Local Regulation of DNA Methylation by the Transcription Factor REST

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Juliane Schmidt

aus Halberstadt, Deutschland

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von Prof. Dr. Dirk Schübeler und Dr. Duncan Odom.

Basel, den 21.06.2016

Prof. Dr. Dirk Schübeler

Prof. Dr. Jörg Schibler

To my parents and my grandfather Benno.

Acknowledgements

First and foremost, I want to thank my doctoral thesis supervisor Prof. Dirk Schübeler for his continuous support over the last years. I am deeply thankful for the opportunity to have worked on this intriguing project. I acknowledge his patience and trust, which has allowed me to become a more independent and critical researcher. Moreover, I am thankful to Dr. Arnaud Krebs for his guidance and feedback on my project. His support was instrumental to the success of the project. I acknowledge essential bioinformatics support by Dr. Michael Stadler, Dr. Altuna Akalin, Dr. Arnaud Krebs, Dr. Anais Bardet and Dr. Lukas Burger. In detail, Dr. Altuna Akalin and Dr. Arnaud Krebs created a computational pipeline for analyzing amplicon Bisulfite sequencing data. Dr. Michael Stadler analyzed RNA-seq data, REST interactor ChIP-seqs and was critically involved in many other data sets. Dr. Anais Bardet analyzed the DNase-I seq experiments. Dr. Lukas Burger contributed to data analysis of ChIP-seq experiments. I am thankful to the FMI Genomics Facility, particularly to Dr. Sophie Dessus-Babus for her support with high-throughput sequencing. I thank Christiane Wirbelauer and Leslie Hoerner for technical support within the Schübeler lab. Thanks to Dr. Nicolas Thomä and Dr. Duncan Odom for offering scientific feedback as my doctoral thesis advisors.

I am very grateful to Boehringer Ingelheim Fonds for generously funding my PhD project as well as attendance to conferences and courses. I have made invaluable professional and personal connections during that time.

Several people have supported me along different steps of my academic career; I want to thank all of them. Special thanks to Prof. Markus Schwaninger, Dr. Gary Parkinson, Dr. Janna Seifried, Dr. Susann Stuttfeld (née Kumpf), Prof. Romeo Ricci, Dr. Stefan Metzler, Prof. Wilhelm Krek and Prof. Dirk Schübeler. Most recently, I thank Dr. Carien Dekker for being an extraordinary mentor as part of the University of Basel Antelope program. More female researchers need to be encouraged by a supportive network.

Personally, I thank Dr. Darko Barisic for being his entertaining and loyal self throughout our time in Basel. And most importantly, thanks to my parents and my brother for everything.

Summary

Transcriptional regulation in eukaryotes is realized through intricate interactions between transcription factors and chromatin. DNA methylation constitutes a chromatin modification that is associated with transcriptional silencing (Deaton and Bird, 2011). Whole-genome methylation profiling in mammals has revealed widespread cytosine methylation with characteristic hypomethylation at *cis*-regulatory elements. Hypomethylation is typically present within CpG islands and distal CpG-poor regions (Stadler et al., 2011). Previous investigations have shown, that some DNA-binding factors like the RE1-silencing transcription factor (REST) directly reduce methylation at these sites. However, how DNA-binding factors mediate such local methylation changes remains largely unknown.

Hence, I studied the regulation of DNA methylation by the transcription factor REST in mouse embryonic stem cells (mESCs). I ectopically expressed different REST mutants and profiled DNA methylation at distal REST binding sites. While the full-length protein is necessary and sufficient to reduce methylation at its binding sites, REST's DNA-binding domain lacks this ability. Instead, hypomethylation at binding sites required DNA-binding factors with interaction domains. The N-terminal REST mutant for example recruits SIN3A to binding sites and shows strong DNA demethylation ability. These experiments suggest that hypomethylation is not an obligatory consequence of protein binding, but rather requires interaction domains, reflecting the potential involvement of cofactors. I inquired whether TET enzymes contribute to reduced methylation within REST binding sites. Complete *Tet1/2/3* deficiency in mouse stem cells caused a strong localized hypermethylation in the immediate vicinity of the REST motif. Whether TET proteins are recruited to REST binding sites through common cofactors or indirect mechanisms remains to be determined. I also characterized chromatin accessibility and nucleosome positioning in the different REST mutant re-expression cells. Interestingly, REST mutants that were competent to decrease DNA methylation also increased chromatin accessibility and nucleosome

positioning. This could potentially link the chromatin remodeling ability of transcription factors to hypomethylation around binding sites.

In summary, the presented study dissected REST induced methylation patterns around binding sites and described several of its required molecular components. This presents an example for a dynamic interplay between genetic and epigenetic information.

Table of Contents

Chapter 1
Introduction

1.1 Thoughts on Epigenetics

Only few questions have continuously puzzled humans similarly as the essence of their own making. For scientists this means addressing the question of how a fertilized egg can develop into a complex multicellular organism. The path from a single cell to a unique human being is long; and a tail of decision-making and information integration.

One question particularly occupies modern biologists: how can the same genome give rise to hundreds of different human cell types and tissues? How is the unambiguous path of cell type determination possible without writing or deleting information? Why does a muscle cell in our body always stay a muscle cell and does not become a hepatocyte? Answers to these questions might be found in the realms of epigenetics. While linguists would abstractly translate “epi-genetics” into the studies of the mechanisms above genetics, modern biologists seem to struggle with a precise terminology. One of the more frequently agreed upon definitions, states that epigenetics is the “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” (Russo V.E.A. et al., 1996). In recent years, biologists have fought over the absolute requirement of heritability and what to call “epigenetic”.

Untroubled by such semantic discussions, the public shows a great interest in this much debated research area. With it come people’s hopes that individuals are not merely the sum of their genes but rather the result of autonomous decisions and environmental experiences. While huge medical promises were based on the sequencing of the human genome, many were dampened when its first draft was presented in 2001. Scientists were left with a book written in four letters, without knowing how to assemble more than word fragments. Epigenetics in a way offers the possibility to organize these words into more meaningful

structures. The advent of next-generation sequencing has indeed allowed us to describe many more functional elements than just open reading frames. International consortia such as ENCODE have gathered unique human and financial resources to profile chromatin modifications throughout different healthy and diseased cell types. The future will tell how many translational benefits will arise directly and indirectly from those efforts.

Well, then what is there to discover yet? With powerful tools like high-throughput sequencing, CRISPR genome editing technologies and single cell assays, possibilities seem limitless. Future's main challenge will be to overcome the difficult cause versus consequence struggle. In chromatin research this means to discern transcriptionally instructive modifications from those that reflect ongoing activity. While we have managed to assemble an impressive amount of information about chromatin, it now becomes decisive to curate it. Eventually, we will also need to distinguish biologically relevant from merely sequenceable chromatin modifications. In any case, it will be an interesting future for epigenetic research. I take it as a sign of progress that this doctoral thesis might be partially outdated by the time it reaches its readers.

Basel, June 2016

1.2 Transcription and its Regulation in Eukaryotes

Spatio-temporal control of gene expression brings about the vast complexity of multicellular organisms. In eukaryotes, regulation of gene expression is often controlled at the step of transcriptional initiation (Sainsbury et al., 2015). Unlike in prokaryotes, DNA in eukaryotes is organized into chromatin and generally not permissive to transcription (Struhl, 1999). Transcriptional initiation is realized through the complex cooperation between several proteins, which recognize and bind promoter DNA to initiate pre-mRNA synthesis.

To initiate transcription, RNA Polymerase II assembles with the general transcription factors TFIIB, TFIID, TFIIE, TFIIF and TFIIH at the promoter to form the pre-initiation complex (PIC) (Sainsbury et al., 2015). The TFIID complex contains the TATA-binding protein (TBP) and binds to the TATA box within the core promoter. Assembly of the general transcriptional machinery however induces only basal levels of transcription. Efficient transcriptional initiation crucially depends on the activity of additional regulatory proteins. Sequence-specific transcription factors bind short DNA motifs within promoter and distal regulatory regions like enhancers. Transcription factors can activate or repress transcription by recruitment of several chromatin-modifying enzymes. Distal enhancers can interact with corresponding proximal promoter regions by DNA looping. The multi-subunit protein complex mediator links specific DNA-binding factors and general transcription factors to the carboxy-terminal domain of RNA Polymerase II (Allen and Taatjes, 2015; Conaway and Conaway, 2011).

Overall, the concerted action of Polymerase II, transcription factors, chromatin-modifying enzymes and mediator induces effective engagement of the transcriptional machinery to the promoter. The process of transcriptional initiation is thereby a well-concerted interplay between *cis*-regulatory information and chromatin.

1.3 Functional Organization of the Genome into Chromatin

Gene regulation in eukaryotes is realized through *trans*-acting factors that bind *cis*-regulatory elements in the context of chromatin. Chromatin is defined as the complex organization of DNA, histones and non-histone proteins. Transcriptional regulation is achieved by modifying nucleosomes and/or DNA. Nucleosomes can be altered in their genomic position, their histone composition or by post-translational histone modifications. Cytosines within CpG dinucleotides are often methylated in mammalian genomes. The interaction between these genetic and epigenetic components is highly relevant for the accurate execution of gene expression. The following project focuses on aspects of transcription factor mediated changes in DNA methylation. Separate chapters will be dedicated to roles of transcription factors and DNA methylation (Chapter 1.4 and Chapter 1.5). First, the following sections will introduce the general concept of chromatin organization and histone modifications.

1.3.1 From Nucleosomes to Higher Order Chromatin Structures

The eukaryotic genome is highly organized and compacted in the nucleus (Figure 1.1). Within the basic unit of chromatin, 146 bp of DNA are wrapped around nucleosomes (Luger et al., 1997). The nucleosomal core particle interacts with the negatively charged DNA backbone through electrostatic interactions (Luger et al., 1997). The latter structure has been termed the 10 nm bead-on-a-string-fiber. Nucleosomes are comprised of an octamer, which consists of the highly conserved histones H2A, H2B, H3 and H4. Nucleosomes are assembled with the help of chaperones from two H2A/H2B dimers and one H3/H4 tetramer (Elsässer and D'Arcy, 2012). Histone variants can replace canonical histones in specific genomic contexts (Maze et al., 2014). Nucleosomes are spaced at defined linker lengths (average linker length in mESCs: 40 bp) and binding of the linker histone H1 is thought to determine the secondary structure of chromatin (Luger, 2003; Teif et al., 2012; Widom and Klug, 1985). The arrangement of chromatin into the 30 nm fiber *in vivo* however is still highly debated (Tremethick, 2007).

Nucleosomes are generally thought to act as a barrier to protein binding (Soufi et al., 2015; Struhl, 1999). Only specific transcription factors, so-called pioneering factors possess the ability to bind their target motifs on nucleosomes (Soufi et al., 2015). The re-positioning of nucleosomes is therefore one way to alter chromatin structure and accessibility. Active promoters often possess a distinct nucleosomal organization, where the vicinity of the TSS is nucleosome-free and the two nucleosomes at the plus one and minus one position are highly positioned (Teif et al., 2012). Nucleosomes can also be phased around distal transcription factor binding sites as observed for the insulator protein CTCF (Fu et al., 2008; Teif et al., 2014). Chromatin remodeling enzymes are the main regulators of nucleosomal organization and location (Clapier and Cairns, 2009).

Higher-order chromatin structures are just starting to be characterized. Recent high-throughput chromatin capture experiments have identified topologically associating domains (TADs) as a substructure within chromatin compartments (Dixon et al., 2012). Local chromatin interactions structure the genome into highly conserved megabase-sized TADs (Dixon et al., 2012). The transcriptional consequences of functional organization into TADs are currently under investigation (Nora et al., 2012). Next to intramolecular interactions, DNA forms interactions with the nuclear envelope. Lamina-associated domains (LADs) are domains of 0.1-10 megabases in size and are characterized by low gene expression levels (Guelen et al., 2008). These two examples demonstrate that mammalian chromosomes are structured and functionally divided into large distinct domains. Individual chromosomes often occupy distinct nuclear territories (Cremer and Cremer, 2010). Further classification then distinguishes chromatin by its DNA staining signal into heterochromatin and euchromatin, where heterochromatin is generally condensed and inactive.

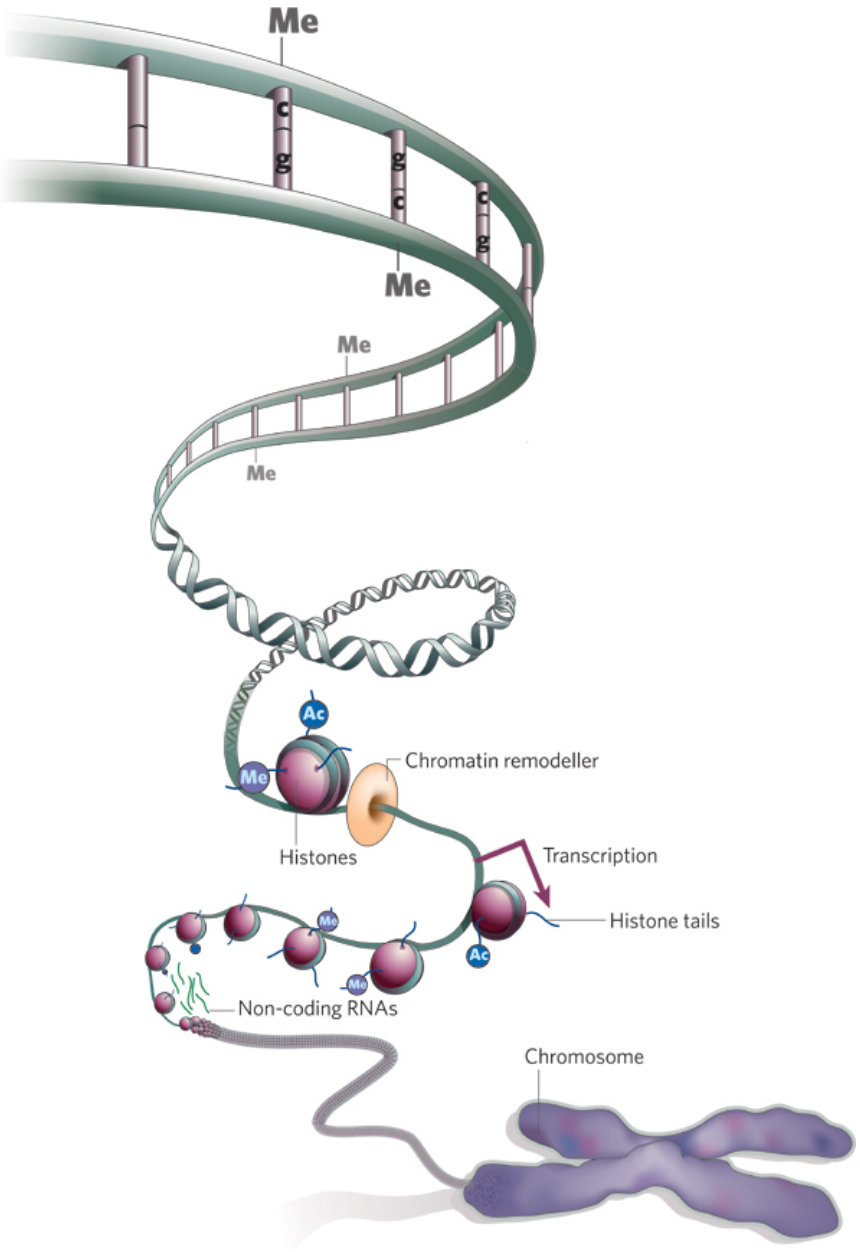

Figure 1.1 | The organization of the mammalian genome. The DNA helix is wrapped around nucleosomes to form the 10 nm bead-on-a-string-fiber. Higher order secondary structures are formed and highest compaction is achieved in metaphase chromosomes. Chromatin, which is the defining unit of DNA, histones and non-histone proteins, can be modified. In mammals, DNA is usually methylated at cytosines within CpG dinucleotides. In addition, nucleosomes can be positioned by chromatin remodeling enzymes. Histones are subjected to many post-translational modifications. The complex interaction between DNA-binding factors and *cis*-regulatory elements enables spatio-temporal regulation of gene expression. (Adapted from Jones et al., 2008; content reproduced with permission of Nature Publishing Group)

1.3.2 Histone Modifications

Covalent modification of histone proteins is one fundamental mechanism of epigenetic regulation. The nucleosomal core histones contain a globular histone fold domain and an N-terminal unstructured domain that protrudes as a tail from the nucleosomal core (Luger et al., 1997). Histone tails in particular show a striking number of post-translational modifications.

Histone modifications are thought to affect gene expression by two main mechanisms (Kouzarides, 2007). The first involves direct structural perturbation of histone and DNA contacts. In this regard, histone phosphorylation and acetylation reduce the positive charge of basic amino acids and thereby lead to a decompaction of chromatin. Increased chromatin accessibility would subsequently facilitate binding of the transcriptional machinery. Allfrey et al.'s pioneering work in the 1960s provided first evidence that histone acetylation reduces the inhibitory effect of histones on transcription (Allfrey et al., 1964). In a second mechanism, histone modifications act through the recruitment of chromatin-modifying proteins (Kouzarides, 2007). Numerous such chromatin factors have been shown to specifically read histone modifications via distinct domains.

Since the first description of histone methylation and acetylation many more histone modifications have been discovered (Allfrey et al., 1964; Bannister and Kouzarides, 2011). With the realization of the number of different histone modifications, Allis et al. suggested the existence of a histone code (Jenuwein and Allis, 2001; Strahl and Allis, 2000). Based on the histone code hypothesis, specific and distinct transcriptional output is realized through the combinatorial power of different chromatin modifications (Jenuwein and Allis, 2001; Strahl and Allis, 2000). Genome-wide profiling of many known chromatin modifications in *Drosophila* has however not substantiated such a complex code (Filion et al., 2010; Schübeler et al., 2004). In fact, the *Drosophila* genome can be segmented into five distinct chromatin types (Filion et al., 2010). Each chromatin type is defined by a specific combination of chromatin proteins and forms domains with unique functional characteristics. Similar efforts in human cell types have described a discrete number of chromatin states and have enabled functional annotation of the human genome (Dunham et al., 2012; Ernst et al., 2011). A

combination of such chromatin features can be used to predict gene expression in different cellular contexts (Dong et al., 2012).

These studies have generally described nucleosomes that are either marked by active or inactive histone modifications. A notable exception from this rule is found at so-called bivalent promoters, which are characterized by co-localization of H3K4me3 and H3K27me3 in embryonic stem cells (Bernstein et al., 2006). Active promoters are marked by H3K4me3 and polymerase binding, while active enhancers possess H3K4me1 and H3K27ac (Bernstein et al., 2005; Guenther et al., 2007; Heintzman et al., 2007). Repressed promoters are either characterized by the Polycomb modification H3K27me3 or by mutually exclusive DNA methylation (Mikkelsen et al., 2007; Mohn et al., 2008). Actively transcribed gene bodies are marked by H3K36me3 (Schaft et al., 2003). Much of the genome is transcriptionally silent with H3K9me3 and DNA methylation (Mikkelsen et al., 2007).

In summary, histone modifications enable additional means to control ubiquitous and cell-type specific gene expression. Still, issues regarding the mechanism of chromatin mediated transcriptional regulation remain unresolved. Many transcriptionally instructive histone modifications still need to be separated from those merely reflecting ongoing transcriptional activity, and the mode of mitotic inheritance of many histone modifications is still unknown.

1.4 Transcription Factors Act on Chromatin

Gene expression in eukaryotes is realized through a complex system of *trans*-acting factors and *cis*-regulatory regions. Correct execution of transcription critically depends on reliable recognition of functional genomic elements. Transcription factors act on chromatinized binding sites and thereby link classical genetic information to epigenetic states. The power of upstream transcription factor regulation was demonstrated in the seminal work of Davis et al., when the expression of the single basic loop-helix factor MYOD1 efficiently converted fibroblasts into myoblasts (Davis et al., 1987). The next chapters will introduce these major players of transcriptional regulation in more detail.

1.4.1 Transcription Factor Classes

Transcription factors present one group of proteins that are essential to the process of gene transcription. The human genome encodes approximately 1400 DNA-binding factors, which constitutes about 6% of all protein-coding genes (Vaquerizas et al., 2009). Organismal complexity of the mammalian lineage was realized through expansion of its *cis*- and *trans*-regulatory repertoire alike. Depending on its involvement in transcriptional initiation, one distinguishes between general and sequence-specific transcription factors. The latter class is predominantly mediating the spatio-temporal dimension of gene regulation.

General Transcription Factors

General transcription factors are involved in the common process of transcription initiation at Polymerase II dependent promoters. They constitute an obligatory component of the basic transcriptional machinery, and often lack a defined DNA-binding domain and undergo structural reorganization upon formation of the pre-initiation complex (PIC) (Sainsbury et al., 2015). The five general transcription factors TFIIB, TFIID, TFIIE, TFIIF and TFIIH act in different steps of the pre-initiation complex formation. Together with RNA Polymerase II they bind to respective promoter regions, enable the unwinding of DNA and allow pre-mRNA synthesis. Release of the initiation factors by elongation factors initiates effective RNA Polymerase II elongation (Sainsbury et al., 2015).

Sequence-specific Transcription Factors

The class of sequence-specific transcription factors is the most studied component of the *trans*-acting module. Their activating or repressive activity enables gene-specific transcription and drives complete cell lineage decisions throughout mammalian development. Transcription factor and DNA interactions vary widely in their specificities. Some factors can bind DNA unspecifically through interactions with the DNA backbone or the minor groove (Aggarwal et al., 1988; Rohs et al., 2009; Zheng et al., 1999). In most cases, sequence-specificity is achieved through DNA-binding domains (DBDs). These domains often form hydrogen bonds with the nucleotides of target sequences (von Hippel and Berg, 1989). Technologies like high-throughput systematic evolution of ligands by exponential enrichment (SELEX) and chromatin immunoprecipitation followed by sequencing (ChIP-seq), have characterized the sequence specificities for hundreds of DNA-binding domains *in vitro* and *in vivo* (Jolma et al., 2013).

The most widely used nomenclature classifies transcription factors according to their annotated DNA-binding domain. The three largest classes in mammals are C_2H_2 zinc fingers, homeodomain and helix-loop-helix factors (Gray et al., 2004; Vaquerizas et al., 2009). Each of these classes has undergone different degrees of expansion during evolution. Zinc finger domain proteins underwent the largest growth in the primate transcription factor repertoire. An alternative approach distinguishes transcription factors according to hierarchy within the transcription factor network. Across diverse developmental pathways, so-called master transcriptional regulators integrate upstream extrinsic signals and control downstream transcription factor binding. Examples for such are the myogenic factor MYOD1 and the transcriptional repressor REST (Chong et al., 1995; Davis et al., 1987; Schoenherr and Anderson, 1995). The ability to remodel the chromatin landscape for downstream effector binding is a shared feature of pioneering factors (Magnani et al., 2011).

1.4.2 The RE1-silencing Transcription Factor

The RE1-silencing transcription factor (REST), also known as neuron-restrictive silencing factor (NRSF), is a functional repressor of neuron-specific genes in non-neuronal cell types (Ooi and Wood, 2007). REST serves as a complex platform for chromatin-modifying proteins, which regulate local chromatin structure and transcription of associated genes. Recent observations have shown REST's association with repositioned nucleosomes, diverse histone modifications and reduced DNA methylation at its binding sites. REST is as an attractive model to study the molecular space between transcription factor binding and chromatin changes that ultimately lead to transcriptional output.

Function

The transcriptional regulator REST was first described to bind a specific DNA sequence known as the repressor element 1 (RE1) (Chong et al., 1995; Schoenherr and Anderson, 1995). The canonical RE1 motif is a well-defined 21 bp long DNA sequence present in strong REST binding sites. REST was the first protein for which comprehensive genome-wide binding maps were generated by chromatin immunoprecipitation followed high-throughput sequencing (ChIP-seq) (Johnson et al., 2007).

Since many RE1 sites were identified in neuron-specific promoters, REST was suggested to be a master regulator of neurogenesis (Bessis et al., 1997; Kallunki et al., 1997; Kraner et al., 1992; Lönnerberg et al., 1996; Mieda et al., 1997; Mori et al., 1992; Wood et al., 1996). The first generated constitutive *Rest* knock out mice supported this notion by showing transcriptional upregulation of neuron-specific tubulin in non-neuronal tissues. The early embryonic lethality at day 11.5 precluded the description of possible neuronal defects (Chen et al., 1998). Investigations have shown rapid degradation of REST protein during *in vitro* neuronal differentiation and further supported a crucial function of REST during neurogenesis (Ballas et al., 2005). However, several studies have reported on more diverse functions of REST in transcriptional regulation of normal and diseased neuronal and non-neuronal cell types (Ooi and Wood, 2007). A debate on the role of REST in the maintenance of pluripotency has been widely resolved (Jørgensen and Fisher, 2010; Jørgensen et al., 2009a; Singh et al., 2008).

Contrary to initial reports, mouse embryonic stem cells (mESCs) deficient of REST show unperturbed pluripotency factor expression and give rise to all three germ layers (Jørgensen et al., 2009a).
Alternative splicing gives rise to the truncated REST4 protein in neuronal cell types. The function of REST4 in repression or de-repression of REST-regulated genes is subject of ongoing investigations (Ovando-Roche et al., 2014; Tabuchi et al., 2002).

Domain Organization

The transcriptional repressor REST belongs to the large family of Kruppel-type zinc finger proteins and functions as a recruitment platform for several macromolecular complexes (Figure 1.2). REST binding to DNA is mediated through a central DNA-binding domain that consists of eight zinc fingers. The first five zinc fingers are essential for DNA-binding, whereas the remaining ones contribute to affinity. The previously mentioned truncated REST4 isoform shows dramatically reduced DNA-binding affinity to the RE1 motif (Shimojo et al., 2001). Zinc finger five does not only contribute to DNA binding, but was suggested to function as a nuclear localization signal (Shimojo, 2006). Stretches of lysine- and proline-rich sequences are adjacent to REST's DNA-binding domain. The molecular implications of these compositional biases are currently unknown.
REST mediated transcriptional repression is largely dependent on its N- and C-terminal interaction domains with the SIN3 and CoREST repressor complexes (Tapia-Ramírez et al., 1997). The N-terminal SIN3 interaction domain has been shown to recruit both mammalian SIN3 paralogues SIN3A and SIN3B. Early studies have reported interactions of REST's N-terminus with the SIN3A paired amphipathic helix domain (PAH) 2 (Grimes et al., 2000). Amino acids 43 to 57 contained in this region were also sufficient for SIN3B recruitment. A published NMR structure indicates specific interactions of REST's hydrophobic alpha-helix (43 – 57 aa) with the hydrophobic cleft of the SIN3B PAH1 domain (Nomura et al., 2005). It remains to be resolved whether the overlap between REST's SIN3A and SIN3B interaction domains implicates simultaneous or alternating recruitment of both repressors.
The C-terminal domain of REST was shown to recruit CoREST via a single zinc finger (Andrés et al., 1999; Tapia-Ramírez et al., 1997). Mutations in this zinc

finger domain abrogate REST's binding to CoREST and abolish REST mediated transcriptional repression of reporter constructs (Andrés et al., 1999; Tapia-Ramírez et al., 1997).

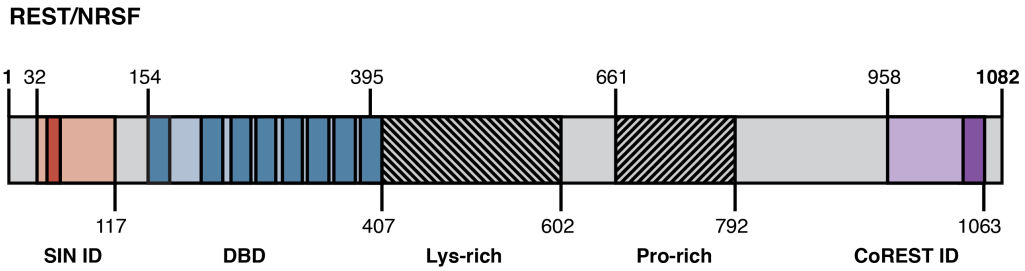

Figure 1.2 | Domain structure of the murine REST protein. The DNA-binding factor REST functions as a bipartite transcriptional repressor of mostly neuronal genes in non-neuronal tissues. It contains an N-terminal interaction domain with the co-repressor complex SIN3, enabling contacts with SIN3A (light red) and SIN3B (dark red). The C-terminus recruits the co-repressor complex CoREST mainly through a single zinc finger (dark purple). DNA binding is realized through a large eight zinc finger (dark blue) containing DNA-binding domain (DBD) (light blue). Adjacent to the DNA-binding domain are stretches of lysines and prolines. Legend: SIN ID (SIN3 interaction domain), DBD (DNA-binding domain), Lys-rich (lysine-rich), Pro-rich (proline-rich), CoREST ID (CoREST interaction domain), amino acid positions are indicated above for the full-length protein and above or below for the domain start and stop, respectively (Figure drawn based on information retrieved from UniProt, The UniProt Consortium, 2017)

Interacting Partners

As described above, the DNA-binding factor REST is characterized by a complex domain architecture that enables its binding to DNA and remodeling of local chromatin (Figure **1**.3). Many REST interacting proteins have been described with different degrees of evidence (Ooi and Wood, 2007). REST functions as a hub for different complexes, containing histone deacetylases (HDACs), nucleosome remodelers and DNA demethylases. The effective size of a completely assembled REST repressosome will likely exceed the megadalton scale. In order to understand REST's effect on transcription and chromatin, it is therefore crucial to delineate the role of its interactors.

Among the best-studied interactions of REST is the recruitment of the SIN3 complex through REST's N-terminal interaction domain. In the mouse, two paralogues of SIN3 exist in the form of SIN3A and SIN3B. Both proteins possess a high sequence similarity and share crucial domains, e.g. the PAH 1- 4 domains

and the Histone Deacetylase Interaction Domain (HID). The SIN3 proteins function as scaffolds within the larger multimeric SIN3 repressor complex (Silverstein and Ekwall, 2005). The transcriptional repression of SIN3 is mediated through its recruitment of class I HDAC enzymes HDAC1 and HDAC2 (Silverstein and Ekwall, 2005). Genome-wide mapping of REST and its cofactors has identified co-occurrence of the SIN3A and SIN3B proteins at REST binding sites (Yu et al., 2011). REST binding to DNA is correlated with significantly reduced levels of H3K9ac and H4K8ac, a likely consequence of histone deacetylase activity (Zheng et al., 2009).

The SIN3 protein does not only recruit histone deacetylases but has been suggested to interact with Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) (Cartron et al., 2013; McDonel et al., 2012; Williams et al., 2011a). Comparisons of genome-wide binding of TET1 and SIN3A in mouse embryonic stem cells indicate largely overlapping binding sites (Williams et al., 2011a). Physical interaction with SIN3A was demonstrated for TET1 and its paralogue TET2 (McDonel et al., 2012). Recent investigations in mouse retina showed interactions of REST with retinal expressed TET3 (Perera et al., 2015). Due to their involvement in the active removal of DNA methylation, the recruitment of these enzymes by transcriptional repressors such as REST and SIN3A seems enigmatic. It remains speculative whether REST activity necessitates DNA demethylation to enable downstream effector binding events.

Beyond the described SIN3 interactions, REST recruits CoREST as a second co-repressor. The complex can be comprised of the RCOR paralogues RCOR1, RCOR2 and RCOR3 (Ooi and Wood, 2007). The RCOR proteins show high sequence similarity but differ in their cell type-specific expression patterns. All three paralogues are expressed in mouse embryonic stem cells and occupy REST binding sites (Yu et al., 2011). When REST becomes downregulated during neuronal differentiation, REST binding sites consequently lose CoREST binding (Westbrook et al., 2008). Reports on the role of CoREST during neurogenesis have shown persisting RCOR1 binding to the Calbindin and BDNF promoter even in the absence of REST (Ballas et al., 2005). Whether this constitutes a genome-wide behavior of CoREST during neurogenesis remains to be determined.

Similar to its N-terminal counterpart, CoREST is sufficient to mediate transcriptional repression of a reporter construct by recruitment of the HDAC1 and HDAC2 enzymes (Andrés et al., 1999; You et al., 2001). Aside from histone deacetylases, the CoREST complex contains many more chromatin modifying enzymes such as G9a, BRG1 and LSD1 (Battaglioli et al., 2002; Lee et al., 2005; Shi et al., 2003, 2005; Zhao et al., 2006). Indeed, REST can mediate transcriptional silencing of neuronal genes not only through histone deacetylation but by methylation of lysine 9 on histone 3 (H3K9). REST's interaction with G9a coincides with localized enrichment of H3K9me2 and H3K9me3 around RE1 sites (Roopra et al., 2004; Zheng et al., 2009). Whether the interaction of C-terminal REST with G9a is direct or indirect as part of the CoREST complex is controversial (Roopra et al., 2004; Shi et al., 2003). On the other hand, the histone demethylase component LSD1 is believed to remove activating histone methylation marks on histone H3. In the context of the CoREST complex, LSD1 is suggested to decrease methylation levels of nucleosomal H3K4me2 and H3K4me3 (Shi et al., 2005; Zheng et al., 2009). Transcriptional repression by the CoREST complex was shown to be dependent on the chromatin remodeling activity of the associated SWI/SNF ATP-dependent nucleosome remodeling factor BRG1 (also known as SMARCA4). BRG1 activity could be reflected in the strong nucleosome positioning around REST binding sites.

Next to the repressive enzymes discussed, the REST motif was shown to be sufficient for deposition of the Polycomb Repressive Complex 2 (PRC2) modification H3K27me3 in neuronal progenitors (Arnold et al., 2013). Other groups have reported interactions of REST with components of the Polycomb complexes (Dietrich et al., 2012; Mozzetta et al., 2014; Tsai et al., 2010). To what extent the transient recruitment of H3K27me3 contributes to transcriptional repression of REST associated promoters is still under debate (McGann et al., 2014).

In sum, the most striking feature of the DNA-binding protein REST is its ability to recruit seemingly redundant repressor complexes. While Bingham et al. report the necessity of particular REST repressor domains for stable repression, overall data on the nature of potential redundancy remains sparse (Bingham et al., 2007). Some of the interacting proteins are purely described by protein-protein interaction data. Structural validation of many REST interactions is still lacking.

It is crucial to consider that our current knowledge is merely a static view on REST and its activity. REST mediated transcriptional regulation however, is the consequence of a multistep process including REST binding, sequential interactor recruitment and chromatin remodeling. The individual composition of REST's interactome is likely dependent on the particular context of genomic location and cell type. Further delineation of those aspects is needed to understand the sequence of events in REST mediated transcriptional repression.

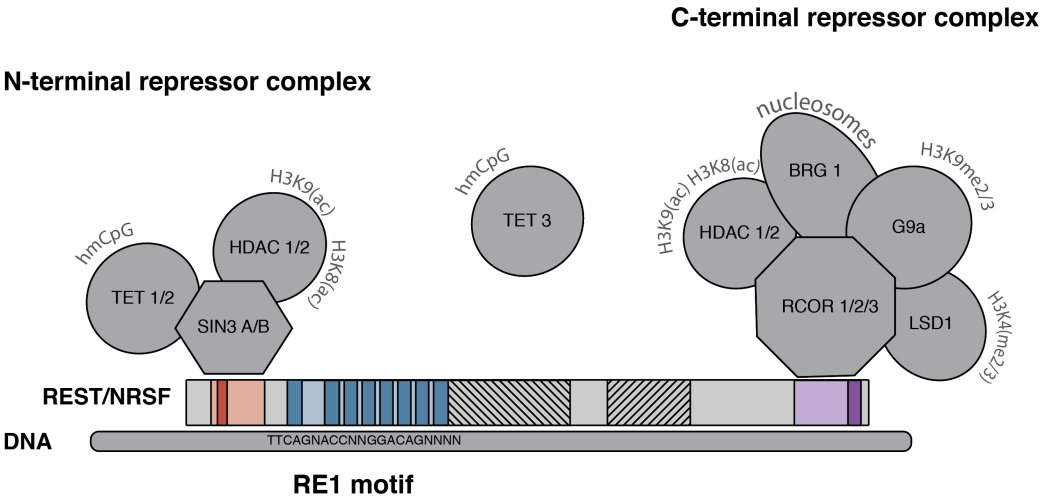

Figure 1.3 I Interaction partners of REST and their chromatin targets. The transcriptional repressor REST binds to the DNA element RE1 and mediates transcriptional repression through N- and C-terminal recruitment of the SIN3 and CoREST repressor complexes. Both serve as platforms for HDAC1 and HDAC2 enzymes, which catalyze the removal of activating histone acetylation marks. SIN3A interacts with the demethylases TET1 and TET2. The nucleosome remodeling factor BRG1, the histone demethylases LSD1 and the histone methyltransferase G9a are further constituents of the CoREST complex**.** Legend: Enzymatic products and substrates (in brackets) are written next to the enzyme in grey (Figure synthesized from references in main text, mainly Ooi and Wood, 2007)

1.5 DNA Methylation in Mammals

DNA methylation is the covalent modification of nucleotides by methyl groups. In vertebrates, methylation occurs almost exclusively at cytosines and in the context of CpG dinucleotides. Due to its maintenance during replication, cytosine methylation is often referred to as DNA's fifth base. DNA methylation was considered the classical example of a stably maintained chromatin modification. Methylation of cytosines is associated with transcriptional silencing of repetitive elements, the inactive X chromosome, imprinted regions and locus-specific promoters (Bird, 2002; De Carvalho et al., 2010; Deaton and Bird, 2011; Goll and Bestor, 2005; Jaenisch and Bird, 2003). The following chapters will focus on mammalian DNA methylation and its interaction with DNA-binding factors.

