 bp, respectively), which further limits the ability to generalize our observations. Clearly, a more comprehensive approach is needed to address the effect of additional DNA-binding factors on DNA methylation.

While the actual mode of demethylation remains to be determined, it seems possible that DNA binding factors recruit TET proteins, which in turn mediate oxidation to 5hmC [41]. However, in light of the generality of the link between LMR formation and 5hmC, this would require a large number of TFs to share such recruitment ability. Alternatively, recruitment might be mediated by general cofactors that are frequently observed at distal regulatory regions such as p300 or by pioneer factors [3,42]. A further scenario could be that a specific nucleosome or DNA organization results from binding of a TF, which in turn triggers TET recruitment [43].

At this point we can only speculate if 5hmC presence at regulatory regions solely reflects active turnover [21,44–48] and how much an active process contributes to the low levels of methylation observed. Moreover, it remains to be shown if presence of hydroxymethylation is actually involved in enhancer regulation. This would require specific readers of this DNA modification. Indeed, several proteins have been suggested to bind 5hmC, including the MBD domain proteins MeCP2 [49] and MBD3 [50]. Our recent functional mapping, however, suggested that genomic binding sites of MBD3 are independent of the presence of hydroxymethylation [51] in agreement with *in vitro* binding [52], making this scenario less likely. In addition, other putative readers of 5hmC were suggested in a proteomics screen, yet only few appear to be selective for 5hmC *in vitro* [52]. Conversely, two recent studies report the accumulation of TET-mediated 5hmC oxidation products 5-formylcytosine and 5-carboxylcytosine at proximal and distal regulatory elements in the absence of TDG [46,47], arguing for the appearance of an active turnover at LMRs. It remains to be determined, whether DNA binding factors, such as CTCF and REST used here, are able to bind to hydroxymethylated regions. While strong CTCF binding sites are devoid of methylation and hydroxymethylation, it is possible that CTCF is able to bind to 5mC as well as to 5hmC at low occupancy sites.

Our findings argue that LMRs do not result solely from a passive loss of methylation during replication, which is in line with the observation that LMRs can be detected in methylomes from non-dividing cells [9] and with recently reported presence of 5-formylcytosine and 5-carboxylcytosine at these elements [46,47]. At this point we lack experimental evidence for the relevance of reduced methylation for the function of distal regulatory regions. It is conceivable, but remains to be shown, that reduced methylation induced by pioneering TFs would enhance binding of other TFs, which are sensitive to DNA methylation even in CpG poor regions [53,54]. Alternatively, but not mutually exclusive, reduced methylation could mediate a chromatin state that functions as a general attractor for DNA binding factors and thus would stabilize the on-state [55].

Materials and Methods

ES cell culture and differentiation

159-2 ES cells were cultured and differentiated as previously described [7,34].

CTCF ChIP-bisulfite sequencing

Chromatin immunoprecipitation (ChIP) assay for CTCF was performed according to the Upstate protocol using the antibody anti-CTCF (SantaCruz #15914). 100 ng of immunoprecipitated DNA were used for subsequent library preparation. DNA fragments were end repaired by incubation at 20°C for 30 minutes with 400 µM dNTP, 3 units of T4 DNA polymerase (NEB #M0203S), 5 units of DNA Polymerase I Lg. Frag. (Klenow) (NEB #M0210S), 10 units of T4 PNK (NEB #M0201S), 1× T4 DNA ligase buffer containing 10 mM ATP (NEB), followed by column purification using QIAquick PCR Purification Kit (QIAGEN #28106). 3′ ends of DNA fragments were adenylated

by incubation at 37°C for 30 minutes with 200 µM dATP, 1×NEB Buffer 2, 5 units Klenow Fragment (3′→5′ exo–) (NEB # M0212L), followed by column purification using MinElute PCR Purification Kit (QIAGEN # 28006). Adapter for single end sequencing were reproduced based on Illumina adapter sequences. Annealed adapters were ligated to the DNA fragments by incubation at room temperature for 15 min in the following mix: 400 nM of annealed adapters, 1× NEB Quick ligase buffer, 2.000 units of T4 Quick ligase (NEB #M2200S), followed by column purification using MinElute PCR Purification Kit. 200 ng of Drosophila DNA (Kc cells) were then added as a carrier. Adapter-ligated DNA of 150–400 bp was selected from 2% agarose gel electrophoresis and purified using MinElute Gel Extraction Kit (QIAGEN #28606). BSA (final concentration 0.5 µg/µl) was added to gel-purified DNA and the mix was then treated with sodium bisulfite using the Imprint DNA Modification Kit (Sigma-Aldrich) as per manufacturer's instructions. DNA was enriched using 18 cycles of PCR with the following reaction composition: 2.5 U of uracil-insensitive *PfuTurboCx* Hotstart DNA polymerase (Stratagene), 5 µl 10× *PfuTurbo* reaction buffer, 25 µM dNTPs, 0.5 µM of Single End Illumina PCR primers (1.1 and 2.1). The thermocycling parameters were: 95°C 2 min, 98°C 30 sec, then 18 cycles of 98°C 15 sec, 65°C 30 sec and 72°C 3 min, ending with one 72°C 5 min step, followed by column purification using the MinElute PCR Purification Kit. DNA was then run on 2% agarose gel to separate the library from adapter dimers and purified using the MinElute Gel Extraction Kit. Quality of the libraries and template size distribution were checked on an Agilent 2100 Bioanalyzer (Agilent Technologies).

RESTko bisulfite sequencing

Library for the shotgun whole-genome BisSeq for RESTko cells was prepared as previously described [7] and sequenced using one lane of Illumina HiSeq 2000.

hMeDIP and MeDIP sequencing library preparation

Genomic DNA was fragmented to 200–1000 bp fragments with a Bioruptor (Diagenode, Sparta, NJ). The protocol for the library preparation was adapted from Illumina Genomic DNA Sample Preparation Guide. Briefly, 7 to 10 µg of fragmented DNA were end repaired and their 3′ ends adenylated. Genomic single end or paired end adapters were annealed. (h)MeDIP was performed as previously described [56] using 4 ug of adapter-ligated DNA and 4 µl of a 1:10 dilution of rabbit polyclonal anti-hmC antibody (Active Motif #39770) for hMeDIP or 10 µl of mouse monoclonal 5mC antibody (Eurogentec #BI-MECY-1000) for 2 hrs, followed by addition of 40 µl of Protein A Dynabeads (Invitrogen, #100.02D, hMeDIP) or Dynabeads M-280 Sheep anti-mouse IgG (Dynal Biotech #112.01) added for another 2 hrs. Immunoprecipitated DNA was amplified by 18 cycles of PCR following the Illumina Genomic DNA Sample Preparation Guide and purified using the MinElute PCR purification kit. Fragments of 250–300 bp (for single end sequencing) or 400–450 bp (for paired end sequencing) were size-selected from 2% agarose gel and purified using the MinElute Gel Extraction Kit. Quality of the libraries and template size distribution were checked on an Agilent 2100 Bioanalyzer (Agilent Technologies).

High-throughput sequencing

(h)MeDIP-seq and ChIP-BisSeq were sequenced using the Illumina HiSeq 2000 as per manufacturer's instructions.

Analysis of sequencing data

The hMeDIP-seq data were analyzed similarly to ChIP-Seq data in Stadler et al. Briefly, the July 2007 M. musculus genome assembly (NCBI37/mm9) provided by NCBI (http://www.ncbi.nlm.nih.gov/genome/guide/mouse/) and the Mouse Genome Sequencing Consortium (http://www.sanger.ac.uk/Projects/M_musculus/) was used as a basis for all analyses. For reads from hMeDIP-seq experiments, alignments to the mouse genome were performed by the software bowtie (version 0.9.9.1) [57] with parameters -v 2 -a -m 100, tracking up to 100 best alignment positions per query and allowing at most two mismatches. Each alignment was weighted by the inverse of the number of hits. All quantifications were based on weighted alignments. Alignments were shifted by 60 bases (estimated fragment length was 120 bp).

In order to identify regions with different signal in hMeDIP-seq between ES and NP, the mouse genome was partitioned into 1 kb sized windows with an overlap of 500 bp. For each window we calculated log2 fold change between NP and ES using in the following way: log2(FC) = log2((n_NP/N_NP *min(-N_ES,N_NP)+p)/(n_ES/N_ES *min(N_ES,N_NP)+p)), where n_ES and n_NP are the summed weights of overlapping ES and NP read alignments, respectively. N_ES and N_NP are the total number of aligned reads in ES and NP samples and p is a pseudocount constant (p = 8) used to regularize enrichments based on low counts that would otherwise be dominated by sampling noise. Windows with log2(FC) bigger than 3 or smaller than −3 in both biological replicates were merged into regions showing the gain and loss of signal in NP, respectively. These regions were used to calculate the enrichment in segment types (constitutive, ES- or NP-specific LMRs, UMRs). Enrichments were calculated as the ratio of observed over expected number of bases of each region class (gain of signal in NP, loss of signal in NP) in a segment type (e.g. ES-specific LMR etc.), where the observed number is the number of bases in regions of a given class that overlap a segment and the expected number is the fraction of genomic bases in that segment type, multiplied with the total number of bases in all regions of that class.

Analysis of ChIP-Seq and bisulfite (ChIP-BisSeq) data, ChIP enrichment calculation and identification of CTCF binding sites were performed as previously described (Stadler et al. 2011). The data from the two CTCF ChIP-BisSeq replicates were pooled for the analysis. Analysis of REST ChIP-Seq data and genome-wide prediction of REST motifs was performed analogously to CTCF. In the case of REST, the inferred weight matrix was extended to allow for a variable linker (0–11 nts in length) after position 9.

Datasets used in this study

Datasets generated for this study, ChIP-BisSeq, hMeDIP-seq, MeDIP-seq and RESTko methylome have been submitted to GEO and are available under the accession number GSE39739. Data for CTCF ChIP-Seq and WG-BisSeq was downloaded from GEO: GSE30206 [7], data for REST ChIP-Seq were downloaded from GSE27148 [58]. Tet1 ChIP-Seq data was downloaded from GEO: GSE26833 [33].

Supporting Information

Figure S1 Genome-wide relation between transcription factor occupancy and methylation states. (A) Correlation of ChIP enrichments between CTCF ChIP-Seq (Stadler et al., Nature 2011) and the two CTCF ChIP-BisSeq replicates used in this study. (B) Correlation of methylation levels at individual CpGs between two CTCF ChIP-BisSeq replicates. Selected cytosines have a minimal coverage of 10 in both replicates. (C) Correlation

of average methylation levels at regions 200 bp around all predicted CTCF sites between WG-BisSeq and a pool of both CTCF ChIP-BisSeq replicates. Selected regions have a minimal coverage of 10 in all cytosines used for the calculation of methylation levels in both WG-BisSeq and ChIP-BisSeq. (D) For individual cytosines within LMRs the methylation difference between ChIP-BisSeq and WG-BisSeq is correlated with the distance to the nearest CTCF motif center.
(PDF)

Figure S2 Relationship between binding strength and DNA methylation within the CTCF motif. (A) CTCF consensus motif used in this study. Here only cytosines are analyzed which are at position 5–6 of the motif. Out of all predicted sites containing a CpG within the motif (24.5% of all predicted sites) 42.2% have a CpG at this position. (B–D) Each point represents one individual CpG at position 5–6 of the PWM. (B) Correlation of methylation and CTCF enrichment identifies three classes of CTCF sites: unbound (light-blue), strongly bound and unmethylated (dark-blue), weakly bound with intermediate levels of methylation (blue). The red line represents a running mean measurement of methylation. (C) Same as B, but only showing cytosines covered in both WG-BisSeq and CTCF ChIP-BisSeq. (D) Same as C but only showing methylation levels derived from CTCF ChIP-BisSeq. In each case bound molecules show the same variation as the entire population. Only cytosines residing within the CTCF binding motif and with a minimal coverage of 10× are shown. In order to prevent over-plotting the points were jittered with a standard deviation of 2%.
(PDF)

Figure S3 5hmC marks LMRs in a cell-type specific fashion. (A) Replicate correlation for hMeDIP-seq. Shown is the log2 fold change of 5hmC between ES and NP in two biological replicates. (B) Correlation of hMeDIPseq and WG-BisSeq at LMRs during neuronal differentiation. Shown are the log2 fold change in 5hmC between ES and NP (y-axis) and change in DNA methylation percentage (x-axis). (C) Correlation of hMeDIP-seq and MeDIP-seq at LMRs during neuronal differentiation. Shown are the log2 fold change in 5hmC between ES and NP (y-axis) and change in DNA methylation percentage (x-axis).
(EPS)

Figure S4 5hmC enrichment at CTCF sites 5hmC enrichment at CTCF sites depends on CTCF binding. Shown are hMeDIP-seq enrichments in ES cells over bound and unbound CTCF motifs.
(EPS)

Table S1 Primer sequences used for qPCR.
(DOCX)

Acknowledgments

We are grateful to Sophie Dessus-Babus, Tim Roloff (FMI) and Ina Nissen, Christian Beisel (ETH BSSE Quantitative Genomics facility Basel) for processing deep sequencing samples. We thank Michael Stadler and members of the Schübeler lab for advice.

Author Contributions

Conceived and designed the experiments: AF RM DS. Performed the experiments: AF RM. Analyzed the data: RI DG LB. Wrote the paper: AF RI RM DS.

References

1. Okita K, Ichisaka T, Yamanaka S (2007) Generation of germline-competent induced pluripotent stem cells. Nature 448: 313–317.
2. Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, et al. (2008) Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol 6: e253.
3. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, et al. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459: 108–112.
4. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, et al. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470: 279–283.
5. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, et al. (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell 132: 958–970.
6. Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, et al. (2011) Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell 44: 17–28.
7. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, et al. (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480: 490–495.
8. Wiench M, John S, Baek S, Johnson TA, Sung MH, et al. (2011) DNA methylation status predicts cell type-specific enhancer activity. EMBO J 30: 3028–3039.
9. Burger L, Gaidatzis D, Schubeler D, Stadler MB (2013) Identification of active regulatory regions from DNA methylation data. Nucleic Acids Res.
10. Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, et al. (2013) Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat Genet.
11. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, et al. (2013) Charting a dynamic DNA methylation landscape of the human genome. Nature 500: 477–481.
12. Bhutani N, Burns DM, Blau HM (2011) DNA demethylation dynamics. Cell 146: 866–872.
13. He YF, Li BZ, Li Z, Liu P, Wang Y, et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333: 1303–1307.
14. Inoue A, Zhang Y (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science 334: 194.
15. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324: 929–930.
16. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324: 930–935.
17. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, et al. (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature 473: 394–397.
18. Serandour AA, Avner S, Oger F, Bizot M, Percevault F, et al. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucleic Acids Res 40: 8255–8265.
19. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE (2011) 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol 12: R54.
20. Szulwach KE, Li X, Li Y, Song CX, Han JW, et al. (2011) Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS Genet 7: e1002154.
21. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, et al. (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the Mammalian genome. Cell 149: 1368–1380.
22. Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, et al. (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One 5: e15367.
23. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, et al. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res 22: 467–477.
24. Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405: 482–485.
25. Xie W, Barr CL, Kim A, Yue F, Lee AY, et al. (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell 148: 816–831.
26. Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, et al. (2012) Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res 22: 1128–1138.
27. Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, et al. (2012) Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res 22: 1120–1127.
28. Rhee HS, Pugh BF Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147: 1408–1419.
29. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, et al. (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One 5: e8888.

30. Jin SG, Kadam S, Pfeifer GP (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res 38: e125.
31. Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, et al. (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 473: 398–402.
32. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, et al. (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 473: 343–348.
33. Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, et al. (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature 473: 389–393.
34. 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.
35. Fedoriw AM, Stein P, Svoboda P, Schultz RM, Bartolomei MS (2004) Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science 303: 238–240.
36. Heath H, Ribeiro de Almeida C, Sleutels F, Dingjan G, van de Nobelen S, et al. (2008) CTCF regulates cell cycle progression of alphabeta T cells in the thymus. EMBO J 27: 2839–2850.
37. Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T (2012) CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. Cell Rep 2: 345–357.
38. Soshnikova N, Montavon T, Leleu M, Galjart N, Duboule D (2010) Functional analysis of CTCF during mammalian limb development. Dev Cell 19: 819–830.
39. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, et al. (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 148: 335–348.
40. Tan L, Xiong L, Xu W, Wu F, Huang N, et al. Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. Nucleic Acids Res.
41. Ding J, Xu H, Faiola F, Ma'ayan A, Wang J (2012) Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res 22: 155–167.
42. Serandour AA, Avner S, Percevault F, Demay F, Bizot M, et al. (2011) Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome Res 21: 555–565.
43. Fu Y, Sinha M, Peterson CL, Weng Z (2008) The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS Genet 4: e1000138.
44. Frauer C, Hoffmann T, Bultmann S, Casa V, Cardoso MC, et al. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. PLoS One 6: e21306.
45. Kubosaki A, Tomaru Y, Furuhata E, Suzuki T, Shin JW, et al. CpG site-specific alteration of hydroxymethylcytosine to methylcytosine beyond DNA replication. Biochem Biophys Res Commun 426: 141–147.
46. Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, et al. (2013) Genome-wide Analysis Reveals TET- and TDG-Dependent 5-methylcytosine Oxidation Dynamics. Cell 153.
47. Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, et al. (2013) Genome-wide profiling of 5-Formylcytosine reveals its roles in epigenetic priming. Cell.
48. Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res 67: 946–950.
49. Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 151: 1417–1430.
50. Yildirim O, Li R, Hung JH, Chen PB, Dong X, et al. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell 147: 1498–1510.
51. Baubec T, Ivanek R, Lienert F, Schubeler D Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. Cell 153: 480–492.
52. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell 152: 1146–1159.
53. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13: 484–492.
54. Schubeler D Molecular biology. Epigenetic islands in a genetic ocean. Science 338: 756–757.
55. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25: 1010–1022.
56. Mohn F, Weber M, Schubeler D, Roloff TC (2009) Methylated DNA immunoprecipitation (MeDIP). Methods Mol Biol 507: 55–64.
57. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.
58. Arnold P, Scholer A, Pachkov M, Balwierz PJ, Jorgensen H, et al. Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. Genome Res 23: 60–73.

Figure S1

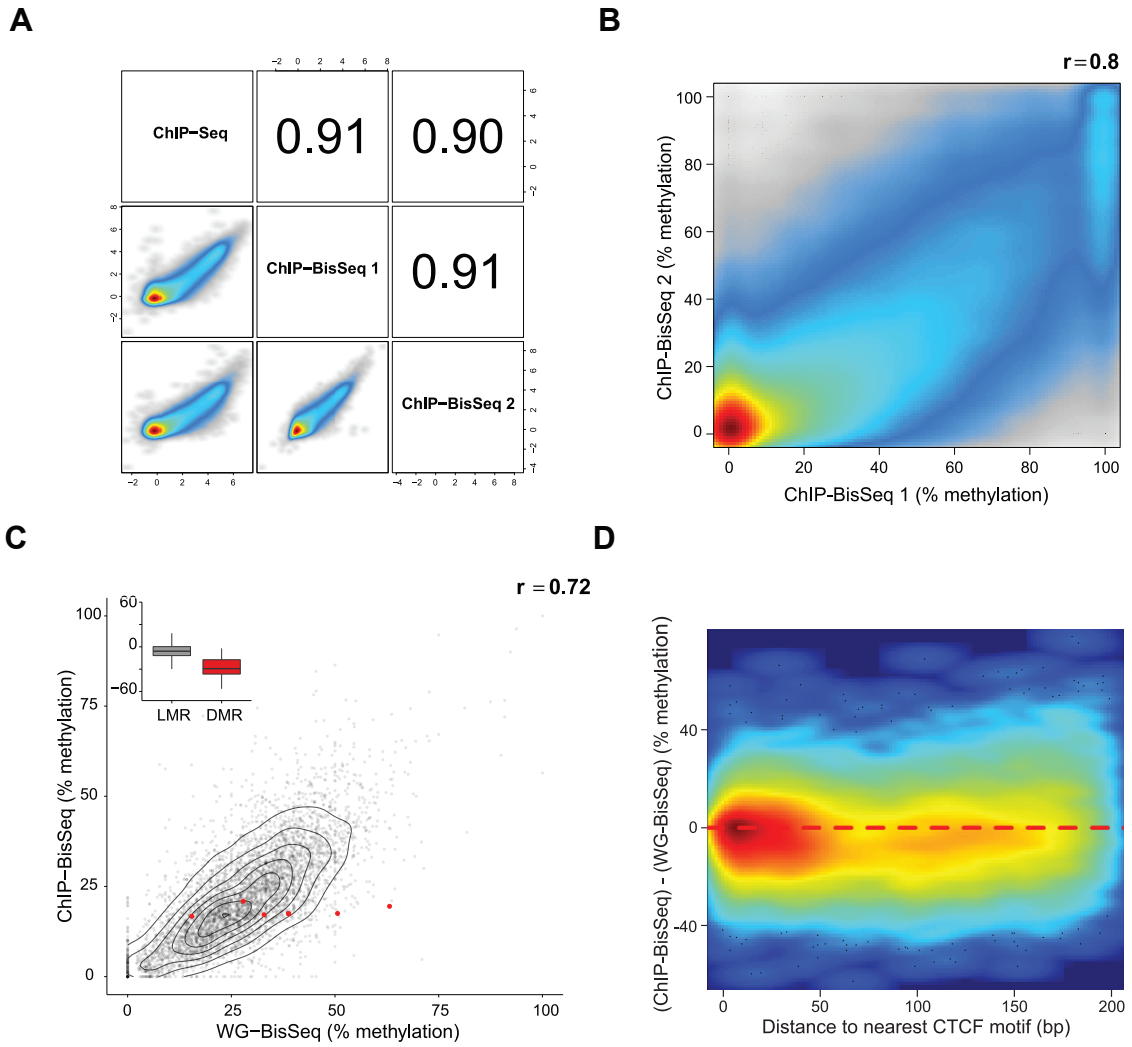
A
ChIP−Seq
0.91
0.90
ChIP−BisSeq 1
0.91
ChIP−BisSeq 2
B
r=0.8
ChIP-BisSeq 2 (% methylation)
ChIP-BisSeq 1 (% methylation)
C
r = 0.72
LMR
DMR
ChIP−BisSeq (% methylation)
WG−BisSeq (% methylation)
D
(ChIP-BisSeq) - (WG-BisSeq) (% methylation)
Distance to nearest CTCF motif (bp)