1.5.1 Chemical Modifications of DNA

Canonical DNA bases can be chemically modified by methyl groups or their oxidation products. DNA methylation is present in several prokaryotic and eukaryotic species. While prokaryotic methylation is directed to cytosine and adenine, the modification is mostly restricted to cytosine in eukaryotes.

In mammals, DNA methylation is predominantly confined to cytosines within the sequence context of 5'CpG3'. Cytosine methylation outside of these CpG dinucleotides is observable in embryonic stem cells and brain tissue, but with a low prevalence (Bird, 2002; Hon et al., 2013; Lister et al., 2009; Ramsahoye et al., 2000). Within these dinucleotides, the methyl group is covalently linked to carbon five of the cytosine pyrimidine ring and resides in the major groove of DNA. Watson-Crick base pairing is not impacted. During the process of active demethylation the methyl group can be oxidized to 5-hydroxymethylcytosine (5hmC) and its derivatives 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Figure 1.4).

In comparison to 5mC, the latter modifications are less frequent and might be considered metastable intermediates (He et al., 2011; Ito et al., 2011; Kriaucionis and Heintz, 2009; Pfaffeneder et al., 2011; Tahiliani et al., 2009).

While the majority of mammalian methylation is dedicated to cytosines, several groups have reported the presence of adenine methylation in eukaryotes (Fu et

al., 2015; Greer et al., 2015; Zhang et al., 2015). In a 2016 publication, Wu et al. report the discovery of N^6-methyladenine (N6mA) in mouse embryonic stem cells (Wu et al., 2016). Functional effects of N6mA have been reported, but comprehensive investigations analyzing this extremely rare modification are yet to come (Wu et al., 2016).

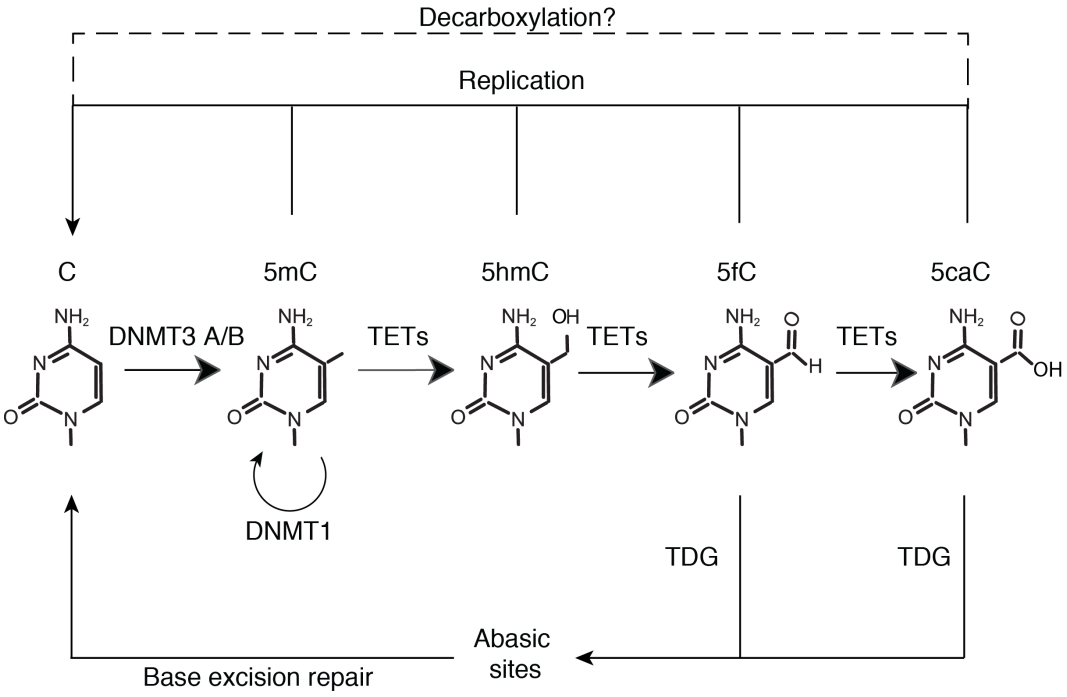

Figure 1.4 | DNA methylation and demethylation pathways. Cytosines within the context of CpG dinucleotides can become *de novo* methylated by the methyltransferases DNMT3A and DNMT3B. Cytosine methylation gets maintained through replication by the methyltransferase DNMT1 and its cofactor UHRF1. The ten eleven translocation (TET) enzymes can iteratively oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Oxidized 5mC bases can be passively lost during replication. Alternatively, 5fC and 5caC can also be actively removed by TDG and the base excision repair pathway. Direct decarboxylation of 5caC has been reported, but an enzyme has so far not been identified. (Adapted from Allis, 2015; content reproduced with permission of Cold Spring Harbor Laboratory Press)

1.5.2 DNA Methylation Machinery

Due to the palindromic nature of the CpG dinucleotide, a simple hypothesis for maintenance of the modification was postulated early on. Decades ago, Holliday, Pugh and Riggs suggested that parental methylated CpG dinucleotides could serve as a template for daughter strands during replication (Holliday and Pugh, 1975; Riggs, 1975). Similar to the canonical DNA bases, DNA methylation would thus be inherited semi-conservatively. Both groups suggested the existence of a maintenance methyltransferase. To this day the classical distinction between DNA methylation maintenance and *de novo* methylation is still widely referred to (Figure 1.4).

The maintenance DNA methyltransferase DNMT1 was discovered in 1983 by Bestor and Ingram (Bestor and Ingram, 1983). In biochemical assays DNMT1 preferably methylates hemi-methylated substrates (i.e. when only one of the two complementary CpGs is methylated). Genetic deletion of DNMT1 in embryonic stem cells causes a genome-wide loss of DNA methylation (Jackson et al., 2004; Li et al., 1992; Liao et al., 2015). DNMT1 is necessary for global DNA methylation, proved however not to be sufficient. To be targeted to hemi-methylated sites during replication, DNMT1 needs to collaborate with its accessory protein UHRF1 (also known as nuclear protein 95) (Achour et al., 2008; Bostick et al., 2007; Sharif et al., 2007). The UHRF1 protein recognizes hemi-methylated, double-stranded DNA by its SET and RING associated (SRA) domain. Crystal structures showed that SRA substrate recognition is realized through a flipping out mechanism of the parental methylated CpG (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008). In comparison to replication of generic DNA bases, maintenance of 5mC is more error prone. Depending on the species, DNMT1 maintenance fidelity ranges between 98.3% and 99.8% (Jackson et al., 2004; Liao et al., 2015).

Complementary to the maintenance system, *de novo* methyltransferases first establish methylation. Homology comparisons to prokaryotic methyltransferases, led to the subsequent discovery of DNMT3A and DNMT3B (Okano et al., 1998, 1999). The related, but catalytically inactive, protein DNMT3L was described to modulate the activity of DNMT3A during gametogenesis (Ooi et al., 2007). Recent studies in mouse embryonic stem cells have shown genome-wide *de*

novo binding of DNMT3A2 and DNMT3B1 to methylated CpG-dense regions and an exclusion from unmethylated sites (Baubec et al., 2015). Global DNA methylation patterns were largely concordant with DNMT binding behavior. Preferential DNMT3B binding to actively transcribed gene bodies was suggested (Baubec et al., 2015; Duymich et al., 2016). Recent crystal structures of DNMT3A and DNMT3L complexes have given mechanistic insights into its exclusion from active promoters. Both studies reported an inhibition of the ATRX-DNMT3-DNMT3L (ADD) domain by H3K4me3 modified histone tails (Guo et al., 2015; Ooi et al., 2007). The structural perspective on DNMT3 activity could explain the genome-wide exclusion of DNA methylation from H3K4me3 positive promoter regions.

Single genetic deletions of the three methyltransferases cause early embryonic (DNMT1, DNMT3B) or postnatal (DNMT3A) lethality in mice (Li et al., 1992; Okano et al., 1999). While loss of DNA methylation is lethal in all reported somatic cell lines, mouse embryonic stem cells are viable and pluripotent even in the absence of all three enzymes (Fan et al., 2001; Jackson et al., 2004; Li et al., 1992; Sen et al., 2010; Trowbridge et al., 2009; Tsumura et al., 2006). Single deletions of DNTM3A and DNMT3B in embryonic stem cells cause only minor changes in DNA methylation, due to possible redundancy (Liao et al., 2015; Okano et al., 1999). In contrast, deletions of both *de novo* enzymes lead to a more pronounced global loss of DNA methylation (Jackson et al., 2004; Liao et al., 2015; Okano et al., 1999). This is likely caused by the imperfect maintenance fidelity of DNMT1, maintenance functions of the DNMT3 enzymes or continuous demethylation by TET enzymes.

1.5.3 DNA Demethylation Pathways

Until recently, DNA methylation has been described as a heritable and stable chromatin modification. Long-term stable repression by DNA methylation is enabled through its semi-conservative maintenance during replication. With identification of an active demethylation machinery, this classical perception has undergone several refinements (Figure 1.4) (He et al., 2011; Ito et al., 2011; Kriaucionis and Heintz, 2009; Pfaffeneder et al., 2011; Tahiliani et al., 2009).

We distinguish passive and active demethylation pathways. Passive demethylation is achieved through successive rounds of replication-dependent dilution of 5-methylcytosine (5mC) in the presence of reduced DNMT1 or UHRF1 activity. In contrast, active demethylation is achieved through direct modification and removal of 5mC. Therefore, the rate of passive demethylation is dependent on the respective replication time. Active DNA demethylation can achieve fast demethylation uncoupled from replication.

First evidence for active demethylation processes was found in mammalian development. During pre-implantation development, maternal and paternal genomes are subjected to global DNA demethylation (Smith and Meissner, 2013). While the maternal genome shows replication-dependent dilution of 5mC, the paternal pronucleus shows fast demethylation within few hours post-fertilization (Mayer et al., 2000; Rougier et al., 1998). Due to the absence of replication, paternal demethylation must be active. In addition to this global process, several locus-specific cases of active demethylation have been reported. For instance, the interleukin-2 enhancer undergoes demethylation within minutes after stimulation of non-replicating T lymphocytes (Bruniquel and Schwartz, 2003). Similarly, postmitotic peripheral monocytes show site-specific active demethylation upon differentiation into dendritic cells (Klug et al., 2010, 2013). Further studies have reported cases of active demethylation at the *Bdnf* promoter in depolarized neurons, *Nanog* and *Oct4* promoters upon cellular fusion and at the pS2 promoter during estrogen activation (Bhutani et al., 2010; Kangaspeska et al., 2008; Martinowich et al., 2003; Métivier et al., 2008).

The above described phenomena have motivated the search for putative DNA demethylases. Transforming work by Kriaucionis, Heintz, Tahiliani et al. has led to the discovery of the active demethylation mark 5-hydoxymethylcytosine

(5hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). The corresponding ten eleven translocation (TET) enzymes were discovered and shown to catalyze the conversion from 5mC to 5hmC (Ito et al., 2010; Tahiliani et al., 2009). The TET enzymes were also reported to oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) *in vitro* and *in vivo* (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011).

Since then, all oxidized forms of 5mC have been quantified and mapped to single base-pair resolution genome-wide (Booth et al., 2013, 2014; Lu et al., 2013; Song et al., 2013; Yu et al., 2012). In general, prevalence of the base modifications decreases with increasing oxidation state. Levels of 5fC and 5caC are at least one order of magnitude lower than 5hmC (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011). The 5hmC modification can be detected across cell types (1% - 5% of 5mC) and is particularly abundant in adult neurons (15- 40% of 5mC) and embryonic stem cells (4% of 5mC) (Globisch et al., 2010; Ito et al., 2011; Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Genome-wide maps have indicated higher 5hmC occurrence within gene bodies, promoters and enhancers (Ficz et al., 2011; Pastor et al., 2011; Robertson et al., 2011, 2012; Song et al., 2011; Stroud et al., 2011; Williams et al., 2011b; Wu et al., 2011a; Xu et al., 2011). The loss of 5hmC within enhancers was shown to be accompanied by hypermethylation and reduced enhancer activity (Hon et al., 2014).

Comparable effort was directed towards the functional characterization of TETs. In mammals, the family of TET proteins consists of the three paralogues TET1, TET2 and TET3. While TET2 and TET3 show broad expression across cell types, TET1 is predominantly expressed in embryonic stem cells and primordial germ cells (Ito et al., 2010; Yamaguchi et al., 2012). All three enzymes require oxygen, iron and alpha-ketoglutarate as substrates and cofactors for 5mC oxidation. Oxidation is catalyzed through TET's C-terminal catalytic domain, that contains a cysteine-rich region and a double-stranded beta-helix (DSBH) (Wu and Zhang, 2014). TET1 and TET3 contain an additional CXXC domain, which could determine their genomic binding to CpGs. TET1 binding is indeed especially enriched over CpG-rich promoters (Williams et al., 2011b).

Tet deficient mice and cell types have been generated to define their function in development and DNA demethylation. Both *Tet1/Tet2* single and double knock out mice are viable, while *Tet3* deficiency leads to neonatal lethality (Dawlaty et

al., 2011, 2013; Gu et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011). Combined *Tet1* and *Tet2* deficiency causes reduced 5hmC levels and increased methylation of certain imprinted loci (Dawlaty et al., 2013). In mouse embryonic stem cells, loss of all three TET enzymes impairs differentiation potential and impacts telomere homeostasis (Dawlaty et al., 2014; Lu et al., 2014a). In line with a previous characterization of *Tet2* knock out mESCs, *Tet* triple knock out cells show preferential hypermethylation of distal enhancer regions (Hon et al., 2014; Lu et al., 2014a). Overall, *Tet* loss of function experiments have uncovered a rather mild methylation phenotype, where hypermethylation predominantly occurs within distal regions of the genome.

Several mechanisms of TET mediated demethylation have been proposed so far. First, it was recognized that DNMT1 activity is less efficient in maintaining hemi-hydroxymethylated cytosines (Hashimoto et al., 2012). Oxidation of 5mC can thereby cause a replication-dependent loss of methylation. Active demethylation can be achieved through consecutive iterative oxidation, DNA glycosylases and base excision repair (Wu and Zhang, 2010). The glycosylase TDG has been reported to efficiently excise 5fC and 5caC bases within DNA (He et al., 2011; Maiti and Drohat, 2011). Consequently, *Tdg* deficiency in mouse embryonic stem cells leads to a significant increase in 5fC and 5caC levels (He et al., 2011; Shen et al., 2013; Song et al., 2013). Repair of TDG induced abasic sites is thought to be completed by the base excision repair pathway (Wu and Zhang, 2010). While direct decarboxylation of 5caC could be conceived, the relevant enzyme has not yet been discovered (Schiesser et al., 2012; Wu and Zhang, 2010).

Altogether, the described discoveries have dramatically impacted our assumption of a stable propagation of DNA methylation. Many ambiguities regarding the extent of active DNA demethylation still need to be resolved. A more detailed molecular description of the DNA demethylation pathway and its players is urgently needed. Nevertheless, the possibility of active demethylation provides additional means to control methylation patterns in a locus-specific manner. Active DNA demethylation could enable increased methylation plasticity, especially within distal CpG-poor regions of the genome.

1.5.4 DNA Methylation Patterns

DNA methylation in mammals is present throughout the genome and was mapped for multiple mouse and human cell types (Gifford et al., 2013; Habibi et al., 2013; Lister et al., 2009; Meissner et al., 2008; Smith et al., 2012; Stadler et al., 2011). These efforts have allowed a refined and more dynamic view on global DNA methylation.

Before describing these methylation patterns in more detail, it is crucial to comment on its substrate distribution. As introduced in the earlier sections, mammalian cytosine methylation mostly occurs within CpG dinucleotides. Due to the increased mutagenicity of methylated cytosines, mammalian genomes have undergone an evolutionary depletion of CpG dinucleotides (Long et al., 2013a). Reduced CpG depletion can be found in so-called CpG islands (CGIs), CpG-dense regions that are unmethylated in the germ line (Deaton and Bird, 2011). CpG islands are frequently located within promoter regions and constitute 60% of all annotated promoters (Larsen et al., 1992; Mohn and Schübeler, 2009).

Cytosine methylation is distributed bimodally in the mouse methylome. While the majority of cytosines show high methylation levels (between 80 - 100%), few regions are almost completely unmethylated. Another fraction of cytosines possesses intermediate methylation levels (10 – 50%). By applying a Hidden Markov Model (HMM), the mouse embryonic stem cell methylome can be segmented into three different clustered methylation states: fully methylated regions (FMRs), low-methylated regions (LMRs) and unmethylated regions (UMRs) (Figure 1.5) (Burger et al., 2013; Stadler et al., 2011). All three methylation categories are functionally distinct from one another, while UMRs and LMRs share high DNA-binding factor occupancy as a common characteristic. Fully methylated cytosines are found throughout the genome within repetitive elements, intergenic and intragenic regions. Completely unmethylated sites overlap with the previously described element of CpG islands. Regions flanking CpG islands have been termed shores (at a distance of 2kb from CGIs) and shelves (2 – 4kb from CGIs) and were assigned distinct methylation characteristics (Irizarry et al., 2009). This classical model has been expanded by the description of LMRs as a separate methylome feature (Stadler et al., 2011).

These regions describe local hypomethylation outside of CpG-dense promoter regions and often demarcate distal regulatory elements.

Whole-genome methylation profiling in somatic cells has also uncovered a more disordered state of some regions (Hon et al., 2012; Lister et al., 2009, 2011; Schroeder et al., 2011). While the classical paradigm above describes local clustering of methylation states, so-called partially methylated domains (PMDs) are lacking this feature. First discovered in human fibroblasts, partially methylated domains harbor average methylation levels lower than 70% and expand over several hundred kilo bases in a disorganized fashion (Lister et al., 2009). Comparisons of published methylomes revealed conservation across different cell types (Gaidatzis et al., 2014). Further indications of a potential functional role of these unique domains are still missing.

While different cell types share common patterns, methylation can undergo large changes during development, somatic differentiation and oncogenic transformation (De Carvalho et al., 2012; Smith and Meissner, 2013; Sproul et al., 2012; Ziller et al., 2013a). Hypomethylation within LMRs is particularly dynamic and often cell-type specific (Burger et al., 2013; Stadler et al., 2011). Methylation data could even be used to determine cell types in mixed populations (Accomando et al., 2014).

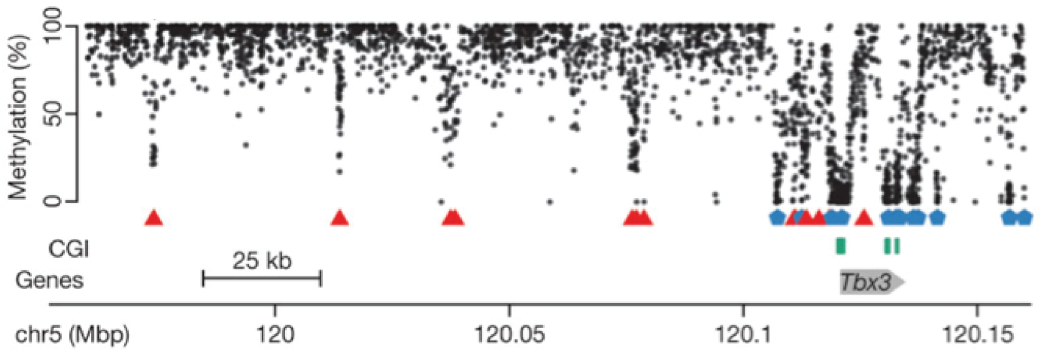

Figure 1.5 | A genomic view at DNA methylation in mouse embryonic stem cells. There is typical clustering of methylation states in mouse embryonic stem cells surrounding the TBX3 gene. The majority of CpGs (depicted as dots) are fully methylated. Local hypomethylation can be found in low-methylated regions (LMRs, red triangles) and unmethylated regions (UMRs, blue pentagons). Some UMRs overlap with CpG islands (CGIs, green bar). (Adapted from Stadler et al., 2011; content reproduced with permission of Nature Publishing Group)

1.5.5 DNA Methylation and its Effect on Transcription

DNA methylation is associated with stable transcriptional silencing of promoters. Experimental evidence for this transcriptional effect was first observed in transfections of retroviral DNA into mammalian cells and *Xenopus* oocytes (Stein et al., 1982; Vardimon et al., 1982).

Deletion of DNMT1 in embryonic stem cells leads to a transcriptional upregulation of repetitive elements (Li et al., 1992). The transcriptional response at endogenous retroviral (ERVs) elements is strongest upon acute loss of DNMT1. A recent study suggested that this derepression is an indirect consequence of UHRF1 mediated inhibition of the H3K9 methyltransferase SETDB1 (Sharif et al., 2016). Since the depletion of DNA methylation is lethal in all somatic cells, not much is known about further transcriptional responses in differentiated cells.

In addition to its role in constitutive silencing of repetitive elements, DNA methylation plays an important role in enabling cell-type specific expression. Only few CpG island promoters are however differentially methylated in embryonic stem cells and terminal neurons (Mohn et al., 2008). During *in vitro* differentiation to neurons, some pluripotency-associated promoters undergo hypermethylation, while neuron-specific genes become demethylated. Methylation of CpG islands proved incompatible with transcription (Mohn et al., 2008). Since then it was progressively uncovered that the majority of methylation changes occur outside of promoter regions (Stadler et al., 2011). Due to the eminent problem of incomplete promoter-enhancer assignment, information on a potential function of distal hypomethylation is still missing.

Models on how the repressive effect of DNA methylation is mediated focus on two scenarios: 1. Attraction of repressive methylation binding proteins and 2. Repulsion of methylation-sensitive activators like transcription factors. While the next section will give an in-depth description of methylation-sensitive factors, the following knowledge has been gained about methyl-binding proteins.

The family of methyl-CpG-binding domain (MBD) proteins comprises the members MECP2, MBD1, MBD2, MBD3 and MBD4 (Bird and Wolffe, 1999). All MBD proteins with the exception of MBD3 preferentially bind methylated genomic sites (Baubec et al., 2013). Binding of these proteins generally increases with methylation density, but methylation-independent targeting can be observed at

inactive enhancers, intermediate CpG content promoters and gene bodies (Baubec et al., 2013)., Non-MBD proteins like the zinc finger protein KAISO can recognize methylated CpGs. How exactly methyl-binding proteins cause effective gene repression of methylated promoters is still under debate. In principle, transcriptional repression by MBDs could be mediated through steric interference with activator binding or through repressor recruitment. Evidence for the latter was first found for MECP2, which was reported to interact with the SIN3A corepressor complex (Jones et al., 1998; Nan et al., 1998). As introduced earlier, SIN3A in turn mediates transcriptional repression through HDAC recruitment (Silverstein and Ekwall, 2005). Similar corepressor recruitment has been suggested for the other MBD proteins (Nan et al., 1998; Ng et al., 1999). Genome-wide binding profiling did not unambiguously confirm SIN3A and NuRD corepressor localization to methylated CpG-dense regions (Baubec et al., 2013; Williams et al., 2011b). These observations challenge the general concept of MBD mediated HDAC recruitment. Individual deletions of MBD proteins have neither shown substantial transcriptional reactivation (Hendrich et al., 2001; Tudor et al., 2002; Zhao et al., 2003). Whether this reflects broad redundancies in MBD activity or a lack of transcriptional activity still remains uncertain. The strongest MBD developmental phenotype so far is described for MeCP2 mutations, which cause the neurological disorder Rett syndrome (Amir et al., 1999). However, *Mecp2* deficient mice only showed subtle transcriptional responses in the brain (Tudor et al., 2002).

In sum, while the description of methylation patterns has made huge advancements, much is still unknown regarding the mechanism of methylation mediated transcriptional repression. Some of the most intriguing questions address the general role of hypomethylation in distal regulatory regions and the function of MBD proteins.

1.5.6 DNA Methylation as a Cause and Consequence of Transcription Factor Binding

Functional genomic elements constitute the space of transcription factor binding and are frequently hypomethylated. Traditionally, research has focused on the transcriptional effects of promoter-associated methylation. DNA methylation was considered a stable chromatin modification that exerts transcriptional repression through repulsion of activators. This paradigm has recently undergone several refinements. In fact, the methylation sensitivity of transcription factors is highly factor-specific. The following paragraphs will describe a more complex relationship between transcription factors and DNA methylation.

Methylation-sensitive Transcription Factors

Structural insights into DNA-protein interactions have suggested that the hydrophobic methyl group of cytosines can directly interfere with base pair recognition by transcription factors (Dantas Machado et al., 2015; Tate and Bird, 1993). In addition, methylation of cytosines alters the shape of DNA and could affect reading by unspecific DNA-binding domains (Dantas Machado et al., 2015; Lazarovici et al., 2013; Rohs et al., 2009). Furthermore, indirect inhibition can be mediated through attraction of methyl-binding proteins like the MBD proteins (Boyes and Bird, 1991; Nan et al., 2007). Which of these instances affect a particular protein-DNA interaction is in most cases not known. However, one can estimate the direct effect by examining known transcription factor motifs. Highly conserved CpG dinucleotides within consensus motifs are more likely to interfere with DNA sequence recognition. In this line, the number of potentially affected transcription factors is high, as about 25 % of the motifs deposited at the JASPAR database contain a conserved CpG dinucleotide (Blattler and Farnham, 2013).

First experimental evidence for interference of cytosine methylation with transcription factor binding has been presented decades ago (Watt and Molloy, 1988). Since then many more methylation-sensitive factors have been described (Campanero et al., 2000; Domcke et al., 2015; Iguchi-Ariga and Schaffner, 1989; Prendergast and Ziff, 1991). One prominent example for methylation-dependent binding can be seen with the parent-of-origin specific activity of CTCF at the

imprinting locus *H19/Igf2* (Bell and Felsenfeld, 2000). Genome-wide binding profiling has uncovered a highly context-dependent behavior, where only a small subset of CTCF sites proves as methylation-sensitive (Maurano et al., 2015; Stadler et al., 2011). Several other high-throughput efforts have been undertaken to denominate the global set of methylation-sensitive transcription factors. One approach has screened for differential binding events in methylation-deficient mouse embryonic stem cells (Domcke et al., 2015). While most protein binding events were not affected by the absence of DNA methylation, the transcription factor NRF1 displayed strong methylation sensitivity. The binding of NRF1 was impacted by the methylation levels of its motif-associated CpG dinucleotides (Figure 1.6A) (Domcke et al., 2015). Complementary, two *in vitro* studies have investigated methylation-dependent protein binding by methylated/unmethylated DNA pull-down coupled mass-spectrometry and a protein microarray, respectively (Hu et al., 2013; Spruijt et al., 2013). Both studies did not only reveal methylation-sensitive binding factors but a large number of previously unknown methyl-binding proteins. The pluripotency factor KLF4 showed context-specific methylation sensitivity in both studies. Overall, these investigations highlight the notion, that DNA methylation represents a highly context-dependent positive or negative protein-binding signal.

Demethylating Transcription Factors

Methylation-sensitivity indirectly implicates DNA methylation upstream of transcription factor binding, yet several observations indicate that hypomethylation of functional elements can be a consequence of protein binding.
In proximal promoters, where DNA methylation is particularly low at CpG islands, transcription factors seem to play a general role in maintenance of the hypomethylated state (Deaton and Bird, 2011; Straussman et al., 2009). Notably, the CXXC-domain containing protein CFP1 specifically binds to unmethylated CpG islands and recruits the chromatin modification H3K4me3 (Thomson et al., 2010). The presence of H3K4me3 on nucleosomes has been reported to inhibit DNMT3A activity *in vitro* and could indirectly protect islands from DNA methylation (Zhang et al., 2010). Deletion of the transcription factor motif associated with SP1 binding, leads to *de novo* methylation of the *Aprt* promoter (Brandeis et al., 1994; Macleod et al., 1994). More systematic studies revealed

the sufficiency of small genetic elements to induce endogenous promoter methylation levels (Krebs et al., 2014; Lienert et al., 2011a). CpG density and transcription factor motifs are critical components of those methylation-determining regions (MDRs) (Lienert et al., 2011a).

Outside of CpG islands, reduced methylation levels can be found within distal CpG-poor regions (Hodges et al., 2011; Stadler et al., 2011). With high-resolution profiling of global DNA methylation in mouse embryonic stem cells and somatic cells, it was progressively realized that hypomethylation is a general feature of protein binding events (Stadler et al., 2011). During differentiation, cell-type specific hypomethylated regions are often occupied by cell-type-specific DNA-binding factors (Hodges et al., 2011; Stadler et al., 2011; Ziller et al., 2013b). The two factors CTCF and REST were indeed experimentally shown to be necessary (REST) and sufficient (CTCF, REST) for reduced methylation at their binding sites (Stadler et al., 2011). REST induced methylation changes were shown to affect binding of downstream methylation-sensitive factors (Figure 1.6B) (Domcke et al., 2015; Stadler et al., 2011). In detail, many distal REST binding sites overlap lowly methylated regions in embryonic stem cells. In the absence of REST expression, cells possess high methylation levels at previously REST bound low-methylated cytosines. Re-introduction of the protein into *Rest* knock out cells re-establishes the hypomethylation of *Rest* wild-type cells. These single-loci experiments strongly indicate REST's ability to bind to methylated sites and reduce methylation levels (Figure 1.6A, Figure 1.6B). In a later study, REST mediated hypomethylation was shown to affect downstream binding of the methylation-sensitive factor NRF1 at selected sites (Figure 1.6B) (Domcke et al., 2015). In a scenario where REST and NRF1 binding sites are adjacent to each other, REST induced methylation changes caused differential binding of downstream methylation-sensitive NRF1 (Domcke et al., 2015). Even though such REST and NRF1 co-occupied sites are rare in the genome, it illustrates a potential hierarchy of methylation-insensitive pioneering factors and downstream methylation-sensitive effector proteins.

While these discussed experiments challenge the concept of stable DNA methylation, little is known about the mechanism of transcription factor-induced methylation changes. Several studies have reported the enrichment of 5hmC at active enhancers and suggested a function in its hypomethylated state

(Feldmann et al., 2013; Hon et al., 2014; Sérandour et al., 2012; Stadler et al., 2011). TET2 recruitment by the transcription factors WT1 and EBF1 may be linked to reduced methylation over binding sites (Guilhamon et al., 2013; Rampal et al., 2014; Wang et al., 2015). By other approaches, EBF1 has been attributed pioneering activity (Boller et al., 2016). Beyond that, more detailed descriptions of the molecular pathways that lead to transcription-factor mediated hypomethylation are missing. It is thereby the focus of this doctoral thesis to describe the molecular components that are necessary for local hypomethylation. Particular focus will be placed on the transcription factor REST.

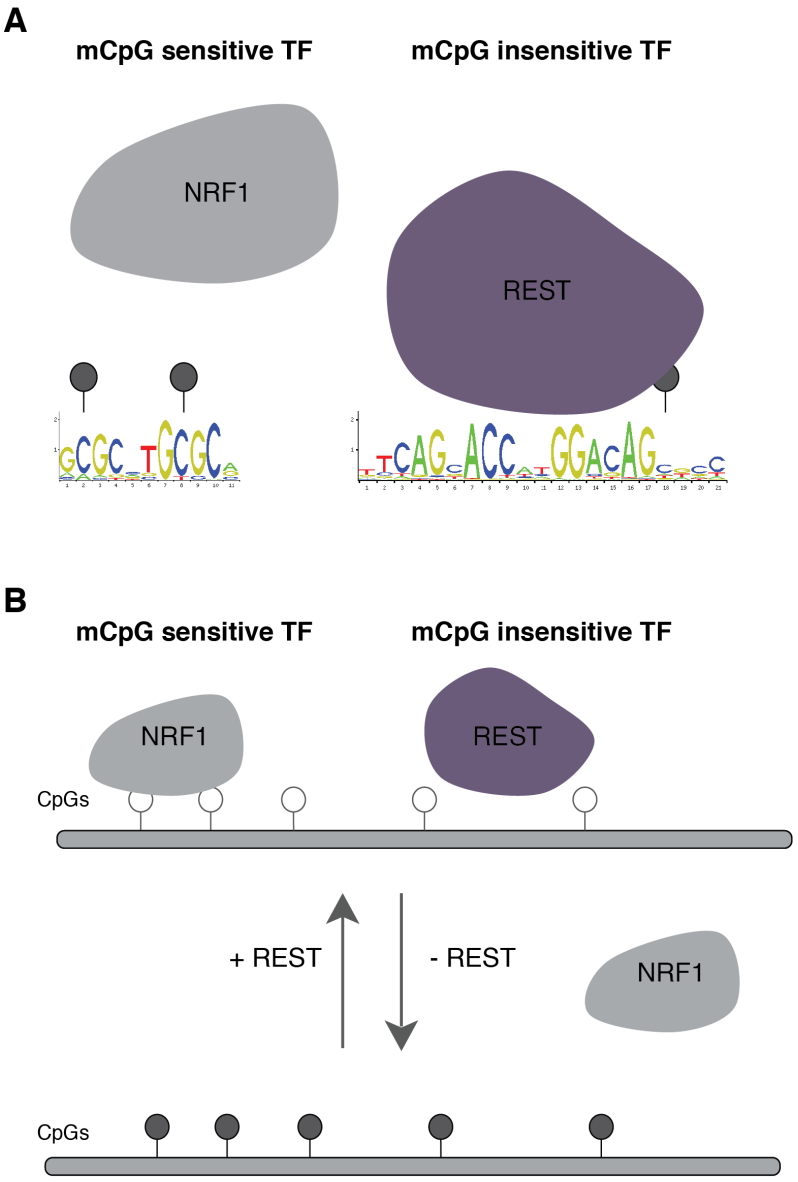

Figure 1.6 | The interplay between transcription factor binding and DNA methylation. A Transcription factors can be sensitive or insensitive to DNA methylation levels of their binding site. In the case of NRF1, methylation of the two highly conserved motif CpGs seems to interfere with binding. The transcription factor REST is suggested to be insensitive to DNA methylation. Note the differences in CpG conservation between the two consensus motifs. **B** REST binding sites are lowly methylated in wild type cells and NRF1 can bind to its adjacent motif. In absence of REST in Rest$^{-/-}$ cells, cytosines around the REST motif become methylated which precludes binding of the methylation sensitive factor NRF1. This single locus illustrates, how hierarchically superior transcription factors enable subsequent methylation sensitive protein binding. Legend: circles illustrate single cytosines within a CpG dinucleotide; empty circles show low methylation, filled circles high methylation; motif logos for NRF1 and REST were taken from the JASPAR database (Portales-Casamar et al., 2010). (Abstracted from Domcke et al., 2015; Stadler et al., 2011; content reproduced with permission of Nature Publishing Group)

Chapter 2
Scope of the Thesis

DNA methylation constitutes a repressive chromatin modification that controls essential processes like retrotransposon silencing, X chromosome inactivation, genomic imprinting and maintenance of cellular identity (Bird, 2002; Jaenisch and Bird, 2003). The mammalian genome is broadly methylated within the context of CpG dinucleotides. Notable exceptions can be found in proximal and distal functional regions, where methylation is characteristically reduced. While unmethylated CpG-rich promoter regions were described a long time ago, hypomethylation within distal CpG-poor regions has only been revealed recently (Deaton and Bird, 2011; Stadler et al., 2011). These low-methylated regions (LMRs) often demarcate active distal regulatory elements such as enhancers. Transcription factor binding was suggested to be a main regulator of local DNA methylation in those regions (Stadler et al., 2011). While enormous progress has been made in mapping DNA methylation in different cell types, our understanding of how DNA-binding factors contribute to hypomethylation is still limited.

I therefore investigated the interaction between transcription factor binding and local methylation in further detail. I used the RE1-silencing transcription factor (REST) as a paradigm to study DNA methylation as a consequence of protein binding. In detail, I asked whether DNA hypomethylation is a consequence of DNA-binding alone. I assayed several REST mutants for their ability to reduce local methylation, alter chromatin accessibility and nucleosome positioning. I integrated genome-wide binding information on REST's interactors to describe the chromatin landscape of REST binding sites. Last, I tested the contribution of an active demethylation and characterized the kinetics of REST induced methylation changes.

Chapter 3

Results

3.1 Targeted Methylation Profiling by Amplicon Bisulfite Sequencing

DNA methylation is subject to large changes during somatic differentiation and contributes to oncogenic transformation (De Carvalho et al., 2012; Sproul et al., 2012; Ziller et al., 2013a). The clinical potential of DNA methylation as a marker for disease diagnosis and progression is high (Schübeler, 2015). Most cytosines in the mammalian genome are however invariably methylated and will not be informative in diagnostics (Ziller et al., 2013b). Several other research questions require precise methylation information for regions of interest. This explains why there is an overall high demand for reliable targeted methylation profiling assays.

In general, bisulfite based assays are considered the gold standard for methylation profiling. Bisulfite treatment of genomic DNA selectively converts unmethylated cytosines to thymines. However, all bisulfite approaches share the disadvantage of miscalling oxidized forms of 5mC as either methylated or unmethylated. Only single-molecule, real-time sequencing (SMRT-seq, *Pacific Biosciences*) so far allows direct detection of DNA base modifications (Flusberg et al., 2010). Several targeted assays based on bisulfite conversion have been developed (Lee et al., 2013). The most widely used approaches are based on arrays (e.g. *Infinium HumanMethylation450 BeadChip, Illumina*), hybridization (e.g. *SureSelect Methyl-seq, Agilent*) or multidroplet Bis-PCRs (*RainDance Technologies*). All of these commercially available techniques are either non-customizable or expensive in their acquisition. The following sections will therefore describe amplicon bisulfite sequencing (amplicon Bis-seq) as a suitable alternative. I will discuss experimental design, implementation and data analysis.

3.1.1 Experimental Implementation of Amplicon Bisulfite Sequencing

We applied amplicon Bis-seq as our method of choice for targeted methylation profiling. Its advantages are time- and cost-efficiency and easy experimental access. Furthermore, amplicon Bis-seq offers the flexibility to be integrated into other experimental assays, such as Nucleosome Occupancy and Methylome Sequencing (NOMe-seq) (Kelly et al., 2012; Pardo et al., 2009, 2011).

In brief, bisulfite-compatible primers are designed for regions of interest in an automated way (Figure 3.1) (Chapter 3.1.2). Genomic DNA is subjected to bisulfite conversion and polymerase chain reaction (PCR). Sequencing libraries are prepared from pooled amplicons and are multiplexed on the Illumina MiSeq platform. Downstream data alignment, quality control and analysis are performed within the R environment (Chapter 3.1.2) (Gaidatzis et al., 2015; Langmead et al., 2009).

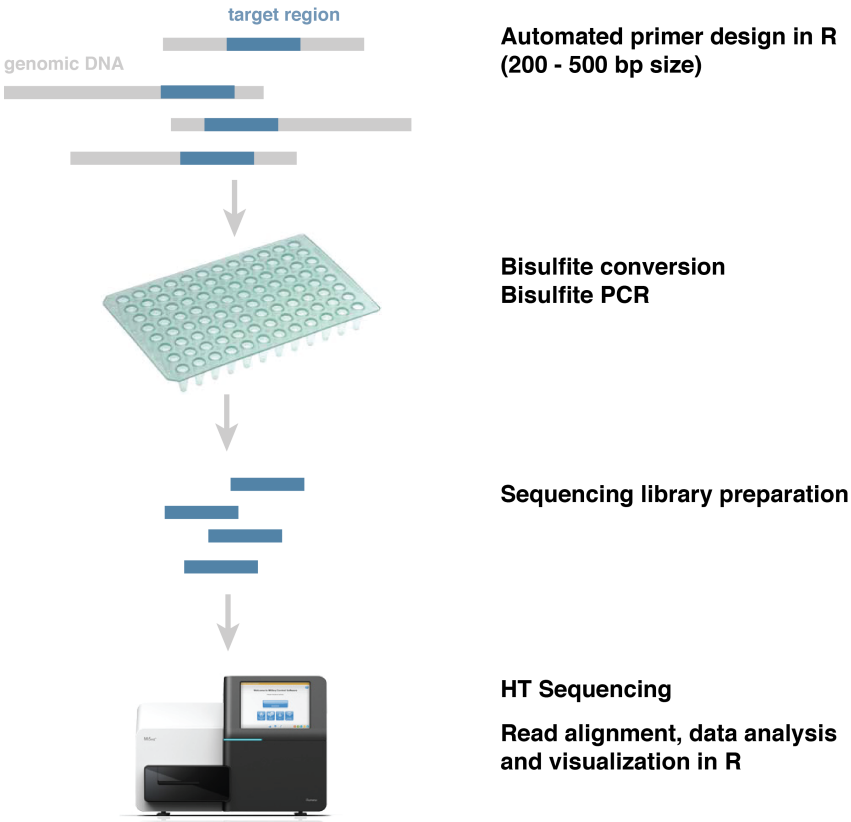

Figure 3.1 | Experimental outline for amplicon Bis-seq experiments. Bisulfite primers are designed for regions of interest and genomic DNA is subjected to bisulfite conversion followed by PCR. DNA libraries are sequenced and next-generation sequencing data is processed and analyzed within the R environment.

3.1.2 An R package for Design and Analysis of Amplicon Bisulfite Sequencing Data

Due to the lack of open-source programs for amplicon Bis-seq assays, we developed the R package AmpliconBiSeq (Akalin, A., AmpliconBiSeq, GitHub repository, https://github.com/BIMSBbioinfo/AmpliconBiSeq). The package is intended for the design, analysis and visualization of amplicon Bis-seq experiments. The AmpliconBiSeq package integrates the functionality of several other R packages, most notably the Bioconductor package QuasR (Quantify and annotate short reads in R) (Gaidatzis et al., 2015; Langmead et al., 2009). The following paragraphs will briefly highlight the core functions of the AmpliconBiSeq package (Figure 3.2).

The critical bottleneck of every amplicon Bis-seq experiment is the primer design. Bisulfite primers need to avoid CpGs in the priming region and anneal to several non-CpG cytosines. The AmpliconBiSeq package enables multiplexed and automated bisulfite primer design for regions of interest. The package hereby interfaces the Primer3 software to design its primers (Koressaar and Remm, 2007). Several parameters like amplicon size and primer melting temperature can be user defined. Primers can be filtered for a minimum of amplicon covered CpGs.

Once high-throughput sequencing data has been generated, raw sequencing reads are aligned within QuasR against the genome of interest. Basic raw data quality analysis can be performed within the package. If *in vitro* methylated T7 DNA and/or unmethylated lambda DNA has been spiked in during the experiment, conversion quality checks can be done within AmpliconBiSeq. Entry point for further methylation analyses in AmpliconBiSeq is the aligned QuasR object, from which an AmpliconViews object is generated. The latter is the central data container within AmpliconBiSeq and contains information on CpG coverage, average CpG methylation, meta-methylation profiles and similarity between adjacent CpGs. To generate the AmpliconViews object, reads can be filtered for defined non-CpG conversion efficiency. From there, average methylation information per CpG and region can be conveniently extracted. Meta-methylation profiles can be plotted to gain information on methylation heterogeneity at a single locus. Together, the QuasR and AmpliconBiSeq packages cover the whole

workflow of classical amplicon Bis-seq, starting with the experimental design to final data visualization.

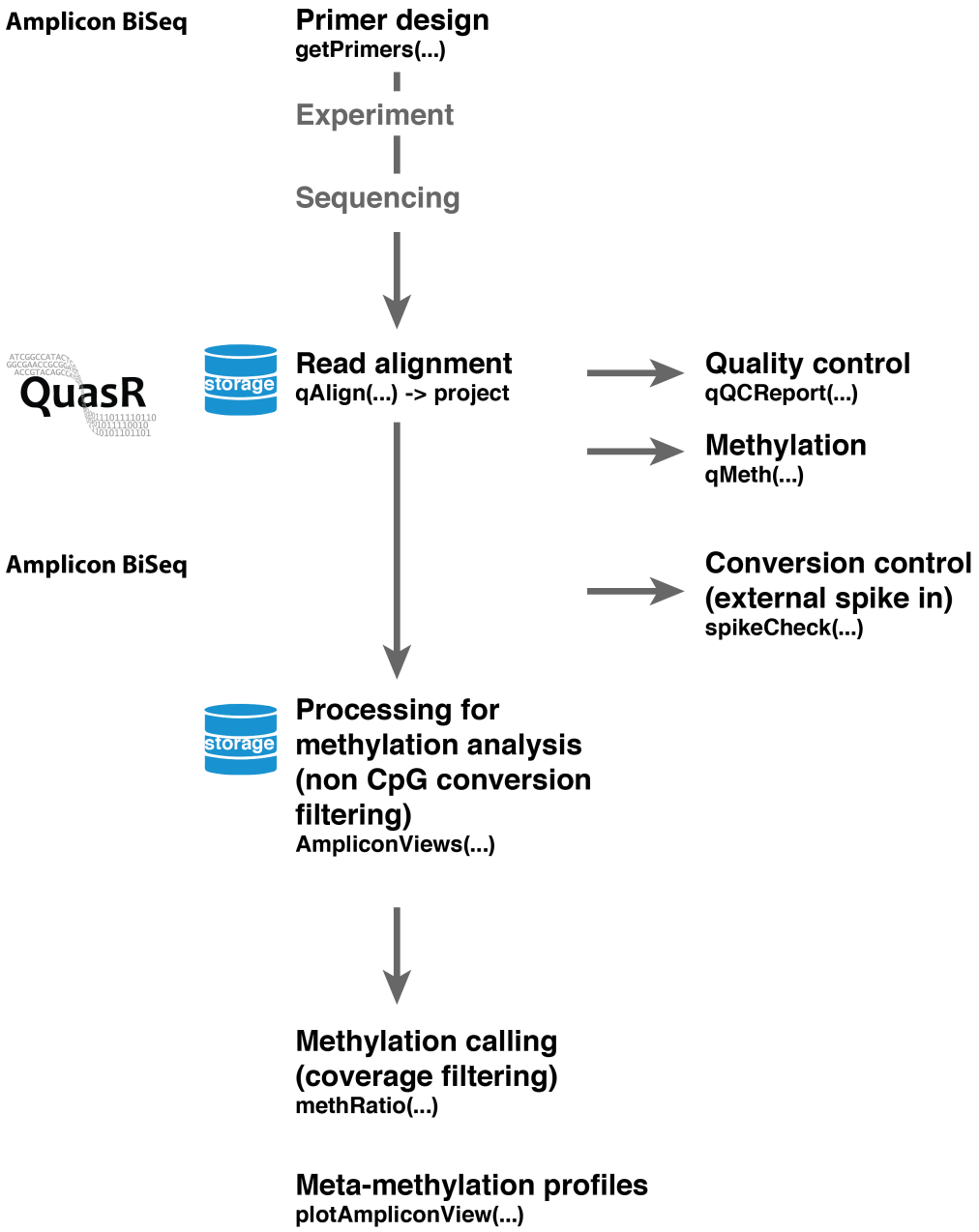

Figure 3.2 | Design and analysis outline for amplicon Bis-seq experiments. Bisulfite primers for subsequent PCRs are designed within the AmpliconBiSeq package. High-throughput sequencing reads are aligned and general quality controls are performed within the Bioconductor package QuasR. Conversion controls of external DNA are handled by AmpliconBiSeq. Further methylation analyses are performed within AmpliconBiSeq to receive information on coverage, methylation averages, meta-methylation profiles and methylation similarity of CpGs.

3.1.3 Amplicon Bisulfite Sequencing for High-coverage Methylation Profiling

To test the validity of our targeted methylation approach, we performed amplicon Bis-seq on previously characterized mouse embryonic stem cells (mESCs) (Stadler et al., 2011). We designed bisulfite primers for 35 distal REST binding sites and 61 control regions, including six regions within the lambda and T7 genome. Amplicons were designed within a size range of 200 to 400 bp and were centered on REST motifs in the case of REST amplicons. Two mouse embryonic stem cell samples were subjected to paired-end sequencing on the Illumina MiSeq platform. Data alignment and analysis were performed within QuasR and AmpliconBiSeq (Akalin, A., AmpliconBiSeq, GitHub repository, https://github.com/BIMSBbioinfo/Amplicon BiSeq; Langmead et al., 2009).

We achieved a high recovery of designed amplicons (80%) in each sample (Table 3.1). Considering both experiments, we were able to detect 88 out of 96 designed amplicons (92%). Coverage was high, exceeding common requirements regarding methylation precision (coverage per CpG > 20). Conversion efficiencies were calculated from external unmethylated lambda DNA and were above 97.7 % in both experiments. Significant overconversion of pre-methylated T7 DNA was not detectable in either sample. Reproducibility was high between technical replicates (r=0.96) and was in line with the previously published whole-genome Bis-seq methylome (r=0.90) (Figure 3.3A, Figure 3.3B) (Stadler et al., 2011). Reproducibility to whole-genome Bis-seq indicates that potential redundant molecule counting in Amplicon Bis-seq does not cause major systematic biases.

We used the AmpliconBiSeq functionalities to calculate meta-methylation profiles and similarities between neighboring CpGs of individual amplicons. I show one representative example of a distal REST binding site with characteristically low methylation levels around the REST motif center (chr4: 153.2819 Mbp) (Figure 3.3C). In this region, DNA methylation is heterogeneous within alleles with four methylation profiles contributing to 92 % of the observable patterns within the population. The calculated Jaccard similarity highlights similar methylation states of individual CpG pairs (e.g. CpGs 5 and 6) or opposite ones (e.g. CpGs 5 and 2). These methylation similarities can become particularly informative to study

local methylation interdependencies. Such phenomena might be observable between nucleosomal linker regions or co-occupied binding sites.

I conclude that amplicon Bis-seq represents a flexible, amenable approach for high-coverage methylation profiling for regions of interest. The presented R package AmpliconBiSeq enables previously not available, open-source design and analysis of such experiments. The experimental framework can be easily integrated into standard laboratories.

	mESCs, replicate 1	**mESCs, replicate 2**
Sequencing type	150bp PE, 1plex MiSeq	150bp PE, 4plex MiSeq
Reads, total	20'338'372	5'065'992
Reads, mapped (mm9)	12'256'010	2'761'522
Number of recovered amplicons	71/90	71/90
Coverage/amplicon, mean	81'024	15'481
Coverage/CpG, mean	150'180	19'581
mCpG (%), methylated T7 control	99.47	99.57
mCpG (%), unmethylated λ control	2.28	0.61

Table 3.1 | Performance metrics of two amplicon bisulfite sequencing experiments. Two samples of mESCs were subjected to amplicon Bis-seq and characterized. Designed bisulfite primers resulted in a high experimental recovery of amplicons. Coverage was saturating experimental needs with paired-end sequencing on the Illumina MiSeq. External spiked in DNA of lambda and pre-methylated T7 DNA allows quantification of conversion quality.

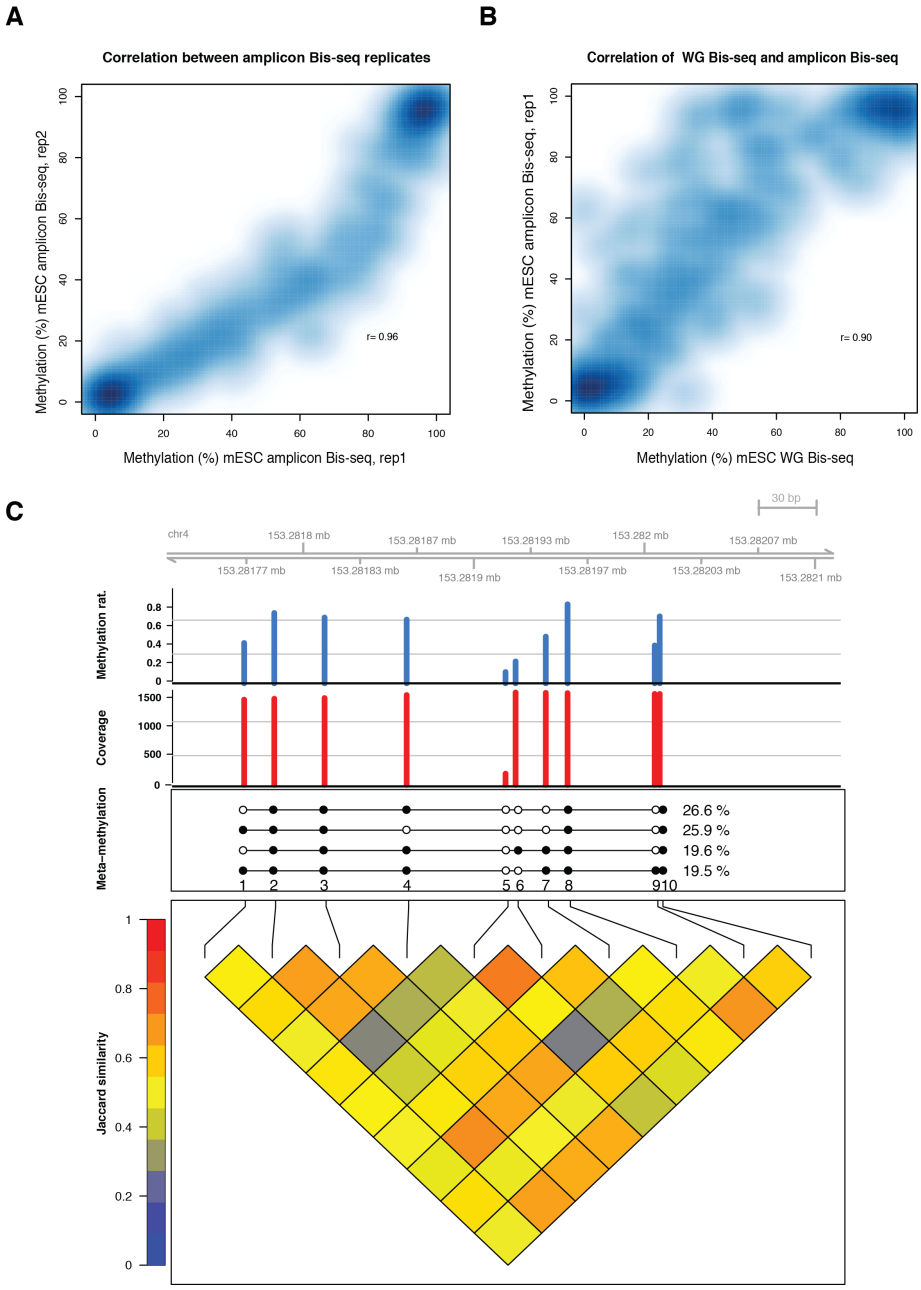

Figure 3.3 | Amplicon bisulfite sequencing yields high-coverage methylation information for regions of interest. A The two-dimensional density plot shows the correlation of single CpG methylation within two biological replicates of mESCs. Overall, reproducibility is high (r= 0.96) with higher variation within intermediately methylated cytosines. **B** Amplicon Bis-seq was benchmarked to whole-genome Bis-seq of previously described mESCs. The density plot shows a high correlation of CpG methylation in both methods (r= 0.90). **C** A single distal REST binding site was analyzed in more detail using the AmpliconBiSeq package. Methylation (upper track, blue) is low around the REST motif (not shown) and coverage consistently high for the cytosines within the designed amplicon (middle track, red). Meta-methylation profiles indicate methylation heterogeneity within alleles (middle track, open circles indicate unmethylated CpGs, filled circles methylated CpGs). The calculated Jaccard similarities quantify methylation interdependencies between CpG pairs (lower track).

3.2 Regulation of DNA Methylation by the RE1-silencing Transcription Factor

Whole-genome bisulfite sequencing has generated precise methylation maps for various cell types (Gifford et al., 2013; Habibi et al., 2013; Lister et al., 2009; Meissner et al., 2008; Smith et al., 2012; Stadler et al., 2011). Frequent methylation patterns have been characterized in detail, distinguishing different hypomethylated regions from the methylated majority of the genome (Chapter 1.5.4). While the description of methylation states has reached an unprecedented scale, knowledge concerning its establishment is still limited. Individual reports have proposed transcription factors to be drivers of local hypomethylation (Hodges et al., 2011; Stadler et al., 2011; Ziller et al., 2013b). Yet, little is known about the mechanism of transcription factor induced methylation changes.

Motivated by this apparent discrepancy, I aim to investigate the mechanism of transcription factor induced hypomethylation within low-methylated regions (LMRs). I concentrated on the transcriptional repressor REST as previous reports indicated REST binding as a cause for local hypomethylation (Chapter 1.5.6) (Stadler et al., 2011). The following sections will describe REST mediated changes in transcription, DNA methylation and other chromatin modifications. We particularly focus on the question, which molecular components of REST are required to induce hypomethylation.

3.2.1 REST is not Required for Pluripotency in Mouse Embryonic Stem Cells

We used previously characterized *Rest* wild type (*Rest wt*) and *Rest* knock out (*Rest ko*) cells to study REST induced chromatin changes (Arnold et al., 2013; Jørgensen et al., 2009b; Stadler et al., 2011). We first investigated REST's transcriptional role in mouse embryonic stem cells and the state of pluripotency in the absence of REST.

Hence, we analyzed previously generated data of total RNA sequencing (RNA-seq) for *Rest wt* and *Rest ko* cells (data unpublished, generated by Michael Stadler). Globally, we detect very similar gene expression in both cell lines (r= 0.992) (Figure 3.4A). We defined potential REST targets as genes, which

possess at least one REST ChIP-seq peak within 10kb of annotated transcriptional start sites (TSSs). In *Rest ko* cells, we observe preferential transcriptional upregulation of REST target genes. Highly expressed REST target and non-target genes are generally less affected by REST deficiency. Though, REST target genes contribute to a large extent to the most strongly upregulated genes. These observations are concordant with REST's function as a known transcriptional repressor (Chen et al., 1998; Chong et al., 1995; Schoenherr and Anderson, 1995).

Next, we addressed a potential role of REST in maintenance of pluripotency. We contrasted gene expression of *Rest wt* and *Rest ko* cells with previously published mouse embryonic stem cells (ESCs) and their *in vitro* differentiated neuronal progenitors (NPs) and terminal neurons (TNs) (Stadler et al., 2011). Here, *Rest wt* and *Rest ko* cells co-cluster with the previous mESC data set (Figure 3.4B). Both NP and TN samples are less correlated, which argues for general cell type similarity among the three stem cell lines.

We selectively compared the expression of the known pluripotency factors OCT4, SOX2, NANOG and KLF4. Again, all three stem cell lines are similar and show high expression of all four transcription factors independent of their genotype (Figure 3.4C). We observe no detectable expression differences in pluripotency factors between *Rest ko* and *Rest wt* cells. In contrast, strong transcriptional downregulation of OCT4, NANOG and KLF4 is apparent in NPs and TNs. Considerable Sox2 expression is still detectable in both NPs and TNs, which is compatible with a reported role of Sox2 during neurogenesis (Amador-Arjona et al., 2015; Favaro et al., 2009).

In summary, *Rest wt* and *Rest ko* cells posses similar global gene expression. REST deficient cells still express high levels of pluripotency markers. Our analysis does not support a role of REST in pluripotency and thereby joins the general consensus in this debate (Jørgensen and Fisher, 2010; Jørgensen et al., 2009a). Furthermore, the data justify the use of *Rest ko* cells to study DNA methylation as a consequence of REST binding. Due to the clear stem cell properties of *Rest ko* cells, I do not anticipate large indirect, cell type associated methylation changes outside of REST binding sites.

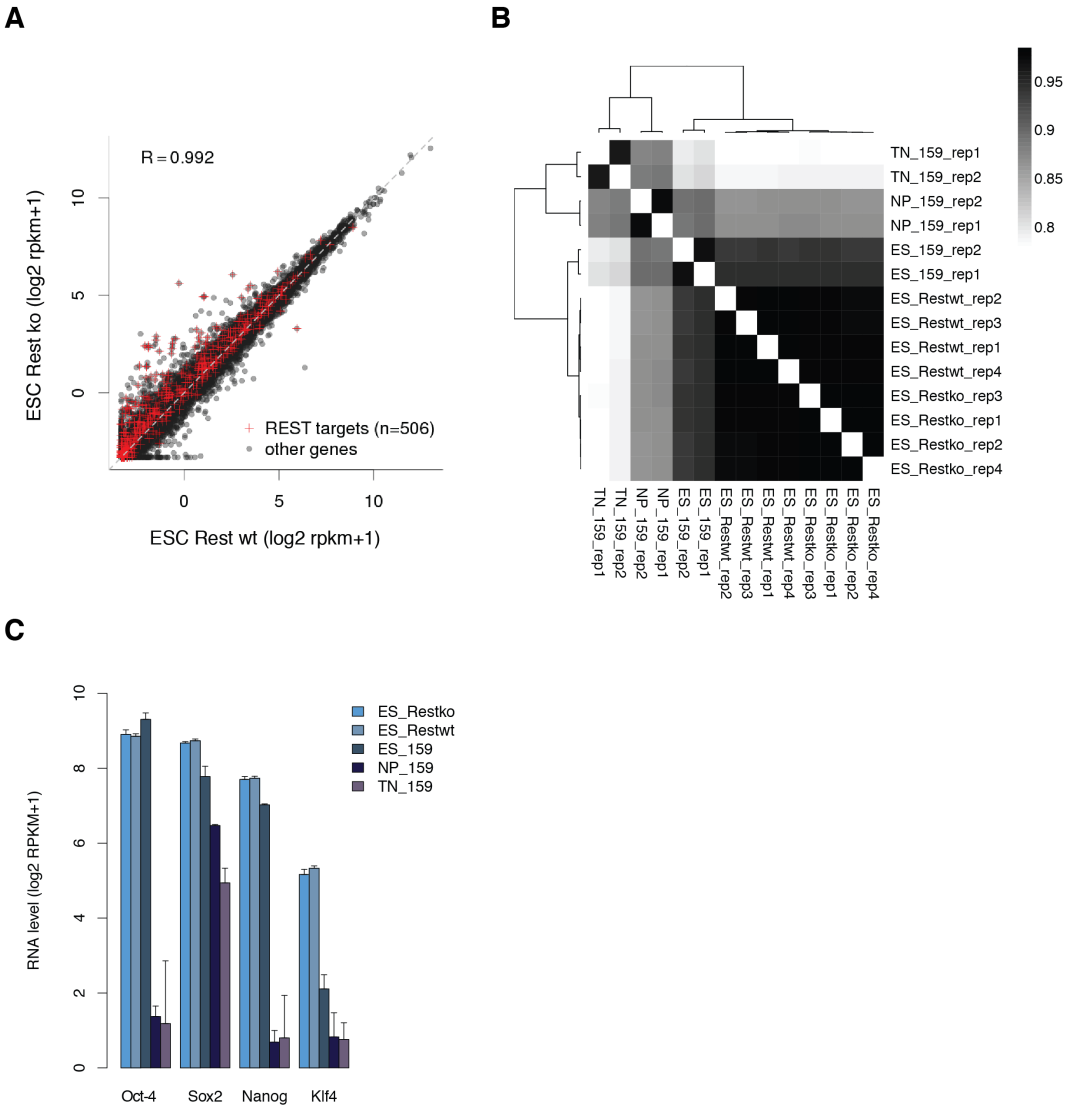

Figure 3.4 | Transcriptional effects of REST deficiency in mouse embryonic stem cells. A The scatter plot shows normalized RNA-seq reads per gene (dots) for putative REST targets (≥ 1 REST peak within 10 kb of TSS, red dots) and other genes (grey dots). Overall, global gene expression is similar between *Rest wt* and *Rest ko* cells. Potential REST target genes show selective transcriptional upregulation. **B** The cross-correlation expression matrix indicates co-clustering of the three embryonic stem (ES) cell lines, irrespective of their genotype. As expected, neuronal progenitors (NPs) and terminal neurons (TNs) are transcriptionally more distant from ESCs. **C** The bar plot depicts similar RNA expression of the classical pluripotency factors in *Rest wt* and *Rest ko* cells.

3.2.2 REST is Necessary and Sufficient for Low Methylation of its Binding Sites

After establishing REST's transcriptional effects in mouse embryonic stem cells, I investigated its interaction with DNA methylation by asking whether REST is necessary and sufficient to establish local hypomethylation.

I compared DNA methylation of distal REST binding sites in the absence or presence of REST protein in previously mentioned *Rest ko* and *Rest wt* cells (Figure 3.5A) (Arnold et al., 2013; Jørgensen et al., 2009b; Stadler et al., 2011). I also stably re-expressed V5-tagged REST in *Rest ko* cells (*Rest*$^{-/-}$*REST*). DNA methylation was profiled by amplicon Bis-seq, which targeted 35 distal REST binding sites and 61 control regions (i.e. not REST bound).

In line with previous reports, REST binding sites show intermediate methylation in *Rest wt* cells (Figure 3.5B). While control regions were highly correlated between *Rest wt* and *Rest ko* cells, REST binding sites became methylated in *Rest ko* cells. Upon re-expression of full-length REST in *Rest ko* cells, REST binding sites were again selectively hypomethylated. However, REST re-expression cells did not recapitulate the full extent of hypomethylation observable in *Rest wt* cells. These results indicate that REST is necessary and sufficient for significant hypomethylation of REST binding sites. It implies that the DNA-binding factor REST is not methylation sensitive as REST binding occurs even at previously methylated REST binding sites (Chapter 3.2.4). These experiments support similar statements made previously based on the study of two REST occupied low-methylated regions (Stadler et al., 2011).

We further tested if the REST motif is sufficient to induce hypomethylation within a random DNA sequence. We replaced the DNA sequence of a previously characterized REST LMR by a random bacterial one and kept the CpG dinucleotides at place (Figure 3.6A) (Stadler et al., 2011). Within this artificial construct we either placed a maximal REST motif (derived from the JASPAR motif database) or a scrambled control motif (Portales-Casamar et al., 2010). Both constructs were then either directly inserted into the mouse beta globin locus or *in vitro* CpG methylated before. Methylation was subsequently profiled by amplicon Bis-seq. After stable genomic insertion, cytosines within both control constructs became fully methylated (Figure 3.6B). In contrast, cytosines were strongly hypomethylated when the REST consensus motif was present.

Methylation was minimal in direct proximity to the REST motif. In case of the unmethylated construct, the CpG within the REST motif became entirely demethylated. The pre-methylated motif construct also showed characteristic hypomethylation, however to a lesser extent. This methylation difference between the unmethylated and pre-methylated motif construct corroborates the above-described observations for full-length re-expressed REST. Whether this difference is biologically meaningful remains speculative at this point. It is possible that analyzed cells have not yet reached methylation equilibrium during the time of cell culture. The results could also imply a certain degree of REST methylation sensitivity or a better maintenance of a prior unmethylated region. I will address the latter possibility of methylation sensitivity in an upcoming section (Chapter 3.2.4). Overall, the REST motif proved sufficient to reduce methylation within a random CpG-poor sequence. The REST motif is the minimal *cis*-component required to establish local hypomethylation.

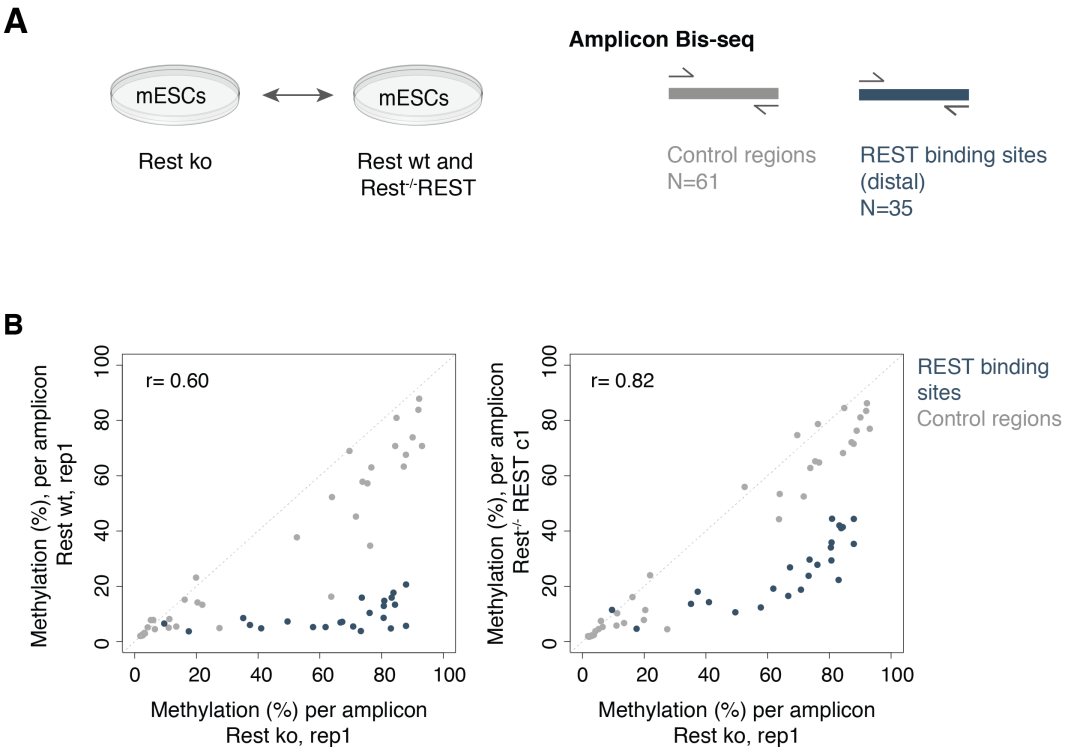

Figure 3.5 | Testing REST's sufficiency for local hypomethylation. A *Rest wt*, *Rest ko* and REST re-expression mESCs are profiled by amplicon Bis-seq. **B** The scatter plots show average methylation levels per amplicon in the three cell lines (tested clones are indicated as *c number*)). Control regions are depicted in grey and distal REST binding sites in blue. Distal REST binding sites show low methylation levels only in the presence of REST protein, i.e. in *Rest wt* and *Rest*$^{-/-}$*REST* cells.

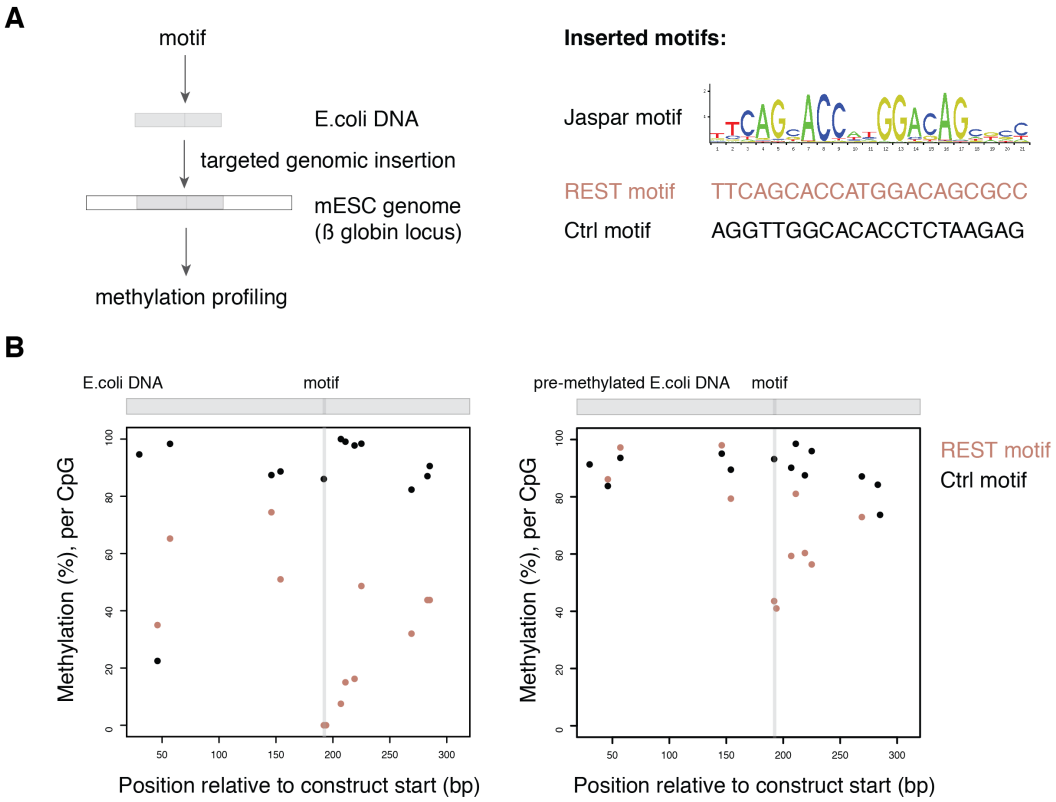

Figure 3.6 | Testing REST motif sufficiency for local hypomethylation. A Artificial bacterial DNA constructs were complemented either with the REST consensus motif or a scrambled control. Constructs were left unmethylated or *in vitro* methylated before stable, targeted insertion into the mouse genome. Methylation was profiled by amplicon Bis-seq. **B** Average methylation per CpG position is shown for control motif (black) or REST motif (red) constructs. In both constructs, cytosines become hypomethylated only in the presence of the REST motif. Methylation reaches a local minimum directly over the REST motif. However, the extent of hypomethylation is less pronounced in the pre-methylated motif construct.

3.2.3 REST Binding Overlaps with Several of its Chromatin Modifying Cofactors

After showing a general sufficiency of REST for local hypomethylation, I was interested which other chromatin modifications are associated with REST binding. We analyzed several known REST interactors and chromatin marks in relationship to REST binding sites.

REST binding sites are frequently overlapping with hypomethylated regions (Stadler et al., 2011). Within REST sites, DNA methylation is negatively correlated with REST ChIP-seq enrichment (Figure 3.7A). On average, methylation is minimal across occupied REST motifs and increases with distance from it (Figure 3.7B). Methylation levels oscillate around the motif with an approximate period of 180 bp, a length compatible with the nucleosomal repeat length (NRL) in mouse embryonic stem cells (Teif et al., 2012). I therefore analyzed the status of nucleosomes over REST binding sites in mESCs using a published MNase-seq data set. In this analysis, nucleosomes are strongly positioned over occupied REST motifs (Figure 3.7B). On the contrary, nucleosomes are not phased over non-bound REST motifs (*data not shown*). We note that the nucleosome and methylation signal are in counter-phase to each other. Local methylation maxima occur over nucleosomal linker regions, while nucleosomal DNA is generally less methylated. Equivalent nucleosome and DNA methylation profiles can be observed over binding sites of the insulator protein CTCF (Fu et al., 2008; Teif et al., 2014). It remains speculative whether both DNA-binding factors mediate nucleosome positioning through active recruitment of chromatin remodeling enzymes. It is possible that REST and CTCF mediate hypomethylation indirectly through nucleosome remodeling.

We extended my analysis of REST binding sites by integrating REST's cofactors and common histone modifications (Figure 3.7C). We categorized all REST motifs by their genomic location into proximal (within 2 kb of an annotated TSS) or distal (outside 2 kb of an annotated TSS), and further into bound or not bound. We contrasted these with REST unbound low-methylated regions (LMRs) and unmethylated regions (UMRs). REST bound proximal and distal sites show lower methylation than unbound REST motifs. Bound distal REST sites are enriched in the putative active demethylation mark 5hmC. We also quantified several chromatin-modifying enzymes that were previously reported to be recruited by

REST. We find evidence for all direct and indirect REST cofactors to bind proximal and distal REST binding sites alike. Bound REST sites show strong enrichment for the N-terminal repressor complex SIN3 (SIN3A, SIN3B) and the C-terminal repressor CoREST (RCOR1, RCOR2, RCOR3). Less, but significant REST cofactor binding is observed in UMRs and LMRs without REST motifs. Particularly SIN3A and SIN3B might be generally shared mESC cofactors that are often recruited by other factors than REST. Concomitantly, we observe REST associated enrichment for indirect interactors like HDAC1, HDAC2, TET1 and TET2. The binding of both TET enzymes is in accordance with the presence of its enzymatic product 5hmC. In line with previous findings, I speculate whether the recruitment of TET1 and TET2 could be indirectly mediated by SIN3A (Cartron et al., 2013; McDonel et al., 2012; Williams et al., 2011a). This could suggest a REST/SIN3A/TET1/2 complex involvement in the putative active demethylation of REST binding sites.