Figure S2

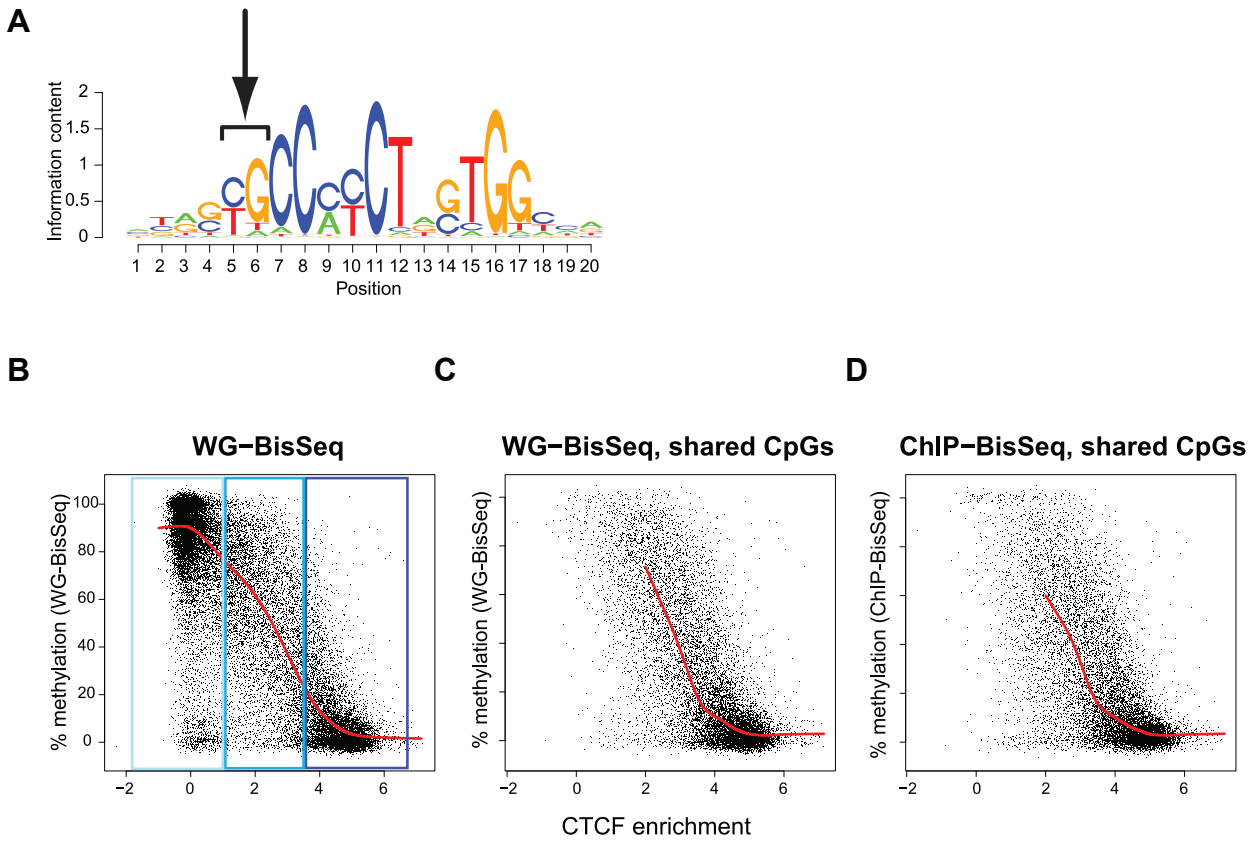
A
Information content
2
1.5
1
0.5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Position
B
C
D
WG−BisSeq
WG−BisSeq, shared CpGs
ChIP−BisSeq, shared CpGs
% methylation (WG-BisSeq)
% methylation (ChIP-BisSeq)
100
80
60
40
20
0
−2 0 2 4 6
CTCF enrichment

Figure S3

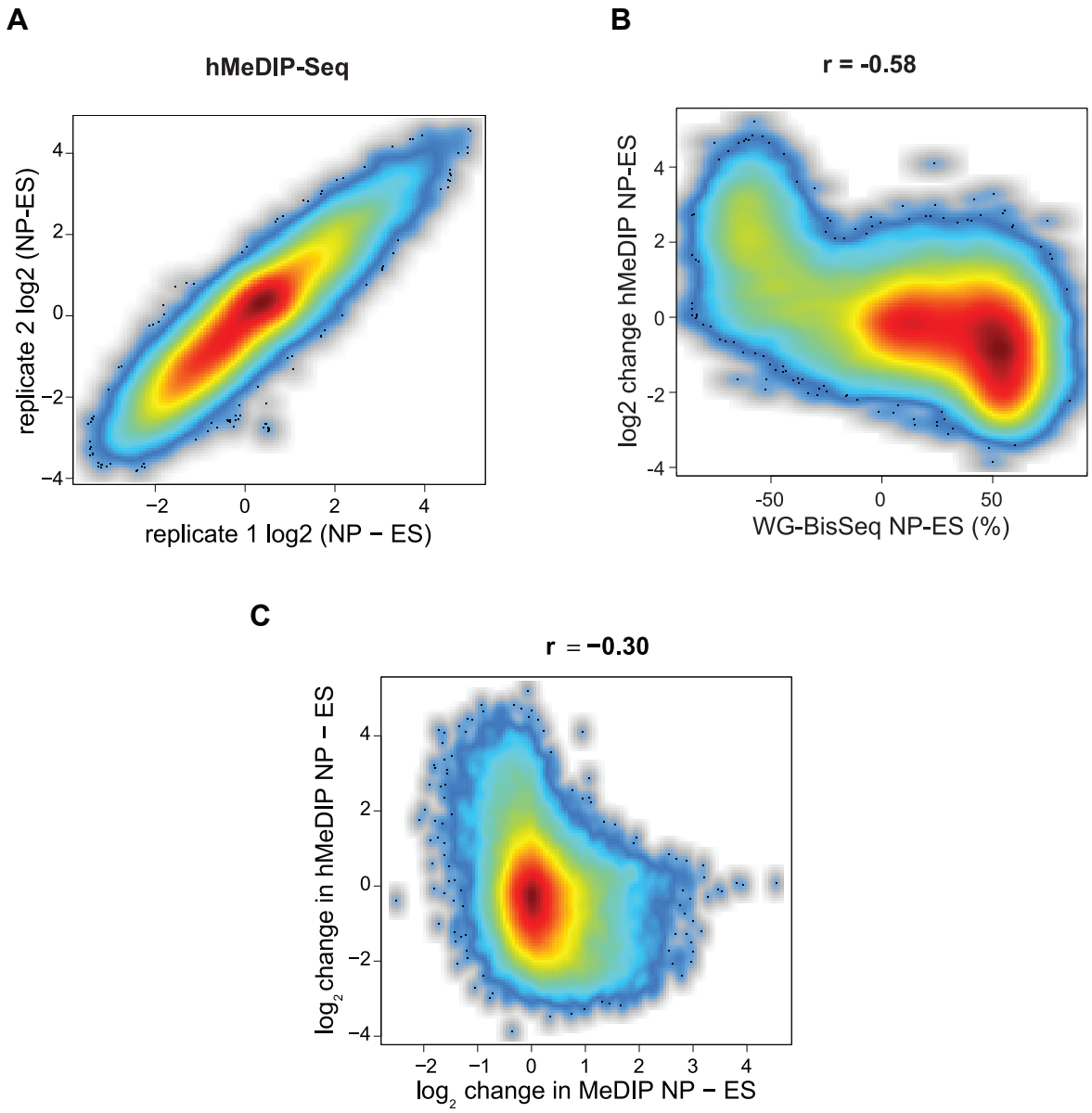
A
hMeDIP-Seq
replicate 2 log2 (NP-ES)
replicate 1 log2 (NP − ES)
B
r = -0.58
log2 change hMeDIP NP-ES
WG-BisSeq NP-ES (%)
C
r = −0.30
log2 change in hMeDIP NP − ES
log2 change in MeDIP NP − ES

Figure S4

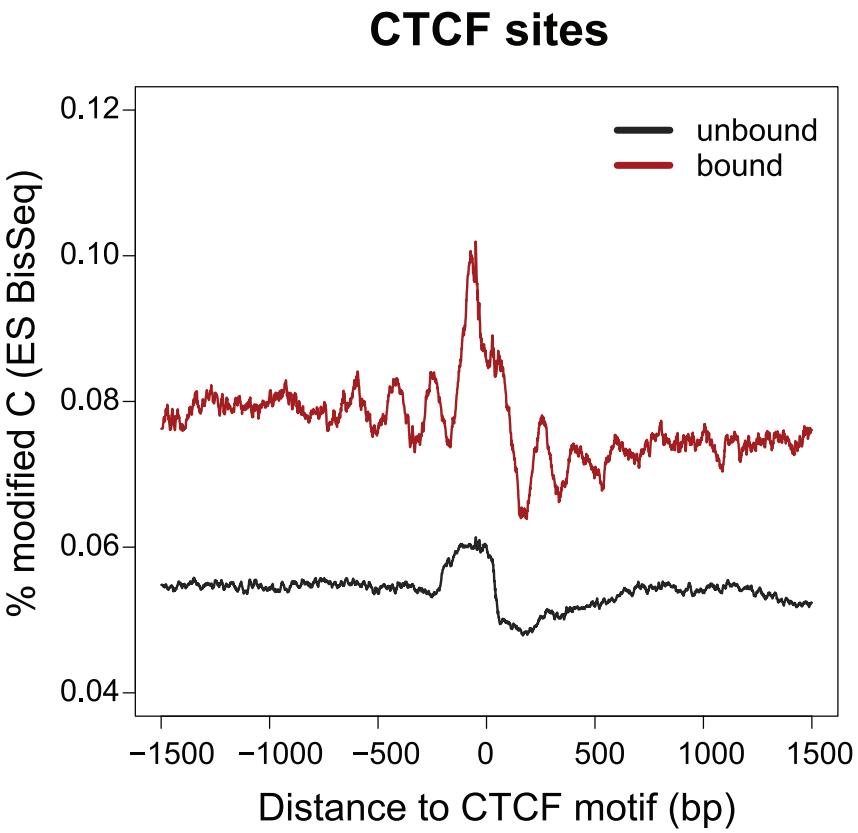
CTCF sites
unbound
bound
% modified C (ES BisSeq)
0.12
0.10
0.08
0.06
0.04
−1500
−1000
−500
0
500
1000
1500
Distance to CTCF motif (bp)

Table S1. Primer Sequences used for qPCR

Genomic region	Forward	Reverse
Camta	GCTTCAGGGCTACAGAGTGC	AGTCAGAGGCTACCCCTGGT
Ciita	GCAAGCTGGAGAAAAAGCAC	TAGGATGAAGCCTGGGTGTC
Crhr2	CGTGGCATTTATCGAAGTCA	GTGGTCAGGAGCTCTCCAAG
Zfp423	CATTTGCTTCTCCGCAGATT	CATGTTTATGTCCGCTGCTG
Muc3	CGGGTAGGAGACATCTCTGG	CCAGAGAGATGATGCTGGAAG
St6ga	CTCTTCTCGGTCACCCATTC	AATCACCCGCTGTGAATCAT
A6300Rik	CCCACGTCTCCATGGTTAAT	TTCTGTGCGTGGCTAAACAG
Zmynd8	GGCGTTTCCTTGATTGACAT	AAGACAGGACCTGGAGGAGA
Interg LMR	CTTTGGCACACTGCCATCTA	CCTTTTCATGAGACCCGAAA
Interg3	ATGCCCCTCAGCTATCACAC	GGACAGACATCTGCCAAGGT
Hprt	CCAAGACGACCGCATGAGAG	CAACGGAGTGATTGCGCATT
Gapdh	CTCTGCTCCTCCCTGTTCC	TCCCTAGACCCGTACAGTGC

3.2 Quantitative analysis of DNA methylation turnover

3.2.1 Abstract

Transcription factor (TF) binding to specific distal regulatory elements, including enhancers and insulators, is crucial for the activation of the correct gene expression program that determines cell type identity (Takahashi and Yamanaka 2006). Upon occupying their target sites, TFs mediate changes in chromatin structure, which in some cases can facilitate the recruitment of additional factors (Zaret and Carroll 2011). Recent evidence from our and others' work suggests that TF binding can mediate local hypomethylation events through triggering a turnover of DNA methylation at their binding sites (Stadler *et al,* 2011; Feldmann *et al,* 2013; Shen *et al,* 2013; Song *et al,* 2013). The specificity and consequence of this turnover for gene regulation as well as its mechanisms and kinetics are not well understood.

Here we inactivate the *de novo* DNA methyltransferases DNMT3A and DNMT3B in embryonic stem cells. This setup prevents the addition of new methylation marks, thus allowing us to study the maintenance and turnover of DNA methylation exclusively. By using a technique allowing for high coverage single-base resolution methylation analysis of specific genomic loci, we show that simultaneous deletion of DNMT3A/B results in a loss of methylation at both low and fully methylated regions. Following methylation changes upon conditional inactivation of DNMT3A/B we are able to determine the precise methylation turnover kinetics for each individual CpG. This analysis reveals accelerated turnover at low methylated cytosines in comparison to fully methylated cytosines.

In summary, we established a controlled system to study DNA methylation turnover. We show that DNMT3A/B-dependent turnover is present in embryonic stem cells. *De novo* DNMTs appear to be required for long-term maintenance of DNA methylation at both low and fully methylated regions. However, our data demonstrate that the turnover predominantly affects cytosines with low DNA methylation thus confirming the link between TF binding and DNA methylation turnover.

3.2.2 Global loss of methylation in DNMT3A/B double knockout embryonic stem cells

We and others previously identified low methylated regions (LMRs) as sites of DNA methylation turnover (Feldmann *et al,* 2013; Shen *et al,* 2013; Song *et al,* 2013). Here we aimed to confirm its presence and further characterize it. 5-hydroxymethylcytosine is an intermediate of active DNA demethylation. Its further processing was reported to occur via facilitated passive demethylation (Valinluck and Sowers 2007; Inoue and Zhang 2011) or direct removal with or without oxidation (He *et al,* 2011; Ito *et al,* 2011; Chen *et al,* 2012). A complete turnover cycle further includes remethylation. Such remethylation would take place independently of maintenance methylation. Maintenance by DNMT1 requires a hemimethylated substrate DNA (Song J. *et al,* 2011; Song *et al,* 2012). Thus, we assumed that if the observed turnover is active it must depend on *de novo* DNA methyltransferase activity (Figure 3-1A). If this hypothesis is true and LMRs are indeed the main targets of DNA methylation turnover, embryonic stem cells lacking DNMT3A and DNMT3B (Dnmt3ab-/- ES cells) should retain wildtype methylation levels at fully (FMRs) and unmethylated regions (UMRs) but lose methylation predominantly at the LMRs.

To test this model we compared DNA methylation levels between Dnmt3ab-/- ES cells (Okano *et al,* 1999) and the corresponding wildtype cell line J1 by a targeted bisulfite PCR approach (Figure 3-1B). We performed PCR on 90 different regions including 24 FMRs, 52 LMRs and 11 UMRs, of which 65 amplicons were covered to at least 50% in both samples (for summary of all datasets see Table 3-1). For all covered amplicons we calculated average methylation levels and compared them between the two cell lines. To our surprise, we found demethylation at all of the analyzed regions in Dnmt3ab-/- ESC, regardless of their initial methylation status. Differences between segments were observed mainly in the extent of this demethylation. While UMRs and LMRs indeed are almost entirely devoid of methylation in Dnmt3ab-/- ESC, FMRs have heterogeneous methylation which spreads over the entire range from 0 to 80% (Figure 3-1C-D). Thus, we conclude that DNMT3A and DNMT3B are required for methylation maintenance at both LMRs and FMRs.

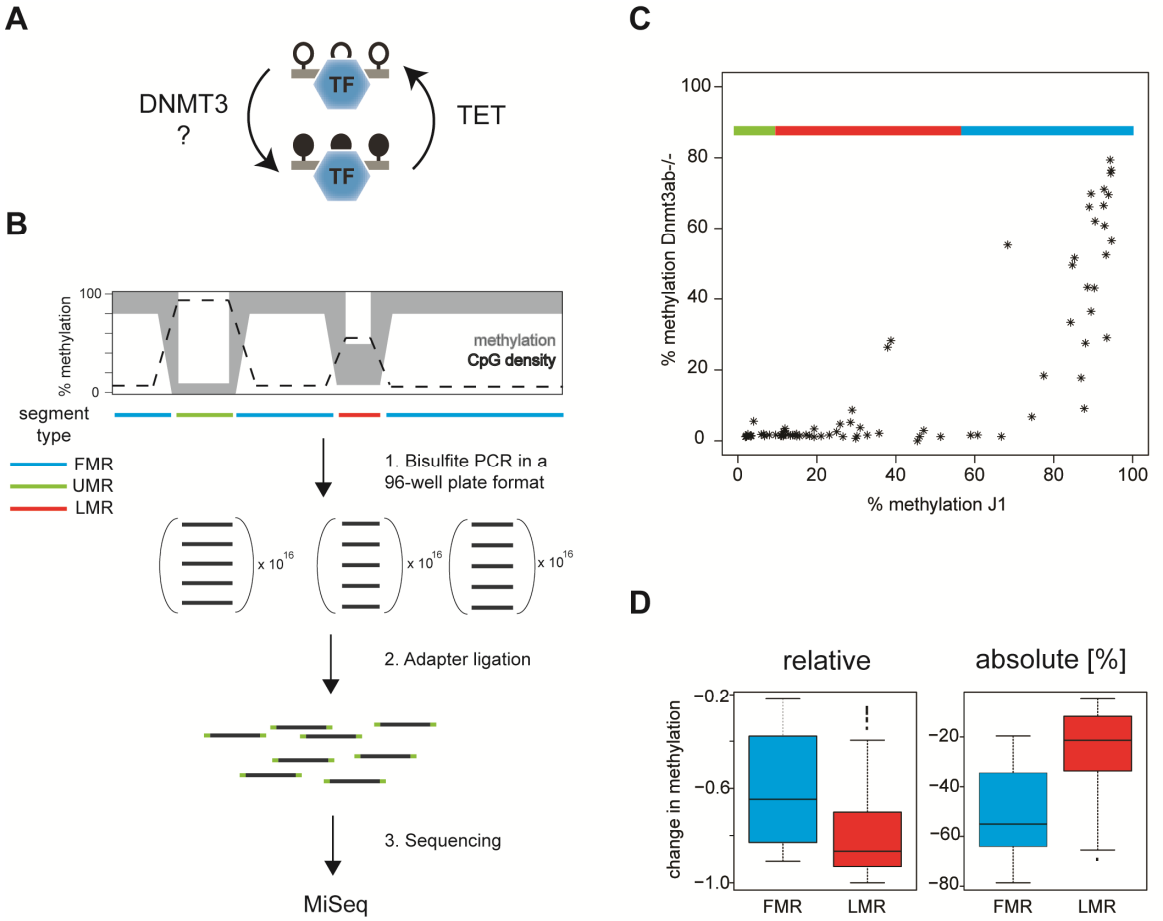

Figure 3-1 Site-specific demethylation in Dnmt3ab-/- ES cells.

(A) In this experiment we tested the dependency of DNA methylation turnover on *de novo* DNA methyltransferases DNMT3A and DNMT3B (DNMT3). (B) Methylation was analyzed using a targeted bisulfite PCR approach, in which regions of interest, residing within LMRs, UMRs or FMRs were amplified from bisulfite converted DNA and subjected to sequencing on the MiSeq platform. (C and D) Fragment methylation was calculated over all cytosines within a fragment covered at least five times in the sequencing sample (C) Comparison of average methylation levels of targeted amplicons between Dnmt3ab-/- ES cells and the corresponding wildtype ES cells J1. All wildtype unmethylated (green) and almost all low methylated fragments (red) are entirely devoid of methylation in DNMT3ab-/- ES cells. Fully methylated fragments also lose their methylation in Dnmt3ab-/- ES cells to a varying degree. (D) Boxplots summarizing changes in DNA methylation between Dnmt3ab-/- and J1. Shown are only fragments overlapping with FMRs and LMRs according to segmentation from Stadler et al. (Stadler *et al,* 2011). Note that all fragments display a loss of methylation in Dnmt3ab-/- ES cells. Loss of methylation was calculated as fraction of methylation loss relative to the wildtype methylation levels (left) and as absolute loss in percentage methylation (right). Note, that while Dnmt3ab-/- ES cells on average lose more than 80% of their starting methylation at the LMRs, their absolute loss only comprises 20% methylation and is smaller than the absolute loss of methylation at FMRs. Boxes show the interquartile range and the median.