While the REST interactome is largely dominated by known transcriptional repressors, we were interested in the corresponding chromatin signature. We detect no characteristic histone modifications on the strongly positioned nucleosomes. In comparison to average LMRs, distal REST binding sites harbor less of the classical enhancer modifications H3K4me1 and H3K27ac. Proximal REST binding sites on the other hand are designated by a much stronger depletion of active histone modifications. In comparison to unbound controls and UMRs, proximal REST binding sites are characterized by a strong reduction in the active promoter marks H3K4me3 and H3K27ac. Reduced levels of H3K27ac are likely reflecting recruitment of the histone deacetylases HDAC1 and HDAC2. It is important to stress that we observe no general enrichment of H3K27me3 at neither distal nor proximal REST binding sites in mESCs. Thereby, the hypomethylation we observe in REST binding sites seems to be in no relationship to this mutually exclusive chromatin mark.

In total, we generated evidence for strong repressive cofactor co-occurrence at distal and proximal REST binding sites. Particularly proximal REST binding sites are depleted for histone marks indicative of active promoters. Both observations are consistent with previous reports on REST's chromatin profile and its function as a transcriptional repressor (Zheng et al., 2009). However, we find no substantial enrichment for the repressive H3K27me3 mark at neither distal nor

proximal sites. The most striking chromatin feature of REST binding sites is the positioning of nucleosomes around the motif. In addition, REST binding sites show signatures that are compatible with active demethylation. I note that nucleosome positioning and active demethylation could be possible mediators of hypomethylation at REST binding sites. Starting from this precise description of the REST binding landscape, I will now dissect its molecular components and test their requirement for local hypomethylation. I later generate REST mutants, which express only parts of the full-length protein domains. I will refer back to the importance of nucleosome positioning and TET recruitment. Ultimately, I aim to define necessary molecular features of DNA-binding factors that are required to induce local hypomethylation.

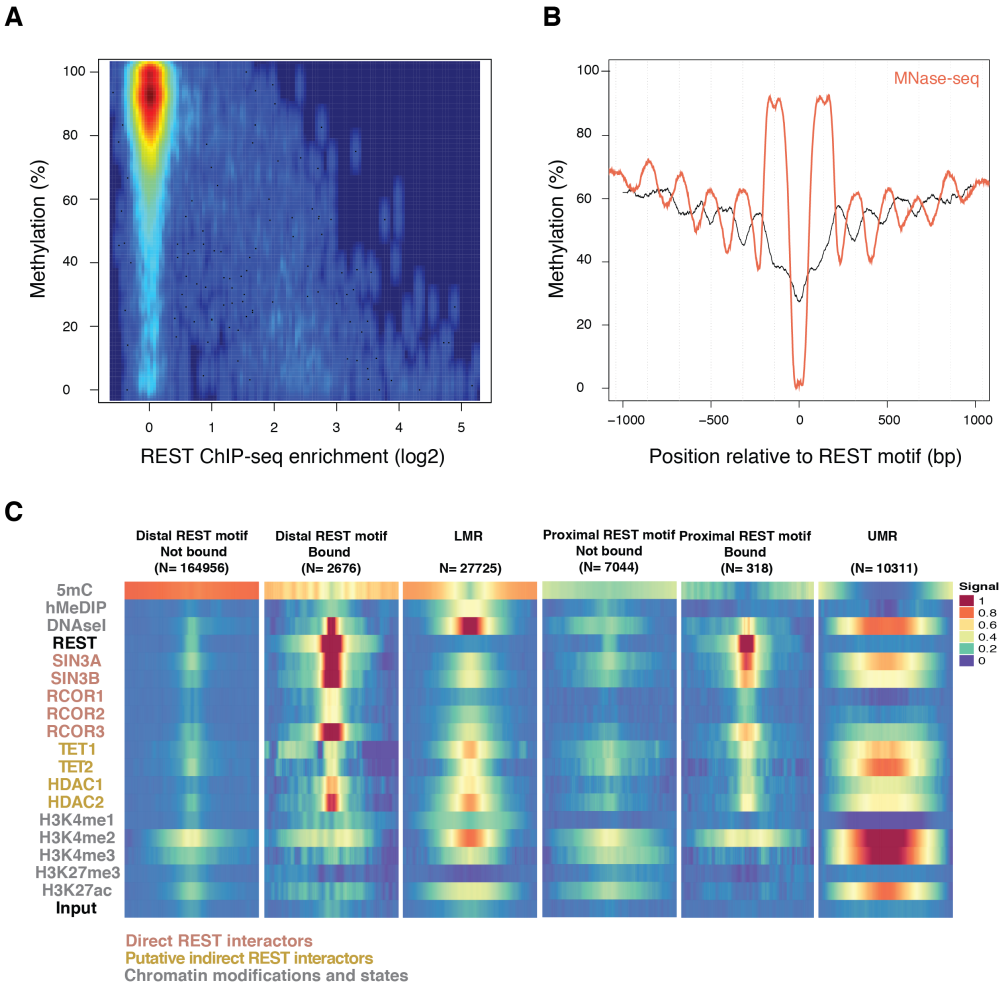

Figure 3.7 | Profiling of REST binding sites for interactors and chromatin modifications. A The two-dimensional heatmap shows average methylation within 200 bp windows around predicted REST motifs as a function of REST ChIP-seq enrichment in the same window. Over REST motifs, methylation is negatively correlated to REST ChIP-seq signal. Strongly bound REST sites show much reduced methylation levels as compared to the genomic average. **B** The composite profile shows average DNA methylation (black) for bound REST sites as a function of motif distance. Nucleosomal positions measured by MNase-seq (red) are superimposed. DNA methylation is minimal directly over the REST motif and increases with distance from it. Nucleosomes are strongly positioned over REST motifs and DNA methylation is higher within nucleosomal linker regions. **C** The heatmap shows normalized log2 enrichments for different chromatin modifications, DNase I and REST interactors. Enrichments are shown within 2 kb windows centered on REST motifs (distal, proximal and bound, not bound) or feature midpoints (LMR, UMR). Direct REST interactors are colored in red, putative indirect interactors in yellow and chromatin features in grey. Bound distal and proximal REST sites show reduced DNA methylation, increased DNase I signal and enrichment for all direct and indirect REST interactors. Significant SIN3A and SIN3B occupancy at UMRs occurs also independent of REST. TET1 and TET2 are possibly recruited to distal REST binding sites by SIN3A, catalyzing the oxidation of 5mC to 5hmC. Concordant with REST's activity as a transcriptional repressor, proximal REST binding sites are characterized by a strong depletion of active promoter modifications (H3K4me3, H3K27ac).

3.2.4 REST Does not Show Signs of Methylation Sensitivity

One obvious requirement for demethylating DNA-binding factors is their ability to interact with initially methylated binding sites. Conversely, all methylation sensitive factors should lack demethylation activity. To test the absence of methylation sensitivity for the transcription factor REST, I asked whether new binding sites are generated in the absence of global DNA methylation.

We compared DNase-I hypersensitivity as a proxy for protein binding over predicted REST motifs in the presence or absence of DNA methylation, using a published DNase-seq data set. Indeed, we observe no discernible differences in DNase signal between *Dnmt wt* and triple knock out (TKO) cells over REST motifs (Figure 3.8A, Figure 3.8B). There are no new hypersensitive REST sites in the absence of DNA methylation. As the canonical REST motif does not contain a strongly conserved CpG dinucleotide, we constricted our analysis to REST motifs containing one CpG (Figure 3.8A). Again, we detected no additional hypersensitivity over REST motifs in *Dnmt TKO* cells.

In summary, we provided additional evidence for methylation insensitivity of REST. In the presence of an unmethylated genome, REST neither explores new motifs nor binds occupied sites stronger.

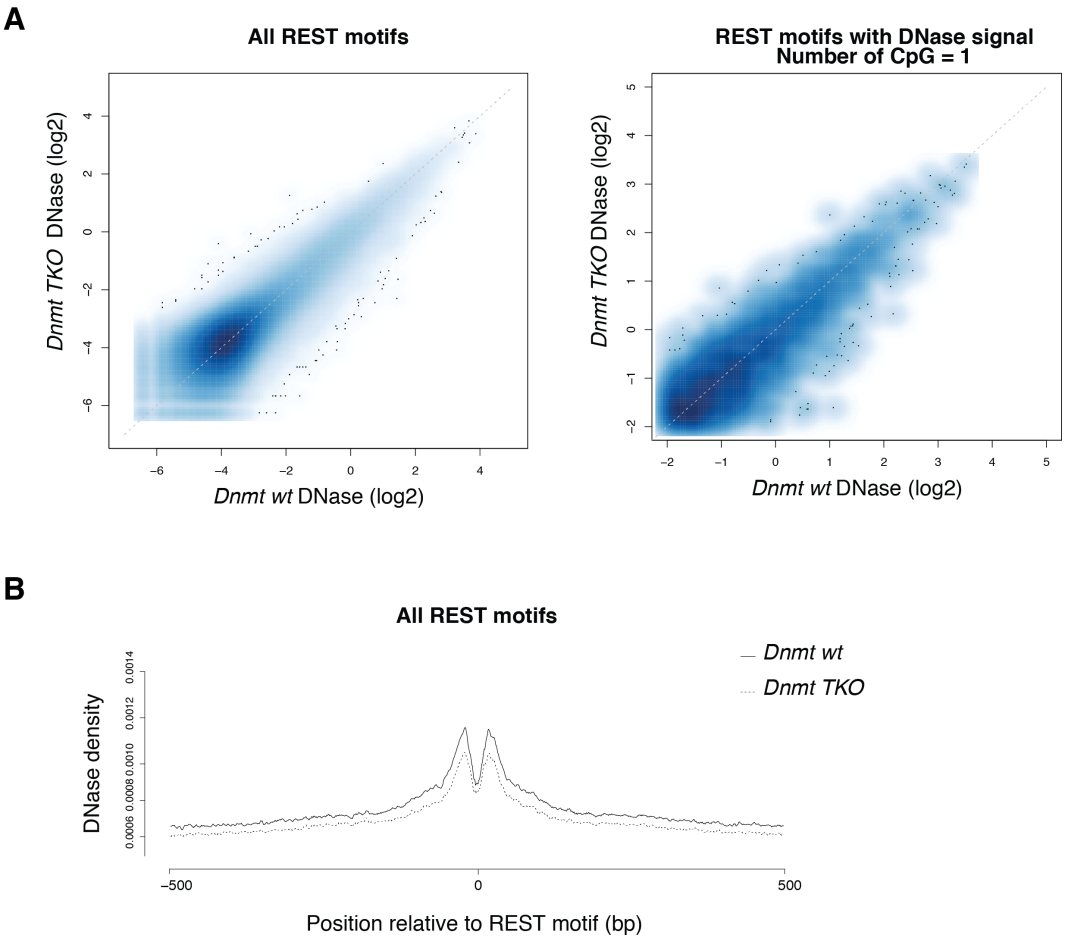

Figure 3.8 | Quantification of DNase hypersensitivity at REST sites in methylation deficient mouse embryonic stem cells. A The density plots show DNase signal within 200 bp windows over predicted REST motifs in *Dnmt wt* and *Dnmt* triple knock out (TKO) cells. In addition, we also limited our analysis to REST motifs that contain one CpG. In both cases, we do not observe significant DNase changes in the absence of DNA methylation. **B** The DNase metaprofiles over REST motifs show similar hypersensitive sites in *Dnmt wt* and TKO cells.

3.2.5 REST's DNA-binding Domain is not Sufficient to Reduce Methylation at Binding Sites

After discussing the necessity for methylation autonomous binding, I examined the possibility that other molecular components aside from a DNA-binding domain are required for hypomethylation. In other words, I ask whether methylation insensitive binding is sufficient to reduce DNA methylation of binding sites.

To address this question, I stably re-expressed V5-tagged mutants of the REST protein into *Rest ko* cells (*referred to as Rest$^{-/-}$protein name*). For this, I utilized the *Rest ko* cells characterized earlier (Chapter 3.2.1, Chapter 3.2.2). I verified expression of each REST mutant by Western Blotting (*data not shown*) and chose three clones per mutant based on high expression levels. Subsequent methylation profiling tested whether methylation of REST binding sites is reduced in either cell line (Figure 3.9A, Figure 3.9B). Based on to the bipartite domain structure of REST, I expressed three different REST deletions. The shortest mutant protein consists of eight zinc finger domains, which define the DNA-binding domain (DBD). The two other mutants comprise either of the two terminal REST repressor domains. The NTE-DBD protein contains the N-terminus up to the DBD, while DBD-CTE contains the DBD and anything C-terminal from it. I also utilized the previously generated REST full-length re-expression cell line. I performed ChIP-seq to verify binding of the re-expressed proteins. DNA methylation was assayed at REST binding sites by using the introduced amplicon Bis-seq design (Chapter 3.1.3).

Binding Profiling of REST Re-expression Cell Lines

I performed V5 ChIP-seq for one clone of each re-expressed REST mutant. Consistent with the assumption that the DNA-binding domain mainly determines DNA-binding specificity, I detected binding for all re-expressed proteins (Figure 3.9C). Approximately half of the peaks observed for full-length REST protein were also detected in the mutant proteins, while no protein possesses unique peaks. I notice the largest peak overlap between DBD-CTE and full-length REST protein.

To avoid binary peak cut offs, I looked at global ChIP-seq signal over predicted proximal and distal REST motifs (Figure 3.10). In comparison to *Rest wt* cells, full-length re-expressed REST shows higher ChIP-seq signal at highly bound sites. This correlation generally supports REST's methylation insensitivity. Further interpretations of subtle differences in ChIP-seq signal need to be avoided due to technical differences in the two experiments. That is, *Rest wt* ChIP-seq was performed with a REST antibody and all REST and REST mutant ChIPs with a V5 antibody. Across all the three mutant proteins I observe high ChIP-seq correlation to the full-length protein. Nonetheless, there are detectable differences between constructs. Similar to the peak analysis, the C-terminal protein DBD-CTE is the most similar to the full-length version. Both shorter mutants, DBD and NTE-DBD bind less strongly. I also asked, whether ChIP-seq signal between constructs is similarly correlated over proximal regions (Figure 3.10). One could speculate, whether increased chromatin accessibility over proximal regions selectively affects mutants like the DNA-binding domain. Quantification of ChIP-seq signal over proximal regions reveals, that all mutants bind similarly stronger.

We next analyzed binding as a function of REST motif strength (Figure 3.11). ChIP-seq enrichment (log2) of all four proteins scales linearly with the score of the associated REST motif. The strongest REST consensus motifs are occupied in all four cell lines. However, the slope of the ChIP-seq log2(enrichment) function is the steepest for full-length and CTE-DBD protein. It is interesting to speculate which components might add additional DNA-binding affinity. I would suggest, that the polylysine stretch immediately C-terminal of REST's DNA-binding domain could increase affinity through unspecific electrostatic interactions with DNA. Such a hypothesis would need to be validated through further experimentation. One could stably express a polylysine-rich stretch in addition to the DNA-binding domain and test its binding strength in ChIP-seq experiments. Randomization of the amino acid sequence within the lysine-rich domain would test the necessity of charged lysines versus specific amino acid motifs.

I conclude, that all four re-expressed REST proteins are able to bind their target sites as defined by the presence of a REST motif. Full-length REST protein and the C-terminal protein DBD-CTE show the highest enrichment over REST binding sites.

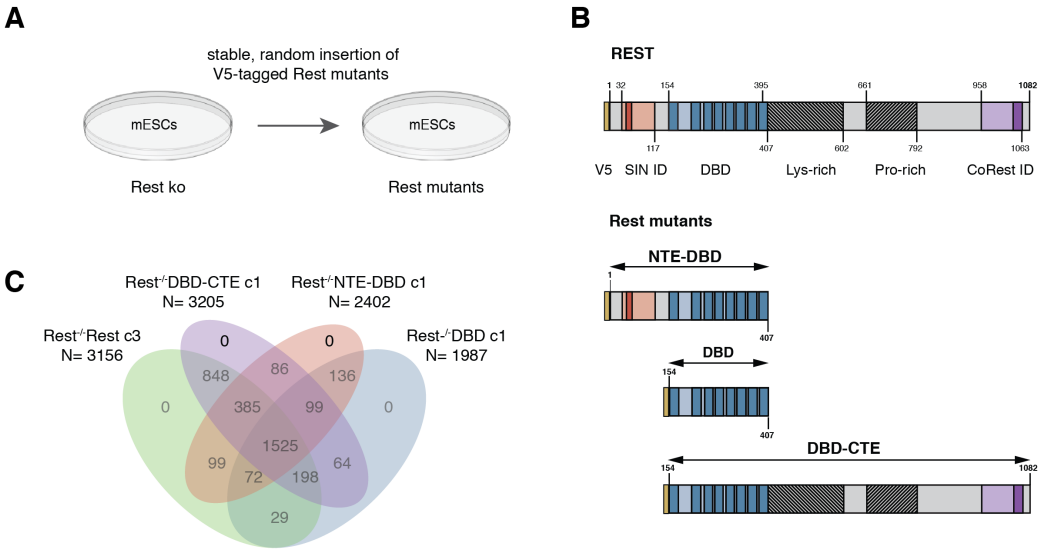

Figure 3.9 I Characterizing binding sites of REST and REST mutants. A Full-length REST and its mutants are stably re-expressed into *Rest ko* cells. **B** I generated three different V5-tagged mutants of the full-length protein. The minimal mutant consists of the DNA-binding domain (DBD) and the other two proteins contain either N-terminal or C-terminal repressor domain (NTE-DBD, DBD-CTE). **C** The Venn diagram illustrates the peaks called from V5 ChIP-seq experiments of all four re-expression cell lines (tested clones are indicated as *c number*). Half of the REST full-length peaks are shared between all proteins. However, the C-terminal protein DBD-CTE shows the largest peak overlap with wild type REST.

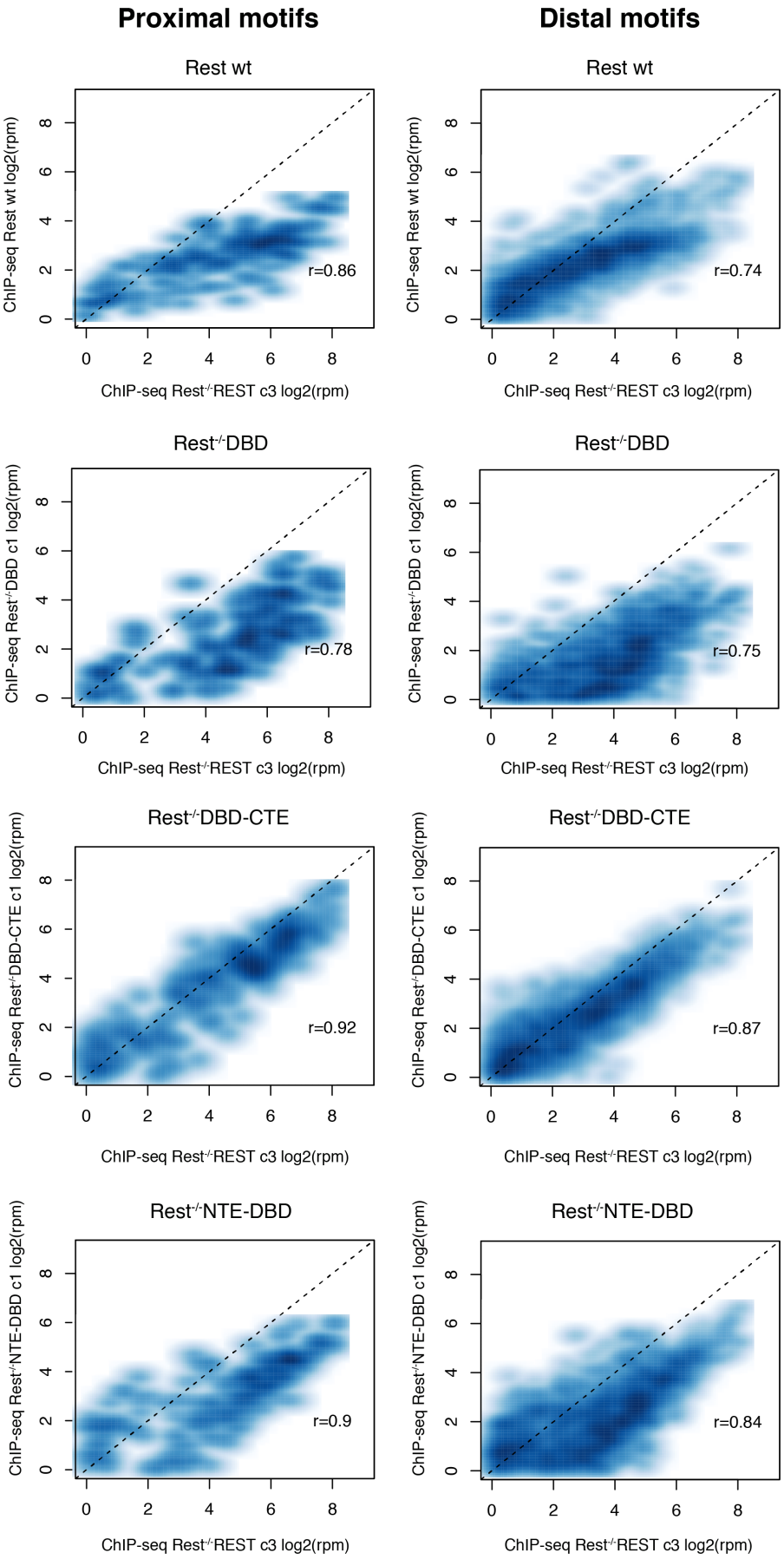
Proximal motifs
Distal motifs
Rest wt
ChIP-seq Rest wt log2(rpm)
r=0.86
r=0.74
Rest-/-DBD
ChIP-seq Rest-/-DBD c1 log2(rpm)
r=0.78
r=0.75
Rest-/-DBD-CTE
ChIP-seq Rest-/-DBD-CTE c1 log2(rpm)
r=0.92
r=0.87
Rest-/-NTE-DBD
ChIP-seq Rest-/-NTE-DBD c1 log2(rpm)
r=0.9
r=0.84
ChIP-seq Rest-/-REST c3 log2(rpm)
0
2
4
6
8

Figure 3.10 | Characterizing binding behavior of REST and REST mutants. Density plots show ChIP-seq signal normalized to mapped library size and quantified within 200 bp windows over predicted REST motifs for *Rest wt* and REST re-expression cells (tested clones are indicated as *c number*). All correlations are shown in respect to full-length re-introduced REST. Motifs are further distinguished based on proximal or distal (> 2kb of TSS) genomic location. While all proteins show similar binding behavior as full-length re-expressed REST, the C-terminal protein is the most correlated. In addition, all proteins bind stronger over proximal sites than distal motifs.

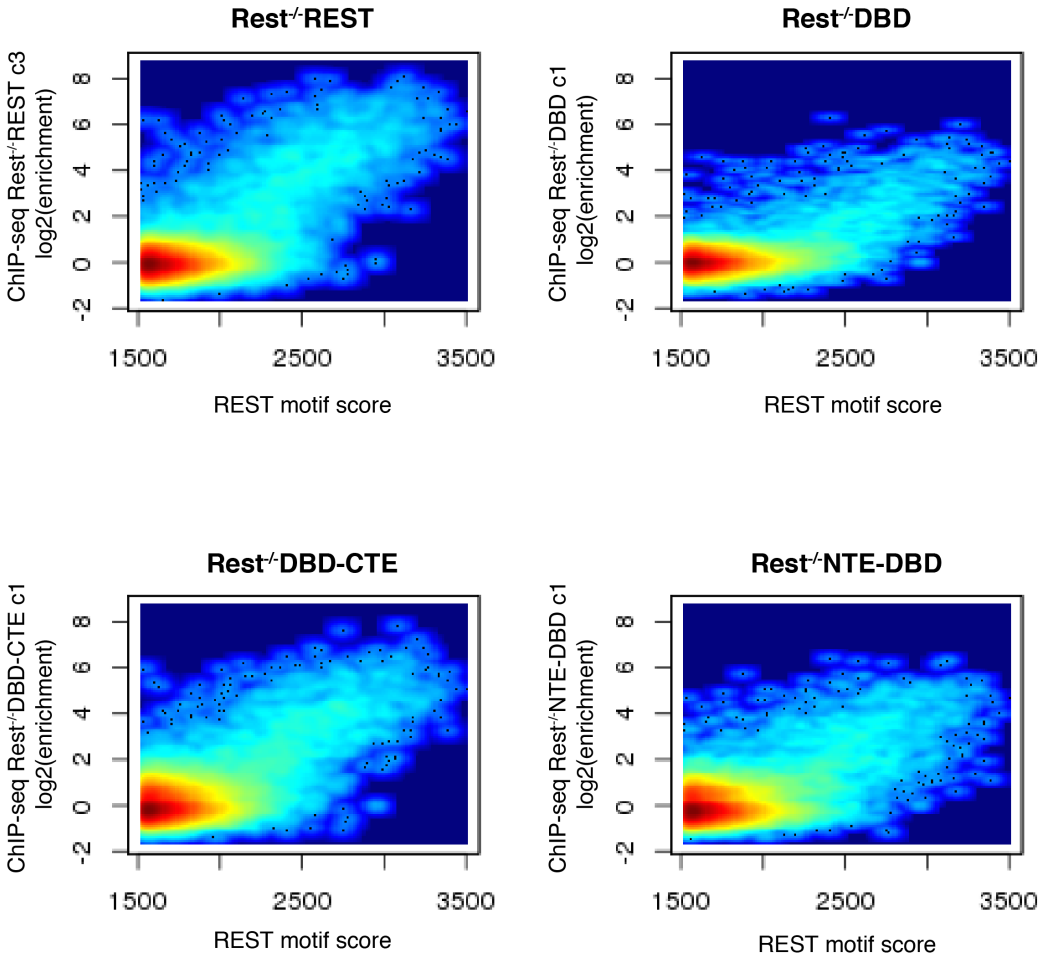

Figure 3.11 | Relationship of protein binding and REST motif score. The heatmaps show the relationship between REST motif score and the associated ChIP-seq enrichment within 200 bp windows over predicted REST motifs, excluding those that overlap with CpG islands. We display all REST re-expression cell lines (clone numbers are indicated as *c number*). In all cell lines, protein binding increases with higher motif strength. The strongest consensus motifs are occupied in all cell lines.

Methylation Profiling of REST Re-expression Cell Lines

After verifying binding of all tested REST mutants, I characterized the methylation states of REST binding sites. I performed targeted methylation profiling by amplicon Bis-seq for 61 control regions and 35 distal REST-binding sites (Figure 3.12A). I chose three different cellular clones for each REST re-expression cell line and three biological replicates for *Rest wt* cells. I analyzed three different clones of *Rest ko* cells that expressed an empty expression plasmid.

I performed hierarchical clustering on all cytosines within REST binding sites that were covered by at least 50 reads in all experiments (Figure 3.12B). This revealed general clustering according to expressed protein, indicating high biological reproducibility at clonal level. Two main vertical clusters separate the cell lines into strongly hypomethylated and weakly hypomethylated/methylated. Different degrees of hypomethylation are associated with all REST re-expression mutants, except for the DNA-binding domain alone (Figure 3.12B). The *REST$^{-/-}$ DBD* cell line shows methylation levels even higher than *Rest ko* cells. On the contrary, all other proteins with at least one REST repressor domain show different degrees of hypomethylation. Lowest methylation levels over REST sites are observed in *Rest wt* cells. Strong hypomethylation over the sites is detected after expression of full-length REST and the N-terminal mutant. The C-terminal protein shows hypomethylation as well, but for fewer cytosines.

To gain more insight into differentially methylated cytosines, I looked into REST motif strength and motif distance. I detect cytosines that show higher methylation levels across all samples (Figure 3.12B, cluster Ib). These cytosines are typically farther away from REST motifs. This general distance effect is in line with the previously described genome-wide methylation average over REST motifs (Chapter 3.2.3, Figure 3.7B). In contrast, I observe cytosines that remain relatively unmethylated in all samples (Figure 3.12B, cluster IIb). Those cytosines are less REST-dependent and their methylation levels increase only slightly in *Rest ko* cells. I find no particularly unique motif feature, but it is possible that the corresponding sites are subject to multiple protein binding events and are thereby less dependent on REST. The remaining two clusters are especially informative as they show hypomethylation in the presence of full-length protein and high methylation in *Rest ko* and *REST$^{-/-}$DBD* cells. Cluster Ia contains cytosines,

which are methylated in *REST$^{-/-}$DBD-CTE* cells and show either similar methylation or hypomethylation in *REST$^{-/-}$NTE-DBD*., The cluster IIa contains cytosines that are demethylated upon expression of both C- and N-terminal proteins or the C-terminal protein alone. Overall, it is interesting to note that some cytosines are selectively responsive to either the N-terminal or the C-terminal protein.

I further investigated cytosines within the context of single REST binding sites (Figure 3.13). In most cases, hypomethylation can be observed for *Rest wt*, *Rest$^{-/-}$REST* and *REST$^{-/-}$NTE-DBD* cells (Figure 3.13A). Hypomethylation is less pronounced at most of the DBD-CTE bound regions and in general cannot be inferred from the ChIP-seq signal. In most regions, cytosines within the *REST$^{-/-}$DBD* cell line are hypermethylated even when strongly bound by the DNA-binding domain. In addition, I identify sites that are selectively demethylated by the full-length and the C-terminal protein (Figure 3.13B). In those cases, the DBD-NTE protein seems to be less bound. I also detect two REST-binding sites, where a single cytosine within the REST motif shows reproducibly reduced levels of methylation even in the *REST$^{-/-}$DBD* cell line (Figure 3.13C). Yet, in all other cases, I do not identify hypomethylation of motif-associated CpGs even at stronger DBD occupied sites. Due to the limited number of binding sites I assay, I cannot make further conclusions on these rare DBD-associated observations.

In summary, I report significant hypomethylation only upon expression of DNA-binding factors that contain at least one protein interaction domain. The possession of a methylation insensitive DNA-binding domain did not prove sufficient to significantly reduce methylation of most occupied sites. The induced hypomethylation did not always scale with protein binding measured by ChIP-seq. These mutant experiments suggest that hypomethylation is not an obligatory consequence of protein binding events, but rather requires interacting domains that could reflect the importance of cofactor recruitment.

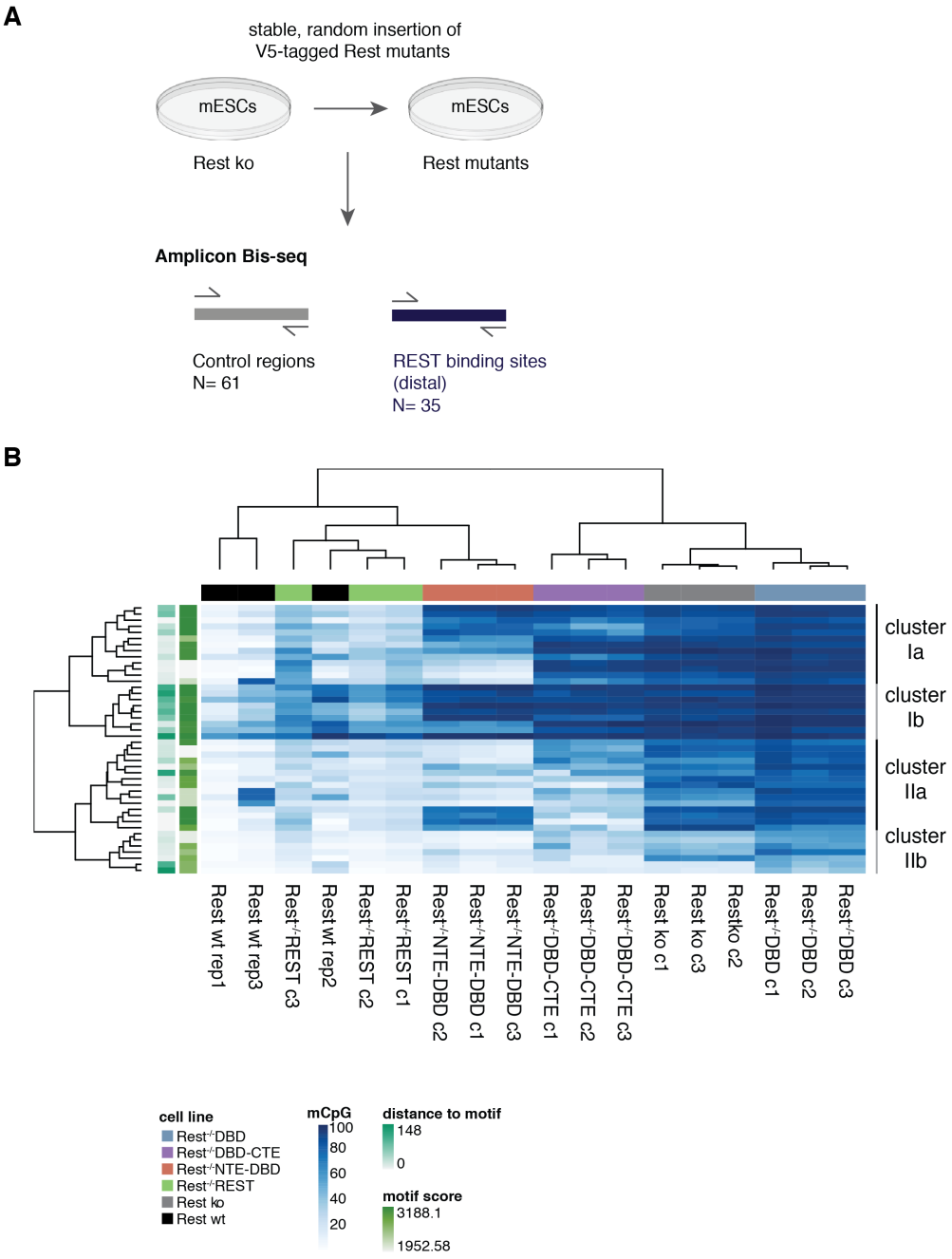

Figure 3.12 I DNA methylation profiling in REST mutant cells. A Stable *Rest ko*, *Rest wt* and REST re-expression cell lines were assayed by amplicon Bis-seq. **B** The heatmap summarizes the methylation results on a single cytosine level (rows) per cell line (columns, colors decode genotype, clone numbers as *c number*, replicates as *rep number*). Increasing intensity of blue indicates increasing average methylation levels. I only depict cytosines within REST binding sites covered by at least 50 reads in all samples. I observe hypomethylation in all cell lines, expressing a REST protein with at least one of the two repressor domains. In the absence of REST (*Rest ko*) or upon expression of the DNA-binding domain, cytosines remain methylated. The latter shows that DNA-binding alone is not sufficient to reduce local methylation, but requires additional interaction domains.

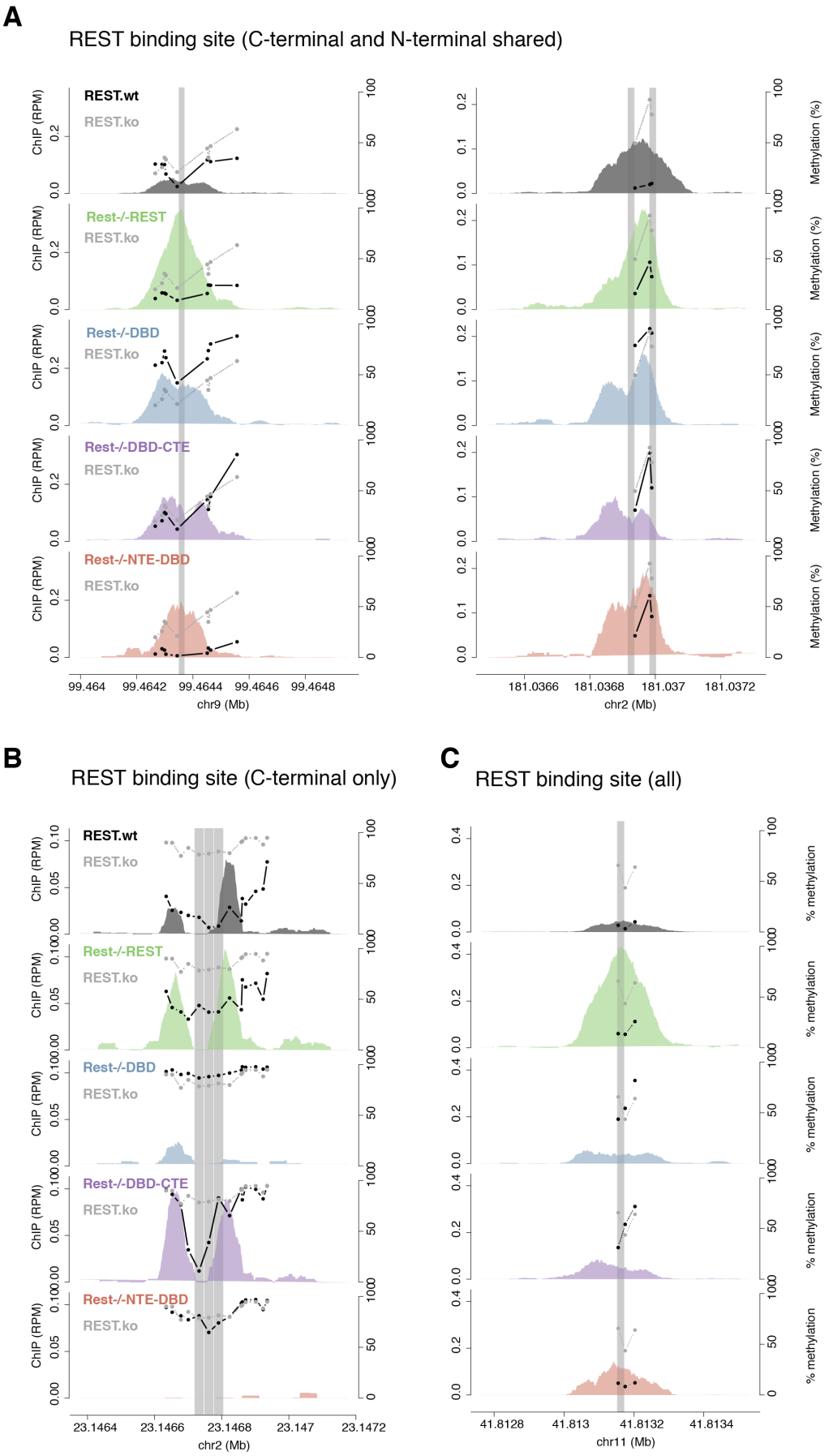
A
REST binding site (C-terminal and N-terminal shared)
REST.wt
REST.ko
Rest-/-REST
Rest-/-DBD
Rest-/-DBD-CTE
Rest-/-NTE-DBD
ChIP (RPM)
Methylation (%)
99.464
99.4642
99.4644
99.4646
99.4648
chr9 (Mb)
181.0366
181.0368
181.037
181.0372
chr2 (Mb)
B
REST binding site (C-terminal only)
23.1464
23.1466
23.1468
23.147
23.1472
chr2 (Mb)
C
REST binding site (all)
% methylation
41.8128
41.813
41.8132
41.8134
chr11 (Mb)

Figure 3.13 | Methylation and protein binding within single REST binding sites. I chose four representative REST binding sites for different hypomethylation classes. Methylation is depicted as dot connected lines (black respective cell line, grey *Rest ko*) and densities depict ChIP-seq signal (rpm) of the respective cell line. Grey vertical rectangles indicate the location of REST motifs (note the existence of multiple REST motifs).

3.2.6 N-terminal REST Protein Reduces DNA Methylation of REST Binding Sites and Recruits SIN3A

The previous chapter has described cytosine methylation as a function of different protein binding events. I further investigated the differences between the DNA-binding domain and the N-terminal protein.

While expression of the DNA-binding domain alone does not induce substantial hypomethylation of binding sites, strong hypomethylation was observed upon expression of N-terminal REST (Figure 3.14A). The methylation changes induced by the NTE-DBD protein are strongly correlated with its binding strength (r= 0.77) (Figure 3.14B). Consequently, highly occupied binding sites undergo stronger reduction in methylation upon expression of NTE-DBD. It is striking, that the expression of 153 amino acids (aa) N-terminal of the DNA-binding domain re-establishes almost the complete REST site hypomethylation observed for full-length protein. I decided to use this confined molecular space as an opportunity to investigate the requirement for certain protein interaction domains. The only annotated domains within the N-terminal stretch (1 – 153 aa) are the interaction domains with the N-terminal corepressor complex SIN3 (Grimes et al., 2000; Nomura et al., 2005). We therefore quantified global SIN3A and SIN3B occupancy over REST binding sites (Figure 3.15A). I notice that SIN3A and SIN3B binding scales with REST ChIP-seq signal over predicted REST binding sites. There is no evidence for selective binding of either SIN3 paralogue, since SIN3A and SIN3B binding is well correlated over REST binding sites. However, due to the population readout of ChIP-seq, I cannot distinguish if SIN3A and SIN3B bind simultaneously or alone on a single allele.

I next asked, whether the N-terminal protein is indeed sufficient to recruit SIN3A to its binding sites. SIN3A ChIP-qPCR was performed over REST binding sites previously assayed in amplicon Bis-seq (Figure 3.15B). I observe selective SIN3A enrichment over REST binding sites in *Rest wt* and *REST$^{-/-}$NTE-DBD* cells. There is reduced or no SIN3A binding at REST binding sites in *Rest ko*, *REST$^{-/-}$DBD* and *REST$^{-/-}$DBD-CTE* cells. REST thereby proved necessary for SIN3A recruitment, while the N-terminal domain was sufficient to recruit SIN3A to REST motifs.

In order to show direct recruitment of SIN3A, we mutated its putative interaction domain within the N-terminal part of REST. Out of the 153 N-terminal amino

acids, positions 40 to 80 show high evolutionary conservation (Figure 3.15C). This conserved part overlaps with reported interaction domains for SIN3A (32 – 117 aa) and SIN3B (43 – 57 aa) (Chapter 1.4.2) (Grimes et al., 2000; Nomura et al., 2005). I introduced eight point mutations within this conserved region that would likely destroy protein-protein contacts (Figure 3.15C). I transiently transfected wild type and mutated NTE-DBD in high and low amounts into *Rest ko* cells. I subsequently performed V5 ChIP-qPCR to verify protein binding and SIN3A ChIP-qPCR to test SIN3A recruitment (Figure 3.15D, Figure 3.15E). While both wild type and mutated NTE-DBD proteins bind their target sites, only the wild type protein leads to sufficient SIN3A recruitment. These experiments indicate that the introduced point mutations interrupt the N-terminal ability to recruit SIN3A to binding sites.

Together, I have shown that N-terminal REST binding leads to a large reduction in methylation of REST binding sites. The extent of NTE-DBD induced demethylation is correlated to its genomic binding. In accordance with its contained interaction domains, N-terminal REST is sufficient to recruit SIN3A to assayed REST binding sites. Point mutations within these interaction domains reduce the ability of REST's N-terminus to recruit SIN3A.

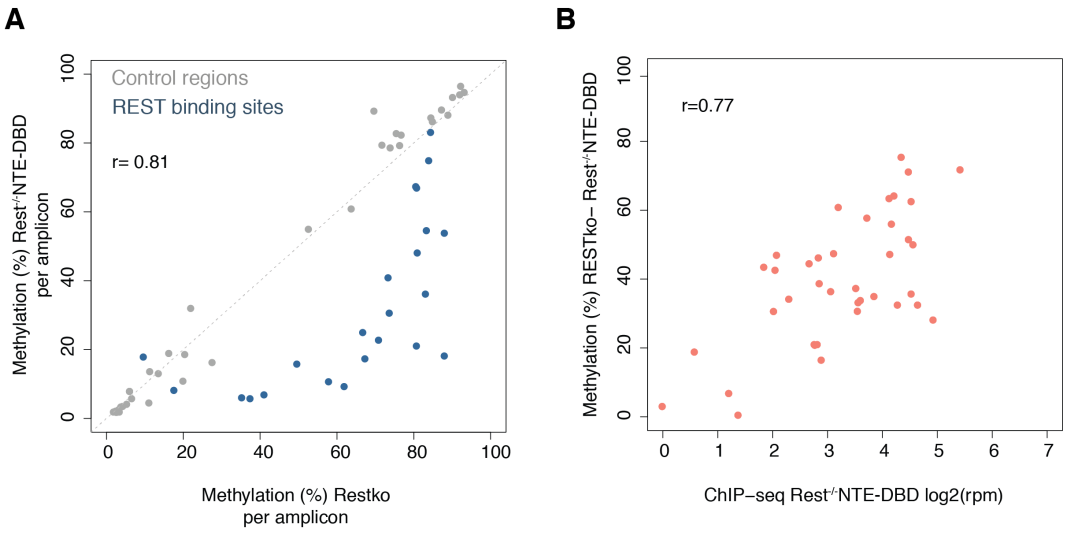

Figure 3.14 | Methylation within the N-terminal REST cell line and its relationship to binding. A The scatterplot shows the average methylation over assayed amplicons for control regions (grey) and REST binding sites (blue). Methylation averages of control regions are highly correlated; only REST binding sites show strong hypomethylation upon expression of NTE-DBD. **B** Methylation delta (*Rest ko – Rest*$^{-/-}$*NTE-DBD*) and ChIP-seq signal is averaged in 200 bp windows over REST motifs. The degree of hypomethylation induced by NTE-DBD scales with its binding.

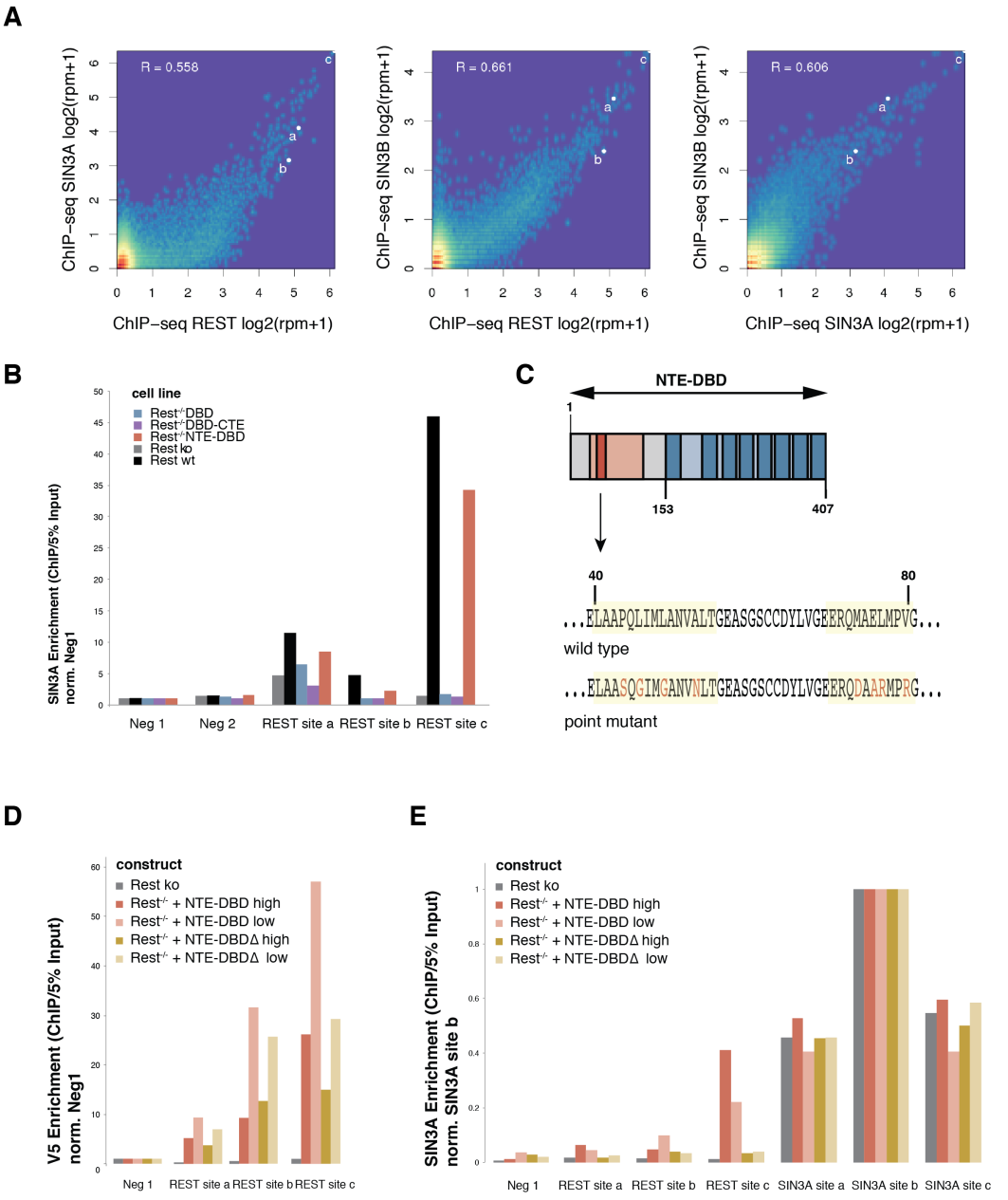

Figure 3.15 I Testing N-terminal REST for its ability to recruit SIN3A. A Two-dimensional heatmaps show correlation between SIN3A, SIN3B and REST ChIP-seq signal over predicted REST motifs. Binding of both SIN3 paralogues is correlated to REST ChIP-seq signal. White labeled dots indicate binding sites assayed by SIN3A ChIP-qPCR. **B** SIN3A ChIP-qPCR shows SIN3A enrichment only at REST binding sites in *Rest wt* and *REST$^{-/-}$NTE-DBD* cells. **C** The N-terminal REST protein NTE-DBD contains the DNA-binding domain and an N-terminal stretch, consisting of 153 amino acids. Here, I introduced eight point mutations at highly conserved residues. Yellow highlighting indicates high degree of conservation and red letters indicate introduced point mutations. **D** Wild type and mutated NTE-DBD were transiently transfected in high and low amounts into *Rest ko* cells. V5 ChIP-qPCR was performed to verify similar binding of wild type and mutant protein. **E** SIN3A ChIP-qPCR indicates that the introduced point mutations within NTE-DBD abrogate recruitment of SIN3A.

3.2.7 Hypomethylation in the Vicinity of the REST Motif is TET dependent

Having verified the ability of the N-terminal domain to recruit SIN3A to its target sites, the question remains how hypomethylation is mediated. SIN3A does not possess catalytic activity on its own but rather functions as an adaptor protein (Pazin and Kadonaga, 1997). Several enzymes are part of the SIN3 repressor complex, amongst it the methylcytosine dioxygenases TET1 and TET2 (Cartron et al., 2013; McDonel et al., 2012; Williams et al., 2011a). The interaction of both TET paralogues with SIN3A opens up the possibility that the strong hypomethylation by NTE-DBD is partially or fully mediated by TET activity. In order to test this hypothesis, we analyzed a published methylome of *Tet* triple knock out (TKO) mouse embryonic stem cells (Lu et al., 2014a). We assayed average methylation over occupied REST motifs in *Tet wt* and *Tet* TKO cells. We detected strong hypermethylation over REST binding sites in the absence of TET proteins. The TET dependency was higher at highly occupied REST sites (*data not shown*). Hypermethylation however is not detected throughout the REST site, but is confined to the direct vicinity of the REST motif (~50 bp ± relative to motif end/start). Cytosines within the motif and the rest of the region remain unchanged. Interestingly, an equivalent localized methylation effect in *Tet TKO* cells can also be observed over occupied CTCF motifs (*data not shown*).

This observation suggests that hypomethylation within the first linker region is molecularly distinct and a consequence of active demethylation. I note, that the TET-dependent hypermethylation effect occurs where I expect binding of REST cofactors like SIN3A. Possibly, TET enzymes are locally recruited through the REST-SIN3A interaction. I do indeed observe a significant hypermethylation effect over SIN3A binding sites (*data not shown*). However, I cannot exclude that TETs are recruited indirectly to this linker sites. Increased chromatin accessibility or looping to TET-bound promoter regions could be alternative reasons for TET binding. Yet, additional TET-independent mechanisms must contribute to hypomethylation outside of REST motif proximal linker regions. In the upcoming chapter, I will discuss nucleosome positioning as such a potential molecular mediator.

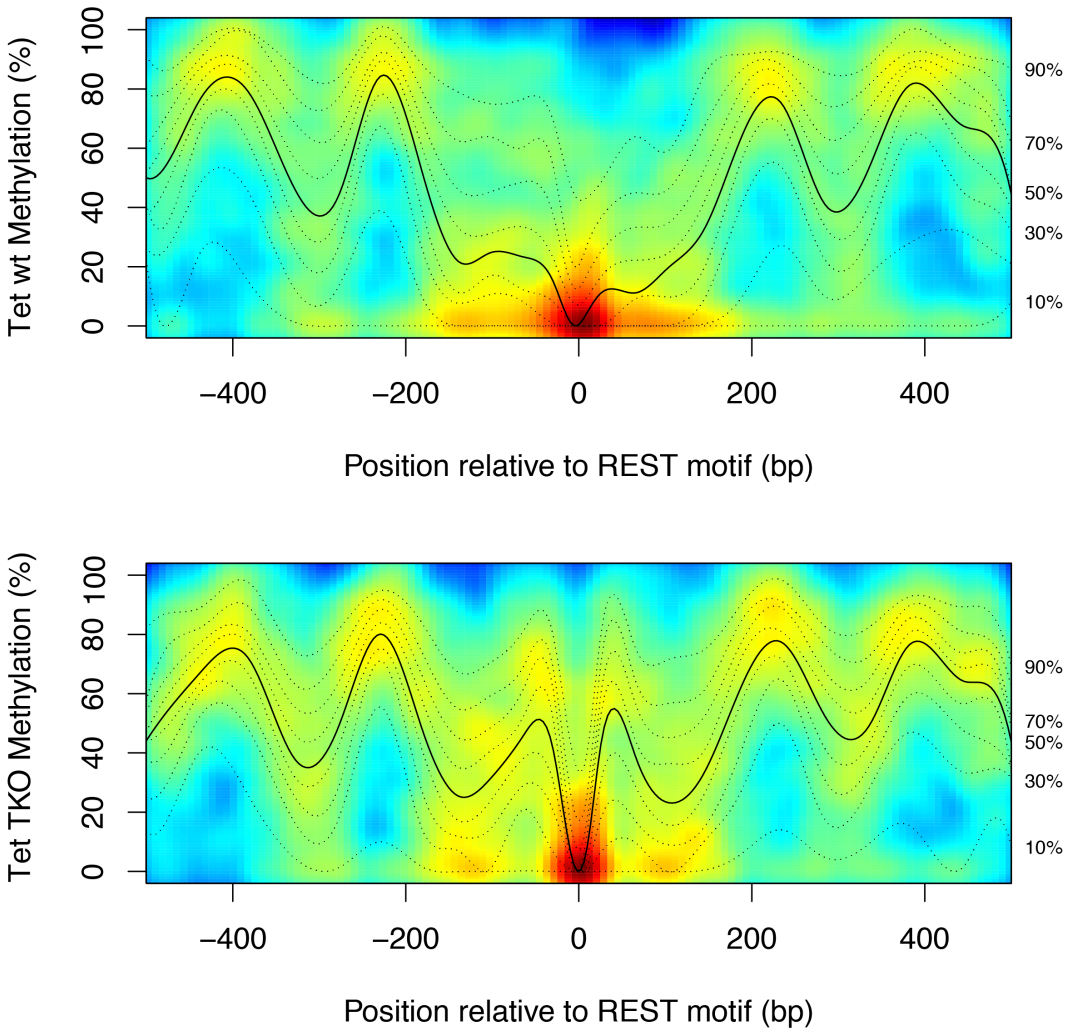

Figure 3.16 | Methylation analysis at REST binding sites in *Tet* deficient cells. Density profiles show cytosine methylation (%) in *Tet wt* and *Tet1/2/3 triple knock out (TKO)* embryonic stem cells as a function of distance to the nearest occupied REST motif. Lines indicate quantiles (solid line depicts median quantile, dashed lines remaining quantiles). Regions overlapping CpG islands or multiple REST motifs were removed from the analysis. Methylation shows characteristic periodicity across the bound REST regions. Only cytosines in direct vicinity of the REST motif gain methylation in the absence of all TET enzymes. Methylation levels over the motif, remaining nucleosomes and linker regions remain unchanged.

3.2.8 Chromatin Accessibility and Nucleosome Positioning are Altered in DNA Hypomethylated Cells

So far, I have described several chromatin differences associated with REST binding and found some of them to be required for reduced methylation. I now ask whether the ability of a protein to reduce methylation is connected to chromatin accessibility changes.

To answer this, we applied Nucleosome Occupancy Methylome sequencing (NOMe-seq) (Kelly et al., 2012; Pardo et al., 2009, 2011). NOMe-seq has the advantage of giving single molecule information about DNA methylation and general chromatin accessibility, from which nucleosome positions and protein occupancies can be inferred (Figure 3.17A). Cell nuclei are treated *in vitro* with a GpC methyltransferase, which preferentially introduces non-mammalian GpC methylation at accessible cytosines. We performed targeted NOMe-seq on previously described REST re-expression cell lines and quantified GpC/CpG methylation by amplicon Bis-seq (Chapter 3.2.5, Figure 3.9B).

In this analysis, increased chromatin accessibility was only detected in cells with reduced methylation at REST sites (Figure 3.17B). In comparison to *Rest ko* cells, all cell lines but *REST$^{-/-}$DBD* show higher GpC methylation attributed to increased accessibility. Noteworthy, again the N-terminal REST expressing cell line most closely resembles *Rest wt* cells.

In addition, we observe distinct nucleosome re-organization over averaged REST motifs (Figure 3.18A). In accordance with the MNase-seq averages in mouse embryonic stem cells, we observe strong nucleosome positioning over REST motifs in *Rest wt* cells (Chapter 3.2.3, Figure 3.7). More specifically, we detect increased chromatin accessibility in the vicinity of REST motifs and a local decrease over lengths compatible with nucleosomal DNA. The locations of nucleosomes defined by NOMe-seq match those of average positions in MNase-seq data sets. We observe no structured nucleosome organization in *Rest ko* and *REST$^{-/-}$DBD* cells. Expression of N-terminal REST again establishes nucleosome positioning equivalent to the one in *Rest wt* cells. We also determined local protein footprints directly over the REST motif. This effect is the most pronounced for full-length protein and NTE-DBD. These observations can also be observed on the level of single loci (Figure 3.18B). Importantly, we observe no coupling between protein binding and reduced CpG methylation within single molecules

(*data not shown*). This observation is in agreement with earlier observations made for the demethylating factor CTCF (Feldmann et al., 2013).

In sum, NOMe-seq has suggested that the ability to reduce local methylation is co-associated with distinct nucleosome positioning. In this scenario, nucleosomes are depleted over REST motifs and adjacent nucleosomes are strongly phased. These nucleosomes could serve as molecular intermediates enabling the spreading of an initial motif-associated signal over larger distances. There is indeed structural evidence, showing that the *de novo* methyltransferases anchor on nucleosomal DNA to methylate adjacent DNA (Guo et al., 2015). Therefore, the positioning of nucleosomes could in fact lead to oscillatory methylation levels as observed around transcription factor binding sites.

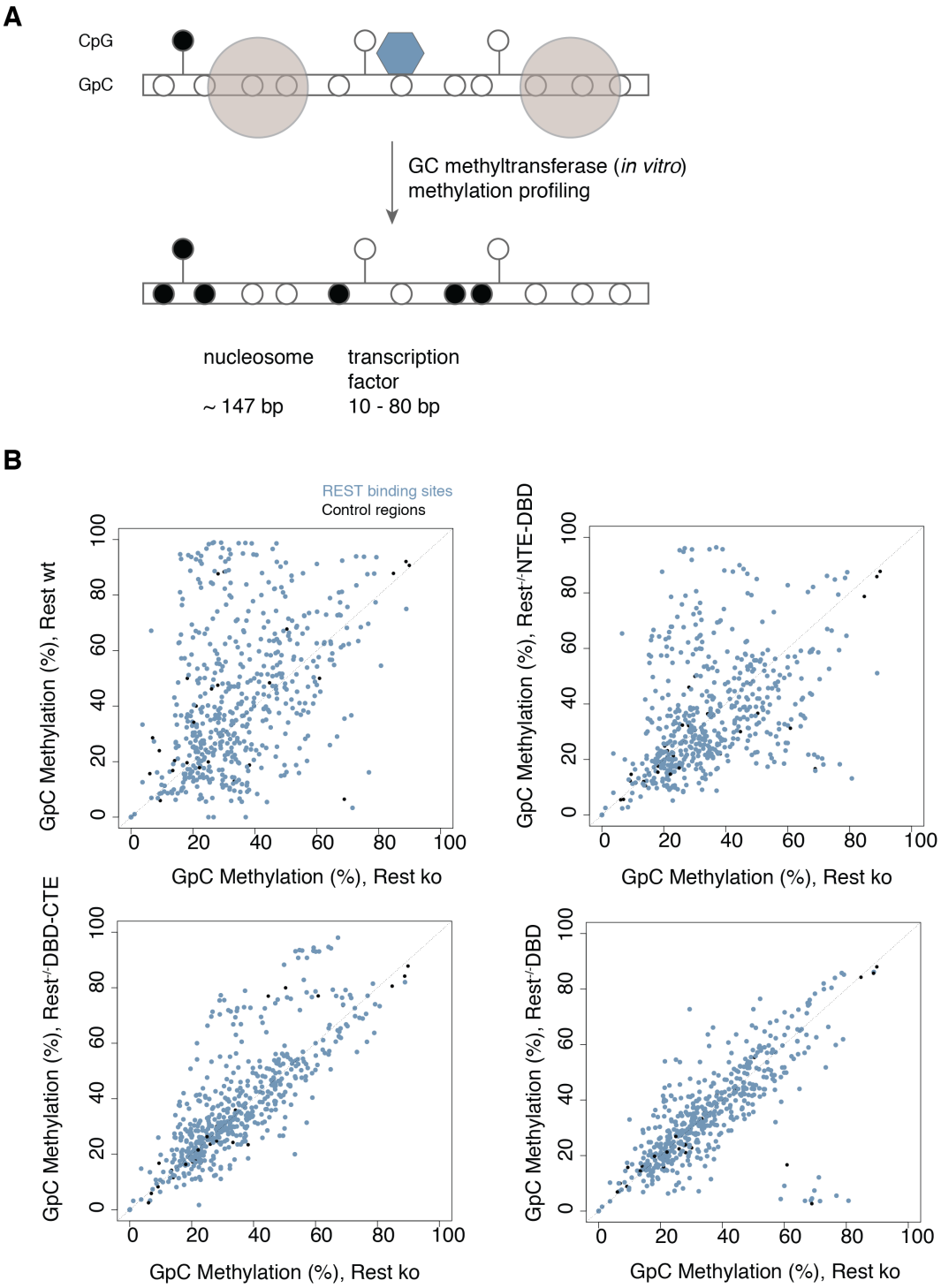

Figure 3.17 | Single-molecule chromatin accessibility profiling in REST cell lines. A Schematic overview of NOMe-seq, which measures chromatin accessibility on a single molecule level. Nucleosomes (grey circles) and DNA-binding factors (blue octagons) lead to decreased GpC methylation. **B** Scatter plots depict correlations between GpC methylation in different cell lines compared to *Rest ko* cells. Blue dots depict GpCs within REST binding sites and black ones within control regions (non REST bound).

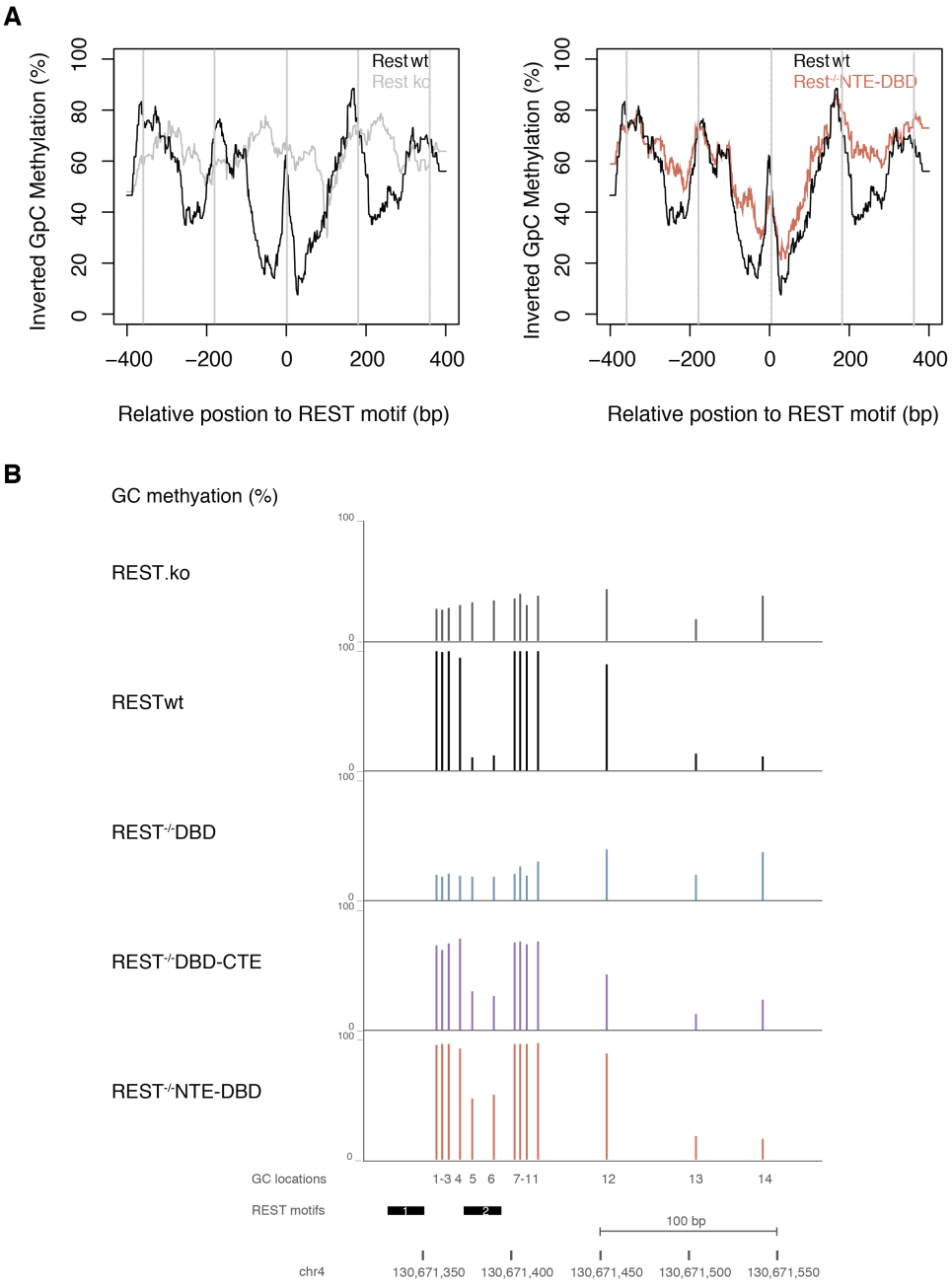

Figure 3.18 I Measuring chromatin accessibility, protein binding and nucleosome positioning in REST mutants. A Composite profiles over REST motifs show protein binding footprints, nucleosomes and increased chromatin accessibility in *Rest wt* and *REST$^{-/-}$NTE-DBD* cells. Note, that we depict inverted GpC methylation (100%-GpC methylation %). Local maxima of inverted GpC methylation denote nucleosomes or DNA-binding factors. Grey vertical lines indicate average nucleosome positions measured by MNase-seq or the REST motif. **B** We depict GpC methylation levels (bars) in assayed *Rest wt*, *Rest ko* and REST re-expression cell lines. In general, we can detect increased chromatin accessibility in vicinity to the second REST motif for *Rest wt*, *REST$^{-/-}$DBD-CTE* and *REST$^{-/-}$NTE-DBD* cells (GpCs 1-4, 7-11). We observe pronounced DNA-binding footprints for full-length REST, N-terminal REST, C-terminal REST and to a lesser degree for the DNA-binding domain (GpCs 5,6). We therefore can determine the second REST motif to be occupied, while in the absence of GpCs no conclusions can be drawn for the first motif. A potentially positioned nucleosome can be detected upstream (GpCs 13,14).

3.2.9 The Kinetics of REST Induced Hypomethylation are Slow at a Majority of Cytosines

Up to this point, I have focused on describing REST associated chromatin changes in equilibrium state. I have identified nucleosome remodeling, TET mediated demethylation and likely steric hindrance over REST motifs as putative molecular intermediates to a hypomethylated state. Additional mechanistic insights could be retrieved by studying the transition of a methylated to hypomethylated state over time.

I decided to transiently transfect full-length REST protein into *Rest ko* cells and assay methylation by amplicon Bis-seq at 8h, 24h and 48 h time intervals (Figure 3.19A). I find no significant methylation changes in control regions, but detect demethylation for cytosines within REST binding sites. On average, hypomethylation is slow across the population. The absolute speed of demethylation decreases over time ($t_{0\text{-}8h}$: 0.25%/h, $t_{0\text{-}24h}$: 0.20%/h, $t_{0\text{-}48h}$: 0.11%/h), in line with an exponential decay.

To resolve potential different kinetic subgroups, I performed hierarchical clustering on the induced methylation changes compared to Rest ko cells (Figure 3.19B). Kinetics are dominated by cytosines that show slow demethylation (cluster II). The second largest cluster (cluster I) demethylates faster, but not with similar speed as the fastest and smallest clusters (cluster IV, cluster V). Surprisingly, I detect hypermethylation for 22 REST cytosines. Hypermethylation is mostly limited to the first eight hours after transfection.

I now compared average methylation kinetics within clusters to two different theoretical scenarios of demethylation (Figure 3.20A). First, I consider passive dilution of the methyl group due to complete inhibition of all DNMTs (equivalent to *Dnmt* triple ko, decay constant 0.50). Second, I present REST mediated, complete inhibition of both *de novo* methyltransferases (equivalent to DNMT3A/B ko, decay constant 0.98). The latter is considered to mainly reflect DNMT1 maintenance fidelity (Jackson et al., 2004). Methylation kinetics faster than complete passive dilution would be obligatory subjects to replication-independent active demethylation. For all theoretical references, I assume standard replication times of mouse embryonic stem cells (t= 16h).

According to these assumptions, all observed kinetics could be in line with a replication-coupled loss of methylation. Average demethylation kinetics within the

main cluster II could be attributed to DNMT3A and DNMT3B inhibition alone. Across all clusters, a substantial fraction of cytosines shows methylation decays that are similar or slower than theoretical decays by DNMT3A/B inhibition ($t_{0\text{-}8h}$: 49%, $t_{0\text{-}24h}$: 38%, $t_{0\text{-}48h}$: 31%). However demethylation within clusters I, IV and V is faster than could be explained through *de novo* DNMT inhibition alone. Especially clusters IV and V contain cytosines that show demethylation kinetics just in line with passive dilution. Though considering that time point zero is the time of transfection and expression initiates several hours later, I cannot conclusively determine these cytosine kinetics to be replication dependent. It would require further experimentation to assign cluster IV and V cytosines to an active or passive demethylation process. Nevertheless, the vast majority of cytosines show kinetics that could be explained by passive dilution of the mark during replication.

I aimed to expand my knowledge by integrating important REST binding site criteria (Chapter 3.2.5, Figure 3.12,). I explored differences between clusters regarding: *Rest ko* methylation levels, REST motif distance and REST ChIP-seq enrichment (Figure 3.20B). There are pronounced differences between the hypermethylated and fast hypomethylated cluster. On average, hypermethylated cytosines have a lower starting methylation in *Rest ko*, are further away from the REST motif and show lower REST ChIP-seq enrichment. The fast hypomethylation cluster shows higher starting methylation and higher REST ChIP-seq enrichment than all other clusters. While I acknowledge the limited cluster sizes, the described differences are in agreement with earlier observations in mESCs (Chapter 3.2.5).

In summary, I have determined REST mediated hypomethylation to be primarily slow. Demethylation for the majority of cytosines is compatible with passive dilution of the modification during replication. REST mediated inhibition of DNMT3A and 3B alone could explain a large fraction of the observed demethylation kinetics. Remaining cytosine demethylation should be attributed to additional DNMT1 inhibition or putative TET activities. The slow demethylation kinetics are in agreement with earlier observations that showed an overall limited methylation effect of TET deficiency (Chapter 3.2.7). Nucleosomal positioning effects could in principle induce slow demethylation through DNMT inhibition (Chapter 3.2.8).

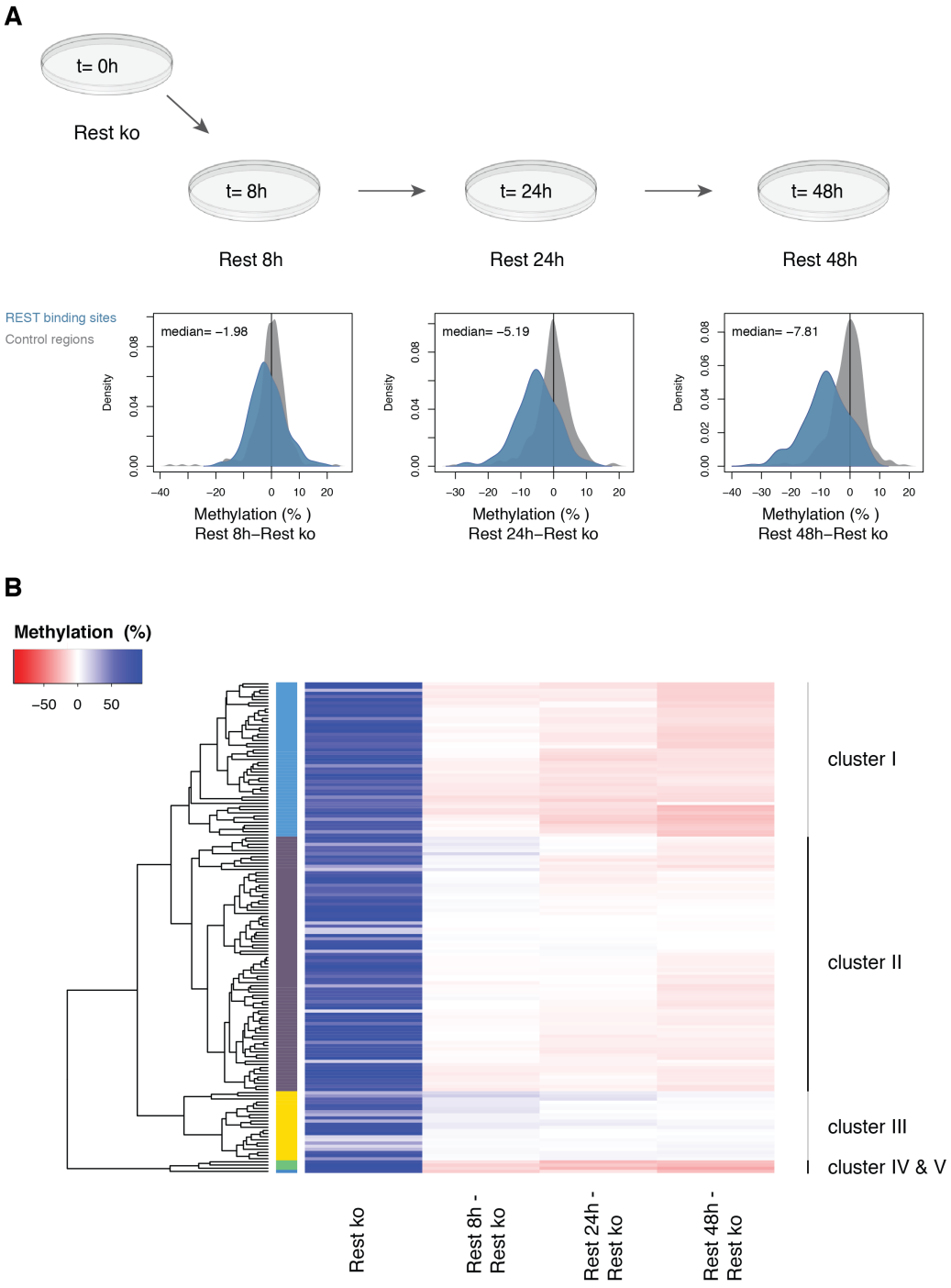

Figure 3.19 I Kinetics of REST associated methylation changes. A I transiently transfected full-length REST protein into *Rest ko* cells and assayed methylation by amplicon Bis-seq at three different time points post transfection. Density plots describe average methylation changes across control regions (grey) and REST binding sites (blue). I observe slow demethylation kinetics for the majority of cytosines. **B** The heatmap summarizes the methylation data for individual cytosines within REST binding sites. Hierarchical clustering was performed on methylation deltas respective to *Rest ko* cells. The heatmap indicates the degree of hypomethylation (red) and hypermethylation (blue) for each individual cytosine (rows) over time (columns 2 – 4). In addition, I included the starting methylation levels in *Rest ko* cells.