To determine whether the turnover is faster within LMRs, we calculated the loss of methylation at LMRs and FMRs using two different methods (Figure 3-1D). As expected, comparison of demethylation normalized to original methylation levels revealed a higher degree of demethylation at LMRs which lost almost 100% of their initial methylation

(Figure 3-1D, left). On the other hand, absolute loss of methylation (i.e. the difference between Dnmt3ab-/- and J1 methylation) is higher at tested FMRs (Figure 3-1D, right). This precludes a direct comparison between the two types of segments.

It is important to note that ES cells used for this experiment have been cultured for 22 passages following deletion of DNMT3A and DNMT3B. Assuming that ES cells divide every 12-18 hours and are transferred on a new plate every second day (Bibel *et al,* 2007), they underwent up to a hundred cell divisions by the time the experiment was performed. This prevents us from drawing conclusions on the timing or kinetics of demethylation at different types of segments.

In summary, our results indicate involvement of DNMT3A and DNMT3B in long-term methylation maintenance at both low and fully methylated regions. However, methylation profiling in stable knockout ES cells does not allow us to ask at which sites the observed demethylation is initiated.

3.2.3 DNMT3A/B dependent DNA demethylation at active regulatory regions

We reasoned that conditional deletion of DNMT3A/B in an ES cell line would allow us to determine the kinetics of demethylation. Therefore, we took advantage of conditional embryonic stem cells. This cell line was derived from mouse embryos containing loxP sites which flank the exons of catalytic domains in both alleles of Dnmt3a and Dnmt3b (Dodge *et al,* 2005). Conditional inactivation of DNMT3A/B was achieved by direct transduction of the HTN-Cre (His-TAT-NLS-Cre) protein (Figure 3-2A). HTN-Cre is tagged with a peptide derived from the human immunodeficiency virus (HIV) transactivator of transcription (TAT) protein and can directly permeabilize the cell membrane (Peitz *et al,* 2002). The advantage of this method is a fast response and a complete lack of leakiness.

Recombination efficiency was measured four days after the transduction by quantitative genotyping and was on average 89.5% and 83.9% for Dnmt3a and Dnmt3b, respectively (Figure 3-2B). When we compared methylation at individual cytosines between ES cells transduced with Cre and control cells, we found that indeed inactivation of DNMT3A/B led to a mild decrease of methylation after four days. Importantly, the loss of methylation was limited to cytosines with low to intermediate methylation levels (Figure 3-2B, day4).

Encouraged by these results, we decided to perform a time course over 15 days. Approximately 10-15% of ES cells in the analyzed population still contained intact

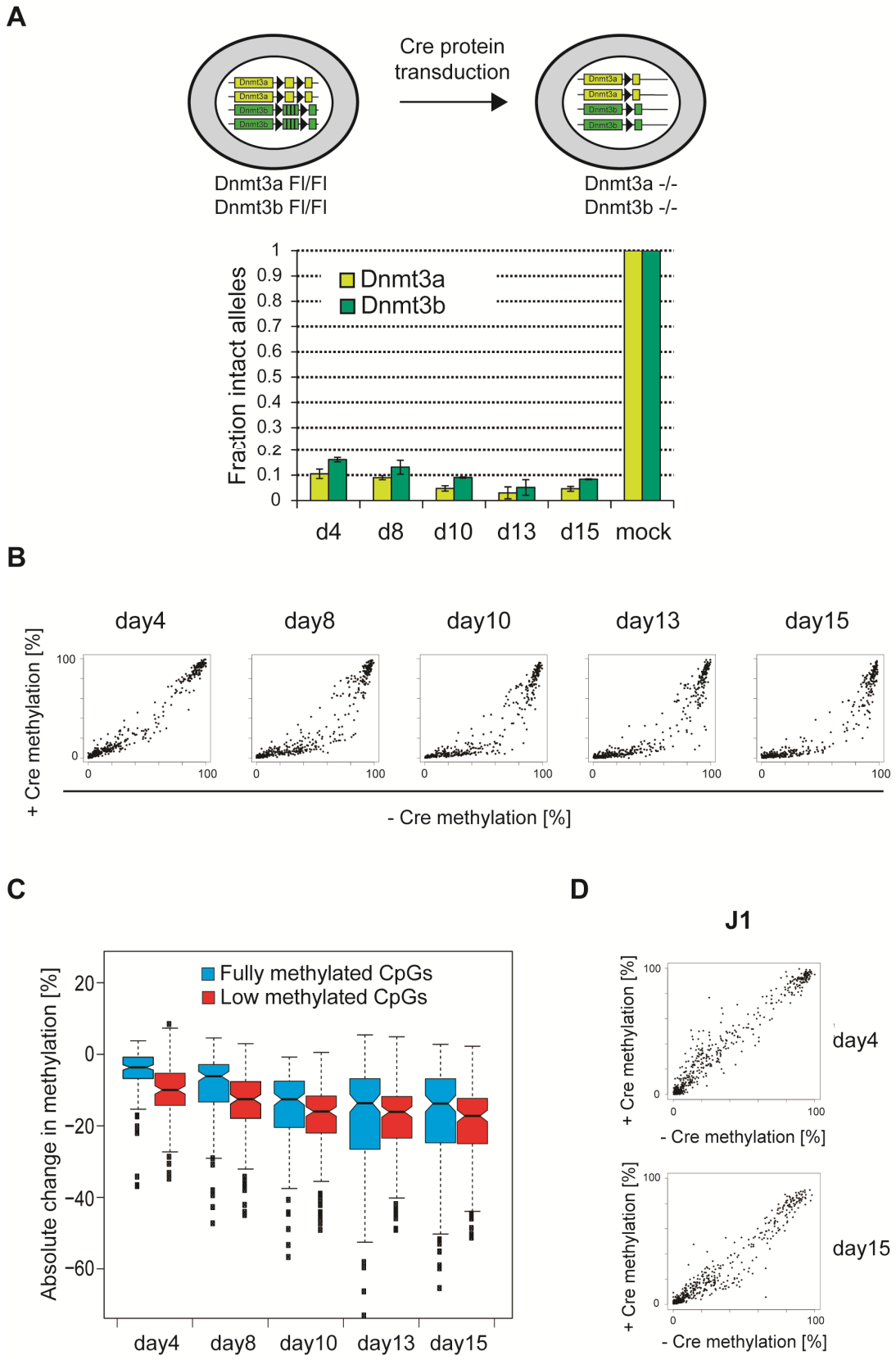

Figure 3-2 Conditional inactivation of DNMT3A/B in ES cells.

(legend continues on the next page)

Figure 3-2 Conditional inactivation of DNMT3A/B in ES cells.

(A) Upper panel: schematic representation of the conditional deletion alleles for Dnmt3a (yellow) and Dnmt3b (green). LoxP sites (black triangles) flank one catalytic exon of the Dnmt3a allele and four of the Dnmt3b allele. The ES cells are homozygous for loxP sites in Dnmt3a and Dnmt3b genes. Please, note that the distances are not true to scale and by far not all exons are presented here. Upon transduction of the HTN-Cre protein the regions flanked by loxP sites are deleted. Bottom panel: the amount of intact alleles in total population of cells was quantified by TaqMan PCR 4, 8, 10, 13 and 15 days after transduction of Cre. Average of two biological replicates is shown. The error bars represent standard deviation. Mock: negative control transduced with Cre buffer alone. (B-D) Only CpGs with coverage of at least ten fold in all presented samples are shown. (B) Correlation of methylation between Cre (+Cre) and corresponding mock transduced (-Cre) conditional knockout ES cells at indicated time points after transduction. Each point represents a single CpG. (C) Absolute loss of methylation relative to the reference sample (mock at day 4) for each time point. For this plot the CpGs were subdivided into fully methylated (60-100% methylation in the reference sample, blue) and low methylated (10-60% methylation in the reference sample, red). The boundaries of the boxes represent the interquartile range and the middle line shows the median. (D) Correlation of methylation between Cre (+Cre) and mock transduced (-Cre) J1 cells. Each point represents a cytosine within a CpG dinucleotide. Note that the level of demethylation achieved after 15 days is much smaller than for the conditional knockout ES cell line.

Dnmt3a/b alleles four days after the Cre transduction. However, these cells do not have a growth advantage in our time course, since the proportion of Dnmt3ab-/- ES cells remained stable and even increased mildly during culturing (Figure 3-2A).

Methylation analysis of the 474 cytosines covered at least ten times in all samples showed that low methylated cytosines started to progressively demethylate before day four after the transduction of Cre. A plateau was reached at day ten, when presumably all methylation was lost from these regions (Figure 3-2B-C). Cytosines with 60-100% methylation, characteristic of fully methylated regions, started to progressively demethylate between day four and day eight post-transduction. This suggests a better capability to maintain their methylation in the absence of DNMT3A/B activity. After 13 days, absolute loss of methylation at fully methylated cytosines almost reached the level observed at low methylated cytosines (Figure 3-2C). Importantly, the effect on DNA methylation was specific to Cre-induced deletion, as Cre transduction in the wildtype cell line J1 did not cause strong demethylation even after 15 days (Figure 3-2D).

We considered the possibility that an increase in the fraction of cells with a deletion in Dnmt3a/b throughout the time course might explain the observed demethylation (Figure 3-2A). Assuming that each Dnmt3ab-/- cell is fully unmethylated and replaces a fully methylated cell in the total population, this would account for a maximum of 5% change in methylation over the time course. This is below the observed changes in methylation and thus does not affect the here measured methylation. Our results from individual CpGs suggest that the turnover is faster at cytosines with low methylation levels.

Analysis of methylation changes within single amplicons across the time course extended our observations from single cytosines to entire regions. While UMRs had stable methylation in ES cells transduced with HTN-Cre, LMRs and FMRs both displayed demethylation. We furthermore note that in the absence of Cre ES cells tend to upregulate their methylation probably as a consequence of prolonged feeder-free culturing (mock in Figure 3-3, and data not shown).

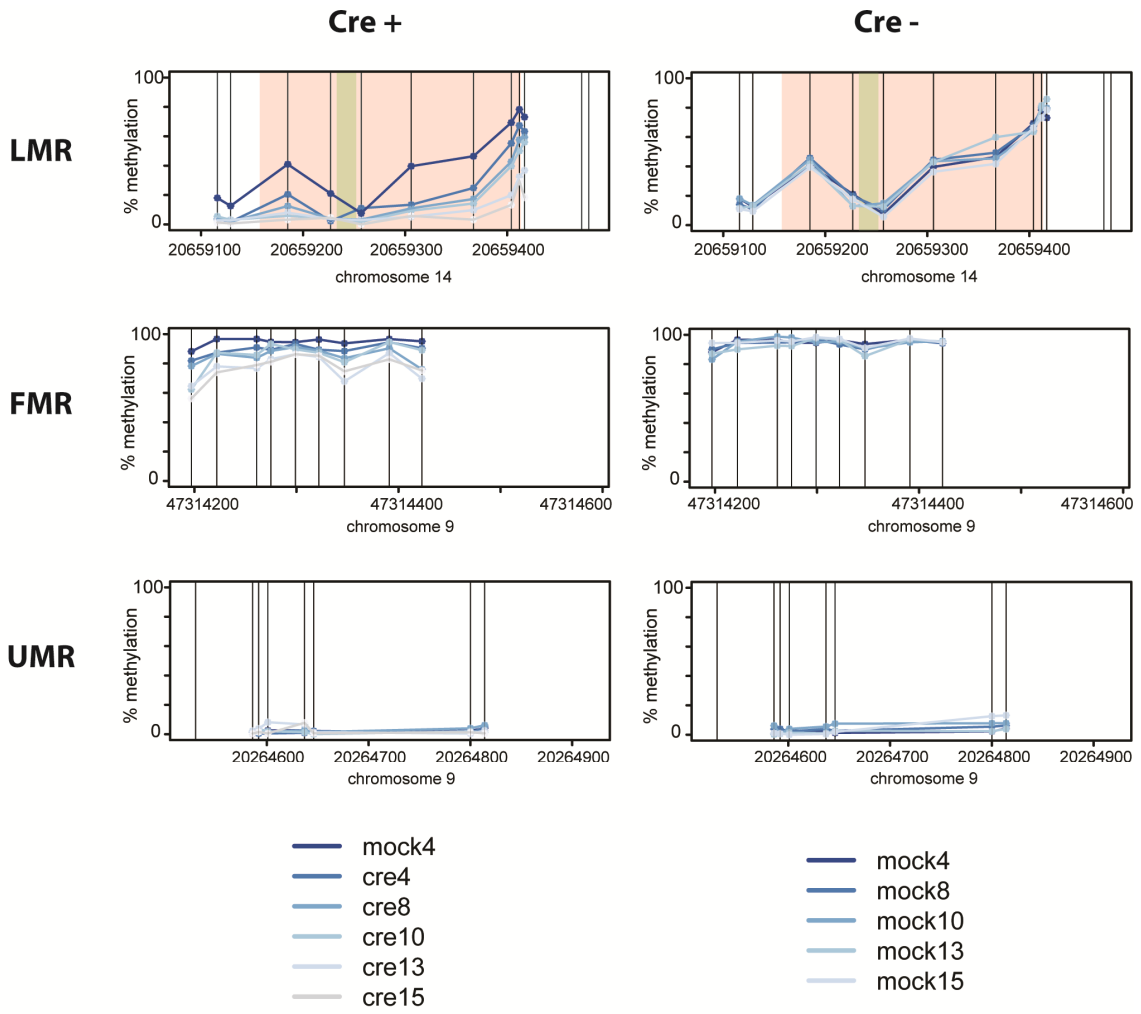

Figure 3-3 Time course of Cre transduction and Mock transduction in representative amplicons.

Note that LMRs and FMRs display high demethylation only upon Cre transduction (Cre +). Some cytosines upregulate their methylation during culturing in the absence of Cre (Cre-). Reference sample for both representations is mock transduction at day four (mock4). The nature (cre/mock) and duration in days (4-15) of the treatment are indicated below the panels. Vertical lack lines mark the position of single CpGs. Red shade: Position of the LMR, green shade: position of a bound CTCF motif. Only cytosines with at least ten fold coverage in all samples are shown.

We asked whether demethylation of a region occurs homogenously or whether it is initiated by certain cytosines. Our targeted PCR approach allowed for analysis of up to 300 basepair long reads thereby enabling us to determine the similarity between cytosines at each time point and detect co-regulatory patterns. This analysis revealed that the heterogeneity in methylation is generally low throughout the time course. Thus, even if demethylation is pioneered by a single cytosine this is not detectable by the here applied resolution (Figure 3-4).

Together, we confirmed the presence of a DNMT3–dependent turnover at low methylated regions. However, in contrast to our expectations, this turnover also affects regions with full methylation.

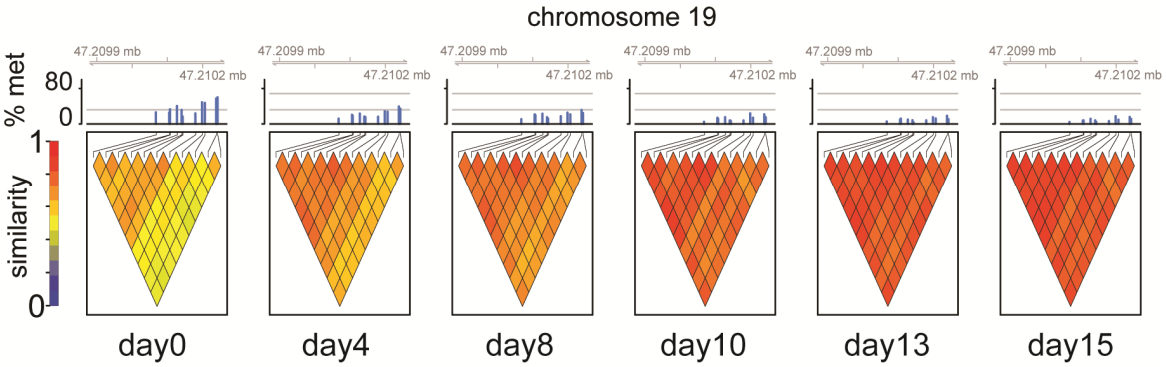

Figure 3-4 Heterogeneity analysis for CpGs within a representative LMR.

Shown are average methylation for each single cytosine (blue bars in the upper panel) and similarity between cytosine pairs (heatmaps). For each pair of CpGs the similarity in methylation was calculated as fraction of reads in which the two cytosines have the same methylation state. Note that the similarity between CpGs never drops below 0.5 and correlates with the degree of demethylation. There is furthermore no focal demethylation detectable.

3.2.4 Unbiased turnover quantification reveals its predominant targeting to cytosines with low methylation

For a better comparison between different types of regions, we next aimed to directly quantify the turnover kinetics and relate it to the level of starting methylation. Therefore, we first determined the methylation decay rate for each individual cytosine in the sample (for details see 3.2.5). To ensure that our methylation measurement is robust, only cytosines covered at least a 100 fold in all samples were considered for this analysis. We performed a non-linear regression for each CpG, assuming an exponential decrease in

methylation (Figure 3-5A). To test whether inferred parameters approximate our measurements, we first correlated the measured methylation with the inferred methylation met(t0) for each cytosine at each time point in the time course (Figure 3-5B). This revealed a very high correlation (r=1 and r=0.98 for replicate 1 and replicate 2, respectively), suggesting a high accuracy of the estimated parameters.

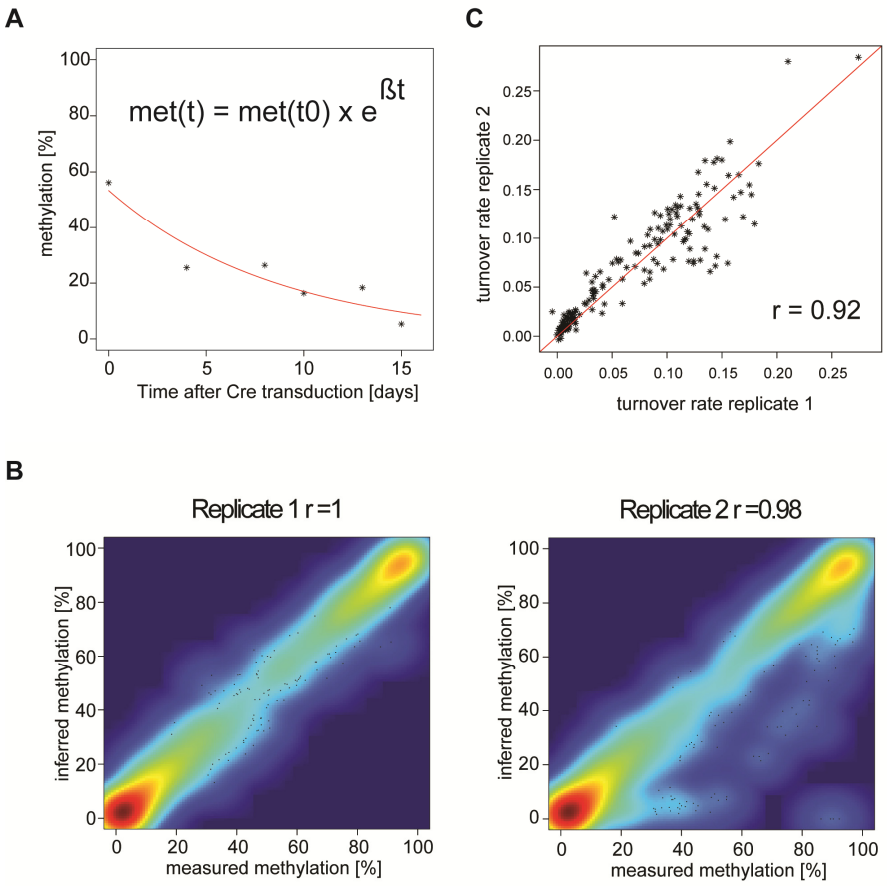

Figure 3-5 Determination of methylation turnover rate

(A) Shown are methylation levels for one representative CpG across the time course. For each cytosine a non-linear regression was performed assuming exponential methylation decay. Methylation at time point t=0 (met(t0)) and the turnover coefficient ß were estimated. Only CpGs covered at least a 100 times were used in this analysis. (B) Methylation for each cytosine and each time point was calculated using met(t0) and ß estimated in A. This inferred methylation rate was correlated with measured methylation for two biological replicates. Each point represents one CpG at one time point during the time course. (C) Correlation of turnover rates (for calculation see 3.2.5) between the two biological replicates. Cytosines with a starting methylation below 10% and a methylation below 5% at four days after Cre transduction were omitted. Each point represents one cytosine.

Next, we asked whether the observed turnover is reproducible across biological replicates. For cytosines below 10% methylation accurate determination of decay rate

cannot be achieved, as a small methylation change would introduce noise in turnover estimation (Figure 3-3 and data not shown). This noise is illustrated in the high spread of turnover rates observed between these cytosines (Figure 3-7). Furthermore, the mean variance of cytosine turnover within UMR amplicons is more than six times higher than the average turnover variance between UMR fragments (8×10^{-3} and 1.3×10^{-3}, respectively). Similar to UMRs, our approach is not suitable for cytosines with a very fast turnover like those entirely losing their methylation four days after Cre transduction. For these cytosines the exact time point of full demethylation cannot be determined and thus their turnover rates would be underestimated. Therefore, we excluded all cytosines with a methylation below 10% at the starting time point and below 5% four days after Cre transduction from further data analysis. For the remaining 168 cytosines covered at least 100 fold in all time points in both experiments we correlated turnover rates from two biological replicates. This revealed a correlation of 0.92, suggesting that the turnover is reproducible (Figure 3-5C, 3-6B). We note, however, that despite this generally high correlation some spread persists for cytosines with a fast turnover.

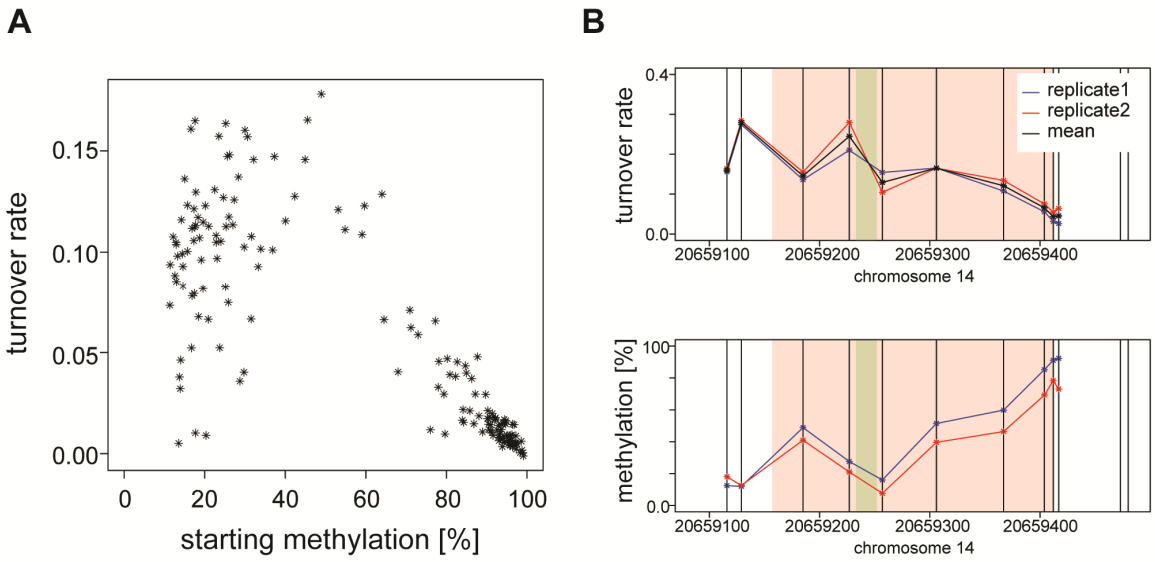

Figure 3-6 Relationship between turnover rate and methylation.

(A) Correlation of turnover rate with methylation reveals higher turnover rates at low methylated cytosines. Shown is the mean turnover rate between two biological replicates. Each point represents one CpG. Only CpGs with coverage of at least 100 in both replicates are shown. Cytosines with a starting methylation of below 10% and methylation of below 5% four days after Cre transduction were omitted. (B) Turnover rate (upper image) and starting methylation (bottom image) in a representative LMR. Blue and red lines show replicate1 and 2, respectively. Black line represents the mean turnover rate of two replicates. Vertical black lines mark CpG positions within the amplicon. Red shade marks the position of the LMR according to previous segmentation (Stadler *et al,* 2011). Green shade: position of a bound CTCF site. For each replicate only CpGs with coverage of at least 100 are shown.

Having this quantitative readout, we decided to reevaluate the relationship between the turnover rates and the starting methylation level of analyzed cytosines. Indeed, within our data-set high turnover rates occur at cytosines with low starting methylation levels, while turnover rates at fully methylated CpGs linearly increase with decreasing methylation (Figure 3-6A). Importantly, this observation is not limited to cytosines residing in different regions, as CpGs within the same genomic region can readily differ in their turnover rates (Figure 3-6B). Overall, little difference was observed in the variance of turnover within and between regions (2.33 x 10^{-3} and 2.00 x 10^{-3}, respectively).

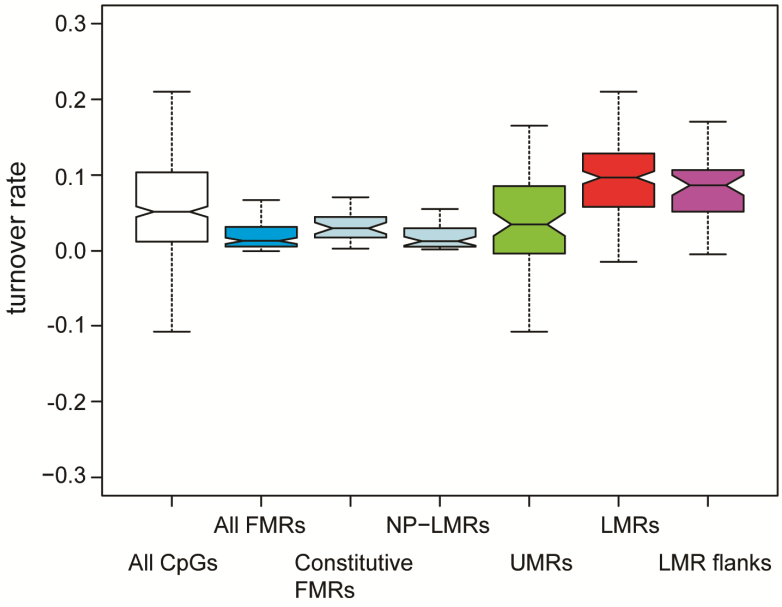

Figure 3-7 Turnover rates within different segments

Turnover rates at all analyzed CpGs (white), CpGs within FMRs (blue), UMRs (green) and LMRs (red) for all individual cytosines covered at least 100 times. Note that constitutive FMRs (light blue, left) do not display lower turnover than NP-specific LMRs which are FMRs in ES cells (light blue, right). The strong almost uniform variation in turnover at UMRs probably reflects measurement limitations at very low methylated cytosines. Note that cytosines within FMRs flanking LMR regions (purple) generally have a higher turnover rate than average FMRs (blue). Boxes show the median and the interquartile range.

Comparison of methylation turnover rates at all analyzed cytosines which reside within LMRs, FMRs or UMRs furthermore shows that turnover is fastest within LMRs (Figure 3-7). Interestingly, FMRs directly adjacent to LMRs in our dataset display enhanced turnover kinetics, suggesting that their methylation could be influenced by TF binding within the LMR regions. No increased methylation turnover could be detected at cell type

specific FMRs which have high hydroxymethylation in neuronal progenitors (NP- LMRs in Figure 3-7) as compared to constitutive FMRs (Constitutive FMRs in Figure 3-7).

In summary, unbiased quantification of turnover reveals its accelerated kinetics at low methylated regions and individual low methylated cytosines.

dataset	total reads	conversion efficiency [%]	protection efficiency [%]	CpGs cov>9	Fragments cov>9
3ab -/- ESC rep1	5415260	98.1	97.6	509	82
3ab -/- ESC rep2	6165442	99.5	99.9	431	69
J1 ESC rep1	5687936	98.2	97.7	525	82
J1 ESC rep2	6368876	99.7	100.0	419	70
CreTD1-d4	15440206	97.8	99.4	502	79
CreTD1-d8	3733392	99.8	99.9	479	75
CreTD1-d10	3728962	98.1	99.7	478	76
CreTD1-d13	4174414	99.3	99.8	476	75
CreTD1-d15	3787318	99.1	99.8	500	80
CreTD2-d4	3358126	99.3	100	447	72
CreTD2-d8	3234086	98.6	99.9	520	81
CreTD2-d10	3245980	98.4	99.9	480	77
CreTD2-d13	2838246	92.3	99.9	480	77
CreTD2-d15	2837296	97.9	99.6	495	78
MockTD1-d4	10003876	99.4	98.2	503	79
MockTD1-d8	4049724	73.5	95.8	504	80
MockTD1-d13	4005596	97.0	99.8	496	79
MockTD1-d15	3194158	99.8	99.6	521	82
MockTD2-d4	3246084	98.8	99.8	495	79
MockTD2-d8	4248638	99.1	99.9	532	82
MockTD2-d10	2723552	95.1	99.9	490	78
MockTD2-d13	2931254	98.5	97.4	509	80
MockTD2-d15	2923594	98.3	98.9	486	76
CreTD-J1-d4	3309530	99.3	99.9	508	79
CreTD-J1-d15	3778564	94.8	99.6	515	79
MockTD-J1-d4	3383836	97.5	99.9	503	79
MockTD-J1-d15	2923594	97.5	99.9	501	78

Table 3-1 Summary of all datasets used in 3.2

TD1: transduction replicate 1; TD2: transduction replicate 2; ESC: embryonic stem cells; d: day; 3ab-/-: Dnmt3ab-/-; rep1/2: replicate1/2; Cre: cre transduction; Mock: control transduction with Cre buffer; cov: coverage

3.2.5 Materials and Methods

Targeted amplicon sequencing

Genomic DNA was isolated from ES cell pellets. For elimination of feeders, the trypsinized ES cell-feeder mix was allowed to settle on a plate for 20-30 minutes before collecting the supernatant. Genomic DNA was isolated as previously described (Mohn *et al,* 2009). Briefly, cell pellets were resuspent in TE and equal volume of lysis buffer containing 20µl of proteinase K (10mg/ml) prior to incubation for at least 5hrs at 55°C. DNA was phenol-chloroform extracted, precipitated with ethanol containing 75mM sodium acetate pH5.2, washed in 70% ethanol and the pellets allowed to dry at RT for 5-10 minutes. Extracted genomic DNA was resuspent in TE buffer containing 20µg/ml RNAse A and incubated for 30min at 37°C while slowly shaking. DNA concentration was determined using Nanodrop (ND-1000 Spectrophotometer, Witec AG), a mean of two measurements was taken. 2-14µg of genomic DNA were spiked with unmethylated lambda DNA and *in vitro* premethylated T7 DNA (1.6pg/µg DNA) and bisulfite converted according to the manufacturer's protocol of the EpiTect Bisulfite kit (QIAgen, #59104), using 1-2µg per conversion reaction. The conversion reactions from one genomic DNA were pooled and used for targeted bisulfite PCR. Targeted bisulfite PCR was prepared in a 96-well plate format with primers described in Table 3-5. For one 25µl reaction mix 1/105 of the converted genomic DNA were mixed with 5µl of a 2µM mix of forward and reverse primers, 2.5µl 10x PCR buffer, 1.25µl DMSO, 1.5µl of 25mM $MgCl_2$, 1.5µl of 2.5mM dNTPs and 0.25µl of 5U/µl AmpliTaq Gold polymerase (Applied Biosystems, # N-8080249) and subjected to PCR according to Table 3-1. Equal volumes of each amplicon were pooled and 100µl of the total mix extracted from a 1.5% agarose gel.

Libraries from extracted amplicon PCR mix were prepared according to the manufacturer's protocol for NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (New England BioLabs, #E6240) for multiplexed libraries. Briefly, the concentration of each sample was measured using NanoDrop 3300 Fluorospectrometer (Witec AG). The samples were end-repaired, dA-tailed and adapter ligated using NEBNext Multiplex Oligos for Illumina (New England BioLabs, #E7335). Following this they were size selected using Agencourt AMPure XP beads (Beckman Coulter, # A63880) and PCR amplified for 12 cycles using appropriate indexed primer for each adapter and cycling conditions according to Illumina recommendations. Adapter-ligated and amplified DNA was eluted using AMPure XP beads, the size distribution was analyzed on Agilent

Bioanalyzer 2100 using Agilent High Sensitivity DNA kit (Agilent technologies, #5067-4626). For multiplex sequencing, libraries were pooled in an equimolar ratio.

Sequencing was performed on Illumina MiSeq, using MiSeq v2 reagent kits 300bp PE (Illumina, #MS-102-2002) according to Illumina recommendations.

Step	T [°C]	t [min:s]
1	95	09:00
2	95	00:30
3	55 to 51	00:30
4	decrease by 0.2°C each cycle	
5	72	00:30
6	repeat steps 2-4 19x	
7	95	00:30
8	51	00:30
9	72	00:30
10	repeat steps 7-9 35x	
11	4	hold

Table 3-2 Bisulfite PCR conditions

Cre protein transduction

ES cells were cultured as previously described (Bibel *et al,* 2007), passaged at least once on feeders prior to trypsinization for Cre protein transduction. Transduction was performed as described (Haupt *et al,* 2007). Briefly, ES cells were trypsinized, resuspent in PBS and quantified. 2.5 x 10^5 cells were transferred into fresh falcon tubes, spinned down and resuspent in 500µl of filtered serum-free medium (Table 3-2) containing either 2µM Cre protein or an equivalent volume of Cre dialysis buffer (2M NaCl, 50mM HEPES pH7.4, 1mM DTT, 1mM EDTA and 5% Glycerol). The cells were plated in 24-well plates pre-coated with feeders and washed twice with PBS. After 16hrs cells were washed twice with PBS and coated with FCS-based ES medium (Bibel *et al,* 2007). ES cells were transferred to gelatin-coated 6-well plates 24hrs and to 10cm plates 72hrs after transduction. Pellets were collected from trypsinized cells at indicated time points and culturing was continued until 15 days post-transduction in feeder-free environment. All

ES cells used for Cre transduction experiments have been cultured in total for five to seven passages prior to Cre transduction.

Reagent	Company	Cat.-No	Stock conc.	End conc.
DMEM/F-12 with HEPES	Life Technologies	31330-038	1x	0.5x
Neurobasal	Life Technologies	21103-049	1x	0.5x
N2 supplement	Life Technologies	17502-048	100x	1x
B27 supplement	Life Technologies	17504-044	50x	1x
MEM Non-Essential amino acids	Life Technologies	11140-050	10mM	100µM
L-glutamin	in house preparation		200mM	2mM
ß-mercaptoethanol	Sigma	M6250	100%	0.01%
LIF	in house preparation		-	-

Table 3-3 Serum-free medium composition

Quantitative genotyping

TaqMan primers and probes were designed using the Primer Express software (Applied Biosystems, v3.0.1) to detect genomic DNA between the two loxP sites and are listed in Table 3-6.

Genomic DNA was diluted to 30ng/µl and subjected to singleplex quantitative PCR using the StepOne Plus cycler (Applied Biosystems) according to conditions described in Tables 3-3 and 3-4. Each PCR was performed in triplicates. For the analysis the amount of template DNA for each PCR was quantified using the absolute standard curve and normalized to Gapdh as reference template. Amount of molecules with intact catalytic exons of Dnmt3a and Dnmt3b was quantified relative to genomic DNA from untreated cells collected at the same time points. Standards with 100%, 20% and 0% of floxed Dnmt3a and Dnmt3b genomic DNA mixed with genomic DNA from an ES cell clone with fully deleted alleles were quantified as controls in each experiment.

reagents	starting concentration (nM)	final concentration (nM)	Volume (µl) per well
TaqMan Universal PCR Master Mix (Applied Biosystems, # 4304437)	2x	1x	12.5
Fwd and Rev primer mix	50000	900	0.45
TaqMan probe	10000	250	0.625
DNA sample	30 ng/ul	75 ng	2.5
ddH2O		-	8.925
total		-	25

Table 3-4 TaqMan PCR reaction mix

Step	T [°C]	t [min:s]
1	50	02:00
2	95	10:00
3	95	00:15
4	60	01:00
5	repeat 3-4 40x	

Table 3-5 TaqMan PCR conditions

Data analysis

The *M. musculus* genome assembly from July 2007 (NCBI37/mm9) provided by NCBI (http://www.ncbi.nlm.nih.gov/genome/guide/mouse/) and the Mouse Genome Sequencing Consortium (http://www.sanger.ac.uk/Projects/M_musculus/) was used as reference genome for all alignments. Alignment of bisulfite sequencing reads was performed using the software package Bismark v0.6.beta2 (Krueger and Andrews 2011) together with Bowtie-0.12.7 (Langmead *et al,* 2009) with the following parameters: -n 3 --non_directional. Sequencing data generated for read 1 and read 2 were aligned separately. Methylation was extracted by Bismark v0.6.beta2 methylation extractor with the parameters --comprehensive –s. Conversion efficiency of unmethylated and the protection efficiency of methylated CpGs were determined by quantification of methylation obtained from spiked-in unmethylated lambda and methylated T7 DNA, respectively (see Table 6-1). To ensure that only fully converted reads are taken into

account for the analysis, the conversion efficiency of non-CpGs within each read was required to be at least 80% to pass the filter. Absolute loss of methylation was calculated as a difference between Dnmt3ab-/- and corresponding wildtype methylation (met(Dnmt3ab-/-) - met(wt)). Relative loss of methylation was calculated as (met(Dnmt3ab-/-) - met(wt)) / met(wt). For similarity analysis for each pair of cytosines the similarity score was quantified as fraction of reads with equal methylation for these two cytosines: reads(equal methylation) / reads(total).

Quantification of turnover kinetics

The turnover was quantified in R version 3.0.2 (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/) using non-linear regression and assuming an exponential decay after observation of all data points. For each cytosine covered at least a 100 times in the datasets a curve was fitted for the equation $met(t) = met(t_0) \times e^{t\beta}$, where t is the time after Cre transduction in days, met is the methylation and ß the turnover coefficient. As reference methylation, the methylation of mock transduced ES cells at t=4 was taken to ensure reproducibility between replicates. The turnover coefficients were estimated for each time course for Cre ($ß_{Cre}$) and the Cre dialysis buffer ($ß_{mock}$) transduced samples. Final turnover rates were calculated by normalizing the turnover coefficients to respective Mock samples and multiplication by -1 to obtain positive turnover rates: $(-1) \times (ß_{Cre} - ß_{mock})$.

Additional Materials

- Multiplate PCR plates 96-well clear (BioRad, #MLP9601)
- PCR salers Microseal ‘B’ Film (BioRad, # MSB1001)
- PCR machines: C1000 Touch Thermal cycler/iCycler (BioRad)
- MicroAmp Fast Optical 96-well reaction plate (Applied Biosystems, # 4346906)
- MicroAmp Fast Optical Adhesive Film (Applied Biosystems # 4311971)

Oligonucleotide sequences:

Region	Forward Primer	Reverse Primer
FMR1	AGTTGTTAGGATTTGAATTTTGGT	CCTCTACTCCTTCTTTCCTAATACA
FMR2	AGGATGGATGTGTTATGTTTTAGT	AAATCTACCTTTCCTTCCAACA
FMR3	TTTTATTGATTGTTATGTGGTGTTTT	ACAACTCCTCTTCTCCAACA
FMR4	AGATGTTTGTTTAGTTTTGGGTT	CCAAAACCCTAACAATCCCC
FMR5	GGGAGGTAGGGGTAGTAAGA	ACACACACACACACACAATA
FMR6	AGAGATTGGTGGGTTGGATT	ACCACCACAAAACAAATACCT
FMR7	GGGAAGTTGAGGTAGGTAGG	CAACCAACCAACCAAAACCT
FMR8	GGGTTTTGTAGGGTGTGAGA	CCACTACCACATCACAATTCC
FMR9	AGTGGAGTGGTGTAGAGGAT	ACCTTAAACCTCTCTCAAAACA
UMR1	GGTTTTGATGGTTGAGGTGT	TTCCCAATCCCCATTTCTCC
UMR2	AGGATTGTTTGGGATGGAAAA	CAAACTCAACCCAACCAACC
UMR3	AGTTAAAGAATGAAATTGAAGTTTGAA	TCCTCTTCATTTTCCCCATCT
UMR4	AGGGATTAGTAGGAAAGGAGTT	CACCTTCCACCCCTCTATTA
UMR5	AGGTATGAGAGTTAGAAATTAAGAGG	AACAACTATACCCACAAATCTCT
UMR6	GTTTTGGTATTTAAGAAAGGTTAGGG	AATTCCCCAACCATTCACCT
UMR7	GGGATTGTTGGGAGGGATAG	CCAAAACAACCAAAACTACACA
UMR8	AGGAGTTAATGAGGGAGAATAAGA	AAACCCCTCCTCCAAAACTC
UMR9	AGTTTTGGTTAATGAAGTAGGAGA	CCCTCATTCCTAACCCCAAT
UMR10	TGGAGGGGAAAAGGGAAAAT	ACACAACAACTACATCAACTAAACT
UMR11	AAGGTTTTGAGGTAATTGAGTGA	TCCCATCTATCTCCTCCACC
metIsland1	GTTGTTAGGGTTAGGTTTTGATT	TCCTATTACTCCCAACAATACCA
Methylated_UMR1	TGGGGTAGAAAAGTTGTTTAGT	ACCACCAAACATAACACACA
Methylated_UMR2	TGTGGGAAAGGTAGTAATAAAATAGA	AACCAACAAACTATCTCATACCA
DMR1	GATTTGGTGGTTGGGAGTTG	AAACTAAACAAACCACCTCAAAA
DMR2	GGTTTTAGAAAGTTGTTTTATTTTGGG	TTCACATCAAAACAACACCTCA
DMR3	AGATGGTGATAGGGGAGAAAA	TCACCCAAATTCAATACCTCAA
ES-specific_LMR1	GGTGGAGGTGGTTTAAAGGT	TACCCAAAACCACCCTAACC
ES-specific_LMR2	TTTAAGATAAGTTGTTGTTGGGTT	TCCTAACCAAAATCCTAAATACCT
ES-specific_LMR3	TGGGATTTGAGATTGTATTAGTTAGG	CAAAACAAATCCCTATCCTCTAAC
ES-specific_LMR4	GGGGTTGGGTAATAGATGGT	ACAATCACACATCAAACCCT
ES-specific_LMR5	TTGTTATTAAGTTGGAGTGGGT	CCATCCACTTATCTCCCACA
ES-specific_LMR6	GGTAGAGTGTTTTAGTTAAATTAAGGG	AACTACCATCCATCCACTCC
ES-specific_LMR7	AGTGTTTAGGTGTATATTAGGAGGT	ACAAAACCCTACCTACTCCT
ES-specific_LMR8	AGGGAGATGATAGATTAGGTGAT	ACCTTCCACTATCCCTACTCA
ES-specific_LMR9	ATTATGTGAGTTAAGATGGGTGT	ACATAAACTTACTTAACCTTATACCCA
NP-specific_LMR1	GGGAGGTAGAGTTGGATTAGTAAA	ACTCCCTATTACCAACTACAATTT
NP-specific_LMR2	TGAGTGGTTTTGTTTGTGAGG	ACTCCCAAACTTTCTTCTATCAC