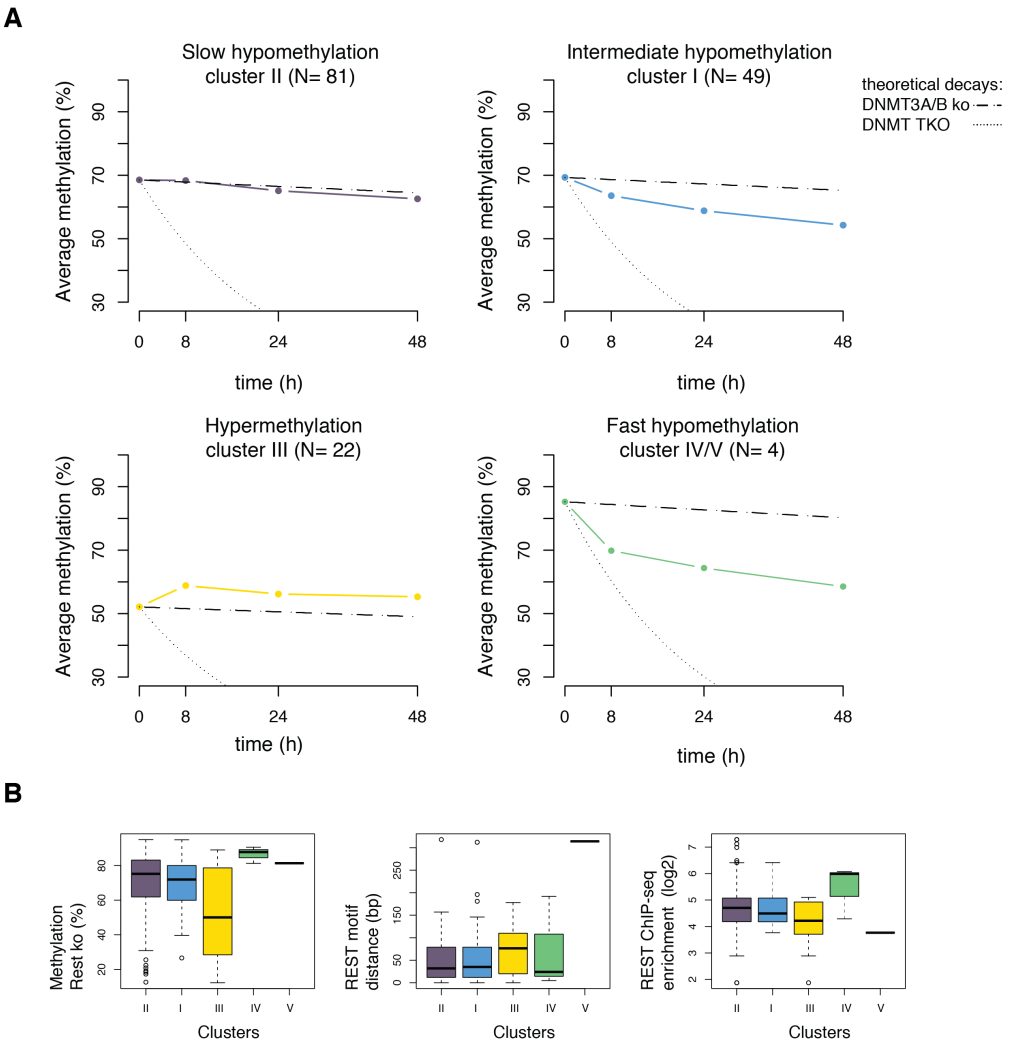

Figure 3.20 | Characteristics of different kinetic demethylation subclasses. A The line plot shows average methylation kinetics within previously described kinetic clusters (same color code). Dotted lines contrast observed demethylation rates with possible, theoretical decays mediated by DNMT3A/B or DNMT3A/B plus DNMT1 inhibition. Theoretical decay rates assume cell cycle length of 16h, DNMT1 maintenance fidelity of 98% and absence of active demethylation. **B** Characterization of the different kinetic clusters regarding methylation in *Rest ko*, REST motif distance and REST ChIP-seq enrichment in *Rest wt* cells.

3.2.10 Methylation Levels of REST Binding Sites Can be Predicted from Binding of REST Interactors

Over the last sections, I have seen several molecular components and chromatin changes that are associated with the methylation state of REST binding sites. To see whether chromatin features are also predictive of methylation, I quantitatively modeled methylation levels of REST binding sites. To that effect, I have built a random forest model predicting methylation in 100 bp windows over distal, bound REST motifs (Dong et al., 2012; Breiman et al., 1998). I restricted my predictions to this category, as promoter structures are inherently more complex and their methylation is less dependent on DNA-binding factors (Krebs et al., 2014; Lienert et al., 2011a). I use several previously described predictors: REST interactor ChIP-seq, nucleosome positioning and genomic location features.

With the random forest model, I was able to predict methylation within REST binding sites relatively well (r= 0.789) (Figure 3.21A). However, the model seems to systematically underestimate methylation in the intermediate to higher range. It is important to note, that the model's performance is significantly worse when using REST ChIP-seq as the only predictor (r= 0.541). REST interactors like SIN3A, SIN3B and TET1 are particularly important variables within the model next to the REST binding signal (Figure 3.21B).

I therefore find evidence that integration of REST interactors improves my current predictions of endogenous methylation levels within distal REST binding sites. This observation strengthens the validity of my previous conclusions drawn from the ectopic re-expression experiments.

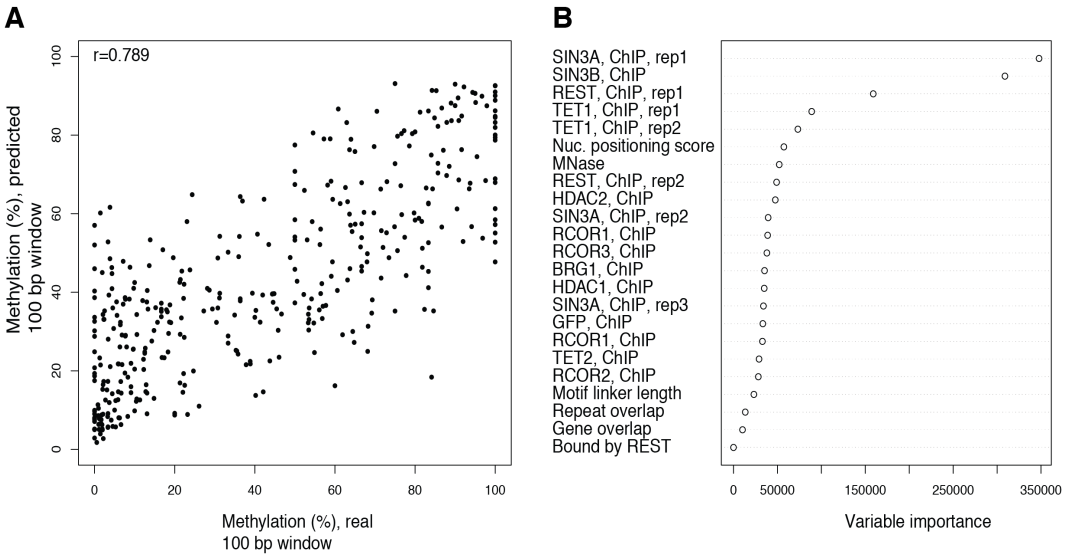

Figure 3.21 | Random Forest prediction of methylation levels within REST binding sites. A Methylation averages within 100bp windows over bound, distal REST motifs were predicted using a random forest model. The model predicts average methylation levels well, but shows systematic underprediction in higher methylation ranges. **B** The plot shows the different quantitative and categorical variables used and their importance for model performance. REST interactors like SIN3A, SIN3B and TET1 are more important variables within the model.

Chapter 4
Discussion and Outlook

Based on its methylation patterns, the mammalian genome can be partitioned into characteristic functional regions (Stadler et al., 2011). The aim of this project was to study this relationship between transcription factor binding and reduced DNA methylation. In detail, I aimed to denominate molecular components required for transcription factor induced hypomethylation. The transcriptional repressor REST served as a paradigm DNA-binding factor in mouse embryonic stem cells. I characterized several REST mutants in their ability to induce local hypomethylation. Information about chromatin accessibility, RNA expression and different chromatin modifications was accordingly integrated.

4.1 REST Binding Sites are Hypomethylated and Show a Distinct Chromatin Landscape

Many initial hypotheses regarding REST induced hypomethylation were generated based on steady state chromatin signatures. REST binding sites frequently overlap with regions of reduced methylation, and low-methylated regions (LMRs) in particular (Stadler et al., 2011). When comparing regions around REST motifs, on average only REST occupied regions show reduced methylation (Figure 3.7C). While these observations indicate a connection between REST binding and methylation at binding sites, they do not establish a cause or consequence relationship. However, I presented three lines of evidence that suggest DNA methylation downstream of REST binding.

First, the degree of hypomethylation over REST binding sites scales with its binding (Figure 3.7A). Regions highly occupied by REST show significantly less methylation than weak binding sites. Second, REST expression proved necessary and sufficient for reduced methylation at the majority of tested distal REST binding sites (Figure 3.5B). In the absence of REST, binding sites became methylated in *Rest ko* cells. In turn, re-expression of full-length REST protein into

Rest ko cells re-established significant hypomethylation. The third argument supporting a REST induced hypomethylated state comes from REST motif experiments. Here, we generated chimeric constructs consisting of random *E.coli* DNA and a REST motif (Figure 3.6A). After targeted insertion into the mouse genome, we observed strong hypomethylation over the REST motif (Figure 3.6B). In contrast, inserted randomized controls became fully methylated. The REST motif is thus the minimal *cis*-element sufficient to introduce hypomethylation within a non-murine DNA sequence. All these experiments further substantiated evidence, that local hypomethylation is a consequence of REST activity (Stadler et al., 2011).

It is important to note, that I focused my analysis on distal REST binding sites, since promoters posses a more complex binding landscape. Multiple proteins simultaneously bind to promoters and additional properties like CpG density contribute to their unmethylated state (Krebs et al., 2014; Lienert et al., 2011a). I therefore speculate, that methylation within proximal genomic regions is likely less dependent on REST.

Next, I explored potential mechanisms by describing REST's binding and chromatin landscape in more detail. I first investigated the relationship between DNA methylation and distance to the REST motif. Average methylation increases with distance from the REST motif, however not linearly but with a prominent periodicity (Figure 3.7B). There is an approximate 180 bp period, which matches the allocated nucleosome repeat length in mouse embryonic stem cells (Teif et al., 2012). Nucleosomal positions and methylation averages are shifted in phase, i.e. methylation levels of linker associated DNA are higher than nucleosomal. It is important to note that this effect is observable across the range of about five positioned nucleosomes in either direction of the REST motif. Absence of nucleosome phasing is observed over non-bound REST motifs and motifs in *Rest ko* cells (Figure 3.18A). In one possible model, nucleosome reorganization could link REST binding to DNA hypomethylation. I further discuss this hypothesis in the following sections.

Next, we compared chromatin states of REST bound motifs to unbound motifs, low-methylated regions (LMRs) and unmethylated regions (UMRs). This analysis supported genomic binding of several known REST interactors to REST binding sites (Figure 3.7C). We detected increased binding of both direct interactors SIN3

and RCOR and all of their paralogues. In line with previous reports, we found no evidence for selective co-occurrence of either SIN3 complex, CoREST complex or their specific paralogues (Figure 3.15A, *data not shown*) (Yu et al., 2011). The potential redundancy of both repressor complexes and their paralogues remains an interesting aspect to be addressed in the future. We observed additional enrichment for SIN3A and SIN3B in REST non-bound UMRs and LMRs, indicating a frequent SIN3 recruitment by other DNA-binding factors or chromatin signals. Furthermore, the indirect interacting proteins HDAC1, HDAC2 and TET1 and TET2 were enriched over REST binding sites. Both HDAC enzymes can be recruited by the SIN3 and CoREST complex, while TET 1 and 2 interactions have been established for SIN3A (Cartron et al., 2013; McDonel et al., 2012; Williams et al., 2011a).

The chromatin signature of REST binding sites is in concordance with occurrence of the above-mentioned enzymes. REST binding sites show enrichment of the TET mediated oxidation product 5hmC, as has been reported before (Figure 3.7C) (Feldmann et al., 2013). I would argue that at least part of this modification could be indirectly mediated through SIN3A. REST dependent 5hmC recruitment has so far been shown on the level of a few binding sites (Feldmann et al., 2013). We have identified no other chromatin modifications that allow an immediate link to DNA methylation. We particularly focused our attention on a potential H3K27me3 enrichment over REST binding sites, but could not substantiate such (Figure 3.7C). Methylation states over REST binding sites are thereby independent of H3K27me3.

REST binding sites revealed a clear depletion of chromatin modifications associated with transcriptional activity. Especially proximal REST binding sites showed decreased H3K27ac, likely reflecting enzymatic activity of co-bound HDACs. REST occupied promoters showed a strong reduction of the active promoter mark H3K4me3 as compared to average UMRs. In contrast, REST associated H3K27ac and H3K4me states were less pronounced for distal REST binding sites. The chromatin states of REST occupied promoters were in great accordance with REST's described function as a transcriptional repressor (Chong et al., 1995; Schoenherr and Anderson, 1995). Further experimentation would be required to delineate transcriptional effects of distal REST binding.

In sum, REST binding sites are characterized by a distinct chromatin landscape that features strong nucleosome positioning, lack of active histone modifications and presence of active demethylation intermediates.

4.2 Molecular Features of REST That Are Associated with its Demethylating Activity

I next investigated molecular components that are required for REST induced binding site hypomethylation. Three different REST truncations were re-expressed into REST deficient cells and methylation states profiled over REST binding sites. The expression of specific REST domains was guided by the well-defined protein domain annotation (Figure 1.2). Subsequent genomics assays were largely motivated by previous observations on REST co-occurring 3chromatin features (Figure 3.7C).

REST's DNA-binding domain possesses several unique features. First and foremost, every demethylating DNA-binding factor is required to bind to initially methylated genomic sites. I presented evidence that REST binding is methylation insensitive. DNase-I signal over REST motifs did not increase in complete absence of DNA methylation (Figure 3.8). REST thereby showed the expected binding behavior of a methylation insensitive factor. I note that, REST's consensus motif does not contain strongly conserved CpG dinucleotides. In general, transcription factors recognizing DNA motifs devoid of high-content CpGs are more likely methylation-insensitive (Blattler and Farnham, 2013).

Second, I have shown that REST's DNA-binding domain is able to bind inaccessible chromatin. Even in the presence of non-positioned nucleosomes, all REST mutants were able to bind a high proportion of their target sites in *Rest ko* cells (Figure 3.9C, Figure 3.9D). It could be possible that REST's DNA-binding domain recognizes partial motifs on the nucleosomal surface. In theory, REST could selectively employ a subset of its zinc finger domains to scan nucleosomes for degenerate motifs. Similar mechanisms have been reported for other pioneering factors (Soufi et al., 2015). The ability to bind target motifs in the context of nucleosomes is thereby a factor-specific ability.

Third, REST's DNA-binding domain is unique in its recognition of a long high-content DNA motif (Chong et al., 1995; Mathelier et al., 2016; Schoenherr and Anderson, 1995). The REST motif length exceeds the mammalian average of 6 – 10 bp by far (Mathelier et al., 2016; Stewart et al., 2012; Wunderlich and Mirny, 2009). Sequence-specific DNA affinities are also a function of the number of recognized DNA nucleotides. The REST motif quality was highly predictive of

REST mutant binding behavior, where the highest scoring REST motifs were all occupied (Figure 3.11). This relationship is far more complex for most other DNA-binding factors, where *in vivo* binding predictions based on motif scores are notoriously difficult (Arvey et al., 2012).

Residence times of DNA-binding domains could also be impacted by motif quality. Although REST's residence time has not been determined, CTCF possesses a long residence time with a similar motif architecture (Sherwood et al., 2014).

It remains speculative, whether the ability to remodel chromatin and induce hypomethylation is a function of all those mentioned qualities inherent to REST's DNA-binding domain. Further experimentation would be needed to study the potential impact of DNA affinity and residence time. In this regard, synthetic DNA-binding factors could address this question. Varying lengths and affinities of TALE Repeat Variable Diresidues (RVDs) could probe the effect of different DNA affinities and residence times on methylation states (Cuculis et al., 2015; Miller et al., 2015).

Within the scope of this project however, I maintained the same DNA-binding domain throughout all experiments. Instead, I evaluated the role of REST's interaction domains in more detail. I contrasted three REST mutants for their ability to induce hypomethylation over REST motifs (Figure 3.9B). One mutant expressed REST's DNA-binding domain (DBD), whereas the remaining two contained an additional repressor domain (NTE-DBD and DBD-CTE).

These REST mutant experiments revealed that substantial, broad hypomethylation requires either REST's N-or C-terminal interaction domains (Figure 3.12B). REST's DNA-binding domain alone showed no significant hypomethylation, with the exception of two motif associated CpGs (Figure 3.13C). In contrast, the two other REST mutants showed varying degrees of hypomethylation over binding sites (Figure 3.13A). The C-terminal REST mutant DBD-CTE showed very similar binding behavior as compared to full-length REST protein (Figure 3.9D). However, binding site associated hypomethylation and nucleosome positioning was less pronounced as in the N-terminal mutant. Further mechanistic delineation of this REST mutant would require extended methylation and chromatin accessibility data.

In contrast, the N-terminal mutant NTE-DBD resembled full-length REST protein in regards to many chromatin characteristics. Binding of N-terminal REST caused strong hypomethylation over the majority of binding sites, but binding was less strong compared to C-terminal REST (Figure 3.14A, Figure 3.9D). The only annotated protein interaction domains within the N-terminal protein are dedicated to the SIN3 paralogues, SIN3A and SIN3B. The SIN3 proteins themselves do not contain enzymatic activity but function as recruitment platforms for several chromatin-modifying enzymes (Pazin and Kadonaga, 1997). Most prominently, SIN3 mediates transcriptional repression by recruiting the histone deacetylases HDAC I and II (Silverstein and Ekwall, 2005). The SIN3A protein has been reported to show striking co-occurrence with the TET1 enzyme in mouse embryonic stem cells (Williams et al., 2011a). Biochemical assays revealed interactions between SIN3A and both TET1 and TET2 (McDonel et al., 2012). In a likely scenario, REST associated hypomethylation could be mediated through its SIN3A and TET interactions. I presented evidence for a partial demethylation by TET enzymes. In the absence of all TET paralogues, REST binding sites show pronounced hypermethylation in the immediate vicinity of REST motifs (Figure 3.16). This localized hypermethylation is in accordance with the expected space of REST associated SIN3A binding. The TET1 CXXC domain could additionally confine TET1 binding to this linker region. Structural studies have indicated, that CXXC domains generally require access to the major and minor groove of DNA and would therefore be restricted to accessible linker DNA (Cierpicki et al., 2010; Long et al., 2013b; Zhou et al., 2012). I see an overall spatial alignment of expected SIN3A associated TET1/2 binding and TET dependent hypomethylation. I can however not exclude an indirect, SIN3 independent recruitment of TET.

While TET recruitment has been shown for other DNA-binding factors before, I note that the adaptor protein SIN3A could allocate demethylating activity to many of its interacting DNA-binding factors (Guilhamon et al., 2013; Rampal et al., 2014; Wang et al., 2015). Further experimentation is needed to unambiguously prove these REST/SIN3A/TET dependencies. First, it would be necessary to show that hypermethylation in TET1/2/3 deficient stem cells is dependent on TET's catalytic activity. Secondly, the question remains whether the NTE-DBD hypomethylation effects are dependent on SIN3A and SIN3B. Methylation

profiling of the described NTE-DBD SIN3 interaction mutant should be informative (Figure 3.15C). In extension, methylome analysis of SIN3A/B loss of function stem cells would allow global statements on the SIN3 and methylation relationship, but could prove difficult in light of reported cellular lethality upon SIN3A and SIN3B loss of function (McDonel et al., 2012).

Aside from its methylation phenotypes, N-terminal REST (NTE-DBD) possessed strong nucleosome positioning abilities. Nucleosomes around REST motifs were phased similarly as in wild type cells (Figure 3.18A). The two other mutants, DBD and DBD-CTE, did not show strong nucleosomal phasing. In light of the strong binding behavior of DBD-CTE, this observation would suggest that protein binding is not sufficient for effective nucleosome positioning. It is therefore possible that NTE-DBD mediates nucleosome positioning through an unknown chromatin remodeling enzyme. Overall, I find the co-occurrence of nucleosome positioning and CpG hypomethylation a striking observation. The antiphasing of methylation and MNase-seq signal in wild type cells further hints towards a connection between both chromatin states. I acknowledge, that nucleosome reorganization could be a possible mechanism to spread transcription factor associated hypomethylation over larger regions than the mere motif. Indeed, the average size of REST-LMRs in the mouse methylome extends well beyond the direct REST binding interface (Stadler et al., 2011).

I have presented evidence that binding site-associated hypomethylation is not a mere function of protein binding. In fact, broadly reduced methylation of binding sites was only observed upon expression of proteins containing interaction domains. The latter likely reflects the role of active cofactors in methylation regulation. Quantitative models predicting methylation states of REST binding sites in fact improve after incorporation of REST's interactor binding signal (Figure 3.21).

In summary, I have thereby described demethylating DNA-binding factors to have defined characteristics. REST possesses a methylation-insensitive DNA-binding domain that recognizes a large, high content motif and binds within regions of not positioned nucleosomes. The presence of interaction domains was crucial to induce a broad demethylation. This likely reflects the necessity of active cofactor recruitment. However, I observed strong hypomethylation effects only in relation with induced nucleosome phasing. I have therefore described a DNA-binding

factor that is mostly ignorant to the chromatin environment of its binding sites. Upon binding, the factor induces substantial chromatin restructuring of its binding sites. Many of these described characteristics are commonly attributed to pioneering transcription factors. It is likely, that the repertoire of mammalian DNA-binding factors can be separated based on their demethylating activities, one class that demethylates and the other being methylation-sensitive. These two classes likely reflect different hierarchical positions within the transcription factor repertoire.

4.3 Molecular Dissection of Methylation within REST Binding Sites

Following above stated observations, I will now dissect methylation states within binding sites into three distinct subregions. Based on relative location to the REST motif I distinguish between the following: the REST motif, direct vicinity of the REST motif (~50 bp ± relative to motif end/start) and the remaining hypomethylated region (Figure 4.1). I will present unique characteristics inherent to either class.

First, REST motif associated methylation is minimal in wild type embryonic stem cells (Figure 3.7B). Similarly, an artificial DNA sequence gets strongly demethylated directly over the inserted motif (Figure 3.6B). Likewise, all REST mutants that induced local hypomethylation showed strongest demethylation over REST motifs (Figure 3.12B, Figure 3.13). The DNA-binding domain alone induced reproducible demethylation over two cytosines within REST motifs. Deficiency of all three TET enzymes does not impact the unmethylated state of REST motifs in REST wild type cells (Figure 3.16). The motif proximal region is distinct from regular linker DNA, where methylation is usually higher than nucleosomal one. Together this collected evidence hints towards a possible inhibition of DNMT1 and/or DNMT3A/B during or after replication. This inhibition could be due to interference with methyltransferase binding. Residence times, DNA affinities and binding dynamics directly after the replication fork would be potential properties likely to affect the extent of this inhibition (Chapter 4.2). Possible DNMT1 inhibition would implicate the necessity of the respective protein to bind faster than UHRF1 and DNMT1 mediated remethylation. Studies in *Drosophila* indicate that not many proteins possess this characteristic (Ramachandran and Henikoff, 2016). At this point, I cannot comment on the specific kinetics of REST binding after replication. On the other hand, strong inhibition of both *de novo* methyltransferases could on its own elicit the full extent of observed hypomethylation. A more comprehensive description of the motif demethylation kinetics would allow to distinguish a DNMT3A/B inhibition alone from the alternative scenarios.

Second, I find the immediate vicinity of the REST motif to be functionally distinct. This region is nucleosome depleted in *Rest wild type*, *Rest*$^{-/-}$*REST* and *Rest*$^{-/-}$

NTE-DBD cells. In contrast to the motif, its vicinity shows a strong dependency on TET enzymes. In the absence of all three TET enzymes, a window of approximately 50 bp upstream and downstream of the motif borders becomes strongly hypermethylated (Figure 3.16). It remains to be seen whether this localized hypermethylation is dependent on TET's catalytic activity. I can however reconcile the localized TET-dependent methylation effect with the expected binding of its interacting partner SIN3A (McDonel et al., 2012; Williams et al., 2011a). It is surprising to see that this relatively small region shows such a unique TET response. It is possible that TET activity ensures methylation-sensitive cofactor or effector binding. Alternatively, REST's cofactors could specifically bind oxidized forms of 5mC that are catalyzed by TETs. A recent study showed at least RCOR2's preferential binding to 5caC (Spruijt et al., 2013).
Third, I characterized the hypomethylated region that starts around the first upstream and downstream positioned nucleosomes. As mentioned earlier, linker DNA within this region shows higher methylation levels than nucleosomal. I would suggest that this local methylation difference could be a consequence of *de novo* methyltransferase targeting. In detail, DNMT1 mediated maintenance is known to be imperfect, while DNMT3A and DNMT3B ensure sufficient remethylation (Jackson et al., 2004). Structural studies indicated that DNMT3A anchors to nucleosomes by its ADD domain interaction with the unmethylated histone H3 tail (Guo et al., 2015). In the presence of stochastically bound nucleosomes, eventually most unmethylated CpGs should undergo remethylation by DNMT3A and DNMT3B. However in the case of well-positioned nucleosomes, cytosines associated with nucleosomes would be less subjected to DNMT3A and DNMT3B remethylation. Consequently, regions with strongly positioned nucleosomes would undergo a slow decrease in methylation over time. Indeed, I detected many cytosines with slow demethylation in my kinetic experiments (Figure 3.19B, Figure 3.20A, cluster II). However, others showed kinetics that would require more than DNMT3A/B inhibition alone. Additional inhibition of DNMT1 maintenance activity would explain why linker DNA levels are still lower than the genomic average methylation. While these statements are compatible with a recent publication on DNTM3 targeting in mouse embryonic stem cells, the effect of nucleosome positioning on DNA methylation is not known (Baubec et al., 2015). Given the current data, I suggest that nucleosome positioning is sufficient

to induce hypomethylation over a broad region. It would be interesting to test DNA-binding domain/chromatin remodeler fusions for their ability to reduce methylation.

The presented kinetic data is largely coherent with the drawn conclusions. On average, cytosines within stronger REST binding sites showed faster demethylation than cytosines within weaker sites at larger distances from the motif. The majority of cytosines show demethylation kinetics in line with replication-coupled demethylation. Based on kinetics, a small number of cytosines could potentially undergo fast active demethylation. We plan to corroborate these ambiguities by larger kinetic methylation data sets. Further experiments should shed light on an active demethylation contribution within REST binding sites. Re-expression of REST in postmitotic REST deficient cells could be similarly informative as re-expression of REST mutants into *Tet* TKO cells.

Overall, I have presented evidence that DNA hypomethylation within REST binding sites can be further dissected into distinct local phenomena (Figure 4.1). The steady state methylation levels of REST binding sites seem to be the effect of multiple chromatin changes. It will be interesting to see, whether all demethylating transcription factors always employ all these mechanisms at the same time. Hence, further descriptions of DNA-binding factor induced hypomethylation would be insightful.

4.4 Functional Implications of REST Associated Hypomethylation

While dissecting mechanistic aspects of REST mediated methylation changes, I have not yet addressed any functional implications. Up to this point it remains a prominent unknown, what the biological effects of hypomethylation within distal regions might be. I will briefly sketch the possible scenarios that could entail a functional role of distal hypomethylation, or a lack thereof.

First, one could argue that hypomethylation within distal regions of the genome merely reflects ongoing DNA-binding events. To that effect, DNA methylation would be a purely concomitant chromatin modification that is non separable from other, functional chromatin alterations. For instance, transcriptionally instructive nucleosome repositioning could in principle always implicate downstream hypomethylation. Effectively, I am currently missing any argument that strictly defies this concept. I want to stress however that we do observe dependencies on TET enzymes, proteins whose described function is to oxidize 5mC. One could question the use of actively catalyzed 5mC oxidation, if DNA methylation were a neutral chromatin modification in the first place.

The other opposing consideration ascribes function to the hypomethylated state of distal regions. This scenario would be more likely if REST on its own was not sufficient to mediate downstream transcriptional effects. The necessity for methylation-sensitive effector binding would even require a methylation-insensitive DNA-binding factor to establish a hypomethylated binding landscape. I have not systematically investigated this possibility, but I refer back to methylation-sensitive NRF1 binding around REST binding sites (Chapter 1.5.6, Figure 1.6). Though in order to claim function for the hypomethylated state of REST binding sites, methylation-sensitive factors implicated in REST's transcriptional activity would need to be identified. Furthermore, the necessity for distal hypomethylation would impose different temporal requirements amongst different transcription factor networks. Systems with fast transcriptional input responses (e.g. DNA damage and the immune system) would require much faster demethylation than observed for the developmental factor REST. It is therefore likely that the characteristics of DNA demethylation could vary based on the specific transcription factor at hand and the kinetics of its output.

4.5 Transcriptional Effects of REST in Mouse Embryonic Stem Cells

Aside from methylation, we have characterized gene expression as a functional consequence of REST deficiency. In general, RNA-seq analysis revealed very similar global gene expression in the presence or absence of REST in mouse embryonic stem cells (Figure 3.4A). We do not observe any differences in expression signatures associated with pluripotency (Figure 3.4C). Together with the normal stem cell morphology, we therefore find no evidence for reduced pluripotency of REST deficient mouse embryonic stem cells. We cannot substantiate an earlier report in this direction (Singh et al., 2008).

The transcriptional changes in *Rest ko* cells are largely occurring at genes with nearby REST binding (Figure 3.4A). The transcriptional response at those putative REST target genes is unidirectional and positive. Transcriptional changes are mostly occurring within genes of lower expression in wild type cells. These observations are in agreement with expected primary transcriptional effects of a repressor like REST (Chong et al., 1995; Schoenherr and Anderson, 1995). On the other hand, this analysis illustrates that the mere absence of a neurogenic repressor is not sufficient to drive pluripotent cells towards the neuronal lineage. Similarly, transcriptional effects are mostly specific to REST target genes, but overall not dramatic.

Due to the difficulties of enhancer-promoter assignments, I have so far mostly described effects of proximal REST binding. A comprehensive delineation of transcriptional impacts of distal REST binding is still lacking. Improvements in enhancer-promoter interaction maps would be a prerequisite for this analysis. Ultimately, the functional role of distal hypomethylation in transcriptional regulation remains to be addressed. Moreover, it would be interesting to investigate the potential redundancy in REST's N- and C-terminal repressor recruitment. Transcriptional analysis of REST re-expression mutants could potentially even help to reveal functions of distal hypomethylation.

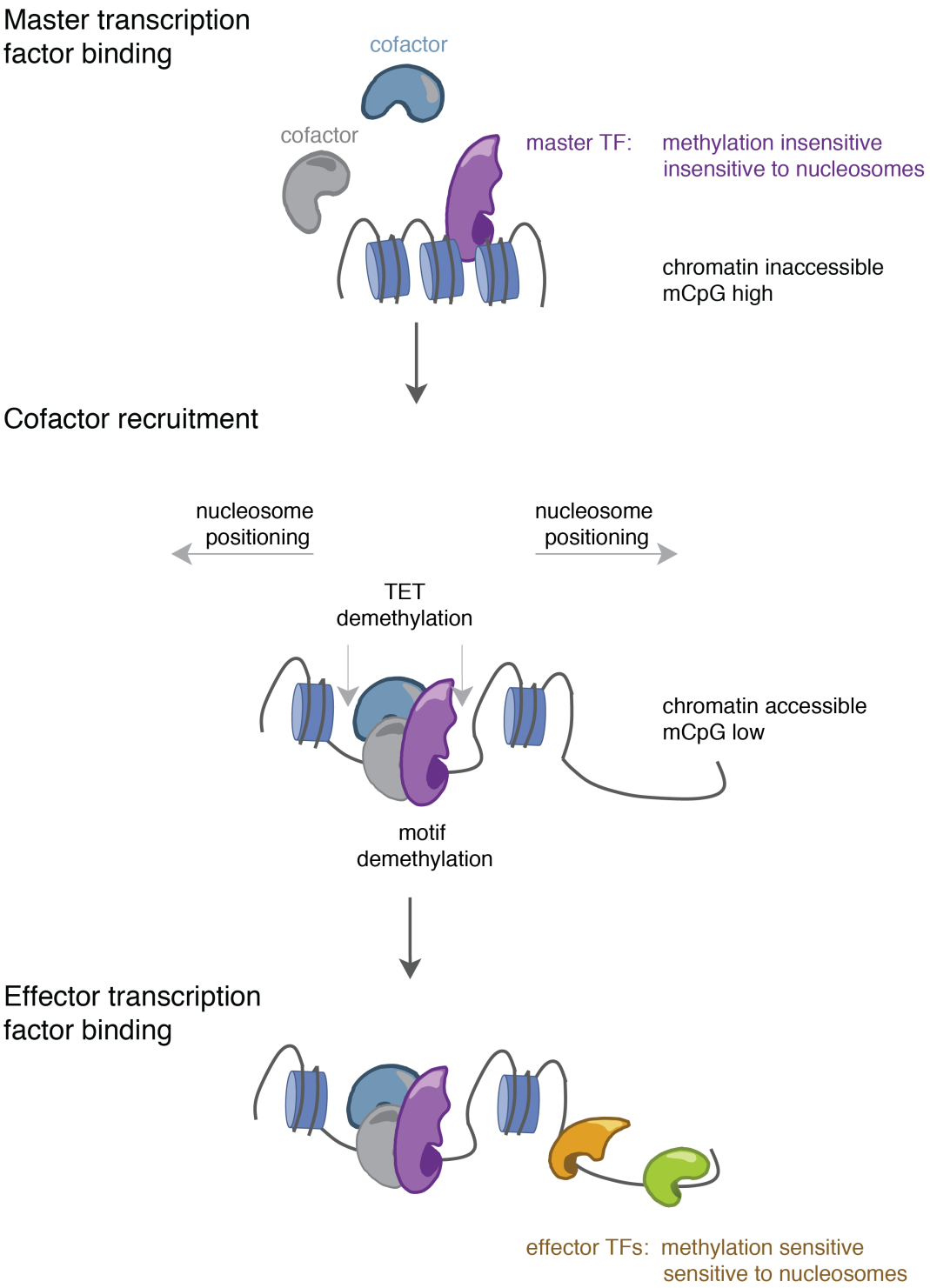

Figure 4.1 | Proposed transcription factor mediated effects on DNA methylation Certain transcription factors like REST have the ability to bind methylated, non-accessible target sites. Upon binding, cofactors are recruited, nucleosomes become positioned and binding sites undergo demethylation. The hypomethylated region adjacent to the REST motif proved to be TET dependent. This large chromatin remodeling of the binding site could enable downstream effector binding.

Chapter 5

Material and Methods

Experiments and analyses have been conducted by Juliane Schmidt, unless otherwise stated. Significant individual contributions by others are stated in italics.

5.1 Published Data Sets Used in Analyses

Sample name	GEO Accession	Experiment	Cell type	Publication
ES_159_rep1	GSM778487	RNA-seq (total)	mESC	Stadler et al., 2011
ES_159_rep2	GSM778488	RNA-seq (total)	mESC	Stadler et al., 2011
NP_159_rep1	GSM778489	RNA-seq (total)	mNP	Stadler et al., 2011
NP_159_rep2	GSM778490	RNA-seq (total)	mNP	Stadler et al., 2011
TN_159_rep1	GSM778491	RNA-seq (total)	mTN	Stadler et al., 2011
TN_159_rep2	GSM778492	RNA-seq (total)	mTN	Stadler et al., 2011
159_WG Bis-seq	GSM748786	Bis-seq (WG)	mESC	Stadler et al., 2011
159_WG Bis-seq	GSM748787	Bis-seq (WG)	mESC	Stadler et al., 2011
Tetwt_Bis-seq	GSM1372647	Bis-seq (WG)	mESC	Lu et al., 2014
TetTKO_Bis-seq	GSM1372649	Bis-seq (WG)	mESC	Lu et al., 2014
Mnase	GSM1004653	Mnase-seq	mESC	Teif et al., 2012
Dnmtwt_DNase-I_rep1	GSM1657364	DNase-I	mESC	Domcke et al., 2015
Dnmtwt_DNase-I_rep2	GSM1657365	DNase-I	mESC	Domcke et al., 2015
DnmtTKO_DNase-I_rep1	GSM1657366	DNase-I	mESC	Domcke et al., 2015
DnmtTKO_DNase-I_rep2	GSM1657367	DNase-I	mESC	Domcke et al., 2015
hMeDIP_rep1	GSM978374	hMedDIP-seq	mESC	Domcke et al., 2015
Input Chip_rep1	GSM671103	ChIP-seq	mESC	Arnold et al., 2013
Input Chip_rep2	GSM747545	ChIP-seq	mESC	Stadler et al., 2011
Input Chip_rep3	GSM747546	ChIP-seq	mESC	Stadler et al., 2011
REST_rep1	GSM671093	ChIP-seq	mESC	Arnold et al., 2013
REST_rep2	GSM671094	ChIP-seq	mESC	Arnold et al., 2013
REST_rep3	GSM671095	ChIP-seq	mESC	Arnold et al., 2013
GFP	*	ChIP-seq	mESC	Yu et al., 2011
SIN3A	GSM698700*	ChIP-seq	mESC	Yu et al., 2011
SIN3B	GSM698701*	ChIP-seq	mESC	Yu et al., 2011
RCOR1	GSM698697*	ChIP-seq	mESC	Yu et al., 2011
RCOR2	GSM698698*	ChIP-seq	mESC	Yu et al., 2011

RCOR3	GSM698699*	ChIP-seq	mESC	Yu et al., 2011
SIN3A (abcam)	GSM611196	ChIP-seq	mESC	Williams et al., 2011
SIN3A (sc)	GSM611197	ChIP-seq	mESC	Williams et al., 2011
RCOR1	GSM845235	ChIP-seq	mESC	Whyte et al., 2012
TET1	GSM659799	ChIP-seq	mESC	Wu et al., 2011
TET2	GSM1023124	ChIP-seq	mESC	Chen et al., 2013

Sample name	**GEO Accession**	**Experiment**	**Cell type**	**Publication**
HDAC1	GSM687277	ChIP-seq	mESC	Whyte et al., 2012
HDAC2	GSM687279	ChIP-seq	mESC	Whyte et al., 2012
BRG1	GSM896923	ChIP-seq	mESC	
H3K4me1_rep1	GSM747542	ChIP-seq	mESC	Stadler et al., 2011
H3K4me1_rep2	GSM594577	ChIP-seq	mESC	Creyghton et al., 2010
H3K4me2	GSM686995	ChIP-seq	mESC	Lienert et al., 2011b
H3K4me3	GSM594581	ChIP-seq	mESC	Creyghton et al., 2010
H3K27ac	GSM594578	ChIP-seq	mESC	Creyghton et al., 2010
H3K27me3_rep1	GSM686992	ChIP-seq	mESC	Lienert et al., 2011b
H3K27me3_rep2	GSM686993	ChIP-seq	mESC	Lienert et al., 2011b
H3K27me3_rep3	GSM686994	ChIP-seq	mESC	Lienert et al., 2011b

Table 5.1 I Overview of published data sets used throughout the project. * Raw data files were not deposited on Geo, but were accessed on: http://genome.crg.eu/~rjohnson/cofactor.chipseq/summary.txt, http://genome.crg.eu/~rjohnson/cofactor.chipseq/raw_chipseq_data.tar.gz

5.2 Cell Culture

I used 159 ESCs (background 129S6/SvEvTac) for benchmarking of Amplicon Bisulfite sequencing. All other experiments were performed with previously described *Rest wt* and *Rest ko* cells (Arnold et al., 2013; Jørgensen et al., 2009a) or REST mutant expressing clones. I cultured mESCs as previously described (Bibel et al., 2007; Mohn et al., 2008). In brief, cells were cultured in Dulbecco's Modified Eagle Medium (Invitrogen), supplemented with 15 % Fetal Calf Serum (Invitrogen), L-Glutamine (Gibco) and Non-essential amino acids (Gibco), beta-mercaptoethanol (Sigma) and homemade leukemia inhibitory factor (LIF). Cells were trypsinized with 0.05% Trypsin-EDTA (Gibco), neutralized with cell culture medium and pelleted prior to plating. Cell culture plates were coated with 0.2 % gelatin beforehand. Cells were passaged every second day and ESC medium was exchanged every day.

5.3 Generation of REST Re-expression Cell Lines

Murine REST cDNA was cloned into the pCDNA6-CAG-V5-MCS-IRES-BlasticidinR (promoter: CAG, N-terminal tag: V5, resistance: Blasticidin) plasmid for stable expression in *Rest ko* cells. Full-length REST contained the entire REST protein (1 – 1082 aa), NTE-DBD positions 1 – 407 aa of REST, DBD positions 154 – 407 aa of REST and DBD-CTE positions 154 – 1082 aa of REST. All proteins therefore expressed a common N-terminal V5 epitope tag. For control cells, *Rest ko* cells were transfected with the empty expression plasmid. All expression plasmids were randomly inserted into the mouse genome. In brief, 50 µg of the respective pCDNA6 plasmid were linearized with FspI (New England BioLabs), precipitated with ethanol and resuspended in TE buffer. Four million cells were electroporated with the plasmid using the Mouse ES Cell Nucleofector Kit (Amaxa). Selection of cells was started two days after transfection with 5ug/ml Blasticidin (Invivogen) and continued for 10-14 days. Individual clones were expanded and tested for expression by Western Blotting probing with a V5 antibody (R960-25, Thermo Fisher Scientific). Three clones of each re-expressed REST construct were chosen based on their highest protein expression in Western Blotting. Further experiments were carried out on three clones per construct (Amplicon Bis-seq) or one clone (ChIP-seq, NOMe-seq) per construct. Amplicon Bis-seq utilized cell lines: Rest wt (three biological replicates); Rest-/-MCS c7, c8, c9; Rest-/-REST c4, c1, c5; Rest-/- DBD c4, c7, c8; Rest-/- NTE-DBD 7c9, 7c5, 7c8 and Rest-/- DBD-CTE 5c11, 5c1, 5c9 cells (c denominates “clone” number; for simplicity reasons clone numbers are converted in main text to c1, c2, c3). ChIP-seq and NOMe-seq was performed for Rest wt, Rest-/-MCS c7, Rest-/-REST c4, Rest-/- DBD c4, Rest-/- NTE-DBD 7c9, Rest-/- DBD-CTE 5c11 cells (for simplicity reasons clone numbers are omitted in main text).

5.4 Amplicon Bisulfite Sequencing

The experimental amplicon Bisulfite sequencing approach was developed by Arnaud Krebs and Juliane Schmidt. The initial bisulfite (Table 5.2) primers were designed by Arnaud Krebs using customized Pearl and R scripts.

Bisulfite primers were designed for amplicon sizes between 200 – 400 bp and primer melting temperatures between 55 °C and 58 °C (Table **5**.2). Desalted oligonucleotides were ordered (Microsynth). Genomic DNA was isolated from cells with the QIAamp DNA Mini Kit (Qiagen). Extracted DNA was treated with 20ug/ml RNase A and incubated for 30 min at 37 °C. DNA concentration was measured on the Nanodrop ND-1000 (ND-1000 Spectrophotometer, Witec AG). Bisulfite conversion was performed with the EpiTect Bisulfite Kit (Quiagen) according to the manufacturer's protocol. In brief, 2ug of genomic DNA, 3.2 pg unmethylated lambda DNA and 3.2 pg pre-methylated T7 DNA were set up for conversion. Three conversions were pooled per sample for subsequent PCR. Bisulfite PCR was performed with 0.05U/ul AmpliTaq Gold polymerase (Thermo Fisher Scientific), 0.4 uM forward and reverse primers (Microsynth), 1 X PCR buffer, 5% DMSO, 1.5mM MgCl2 and 0.15 mM dNTPs (Thermo Fisher Scientific). PCR was performed on a 96-well format (Bio-Rad) with one primer pair per well according to described settings (Table 5.3). PCR products were pooled per sample at equal volume and purified with Agencourt AMPure XP beads (Beckman Coulter) and concentration measured with Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

Next-generation sequencing libraries were prepared from 20 ng pooled bisulfite PCR product using the Next ChIP-seq Library Prep Master Mix Set for Illumina (New England BioLabs). Up to twelve samples were multiplexed in one MiSeq sequencing run using NEBNext Multiplex Oligos for Illumina (New England BioLabs). According to the manufacturer's protocol, samples were end-repaired, dA-tailed and adaptor ligated. Ligated fragments were size selected with Agencourt AMPure XP beads (Beckman Coulter) and amplified for twelve PCR cycles using the NEBNext High-Fidelity 2X PCR Master Mix (New England BioLabs). Quality control of the amplified libraries was performed with the Agilent High Sensitivity DNA kit (Agilent technologies) on the Bioanalyzer 2100 (Agilent technologies). Libraries were quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) and pooled at equimolar ratio. High-throughput sequencing was done on the Illumina MiSeq platform. I performed 150 bp paired-end sequencing, using the MiSeq v2 reagent kit 300bp PE (Illumina). Libraries were spiked in with 10% PhiX Sequencing Control V3 (Illumina) and a 13pM library was sequenced.

Region	Forward Sequence	Reverse Sequence
lambda	TGTGTTGGTTGGAAGAGGTT	ACTATCACTCTTCTCCTCCTCT
lambda	TGTTGTTGGTTGATTTTGATGAG	TCCTCTTTCAACTCTACCACA
lambda	TTGGATGTATTGGAGAAGTATGAT	CCACCATACTAATAATCAAATCTAACA
T7	AGTGAGGGTATTGATTTTGAGT	ACCTTAAATCTATCACTCAACAAATTC
T7	GGGATGGTGAGTTTGTTGAA	CCTAATACATCTACAACTACCTCAT
T7	TGATTAGTTGAAGGATTGGAAGT	TCCCCATCAAACATAAAACCA
FMR	AGTTGTTAGGATTTGAATTTTGGT	CCTCTACTCCTTCTTTCCTAATACA
FMR	AGGATGGATGTGTTATGTTTTAGT	AAATCTACCTTTCCTTCCAACA
FMR	TTTTATTGATTGTTATGTGGTGTTTT	ACAACTCCTCTTCTCCAACA
FMR	AGATGTTTGTTTAGTTTTGGGTT	CCAAAACCCTAACAATCCCC
FMR	GGGAGGTAGGGGTAGTAAGA	ACACACACACACACACAATA
FMR	AGAGATTGGTGGGTTGGATT	ACCACCACAAAACAAATACCT
FMR	GGGAAGTTGAGGTAGGTAGG	CAACCAACCAACCAAAACCT
FMR	GGGTTTTGTAGGGTGTGAGA	CCACTACCACATCACAATTCC
FMR	AGTGGAGTGGTGTAGAGGAT	ACCTTAAACCTCTCTCAAAACA
UMR	GGTTTTGATGGTTGAGGTGT	TTCCCAATCCCCATTTCTCC
UMR	AGGATTGTTTGGGATGGAAAA	CAAACTCAACCCAACCAACC
UMR	AGTTAAAGAATGAAATTGAAGTTTGAA	TCCTCTTCATTTTCCCCATCT
UMR	AGGGATTAGTAGGAAAGGAGTT	CACCTTCCACCCCTCTATTA
UMR	AGGTATGAGAGTTAGAAATTAAGAGG	AACAACTATACCCACAAATCTCT
UMR	GTTTTGGTATTTAAGAAAGGTTAGGG	AATTCCCCAACCATTCACCT
UMR	GGGATTGTTGGGAGGGATAG	CCAAAACAACCAAAACTACACA
UMR	AGGAGTTAATGAGGGAGAATAAGA	AAACCCCTCCTCCAAAACTC
UMR	AGTTTTGGTTAATGAAGTAGGAGA	CCCTCATTCCTAACCCCAAT
UMR	TGGAGGGGAAAAGGGAAAAT	ACACAACAACTACATCAACTAAACT
UMR	AAGGTTTTGAGGTAATTGAGTGA	TCCCATCTATCTCCTCCACC
MUMR	GTTGTTAGGGTTAGGTTTTGATT	TCCTATTACTCCCAACAATACCA
MUMR	TGGGGTAGAAAAGTTGTTTAGT	ACCACCAAACATAACACACA
MUMR	TGTGGGAAAGGTAGTAATAAAATAGA	AACCAACAAACTATCTCATACCA
DMR	GATTTGGTGGTTGGGAGTTG	AAACTAAACAAACCACCTCAAAA
DMR	GGTTTTAGAAAGTTGTTTTATTTTGGG	TTCACATCAAAACAACACCTCA
DMR	AGATGGTGATAGGGGAGAAAA	TCACCCAAATTCAATACCTCAA
ES	GGTGGAGGTGGTTTAAAGGT	TACCCAAAACCACCCTAACC
ES	TTTAAGATAAGTTGTTGTTGGGTT	TCCTAACCAAAATCCTAAATACCT
ES	TGGGATTTGAGATTGTATTAGTTAGG	CAAAACAAATCCCTATCCTCTAAC
ES	GGGGTTGGGTAATAGATGGT	ACAATCACACATCAAACCCT
ES	TTGTTATTAAGTTGGAGTGGGT	CCATCCACTTATCTCCCACA
ES	GGTAGAGTGTTTTAGTTAAATTAAGGG	AACTACCATCCATCCACTCC
ES	AGTGTTTAGGTGTATATTAGGAGGT	ACAAAACCCTACCTACTCCT
ES	AGGGAGATGATAGATTAGGTGAT	ACCTTCCACTATCCCTACTCA
ES	ATTATGTGAGTTAAGATGGGTGT	ACATAAACTTACTTAACCTTATACCCA
NP	GGGAGGTAGAGTTGGATTAGTAAA	ACTCCCTATTACCAACTACAATTT
NP	TGAGTGGTTTTGTTTGTGAGG	ACTCCCAAACTTTCTTCTATCAC
NP	TGTTGGAAGTTGATATATTGTAGTTGA	ACCTCAAACTCAACTCACACT
NP	TGGTTTTAGTTTAAGAAAAGGAAAGT	TTCACTTCATTTACTCCTCTCTT
NP	TTGGAGGGAGTAGGGGAG	ACCCAATCAACAATATTACATATCCA
NP	GGTTGAGTTTAAATAGAGGTTAGGG	ACTTCTATTTCCACTAAACCTACA
NP	AGAGTAAGGTTTTGAGGTGAGT	CCCAACCTCTTAACTTCCCA
NP	TGATGGGAGAGAAAGAGTGAG	ACTCTCCAATTCATTTAATAAAACTCT
NP	AGGTTATTTTAGAGGTTTGTTAGGT	ACATCACAAACCCTTTTCAAAA
NP	TGTGAGAAGGTAAGAGGTGTG	CCTATCAAACTAACCAACTACCT
NP	AGGTAGTTGGTTAGTTTGATAGG	ACAAAACAAAACAATACCAACCA

NP	TGGTTGAGTAATGAGATAGGTTT	ACCCCATAATTATCTCAAATCTCA
CTCF	TGTTTTGGTATGAAAGTTTTGGT	CCTCAACCTAACCTAAACCCA
CTCF	AGTTTTGTTTTGTATTTGGTTGTTAA	ACTCAATCATTTCCATTCCAAAA
CTCF	TGGGGAGGGATGTGGTATAA	ACTTCACTTCCACCTAAAACTT
CTCF	TGGGAGAGGAAGTGTGTTTT	ATCAACAACCACCTCCAAAA
CTCF	AAGGTAAGTTTGATTTAGAGAATTGA	ACCACTATCCAAACCCAAACT
CTCF	AATAGTAGAGGTGGATTTGATTATAGA	CAAACCACACTAAACCTCACA
CTCF	GGTTATGTTATTGTAGTGAGTGGT	ACTCTCAACAACCAATACTCCA
CTCF	GGTTATGTTATTGTAGTGAGTGGT	ACTCTCAACAACCAATACTCCA
REST	TGTAGTTTGGAATTAGAAGTGTTATT	CTCTAAATCTAAACTCTCTATTCAACA
REST	AGAGAGTTGAGATTAGAGGGGA	ACTCAACTCCACAACCAAAC
REST	TGTTAGGAGTGTAATAGTTAAGTGG	ACCAAAATTCAAACCCCAACA
REST	GGGGTAGTAAGATAAATAGTAGGGA	TCTAACTACATAACCTCAAACCAA
REST	AGTGTTGGTAGTAGGTATTGGT	ACCTCTAATAACAAAATTACTCAACA
REST	TTTAGGATTAGGGATAGTAGTAAAGTT	ACCTTCCAACTCCCAAACAT
REST	GAATTGTAGGGAAAAGGTGAGT	ACACCTCAAATTTCAACACCA
REST	AAGTTTGTTAAAATGAGATTAGGATTG	ACCTATTATAAACTCCAACCTACAA
REST	AGGATGGTGTTGAAAATTGTTATT	CCCTACTTATAATACTCCTTAAACAAA
REST	TTGGGGAAGGTTTGTTGGTT	CTCTCTCAACCTTACTTCCAAAA
REST	GTGGAGATAATTGTTTTAGTGTTTGA	ACCACAACTAACATTCCCCA
REST	TTGGGGAAGGTTTGTTGGTT	CTCTCTCAACCTTACTTCCAAAA
REST	TGTTGTATTTTGGTTTAGTGGTTTG	ACACCTAAACTTTCAATCAACCA
REST	TGGTTAGGGGTAAGGTTGTG	AACCACAAACCCAACAATCC
REST	AGAATTGTAGGGAAAATGTGAGT	CCACACCTCAAATTCCAACA
REST	GTGGAGATAATTGTTTTAGTGTTTGA	ACCACAACTAACATTCCCCA
REST	GGAATGGTTTTGGTTGAGGT	CCAATACCTACCAAACAACCA
REST	TGTGTGAGGTTTGGTATGTAGT	AATAACACCACACATCAACCT
REST	TGAGATAAGGTTAGTATTATGGATAGT	ACTAATTTCTTAACTACATCACCAACT
REST	TGTATTTTGGGGATTTTAGGTAGG	TCCTCATAACAACCCAAAACT
REST	GGGATGGTGGTTGTTTGTTA	ATCTACCCAAACCTCCTCCT
REST	TGGATAGTAGGATTTGGGTTTGT	ACAACCTAACAAACATCAATTCCA
REST	TTGTTTAGGGAGGGGATTGG	CCCCAACCCTATAAACAAACA
REST	GATAGTGTGGGGAGTGGATT	AACCATACCTCCAAACTTTACA
REST	AGGTTTATGGGTTGGAAGTTT	AACACCACAACATCTCAACC
REST	TGTAGTGATTTTAGGATTTTGAGTGT	TCAATAAACTCTCCTACAAAATAAACT
REST	TGTGTGAGGTGTAATGTGTG	ACTACACAAACAAAACCCAACA
REST	AATGGGAAAGTAAGGTGAAGG	TCACTTCAACAAACTTTCCCC
REST	AGGGTAGAATGATTGTTTTAGTGT	TCTAAACTCTTAATACCTACCCAAAC
REST	AGGAGTATTTGGTTTGGAGTGA	ACTTTCATACACTTCCCACATTT
REST	TTTTAAGGTTGGAAGAGTGAAAGT	ACTTCAATCCAACCATCCTCC
REST	GTTGATAGTGGATGTAGTTAAAGGT	ACCCAAACAAATAAATCAAACCT
REST	AGATAAATTAGGGAGTGAAGGGA	AAAACTCTAACCACCACACCT
REST	TGTTTTATGGAATATTTGGGTTATGT	CAACCAACCATCCAACTAACA
REST	GTTTGGTTGGGGTAAAGTTAGT	AACCAAAATCATATCACAAATCCA

Table 5**.2 l Primer sequences for amplicon bisulfite sequencing.** Table lists primers that were used for all amplicon bisulfite sequencing experiments. Methylation region types are named are as follows: REST: REST-LMRs (LMRs with REST binding), CTCF: CTCF-LMRS (LMRs with CTCF binding), lambda: lambda control, T7: T7 control, FMR: fully methylated region, UMR: unmethylated region, MUMR: methylated UMR, DMR: differentially methylated region, NP: neuronal progenitor-specific LMR, ES: embryonic stem cell-specific LMR

PCR cycle	T [°C]	t [min:s]	Cycle number
1	95	09:00	1
2	95	00:30	20
	55 to 51*	00:30	
	72	00:30	
3	95	00:30	36
	51	00:30	
	72	00:30	
4	72	07:00	1
5	4	hold	1

Table 5.3 I Bisulfite PCR settings. *reduce $T_{annealing}$ by 0.2°C every cycle

5.5 Amplicon Bisulfite Sequencing Analysis

Analysis was performed by Juliane Schmidt. Arnaud Krebs set up the initial computational pipeline for Amplicon Bisulfite sequencing analysis. The subsequent AmpliconBiSeq package was mainly developed by Altuna Akalin.

Analysis of amplicon bisulfite sequencing experiments was performed with R version 3.3.0 (R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.), using the Bioconductor package QuasR_1.12.0 (Gaidatzis et al., 2015; Langmead et al., 2009) and AmpliconBiSeq_0.1.17 (Akalin, A., AmpliconBiSeq, GitHub repository, https://github.com/BIMSBbioinfo/AmpliconBiSeq).

In brief, FASTQ files were aligned within QuasR using Rbowtie v1.12.0 (parameters: -m 1 --best --strata) (Langmead et al., 2009). The July 2007 Mus musculus genome assembly (NCBI37/mm9) provided by the National Center for Biotechnology Information (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 for all future analyses. Raw alignment statistics and quality control were performed within QuasR. Conversion calculation of unmethylated lambda and pre-methylated T7 external spike-in DNA was performed within AmpliconBiSeq. Further data processing was

performed within AmpliconBiSeq, where individual reads were filtered for at least 80% conversion efficiencies (derived from non-CpG methylation percentage per read). Methylation was calculated from reads passing filter and with a minimum coverage of 50 reads per CpG. Methylation averages were calculated for single CpGs or over the covered amplicon windows. Jaccard similarities and meta-methylation profiles were called within AmpliconBiSeq with expected variance settings of 90 %. Further data visualization was performed within R version 3.3.0. The methylation heatmap (Figure 3.12) was created with the R package NMF (Gaujoux and Seoighe, 2010).

5.6 RNA-seq Analysis

Cell sample preparation was done by Juliane Schmidt. The sequencing was done by the lab of Michael Stadler and the data was analyzed by Michael Stadler.

Total RNA-seq was performed for four biological replicates of *Rest wt* and *Rest ko* cells (Arnold et al., 2013; Jørgensen et al., 2009a). Data alignment and analysis was done within QuasR (Gaidatzis et al., 2015; Langmead et al., 2009). RNA-seq signal was quantified as reads per kilobase of exon per million mapped reads (rpkm) and a pseudocount of 1 was added. REST targets were defined as genes, where at least one REST ChIP-seq peak is localized within a window of 10 kb of the corresponding TSS.

5.7 REST Motif Insertions

The experiment was designed by Arnaud Krebs with collaboration with Juliane Schmidt. Both performed the experiments. Arnaud Krebs performed the methylation analysis.

REST's position weight matrix (PWM) was downloaded from JASPAR 2010 (Portales-Casamar et al., 2010) and used as a reference to derive the sequence matching the maximum score motif. For the control motif, the sequence was iteratively randomized (10,000 times) and the sequence with a minimal predicted binding strength, while maintaining CG composition was chosen for insertion. Pairs of oligonucleotides were ordered for the REST motif and control motif and

annealed. The receiving construct was designed based on a previously characterized REST-LMR (Stadler et al., 2011). The underlying sequence was exchanged against random *E.coli* DNA, while CpG positions were maintained. The annealed REST motif and control oligos were cloned into the middle of the *E.coli* fragment and inserted into the beta globin locus of mESCs as previously described (Lienert et al., 2011a). Amplicon bisulfite sequencing was performed for mESCs as previously described (Chapter 5.4). The Bisulfite PCR was performed with primers flanking the *E.coli* construct. Sequencing was done on the Illumina MiSeq platform. Data analysis was performed using customized scripts in Perl and R.

5.8 ChIP-seq of REST Re-expression Cells

Chromatin immunoprecipitation (ChIP) was performed for one clone of each individual REST mutant re-expression cell line (Rest-/-REST c4, Rest-/-DBD c4, Rest-/-NTE-DBD 7c9, Rest-/-DBD-CTE 5c11; for simplicity reasons clone numbers are omitted in main text). The protocol was carried out as previously described (Weber et al., 2007).

Cells were cross-linked in medium with 1% formaldehyde for 10 min at room temperature. Formaldehyde was quenched by 150 mM glycine and washed twice with cold PBS. Cells were harvested in cold PBS, pelleted and resuspended in cold Buffer 1 (10 mM Tris pH 8.0, 10 mM EDTA pH 8.0, 0.5 mM EGTA, 0.25% Triton X-100) and incubated on ice for 5 min. Cells were pelleted and resuspended in Buffer 2 (10mM Tris pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl) and incubated on ice for 5 min. Cells were re-pelleted and resuspended in lysis buffer (50 mM HEPES/KOH pH 7.5, 500mM NaCl, 1mM EDTA, 1% Triton X-100, 1x Protease Inhibitor cocktail (Roche)) and incubated on ice for an hour. Chromatin lysate was sonicated on a Bioruptor (Diagenode) for 35 cycles (high setting, 30 sec on, 45 sec off).

100 ug sonicated chromatin was used for subsequent immunoprecipitation. Chromatin was pre-cleared for an hour with Dynabeads Protein A (Invitrogen), blocked prior with BSA and tRNA. 5% of pre-cleared lysate was used as an input control. Remaining lysate was incubated with 5ul of V5 antibody (R960-25, Thermo Fisher Scientific) rotating overnight at 4°C. The next day, lysate was

incubated with 30 µl Dynabeads Protein A (Invitrogen) for three hours. Beads were washed in subsequent 5 min rounds: twice lysis buffer and once DOC buffer (10mM Tris Ph 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% DOC, 1 mM EDTA). Chromatin was eluted in buffer (1% SDS, 100 mM NaHCO3). Reverse-crosslinking of chromatin was performed overnight (50 µg/ml RNase A, 30min at 37°C; 200 µg/ml Proteinase K, 3h at 55°C, overnight at 65°C). The next day, DNA was extracted by phenol/chlorophorm, precipitated in ethanol and resuspended in 40 ul TE buffer.

qPCR was performed for REST binding sites and negative controls using SYBR Green PCR Master Mix (Thermo Fisher Scientific). Measurements were normalized to input controls and negative controls. Up to ten ChIPs were pooled to generate libraries for high-throughput sequencing. Sequencing libraries were prepared according to manufacturer's protocol (Next ChIP-seq Library Prep Master Mix Set for Illumina, New England BioLabs) and sequenced on the Illumina HiSeq (50 bp, single-end).

5.9 ChIP-seq Analysis of REST Re-expression Cells

Analysis of ChIP-seq experiments was performed with R version 3.3.0 (R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.), using the Bioconductor package QuasR_1.12.0 (Gaidatzis et al., 2015; Langmead et al., 2009). In brief, FASTQ files were aligned within QuasR using Rbowtie v1.12.0 (parameters: -m 1 --best --strata) (Langmead et al., 2009) against mouse genome assembly mm9. Raw alignment statistics and quality was controlled within QuasR.

Reads were shifted by 75 bp and quantified in 200 bp windows over predicted REST motifs and normalized to the mapped library size in million. Peaks were called with MACS-1.4.2 (Zhang et al., 2008) using standard settings and providing a common input control. Venn diagrams were produced within the Bioconductor package ChIPpeakAnno (Zhu et al., 2010).

5.10 SIN3A ChIP-qPCR

ChIP was performed as described above (Chapter 5.8), using 10 ul of SIN3A antibody (K-20, sc-994, Santa Cruz Biotechnologies). Primer sequences are listed below (Table 5.4).

Region	Forward sequence	Reverse sequence
REST site a	CACCTGGCCTCTCTTCTCTC	CAACGTGGTCCCTACAAAGC
REST site b	TCAGTACCATGGACAGCGTT	GCTGGCCAGTTCAAGAAGAC
REST site c	AGGATGTCACTCCTGTTGGG	GAGCTACACCCTACCACCAG
SIN3A site a	CAGACAGACACCTACCCCTG	GCGTTGCTAGGAGAGAAGGA
SIN3A site b	CCCGGGACAGGACTGAAG	CGCCGCTTACCTCTGTTAAC
SIN3A site c	CAAGGCACAGCACGGAAG	CGCATTGCCTGTCGTCATTA

Table 5.4 I Primer sequences for SIN3A ChIP-qPCR. Primer regions are REST sites, i.e. REST binding sites with expected SIN3A enrichment. In contrast, SIN3A sites are enriched for SIN3A but not for REST.

5.11 Test of NTE-DBD Interaction Mutant after Transient Transfection

Point mutations were introduced into the N-terminal part (1 – 153 aa of REST) of the NTE-DBD construct directed by high conservation. The following mutations were introduced between positions 40 – 80 aa of REST (conserved amino acids in grey, mutations in red):

```
Wild type   ELAAPQLIMLANVALTGEASGSCCDYLVGEERQMAELMPVG
Mutant      ELAASQGIMGANVNLTGEASGSCCDYLVGEERQDAARMPRG
```

Plasmids expressing NTE-DBD and NTE-DBDmutant were transiently transfected into *Rest ko* cells. GFP was co-transfected visually inspect transfection efficiencies. Transfection was done using Lipofectamine 2000 (Thermo Fisher Scientific) according to manufacturer's protocol. High (protocol's recommended DNA amount) and low amounts (10% of protocol's recommended DNA amount) of plasmids were used during transfection. Transfected cells and *Rest ko* cells were subjected to SIN3A and V5 ChIP-qPCR the next day (Chapter 5.10, Chapter 5.8).

5.12 NOMe-seq

The initial NOMe-seq protocol was optimized and established in the laboratory by Arnaud Krebs. Experiments and analyses were performed by Juliane Schmidt and Arnaud Krebs together.

NOMe-seq was performed for one clone of each REST re-expression cell line. The protocol was based on previous publications (Kelly et al., 2010; Wolff et al., 2010). In brief, cells were trypsinized, washed with cold PBS and 250,000 cells were resuspended in 1 ml lysis buffer (10mM Tris pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1mM EDTA, 0.5% NP-40) and incubated for 5 to 10 min on ice. Next, cells were centrifuged for 5 min, 3000 rpm at 4°C and the supernatant discarded. Nuclei were resuspended in 250µl cold wash buffer (10mM Tris pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1mM EDTA). Samples were again centrifuged for 5 min, 3000 rpm at 4°C and supernatant discarded. Nuclei were resuspended in 94.5µl 1X M.GpC buffer. Samples were treated with GpC methyltransferase as follows: 94.5 µl nuclei, 20 µl 10X M.GpC buffer, 6 µl 32 mM SAM, 90 µl 1 M sucrose, 50 µl GpC methyltransferase (M.CviPI, 200 U of 4U/ul, New England BioLabs), 34 µl H2O. Nuclei were incubated for 7.5 min at 37°C. 300 µl stop solution (20 mM Tris-HCl, pH 7.9, 600 mM NaCl, 1% SDS, 10 mM EDTA) was added to the final reaction. Samples were incubated with 200 µg/ml proteinase K at 55°C for 16 h. DNA was isolated by phenol/chloroform extraction and ethanol precipitation. From this step on, the protocol followed the regular amplicon Bis-seq steps (Chapter 5.4). Bisulfite primers were designed over REST binding sites and control sites (Table 5.5).

Analysis of NOMe-seq was performed within R version 3.3.0 (R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.), using the Bioconductor package QuasR_1.12.0 (Gaidatzis et al., 2015). The only difference to regular amplicon Bis-seq is the additional methylation quantification over GpCs. GpCpGs were eliminated from both, CG and GC, methylation analyses due to unambiguity of the calls.