NP-specific_LMR3	TGTTGGAAGTTGATATATTGTAGTTGA	ACCTCAAACTCAACTCACACT
NP-specific_LMR4	TGGTTTTAGTTTAAGAAAAGGAAAGT	TTCACTTCATTTACTCCTCTCTT
NP-specific_LMR5	TTGGAGGGAGTAGGGGAG	ACCCAATCAACAATATTACATATCCA
NP-specific_LMR6	GGTTGAGTTTAAATAGAGGTTAGGG	ACTTCTATTTCCACTAAACCTACA
NP-specific_LMR7	AGAGTAAGGTTTTGAGGTGAGT	CCCAACCTCTTAACTTCCCA
NP-specific_LMR8	TGATGGGAGAGAAAGAGTGAG	ACTCTCCAATTCATTTAATAAAACTCT
NP-specific_LMR9	AGGTTATTTTAGAGGTTTGTTAGGT	ACATCACAAACCCTTTTCAAAA
NP-specific_LMR10	TGTGAGAAGGTAAGAGGTGTG	CCTATCAAACTAACCAACTACCT
NP-specific_LMR11	AGGTAGTTGGTTAGTTTGATAGG	ACAAAACAAAACAATACCAACCA
NP-specific_LMR12	TGGTTGAGTAATGAGATAGGTTT	ACCCCATAATTATCTCAAATCTCA
Constitutive_LMR-CTCF1	TGTTTTGGTATGAAAGTTTTGGT	CCTCAACCTAACCTAAACCCA
Constitutive_LMR-CTCF2	AGTTTTGTTTTGTATTTGGTTGTTAA	ACTCAATCATTTCCATTCCAAAA
Constitutive_LMR-CTCF3	TGGGGAGGGATGTGGTATAA	ACTTCACTTCCACCTAAAACTT
Constitutive_LMR-CTCF4	TGGGAGAGGAAGTGTGTTTT	ATCAACAACCACCTCCAAAA
Constitutive_LMR-CTCF5	AAGGTAAGTTTGATTTAGAGAATTGA	ACCACTATCCAAACCCAAACT
Constitutive_LMR-CTCF6	AATAGTAGAGGTGGATTTGATTATAGA	CAAACCACACTAAACCTCACA
Constitutive_LMR-CTCF7	GGTTATGTTATTGTAGTGAGTGGT	ACTCTCAACAACCAATACTCCA
Constitutive_LMR-CTCF8	GGTTATGTTATTGTAGTGAGTGGT	ACTCTCAACAACCAATACTCCA
Constitutive_LMR-REST1	TGTAGTTTGGAATTAGAAGTGTTATT	CTCTAAATCTAAACTCTCTATTCAACA
Constitutive_LMR-REST2	AGAGAGTTGAGATTAGAGGGGA	ACTCAACTCCACAACCAAAC
Constitutive_LMR-REST3	TGTTAGGAGTGTAATAGTTAAGTGG	ACCAAAATTCAAACCCCAACA
Constitutive_LMR-REST4	GGGGTAGTAAGATAAATAGTAGGGA	TCTAACTACATAACCTCAAACCAA
Constitutive_LMR-REST5	AGTGTTGGTAGTAGGTATTGGT	ACCTCTAATAACAAAATTACTCAACA
Constitutive_LMR-REST6	TTTAGGATTAGGGATAGTAGTAAAGTT	ACCTTCCAACTCCCAAACAT
Constitutive_LMR-REST7	GAATTGTAGGGAAAAGGTGAGT	ACACCTCAAATTTCAACACCA
Constitutive_LMR-REST8	AAGTTTGTTAAAATGAGATTAGGATTG	ACCTATTATAAACTCCAACCTACAA
Constitutive_LMR-REST9	AGGATGGTGTTGAAAATTGTTATT	CCCTACTTATAATACTCCTTAAACAAA
Constitutive_LMR-REST10	TTGGGGAAGGTTTGTTGGTT	CTCTCTCAACCTTACTTCCAAAA
Constitutive_LMR-REST11	GTGGAGATAATTGTTTTAGTGTTTGA	ACCACAACTAACATTCCCCA
Constitutive_LMR-REST12	TTGGGGAAGGTTTGTTGGTT	CTCTCTCAACCTTACTTCCAAAA
Constitutive_LMR-REST13	TGTTGTATTTTGGTTTAGTGGTTTG	ACACCTAAACTTTCAATCAACCA
Constitutive_LMR-REST14	TGGTTAGGGGTAAGGTTGTG	AACCACAAACCCAACAATCC
Constitutive_LMR-REST15	AGAATTGTAGGGAAAATGTGAGT	CCACACCTCAAATTCCAACA
Constitutive_LMR-REST16	GTGGAGATAATTGTTTTAGTGTTTGA	ACCACAACTAACATTCCCCA
Constitutive_LMR-REST17	GGAATGGTTTTGGTTGAGGT	CCAATACCTACCAAACAACCA
Constitutive_LMR-REST18	TGTGTGAGGTTTGGTATGTAGT	AATAACACCACACATCAACCT
Constitutive_LMR-REST19	TGAGATAAGGTTAGTATTATGGATAGT	ACTAATTTCTTAACTACATCACCAACT
Constitutive_LMR-REST20	TGTATTTTGGGGATTTTAGGTAGG	TCCTCATAACAACCCAAAACT
Constitutive_LMR-REST21	GGGATGGTGGTTGTTTGTTA	ATCTACCCAAACCTCCTCCT
Constitutive_LMR-REST22	TGGATAGTAGGATTTGGGTTTGT	ACAACCTAACAAACATCAATTCCA

Constitutive_LMR-REST23	TTGTTTAGGGAGGGGATTGG	CCCCAACCCTATAAACAAACA
Constitutive_LMR-REST24	GATAGTGTGGGGAGTGGATT	AACCATACCTCCAAACTTTACA
Constitutive_LMR-REST25	AGGTTTATGGGTTGGAAGTTT	AACACCACAACATCTCAACC
Constitutive_LMR-REST26	TGTAGTGATTTTAGGATTTTGAGTGT	TCAATAAACTCTCCTACAAAATAAACT
Constitutive_LMR-REST27	TGTGTGAGGTGTAATGTGTG	ACTACACAAACAAAACCCAACA
Constitutive_LMR-REST28	AATGGGAAAGTAAGGTGAAGG	TCACTTCAACAAACTTTCCCC
Constitutive_LMR-REST29	AGGGTAGAATGATTGTTTTAGTGT	TCTAAACTCTTAATACCTACCCAAAC
Constitutive_LMR-REST30	AGGAGTATTTGGTTTGGAGTGA	ACTTTCATACACTTCCCACATTT
Constitutive_LMR-REST31	TTTTAAGGTTGGAAGAGTGAAAGT	ACTTCAATCCAACCATCCTCC
Constitutive_LMR-REST32	GTTGATAGTGGATGTAGTTAAAGGT	ACCCAAACAAATAAATCAAACCT
Constitutive_LMR-REST33	AGATAAATTAGGGAGTGAAGGGA	AAAACTCTAACCACCACACCT
Constitutive_LMR-REST34	TGTTTTATGGAATATTTGGGTTATGT	CAACCAACCATCCAACTAACA
Constitutive_LMR-REST35	GTTTGGTTGGGGTAAAGTTAGT	AACCAAAATCATATCACAAATCCA
Lamda Control1	TGTGTTGGTTGGAAGAGGTT	ACTATCACTCTTCTCCTCCTCT
Lamda Control2	TGTTGTTGGTTGATTTTGATGAG	TCCTCTTTCAACTCTACCACA
Lamda Control3	TTGGATGTATTGGAGAAGTATGAT	CCACCATACTAATAATCAAATCTAACA
T7 Control1	AGTGAGGGTATTGATTTTGAGT	ACCTTAAATCTATCACTCAACAAATTC
T7 Control2	GGGATGGTGAGTTTGTTGAA	CCTAATACATCTACAACTACCTCAT
T7 Control3	TGATTAGTTGAAGGATTGGAAGT	TCCCCATCAAACATAAAACCA

Table 3-6 Bisulfite primer sequences

PCR	Fwd	Rev	probe	5' dye	3' quencher
Dnmt3a	GCAGAAGGTACCAGTTTAGAAAGCA	TGCCCGCAAGGGACTTTAT	AGGAGGGCACCTTAC	6FAM	MGBNFQ
Dnmt3b	GCTGTGCAGGCAACATATGG	CCTTACGTGACCGAGCTGTCT	CAACTAACCGGAGGTTC	NED	MGBNFQ
Gapdh	GAGCCCCAGGCTATCTCATG	GTTCTCCACACCTATGGTGCAA	TCTTCAGAGTGGAATACT	VIC	MGBNFQ

Table 3-7 Probes and primers used for quantitative genotyping (TaqMan).

Fwd: forward primer, Rev: reverse primer, MGBNFQ=minor-groove binder non-fluorescent quencher. All reagents were ordered from Life Technologies.

3.2.6 Discussion

Active distal regulatory regions have reduced levels of DNA methylation which critically depend on the binding of transcription factors. Recent evidence suggests that maintenance and reprogramming of hypomethylated states is achieved through a turnover of DNA methylation (see 3.1). However, mechanisms underlying this turnover are not well understood. By comparing DNA methylation in embryonic stem cells with and without *de novo* methyltransferases activity we show that the maintenance of methylation states at regulatory regions requires *de novo* methylation. This suggests that the observed DNA methylation turnover depends on DNMT3A/B for remethylation and thus deletion of *de novo* methyltransferases can be used for turnover studies. Surprisingly, maintenance of methylation states requires *de novo* methylation also at regions which are fully methylated. By profiling DNA methylation at various time points following conditional deletion of DNMT3A/B we demonstrate that increased turnover is characteristic of low methylated active regulatory regions. Moreover, this accelerated turnover at regulatory regions and single cytosines with reduced methylation is conserved between biological replicates. Our results provide evidence that DNMT3A/B dependent DNA methylation turnover is involved in maintaining a low methylation level at active distal regulatory elements.

Active or passive demethylation?

The accelerated turnover observed at regulatory regions does not seem to be a random event, as it was reproduced across biological replicates at the level of single CpGs (Figure 3-5 and 3-6). This observation argues for an active demethylation-remethylation process. However, using a replicating cellular system we cannot exclude the involvement of a replication-dependent passive demethylation mechanism. The observed demethylation is likely to be active if measured turnover rates are higher than those inferred from passive demethylation. The following simple calculations are used to estimate the mode of demethylation during the observed turnover.

In case of an exclusively passive demethylation DNA methylation would be diluted at each cell cycle by half in the absence of *de novo* methyltransferases. Assuming one cell division in 18 hours - which is rather slow for ES cells (Welham *et al,* 2011) - a fully methylated cytosine would appear almost completely unmethylated after four days (Figure 3-8). The inferred turnover coefficient in this case (ß = -0.92) exceeds the highest turnover coefficient determined in our time courses (ß = -0.31) by three fold. Thus, the

observed turnover is slower than in case of a solely passive demethylation and cannot be unambiguously called active. While it is feasible that the turnover does not affect all cytosines in a given region simultaneously, it is likely that it involves both active and passive demethylation (Hsieh 1999). Turnover studies in postmitotic cells will be required to clarify the contribution of active demethylation.

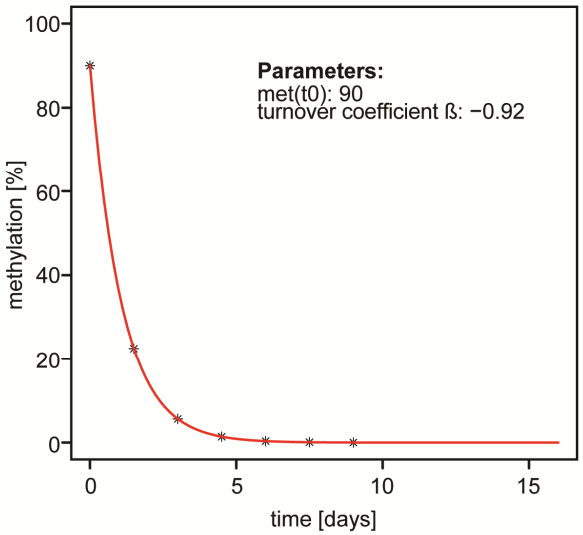

Figure 3-8 Estimation of the turnover coefficient for passive demethylation.

It was assumed that the ES cells divide once in 18 hours and the starting methylation is 90%. Note that the parameter estimated here is the turnover coefficient (not the normalized turnover rate).

DNMT3A/B targeting to sites of increased turnover

Our data indicate that that the remethylation events contributing to the observed turnover depend on DNMT3A/B (see Figure 3-1 and 3-2). Such DNMT3 dependent turnover could be mediated by different non-mutually exclusive mechanisms.

It has been suggested that DNMT3A/B could fill in the gaps left by DNMT1 during replication (Chen *et al,* 2003). One could hypothesize that if the activity of DNMT1 is reduced at some regions, this could result in increased dependency of these regions on DNMT3A/B for maintaining their methylation. DNMT1 could be prevented from maintaining DNA methylation at newly replicated DNA by hydroxymethylation (Inoue and Zhang 2011) or TF occupancy (Matsuo *et al,* 1998). It is thus conceivable that active regulatory regions that display both high 5hmC and TF enrichments have lower DNMT1 activity and are therefore predominantly targeted by DNMT3-dependent turnover.

Targeting of DNMT3A/B could furthermore depend on the sequence (Luu *et al,* 2013) or simply on the amount of demethylation within a given region. We analyzed two types of

inactive fully methylated regions (FMRs) in ES cells - neuronal progenitor specific low methylated regions (LMRs) and constitutive FMRs (Figure 3-7). Assuming that only cell type specific active regulatory regions are subject to increased turnover, neuronal progenitor specific LMRs must have accelerated kinetics in NPs, but not in ES cells. On the other hand, if the sequence composition is critical for turnover determination, these LMRs must have similarly high turnover rates in any cell type. We could not detect an accelerated turnover in progenitor-specific LMRs, when analyzed in ES cells, suggesting that DNMT3A/B targeting is not sequence specific. A kinetic analysis in NPs will be required for a definite proof of this hypothesis.

Together, our data clearly demonstrate the dependency of DNA methylation turnover on *de novo* DNA methyltransferases but do not allow drawing conclusions about the actual mode of DNMT3A/B targeting.

Turnover kinetics

Using an unbiased approach we calculated the turnover rates for cytosines residing within different regions. Upon deletion of DNMT3A/B complete demethylation is achieved on average after eight days at low methylated regulatory regions and after varying time spans at fully methylated non-regulatory sites (see Figure 3-2). While our analysis establishes hypomethylated active regulatory regions as sites of enhanced turnover, further experiments will be required in order to fully quantify its kinetics.

It is important to note that the determined turnover rates likely underestimate the speed of turnover in ES cells for two reasons. First, bisulfite sequencing does not distinguish between methylated and hydroxymethylated cytosines (Huang *et al,* 2010), so that a fraction of “methylated” cytosines could have already undergone a turnover through oxidation. In this case we expect the turnover to be even faster at low methylated regions as sites of increased hydroxymethylation (see 3.1.2). Further basepair resolution analysis of hydroxymethylation will be required to quantify this phenomenon (Booth *et al,* 2012; Yu *et al,* 2012). Second, our conditional deletion approach does not allow for analysis of a homogenous population of cells. Throughout the time course, 5-15% of the analyzed cells still display activity of DNMT3A/B and could account for the plateau observed in Figure 3-2. An experimental system which allows for a complete, homogenous and rapid disruption of DNMT3A/B function will be required for a better estimation of turnover. Due to rapidly occurring demethylation, this system must enable a tight control of DNMT activity. Such experimental approach could utilize a tightly regulated Cre-inducible or a protein degradation based system (Feil *et al,* 1997; Chu *et*

al, 2008). It is furthermore likely that the turnover is slower in postmitotic cells in the absence of passive demethylation.

We note that the turnover kinetics shows some variation among highly methylated cytosines with a linear dependency on the methylation level (see Figure 3-6). It is possible that cytosines with slightly lower methylation levels (below 80%) are generally less accurately maintained by DNMT1 and thus always rely on DNMT3A/B for maintenance. This could be affected by the surrounding sequence (Luu *et al,* 2013) or histone modifications. For example, UHRF1 binding to H3K9me3 could reinforce DNA methylation maintenance by DNMT1 (Rothbart *et al,* 2012; Liu *et al,* 2013). DNMT3A/B could furthermore be directed to hypermethylated regions, as both enzymes have been shown to tether to methylated nucleosomes in a cancer cell line (Jeong *et al,* 2009; Sharma *et al,* 2011). It is conceivable that a combination of both mechanisms results in increased methylation maintenance of certain regions.

While the turnover is generally faster at cytosines with low methylation levels, the considerable spread in turnover rates present among these CpGs precludes predicting the actual turnover kinetics solely from the starting methylation level (Figure 3-5). This suggests that additional parameters account for the accelerated turnover. Such parameters could include transcription factor binding or nucleosome positioning and are discussed in more details below (see paragraph 4).

4 General Discussion and Conclusions

Active CpG poor regulatory regions display reduced DNA methylation levels which are largely dependent on the occupancy by DNA binding factors (Stadler *et al,* 2011). However, the mechanisms underlying the reduction of DNA methylation at these regions are poorly understood. During my PhD studies I investigated maintenance and reprogramming of DNA methylation at sites of transcription factor binding.

Genome-wide profiling of the intermediate of active demethylation 5-hydroxymethylcytosine in stem and neuronal progenitor cells revealed its enrichment at active regulatory regions. Together with the observation that CTCF can bind both methylated and unmethylated cytosines, our data suggest that reduced methylation can be achieved through a transcription factor mediated turnover of DNA methylation. By deleting *de novo* DNA methyltransferases we show that this turnover likely depends on DNMT3A/B and predominantly affects active regulatory regions and cytosines with low methylation levels.

Below I will discuss the main findings and elaborate on the implications of these results for our understanding of the role of DNA methylation in gene regulation.

4.1 Transcription factor mediated demethylation

We used CTCF as example to demonstrate that binding of factors to DNA can mediate active turnover of DNA methylation during maintenance and reprogramming of correct methylation at distal regulatory regions. Genome-wide mapping of hydroxymethylation revealed that its upregulation during cellular differentiation coincides with loss of methylation and vice versa. Up to 20% of differentiation-associated changes in hydroxymethylation occur at low methylated regions (LMRs), suggesting that this effect is characteristic of active regulatory regions. These results are in line with previously reported hydroxymethylation dynamics (Serandour *et al,* 2012; Tan *et al,* 2013). Furthermore, we provide evidence for a direct dependency of 5hmC at a given site on the presence of DNA binding factors, as deletion of the TF REST leads to upregulation of methylation and decrease of hydroxymethylation at analyzed sites in ES cells. Our data mechanistically link transcription factor occupancy to active demethylation. However,

since only two factors out of the estimated 1400 in the mammalian genomes were studied, it is premature to generalize these findings.

CTCF can occupy any methylation state in CpG poor sites which are furthermore enriched for TET1 and 5hmC. Together with the observed TF binding dependent hypomethylated state of these regions, our results are compatible with a scenario where TF binding mediates TET-dependent active demethylation. We furthermore show for CTCF that its enrichments as determined by ChIP-Seq directly relate to the level of demethylation at bound CpGs. The likelihood of being demethylated correlates with the frequency of binding for a cytosine within the CTCF binding motif. This relationship could be a consequence of accelerated turnover and will be discussed in detail below (4.3). At this point it is important to note that the factors CTCF and REST analyzed in this study are particular in terms of the length and strength of their binding motifs. In order to extend our findings to other factors, a more comprehensive approach will be required.

Currently, any proposed demethylation mechanism would be speculative and based on correlative findings. It seems realistic that such demethylation occurs via TF mediated recruitment of TET proteins and subsequent hydroxymethylation (Ding *et al,* 2012). However, since hypomethylation and hydroxymethylation are general characteristics of active regulatory regions, this would require physical interactions of many different TFs with TET proteins. Alternative scenarios involve TET recruitment by factors frequently present at regulatory elements, such as p300 or pioneer transcription factors (Heintzman *et al,* 2009; Serandour *et al,* 2011). Furthermore, TF binding at regulatory regions could cause a change in chromatin conformation which attracts TET proteins (Fu *et al,* 2008).

Two explanations for the increased hydroxymethylation at regulatory regions are possible. First, hydroxymethylation could solely be an intermediate of active turnover (Valinluck and Sowers 2007; Frauer *et al,* 2011; Kubosaki *et al,* 2012; Yu *et al,* 2012; Shen *et al,* 2013; Song *et al,* 2013). However, it remains to be determined how much active turnover contributes to the hypomethylated states at these sites. Second, hydroxymethylation might be involved in enhancer function. In this case specific readers of 5hmC would be required, as has been suggested for MBD3 and MeCP2 (Yildirim *et al,* 2011; Mellen *et al,* 2012). However, other studies could not detect interaction between 5hmC and these proteins (Valinluck *et al,* 2004; Baubec *et al,* 2013). Furthermore counteracting the hypothesis of an autonomous function as a signaling module, 5hmC is accumulated in neurons suggesting that it is diluted during replication in dividing cell types (Lister *et al,* 2013). In comparison to 5hmC a larger amount of specific readers has been determined for further oxidation products 5fC and 5caC (Spruijt *et al,* 2013).

Indeed, active regulatory regions are enriched for these modifications in the absence of TDG (Shen *et al,* 2013; Song *et al,* 2013). Again, this argues for a complete processing of 5hmC at regulatory sites and thus for active turnover. Based on observations from our group and others we favor the hypothesis that 5hmC represents an intermediate in DNA methylation turnover at active regulatory regions.

Together, our findings argue that reduced methylation at regulatory regions is not solely a product of passive demethylation. Whether reduced methylation is relevant for the activity of regulatory elements remains to be determined. For CTCF it is evident that it can bind to methylated CpGs in CpG poor regions (Stadler *et al,* 2011). It is possible, but requires experimental validation, that hypomethylation established by pioneering factors is necessary to facilitate access for methylation-sensitive factors (Schubeler 2012).

4.2 DNMT3-dependent remethylation

Using embryonic stem cells with deletion of both *de novo* DNA methyltransferases we provide evidence for the presence of a turnover of DNA methylation preferentially at active regulatory sites. We show that DNMT3A/B are required for methylation maintenance at both hypomethylated regulatory regions and fully methylated regions. This finding allowed us to use conditional inactivation of DNMT3A/B to quantify turnover kinetics in ES cells. Profiling of DNA methylation throughout a time course upon deletion of DNMT3A/B revealed increased demethylation at active regulatory regions. This is reproducible between biological replicates at the level of single cytosines, arguing for an active targeting. Our results confirm the presence of a DNA methylation turnover predominantly at active regulatory regions and furthermore establish its dependency on DNMT3A/B. These findings are compatible with the observations made in ES cells cultured in 2i medium, in which DNMT3A and DNMT3B are downregulated (Habibi *et al,* 2013).

We show that in the absence of DNMT3A/B methylation is preferentially lost from active regulatory regions. In contrast to DNMT1, DNMT3A/B mainly function outside of replication (Chen *et al,* 2003), opening up the possibility that the observed turnover is active. However, the kinetics of demethylation is not fast enough to unambiguously argue for active turnover (for details see 3.2.6). More conclusive analysis in postmitotic cells will be required in order to clarify the contribution of active and passive demethylation events in the turnover process. Importantly, it is possible that both modes

coexist during demethylation of a given region as has been suggested for EBNA1 mediated demethylation (Hsieh 1999).

Increased turnover at regulatory regions is reproducible between biological replicates arguing for a regulated targeting of *de novo* DNA methyltransferases to specific sites. How exactly DNMTs are targeted to the sites of turnover and whether this is an active or a passive process remains to be determined. A sequence-specific targeting has been proposed by several studies (Lin *et al,* 2002; Handa and Jeltsch 2005). We observed that neuronal progenitor specific low methylated regions showing turnover in NPs (see 3.1) do not represent preferential turnover targets in embryonic stem cells (Figure 3-7). Together with the evidence that these sites are enriched for hydroxymethylation in NPs (see 3.1), these suggests that the targeting mechanism for DNMT3A/B is not sequence dependent. Alternatively, DNMT3A/B might be required to fill in the gaps left out by incomplete DNMT1 activity during replication (Chen *et al,* 2003). In this scenario, maintenance methylation would be compromised at sites of increased turnover in a process involving for instance TF binding (Matsuo *et al,* 1998), hydroxymethylation (Inoue and Zhang 2011) or both. A third possibility implies replication-independent demethylation followed by DNMT3 dependent remethylation. At the moment we cannot exclude a function of DNMT1 in turnover. However, this would require preceding strand-specific demethylation (Song J. *et al,* 2011; Song *et al,* 2012). Further analysis in a postmitotic system will shed light on the role of DNMT1 in the observed turnover events.

In conclusion, we confirmed the presence of a DNA methylation turnover in embryonic stem cells with a preference for active regulatory regions. This turnover depends at least in part on *de novo* DNA methyltransferases.

4.3 Transcription factor binding and turnover kinetics

We established a relationship between transcription factor binding and induction of a methylation turnover at regulatory sites. Moreover, by quantifying the kinetics of demethylation upon conditional deletion of DNMT3A/B we demonstrate that active hypomethylated regions are indeed the preferred turnover targets. While exact numbers will have to be corrected upon basepair resolution 5hmC profiling (Yu *et al,* 2012) and turnover analysis in postmitotic cells (see 3.2.6 for details), our quantification allows for a comparison between regions.

We observed differences in turnover kinetics which, only partly, could be attributed to differences in methylation levels of individual cytosines (Figures 3-6 and 3-7). These differences could result from distinct histone modifications (Rothbart *et al,* 2012) or nucleosome organization (Fu *et al,* 2008) at turnover sites. Alternatively, but not mutually exclusive, the turnover kinetics at a given site changes with differences in the type and strength of TF binding at this site. Several scenarios are possible. For instance, methylation turnover could correlate with the frequency of transcription factor binding (Lickwar *et al,* 2012). Furthermore, TFs which remain bound to DNA during replication (Zaidi *et al,* 2003; Kadauke *et al,* 2012; Caravaca *et al,* 2013) would add a passive demethylation component and thus potentially increase the turnover rates at their sites. Turnover kinetics of an individual cytosine could also be influenced by its relative position to transcription factor sequence motifs. Indeed, in line with a recent study (Jeong *et al,* 2014) we observe an accelerated turnover at cytosines located at the borders of regulatory regions.

In summary, turnover kinetics differs between distinct cytosines and regions. Systematic analyses of a collection of different transcription factors and / or sequence motifs will be required in order to gain conclusive insight into the relationship between occupancy and turnover kinetics.

4.4 Regulatory role of DNA methylation turnover

Our data argue for a regulated turnover of DNA methylation at transcription factor occupied sites. However, whether active turnover or even demethylation has a role in enhancer function is not well understood.

Instructive function of reduced methylation requires methylation-sensitive transcription factors. Attempts to categorize transcription factors by their sensitivity to DNA methylation have been mostly performed *in vitro* (Bartke *et al,* 2010; Hu *et al,* 2013; Iurlaro *et al,* 2013; Spruijt *et al,* 2013). Both methylation sensitive and insensitive factors have been reported in these studies. Thus, while a subset of TFs can occupy their cognate sequences independently of their methylation state, demethylation might be necessary for recruitment of other factors. It is possible that a small subset of DNA binding factors has a pioneering function in induction of demethylation at a certain locus (Zaret and Carroll 2011), thus rendering it accessible for methylation-sensitive factors. Repeated remethylation would then preclude promiscuous binding of the latter TFs and

thereby reduce transcriptional noise (Chen *et al,* 2013b). However, in the absence of experimental validation this scenario remains speculative.

High enrichments of TET1 and hydroxymethylation at TF-bound sites suggest a role of TET proteins in active demethylation of these regions. We furthermore demonstrate that DNMT3A/B are required for maintenance of their methylation states. However, ES cells lacking DNMTs (Sakaue *et al,* 2010) or TET proteins (Dawlaty *et al,* 2013) largely retain their transcriptional signatures, suggesting that turnover is dispensable for the maintenance of pluripotent states. Moreover, TET1 and TET2 double knockout mice are both viable and fertile (Dawlaty *et al,* 2013), while gene expression remains stable throughout neuronal differentiation of ES cells deficient for DNMT3A/B (data not shown). Taken together, data obtained from DNMT and TET knockout animals and cells argue against an essential role of DNA methylation turnover in gene regulation.

4.5 Implications of turnover

We show that DNA methylation is continuously removed and reestablished at sites of transcriptional regulation. This observation contrasts the widely accepted view of high heritability of DNA methylation states. At least for CpG poor sites, methylation appears largely instable and thus is unlikely to be involved in the propagation of repressive states. Furthermore, increased instability of this mark might limit a direct involvement in gene regulation at these sites. Importantly, we cannot generalize these findings to other genomic regions. Stably silenced genomic regions might utilize additional mechanisms to propagate DNA methylation, as has been reported for H3K9me3 (Rothbart *et al,* 2012; Liu *et al,* 2013).

In conclusion, we demonstrate the presence of enhanced turnover at transcription factor bound distal regulatory regions with reduced methylation. Our studies suggest that transcription factor occupancy can trigger demethylation and thus mediate turnover of DNA methylation. Mechanistically this turnover appears to be executed by opposed activities of TET proteins and *de novo* DNA methyltransferases in response to transcription factor binding. Involvement of the active mode of demethylation during this turnover is supported by several lines of evidence. However, whether this is a necessary and sufficient mode to create a demethylated region at TF-bound sites requires further proof in a postmitotic system.

5 References

Adams C.C. and Workman J.L. (1995) Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol* **15**: 1405-1421.

Ahmad K. and Henikoff S. (2002) The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell* **9**: 1191-1200.

Allfrey V.G., Faulkner R. and Mirsky A.E. (1964) Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* **51**: 786-794.

Allis C.D.J., T.; Reinberg, D.; Caparros, M.-L. (2007) Epigenetics. *Cold Spring Harbor Laboratory Press*.

Arand J., Spieler D., Karius T., Branco M.R., Meilinger D., Meissner A., Jenuwein T., Xu G., Leonhardt H., Wolf V. and Walter J. (2012) In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet* **8**: e1002750.

Arnold P., Scholer A., Pachkov M., Balwierz P.J., Jorgensen H., Stadler M.B., van Nimwegen E. and Schubeler D. (2013) Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome Res* **23**: 60-73.

Arnosti D.N. and Kulkarni M.M. (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* **94**: 890-898.

Badis G., Chan E.T., van Bakel H., Pena-Castillo L., Tillo D., Tsui K., Carlson C.D., Gossett A.J., Hasinoff M.J., Warren C.L., Gebbia M., Talukder S., Yang A., Mnaimneh S., Terterov D., Coburn D., Li Yeo A., Yeo Z.X., Clarke N.D., Lieb J.D., Ansari A.Z., Nislow C. and Hughes T.R. (2008) A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol Cell* **32**: 878-887.

Banerji J., Rusconi S. and Schaffner W. (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299-308.

Bannister A.J. and Kouzarides T. (2011) Regulation of chromatin by histone modifications. *Cell Res* **21**: 381-395.

Bardet A.F. (2012) Computational analyses of transcription factor binding across species and tissues.

Bartke T., Vermeulen M., Xhemalce B., Robson S.C., Mann M. and Kouzarides T. (2010) Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**: 470-484.

Baubec T., Ivanek R., Lienert F. and Schubeler D. (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**: 480-492.

Bell A.C., West A.G. and Felsenfeld G. (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**: 387-396.

Bell A.C. and Felsenfeld G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**: 482-485.

Bell A.C., West A.G. and Felsenfeld G. (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**: 447-450.

Bestor T., Laudano A., Mattaliano R. and Ingram V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal

domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* **203**: 971-983.

Bestor T.H. (1990) DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philos Trans R Soc Lond B Biol Sci* **326**: 179-187.

Bestor T.H. (1999) Sex brings transposons and genomes into conflict. *Genetica* **107**: 289-295.

Bhaumik S.R., Smith E. and Shilatifard A. (2007) Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol* **14**: 1008-1016.

Bhutani N., Brady J.J., Damian M., Sacco A., Corbel S.Y. and Blau H.M. (2010) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**: 1042-1047.

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

Bird A., Taggart M., Frommer M., Miller O.J. and Macleod D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**: 91-99.

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

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

Bird A.P. and Southern E.M. (1978) Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. *J Mol Biol* **118**: 27-47.

Bird A.P. (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* **8**: 1499-1504.

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

Bird A.P. and Wolffe A.P. (1999) Methylation-induced repression--belts, braces, and chromatin. *Cell* **99**: 451-454.

Blackledge N.P., Long H.K., Zhou J.C., Kriaucionis S., Patient R. and Klose R.J. (2012) Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA. *Nucleic Acids Res* **40**: e32.

Booth M.J., Branco M.R., Ficz G., Oxley D., Krueger F., Reik W. and Balasubramanian S. (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**: 934-937.

Borgel J., Guibert S., Li Y., Chiba H., Schubeler D., Sasaki H., Forne T. and Weber M. (2010) Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* **42**: 1093-1100.

Boyer L.A., Lee T.I., Cole M.F., Johnstone S.E., Levine S.S., Zucker J.P., Guenther M.G., Kumar R.M., Murray H.L., Jenner R.G., Gifford D.K., Melton D.A., Jaenisch R. and Young R.A. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**: 947-956.

Boyes J. and Bird A. (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**: 327-333.

Brandeis M., Frank D., Keshet I., Siegfried Z., Mendelsohn M., Nemes A., Temper V., Razin A. and Cedar H. (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**: 435-438.

Brennecke J., Aravin A.A., Stark A., Dus M., Kellis M., Sachidanandam R. and Hannon G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**: 1089-1103.

Brinkman A.B., Gu H., Bartels S.J., Zhang Y., Matarese F., Simmer F., Marks H., Bock C., Gnirke A., Meissner A. and Stunnenberg H.G. (2012) Sequential ChIP-bisulfite

sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res* **22**: 1128-1138.

Brownell J.E. and Allis C.D. (1996) Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr Opin Genet Dev* **6**: 176-184.

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

Burger L., Gaidatzis D., Schubeler D. and Stadler M.B. (2013) Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res* **41**: e155.

Campanero M.R., Armstrong M.I. and Flemington E.K. (2000) CpG methylation as a mechanism for the regulation of E2F activity. *Proc Natl Acad Sci U S A* **97**: 6481-6486.

Caravaca J.M., Donahue G., Becker J.S., He X., Vinson C. and Zaret K.S. (2013) Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev* **27**: 251-260.

Carlson L.L., Page A.W. and Bestor T.H. (1992) Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes Dev* **6**: 2536-2541.

Carr A. and Biggin M.D. (1999) A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos. *EMBO J* **18**: 1598-1608.

Carroll J.S., Meyer C.A., Song J., Li W., Geistlinger T.R., Eeckhoute J., Brodsky A.S., Keeton E.K., Fertuck K.C., Hall G.F., Wang Q., Bekiranov S., Sementchenko V., Fox E.A., Silver P.A., Gingeras T.R., Liu X.S. and Brown M. (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38**: 1289-1297.

Challen G.A., Sun D., Jeong M., Luo M., Jelinek J., Berg J.S., Bock C., Vasanthakumar A., Gu H., Xi Y., Liang S., Lu Y., Darlington G.J., Meissner A., Issa J.P., Godley L.A., Li W. and Goodell M.A. (2012) Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet* **44**: 23-31.

Chambeyron S. and Bickmore W.A. (2004) Does looping and clustering in the nucleus regulate gene expression? *Curr Opin Cell Biol* **16**: 256-262.

Chen C.C., Wang K.Y. and Shen C.K. (2012) The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J Biol Chem* **287**: 33116-33121.

Chen C.C., Wang K.Y. and Shen C.K. (2013a) DNA 5-methylcytosine demethylation activities of the mammalian DNA methyltransferases. *J Biol Chem* **288**: 9084-9091.

Chen C.C., Xiao S., Xie D., Cao X., Song C.X., Wang T., He C. and Zhong S. (2013b) Understanding variation in transcription factor binding by modeling transcription factor genome-epigenome interactions. *PLoS Comput Biol* **9**: e1003367.

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

Chu B.W., Banaszynski L.A., Chen L.C. and Wandless T.J. (2008) Recent progress with FKBP-derived destabilizing domains. *Bioorg Med Chem Lett* **18**: 5941-5944.

Cirillo L.A., Lin F.R., Cuesta I., Friedman D., Jarnik M. and Zaret K.S. (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**: 279-289.

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

Collings C.K., Waddell P.J. and Anderson J.N. (2013) Effects of DNA methylation on nucleosome stability. *Nucleic Acids Res* **41**: 2918-2931.

Colquitt B.M., Allen W.E., Barnea G. and Lomvardas S. (2013) Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc Natl Acad Sci U S A* **110**: 14682-14687.

Corden J., Wasylyk B., Buchwalder A., Sassone-Corsi P., Kedinger C. and Chambon P. (1980) Promoter sequences of eukaryotic protein-coding genes. *Science* **209**: 1406-1414.

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

Coulondre C., Miller J.H., Farabaugh P.J. and Gilbert W. (1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature* **274**: 775-780.

Cross S.H., Charlton J.A., Nan X. and Bird A.P. (1994) Purification of CpG islands using a methylated DNA binding column. *Nat Genet* **6**: 236-244.

Davidson E.H. (2010) Emerging properties of animal gene regulatory networks. *Nature* **468**: 911-920.

Dawlaty M.M., Breiling A., Le T., Raddatz G., Barrasa M.I., Cheng A.W., Gao Q., Powell B.E., Li Z., Xu M., Faull K.F., Lyko F. and Jaenisch R. (2013) Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev Cell* **24**: 310-323.

De Koning L., Corpet A., Haber J.E. and Almouzni G. (2007) Histone chaperones: an escort network regulating histone traffic. *Nat Struct Mol Biol* **14**: 997-1007.

Deal R.B., Henikoff J.G. and Henikoff S. (2010) Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**: 1161-1164.

Dean W., Santos F., Stojkovic M., Zakhartchenko V., Walter J., Wolf E. and Reik W. (2001) Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc Natl Acad Sci U S A* **98**: 13734-13738.

Deaton A.M. and Bird A. (2011) CpG islands and the regulation of transcription. *Genes Dev* **25**: 1010-1022.

Ding J., Xu H., Faiola F., Ma'ayan A. and Wang J. (2012) Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Res* **22**: 155-167.

Dion M.F., Kaplan T., Kim M., Buratowski S., Friedman N. and Rando O.J. (2007) Dynamics of replication-independent histone turnover in budding yeast. *Science* **315**: 1405-1408.

Dodge J.E., Okano M., Dick F., Tsujimoto N., Chen T., Wang S., Ueda Y., Dyson N. and Li E. (2005) Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *J Biol Chem* **280**: 17986-17991.

Elgin S.C. (1996) Heterochromatin and gene regulation in Drosophila. *Curr Opin Genet Dev* **6**: 193-202.

Elliott E., Ezra-Nevo G., Regev L., Neufeld-Cohen A. and Chen A. (2010) Resilience to social stress coincides with functional DNA methylation of the Crf gene in adult mice. *Nat Neurosci* **13**: 1351-1353.

Falvo J.V., Thanos D. and Maniatis T. (1995) Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* **83**: 1101-1111.

Feil R., Wagner J., Metzger D. and Chambon P. (1997) Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* **237**: 752-757.

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

Feldmann A., Ivanek R., Murr R., Gaidatzis D., Burger L. and Schubeler D. (2013) Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet* **9**: e1003994.

Feng J., Zhou Y., Campbell S.L., Le T., Li E., Sweatt J.D., Silva A.J. and Fan G. (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat Neurosci* **13**: 423-430.

Filion G.J., Zhenilo S., Salozhin S., Yamada D., Prokhortchouk E. and Defossez P.A. (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol* **26**: 169-181.

Filion G.J., van Bemmel J.G., Braunschweig U., Talhout W., Kind J., Ward L.D., Brugman W., de Castro I.J., Kerkhoven R.M., Bussemaker H.J. and van Steensel B. (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**: 212-224.

Franchini D.M., Schmitz K.M. and Petersen-Mahrt S.K. (2012) 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu Rev Genet* **46**: 419-441.

Fraser P. (2006) Transcriptional control thrown for a loop. *Curr Opin Genet Dev* **16**: 490-495.

Frauer C., Hoffmann T., Bultmann S., Casa V., Cardoso M.C., Antes I. and Leonhardt H. (2011) Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One* **6**: e21306.

Freitag M., Williams R.L., Kothe G.O. and Selker E.U. (2002) A cytosine methyltransferase homologue is essential for repeat-induced point mutation in Neurospora crassa. *Proc Natl Acad Sci U S A* **99**: 8802-8807.

Fromm M. and Berg P. (1983) Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. *Mol Cell Biol* **3**: 991-999.

Frommer M., McDonald L.E., Millar D.S., Collis C.M., Watt F., Grigg G.W., Molloy P.L. and Paul C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* **89**: 1827-1831.

Fu Y., Sinha M., Peterson C.L. and Weng Z. (2008) The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet* **4**: e1000138.

Furey T.S. (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* **13**: 840-852.

Furner I.J. and Matzke M. (2011) Methylation and demethylation of the Arabidopsis genome. *Curr Opin Plant Biol* **14**: 137-141.

Gaidatzis D., Burger L., Murr R., Lerch A., Dessus-Babus S., Schubeler D. and Stadler M.B. (2014) DNA sequence explains seemingly disordered methylation levels in partially methylated domains of Mammalian genomes. *PLoS Genet* **10**: e1004143.

Gannon F., O'Hare K., Perrin F., LePennec J.P., Benoist C., Cochet M., Breathnach R., Royal A., Garapin A., Cami B. and Chambon P. (1979) Organisation and sequences at the 5' end of a cloned complete ovalbumin gene. *Nature* **278**: 428-434.

Gaudet F., Rideout W.M., 3rd, Meissner A., Dausman J., Leonhardt H. and Jaenisch R. (2004) Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. *Mol Cell Biol* **24**: 1640-1648.

Gehring M., Reik W. and Henikoff S. (2009) DNA demethylation by DNA repair. *Trends Genet* **25**: 82-90.

Gillies S.D., Morrison S.L., Oi V.T. and Tonegawa S. (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**: 717-728.