Region	Forward sequence	Reverse sequence
lambda	TGTGTTGGTTGGAAGAGGTT	ACTATCACTCTTCTCCTCCTCT
lambda	TGTTGTTGGTTGATTTTGATGAG	TCCTCTTTCAACTCTACCACA
T7	AGTGAGGGTATTGATTTTGAGT	ACCTTAAATCTATCACTCAACAAATTC
T7	GGGATGGTGAGTTTGTTGAA	CCTAATACATCTACAACTACCTCAT
UMR	AGGATTGTTTGGGATGGAAAA	CAAACTCAACCCAACCAACC
UMR	TGGAGGGGAAAAGGGAAAAT	ACACAACAACTACATCAACTAAACT
ES	AGGGAGATGATAGATTAGGTGAT	ACCTTCCACTATCCCTACTCA
CTCF	TGGGAGAGGAAGTGTGTTTT	ATCAACAACCACCTCCAAAA
REST	TGTAGTTTGGAATTAGAAGTGTTATT	CTCTAAATCTAAACTCTCTATTCAACA
REST	GATAGTGTGGGGAGTGGATT	AACCATACCTCCAAACTTTACA
REST	AGGGTAGAATGATTGTTTTAGTGT	TCTAAACTCTTAATACCTACCCAAAC
REST	AGATAAATTAGGGAGTGAAGGGA	AAAACTCTAACCACCACACCT
REST	GGAAGATAGGAGATGGGTGTT	TCCCTAACTTCTAATCCTCTATCT
REST	GTGATTGTTAAGAAGAGGTAGTTTT	AAAACACCCATCTCCTATCTTC
REST	GGAGAGAGGTTTTGAGTTTAATTTT	AAAACACTCCATATCTTCTTATCAAA
REST	AGATAGAGGATTAGAAGTTAGGGA	CACCATATAAAATTAAACTCAAAACCT
REST	TGAGTGTGGATGATTAAGAAGAAG	CCTCAATTCCTCACTTTTCAATATAAA
REST	TTTGGATTAGATAGTAGGAGATTTAGA	CACCTTCTTCTTAATCATCCACA
REST	GAAGATTAGAAGAGGGTGTTTGA	TCTAAATCTCCTACTATCTAATCCAAA
REST	TTGGTTTATAAAGTGAGTTTTAGGAT	ACCCTCTTCTAATCTTCTAATCACA
REST	TGTGGGATAGTGATTTATGGGA	AAACCCTCCTCTCCAATCTC
REST	TGGAAGGTTATTTTGGGGTTTAG	CCATTCCCTATCTATCCACTCC
REST	GAGATTGGAGAGGAGGGTTT	TTCTTACTTTTCATTCCTTCTAAACT
REST	AGTGTGAGGGATAAAATTAGGGA	CTATAACCCTCTCAATAAAATATTCCA
REST	AGTTTAGAAGGAATGAAAAGTAAGAA	TCCTTCCAATCCAAATATAAAACTCT
REST	GGTAATTTTGGATTTGAGAATTTAGGA	AAATACCCTTCCCATAAATCACT
REST	GGTGGAAGATTAGGGTGTGA	TCTCAAAACTCATAAACTCAAAACT
REST	ATTTGGATTGTTAATTTGAGTGGA	CACAACAAACAACTCACTCAA
REST	TGGTATGGTTTAGTATGGATTAGTG	CACCCTAATCTTCCACCTTATCT
REST	GGTGGAAGATTAGGGTGTGA	TCTCAAAACTCATAAACTCAAAACT
REST	ATTTGGATTGTTAATTTGAGTGGA	CACAACAAACAACTCACTCAA
REST	TGGTATGGTTTAGTATGGATTAGTG	CACCCTAATCTTCCACCTTATCT
REST	GGTGGAAGATTAGGGTGTGA	TCTCAAAACTCATAAACTCAAAACT
REST	TGGTATGGTTTAGTATGGATTAGTG	CACCCTAATCTTCCACCTTATCT
REST	GGGATATAGGTGGAAGGTGTATA	CACACCTAATAACTCATTCACTCA
REST	TGTTTGGTTAAAGTTTAATTTGGAAA	TCACTTATCCCACACCACATAA
REST	GAGTTGTGGGTATATGGAATTAGT	CCTAAACCTTACCAACCTAATATCTAA
REST	GGATGTTTATTTAGATATTAGGTTGGT	ACAAATTTACTAAATATACACCTTCCA
REST	AGATAGAGGGTTAGTTAGGAGAGA	CCACAAAATTAACAAAATCCCAAATTA
REST	GGATTTTGAGTTGAAGTAGTGGATT	TCATCCCTAAAATAAAACCCACAA
REST	TGGGTTTTATTTTAGGGATGAGG	AACACAAAACCCCACTTAATAATAA
REST	GAGATGGTGTAGGAGAGAAATTT	AACCAAATCCTATCCCTAAAATAAAA

REST	GTTTGATTTATATAAAAGAATGGGAGG	CAAATTAAATTTCTCTCCTACACCA
REST	AGAATGATTGTTTAAGGGATTGGA	CAATAATCACCTCAATATCACCCA
REST	AGATGAGAAAATTGAGGTTTAGATG	TCCAATCCCTTAAACAATCATTCT
REST	ATTATTTGTAATGGGATTTGATGTTTT	AAAACCTAAATCCCACCCCT
REST	AAATTATAAGGGAGGAGAGAGGA	CTCCATTTCCTTACTAATCCAATCA
REST	TGTTAAATATGGGAAGTTTGGTTT	ACCATATCTATTACTCCATTTCCTT
REST	TGAGTGGAAATTTATGGTTAGGT	AAACATCAAATCCCATTACAAATAATT
REST	GGGTGGGGATGAATAGGTTT	CCACTATCATCCCAACAATCT
REST	AGATTGTTGGGATGATAGTGG	ACTTCCTAATAATAACCCACATCAATA
REST	TTGGGGATAGTTTAAATTATGAGAAAA	CCTATTCTAACAATACAAACCCAAAA
REST	TGGGATAGGTTATTTATTATTTTGTGT	TCTCTAAACACTAAAACAATCATTCT
REST	GAATTTTGGGTTTGTATTGTTAGAATA	CCTCTCTAAATAAAACTACTACCCATA
REST	TGAGTTTTATTGTGGGTAATTGGA	TCACCTATCTCAATTCTCTTTCCT
REST	TTTTGGGTGGGAATTTTGTTTT	TTAAATTCCCACCCTTATTACAAAA
REST	GTTTGTAATGGGTTATTTAAATGGAGT	AACACCAAAATTCCAATTACCC
REST	TGTTTTGGTTGAGAAAAGAGATAGT	CCTCTCATAAAACACTATCTTTCTATC
REST	AGTTTTGGATAAGATGAATTTGAGT	ACTCCATTTAAATAACCCATTACAAAC
REST	TGTAATAAGGGTGGGAATTTAATTTT	AAACTCCATTTAAATAACCCATTACA
REST	AGTATAGTAGTTAAGGGTGATTTTGA	ACAAAACCACCCAATCTCAA
REST	TTAGATAGTTGTTTTAGTTAGGGTTTT	CCTTCAAAATCACCCTTAACTACT
REST	TGGTTTATAGAGTGAGTTTTAGGATAG	AAAACCCTAACTAAAACAACTATCTAA
REST	TTTGTGATAGGAAGAGATAGATTGA	TCAAACACTTAACTAATATCCCCT
REST	TGTAGGAGAGGAATATTTAAAAGGG	TCTATCAATCTATCTCTTCCTATCAC
REST	TTGATGGAGGAGGAGTGATT	CACTAATCCTCAAACTACCCCA
REST	GGGTATAAGATAGTGTGAAGTGATTT	ATCACTCCTCCTCCATCAAA
REST	TGGGGTAGTTTGAGGATTAGTG	CACCACAACATTCCATATCTATCTAA
REST	GGATGATTGTGTTGGTAAAGAATT	TCAACAATTCCACATCCTTCTAAA
REST	TTTGGAGAGGGATTATGGGG	AAAATCCATCCCCTTTCCCT
REST	AGAGTTATAGGGAAGGATTTAAGAAA	ACCTAATACTAAACCTACCCCAAAT
REST	TGTGAGATTTTGAAGGATAGATAGG	TTTCTTAAATCCTTCCCTATAACTCT
REST	TTGTTATTGTGGGTGGAGGG	CACTCAAACTAACAAAACCCCA
REST	TGTGGATTTTAGTTTTGGGGTATT	CCCCTCCACCCACAATAA
REST	TGATGGGGTTTTGTTAGTTTGA	TCTACCAAACACCTCATTAAAACT
REST	AGTGGGTAGATATTGGGGTTATT	TCTCTCTCTCTCTCTCTCTATATA
REST	GTTTTGTTAGTTTGAGTGAGTGTT	TTACTAATAACCCCAATATCTACCC
REST	GGAAGGGTATTTTGGGATTTGG	TCCAATTCCCCACTATCCAC
REST	GGAGTGGATAGTGGGGAATT	AAAATCCTCCTTCCTACCTCA
REST	GGAAGGGTATTTTGGGATTTGG	TCCAATTCCCCACTATCCAC
REST	GGAGTGGATAGTGGGGAATT	AAAATCCTCCTTCCTACCTCA
REST	AGAGTTTAAAGGAATTAGAAAGAGTGA	TCCCAAAATACCCTTCCATCT
REST	AGAGATTGGAGTTTTAGGGGT	CCTCCAATCTCCAATCTCATCT
REST	TGGAGGAATTAGAGATTTAATGGAG	ACCCCTAAAACTCCAATCTCT
REST	TGTGTTATTGTGTTTGTTTGTGAAT	ACTTAAACACCTAATCCTCCTCT

REST	GGGTGAAAGGGGAATTTTGA	CCACACTCCAATCACACTAAAA
REST	GGGATGAGAATTTTGGGAAGAG	CCAAACCCACCAATTTCCTAA
REST	TGTTTGGAGTTTGGTTTAGTAGG	AAACCCTAAAATCTCACTAACCA
REST	TGGTTAGTGAGATTTTAGGGTTT	CAAATATCAACCTCAAAACATCTCA
REST	GGAGAGATTTATTTAATTTTGGGATTT	CCTTCCCTCCAAAATCTTCTAT
REST	TTTGTGGTGGTGGGAGATAA	AACCCAAAATCTCCACTCCC
REST	AGGAGGGAGTGGAGATTTTG	ACATCCACTAAATCATCTCCTATAA
REST	AGATTGGTAGATATAATTATGGGAGG	ATTAAACATCTTCCTTCCATTCAAA
REST	GGGAGTTTAGGGATGGATTGTT	CCCAAATTATATCATCTCTTCTCTTTT
REST	TTGGGGTGGGTATTAGAATTTT	AAAATTAATCCTACAACACTACCATT
REST	TTAGGATATTTGAGAAAGTTGAGTG	AAAACTTAATACAAATTCCTCCCC

Table 5.5 I Primer sequences for NOMe-seq. Table lists primers that were used for targeted NOMe-seq experiments. Methylation region types are named are as follows: REST: REST-LMRs (LMRs with REST binding), CTCF: CTCF-LMRS (LMRs with CTCF binding), lambda: lambda control, T7: T7 control, UMR: unmethylated region, ES: embryonic stem cell-specific LMR

5.13 Kinetics of REST Binding Site Methylation after Transient REST Transfection

Rest ko cells were transiently transfected with full-length REST expressing plasmid. Transfection was performed with Lipofectamine 2000 (Thermo Fisher Scientific) according to manufacturer's instructions. Cells were harvested at 8h, 24h and 48h after addition of the liposomal DNA complexes to cells. GFP was co-transfected to visually inspect transfection efficiencies. Genomic DNA was isolated from cells and amplicon Bis-seq performed and analyzed as described above (Chapter 5.4).

5.14 Random Forest Model to Predict Methylation Levels of Distal REST Binding Sites

The analysis was done in R version 3.3.0 (R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.) using the package randomForest_4.6-12 (Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.).

The following data sets were used for the analysis: GSM748786, GSM748787, GSM1004653, GSM978374, GSM671094, GSM671095, GSM698700, GSM698701, GSM698697, GSM698698, GSM698699, GSM611196, GSM611197, GSM845235, GSM659799, GSM1023124, GSM687277, GSM687279, GSM896923, GFP (Table 5.1).

A random forest model was built to predict average methylation over 100 bp windows around bound REST motifs. Only occupied distal (further than 2kb from annotated TSS) REST motifs were included in the prediction. Several quantitative and categorical predictors were used. In brief, ChIP-seq experiments for REST and its direct and indirect interactors were used as quantitative predictors. ChIP-seq signal was quantified over 500 bp windows around distal bound REST motifs and normalized to mapped library size. Overlap with genomic features was treated as a categorical predictor (repeat overlap: within 500 bp of annotated repeat, gene overlap: within 500 bp of annotated gene). In addition, I calculated a nucleosome positioning score based on the following calculation: log2(mean(MNase reads(-165 bp – -115 bp of REST motif center), MNase reads(115 bp – 165 bp of REST motif center))+10)- log2(MNase reads(-25 – 25 bp of REST motif center)+10). A high score therefore indicates high positioning of nucleosomes at plus one and minus one position around the REST motif. I used 80 % of the data to train the random forest model (parameters: mtry= 23, ntree= 1000) and the remaining data as a test set. Variable importance evaluation was done within the randomForest package.

List of Abbreviations

aC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
aa	Amino acid
ADD	ATRX-DNMT3-DNMT3L
Amplicon Bis-seq	Amplicon bisulfite sequencing
BDNF	Brain-derived neurotrophic factor
CGI	CpG island
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
DBD	DNA-binding domain
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DSBH	Double-stranded beta-helix
ENCODE	Encyclopedia of DNA Elements
ERV	Endogenous retrovirus
ES(C)	Embryonic stem (cell)
FMR	Fully methylated region
GFP	Green fluorescent protein
HDAC	Histone deacetylase
HID	Histone deacetylase interaction domain
HMM	Hidden Markov Model
ID	Interaction domain
KLF4	Krueppel-like factor 4
LIF	Leukemia inhibitory factor
LMR	Low-methylated region
LSD1	Lysine-specific histone demethylase 1
MBD	Methyl-binding domain
mESC	Mouse embryonic stem cell

mNP	Mouse neuronal progenitor
mTN	Mouse terminal neuron
NMR	Nuclear magnetic resonance
NOMe-seq	Nucleosome Occupancy Methylome-sequencing
NRF1	Nuclear respiratory factor 1
NRL	Nucleosomal repeat length
NRSF	Neuron-restrictive silencing factor
PAH	Paired amphipathic helix
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PMD	Partially methylated domain
PRC2	Polycomb-repressive complex 2
PWM	Position weight matrix
QuasR	Quantify and annotate short reads in R
RCOR	REST corepressor 1
REST	RE1-silencing transcription factor
RNA-seq	RNA sequencing
SELEX	Systematic evolution of ligands by exponential enrichment
SRA	SET and RING finger associated
TALE	Transcription activator–like effector
TBP	TATA-binding protein
TDG	Thymine DNA glycosylase
TET	Ten-eleven Translocation
TF	Transcription factor
TKO	Triple knock out
TSS	Transcription start site
UMR	Unmethylated region

Protein names are in capital letters irrespective of species and gene names are in italics, non-capitalized.

List of Figures

Individual contributions to figure generation and copyright permission is stated below figure titles in italic.

References

Accomando, W.P., Wiencke, J.K., Houseman, E., Nelson, H.H., and Kelsey, K.T. (2014). Quantitative reconstruction of leukocyte subsets using DNA methylation. Genome Biol. *15*, R50.

Achour, M., Jacq, X., Rondé, P., Alhosin, M., Charlot, C., Chataigneau, T., Jeanblanc, M., Macaluso, M., Giordano, A., Hughes, A.D., et al. (2008). The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. Oncogene *27*, 2187–2197.

Aggarwal, A., Rodgers, D., Drottar, M., Ptashne, M., and Harrison, S. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science *242*, 899–907.

Akalin, A., AmpliconBiSeq. GitHub repository, https://github.com/BIMSBbioinfo/AmpliconBiSeq.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

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. Epigenetics (Cold Spring Harbor, New York: CSH Press, Cold Spring Harbor Laboratory Press). 2015

Amador-Arjona, A., Cimadamore, F., Huang, C.-T., Wright, R., Lewis, S., Gage, F.H., and Terskikh, A.V. (2015). SOX2 primes the epigenetic landscape in neural precursors enabling proper gene activation during hippocampal neurogenesis. Proc. Natl. Acad. Sci. *112*, E1936–E1945.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. *23*, 185–188.

Andrés, M.E., Burger, C., Peral-Rubio, M.J., Battaglioli, E., Anderson, M.E., Grimes, J., Dallman, J., Ballas, N., and Mandel, G. (1999). CoREST: a functional corepressor required for regulation of neural-specific gene expression. Proc. Natl. Acad. Sci. U. S. A. *96*, 9873–9878.

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

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.

Arvey, A., Agius, P., Noble, W.S., and Leslie, C. (2012). Sequence and chromatin determinants of cell-type-specific transcription factor binding. Genome Res. *22*, 1723–1734.

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

Ballas, N., Grunseich, C., Lu, D.D., Speh, J.C., and Mandel, G. (2005). REST and Its Corepressors Mediate Plasticity of Neuronal Gene Chromatin throughout Neurogenesis. Cell *121*, 645–657.

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

Battaglioli, E., Andrés, M.E., Rose, D.W., Chenoweth, J.G., Rosenfeld, M.G., Anderson, M.E., and Mandel, G. (2002). REST repression of neuronal genes requires components of the hSWI.SNF complex. J. Biol. Chem. *277*, 41038–41045.

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

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243–247.

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

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169–181.

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

Bessis, A., Champtiaux, N., Chatelin, L., and Changeux, J.P. (1997). The neuron-restrictive silencer element: a dual enhancer/silencer crucial for patterned expression of a nicotinic receptor gene in the brain. Proc. Natl. Acad. Sci. U. S. A. *94*, 5906–5911.

Bestor, T.H., and Ingram, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc. Natl. Acad. Sci. U. S. A. *80*, 5559–5563.

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.

Bingham, A.J., Ooi, L., Kozera, L., White, E., and Wood, I.C. (2007). The Repressor Element 1-Silencing Transcription Factor Regulates Heart-Specific Gene Expression Using Multiple Chromatin-Modifying Complexes. Mol. Cell. Biol. *27*, 4082–4092.

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

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

Blattler, A., and Farnham, P.J. (2013). Cross-talk between site-specific transcription factors and DNA methylation states. J. Biol. Chem. *288*, 34287–34294.

Boller, S., Ramamoorthy, S., Akbas, D., Nechanitzky, R., Burger, L., Murr, R., Schübeler, D., and Grosschedl, R. (2016). Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. Immunity *44*, 527–541.

Booth, M.J., Ost, T.W.B., Beraldi, D., Bell, N.M., Branco, M.R., Reik, W., and Balasubramanian, S. (2013). Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat. Protoc. *8*, 1841–1851.

Booth, M.J., Marsico, G., Bachman, M., Beraldi, D., and Balasubramanian, S. (2014). Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Nat. Chem. *6*, 435–440.

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

Boyes, J., and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell *64*, 1123–1134.

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

Breiman. Classification and regression trees (Boca Raton: Chapman & Hall [u.a.]). 1998

Bruniquel, D., and Schwartz, R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nat. Immunol. *4*, 235–240.

Burger, L., Gaidatzis, D., Schübeler, 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.

Cartron, P.-F., Nadaradjane, A., LePape, F., Lalier, L., Gardie, B., and Vallette, F.M. (2013). Identification of TET1 Partners That Control Its DNA-Demethylating Function. Genes Cancer *4*, 235–241.

De Carvalho, D.D., You, J.S., and Jones, P.A. (2010). DNA methylation and cellular reprogramming. Trends Cell Biol. *20*, 609–617.

De Carvalho, D.D., Sharma, S., You, J.S., Su, S.-F., Taberlay, P.C., Kelly, T.K., Yang, X., Liang, G., and Jones, P.A. (2012). DNA Methylation Screening Identifies Driver Epigenetic Events of Cancer Cell Survival. Cancer Cell *21*, 655–667.

Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. Nature *493*, 561–564.

Chen, Z.F., Paquette, A.J., and Anderson, D.J. (1998). NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. Nat. Genet. *20*, 136–142.

Chong, J.A., Tapia-Ramírez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuller, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995). REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. Cell *80*, 949–957.

Cierpicki, T., Risner, L.E., Grembecka, J., Lukasik, S.M., Popovic, R., Omonkowska, M., Shultis, D.D., Zeleznik-Le, N.J., and Bushweller, J.H. (2010). Structure of the MLL CXXC domain–DNA complex and its functional role in MLL-AF9 leukemia. Nat. Struct. Mol. Biol. *17*, 62–68.

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

Conaway, R.C., and Conaway, J.W. (2011). Function and regulation of the Mediator complex. Curr. Opin. Genet. Dev. *21*, 225–230.

Cremer, T., and Cremer, M. (2010). Chromosome territories. Cold Spring Harb. Perspect. Biol. *2*, a003889.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Cuculis, L., Abil, Z., Zhao, H., and Schroeder, C.M. (2015). Direct observation of TALE protein dynamics reveals a two-state search mechanism. Nat. Commun. *6*, 7277.

Dantas Machado, A.C., Zhou, T., Rao, S., Goel, P., Rastogi, C., Lazarovici, A., Bussemaker, H.J., and Rohs, R. (2015). Evolving insights on how cytosine methylation affects protein-DNA binding. Brief. Funct. Genomics *14*, 61–73.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.-C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.-W., Page, D.C., et al. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell *9*, 166–175.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., et al. (2013). Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. Dev. Cell *24*, 310–323.

Dawlaty, M.M., Breiling, A., Le, T., Barrasa, M.I., Raddatz, G., Gao, Q., Powell, B.E., Cheng, A.W., Faull, K.F., Lyko, F., et al. (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Dev. Cell *29*, 102–111.

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

Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Södersten, E., and Hansen, K. (2012). REST–Mediated Recruitment of Polycomb Repressor Complexes in Mammalian Cells. PLoS Genet. *8*, e1002494.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. Nature *528*, 575–579.

Dong, X., Greven, M.C., Kundaje, A., Djebali, S., Brown, J.B., Cheng, C., Gingeras, T.R., Gerstein, M., Guigó, R., Birney, E., et al. (2012). Modeling gene expression using chromatin features in various cellular contexts. Genome Biol. *13*, R53.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Duymich, C.E., Charlet, J., Yang, X., Jones, P.A., and Liang, G. (2016). DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat. Commun. *7*, 11453.

Elsässer, S.J., and D'Arcy, S. (2012). Towards a mechanism for histone chaperones. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1819*, 211–221.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Fan, G., Beard, C., Chen, R.Z., Csankovszki, G., Sun, Y., Siniaia, M., Biniszkiewicz, D., Bates, B., Lee, P.P., Kuhn, R., et al. (2001). DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. J. Neurosci. Off. J. Soc. Neurosci. *21*, 788–797.

Favaro, R., Valotta, M., Ferri, A.L.M., Latorre, E., Mariani, J., Giachino, C., Lancini, C., Tosetti, V., Ottolenghi, S., Taylor, V., et al. (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. Nat. Neurosci. *12*, 1248–1256.

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

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398–402.

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., et al. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212–224.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods *7*, 461–465.

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.

Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Doré, L.C., et al. (2015). N6-Methyldeoxyadenosine Marks Active Transcription Start Sites in Chlamydomonas. Cell *161*, 879–892.

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

Gaidatzis, D., Lerch, A., Hahne, F., and Stadler, M.B. (2015). QuasR: quantification and annotation of short reads in R. Bioinforma. Oxf. Engl. *31*, 1130–1132.

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149–1163.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, 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.

Gray, P.A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D.-I., Tsung, E.F., Cai, Z., et al. (2004). Mouse brain organization revealed through direct genome-scale TF expression analysis. Science *306*, 2255–2257.

Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.-H., Aravind, L., He, C., and Shi, Y. (2015). DNA Methylation on N6-Adenine in C. elegans. Cell *161*, 868–878.

Grimes, J.A., Nielsen, S.J., Battaglioli, E., Miska, E.A., Speh, J.C., Berry, D.L., Atouf, F., Holdener, B.C., Mandel, G., and Kouzarides, T. (2000). The co-repressor mSin3A is a functional component of the REST-CoREST repressor complex. J. Biol. Chem. *275*, 9461–9467.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606–610.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

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

Guilhamon, P., Eskandarpour, M., Halai, D., Wilson, G.A., Feber, A., Teschendorff, A.E., Gomez, V., Hergovich, A., Tirabosco, R., Fernanda Amary, M., et al. (2013). Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. Nat. Commun. *4*, 2166.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature *517*, 640–644.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369.

Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E., and Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature *455*, 826–829.

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.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (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., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

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

Von Hippel, P.H., and Berg, O.G. (1989). Facilitated target location in biological systems. J. Biol. Chem. *264*, 675–678.

Hodges, E., Molaro, A., Dos Santos, C.O., Thekkat, P., Song, Q., Uren, P.J., Park, J., Butler, J., Rafii, S., McCombie, W.R., et al. (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.

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res. *22*, 246–258.

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.

Hon, G.C., Song, C.-X., Du, T., Jin, F., Selvaraj, S., Lee, A.Y., Yen, C., Ye, Z., Mao, S.-Q., Wang, B.-A., et al. (2014). 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. Mol. Cell *56*, 286–297.

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., et al. (2013). DNA methylation presents distinct binding sites for human transcription factors. eLife *2*.

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.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.

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.

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.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33 Suppl*, 245–254.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Johnson, D.S., Mortazavi, A., Richard M. Myers, and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.

Jones, P.A., Archer, T.K., Baylin, S.B., Beck, S., Berger, S., Bernstein, B.E., Carpten, J.D., Clark, S.J., Costello, J.F., Doerge, R.W., et al. (2008). Moving AHEAD with an international human epigenome project. Nature *454*, 711–715.

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.

Jørgensen, H.F., and Fisher, A.G. (2010). Can controversies be put to REST? Nature *467*, E3–E4.

Jørgensen, H.F., Terry, A., Beretta, C., Pereira, C.F., Leleu, M., Chen, Z.-F., Kelly, C., Merkenschlager, M., and Fisher, A.G. (2009a). REST selectively represses a subset of RE1-containing neuronal genes in mouse embryonic stem cells. Dev. Camb. Engl. *136*, 715–721.

Jørgensen, H.F., Chen, Z.-F., Merkenschlager, M., and Fisher, A.G. (2009b). Is REST required for ESC pluripotency? Nature *457*, E4–E5.

Kallunki, P., Edelman, G.M., and Jones, F.S. (1997). Tissue-specific expression of the L1 cell adhesion molecule is modulated by the neural restrictive silencer element. J. Cell Biol. *138*, 1343–1354.

Kangaspeska, S., Stride, B., Métivier, 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.

Kelly, T.K., Miranda, T.B., Liang, G., Berman, B.P., Lin, J.C., Tanay, A., and Jones, P.A. (2010). H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. Mol. Cell *39*, 901–911.

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.

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.

Klug, M., Schmidhofer, S., Gebhard, C., Andreesen, R., and Rehli, M. (2013). 5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes. Genome Biol. *14*, R46.

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. Bioinforma. Oxf. Engl. *23*, 1289–1291.

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

Kraner, S.D., Chong, J.A., Tsay, H.J., and Mandel, G. (1992). Silencing the type II sodium channel gene: a model for neural-specific gene regulation. Neuron *9*, 37–44.

Krebs, A.R., Dessus-Babus, S., Burger, L., and Schübeler, D. (2014). High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. eLife *3*.

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

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.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095–1107.

Lazarovici, A., Zhou, T., Shafer, A., Dantas Machado, A.C., Riley, T.R., Sandstrom, R., Sabo, P.J., Lu, Y., Rohs, R., Stamatoyannopoulos, J.A., et al. (2013). Probing DNA shape and methylation state on a genomic scale with DNase I. Proc. Natl. Acad. Sci. U. S. A. *110*, 6376–6381.

Lee, E.-J., Luo, J., Wilson, J.M., and Shi, H. (2013). Analyzing the cancer methylome through targeted bisulfite sequencing. Cancer Lett. *340*, 171–178.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature *437*, 432–435.

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

Li, Z., Cai, X., Cai, C.-L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.-C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509–4518.

Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C., et al. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat. Genet. *47*, 469–478.

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schübeler, D. (2011a). Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet.

Lienert, F., Mohn, F., Tiwari, V.K., Baubec, T., Roloff, T.C., Gaidatzis, D., Stadler, M.B., and Schübeler, D. (2011b). Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. PLoS Genet. *7*, e1002090.

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., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68–73.

Long, H.K., Sims, D., Heger, A., Blackledge, N.P., Kutter, C., Wright, M.L., Grützner, F., Odom, D.T., Patient, R., Ponting, C.P., et al. (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.

Lönnerberg, P., Schoenherr, C.J., Anderson, D.J., and Ibáñez, C.F. (1996). Cell type-specific regulation of choline acetyltransferase gene expression. Role of the neuron-restrictive silencer element and cholinergic-specific enhancer sequences. J. Biol. Chem. *271*, 33358–33365.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014a). Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev. *28*, 2103–2119.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014b). Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev. *28*, 2103–2119.

Lu, X., Song, C.-X., Szulwach, K., Wang, Z., Weidenbacher, P., Jin, P., and He, C. (2013). Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. J. Am. Chem. Soc. *135*, 9315–9317.

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

Luger, K., Mäder, 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.

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.

Magnani, L., Eeckhoute, J., and Lupien, M. (2011). Pioneer factors: directing transcriptional regulators within the chromatin environment. Trends Genet. *27*, 465–474.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J. Biol. Chem. *286*, 35334–35338.

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.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *44*, D110–D115.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep. *12*, 1184–1195.

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

Maze, I., Noh, K.-M., Soshnev, A.A., and Allis, C.D. (2014). Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nat. Rev. Genet. *15*, 259–271.

McDonel, P., Demmers, J., Tan, D.W.M., Watt, F., and Hendrich, B.D. (2012). Sin3a is essential for the genome integrity and viability of pluripotent cells. Dev. Biol. *363*, 62–73.

McGann, J.C., Oyer, J.A., Garg, S., Yao, H., Liu, J., Feng, X., Liao, L., Yates, J.R., and Mandel, G. (2014). Polycomb- and REST-associated histone deacetylases are independent pathways toward a mature neuronal phenotype. eLife *3*.

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

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

Mieda, M., Haga, T., and Saffen, D.W. (1997). Expression of the rat m4 muscarinic acetylcholine receptor gene is regulated by the neuron-restrictive silencer element/repressor element 1. J. Biol. Chem. *272*, 5854–5860.

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

Miller, J.C., Zhang, L., Xia, D.F., Campo, J.J., Ankoudinova, I.V., Guschin, D.Y., Babiarz, J.E., Meng, X., Hinkley, S.J., Lam, S.C., et al. (2015). Improved specificity of TALE-based genome editing using an expanded RVD repertoire. Nat. Methods *12*, 465–471.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*, 393–396.

Mohn, F., and Schübeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet. TIG *25*, 129–136.

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

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell *20*, 11–24.

Mori, N., Schoenherr, C., Vandenbergh, D.J., and Anderson, D.J. (1992). A common silencer element in the SCG10 and type II Na+ channel genes binds a factor present in nonneuronal cells but not in neuronal cells. Neuron *9*, 45–54.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing. Mol. Cell *53*, 277–289.

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.

Nan, X., Cross, S., and Bird, A. (2007). Gene Silencing by Methyl-CPG-Binding Proteins. In Novartis Foundation Symposia, D.J. Chadwick, and G. Cardew, eds. (Chichester, UK: John Wiley & Sons, Ltd.), pp. 6–21.

Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat. Genet. *23*, 58–61.

Nomura, M., Uda-Tochio, H., Murai, K., Mori, N., and Nishimura, Y. (2005). The neural repressor NRSF/REST binds the PAH1 domain of the Sin3 corepressor by using its distinct short hydrophobic helix. J. Mol. Biol. *354*, 903–915.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

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, L., and Wood, I.C. (2007). Chromatin crosstalk in development and disease: lessons from REST. Nat. Rev. Genet. *8*, 544–554.

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

Ovando-Roche, P., Yu, J.S.L., Testori, S., Ho, C., and Cui, W. (2014). TRF2-mediated stabilization of hREST4 is critical for the differentiation and maintenance of neural progenitors. Stem Cells Dayt. Ohio *32*, 2111–2122.

Pardo, C., Hoose, S.A., Pondugula, S., and Kladde, M.P. (2009). DNA methyltransferase probing of chromatin structure within populations and on single molecules. Methods Mol. Biol. Clifton NJ *523*, 41–65.

Pardo, C.E., Darst, R.P., Nabilsi, N.H., Delmas, A.L., and Kladde, M.P. (2011). Simultaneous single-molecule mapping of protein-DNA interactions and DNA methylation by MAPit. Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al *Chapter 21*, Unit 21.22.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394–397.

Pazin, M.J., and Kadonaga, J.T. (1997). What's up and down with histone deacetylation and transcription? Cell *89*, 325–328.

Perera, A., Eisen, D., Wagner, M., Laube, S.K., Künzel, A.F., Koch, S., Steinbacher, J., Schulze, E., Splith, V., Mittermeier, N., et al. (2015). TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression. Cell Rep. *11*, 283–294.

Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeier, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew. Chem. Int. Ed Engl. *50*, 7008–7012.

Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W., and Sandelin, A. (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res. *38*, D105–D110.

Prendergast, G.C., and Ziff, E.B. (1991). Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science *251*, 186–189.

Ramachandran, S., and Henikoff, S. (2016). Transcriptional Regulators Compete with Nucleosomes Post-replication. Cell *165*, 580–592.

Rampal, R., Alkalin, A., Madzo, J., Vasanthakumar, A., Pronier, E., Patel, J., Li, Y., Ahn, J., Abdel-Wahab, O., Shih, A., et al. (2014). DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. Cell Rep. *9*, 1841–1855.

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.

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

Robertson, A.B., Dahl, J.A., Vågbø, C.B., Tripathi, P., Krokan, H.E., and Klungland, A. (2011). A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. *39*, e55.

Robertson, A.B., Dahl, J.A., Ougland, R., and Klungland, A. (2012). Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. Nat. Protoc. *7*, 340–350.

Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009). The role of DNA shape in protein–DNA recognition. Nature *461*, 1248–1253.

Roopra, A., Qazi, R., Schoenike, B., Daley, T.J., and Morrison, J.F. (2004). Localized domains of G9a-mediated histone methylation are required for silencing of neuronal genes. Mol. Cell *14*, 727–738.

Rougier, N., Bourc'his, D., Gomes, D.M., Niveleau, A., Plachot, M., Pàldi, A., and Viegas-Péquignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. Genes Dev. *12*, 2108–2113.

Russell, P.J. (2010). iGenetics: a molecular approach (San Francisco: Benjamin Cummings).

Russo V.E.A., Martienssen R.A., and Riggs A.D. (1996). Epigenetic mechanisms of gene regulation (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 129–143.

Schaft, D., Roguev, A., Kotovic, K.M., Shevchenko, A., Sarov, M., Shevchenko, A., Neugebauer, K.M., and Stewart, A.F. (2003). The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. Nucleic Acids Res. *31*, 2475–2482.

Schiesser, S., Hackner, B., Pfaffeneder, T., Müller, M., Hagemeier, C., Truss, M., and Carell, T. (2012). Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angew. Chem. Int. Ed Engl. *51*, 6516–6520.

Schoenherr, C.J., and Anderson, D.J. (1995). The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. Science *267*, 1360–1363.

Schroeder, D.I., Lott, P., Korf, I., and LaSalle, J.M. (2011). Large-scale methylation domains mark a functional subset of neuronally expressed genes. Genome Res. *21*, 1583–1591.

Schübeler, D. (2015). ESCI award lecture: regulation, function and biomarker potential of DNA methylation. Eur. J. Clin. Invest. *45*, 288–293.

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

Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature *463*, 563–567.

Sérandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Péron, C.L., et al. (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., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908–912.

Sharif, J., Endo, T.A., Nakayama, M., Karimi, M.M., Shimada, M., Katsuyama, K., Goyal, P., Brind'Amour, J., Sun, M.-A., Sun, Z., et al. (2016). Activation of Endogenous Retroviruses in Dnmt1–/– ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. Cell Stem Cell.

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.

Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., Karun, V., Jaakkola, T., and Gifford, D.K. (2014). Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat. Biotechnol. *32*, 171–178.

Shi, Y., Sawada, J., Sui, G., Affar, E.B., Whetstine, J.R., Lan, F., Ogawa, H., Luke, M.P.-S., Nakatani, Y., and Shi, Y. (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. Nature *422*, 735–738.

Shi, Y.-J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. Mol. Cell *19*, 857–864.

Shimojo, M. (2006). Characterization of the nuclear targeting signal of REST/NRSF. Neurosci. Lett. *398*, 161–166.

Shimojo, M., Lee, J.H., and Hersh, L.B. (2001). Role of zinc finger domains of the transcription factor neuron-restrictive silencer factor/repressor element-1 silencing transcription factor in DNA binding and nuclear localization. J. Biol. Chem. *276*, 13121–13126.

Silverstein, R.A., and Ekwall, K. (2005). Sin3: a flexible regulator of global gene expression and genome stability. Curr. Genet. *47*, 1–17.

Singh, S.K., Kagalwala, M.N., Parker-Thornburg, J., Adams, H., and Majumder, S. (2008). REST maintains self-renewal and pluripotency of embryonic stem cells. Nature *453*, 223–227.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339–344.

Song, C.-X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., et al. (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., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell *153*, 678–691.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. Cell *161*, 555–568.

Sproul, D., Kitchen, R.R., Nestor, C.E., Dixon, J.M., Sims, A.H., Harrison, D.J., Ramsahoye, B.H., and Meehan, R.R. (2012). Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. Genome Biol. *13*, R84.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (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., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Stein, R., Razin, A., and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proc. Natl. Acad. Sci. U. S. A. *79*, 3418–3422.

Stewart, A.J., Hannenhalli, S., and Plotkin, J.B. (2012). Why transcription factor binding sites are ten nucleotides long. Genetics *192*, 973–985.

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

Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., Simon, I., Yakhini, Z., and Cedar, H. (2009). Developmental programming of CpG island methylation profiles in the human genome. Nat. Struct. Mol. Biol. *16*, 564–571.

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., and Jacobsen, S.E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. *12*, R54.

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

Tabuchi, A., Yamada, T., Sasagawa, S., Naruse, Y., Mori, N., and Tsuda, M. (2002). REST4-mediated modulation of REST/NRSF-silencing function during BDNF gene promoter activation. Biochem. Biophys. Res. Commun. *290*, 415–420.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (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.

Tapia-Ramírez, J., Eggen, B.J., Peral-Rubio, M.J., Toledo-Aral, J.J., and Mandel, G. (1997). A single zinc finger motif in the silencing factor REST represses the neural-specific type II sodium channel promoter. Proc. Natl. Acad. Sci. U. S. A. *94*, 1177–1182.

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

Teif, V.B., Beshnova, D.A., Vainshtein, Y., Marth, C., Mallm, J.-P., Höfer, T., and Rippe, K. (2014). Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. Genome Res. *24*, 1285–1295.

The Uniprot Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, 158-169

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

Tremethick, D.J. (2007). Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. Cell *128*, 651–654.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell *5*, 442–449.

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.

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

Tudor, M., Akbarian, S., Chen, R.Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proc. Natl. Acad. Sci. U. S. A. *99*, 15536–15541.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

Vardimon, L., Kressmann, A., Cedar, H., Maechler, M., and Doerfler, W. (1982). Expression of a cloned adenovirus gene is inhibited by in vitro methylation. Proc. Natl. Acad. Sci. U. S. A. *79*, 1073–1077.

Wang, Y., Xiao, M., Chen, X., Chen, L., Xu, Y., Lv, L., Wang, P., Yang, H., Ma, S., Lin, H., et al. (2015). WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. Mol. Cell *57*, 662–673.

Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. *2*, 1136–1143.

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

Westbrook, T.F., Hu, G., Ang, X.L., Mulligan, P., Pavlova, N.N., Liang, A., Leng, Y., Maehr, R., Shi, Y., Harper, J.W., et al. (2008). SCFβ-TRCP controls oncogenic transformation and neural differentiation through REST degradation. Nature *452*, 370–374.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221–225.

Widom, J., and Klug, A. (1985). Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. Cell *43*, 207–213.

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

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

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T., and Migeon, B.R. (1984). Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc. Natl. Acad. Sci. U. S. A. *81*, 2806–2810.

Wolff, E.M., Byun, H.-M., Han, H.F., Sharma, S., Nichols, P.W., Siegmund, K.D., Yang, A.S., Jones, P.A., and Liang, G. (2010). Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS Genet. *6*, e1000917.

Wood, I.C., Roopra, A., and Buckley, N.J. (1996). Neural specific expression of the m4 muscarinic acetylcholine receptor gene is mediated by a RE1/NRSE-type silencing element. J. Biol. Chem. *271*, 14221–14225.

Wu, H., and Zhang, Y. (2014). Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. Cell *156*, 45–68.

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

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011a). 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, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011b). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389–393.

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. Nature *532*, 329–333.

Wunderlich, Z., and Mirny, L.A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. Trends Genet. TIG *25*, 434–440.

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

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. Nature *492*, 443–447.

You, A., Tong, J.K., Grozinger, C.M., and Schreiber, S.L. (2001). CoREST is an integral component of the CoREST- human histone deacetylase complex. Proc. Natl. Acad. Sci. *98*, 1454–1458.

Yu, H.-B., Johnson, R., Kunarso, G., and Stanton, L.W. (2011). Coassembly of REST and its cofactors at sites of gene repression in embryonic stem cells. Genome Res. *21*, 1284–1293.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368–1380.

Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., et al. (2015). N6-Methyladenine DNA Modification in Drosophila. Cell *161*, 893–906.

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

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., et al. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. *38*, 4246–4253.

Zhao, L.-J., Subramanian, T., and Chinnadurai, G. (2006). Changes in C-terminal binding protein 2 (CtBP2) corepressor complex induced by E1A and modulation of E1A transcriptional activity by CtBP2. J. Biol. Chem. *281*, 36613–36623.

Zhao, X., Ueba, T., Christie, B.R., Barkho, B., McConnell, M.J., Nakashima, K., Lein, E.S., Eadie, B.D., Willhoite, A.R., Muotri, A.R., et al. (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. Proc. Natl. Acad. Sci. U. S. A. *100*, 6777–6782.

Zheng, D., Zhao, K., and Mehler, M.F. (2009). Profiling RE1/REST-mediated histone modifications in the human genome. Genome Biol. *10*, R9.

Zheng, N., Fraenkel, E., Pabo, C.O., and Pavletich, N.P. (1999). Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. Genes Dev. *13*, 666–674.

Zhou, J.C., Blackledge, N.P., Farcas, A.M., and Klose, R.J. (2012). Recognition of CpG island chromatin by KDM2A requires direct and specific interaction with linker DNA. Mol. Cell. Biol. *32*, 479–489.

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237.

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

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.-Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013b). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477–481.