Globisch D., Munzel M., Muller M., Michalakis S., Wagner M., Koch S., Bruckl T., Biel M. and Carell T. (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**: e15367.

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

Gorovsky M.A., Hattman S. and Pleger G.L. (1973) (6 N)methyl adenine in the nuclear DNA of a eucaryote, Tetrahymena pyriformis. *J Cell Biol* **56**: 697-701.

Goto K., Numata M., Komura J.I., Ono T., Bestor T.H. and Kondo H. (1994) Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. *Differentiation* **56**: 39-44.

Gowher H., Loutchanwoot P., Vorobjeva O., Handa V., Jurkowska R.Z., Jurkowski T.P. and Jeltsch A. (2006) Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. *J Mol Biol* **357**: 928-941.

Gregory T.R. (2014) Animal Genome Size Database. http://www.genomesize.com.

Grewal S.I. and Rice J.C. (2004) Regulation of heterochromatin by histone methylation and small RNAs. *Curr Opin Cell Biol* **16**: 230-238.

Grosschedl R., Wasylyk B., Chambon P. and Birnstiel M.L. (1981) Point mutation in the TATA box curtails expression of sea urchin H2A histone gene in vivo. *Nature* **294**: 178-180.

Gu T.P., Guo F., Yang H., Wu H.P., Xu G.F., Liu W., Xie Z.G., Shi L., He X., Jin S.G., Iqbal K., Shi Y.G., Deng Z., Szabo P.E., Pfeifer G.P., Li J. and Xu G.L. (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**: 606-610.

Guo J.U., Su Y., Zhong C., Ming G.L. and Song H. (2011a) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**: 423-434.

Guo J.U., Ma D.K., Mo H., Ball M.P., Jang M.H., Bonaguidi M.A., Balazer J.A., Eaves H.L., Xie B., Ford E., Zhang K., Ming G.L., Gao Y. and Song H. (2011b) Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat Neurosci* **14**: 1345-1351.

Habibi E., Brinkman A.B., Arand J., Kroeze L.I., Kerstens H.H., Matarese F., Lepikhov K., Gut M., Brun-Heath I., Hubner N.C., Benedetti R., Altucci L., Jansen J.H., Walter J., Gut I.G., Marks H. and Stunnenberg H.G. (2013) Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**: 360-369.

Hackett J.A., Reddington J.P., Nestor C.E., Dunican D.S., Branco M.R., Reichmann J., Reik W., Surani M.A., Adams I.R. and Meehan R.R. (2012) Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* **139**: 3623-3632.

Hajkova P., Erhardt S., Lane N., Haaf T., El-Maarri O., Reik W., Walter J. and Surani M.A. (2002) Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**: 15-23.

Handa V. and Jeltsch A. (2005) Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *J Mol Biol* **348**: 1103-1112.

Hark A.T., Schoenherr C.J., Katz D.J., Ingram R.S., Levorse J.M. and Tilghman S.M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**: 486-489.

Harrington M.A., Jones P.A., Imagawa M. and Karin M. (1988) Cytosine methylation does not affect binding of transcription factor Sp1. *Proc Natl Acad Sci U S A* **85**: 2066-2070.

Hashimoto H., Liu Y., Upadhyay A.K., Chang Y., Howerton S.B., Vertino P.M., Zhang X. and Cheng X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* **40**: 4841-4849.

Haupt S., Edenhofer F., Peitz M., Leinhaas A. and Brustle O. (2007) Stage-specific conditional mutagenesis in mouse embryonic stem cell-derived neural cells and postmitotic neurons by direct delivery of biologically active Cre recombinase. *Stem Cells* **25**: 181-188.

He Y.F., Li B.Z., Li Z., Liu P., Wang Y., Tang Q., Ding J., Jia Y., Chen Z., Li L., Sun Y., Li X., Dai Q., Song C.X., Zhang K., He C. and Xu G.L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**: 1303-1307.

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

Heintzman N.D. and Ren B. (2009) Finding distal regulatory elements in the human genome. *Curr Opin Genet Dev* **19**: 541-549.

Heintzman N.D., Hon G.C., Hawkins R.D., Kheradpour P., Stark A., Harp L.F., Ye Z., Lee L.K., Stuart R.K., Ching C.W., Ching K.A., Antosiewicz-Bourget J.E., Liu H., Zhang X., Green R.D., Lobanenkov V.V., Stewart R., Thomson J.A., Crawford G.E., Kellis M. and Ren B. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108-112.

Hodges E., Molaro A., Dos Santos C.O., Thekkat P., Song Q., Uren P.J., Park J., Butler J., Rafii S., McCombie W.R., Smith A.D. and Hannon G.J. (2011) Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell* **44**: 17-28.

Holliday R. and Pugh J.E. (1975) DNA modification mechanisms and gene activity during development. *Science* **187**: 226-232.

Hollister J.D. and Gaut B.S. (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* **19**: 1419-1428.

Hon G.C., Rajagopal N., Shen Y., McCleary D.F., Yue F., Dang M.D. and Ren B. (2013) Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* **45**: 1198-1206.

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

Hsieh C.L. (1994) Dependence of transcriptional repression on CpG methylation density. *Mol Cell Biol* **14**: 5487-5494.

Hsieh C.L. (1999) Evidence that protein binding specifies sites of DNA demethylation. *Mol Cell Biol* **19**: 46-56.

Hu S., Wan J., Su Y., Song Q., Zeng Y., Nguyen H.N., Shin J., Cox E., Rho H.S., Woodard C., Xia S., Liu S., Lyu H., Ming G.L., Wade H., Song H., Qian J. and Zhu H. (2013) DNA methylation presents distinct binding sites for human transcription factors. *Elife* **2**: e00726.

Huang Y., Pastor W.A., Shen Y., Tahiliani M., Liu D.R. and Rao A. (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* **5**: e8888.

Huang Y., Chavez L., Chang X., Wang X., Pastor W.A., Kang J., Zepeda-Martinez J.A., Pape U.J., Jacobsen S.E., Peters B. and Rao A. (2014) Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **111**: 1361-1366.

Huangfu D., Maehr R., Guo W., Eijkelenboom A., Snitow M., Chen A.E. and Melton D.A. (2008) Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol* **26**: 795-797.

Iguchi-Ariga S.M. and Schaffner W. (1989) CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev* **3**: 612-619.

Inoue A. and Zhang Y. (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**: 194.

Ioshikhes I.P., Albert I., Zanton S.J. and Pugh B.F. (2006) Nucleosome positions predicted through comparative genomics. *Nat Genet* **38**: 1210-1215.

Ito S., D'Alessio A.C., Taranova O.V., Hong K., Sowers L.C. and Zhang Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**: 1129-1133.

Ito S., Shen L., Dai Q., Wu S.C., Collins L.B., Swenberg J.A., He C. and Zhang Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**: 1300-1303.

Iurlaro M., Ficz G., Oxley D., Raiber E.A., Bachman M., Booth M.J., Andrews S., Balasubramanian S. and Reik W. (2013) A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol* **14**: R119.

Iyer V.R., Horak C.E., Scafe C.S., Botstein D., Snyder M. and Brown P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**: 533-538.

Izzo A. and Schneider R. (2010) Chatting histone modifications in mammals. *Brief Funct Genomics* **9**: 429-443.

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

Jahner D., Stuhlmann H., Stewart C.L., Harbers K., Lohler J., Simon I. and Jaenisch R. (1982) De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* **298**: 623-628.

Jahner D. and Jaenisch R. (1985) Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity. *Nature* **315**: 594-597.

Jamai A., Imoberdorf R.M. and Strubin M. (2007) Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. *Mol Cell* **25**: 345-355.

James T.C., Eissenberg J.C., Craig C., Dietrich V., Hobson A. and Elgin S.C. (1989) Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila. *Eur J Cell Biol* **50**: 170-180.

Jeong M., Sun D., Luo M., Huang Y., Challen G.A., Rodriguez B., Zhang X., Chavez L., Wang H., Hannah R., Kim S.B., Yang L., Ko M., Chen R., Gottgens B., Lee J.S., Gunaratne P., Godley L.A., Darlington G.J., Rao A., Li W. and Goodell M.A. (2014) Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet* **46**: 17-23.

Jeong S., Liang G., Sharma S., Lin J.C., Choi S.H., Han H., Yoo C.B., Egger G., Yang A.S. and Jones P.A. (2009) Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. *Mol Cell Biol* **29**: 5366-5376.

John S., Sabo P.J., Thurman R.E., Sung M.H., Biddie S.C., Johnson T.A., Hager G.L. and Stamatoyannopoulos J.A. (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* **43**: 264-268.

Johnson A.D., Meyer B.J. and Ptashne M. (1979) Interactions between DNA-bound repressors govern regulation by the lambda phage repressor. *Proc Natl Acad Sci U S A* **76**: 5061-5065.

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

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

Joseph R., Orlov Y.L., Huss M., Sun W., Kong S.L., Ukil L., Pan Y.F., Li G., Lim M., Thomsen J.S., Ruan Y., Clarke N.D., Prabhakar S., Cheung E. and Liu E.T. (2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. *Mol Syst Biol* **6**: 456.

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

Kadauke S., Udugama M.I., Pawlicki J.M., Achtman J.C., Jain D.P., Cheng Y., Hardison R.C. and Blobel G.A. (2012) Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell* **150**: 725-737.

Kanduri C., Pant V., Loukinov D., Pugacheva E., Qi C.F., Wolffe A., Ohlsson R. and Lobanenkov V.V. (2000) Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol* **10**: 853-856.

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

Kaplan N., Moore I.K., Fondufe-Mittendorf Y., Gossett A.J., Tillo D., Field Y., LeProust E.M., Hughes T.R., Lieb J.D., Widom J. and Segal E. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**: 362-366.

Kaplan T., Li X.Y., Sabo P.J., Thomas S., Stamatoyannopoulos J.A., Biggin M.D. and Eisen M.B. (2011) Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. *PLoS Genet* **7**: e1001290.

Kellinger M.W., Song C.X., Chong J., Lu X.Y., He C. and Wang D. (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat Struct Mol Biol* **19**: 831-833.

Kelly T.K., Liu Y., Lay F.D., Liang G., Berman B.P. and Jones P.A. (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res* **22**: 2497-2506.

Kharchenko P.V., Alekseyenko A.A., Schwartz Y.B., Minoda A., Riddle N.C., Ernst J., Sabo P.J., Larschan E., Gorchakov A.A., Gu T., Linder-Basso D., Plachetka A., Shanower G., Tolstorukov M.Y., Luquette L.J., Xi R., Jung Y.L., Park R.W., Bishop E.P., Canfield T.K., Sandstrom R., Thurman R.E., MacAlpine D.M., Stamatoyannopoulos J.A., Kellis M., Elgin S.C., Kuroda M.I., Pirrotta V., Karpen G.H. and Park P.J. (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**: 480-485.

Kidwell M.G. (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**: 49-63.

Kim J., Kollhoff A., Bergmann A. and Stubbs L. (2003) Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. *Hum Mol Genet* **12**: 233-245.

Klose R.J. and Bird A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**: 89-97.

Klug M., Heinz S., Gebhard C., Schwarzfischer L., Krause S.W., Andreesen R. and Rehli M. (2010) Active DNA demethylation in human postmitotic cells correlates with activating histone modifications, but not transcription levels. *Genome Biol* **11**: R63.

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

Kriaucionis S. and Heintz N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**: 929-930.

Krogan N.J., Dover J., Wood A., Schneider J., Heidt J., Boateng M.A., Dean K., Ryan O.W., Golshani A., Johnston M., Greenblatt J.F. and Shilatifard A. (2003) The

Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* **11**: 721-729.

Krueger F. and Andrews S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571-1572.

Kubosaki A., Tomaru Y., Furuhata E., Suzuki T., Shin J.W., Simon C., Ando Y., Hasegawa R., Hayashizaki Y. and Suzuki H. (2012) CpG site-specific alteration of hydroxymethylcytosine to methylcytosine beyond DNA replication. *Biochem Biophys Res Commun* **426**: 141-147.

La Salle S., Mertineit C., Taketo T., Moens P.B., Bestor T.H. and Trasler J.M. (2004) Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. *Dev Biol* **268**: 403-415.

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 Biol* **10**: R25.

LaPlant Q., Vialou V., Covington H.E., 3rd, Dumitriu D., Feng J., Warren B.L., Maze I., Dietz D.M., Watts E.L., Iniguez S.D., Koo J.W., Mouzon E., Renthal W., Hollis F., Wang H., Noonan M.A., Ren Y., Eisch A.J., Bolanos C.A., Kabbaj M., Xiao G., Neve R.L., Hurd Y.L., Oosting R.S., Fan G., Morrison J.H. and Nestler E.J. (2010) Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. *Nat Neurosci* **13**: 1137-1143.

Lee J., Inoue K., Ono R., Ogonuki N., Kohda T., Kaneko-Ishino T., Ogura A. and Ishino F. (2002) Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* **129**: 1807-1817.

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

Lenhard B., Sandelin A. and Carninci P. (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* **13**: 233-245.

Levine M. and Tjian R. (2003) Transcription regulation and animal diversity. *Nature* **424**: 147-151.

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

Li G., Margueron R., Hu G., Stokes D., Wang Y.H. and Reinberg D. (2010) Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo. *Mol Cell* **38**: 41-53.

Lickwar C.R., Mueller F., Hanlon S.E., McNally J.G. and Lieb J.D. (2012) Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**: 251-255.

Lienert F., Wirbelauer C., Som I., Dean A., Mohn F. and Schubeler D. (2011) Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* **43**: 1091-1097.

Lin I.G., Tomzynski T.J., Ou Q. and Hsieh C.L. (2000) Modulation of DNA binding protein affinity directly affects target site demethylation. *Mol Cell Biol* **20**: 2343-2349.

Lin I.G., Han L., Taghva A., O'Brien L.E. and Hsieh C.L. (2002) Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. *Mol Cell Biol* **22**: 704-723.

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

Lister R., Pelizzola M., Dowen R.H., Hawkins R.D., Hon G., Tonti-Filippini J., Nery J.R., Lee L., Ye Z., Ngo Q.M., Edsall L., Antosiewicz-Bourget J., Stewart R., Ruotti V., Millar A.H., Thomson J.A., Ren B. and Ecker J.R. (2009) Human DNA

methylomes at base resolution show widespread epigenomic differences. *Nature* **462**: 315-322.

Lister R., Mukamel E.A., Nery J.R., Urich M., Puddifoot C.A., Johnson N.D., Lucero J., Huang Y., Dwork A.J., Schultz M.D., Yu M., Tonti-Filippini J., Heyn H., Hu S., Wu J.C., Rao A., Esteller M., He C., Haghighi F.G., Sejnowski T.J., Behrens M.M. and Ecker J.R. (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* **341**: 1237905.

Liu X., Lee C.K., Granek J.A., Clarke N.D. and Lieb J.D. (2006) Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. *Genome Res* **16**: 1517-1528.

Liu X., Gao Q., Li P., Zhao Q., Zhang J., Li J., Koseki H. and Wong J. (2013) UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat Commun* **4**: 1563.

Long H.K., Sims D., Heger A., Blackledge N.P., Kutter C., Wright M.L., Grutzner F., Odom D.T., Patient R., Ponting C.P. and Klose R.J. (2013a) Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife* **2**: e00348.

Long H.K., Blackledge N.P. and Klose R.J. (2013b) ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem Soc Trans* **41**: 727-740.

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

Luger K. (2003) Structure and dynamic behavior of nucleosomes. *Curr Opin Genet Dev* **13**: 127-135.

Lupien M., Eeckhoute J., Meyer C.A., Wang Q., Zhang Y., Li W., Carroll J.S., Liu X.S. and Brown M. (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**: 958-970.

Luu P.L., Scholer H.R. and Arauzo-Bravo M.J. (2013) Disclosing the crosstalk among DNA methylation, transcription factors, and histone marks in human pluripotent cells through discovery of DNA methylation motifs. *Genome Res* **23**: 2013-2029.

Lynch M.D., Smith A.J., De Gobbi M., Flenley M., Hughes J.R., Vernimmen D., Ayyub H., Sharpe J.A., Sloane-Stanley J.A., Sutherland L., Meek S., Burdon T., Gibbons R.J., Garrick D. and Higgs D.R. (2012) An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J* **31**: 317-329.

Macleod D., Charlton J., Mullins J. and Bird A.P. (1994) Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev* **8**: 2282-2292.

Maeder M.L., Angstman J.F., Richardson M.E., Linder S.J., Cascio V.M., Tsai S.Q., Ho Q.H., Sander J.D., Reyon D., Bernstein B.E., Costello J.F., Wilkinson M.F. and Joung J.K. (2013) Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31**: 1137-1142.

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

Matsuo K., Silke J., Georgiev O., Marti P., Giovannini N. and Rungger D. (1998) An embryonic demethylation mechanism involving binding of transcription factors to replicating DNA. *EMBO J* **17**: 1446-1453.

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

McGhee J.D. and Ginder G.D. (1979) Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature* **280**: 419-420.

McKay D.J. and Lieb J.D. (2013) A common set of DNA regulatory elements shapes Drosophila appendages. *Dev Cell* **27**: 306-318.

Mellen M., Ayata P., Dewell S., Kriaucionis S. and Heintz N. (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**: 1417-1430.

Mendenhall E.M., Koche R.P., Truong T., Zhou V.W., Issac B., Chi A.S., Ku M. and Bernstein B.E. (2010) GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* **6**: e1001244.

Merika M., Williams A.J., Chen G., Collins T. and Thanos D. (1998) Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Mol Cell* **1**: 277-287.

Metivier R., Gallais R., Tiffoche C., Le Peron C., Jurkowska R.Z., Carmouche R.P., Ibberson D., Barath P., Demay F., Reid G., Benes V., Jeltsch A., Gannon F. and Salbert G. (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**: 45-50.

Miller J.A. and Widom J. (2003) Collaborative competition mechanism for gene activation in vivo. *Mol Cell Biol* **23**: 1623-1632.

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

Mohn F., Weber M., Schubeler D. and Roloff T.C. (2009) Methylated DNA immunoprecipitation (MeDIP). *Methods Mol Biol* **507**: 55-64.

Monk M., Boubelik M. and Lehnert S. (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**: 371-382.

Morgan H.D., Dean W., Coker H.A., Reik W. and Petersen-Mahrt S.K. (2004) Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem* **279**: 52353-52360.

Morgan H.D., Santos F., Green K., Dean W. and Reik W. (2005) Epigenetic reprogramming in mammals. *Hum Mol Genet* **14 Spec No 1**: R47-58.

Morris S.A., Baek S., Sung M.H., John S., Wiench M., Johnson T.A., Schiltz R.L. and Hager G.L. (2014) Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nat Struct Mol Biol* **21**: 73-81.

Mueller-Planitz F., Klinker H. and Becker P.B. (2013) Nucleosome sliding mechanisms: new twists in a looped history. *Nat Struct Mol Biol* **20**: 1026-1032.

Mullen A.C., Orlando D.A., Newman J.J., Loven J., Kumar R.M., Bilodeau S., Reddy J., Guenther M.G., DeKoter R.P. and Young R.A. (2011) Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**: 565-576.

Nabel C.S., Jia H., Ye Y., Shen L., Goldschmidt H.L., Stivers J.T., Zhang Y. and Kohli R.M. (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol* **8**: 751-758.

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

Ng H.H., Robert F., Young R.A. and Struhl K. (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11**: 709-719.

Ng R.K. and Gurdon J.B. (2008a) Epigenetic inheritance of cell differentiation status. *Cell Cycle* **7**: 1173-1177.

Ng R.K. and Gurdon J.B. (2008b) Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat Cell Biol* **10**: 102-109.

Nguyen S., Meletis K., Fu D., Jhaveri S. and Jaenisch R. (2007) Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Dev Dyn* **236**: 1663-1676.

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

Okamoto I. and Heard E. (2009) Lessons from comparative analysis of X-chromosome inactivation in mammals. *Chromosome Res* **17**: 659-669.

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

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

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

Oswald J., Engemann S., Lane N., Mayer W., Olek A., Fundele R., Dean W., Reik W. and Walter J. (2000) Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* **10**: 475-478.

Pastor W.A., Pape U.J., Huang Y., Henderson H.R., Lister R., Ko M., McLoughlin E.M., Brudno Y., Mahapatra S., Kapranov P., Tahiliani M., Daley G.Q., Liu X.S., Ecker J.R., Milos P.M., Agarwal S. and Rao A. (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**: 394-397.

Pastor W.A., Aravind L. and Rao A. (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* **14**: 341-356.

Payer B. and Lee J.T. (2008) X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**: 733-772.

Peitz M., Pfannkuche K., Rajewsky K. and Edenhofer F. (2002) Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. *Proc Natl Acad Sci U S A* **99**: 4489-4494.

Phillips-Cremins J.E. and Corces V.G. (2013) Chromatin insulators: linking genome organization to cellular function. *Mol Cell* **50**: 461-474.

Plass C., Pfister S.M., Lindroth A.M., Bogatyrova O., Claus R. and Lichter P. (2013) Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat Rev Genet* **14**: 765-780.

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

Popp C., Dean W., Feng S., Cokus S.J., Andrews S., Pellegrini M., Jacobsen S.E. and Reik W. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**: 1101-1105.

Pospisil V., Vargova K., Kokavec J., Rybarova J., Savvulidi F., Jonasova A., Necas E., Zavadil J., Laslo P. and Stopka T. (2011) Epigenetic silencing of the oncogenic miR-17-92 cluster during PU.1-directed macrophage differentiation. *EMBO J* **30**: 4450-4464.

Pribnow D. (1975) Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc Natl Acad Sci U S A* **72**: 784-788.

Probst A.V., Dunleavy E. and Almouzni G. (2009) Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol* **10**: 192-206.

Radman-Livaja M., Verzijlbergen K.F., Weiner A., van Welsem T., Friedman N., Rando O.J. and van Leeuwen F. (2011) Patterns and mechanisms of ancestral histone protein inheritance in budding yeast. *PLoS Biol* **9**: e1001075.

Ramsahoye B.H., Biniszkiewicz D., Lyko F., Clark V., Bird A.P. and Jaenisch R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A* **97**: 5237-5242.

Rangam G., Schmitz K.M., Cobb A.J. and Petersen-Mahrt S.K. (2012) AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS One* **7**: e43279.

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

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

Ringrose L. and Paro R. (2004) Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet* **38**: 413-443.

Rothbart S.B., Krajewski K., Nady N., Tempel W., Xue S., Badeaux A.I., Barsyte-Lovejoy D., Martinez J.Y., Bedford M.T., Fuchs S.M., Arrowsmith C.H. and Strahl B.D. (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* **19**: 1155-1160.

Rufiange A., Jacques P.E., Bhat W., Robert F. and Nourani A. (2007) Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Mol Cell* **27**: 393-405.

Sado T., Fenner M.H., Tan S.S., Tam P., Shioda T. and Li E. (2000) X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol* **225**: 294-303.

Sakaue M., Ohta H., Kumaki Y., Oda M., Sakaide Y., Matsuoka C., Yamagiwa A., Niwa H., Wakayama T. and Okano M. (2010) DNA methylation is dispensable for the growth and survival of the extraembryonic lineages. *Curr Biol* **20**: 1452-1457.

Santos-Rosa H., Schneider R., Bannister A.J., Sherriff J., Bernstein B.E., Emre N.C., Schreiber S.L., Mellor J. and Kouzarides T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* **419**: 407-411.

Santos-Rosa H., Schneider R., Bernstein B.E., Karabetsou N., Morillon A., Weise C., Schreiber S.L., Mellor J. and Kouzarides T. (2003) Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol Cell* **12**: 1325-1332.

Santos F., Hendrich B., Reik W. and Dean W. (2002) Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* **241**: 172-182.

Satchwell S.C., Drew H.R. and Travers A.A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol* **191**: 659-675.

Schaller H., Gray C. and Herrmann K. (1975) Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proc Natl Acad Sci U S A* **72**: 737-741.

Schiesser S., Pfaffeneder T., Sadeghian K., Hackner B., Steigenberger B., Schroder A.S., Steinbacher J., Kashiwazaki G., Hofner G., Wanner K.T., Ochsenfeld C. and Carell T. (2013) Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J Am Chem Soc* **135**: 14593-14599.

Schilling E. and Rehli M. (2007) Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics* **90**: 314-323.

Scholer H.R. and Gruss P. (1984) Specific interaction between enhancer-containing molecules and cellular components. *Cell* **36**: 403-411.

Schones D.E., Cui K., Cuddapah S., Roh T.Y., Barski A., Wang Z., Wei G. and Zhao K. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**: 887-898.

Schorderet D.F. and Gartler S.M. (1992) Analysis of CpG suppression in methylated and nonmethylated species. *Proc Natl Acad Sci U S A* **89**: 957-961.

Schubeler D., Lorincz M.C., Cimbora D.M., Telling A., Feng Y.Q., Bouhassira E.E. and Groudine M. (2000) Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation. *Mol Cell Biol* **20**: 9103-9112.

Schubeler D. (2012) Molecular biology. Epigenetic islands in a genetic ocean. *Science* **338**: 756-757.

Seisenberger S., Peat J.R. and Reik W. (2013) Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells. *Curr Opin Cell Biol* **25**: 281-288.

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

Serandour A.A., Avner S., Percevault F., Demay F., Bizot M., Lucchetti-Miganeh C., Barloy-Hubler F., Brown M., Lupien M., Metivier R., Salbert G. and Eeckhoute J. (2011) Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res* **21**: 555-565.

Serandour A.A., Avner S., Oger F., Bizot M., Percevault F., Lucchetti-Miganeh C., Palierne G., Gheeraert C., Barloy-Hubler F., Peron C.L., Madigou T., Durand E., Froguel P., Staels B., Lefebvre P., Metivier R., Eeckhoute J. and Salbert G. (2012) Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res* **40**: 8255-8265.

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

Sharma S., De Carvalho D.D., Jeong S., Jones P.A. and Liang G. (2011) Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. *PLoS Genet* **7**: e1001286.

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

Shen L., Wu H., Diep D., Yamaguchi S., D'Alessio A.C., Fung H.L., Zhang K. and Zhang Y. (2013) Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**: 692-706.

Shirohzu H., Kubota T., Kumazawa A., Sado T., Chijiwa T., Inagaki K., Suetake I., Tajima S., Wakui K., Miki Y., Hayashi M., Fukushima Y. and Sasaki H. (2002) Three novel DNMT3B mutations in Japanese patients with ICF syndrome. *Am J Med Genet* **112**: 31-37.

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

Siggers T., Duyzend M.H., Reddy J., Khan S. and Bulyk M.L. (2011) Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol Syst Biol* **7**: 555.

Simpson R.T. and Shindo H. (1979) Conformation of DNA in chromatin core particles containing poly(dAdT)-poly(dAdT) studied by 31 P NMR spectroscopy. *Nucleic Acids Res* **7**: 481-492.

Song C.X., Szulwach K.E., Fu Y., Dai Q., Yi C., Li X., Li Y., Chen C.H., Zhang W., Jian X., Wang J., Zhang L., Looney T.J., Zhang B., Godley L.A., Hicks L.M., Lahn B.T., Jin P. and He C. (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* **29**: 68-72.

Song C.X., Szulwach K.E., Dai Q., Fu Y., Mao S.Q., Lin L., Street C., Li Y., Poidevin M., Wu H., Gao J., Liu P., Li L., Xu G.L., Jin P. and He C. (2013) Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**: 678-691.

Song J., Rechkoblit O., Bestor T.H. and Patel D.J. (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**: 1036-1040.

Song J., Teplova M., Ishibe-Murakami S. and Patel D.J. (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* **335**: 709-712.

Song L., Zhang Z., Grasfeder L.L., Boyle A.P., Giresi P.G., Lee B.K., Sheffield N.C., Graf S., Huss M., Keefe D., Liu Z., London D., McDaniell R.M., Shibata Y., Showers K.A., Simon J.M., Vales T., Wang T., Winter D., Zhang Z., Clarke N.D., Birney E., Iyer V.R., Crawford G.E., Lieb J.D. and Furey T.S. (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* **21**: 1757-1767.

Spitz F. and Furlong E.E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**: 613-626.

Spruijt C.G., Gnerlich F., Smits A.H., Pfaffeneder T., Jansen P.W., Bauer C., Munzel M., Wagner M., Muller M., Khan F., Eberl H.C., Mensinga A., Brinkman A.B., Lephikov K., Muller U., Walter J., Boelens R., van Ingen H., Leonhardt H., Carell T. and Vermeulen M. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**: 1146-1159.

Stadler M.B., Murr R., Burger L., Ivanek R., Lienert F., Scholer A., van Nimwegen E., Wirbelauer C., Oakeley E.J., Gaidatzis D., Tiwari V.K. and Schubeler D. (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490-495.

Statham A.L., Robinson M.D., Song J.Z., Coolen M.W., Stirzaker C. and Clark S.J. (2012) Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res* **22**: 1120-1127.

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

Strahl B.D. and Allis C.D. (2000) The language of covalent histone modifications. *Nature* **403**: 41-45.

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

Struhl K. and Segal E. (2013) Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**: 267-273.

Suzuki M.M. and Bird A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**: 465-476.

Tahiliani M., Koh K.P., Shen Y., Pastor W.A., Bandukwala H., Brudno Y., Agarwal S., Iyer L.M., Liu D.R., Aravind L. and Rao A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930-935.

Takahashi K. and Yamanaka S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663-676.

Tan L., Xiong L., Xu W., Wu F., Huang N., Xu Y., Kong L., Zheng L., Schwartz L., Shi Y. and Shi Y.G. (2013) Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic Acids Res* **41**: e84.

Tate P.H. and Bird A.P. (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev* **3**: 226-231.

Taverna S.D., Ilin S., Rogers R.S., Tanny J.C., Lavender H., Li H., Baker L., Boyle J., Blair L.P., Chait B.T., Patel D.J., Aitchison J.D., Tackett A.J. and Allis C.D. (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**: 785-796.

Taylor I.C., Workman J.L., Schuetz T.J. and Kingston R.E. (1991) Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. *Genes Dev* **5**: 1285-1298.

Teif V.B., Vainshtein Y., Caudron-Herger M., Mallm J.P., Marth C., Hofer T. and Rippe K. (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol* **19**: 1185-1192.

Thomassin H., Flavin M., Espinas M.L. and Grange T. (2001) Glucocorticoid-induced DNA demethylation and gene memory during development. *EMBO J* **20**: 1974-1983.

Thomson J.P., Skene P.J., Selfridge J., Clouaire T., Guy J., Webb S., Kerr A.R., Deaton A., Andrews R., James K.D., Turner D.J., Illingworth R. and Bird A. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**: 1082-1086.

Trojer P. and Reinberg D. (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* **28**: 1-13.

Trompouki E., Bowman T.V., Lawton L.N., Fan Z.P., Wu D.C., DiBiase A., Martin C.S., Cech J.N., Sessa A.K., Leblanc J.L., Li P., Durand E.M., Mosimann C., Heffner G.C., Daley G.Q., Paulson R.F., Young R.A. and Zon L.I. (2011) Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**: 577-589.

Tsai M.C., Manor O., Wan Y., Mosammaparast N., Wang J.K., Lan F., Shi Y., Segal E. and Chang H.Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**: 689-693.

Tsankov A.M., Thompson D.A., Socha A., Regev A. and Rando O.J. (2010) The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol* **8**: e1000414.

Tsukahara S., Kobayashi A., Kawabe A., Mathieu O., Miura A. and Kakutani T. (2009) Bursts of retrotransposition reproduced in Arabidopsis. *Nature* **461**: 423-426.

Tsumura A., Hayakawa T., Kumaki Y., Takebayashi S., Sakaue M., Matsuoka C., Shimotohno K., Ishikawa F., Li E., Ueda H.R., Nakayama J. and Okano M. (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**: 805-814.

Tucker K.L., Beard C., Dausmann J., Jackson-Grusby L., Laird P.W., Lei H., Li E. and Jaenisch R. (1996) Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev* **10**: 1008-1020.

Tweedie S., Charlton J., Clark V. and Bird A. (1997) Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol* **17**: 1469-1475.

Valinluck V., Tsai H.H., Rogstad D.K., Burdzy A., Bird A. and Sowers L.C. (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res* **32**: 4100-4108.

Valinluck V. and Sowers L.C. (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* **67**: 946-950.

Velasco G., Hube F., Rollin J., Neuillet D., Philippe C., Bouzinba-Segard H., Galvani A., Viegas-Pequignot E. and Francastel C. (2010) Dnmt3b recruitment through E2F6

transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* **107**: 9281-9286.

Vermeulen M., Mulder K.W., Denissov S., Pijnappel W.W., van Schaik F.M., Varier R.A., Baltissen M.P., Stunnenberg H.G., Mann M. and Timmers H.T. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**: 58-69.

Verzijlbergen K.F., van Welsem T., Sie D., Lenstra T.L., Turner D.J., Holstege F.C., Kerkhoven R.M. and van Leeuwen F. (2011) A barcode screen for epigenetic regulators reveals a role for the NuB4/HAT-B histone acetyltransferase complex in histone turnover. *PLoS Genet* **7**: e1002284.

Voss T.C., Schiltz R.L., Sung M.H., Yen P.M., Stamatoyannopoulos J.A., Biddie S.C., Johnson T.A., Miranda T.B., John S. and Hager G.L. (2011) Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* **146**: 544-554.

Waddington C.H. (2012) The epigenotype. 1942. *Int J Epidemiol* **41**: 10-13.

Walsh C.P., Chaillet J.R. and Bestor T.H. (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**: 116-117.

Wang R.Y., Gehrke C.W. and Ehrlich M. (1980) Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res* **8**: 4777-4790.

Watanabe D., Uchiyama K. and Hanaoka K. (2006) Transition of mouse de novo methyltransferases expression from Dnmt3b to Dnmt3a during neural progenitor cell development. *Neuroscience* **142**: 727-737.

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

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

Welham M.J., Kingham E., Sanchez-Ripoll Y., Kumpfmueller B., Storm M. and Bone H. (2011) Controlling embryonic stem cell proliferation and pluripotency: the role of PI3K- and GSK-3-dependent signalling. *Biochem Soc Trans* **39**: 674-678.

Whitehouse I. and Tsukiyama T. (2006) Antagonistic forces that position nucleosomes in vivo. *Nat Struct Mol Biol* **13**: 633-640.

Wiench M., John S., Baek S., Johnson T.A., Sung M.H., Escobar T., Simmons C.A., Pearce K.H., Biddie S.C., Sabo P.J., Thurman R.E., Stamatoyannopoulos J.A. and Hager G.L. (2011) DNA methylation status predicts cell type-specific enhancer activity. *EMBO J* **30**: 3028-3039.

Wigler M., Levy D. and Perucho M. (1981) The somatic replication of DNA methylation. *Cell* **24**: 33-40.

Williams K., Christensen J., Pedersen M.T., Johansen J.V., Cloos P.A., Rappsilber J. and Helin K. (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**: 343-348.

Wippo C.J., Israel L., Watanabe S., Hochheimer A., Peterson C.L. and Korber P. (2011) The RSC chromatin remodelling enzyme has a unique role in directing the accurate positioning of nucleosomes. *EMBO J* **30**: 1277-1288.

Workman J.L. and Kingston R.E. (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem* **67**: 545-579.

Wossidlo M., Nakamura T., Lepikhov K., Marques C.J., Zakhartchenko V., Boiani M., Arand J., Nakano T., Reik W. and Walter J. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* **2**: 241.

Wu H., D'Alessio A.C., Ito S., Wang Z., Cui K., Zhao K., Sun Y.E. and Zhang Y. (2011) Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev* **25**: 679-684.

Wu S.C. and Zhang Y. (2010) Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* **11**: 607-620.

Wyatt G.R. (1951) Recognition and estimation of 5-methylcytosine in nucleic acids. *Biochem J* **48**: 581-584.

Wyatt G.R. and Cohen S.S. (1953) The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *Biochem J* **55**: 774-782.

Xie W., Schultz M.D., Lister R., Hou Z., Rajagopal N., Ray P., Whitaker J.W., Tian S., Hawkins R.D., Leung D., Yang H., Wang T., Lee A.Y., Swanson S.A., Zhang J., Zhu Y., Kim A., Nery J.R., Urich M.A., Kuan S., Yen C.A., Klugman S., Yu P., Suknuntha K., Propson N.E., Chen H., Edsall L.E., Wagner U., Li Y., Ye Z., Kulkarni A., Xuan Z., Chung W.Y., Chi N.C., Antosiewicz-Bourget J.E., Slukvin I., Stewart R., Zhang M.Q., Wang W., Thomson J.A., Ecker J.R. and Ren B. (2013) Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**: 1134-1148.

Xu J., Watts J.A., Pope S.D., Gadue P., Kamps M., Plath K., Zaret K.S. and Smale S.T. (2009) Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev* **23**: 2824-2838.

Xu Y., Wu F., Tan L., Kong L., Xiong L., Deng J., Barbera A.J., Zheng L., Zhang H., Huang S., Min J., Nicholson T., Chen T., Xu G., Shi Y., Zhang K. and Shi Y.G. (2011) Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell* **42**: 451-464.

Xu Y., Xu C., Kato A., Tempel W., Abreu J.G., Bian C., Hu Y., Hu D., Zhao B., Cerovina T., Diao J., Wu F., He H.H., Cui Q., Clark E., Ma C., Barbara A., Veenstra G.J., Xu G., Kaiser U.B., Liu X.S., Sugrue S.P., He X., Min J., Kato Y. and Shi Y.G. (2012) Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell* **151**: 1200-1213.

Yamazaki Y., Mann M.R., Lee S.S., Marh J., McCarrey J.R., Yanagimachi R. and Bartolomei M.S. (2003) Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning. *Proc Natl Acad Sci U S A* **100**: 12207-12212.

Yildirim O., Li R., Hung J.H., Chen P.B., Dong X., Ee L.S., Weng Z., Rando O.J. and Fazzio T.G. (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* **147**: 1498-1510.

Yu M., Hon G.C., Szulwach K.E., Song C.X., Zhang L., Kim A., Li X., Dai Q., Shen Y., Park B., Min J.H., Jin P., Ren B. and He C. (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**: 1368-1380.

Yuan G.C., Liu Y.J., Dion M.F., Slack M.D., Wu L.F., Altschuler S.J. and Rando O.J. (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. *Science* **309**: 626-630.

Zacharias H. (1995) Emil Heitz (1892-1965): chloroplasts, heterochromatin, and polytene chromosomes. *Genetics* **141**: 7-14.

Zaidi S.K., Young D.W., Pockwinse S.M., Javed A., Lian J.B., Stein J.L., van Wijnen A.J. and Stein G.S. (2003) Mitotic partitioning and selective reorganization of tissue-specific transcription factors in progeny cells. *Proc Natl Acad Sci U S A* **100**: 14852-14857.

Zaret K.S. and Carroll J.S. (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**: 2227-2241.

Zemach A. and Zilberman D. (2010) Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol* **20**: R780-785.

Zemach A., McDaniel I.E., Silva P. and Zilberman D. (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**: 916-919.

Zhang X., Yazaki J., Sundaresan A., Cokus S., Chan S.W., Chen H., Henderson I.R., Shinn P., Pellegrini M., Jacobsen S.E. and Ecker J.R. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**: 1189-1201.

Zhang Y., Moqtaderi Z., Rattner B.P., Euskirchen G., Snyder M., Kadonaga J.T., Liu X.S. and Struhl K. (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol* **16**: 847-852.

Zhang Z., Wippo C.J., Wal M., Ward E., Korber P. and Pugh B.F. (2011) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* **332**: 977-980.

Zhu B., Zheng Y., Hess D., Angliker H., Schwarz S., Siegmann M., Thiry S. and Jost J.P. (2000) 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proc Natl Acad Sci U S A* **97**: 5135-5139.

Zilberman D., Gehring M., Tran R.K., Ballinger T. and Henikoff S. (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61-69.

Ziller M.J., Gu H., Muller F., Donaghey J., Tsai L.T., Kohlbacher O., De Jager P.L., Rosen E.D., Bennett D.A., Bernstein B.E., Gnirke A. and Meissner A. (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**: 477-481.

6 Curriculum vitae

Angelika Feldmann

Education

Nov 2009 to Apr 2014	**PhD studies in Biology** Friedrich-Miescher Institute for Biomedical Research (FMI), Basel, Switzerland Supervisor: Prof. Dr. Dirk Schübeler Title: Turnover and function of DNA methylation at transcription factor binding sites
Sep 2009	**Diploma in Molecular Medicine** Albert-Ludwigs-University Freiburg, Germany
Oct 2008 to Sep 2009	**Diploma thesis in Genetics and Bioinformatics** Albert-Ludwigs-University/ZBSA Freiburg, Germany Supervisor: Prof. Dr. Ralf Baumeister Title: Establishment of a Tet-on System in *C. elegans*
Oct 2004 to Sep 2009	**Studies in Molecular Medicine** Albert-Ludwigs-University Freiburg, Germany

Research experience

May to June 2008	**Internship in Molecular Neurosurgery** University Clinical Centre Freiburg, Germany Supervisor: Prof. Dr. Guido Nikkhah Project: Establishment of brain-tumour stem-cell system, analysis of neuronal stem-cell transplants in rat brain
Nov 2007 to Oct 2008	**Research Assistant in Genetics and Bioinformatics** Albert-Ludwigs-University Freiburg, Germany Supervisor: Prof. Dr. Ralf Baumeister Project: Tet-inducible systems in *C. elegans*
Oct to Nov 2007	**Internship in Genetics and Bioinformatics** Albert-Ludwigs-University Freiburg, Germany Supervisor: Prof. Dr. Ralf Baumeister Project: Establishment of a baculovirus expression system, *C. elegans* model of Parkinson's Disease
May 2006 to Nov 2007	**Research Assistant in Immunology** Max-Planck-Institute of Immunology (MPI), Freiburg, Germany Supervisor: Prof. Dr. Michael Reth, Dr. Elias Hobeika Project: Characterization of a mouse model for B-cell development

Publications

Feldmann A., Ivanek R., Murr R., Gaidatzis D., Burger L. and Schubeler D. (2013) Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. PLoS Genet 9: e1003994.

Di Cerbo V., Mohn F., Ryan D.P., Montellier E., Kacem S., Tropberger P., Kallis E., Holzner M., Hoerner L., **Feldmann A.**, Richter F.M., Bannister A.J., Mittler G., Michaelis J., Khochbin S., Feil R., Schuebeler D., Owen-Hughes T., Daujat S. and Schneider R. (2014) Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. Elife 3: e01632.

Grants and prizes

Oct 2013	Poster prize (second, FEBS workshop, Capri)
March 2013	Poster prize (equal third, Epigenetics and Chromatin Conference, Boston)
March 2011	Prize for best presentation and participation (7th Course of Epigenetics, Paris)
Jan 2010 to Dec 2012	Fellow of the Marie-Curie Initial Training Network “Nucleosome 4D”
2004	“Carl-von-Fischer” student prize in biology